# Long-term developments of respondent financial product portfolios in the EU: A multilevel latent class analysis

Leonard J. Paas, School of Communication, Journalism and Marketing, Massey University, Private Bag 102904, North Shore Auckland 0745, New Zealand; Department of Marketing, VU University Amsterdam, De Boelelaan 1105, 1081HV, Amsterdam, The Netherlands. L.J.Paas@massey.ac.nz.

Tammo H.A. Bijmolt, Department of Marketing, Faculty of Economics & Business, University of Groningen, Landleven 5, 9747 AD Groningen, The Netherlands, t.h.a.bijmolt@rug.nl, +31-50-3637319.

Jeroen K. Vermunt, Department of Methodology and Statistics, Tilburg University,
PO Box 90153, 5000 LE Tilburg, The Netherlands, j.k.vermunt@uvt.nl, +31-13-4662748.

# Long-term developments of respondent financial product portfolios in the EU: A multilevel latent class analysis

**Abstract.** Segmentation structures can be unstable over time. Therefore, previous research has analyzed panel data for providing insight into changes in segmentation structures or switches by individuals between segments. Unfortunately, panel data are often unavailable when analyzing developments across different countries and over longer time-periods. The analysis reported in this paper makes it possible to analyze differences in segmentation structures across countries at different time-points using multiple cross sectional datasets. This provides indications into long-term developments of segmentation structures across different countries. In the utilized multilevel latent class analysis model respondents are the lower level units. Data from the same country and time-point are treated as the higher level units of analysis. As an illustrative and salient empirical example we assess similarities and differences in consumer financial product portfolios across 14 EU countries from 1969 to 2003, based on three disaggregate cross-sectional databases.

**Keywords:** Cross-country segmentation, multilevel latent class analysis, financial services, long-term developments, European Union

## 1. Introduction

Segmentation analysis aims to classify units, such as respondents, in categories. However, segmentation structures are often unstable over time, because the classified units may switch between classes and the definitions of classes can change [38]. To accommodate for such dynamics previous research has applied latent Markov models [3,27] to panel data. However, panel data in which the same respondent is interviewed at multiple time-points are often unavailable, in particular when analyzing different countries over longer time-periods. The analysis method presented below, derives insights into similarities and differences between segmentation structures across countries and times. The analysed cross sectional datasets have been collected in different countries at different time-points. The derived insight facilitates assessment of long-term cross-country developments of segmentation structures.

To analyse the cross-sectional datasets mentioned above we apply multilevel Latent Class Analysis (LCA) [6,13,24], an extension of the more conventional LCA model [14,19]. In conventional LCA there is one type of measurement unit, typically the individual respondent, whereas multilevel LCA includes: (1) lower level units, e.g., respondents, and (2) higher level units, e.g., countries to which respondents are allocated. Furthermore, data from the same country and time-point lead to the definition of higher level units of analysis. For example, in our dataset The Netherlands is represented at three time points, namely 1969, 1990 and 2003. Dutch respondents that were interviewed in 1969 are coded as belonging to the higher level unit Netherlands_1969. Respondents from The Netherlands that were interviewed in 1990 are allocated to the higher level unit Netherlands_1990, while respondents from the Netherlands interviewed in 2003 are in Netherlands_2003. This approach is followed for all 14 countries in our dataset at the three measurement occasions, 1969, 1990 and 2003, which results in 42 higher level measurement units. A formal definition of the multilevel LCA model is provided below. The model is not new, but the coding of higher-level units based on country in addition to time is a novel application that provides insight into similarities and differences across countries at different points in time. This application is salient due to the commonly occurring absence of longitudinal data in empirical research conducted across different countries.

We apply the multilevel LCA model for analyzing similarities and differences between consumer financial product portfolios across 14 European countries at three time-points. This is an important application, because in various disciplines, empirical studies have been published on financial product portfolios, e.g., economic psychology [23,39], statistics [27,28], operations research [29], marketing research [6,18] and services marketing [25]. Insight into consumer financial product portfolios is relevant for marketing activities such as segmentation and cross-selling products [18,26]. In internationalized markets insights into the cross-national similarities and differences between financial product portfolios may reveal fruitful directions for international marketing strategy formulation [6]. In such a segmentation model, respondents with similar financial product portfolios are allocated to the same respondent-level segment and countries that are more similar in the occurrence of the various respondent-level segments are relatively likely to be allocated to the same country-level segment.

