



This is a postprint of:

Extending dynamic segmentation with lead generation: A latent class Markov analysis of financial product portfolios

L.J. Paas, T.H.A. Bijmolt & J.K. Vermunt

CentER Discussion Paper, 2004, no. 2004-1, pp. 46

For citation:

Please use the reference above

Link to the postprint file:

<http://arno.uvt.nl/show.cgi?fid=10506>

More information about this publication:

<http://repository.uvt.nl/id/ir-uvt-nl:oai:wo.uvt.nl:138176>



No. 2004-01

**EXTENDING DYNAMIC SEGMENTATION WITH LEAD
GENERATION: A LATENT CLASS MARKOV ANALYSIS OF
FINANCIAL PRODUCT PORTFOLIOS**

By L.J. Paas, T.H.A. Bijmolt, J.K. Vermunt

January 2004

ISSN 0924-7815

Extending Dynamic Segmentation with Lead Generation: A Latent Class Markov Analysis of Financial Product Portfolios

Leonard J. Paas

Tammo H.A. Bijmolt

Jeroen K. Vermunt

December 18th, 2003

Leonard J. Paas and Tammo H.A. Bijmolt are Assistant Professor and Professor of Marketing Research, respectively, at the Department of Marketing, Faculty of Economics and Business Administration, Tilburg University. Jeroen K. Vermunt is Professor of Methodology and Statistics, Faculty of Social Sciences, Tilburg University. We would like to express our gratitude towards GfK in the Netherlands for making the data set available, and to Harald van Heerde for the useful suggestions on a previous draft of the paper. Contact address: PO Box 90153, 5000 LE, Tilburg, the Netherlands; l.j.paas@uvt.nl; +31-13-4663022.

Extending Dynamic Segmentation with Lead Generation: A Latent Class Markov Analysis of Financial Product Portfolios

ABSTRACT

A recent development in marketing research concerns the incorporation of dynamics in consumer segmentation. This paper extends the latent class Markov model, a suitable technique for conducting dynamic segmentation, in order to facilitate lead generation. We demonstrate the application of the latent Markov model for these purposes using a database containing information on the ownership of twelve financial products and demographics for explaining (changes in) consumer product portfolios. Data were collected in four bi-yearly measurement waves in which a total of 7676 households participated. The proposed latent class Markov model defines dynamic segments on the basis of consumer product portfolios and shows the relationship between the dynamic segments and demographics. The paper demonstrates that the dynamic segmentation resulting from the latent class Markov model is applicable for lead generation.

INTRODUCTION

Much like other markets, the market structure of financial products is non-stable over time. Financial product ownership, for example, will vary between stages of the product life cycle and between households, depending on the family life cycle. A segmentation studies on such dynamics (i.e., dynamic segmentation) may yield highly actionable information. In particular, dynamic segmentation based on product ownership will imply directions for lead generation: identification of consumers who currently do not own a certain product, but have a high propensity to purchase the product in the next period.

Dynamic segmentation has received limited attention in the marketing literature (Wedel and Kamakura, 2000, chapter 10). Furthermore, extant literature on dynamic segmentation concerns fast-moving consumer goods and does not address lead generation as a logical extension of dynamic segmentation (Böckenholt and Langeheine, 1996; Böckenholt and Dillon, 2000; Kamakura, Kim and Lee, 1996; Poulsen, 1990; Ramaswamy, 1997; Seetharam, 2003; Wedel, Kamakura, DeSarbo and Hofstede, 1995). However, dynamic segmentation and lead generation are of key importance in assessing the major threats and opportunities facing companies operating in the financial services sector. First, it is unrealistic to assume stationary segments in this market, due to the dynamics in consumer needs and product portfolios (Browning and Lusardi, 1996; Wärneryd, 1999). Second, lead generation is an important issue in this market because of the great diversity in financial products that can be offered to consumers (Kamakura, Ramaswami and Srivastava, 1991; Knott, Hayes and Neslin, 2002; Winer, 2001). The magnitude of assets in this

market—billions of dollars—also suggests that this is an interesting field of study (Kamakura et al. (1991).

As in previous studies in the financial services market (e.g., Cohn, Lewellen, Lease, and Scharlbaum, 1975; Ramaswami, Srivastava and McInish, 1992; Ramaswamy, Chatterjee and Chen, 1996; Srivastava, McInish and Price, 1984), we will concentrate on consumer product portfolios. Given the wide range of available financial products and the dynamics in consumer product portfolios, it is surprising that most previous studies concern a stationary analysis. In addition the few studies delving into the dynamics of consumers' financial product portfolios are based on cross-sectional data (Dickinson and Kirzner, 1986; Kamakura, et al., 1991; Paas, 1998; Soutar and Cornish-Ward, 1997; Stafford, Kasulis and Lusch, 1982). Cross-sectional data actually give limited insight into the dynamics of product portfolios and may possibly confound consumer-specific effects and time effects.

The purpose of this paper is twofold. The first contribution is methodological: we extend and utilize the latent class Markov model for dynamic segmentation with formulas supporting lead generation. The second contribution of this paper is that it presents an empirical study on the dynamics of consumer financial product portfolios using longitudinal data. The latent class Markov model itself has previously been applied successfully in marketing (Böckenholt and Dillon, 2000; Böckenholt and Langeheine, 1996; Brangule-Vlagsma, Pieters and Wedel, 2002; Poulsen, 1990; Ramaswamy, 1997). We contribute to extant literature through new empirical findings on consumer financial product portfolios and changes therein, showing how these findings can be used for dynamic segmentation and lead generation.

As for the organization of the paper, the next section reviews the literature on segmentation and lead generation in the financial services market. Then we present the latent class Markov model for dynamic segmentation and extend this model with lead generation formulas. Next follows an analysis of a dataset with ownership information on 12 financial products by 7676 households collected in four bi-yearly waves. We report the dynamic segmentation and the lead generation resulting from the application of the latent class Markov model to the data. The paper concludes by discussing the utility of the model for dynamic segmentation, lead generation and the implications of the main findings.

LITERATURE REVIEW

Segmentation studies in the financial services market are frequently based on the financial products owned by households. Segmentation based on product ownership is domain-specific, because it is founded on observable behavior. Such a foundation is more likely to result in effective segmentation than are general consumer characteristics or variables that can only be measured indirectly, such as consumer attitudes or beliefs (Van Raaij and Verhallen, 1994; Wedel and Kamakura, 2000).

The two main approaches for studying consumer product ownership of financial products are based either on consumer product portfolios, or on the order in which consumers acquire financial products. The first approach investigates the combinations in which consumers own financial products (e.g. Cohn et al., 1975; Ramaswami et al., 1992; Ramaswamy et al., 1996; Srivastava et al., 1984). This approach gives insight into product complementarity. An important limitation is that

this approach is a stationary form of segmentation; it does not model dynamics of consumer product portfolios and, thus, provides limited to no insight herein.

The alternative approach for studying consumer ownership of financial products is called acquisition pattern analysis. This approach assumes that households have various financial objectives (Hauser and Urban, 1986; Kasulis, Lusch and Stafford, 1979; Paroush, 1965). These objectives cannot be fulfilled at once, however, because acquisitions of financial products usually imply major investments or long-term contractual obligations. Finite resources and the choices that must be made between attending to different financial objectives will lead to a priority structure of objectives. Empirical acquisition pattern analyses typically assume that cross-sectional data can be used to find the common order in which households acquire products. Investigations on ownership of financial products (Dickinson and Kirzner, 1986; Kamakura et al., 1991; Paas, 1998; Soutar and Cornish-Ward, 1997; Stafford et al., 1982) have shown that a common order of acquisition indeed applies for financial product markets. Consumers generally acquire products that are related to more basic objectives, such as products related to liquidity, cash reserve or risk management (insurance policies), before products for higher order objectives, such as investing and speculation.

Obviously, insight into common orders of acquisition is useful for segmentation and lead generation. If a common order of acquisition exists, consumers can be allocated to segments on the basis of the set of products they currently own. For lead generation, consumers owning fewer products are considered to be financially less mature, and are expected to acquire relatively basic products. As consumers own more products, they are allocated to segments in which individuals are considered to be more financially mature, and where acquisitions of more

sophisticated financial products are more likely. The latter obviously has implications for lead generation (Kamakura et al., 1991).

Despite the potential benefits, acquisition pattern analysis based on cross-sectional data has three important limitations. First, consumers may not always follow the most common order of acquisition, while cross-sectional investigations give no insight into divergent acquisition orders. This suggests that segmenting consumers on the basis of a single order may lead to imprecise dynamic segmentation and inaccurate lead generation predictions. Second, it is interesting to know not only which product should be offered next to specific consumers, but also in which period offers should be made. Cross-sectional data do not provide such time-related insights. Third, for cross-sectional data, observed differences between product portfolios may reflect consumer-specific effects and time-specific effects, but investigation based on such data only assumes one common order of acquisition. Time specific effects and consumer specific effects are, thus, confounded.

Here we suggest a third approach to study ownership of financial products: Analysis of longitudinal data on financial product portfolios. This new approach combines the merits of product portfolio analysis and acquisition pattern analysis. Research into the dynamics of product portfolios will reveal both complementarity and substitutability relations between products. Furthermore, our approach does not require the assumption of a common acquisition order. Finally, since product ownership varies between consumers and for individual consumers over time, analysis of longitudinal data allows separating effects of time factors and consumer characteristics.

MODEL FOR DYNAMIC SEGMENTATION

Specification of the latent class Markov model

The (dynamics of) consumer product portfolios can be analyzed by means of latent class Markov models with concomitant variables (Böckenholt and Dillon, 2000; Poulsen, 1990; Ramaswamy, 1997; Van de Pol and Langeheine, 1990; Vermunt, Langeheine, and Böckenholt, 1999). We first introduce some notation in Table 1.

INSERT TABLE 1 HERE

The latent class Markov model uses information on ownership of J products by I consumers at T measurement occasions, and the values of these I consumers on K covariates at the T measurement occasions. Such models are based on the following three components:

- (1) A latent class structure for defining a segmentation based on the observed product portfolios at each measurement occasion, represented by

$$\prod_{s=1}^S \prod_{j=1}^J P(Y_{ijt} = 1 | X_{it} = s_t).$$

This component defines the probability that consumer i owns a specific combination of the J products at t , given membership probabilities for each of the S segments at t for consumer i .

- (2) A regression structure for studying covariate effects on the initial state,

$$P(X_{i,t=1} = s_{t=1} | \mathbf{Z}_{i,t=1} = \mathbf{z}_{t=1}).$$

This component defines probabilities for consumer i being allocated to each segment s at the first measurement occasion, given this person's covariate values at the first occasion of measurement.

(3) A regression structure for modeling the first-order Markov transitions between portfolios across measurement occasions and covariate effects on

these transitions, $\prod_{t=2}^T P(X_{it} = s_t | X_{i,t-1} = s_{t-1}, \mathbf{Z}_{it} = \mathbf{z}_t)$.

On the basis of these components, the latent Markov model can be defined as in equation [1].

$$[1] P(\mathbf{Y}_i = \mathbf{y} | \mathbf{Z}_i = \mathbf{z}) = \sum_{s_{t=1}=1}^S \sum_{s_{t=2}=1}^S \dots \sum_{s_{t=T}=1}^S [P(X_{i,t=1} = s_{t=1} | \mathbf{Z}_{i,t=1} = \mathbf{z}_{t=1}) \\ \prod_{t=2}^T P(X_{it} = s_t | X_{i,t-1} = s_{t-1}, \mathbf{Z}_{it} = \mathbf{z}_t) \\ \prod_{s_{t=1}=1}^S \prod_{j=1}^J P(Y_{ijt} = 1 | X_{it} = s_t)]$$

The three basic assumptions of this model are as follows: (1) ownership indicators, Y_{ijt} , are mutually independent, given the time-specific latent states (called local independence). This measurement model is comparable to the conventional latent class analysis measurement model; (2) the latent transition structure, $\prod_{t=2}^T P(X_{it} = s_t | X_{i,t-1} = s_{t-1}, \mathbf{Z}_{it} = \mathbf{z}_t)$, has the form of a first-order Markov chain, meaning that besides the values on the covariates at t , \mathbf{z}_t , X_{it} depends only on $X_{i,t-1}$ and not on segment membership at earlier measurement occasions; (3) covariates may affect the latent states, X_{it} , but not the observed states, Y_{ijt} .

Consumers belong to only one segment. However, this segment cannot be established with certainty. Thus, consumers are probabilistically assigned to each of the segments. Equation [1] specifies the probabilities for the occurrence of consumer i 's manifest data pattern, $P(\mathbf{Y}_i = \mathbf{y} | \mathbf{Z}_i = \mathbf{z})$, given the three above-mentioned main elements of the latent class Markov model.

Model estimation

Model parameters are estimated using maximum likelihood estimation. The measurement component of the Markov model, component 1, is modeled by logit equations. These allow us to estimate the probability for each consumer i , to own product j at measurement occasion t , conditional on the segment to this consumer i is allocated to at t . Equation [2] describes this type of relationship.

$$[2] P(Y_{ijt} = 1 | X_{it} = s) = \frac{\exp(\beta_{js})}{1 + \exp(\beta_{js})}$$

where β_{js} represents the logistic regression coefficients linking the probability for owning product j to membership of segment s . For estimating relevant parameters in components 2 and 3, defined above, we use logit equations with concomitant variables (Dayton and MacReady, 1988; Gupta and Chintagunta, 1994).

$$[3] P(\mathbf{X}_{i,t=1} = \mathbf{s}_{t=1} | \mathbf{Z}_{i,t=1} = \mathbf{z}_t = 1) = \frac{\exp(\gamma_{0s,t=1} + \sum_{k=1}^K \gamma_{0s,t=1} z_{i,t=1,k})}{\sum_{s=1}^S \exp(\gamma_{0s,t=1} + \sum_{k=1}^K \gamma_{0s,t=1} z_{i,t=1,k})}$$

Maximum likelihood (ML) estimates of the logit equations are typically obtained by means of the Expectation Maximization (EM) algorithm (Vermunt, Langeheine and Böckenholt, 1999). For application of the latent class Markov model to our dataset, we adapt a variant of the Expectation Maximization algorithm (EM-algorithm) called the forward-backward algorithm (Baum, Petrie, Soules and Weiss, 1970). Instead of computing the entries in the joint posterior latent distribution, as the standard EM algorithm, the forward-backward algorithm obtains the entries in bivariate marginal posterior distribution corresponding with adjacent points in time. Therefore, the forward-backward algorithm is used when there are many time-points and segments

to be estimated. The more conventional application of the EM-algorithm, based on Dempster, Laird and Rubin (1977), is less feasible in such situations (McDonald and Zucchini, 1997). More important for our purposes is that the forward-backward algorithm, as it is reformulated in the Appendix, allows us to define equations for predicting ownership of products at $t+1$ on the basis of all information available up to measurement occasion t . This is discussed below.

FORMULAS FOR LEAD GENERATION

For lead generation, we assess the probability of ownership of each product j at time point $t+1$ given all information available at t . This available information concerns product portfolios, values on covariates and the model parameters at t and before t . The probability of owning each product j at measurement occasion $t+1$ is denoted as $P(Y_{ij,t+1} = 1 | \mathbf{Y}_{it-} = \mathbf{y}_{t-}, \mathbf{Z}_{it-} = \mathbf{z}_{t-})$, where the symbol $t-$ is used to refer to time point t and all preceding time points. A lead is found when a household does not own a product j at t , but has a high predicted probability of owning this product at $t+1$.

The predicted ownership probabilities $P(Y_{ij,t+1} = 1 | \mathbf{Y}_{it-} = \mathbf{y}_{t-}, \mathbf{Z}_{it-} = \mathbf{z}_{t-})$, are obtained in three steps. The first step is based on the forward probabilities, $\alpha_{it}(X_{it} = s_t)$. The forward probabilities are a result of the adjusted forward-backward algorithm presented in the Appendix. They define the posterior probabilities that individual consumers belong to segment s at measurement occasion t , and the probabilities for owning all products at t and all preceding measurement occasions. In the first step we compute the posterior segment membership probabilities for

measurement occasion t for subject i , given all observed information up to t , $P(X_{it} = s_t | \mathbf{Y}_{it-} = \mathbf{y}_{t-}, \mathbf{Z}_{it-} = \mathbf{z}_{t-})$ by normalizing the forward probabilities:

$$[4] P(X_{it} = s_t | \mathbf{Y}_{it-} = \mathbf{y}_{t-}, \mathbf{Z}_{it-} = \mathbf{z}_{t-}) = \frac{\alpha_{it}(X_{it} = s_t)}{\sum_{s_t=1}^S \alpha_{it}(X_{it} = s_t)}.$$

The second step consists of calculating prior segment membership probabilities at time $t+1$, given the observed information up to t , $P(X_{it+1} = s_{t+1} | \mathbf{Y}_{it-} = \mathbf{y}_{t-}, \mathbf{Z}_{it-} = \mathbf{z}_{t-})$.