Previously published empirical studies generally analyze cross-sectional datasets on consumer financial product portfolios within a single country. To the best of our knowledge only three previous studies analyzed cross-country data on such product portfolios [6,24,28], but these were based on a single cross section. For single country segmentation research, analysis of time dynamics has been accepted as a key element to be studied [38]. Furthermore, previously published papers report the application of segmentation analysis to longitudinal data on consumer financial product portfolios, e.g., Paas et al. [27] and Prinzie and Van den Poel [29]. However, dynamics in financial product portfolio based segments across countries have been ignored. Given the dynamic nature of international financial markets one may anticipate that international segmentation structures, representing respondent financial product portfolios, will change over time. Financial firms require insight into such changes and will have to react to trends that occur across countries [6].

Below we analyze disaggregate data on financial product portfolios that occur across the 14 analyzed EU countries in 1969, 1990 and 2003. At each time-point a different random sample of respondents was interviewed. The presented multilevel LCA model facilitates analyzing the similarities and differences in segmentation structures across these 14 countries and over the three time-points. Such similarities and differences provide indications into long-term developments of segmentation structures across the 14 EU countries. Note that a latent Markov modeling approach [3] is less feasible for the data analyzed in this paper, because this model assumes longitudinal data in which the same respondent is interviewed at multiple measurement occasions. We analyze multiple cross sectional datasets collected at different time-points.

The contribution of the paper is twofold. First and foremost, the paper has a methodological contribution by providing an approach for using multiple international cross-national datasets when no longitudinal cross-sectional data are available. This approach can be applied in future empirical studies. Second, the presented application of multilevel LCA contributes to the extensive literature on consumer financial product portfolios [e.g., 6,15,18,23,24,25,27,28,29], as cross-national differences over time in such product portfolios have not been studied previously.

As for the organization of the paper, in section 2 we discuss theories that are relevant for explaining similarities and differences in developments in consumer financial product portfolios across countries. Section 3 presents the applied multilevel LCA model. In the empirical study, reported in sections 4 and 5, we assess country specificity of segmentation structures based on financial product ownership by individual respondents. The paper is concluded with a discussion on the theoretical and managerial implications, section 6.

## 2. International segmentation of financial product portfolios

The application of the multilevel LCA model to consumer financial product portfolios is of theoretical interest as there are different possible outcomes. At one extreme we would expect highly similar segmentation structures across different countries; while at the other extreme countries can be very different in terms of segmentation structures representing consumer financial product portfolios. Two contradicting theories lead to such different expectations.

First, theory on the development of consumer financial product portfolios within a single country suggests that similar segmentation structures of such product portfolios occur within and across countries. That is, consumer financial product portfolios are considered to be the observable outcomes of the respondent lifecycle and the resulting savings needs that respondents are likely to find relevant [18]. Concerning the latter, four hierarchically ordered saving motives have been studied extensively in the field of economic psychology [15,20,39]. The most basic is called the cash management motive, involving short-term financial issues, such as direct payment for transactions. At the second level, represented by the precautionary motive, respondents develop a financial reserve for unexpected expenditures. This is followed by the down-payment motive at the third level, i.e., accumulating a financial deposit for a house, saving for a car, an extensive vacation or durables. Fourth, wealth management consists of enterprise and investing assets. The second construct, the lifecycle hypothesis and the related permanent income hypothesis [7], assumes acquisitions result from consumer circumstances, such as the lifecycle phase and income. Young consumers require financial products for borrowing or investing small amounts of assets. Later in life, when income and assets increase, consumers require more sophisticated products for purposes such as speculation and asset accumulation.

Kamakura et al. [18] suggest that the lifecycle theory and saving needs are interrelated and lead to a common order for acquiring financial products. Such an acquisition order reflects assets that are invested and the risk levels of the different products. Consumers in early life-cycle stages have higher priorities for financial products related to basic motives in the saving motive hierarchy, involving fewer assets and relatively low risk levels, e.g., saving accounts. Products that are relevant for higher order motives, e.g., shares, will often be acquired by consumers with more assets and higher levels of financial knowledge. We conjecture that the tendency of consumers to acquire more basic financial products before products satisfying higher order needs may occur across countries. The latter would imply that consumers in different countries tend to own higher order products when they also own products relevant for the satisfaction of more basic needs, resulting in similar financial product-portfolios across countries, i.e., similar segmentation structures.

Contradicting expectations can be forwarded, due to country-specific situational factors. Cultural factors may be particularly relevant herein, such as those captured through the well-known Hofstede dimensions [17]. For example, cross-country differences on the uncertainty avoidance dimension in Hofstede's model may be reflected in risk aversion of consumers in the financial market, resulting in country specificity in the development of respondent financial product portfolios. In low uncertainty avoidance cultures the consumer may less often develop a financial buffer in safe assets, such as savings accounts, but may first invest in risky assets, such as shares. Cultural differences, on for example uncertainty avoidance and on the other Hofstede dimensions [11], persist between EU countries, the region that is analyzed in our empirical study. Another relevant issue concerns GDP differences across countries. Extant research has consistently shown that consumers with higher income levels are more likely to own various financial products [7,39]. We conjecture that this may also lead to differences in consumer financial product portfolios across countries with varying GDP's.

In sum, the lifecycle hypothesis [7] and the savings need hierarchy [15,20,39] would suggest that the segmentation structures derived from consumer financial product portfolios are likely to be similar across countries. Contrarily, the literature on persisting cultural differences across countries in the EU [17] and the discussion above on the relevance of income for financial product portfolios suggest the occurrence of differences between segmentation structures across EU countries. This contradiction implies that the application of the multilevel LCA presented below is theoretically relevant.

## 3. Model specification

In this paper a multilevel LCA is employed in a novel way. Bijmolt et al. [6] and Paccagnella and Varriale [28] previously used multilevel LCA for assessing financial product portfolios across countries at one time-point. We apply this model to data collected in multiple countries at various time-points, providing insight into the differences in consumer financial product portfolios across countries and time-points. These differences allow assessment of changes in product portfolios in the same country across multiple time-points and whether developments are similar or different across countries.

We define the applied model using the notation that was presented previously in Bijmolt et al [6]. Assume that data are available on multiple respondents, denoted $i = 1,.., I$. The $I$ respondents are from a set of countries, denoted $j = 1,..., J$, and were interviewed at various time-points, $t = 1,\ldots, T$. For each respondent $i$, product ownership indications are available on a set of products, denoted $k = 1,..., K$, where $Y_{ijtk} = 1$, if respondent $i$ from country $j$ interviewed at time-point $t$ owns product $k$, else $Y_{ijtk} = 0$. All ownership indications for respondent $i$ are represented in the vector $Y_{ijt.}$. Furthermore, $Y_{jt}$ is a matrix containing ownership data of all respondents of country $j$ at time-point $t$. The model assumes $S$ respondent-level latent classes, denoted $s = 1 ,\ldots, S$. The analyzed countries belong to higher level latent classes at the different time-points, denoted $h = 1 ,\ldots, H$. In our application, the dataset analyzed contains information from 14 countries at three time-points resulting in 42 time-by-country combinations. These 42 higher level units are placed in $H$ higher-level latent classes. The categorical latent variable $X_{ijt}$ represents respondent latent class membership probabilities and $Z_{jt}$ represents higher level latent class membership probabilities. Note that we formulate a two-level model involving all possible time-country combinations. It would also be possible to formulate a three level model, with times denoting the second level and countries the highest level three. A two-level formulation facilitates a more flexible allocation of countries at different points in time to the same higher-level segment. Furthermore, the number of higher level units is too limited for the formulation of a three-level model.

Multilevel LCA models [6,34] consist of two model equations, one at the respondent level (Equation 1) and another at the higher level, i.e., the country-by-time level in our study (Equation 2). For respondents, probabilities of product ownership, $Y_{ijt}$, for respondent $i$ from country $j$ at time-point $t$, is conditional on membership of country $j$ at time-point $t$ to the higher-level latent classes, $h$, i.e.:

$$(1) \qquad P(Y_{ijt} \mid Z_{jt} = h) = \sum_{s=1}^{S} P\left(X_{ijt} = s \mid Z_{jt} = h\right) \prod_{k=1}^{K} P\left(Y_{ijtk} \mid X_{ijt} = s\right) .$$

Equation (1) can be interpreted as a regular LCA model. However, in this multilevel LCA equation the relative sizes of the $S$ latent classes at the respondent level are affected by the higher-level latent class membership probabilities of the 42 higher-level units. As in the conventional LCA model it is assumed that within a latent class $s$, the ownership probabilities of the various financial products are independent, i.e., local stochastic independence.