This involves combining the posteriors from the first step with the transition probabilities $P(X_{it+1} = s_{t+1} | X_{it} = s_t, \mathbf{Z}_{it+1} = \mathbf{z}_{t+1})$ as follows:

$$[5] P(X_{it+1} = s_{t+1} | \mathbf{Y}_{it-} = \mathbf{y}_{t-}, \mathbf{Z}_{it-} = \mathbf{z}_{t-}) = \sum_{s_t=1}^S [P(X_{it} = s_t | \mathbf{Y}_{it-} = \mathbf{y}_{t-}, \mathbf{Z}_{it-} = \mathbf{z}_{t-}) P(X_{it+1} = s_{t+1} | X_{it} = s_t, \mathbf{Z}_{it+1} = \mathbf{z}_{t+1})]$$

In the third step we obtain the predicted ownership probabilities, $P(Y_{ij,t+1} = 1 | \mathbf{Y}_{it-} = \mathbf{y}_{t-}, \mathbf{Z}_{it-} = \mathbf{z}_{t-})$, from the prior segment membership probabilities and the segment-specific ownership probabilities:

$$[6] P(Y_{ij,t+1} = 1 | \mathbf{Y}_{it-} = \mathbf{y}_{t-}, \mathbf{Z}_{it-} = \mathbf{z}_{t-}) = \sum_{s_{t+1}=1}^S [P(X_{it+1} = s_{t+1} | \mathbf{Y}_{it-} = \mathbf{y}_{t-}, \mathbf{Z}_{it-} = \mathbf{z}_{t-}) P(Y_{ij,t+1} = 1 | X_{it+1} = s_{t+1})].$$

The third step results in an $I \times J$ table for each measurement occasion, with I rows and J columns. In the lead generation table, element $\{i,j\}$ reports the predicted probability that consumer i owns products j at $t+1$ on the basis of all information available at measurement occasion t . For lead generation, the aim is to find (for each of the J products) households with the highest probability of owning a product j at $t+1$, amongst the households that do not own j at t .

EMPIRICAL STUDY: DATA AND METHOD

The Dutch division of the international market research company, GfK, conducts a large bi-yearly empirical study on consumer financial product ownership in the Netherlands, known as “Total Investigation Financial Services”. The retrieved information concerns household ownership of 16 financial products in 1996, 1998, 2000 and 2002. Interviews are conducted face-to-face, and respondents show their financial papers to verify answers. A total of 7676 households, that form a representative sample of the Dutch population, participated in the research.

Not all households participated in each panel wave, as a result of attrition or signing up with the panel after 1996. Fortunately, it is straightforward to estimate the latent class Markov model with partially observed data (Vermunt, 1997), under the assumption that the missing data are missing at random. This is clearly preferred to omitting cases with missing values from the analysis, which may seriously bias results (Verbeek and Nijman, 1992; Winer, 1983).

Table 2 presents the penetration rates of the twelve products included in the analysis (four products with penetrations above 99% at each measurement occasion were excluded from the analysis). The twelve products are divided over three types. The lower part of Table 2 presents the most basic products (called foundation products); then, there are income management products for securing long-term income; at the top are the risky investments. Beside information on product ownership, there is also information on the following: (1) net household income, (2) age of the household head, and (3) household assets. These demographics will be the covariates for explaining the structure of product portfolios and changes therein. Previous research shows that these variables are strongly related to consumer

financial product portfolios (Gunnarsson and Wahlund, 1997; Ramaswamy et al., 1996; Wärneryd, 1999).

INSERT TABLE 2 ABOUT HERE

Parameters are estimated with an experimental version of the Latent Gold program (Vermunt and Magidson, 2000), which implements the latent class Markov model described above. We use an unrestricted measurement model. Following convention (Brangule-Vlagsma et al., 2002; Vermunt, 2001), we started with a one-segment model and added more segments until the value on the BIC statistic began to increase. To overcome the potential problem of attaining sub-optimal solutions, we conducted all analyses with several different sets of random starting values. Below, we first present the dynamic segmentation results and then discuss application of the results for lead generation on the basis of equations [3] to [5].

DYNAMIC SEGMENTATION RESULTS

Measurement model

Alternative latent class Markov models can be formulated by modeling various types of change (Böckenholt and Langeheine, 1996; Brangule-Vlagsma et al., 2002; Wedel and Kamakura, 2000). Particularly important is the distinction between manifest change and latent change. Manifest change refers to dynamics in the measurement model (i.e. ownership probabilities for products are not constant per segment over time). Latent change refers to dynamics in segment sizes and switching between segments by individual consumers. We chose to develop a model that assumes a time-constant measurement model and allows for switching between segments. This type

of model allows for latent change but not for manifest change. Such models are most suitable for segmentation purposes, as the structure of segments remains the same over time. Changing segment structures would lead to a continuous reformulation of segment specific marketing strategies.

We established the appropriate number of segments (S) for the latent class Markov model by increasing the number of segments. This showed that a model with a nine-segment measurement model is most suitable, with 171 parameters and $BIC=136478$. Models with fewer (or more) segments result in higher BIC-values. This nine-segment model, which will be called the final model, assumes time-constant segments (a time constant measurement model) and time-constant transition probabilities. Thus, in our data the same measurement model applies for 1996, 1998, 2000 and 2002. Also, in the final model the probability to switch from segment s , at t , to segment s' , at $t+1$, is the same as the probability to switch from s to s' between $t+1$ and $t+2$. This applies for all s and all t .

We examined the relative fit of alternative model specifications. The first benchmark model has a time-constant measurement model, like the final model has. Like the final model this first benchmark allows switching between segments (latent change). However, in the first benchmark model, switching probabilities are time varying. This first nine-segment benchmark model has 315 parameters and $BIC=137567$. The BIC resulting from the first benchmark is higher than the BIC for the final model, implying it is realistic to assume time-constant transition probabilities.

We specified a second benchmark model with a time-varying measurement model (manifest change) that embraces time-constant transition probabilities. With

this model we tested whether it is feasible to assume latent change instead of manifest change. The second nine-segment benchmark model has 468 parameters and a BIC of 142063, which is higher than the BIC of the final model. As a third benchmark, we considered fit of a stationary model. This model has a time constant measurement model and assumes respondents stay in the same segment over time. Thus, there is neither manifest nor latent change. The resulting nine-segment model has 99 parameters and a BIC of 136799. This is also higher than the BIC of the final model, suggesting it is correct to assume change occurs. Summarized, the model specification as presented in this paper empirically outperforms, in terms of lower BIC, each of these alternative model specifications.

Table 3 presents the final nine-segment measurement model. Because there is no manifest change, this measurement model is relevant for each of the four measurement occasions. To enhance interpretation, we ranked segments one to seven in ascending order of product penetration rates across the twelve products. Furthermore, the two segments for which penetration rates for some of the more sophisticated products are high, but with low penetration rates for products related to ownership of a house (mortgages and house insurance), are placed after the other segments (segments eight and nine). The segments can be characterized as: (1) inactives who do not own a car insurance; (2) inactives who do own a car insurance; (3) homeowners without a mortgage; (4) homeowners with a mortgage; (5) homeowner income managers without credit cards; (6) homeowner income managers with credit cards; (7) actives; (8) credit card-oriented light income managers; (9) loan-oriented light income managers.

| |
|---------------------------|
| INSERT TABLE 3 ABOUT HERE |
|---------------------------|

Dynamics of the segmentation

Table 4 presents the sizes of the segments at each measurement occasion. Segment sizes are stable, except for segments five, six, seven and nine. The average probability for consumers to be allocated to segments five and nine decreased in the 1996-2002 period. Segments six and seven became larger.

INSERT TABLES 4 AND 5 HERE

Table 5, the latent transition matrix, presents the probabilities for switching between the segments in the measurement model. A large percentage of the households remained in the same segment over the two-year periods between consecutive panel waves, as indicated by the high proportions in the cells on the diagonal of the transition matrix. Nevertheless, the switching that does occur can explain changes in the size of segments. The increase of the size of segment six is probably due to switches from segment five to segment six. This switch has a 0.12 probability of occurrence in each two-year period between measurement occasions. It probably results from the growing popularity of the credit card in the period 1996-2002, as Table 3 shows that ownership of the credit card is the main difference between these two segments. The decreasing membership in segment nine is due to switching towards segments five and six (Table 5). This may reflect households who buy a house, as ownership probabilities for mortgages and home insurance policies increases substantially when switching from segment nine to five or six. Another development is the increasing size of segment seven, indicating that the percentage of highly active consumers in the financial services market is increasing, as segment seven members have relatively high probabilities for owning most products.

Also evident is switching behavior between segments, unrelated to overall segment size. Switching occurs into and out of segments one, two, three, four and eight, while these segments do not change in size. Thus, switching may result from the needs of individual households, not just from general changes in the use of specific products in the entire population.