At the country-by-time level, a second equation is specified:

$$(2) \qquad P(Y_{jt}) = \sum_{h=1}^{H} P\left(Z_{jt} = h\right) \prod_{i=1}^{N_{jt}} P\left(Y_{ijt} \middle| Z_{jt} = h\right),$$

where $N_{jt}$ represents the sample size of country $j$ as interviewed at time-point $t$. A combination of equations (1) and (2) leads to the following:

$$(3) \qquad P(Y_{jt}) = \sum_{h=1}^{H} \left[ P\left(Z_{jt} = h\right) \prod_{i=1}^{N_{jt}} \left[ \sum_{s=1}^{S} P\left(X_{ijt} = s \middle| Z_{jt} = h\right) \prod_{k=1}^{K} P\left(Y_{ijkt} \middle| X_{ijt} = s\right) \right] \right] .$$

In equation (3) the outcome $P(Y_{jt})$ results from: (a) the probability that country $j$ at time-point $t$ belongs to the higher-level latent class $h$, (b) the probability respondent $i$ belongs to consumer latent class $s$, given the higher-level latent class membership probabilities, $h$, and (c) the probability respondent $i$ owns product $k$, given the respondent latent class membership probability, $s$. Component (c) of equation (3) involves the differences between the $S$ respondent-level latent classes, which are the conditional probabilities that a respondent owns a product $k$, given this unit's segment membership probabilities. For parameter estimation, these probabilities are parameterized through a logit equation:

$$(4) \qquad P\left(Y_{ijtk} = 1 \middle| X_{ijt} = s\right) = \frac{\exp\left(\beta_{ks}\right)}{1 + \exp\left(\beta_{ks}\right)},$$

where $\beta_{ks}$ is the logit of ownership of product $k$ by individuals belonging to class $s$**.** Component (b) of equation (3) assesses key differences between the $H$ higher-level latent classes. This is based on the relative sizes of respondent-level latent classes in country $j$ at time point $t$, i.e.:

$$(5) \qquad P\left(X_{ijt} = s \middle| Z_{jt} = h\right) = \frac{\exp\left(\gamma_{sh}\right)}{\sum_{s'=1}^{S} \exp\left(\gamma_{s'h}\right)},$$

where $\gamma_{sh}$ is a multinomial logit coefficient. We also anticipate effects of demographics on respondent latent class membership probabilities. These covariate effects are modeled using the concomitant approach [1,9,12], which was applied previously by Bijmolt et al [6] for retrieving covariate effects in multilevel LCA. The model for the lower-level classes becomes

$$(6)\ P(X_{ijt} = s|Z_{jt} = h, W_{ijt}) = \frac{\exp(\gamma_{sh0} + \sum_{p=1}^{P} \gamma_{sp} W_{ijtp})}{\sum_{s'=1}^{S} \exp(\gamma_{s'h0} + \sum_{p=1}^{P} \gamma_{s'p} W_{ijtp})},$$

where $W_{ijtp}$ is one the $P$ covariate and $\gamma$ are logistic regression parameters.

Model parameters can be estimated through maximum likelihood using an Expectation-Maximization (EM) algorithm [5,10]. For this purpose, a special implementation of the E step of the EM algorithm is required, which was described in detail by Vermunt [34,35,36], who called it an upward-downward algorithm. In the upward step, one obtains the posterior probabilities for the higher-level class memberships based on all information of the higher-level country-time unit concerned; that is, $P(Z_{jt} = h|Y_{ij})$. In the downward step, one obtains the bivariate posteriors $P(Z_{jt} = h, X_{ijt} = s|Y_{ij})$. With $P(Z_{jt} = h|Y_{ij})$ and $P(Z_{jt} = h, X_{ijt} = s\big|Y_{ij}\big)$, one can construct the expected complete data log-likelihood which is maximized in the M step of the EM algorithm. For the latter, standard Newton-type algorithms for logistic regression can be used. We used the Latent GOLD 5.0 program [37] for our analysis. The R package MultiCIRT [4] also implements the multilevel LCA models.

Reweighting was required because cross-country respondent-level data are generally based on national samples that are disproportional for the actual population size. That is, in the sample a larger percentage of respondents were selected from those countries with smaller populations. For example, the 2003 dataset includes 493 respondents from Luxembourg with a population of approximately half a million. The number of respondents from Great Britain is approximately double that of Luxembourg in 2003, namely 1054. This is not proportionate to the larger population of Great Britain, almost 60 million in 2003. In a previously reported application of multilevel LCA, Bijmolt et al. [6] employed the reweighting procedure suggested by Ter Hofstede et al. [33], in which respondents from countries with smaller populations are down-weighted, while respondents from countries with larger populations are up-weighted. In the current paper, we use a different kind of weighting which fits better to the multilevel nature of our data and model. It involves using weights at both the respondent level and the higher country-time level. The approach reweights (down-weights) the individual-level data in such a manner that each country-time combination contains the equivalent of 100 respondents, which corresponds with a multilevel analysis with rather large (and equally sized) groups. At the country level, weights are applied to accommodate for population size differences between countries and time points: Countries with larger populations receive higher weights, where for each country-time combination the sum of the product of the lower- and higher-level weights are proportional to the population sizes. Similar weighting schemes with products of multilevel weights proportional to the total number of lower-level units per higher-level units in the population have been proposed for other types of multilevel level models by Asparouhov and Muthen [2] and Rabe-Hesketh and Skrondal [30].

Relative fit of alternative model specifications is evaluated using the minimum Bayesian Information Criterion (BIC) rule [31]. In the comparison of the relative fit of alternative models we use the three-step procedure introduced by Lukociene et al. [21]. In the first step of this procedure, one estimates a conventional LCA model [14,19] with a single level of latent classes for segmenting the respondents. Here we use the BIC based on the number of individuals. Next the number of lower-level latent classes is fixed to the value that has been attained through step 1 and the appropriate number of higher-level classes is determined using the minimum BIC rule, but now using the number of country-time combinations as the sample size (step 2). In the final step 3 the number of higher-level classes is fixed to the value attained in step 2 and the number of lower-level classes is determined again using BIC based on the lower-level sample size. To account for sub-optimal solutions we estimated each model 160 times with different random starting values, retaining the best solution.

## 4. Data and analysis

We analyze multiple cross-sectional datasets that were collected in: 1969, 1990 and 2003. The 1969 and 1990 data were collected for the EURODATA project of the Readers Digest Respondent Survey. The 1969 sample contains 24,180 respondents from 16 countries and the 1990 sample 22,339 individuals from 17 countries. Respondents were 18 years or older and were living in private respondents. The 2003 data, Eurobarometer 60.2, were collected by a consortium of market research agencies at request of the European Commission, Directorate-General Press and Communication, Opinion Polls. It covers the population (aged 15 years and over) of the 15 EU member states in 2003 and consists of 16,200 respondents. In all years interviews were face-to-face and in the appropriate national language.

The empirical study includes those 14 countries that are represented at all three time-points. Also, respondents aged between 15 and 17 were excluded from the 2003 dataset. This leads to a database with 51,674 respondents, aged 18 years and older, from Austria, Belgium, Denmark, Finland, France, West Germany, Great Britain (including Northern Ireland), Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain and Sweden. Thus, the analyzed dataset covers all countries that joined the EU prior to the ten-country extension in 2004, except Greece and former East Germany. In total we have 42 country-time combinations, which are the higher level units in our multilevel LCA. We reweighted the individual observations to equal 100 per group, as outlined in section 3.

Because we analyze financial product portfolios in 1969, 1990 and 2003, we can assess country-level developments in respondent financial product portfolios. Note that a long time-span is analyzed, because effects of internationalization are expected to take place over longer periods [11]. Also, financial product portfolios are developed over long-term respondent lifecycles [7,25,39]. The disaggregate nature of the dataset implies insight is provided into similarities and differences across respondents within the same country and also between respondents across countries [6] and across measurement occasions.