Covariate effects

Next, we assessed covariate effects on segment membership in the initial state (1996) and on the transition probabilities. Households in our sample were allocated to one of the following categories of net monthly income in Euros: <1000, 1000-1500, 1500-2000 or ≥ 2000 . Age refers to the head of household, and consists of the following categories: <35, 35-49, 50-64 or ≥ 65 . Household assets, the total amount of household savings, consists of the following categories: <5000, 5000-20000 or >20000 Euros. All three covariates have a significant effect on segment membership in 1996 (income: Wald=619.67, d.f.=24, $p < 0.001$; Age: Wald=594.79, d.f.=24, $p < 0.001$; Household assets: Wald=230.90, d.f.=16, $p < 0.001$).

The effects are presented in Table 6. Effect coding was used to identify coefficients. This implies that parameters sum to zero over segments and covariate levels (Alba, 1987). For each covariate-segment combination, there is a coefficient indicating whether (controlling for the other covariates) membership in that segment is more (or less) likely than average. For example, Table 6 shows that households with an income <2500 are more likely to be found in segment 1 than in other segments, while the opposite applies for households with income levels 3500-4999 and ≥ 5000 .

INSERT TABLES 6 AND 7 HERE

The general tendency reported in Table 6 is consistent with the results of previous research (Browning and Lusardi, 1996; Gunnarson and Wahlund, 1997; Soutar and Cornish-Ward, 1997; Wärneryd, 1999). The higher the income of a given household, the more likely it is that the household will be allocated to a segment with higher product-penetrations. This also applies for household assets. Moreover, the age-effect is consistent with the lifecycle hypothesis. Households with heads aged 65 or older are over-represented in the segments with low probabilities of product ownership (segments one and two). Households with younger heads (age<65) are more likely to belong to the more active segments five and six. However, these age groups do not have a significantly greater probability of belonging to the most active segment seven.

All three covariates also significantly affect switching probabilities (income: Wald=127.47, d.f.=24, p<0.001; age: Wald=247.99, d.f.=24, p<0.001; household assets: Wald=179.83, d.f.=16, p<0.001). Where values on covariates imply a greater probability to belong to an initial state, the model generally showed that these covariate values also imply a greater probability for switching into this state. The components of the model that describe changes between panel waves are presented in Table 7 and are again effect coded.

Summarizing, we presented a highly interpretable dynamic measurement model that describes segments based on product ownership. We showed that covariates in the latent class Markov model provide insight into the characteristics of members of segments and into switching between segments.

EVALUATION OF THE LEAD GENERATION RESULTS

Methodology for evaluation

We determine the predictive validity of the lead generation equations, formulas [4] to [6], by assessing how well the lead generation equations based on measurement occasion t predict acquisitions of products by households between measurement occasions t and $t+1$. Only those households that participated in the survey at both measurement occasions, t and $t+1$, are included in this evaluation. The predictive validity of the lead generation equations is assessed by evaluating to what extent these equations can distinguish households that own product j at $t+1$ from households that do not own product j at $t+1$. This distinction is made amongst households that do not own product j at measurement occasion t . Predictive validity is considered higher when the lead generation equations, based on the information available at t , better predict which households not owning j at t acquire this product between t and $t+1$.

Predictive validity of the lead generation equations is assessed using *Gini*, a measure of concentration. Kamakura, Wedel, De Rosa and Mazzon (2003) previously applied *Gini* for evaluating models that predict ownership of products. Various notations of *Gini* are available. We use the notation by Sen (1997), as this notation is generally accepted. Sen (1997) uses *Gini* for its original purpose, evaluating income inequality. Income inequality can be viewed as a Lorenz curve. The cumulative percentages of the population, arranged from richest to poorest, are presented on the horizontal axis. The cumulative percentages of the total income of the population, on the vertical axis, are presented on the vertical axis (see Figure 1).

INSERT FIGURE 1 HERE

In Figure 1, 0% of the population enjoys 0% of the total income received by the entire population, and 100% receive 100% of this total income. In a situation of perfect equality, the Lorenz curve is the diagonal; the cumulative percentages of income always equal the cumulative percentage of the population, and *Gini* equals 0. Under conditions of perfect inequality, one person receives all income, and *Gini* equals 1. Here, the Lorenz curve goes straight up at the origin and along the Y-axis, and then to the right. In reality, Lorenz curves are between these two extremes. Here the value on *Gini* is determined by dividing (1) the surface between the straight diagonal line and the Lorenz curve through (2) the complete area above the diagonal line (which is the area between the straight diagonal line and the Lorenz curve obtained when one person receives the entire income of the population). Values on this quotient are determined using equation [7] (Sen, 1997).

$$[7] \text{ Gini} = 1 + 1/n - \frac{2}{n^2 \mu} \sum_{i=1}^n r_i y_i$$

where n is the number of persons in the population, and μ is the mean income; r_i is the rank of person i with regard to income, and y_i is the income of person i .

Gini is applicable for evaluating the accuracy of lead generation equations, by assuming that in equation [7] n represents all persons not owning product j at t ; μ is the percentage of these n persons that own product j at $t+1$; r_i is the rank of person i with regard to the predicted probability of owning product j at $t+1$. Note that this predicted probability is calculated through equations [4] to [6] and is based on all information available at t . The person with the highest predicted probability receives rank 1, the person with second highest predicted probability rank 2, etc. Finally, y_i equals 1 when person i owns product j at $t+1$; otherwise, it equals 0.

Results

Values on *Gini* were determined for all products at each measurement occasion (except for 2002, for which no $t+1$ is available in our data; 2002 is the last occasion of measurement). Below we discuss two results in more detail. The first concerns predictions for ownership of the mortgage in 2002 ($t+1$) by respondents that do not own this product in 2000 (t). The second concerns formulas for predicting which respondents, of those that do not own bonds in 1998 (t), own bonds in 2000 ($t+1$).

In Panel A of Figure 2, the X-axis represents the cumulative percentage of households not owning mortgages in 2000. These households are ordered on the basis of the predicted probability that they will own a mortgage in 2002. Closer to the origin are households with large predicted probabilities, and further from this point are those households with smaller predicted probabilities. The Y-axis displays the cumulative percentage of households actually owning mortgages in 2002. Consider the 10% of respondents without a mortgage in 2000, and with the highest predicted probability to own this product in 2002. We find 37% of all the respondents that *do not* own a mortgage in 2000 but *do* own this product in 2002 among this 10% group, as displayed in Panel A of Figure 2. This is considerably better than random predictions represented by the diagonal line.

INSERT FIGURE 2 HERE

The power-curve of Panel A in Figure 2 leads to a *Gini*-value of 0.49, implying that almost half of the total surface above the diagonal falls between the power-curve and the diagonal. To understand the implications of this value on *Gini*, consider that, on average, 5.9% of the non-owners of mortgages in 2000 actually own this product in 2002. We find that 21.7% of the top 10% group actually own a

mortgage in 2002. Yet, of the 50% non-owners of a mortgage in 2000, with the lowest probabilities to own this product in 2002, only 2.1% do. Thus, the probability that members of the top 10% group acquire a mortgage is about ten times greater than the corresponding probability for the 50% with the lowest probability to make this acquisition.

Panel B of Figure 2 shows an example of another power-curve, which concerns respondents who do not own bonds in 1998. This power-curve leads to a *Gini*-value of 0.29. Here 4.4% of the households in the top 10% group own bonds in 2000, while only 0.9% of the bottom 50% group do. The average probability that non-owners of bonds in 1998 do own this product in 2000 equals 1.2%.

Values on *Gini* for all lead generation equations are given in Table 8. The following seven products have power-curves similar to those in Figure 2: bonds, shares, investment trusts, unemployment insurance, life insurance, house insurance and mortgages. Values on *Gini* are somewhat lower for four products: pension funds, loans, credit cards and savings accounts. Forecasting accuracy is much lower for acquisitions of car insurances.

INSERT TABLE 8 ABOUT HERE

An important point concerns the consistency of the values on *Gini*. According to the contents of Table 8, the forecasting accuracy for the acquisition of products in the period between the 1996 and the 1998 panel waves closely resembles the forecasting accuracy of products in the two other periods. For example, *Gini-values* of the lead generation equation of 1996 for mortgages in the 1996-1998 period equals 0.41. The *Gini* for the 1998 equation, for acquiring the mortgage in the 1998-2000 period, equals 0.40; for the 2000 formulas, predicting acquisitions of the mortgage in

the 2000-2002 period, the value on *Gini* is 0.49. Small differences between values on *Gini* in the three relevant periods imply high stability of the lead generation equations across measurement occasions.

DISCUSSION

This paper has shown that insight into consumer product portfolios and changes therein provide marketers with useful information for dynamic segmentation and lead generation. We proposed that research into (changes in) financial product portfolios should be based on the latent class Markov model. In particular, we find that a dynamic segmentation that assumes the segmentation structure to be constant over time, but allows for changes in segment size and switching between segments, to be most suitable for the data analyzed in this paper.

Study of longitudinal data on financial product portfolios has yielded a number of insights that could not have been obtained through a stationary segmentation approach or through the acquisition pattern analysis approach. Stationary segmentation would not have revealed changes in segment size and the probabilities that consumers switch from one segment to another. Concerning acquisition pattern analysis, the dynamic segmentation approach that we have proposed offers insight into divergent orders of acquisition. This insight would not be gained through acquisition pattern analysis, as the latter assumes a single common order of acquisition. For example, through acquisition pattern analysis we would not have found that there is a segment nine, with consumers who have a high probability to own life insurance policies and pension funds, but not a mortgage or home insurance policy, whereas the opposite holds for segment four. Also, we found that

consumers in segment nine are likely to switch to segment four, whereas the opposite switching direction is very improbable. Such information on divergence from the common order of acquisition and the related information into switching behavior would probably be confounded and be represented in a simplified manner when using acquisition pattern analysis instead of dynamic segmentation. A last obvious advantage of the dynamic segmentation approach is that our results give the probabilities that certain events, such as switching between segments, will occur between measurement occasions, which is a two-year period.

Another important point is that the obtained dynamic segmentation has high face-validity and our findings are consistent, albeit more detailed due to the incorporation of dynamics, to results of previous studies (Browning and Lusardi, 1996; Gunnarson and Wahlund, 1997; Soutar and Cornish-Ward, 1997; Wärneryd, 1999). The product-portfolios-based segments are highly interpretable, the changes therein are also plausible, and the relationships with covariates are consistent with the well-known lifecycle model for consumer financial behavior and other findings of previous research conducted in the financial services market. This interpretability enhances application for marketing purposes; it is relatively easy to formulate a marketing mix for interpretable segments.

As for the second contribution of the paper, we deduced lead generation formulas from the developed latent class Markov model (equations [4] to [6]) that can be used to allocate probabilities for acquiring specific products to households. Results of the evaluation of the lead generation formulas in terms of forecasting accuracy suggest that the formulated dynamic segmentation has considerable predictive validity. Predictive power was particularly precise for the acquisition of products used for asset accumulation purposes. Important in this regard is that the lead generation

results are consistent over the six-year period (1996 to 2002), during which our panel was interviewed four times. This implies that the lead generation, proposed in this paper, is highly consistent over time in our illustration. Lead generation formulas are thus likely to be applicable over a longer period.

Given the reported consistency in forecasting accuracy and precision, managers can distinguish individual consumers on the basis of the probability that they will acquire various products. This is useful marketing information. Consumers that are likely to acquire a product should be made an offer to do so, before competing firms successfully offer the product of interest. Those consumers with relatively low propensities to acquire a product should, then, not be offered this product, as marketing effort and funds are more likely to be employed profitably when consumers who are likely to acquire a product are made offers.

There are some indications for limitations of the proposed dynamic segmentation approach. Lead generation results are somewhat less effective for credit products (loans and credit cards). Further investigation should explore whether acquisitions of credit products can be modeled more effectively if other covariates are incorporated in the latent class Markov model. We are not certain about this; it may be possible that modeling the ownership and acquisitions of credit products is more difficult than modeling products related to asset accumulation. This issue is open for further investigation.

Lead generation is also less effective for savings accounts. This product has a very high penetration at all measurement occasions; at least 93% of the respondents own this product at each occasion (see Table 2). It is possible that the lead generation equations are less effective for predicting acquisitions of such commonly owned

products. For car insurance and pension funds, the lead generation formulas are the least effective, according to values on *Gini* (see Table 8). It is possible that car insurance is redundant for many households in the Netherlands, due to lease cars supplied by employers. Such conditions of labor are apparently unrelated to household behavior in the financial services market, and ownership of this product may therefore be inadequately modeled in combination with ownership of other financial products. Similar labor conditions exist for pension funds.

There are two other directions for further research that are not a direct consequence of the empirical results reported in this paper. The first concerns the type of data to which we applied the latent class Markov model and the lead generation equations. Our empirical research is based on survey data. Currently, marketing analysts at banks often scrutinize so-called database marketing data, deduced from interactions between clients and the company. Such data contain information on the ownership of products only at the own bank, not at the competitor. This partial information on product portfolios can obviously result in inaccurate lead generation. Kamakura et al. (2003) developed a model based on factor analysis with non-normal data (Kamakura and Wedel, 2000) that is suitable for augmenting database-marketing data. Future research could aim at integrating the modeling approach discussed in the current paper and the Kamakura et al. model. Second, in this paper we applied the latent class Markov model to one set of products in a single country. Applications in other countries and to other types of products, such as durable products, could explore the general applicability of the lead generation equations.

APPENDIX

The forward-backward algorithm was originally proposed for latent class Markov models without covariates, with a single indicator per occasion. The only difference between the forward-backward algorithm and the standard EM-algorithm concerns the implementation of the E-step. Baum et al. (1970) developed the forward-backward algorithm for single indicator latent class Markov models. We extend the model for application to multiple indicator latent class Markov models with covariates. More important for our purposes: we show that equations resulting from the forward-backward algorithm are applicable for lead generation. These lead generation formulas are discussed in the main text.

The standard variant of the E-step involves computing the expected value of the complete data log-likelihood (Dempster, Laird and Rubin, 1977). The contribution for case i to this “completed” data log-likelihood is as follows:

$$\begin{aligned}
 \text{[A.1] } \log L_i = & [P_i(X_{it=1} = s_{t=1}) \log P(X_{it=1} = s_{t=1} | \mathbf{Z}_{it} = \mathbf{z}_{t=1}) + \\
 & \sum_{t=2}^T P_i(X_{i,t-1} = s_{t-1}, X_{it} = s_t) \log P(X_{i,t-1} = s_{t-1} | X_{it} = s_t, \mathbf{Z}_{it} = \mathbf{z}_t) + \\
 & \sum_{t=1}^T \sum_{k=1}^K P_i(X_{it} = s_t) \log P(Y_{ijt} = 1 | X_t = s_t)]
 \end{aligned}$$

The E-step, both for the standard calculation manner (Dempster et al., 1977) and for the forward-backward algorithm, amounts to obtaining the univariate and bivariate posterior membership probabilities, $P_i(X_{it} = s_t)$ and $P_i(X_{i,t-1} = s_{t-1}, X_{it} = s_t)$, for person i . The forward-backward algorithm computes these quantities in an efficient manner, without processing the joint posterior latent distribution for all time points. In the M-step of EM, the standard complete data methods for logistic regression can be used to update model parameters. Convergence of the EM-algorithm is the same when

using the forward-backward algorithm, as when using the standard procedures for the E-step (McDonald and Zucchini, 1997).

Let $\alpha_{it}(X_{it} = s_t)$ and $\beta_{it}(X_{it} = s_t)$ be defined as follows:

$$[A.2] \alpha_{it}(X_{it} = s_t) = P(X_{it} = s_t, \mathbf{Y}_{it-} = \mathbf{y}_{t-} \mid \mathbf{Z}_{it-} = \mathbf{z}_{t-}),$$

$$[A.3] \beta_{it}(X_{it} = s_t) = P(\mathbf{Y}_{it+} = \mathbf{y}_{t+} \mid X_{it} = s_t, \mathbf{Z}_{it+} = \mathbf{z}_{t+}),$$

where $t-$ refers to time point t and all earlier time points, and $t+$ to all time points after t . The relationship between these two quantities and the relevant posteriors is described as follows:

$$[A.4] P(X_{it} = s_t) = \frac{\alpha_{it}(X_{it} = s_t)\beta_{it}(X_{it} = s_t)}{P(\mathbf{Y}_i = \mathbf{y})}, \text{ and}$$

$$[A.5] P_i(X_{i,t-1} = s_{t-1}, X_{it} = s_t) = [\alpha_{i,t-1}(X_{i,t-1} = s_{t-1}, \mathbf{Z}_i = \mathbf{z})P(X_{it} = s_t \mid X_{i,t-1} = s_{t-1}, \mathbf{Z}_i = \mathbf{z})P(\mathbf{Y}_{it} = \mathbf{y} \mid X_{it} = s_t)\beta_{it}(X_{it} = s_t)] / P(\mathbf{Y}_i),$$

where $P_i(\mathbf{Y}_{it} = \mathbf{y} \mid X_{it} = s_t) = \prod_{j=1}^J P(Y_{ijt} = 1 \mid X_{it} = s_t)$, and $P_i(\mathbf{Y}_i = \mathbf{y})$ is the likelihood

of the response vector of person i . Note that $P(\mathbf{Y}_i = \mathbf{y}) = \left[\sum_{X_T} \alpha_{iT}(X_{iT} = s_T) \right]$.

Thus, once we have $\alpha_{it}(X_{it} = s_t)$ and $\beta_{it}(X_{it} = s_t)$, it is straightforward to obtain the relevant posteriors $P(X_{it} = s_t)$ and $P(X_{i,t-1} = s_{t-1}, X_{it} = s_t)$ (Baum et al., 1970). Adapted to the more general latent class Markov model used in this paper, we get

$$[A.6] \alpha_{i1}(X_{i1} = s_1) = P(X_{i1} = s_1 \mid \mathbf{Z}_{i1} = \mathbf{z}_1)P(\mathbf{Y}_{i1} = \mathbf{y}_1 \mid X_{i1} = s_1), \text{ and}$$

$$[A.7] \alpha_{it}(X_{it} = s_t) = \sum_{s_{t-1}=1}^S [\alpha_{it-1}(X_{it-1} = s_{t-1})P(X_{it} = s_t | X_{it-1} = s_{t-1}, \mathbf{Z}_{it} = \mathbf{z}_t)P(\mathbf{Y}_{it} = \mathbf{y}_t | X_{it} = s_t)],$$

for $1 < t \leq T$, and

$$[A.8] \beta_{iT}(X_T = s_T) = 1, \text{ and}$$

$$[A.9] \beta_{it}(X_{it} = s_t) = \sum_{s_{t+1}=1}^S \beta_{it+1}(X_{i,t+1} = s_{t+1})P(X_{i,t+1} = s_{t+1} | X_{it} = s_t, \mathbf{Z}_{i,t+1} = \mathbf{z}_{t+1})P(\mathbf{Y}_{i,t+1} = \mathbf{y}_{t+1} | X_{i,t+1} = s_{t+1})$$

for $T > t \geq 1$.

Note that the step leading to $\alpha_{it}(X_{it} = s_t)$ is defined as the forward recursion step, and the step leading to $\beta_{it}(X_{it} = s_t)$ as the backward recursion step. Furthermore, all $\alpha_{it}(X_{it} = s_t)$ are referred to as forward probabilities, and $\beta_{it}(X_{it} = s_t)$ as backwards probabilities. In this paper the forward probabilities are used for lead generation purposes.

TABLE 1

Definition of notation

| Notation | Description |
|--|--|
| $i = 1 \dots I$ | index of subjects |
| $j = 1 \dots J$ | index of products |
| $k = 1 \dots K$ | index of covariates |
| $s = 1 \dots S$ | index of segments |
| $t = 1 \dots T$ | index of measurement occasions |
| $P(\mathbf{Y}_i=\mathbf{y} \mathbf{Z}_i=\mathbf{z})$ | denotes the probability that consumer i has ownership pattern \mathbf{y} across all measurement occasions, given that the covariate vector \mathbf{Z}_i of this person takes on the combination of values \mathbf{z} |
| $X_{ist} = s_t$ | implies that consumer i is member of segment s at measurement occasion t |
| \mathbf{Y}_{it} | denotes the $(1 \times J)$ vector of J product ownership indications for consumer i at measurement occasion t |
| $\mathbf{Y}_{it} = y_t$ | denotes the actual values on the $(1 \times J)$ vector of J product ownership indications for consumer i at measurement occasion t |
| Y_{ijt} | denotes the ownership indication for product j of consumer i at measurement occasion t , $Y_{ijt}=1$ if subject i owns j at t ; otherwise, $Y_{ijt}=0$ |
| \mathbf{Z}_{it} | denotes the $(1 \times K)$ vector of values that consumer i has on each of the K covariates at measurement occasion t |
| $\mathbf{Z}_{it} = \mathbf{z}_t$ | denotes the actual values on the $(1 \times K)$ vector of covariates for consumer i at measurement occasion t |

TABLE 2

Ownership levels of the analyzed products in each panel wave

| Product | 1996 | 1998 | 2000 | 2002 |
|---------------------------------|------|------|------|------|
| <u>Risky Investments:</u> | | | | |
| (1) Corporate/Government Bonds | .04 | .04 | .03 | .03 |
| (2) Shares | .08 | .10 | .11 | .11 |
| (3) Investment Trusts | .11 | .20 | .25 | .21 |
| <u>Income Management:</u> | | | | |
| (4) Unemployment Insurance | .10 | .21 | .22 | .22 |
| (5) Life Insurance | .59 | .60 | .59 | .55 |
| (6) Pension Fund | .62 | .67 | .64 | .59 |
| <u>Foundation:</u> | | | | |
| (7) Loan | .24 | .20 | .20 | .15 |
| (8) Credit Card | .29 | .30 | .35 | .41 |
| (9) Mortgage | .52 | .53 | .54 | .53 |
| (10) House Insurance (building) | .62 | .64 | .64 | .66 |
| (11) Car Insurance | .76 | .77 | .77 | .78 |
| (12) Savings Account | .93 | .95 | .96 | .96 |

TABLE 3*
Measurement model

| <u>MEASUREMENT MODEL</u> | | | | | | | | | |
|---------------------------|------------|-------------|------------|------------|------------|------------|------------|------------|------------|
| Segment | | | | | | | | | |
| Product | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| <u>Risky Investments:</u> | | | | | | | | | |
| (1) Bonds | .01 | .00 | .08 | .02 | .01 | .02 | .19 | .08 | .00 |
| (2) Shares | .01 | .01 | .14 | .06 | .03 | .08 | .55 | .26 | .02 |
| (3) Investment Trust | .04 | .06 | .19 | .11 | .12 | .21 | .76 | .43 | .08 |
| <u>Income Management:</u> | | | | | | | | | |
| (4) Unemploy. Ins. | .00 | .00 | .02 | .01 | .32 | .39 | .30 | .13 | .24 |
| (5) Life Insurance | .18 | .14 | .25 | .48 | .97 | .97 | .78 | .43 | .77 |
| (6) Pension Fund | .29 | .36 | .29 | .39 | .95 | .93 | .80 | .63 | .89 |
| <u>Foundation:</u> | | | | | | | | | |
| (7) Loan | .10 | .10 | .03 | .12 | .27 | .34 | .12 | .13 | .51 |
| (8) Credit Card | .06 | .05 | .18 | .32 | .04 | .98 | .69 | .71 | .32 |
| (9) Mortgage | .00 | .00 | .02 | .97 | .99 | .99 | .80 | .01 | .00 |
| (10) House Ins. | .00 | .01 | .98 | .93 | .99 | .97 | .99 | .01 | .02 |
| (11) Car Insurance | .00 | 1.00 | .76 | .74 | .91 | .94 | .94 | .87 | .84 |
| (12) Savings Acc. | .82 | .87 | .95 | .90 | .99 | .99 | .97 | .98 | .94 |

* To enhance interpretation of the measurement model, proportions ≥ 0.50 appear in bold, and those 0.30 to 0.49 in italics.

TABLE 4

Proportion in each segment

| Panel Wave | <u>Segment</u> | | | | | | | | |
|------------|----------------|-----|-----|-----|-----|-----|-----|-----|-----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1996 | .11 | .09 | .06 | .12 | .23 | .14 | .05 | .04 | .16 |
| 1998 | .11 | .09 | .06 | .12 | .20 | .17 | .06 | .04 | .15 |
| 2000 | .11 | .09 | .06 | .12 | .18 | .19 | .07 | .04 | .15 |
| 2002 | .11 | .09 | .06 | .12 | .16 | .21 | .08 | .04 | .14 |

TABLE 5

Transition Matrix

| Segment at t | Segment at t+1 | | | | | | | | |
|--------------|----------------|------------|------------|------------|------------|------------|-------------|------------|------------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1 | .92 | .02 | .00 | .01 | .00 | .00 | .00 | .01 | .04 |
| 2 | .05 | .92 | .00 | .01 | .01 | .00 | .00 | .01 | .00 |
| 3 | .01 | .01 | .95 | .03 | .00 | .00 | .00 | .00 | .00 |
| 4 | .00 | .00 | .02 | .95 | .00 | .01 | .01 | .01 | .00 |
| 5 | .00 | .01 | .00 | .01 | .85 | .12 | .00 | .00 | .01 |
| 6 | .00 | .00 | .00 | .00 | .00 | .96 | .03 | .00 | .01 |
| 7 | .00 | .00 | .00 | .00 | .00 | .00 | 1.00 | .00 | .00 |
| 8 | .00 | .00 | .01 | .00 | .00 | .03 | .01 | .95 | .00 |
| 9 | .00 | .00 | .00 | .00 | .04 | .03 | .00 | .01 | .92 |

* To enhance interpretation of the transition matrix, all values above 0.02 appear in boldface type.

TABLE 6

Effects of covariates on segment membership at t (1996)

| | Segment | | | | | | | | |
|--------------------------|---------|--------|--------|--------|--------|--------|--------|--------|--------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| <u>Income:</u> | | | | | | | | | |
| <2500 | 2.51* | 1.95* | 0.79* | 0.07 | -1.90* | -2.72* | -1.32* | 0.75* | -0.12 |
| 2500-3499 | 0.29 | 0.45* | 0.27* | 0.15 | -0.05 | 0.02 | -0.30 | -0.43 | -0.39 |
| 3500-4999 | -1.54* | -0.59* | -0.17 | 0.04 | 0.84* | 1.25* | 0.48* | -0.41 | 0.11 |
| ≥5000 | -1.25* | -1.80* | -0.90* | -0.25 | 1.11* | 1.46* | 1.15* | 0.09 | 0.40 |
| <u>Age:</u> | | | | | | | | | |
| <35 | -0.27 | 0.07 | 0.76* | 0.56* | 1.03* | 2.02* | 0.71* | -3.36* | -1.52* |
| 35-49 | -1.49* | -1.50* | 0.15 | 0.27 | 1.09* | 2.01* | 0.87* | -0.75 | -0.65 |
| 50-64 | -0.45 | -0.88* | 0.48 | -0.31 | 0.39 | 1.04* | 0.12 | -0.51 | 0.13 |
| ≥65 | 2.21* | 2.32* | -1.38* | -0.52* | -2.51* | -5.07* | -1.71* | 4.63* | 2.04* |
| <u>Household assets:</u> | | | | | | | | | |
| <10000 | 1.06* | -0.18 | 0.08 | 0.10 | -0.23 | 0.23 | -0.85* | -0.40* | 0.19 |
| 10000-50000 | -0.64* | 0.05 | 0.38* | 0.30 | 0.49* | 0.60* | -0.42* | -0.98* | 0.23 |
| >50000 | -0.42* | 0.13 | -0.46* | -0.41 | -0.26* | -0.83* | 1.28* | 1.38* | -0.42 |

* implies significance at the 0.05-level

TABLE 7

Effects of covariates on changes in segment membership

| | Segment at t+1 | | | | | | | | |
|----------------------------------|----------------|--------|-------|--------|--------|--------|--------|--------|-------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| <u>Income:</u> | | | | | | | | | |
| <2500 | 0.38 | 0.90* | 0.17 | 0.11 | -0.70* | -0.57* | -0.79* | 1.22* | -0.72 |
| 2500-3499 | -0.21 | 0.07 | 0.14 | -0.04 | 0.15 | -0.15 | 0.36 | -0.32 | 0.02 |
| 3500-4999 | 0.16 | -0.30* | 0.02 | -0.10 | 0.07 | 0.12 | 0.11 | -0.23 | 0.15 |
| ≥5000 | -0.33 | -0.67* | -0.33 | 0.03 | 0.49* | 0.61* | 0.32 | -0.66* | 0.55* |
| <u>Age:</u> | | | | | | | | | |
| <35 | -1.09* | -0.84 | -0.74 | 0.00 | 1.90* | 2.15* | 1.79* | -1.79* | -1.36 |
| 35-49 | -0.79 | -1.37* | -0.71 | -0.25 | 0.81* | 1.57* | 1.14* | 0.51* | -0.91 |
| 50-64 | 0.74 | 0.62 | 0.69 | 0.55 | -2.24* | -1.73* | -0.76 | 1.22* | 0.91 |
| ≥65 | 1.14* | 1.59* | 0.77 | -0.30 | -0.46* | -1.98* | -2.16* | 0.06 | 1.36 |
| <u>Household assets (x1000):</u> | | | | | | | | | |
| <10 | 1.79* | 0.45 | 0.52 | 0.82* | -0.28 | 0.44 | -0.85* | -3.18* | 0.30 |
| 10-50 | -1.10* | -0.36 | -0.19 | -0.11 | 0.03 | 0.02 | 0.13 | 1.77* | -0.17 |
| >50 | -0.68* | -0.09 | -0.33 | -0.70* | 0.26 | -0.46 | 0.73* | 1.42* | -0.13 |

* implies significance at the 0.05-level

TABLE 8

Gini values

| Product | Measurement occasions | | |
|----------------------|-----------------------|-------------|-------------|
| | 1996 → 1998 | 1998 → 2000 | 2000 → 2002 |
| (1) Bonds | 0.34 | 0.29 | 0.50 |
| (2) Shares | 0.34 | 0.20 | 0.31 |
| (3) Investment Trust | 0.28 | 0.23 | 0.26 |
| (4) Unemploy. Ins. | 0.36 | 0.42 | 0.36 |
| (5) Life Insurance | 0.36 | 0.40 | 0.32 |
| (6) Pension Fund | 0.09 | 0.18 | 0.07 |
| (7) Loan | 0.22 | 0.22 | 0.17 |
| (8) Credit Card | 0.21 | 0.21 | 0.19 |
| (9) Mortgage | 0.41 | 0.40 | 0.49 |
| (10) House Ins. | 0.43 | 0.34 | 0.44 |
| (11) Car Insurance | -0.22 | -0.26 | -0.23 |
| (12) Savings Acc. | 0.15 | 0.18 | 0.11 |

Figure 1.

Traditional Lorenz Curve

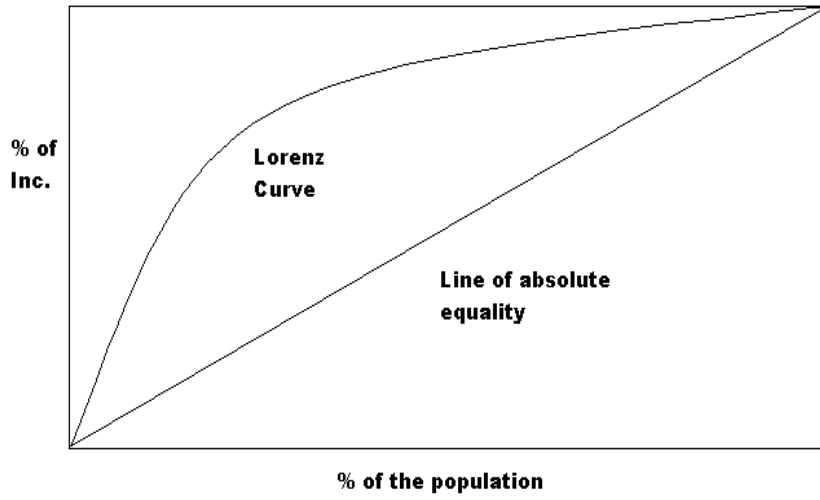
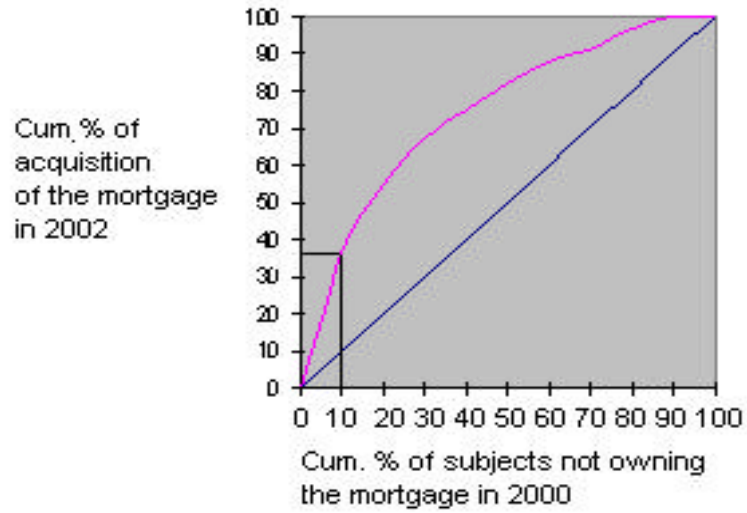


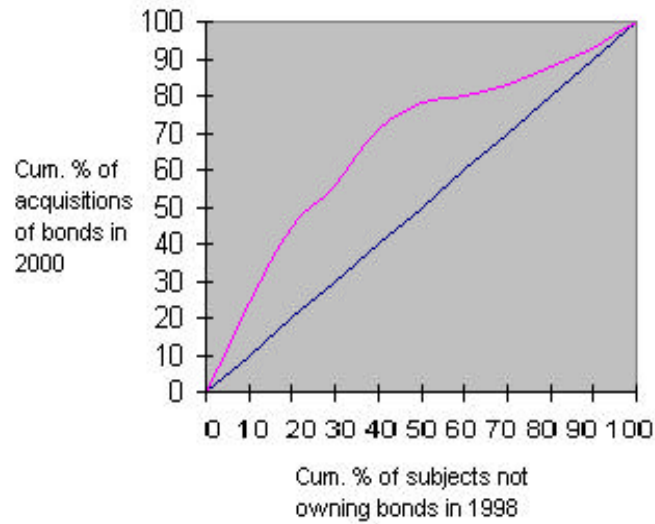
FIGURE 2

Power curves

Panel A. Power curve mortgage 2000 → 2002



Panel B. Power curve Bonds 1996 → 1998



REFERENCES

- Alba, Richard D. (1987), "Interpreting the parameters of log-linear models," *Sociological Methods and Research*, 16 (1), 45-77.
- Baum, Leonard E., Ted Petrie, George Soules and Norman Weiss (1970), "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Annals of Mathematical Statistics*, 41, 164-171.
- Böckenholt, Ulf and William R. Dillon (2000), "Inferring latent brand dependencies," *Journal of Marketing Research*, 37 (February), 72-87.
- Böckenholt, Ulf and Rolf Langeheine (1996), "Latent change in recurrent choice data," *Psychometrika*, 61 (2), 285-302.
- Brangule-Vlagsma, Kristine, Rik G.M. Pieters and Michel Wedel (2002), "The dynamics of value segments: Modelling framework and empirical illustration," *International Journal of Research in Marketing*, 19 (3), 267-285.
- Browning, Martin and Annamaria Lusardi (1996), "Household savings: Micro theories and micro facts," *Journal of Economic Literature*, 34 (4), 1797-1855.
- Cohn, Richard, Wilbur G. Lewellen, Ronald C. Lease and Gary G. Scharlbaum (1975), "Individual investor risk aversion and investment portfolio composition," *Journal of Finance*, 30 (May), 605-620.
- Dayton, C.M. and G.B. MacReady (1988), "Concomitant-variable latent-class models," *Journal of the American Statistical Association*, 83 (March), 173-178.
- Dempster, A.P., N.M. Laird, and D.B. Rubin (1977), "Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion)," *Journal of the Royal Statistical Society, Ser. B.*, 39 (1), 1-38.

- Dickenson, John R. and Eric Kirzner (1986), "Priority patterns of acquisition of financial assets," *Journal of the Academy of Marketing Science*, 14 (Summer), 43-49.
- Gunnarsson, Jonas and Richard Wahlund (1997), "Household financial strategies in Sweden: An exploratory study," *Journal of Economic Psychology*, 18 (2-3), 201-233.
- Gupta, Sachin and Pradeep K. Chintagunta (1994), "On using demographic variables to determine segment membership in logit mixture models," *Journal of Marketing Research*, 31 (February), 128-136.
- Hauser, John R. and Glen L. Urban (1986), "The value priority hypotheses for consumer budget plans," *Journal of Consumer Research*, 12 (March), 446-462.
- Kamakura, Wagner A., Kim, Byung-Do and Jonathan Lee (1996), "Modeling preference and structural heterogeneity," *Marketing Science*, 15 (2) 152-172.
- Kamakura, Wagner A., Sridhar N. Ramaswami and Rajendra K. Srivastava (1991), "Applying latent trait analysis in the evaluation of prospects for cross-selling of financial services," *International Journal of Research in Marketing*, 8 (4), 329-349.
- Kamakura, Wagner A. and Michel Wedel (2000), "Factor analysis and missing data," *Journal of Marketing Research*, 37 (November), 490-498.
- Kamakura, Wagner A., Michel Wedel, Fernando De Rosa and Jose A. Mazzon (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.

- Kasulis, Jack L., Robert F. Lusch and Edward F. Stafford (1979), "Consumer acquisition patterns for durable goods," *Journal of Consumer Research*, 6 (June), 47-57.
- Knott, Aaron, Andrew Hayes and Scott A. Neslin (2002), "Next-product-to-buy models for cross-selling applications," *Journal of Interactive Marketing*, 16 (Summer), 59-75.
- McDonald, Ian L. and Walter Zucchini (1997), "*Hidden Markov and other models for discrete-valued time series (Monographs on statistics and applied probability, number 70)*," London: Chapman and Hall.
- Paas, Leonard J. (1998), "Mokken scaling characteristic sets and acquisition patterns of durable and financial products," *Journal of Economic Psychology*, 19 (3), 353-376.
- Paroush, Jacob (1965), "The order of acquisition of consumer durables," *Econometrica*, 33 (1), 225-235.
- Poulsen, Carsten S. (1990), "Mixed Markov and latent class Markov modeling applied to brand choice behavior," *International Journal of Research in Marketing*, 7 (1), 5-19.
- Ramaswami, Sridhar N., Rajenda K. Srivastava and Thomas H. McInish (1992), "An exploratory study of portfolio objectives and asset holding," *Journal of Economic Behavior and Organization*, 19 (3), 285-306.
- Ramaswamy, Venkatram (1997), "Evolutionary preference segmentation with panel study data: An application to new products," *International Journal of Research in Marketing*, 14 (1), 57-80.

- Ramaswamy, Venkatram, Rabikar Chatterjee and Steven H. Chen (1996), "Joint segmentation on distinct interdependent bases with categorical data," *Journal of Marketing Research*, 33 (August), 337-350.
- Seetharaman, P.B. (2003), "Probabilistic versus random-utility models of state dependence: An empirical comparison," *International Journal of Research in Marketing*, 20 (1), 87-96.
- Sen, Amartya (1997), *On economic inequality (Expanded edition with a substantial annexe by James E. Foster and Amartya Sen)*, Oxford: Clarendon Press.
- Soutar, Geoffrey N. and Steven T. Cornish-Ward (1997), "Ownership patterns for durable goods and financial assets: A Rasch analysis," *Applied Economics*, 29 (11), 903-911.
- Srivastava, Rajendra K., Thomas H. McInish and Linda L. Price (1984), "Information costs and portfolio selection", *Journal of Banking and Finance*, 8 (3), 417-429.
- Stafford, Edward F., Jack J. Kasulis and Robert L. Lusch (1982), "Consumer behavior in accumulating household financial assets," *Journal of Business Research*, 10 (December), 397-417.
- Van de Pol, Frank and Rolf Langeheine (1990), "Mixed Markov latent class models," *Sociological Methodology*, 20, 213-247.
- Van de Pol, F. and J. de Leeuw (1986), "A Latent class Markov model to correct for measurement error," *Sociological Methods and Research*, 15, 1118-1141.
- Van Raaij, W. Fred and Theo M.M. Verhallen (1994), "Domain-specific market segmentation", *European Journal of Marketing*, 28 (10), 49-66.
- Verbeek, Marna and Theo Nijman (1992), "Testing for selectivity bias in panel data models," *International Economic Review*, 33 (3), 681-703.

- Vermunt, Jeroen K. (1997), *Log-linear models for event histories*, Thousand Oakes: Series QASS, vol. 8. Sage Publications.
- Vermunt, Jeroen K. (2001), "The use of restricted latent class models for defining and testing nonparametric and parametric item response theory models," *Applied Psychological Measurement*, 25 (3), 283-294.
- Vermunt, Jeroen K., and Jay Magidson (2000), *Latent GOLD 2.0 User's Guide*, Belmont, MA: Statistical Innovations Inc.
- Vermunt, Jeroen K., Rolf Langeheine and Ulf Böckenholt (1999), "Latent class Markov models with time-constraints and time-varying covariates," *Journal of Educational and Behavioral Statistics*, 24, 178-205.
- Wärneryd, Karl-Erik (1999), *The Psychology of Saving: A Study of Economic Psychology*, Northampton: Edward Elgar Publishing.
- Wedel, Michel, Wagner A. Kamakura, Wayne S. DeSarbo and Frenkel ter Hofstede (1995), "Implications for asymmetry, nonproportionality and heterogeneity in brand switching from piece-wise exponential hazard models," *Journal of Marketing Research*, 32 (November), 457-462.
- Wedel, Michel and Wagner A. Kamakura (2000), *Market segmentation: Conceptual and methodological foundations (second edition)*. Boston, Dordrecht, London: Kluwer Academic Publishers.
- Winer, Russell S. (1983), "Attrition bias in econometric models estimated with panel data," *Journal of Marketing Research*, 20 (May), 177-186.
- Winer, Russell S. (2001), "A framework for customer relationship management", *California Management Review*, 43 (4), 89-105.