We analyze the following products: savings account, life insurance, bonds, investment trusts, and shares. Table 1 presents the time-specific penetrations of these financial products, ranked according to decreasing risk levels [23], across all analyzed countries. Shares are the most risky assets, depending on highly volatile exchange values on the stock market. Investment trusts concern a mix of different shares and corporate/government bonds and are subjected to smaller value fluctuations. Next are the corporate/government bonds. Life insurances are less risky again. Like investment trusts these concern a mix of shares and bonds. However, investments are over longer periods and, therefore, involve lower risks [23]. Besides this, life insurance policies have a guaranteed minimum pay-off. Least risky is the savings account.

INSERT TABLE 1 HERE

The database also includes three respondent-level demographics: age, income and marital status. Previous research has shown these variables to be of key importance for respondent financial product portfolios [7,21,39]. The demographic variables are used to profile the financial product portfolio based segments derived from the dataset, by being included in the model as covariates.

## 5. Empirical results

### *5.1. Model selection*

As discussed above, we employed the three-step procedure [21] for determining an appropriate number of lower- and higher-level latent classes in the employed multilevel LCA. Table 2 presents the BIC values for the alternative models in each step. In the first step, which ignores the multilevel structure in the LCA model, we find that the LCA with five latent classes results in the lowest BIC. In the second step the number of lower-level latent classes consequently is fixed to five and the number of higher-level latent classes is increased from 1 to 10, while estimating the multilevel LCA. The model with nine higher-level classes leads to the lowest BIC, see Table 2. In step 3 the number of higher-level latent classes is set to nine and the number of lower-level latent classes is again investigated using multilevel LCA. For this purpose, we estimated models with one class less and one class more than five. We find again that five lower-level latent classes result in the lowest value on BIC. Thus, the model with five lower level latent classes for segmenting respondents and nine higher level classes for segmenting countries is selected.

INSERT TABLE 2 HERE

### *5.2. The segmentation solution*

Table 3 presents latent class specific product penetrations. In the first latent class, ‘in-actives’ ($s$=1), penetrations of all products are close to zero. In the second class, ‘savers’ ($s$=2), only the savings account has a high penetration. In the third class, ‘life-savers’ ($s$=3), respondents own the savings account and the life insurance, but not the three other financial products in Table 3. In the fourth class, ‘life-savers with bonds’ ($s$=4), we find respondents that generally own a savings account, and relatively often have a life insurance and bonds. In the fifth latent class, i.e., ‘actives’ ($s$=5), are respondents that often own savings accounts, life insurances, investment trusts and shares. Sometimes they also own bonds.

INSERT TABLE 3 HERE

The results in Table 3 suggest that respondents have followed similar orders for acquiring financial products. That is, in each latent class where the other four products have high penetrations the savings account is also commonly owned, implying that respondents usually acquire savings accounts before the other products in our dataset [6,27]. Next they are likely to acquire a life insurance, because this product has high penetrations in classes where the bonds, investment trust and shares also are commonly owned. In $s$=4 we find a high penetration for bonds and lower penetrations for investment trusts and shares. Contrarily, in latent class $s$=5 we find higher pemetrations for investment trusts and shares than for bonds. This difference implies some respondents owning a savings account and life insurances chose between acquiring bonds, on the one hand, and some chose to acquire investments trust and/or shares first.

Table 4 reports the covariate effects of the demographics, has partner (y/n), age category and income above median (y/n), on the membership probabilities for the $S$ lower level latent classes. To summarize the findings in Table 4, effects of income, age and having a partner are all significant ($p$<0.01). Respondents with an income above the median are less often in the segments in which few products have high penetrations, i.e., in-actives ($s$=1) and savers ($s$=2). This also applies for respondents with a head of intermediate or higher age (51+) and for respondents in which the head has a partner. These findings are mostly consistent with extant theory and previous research [6,35] and thereby support validity of our results.

INSERT TABLE 4 HERE

Part 2 of Table 3 presents the *H* higher-level latent classes which represent the commonly occurring combinations of the *S* lower-level latent classes within the 42 different higher-level units. Part 2 of Table 3 shows some higher-level latent classes are characterized by respondent level latent class structures that are not very advanced in terms financial product portfolios. For example, higher-level latent class $h$=1 contains countries at a time that many of the respondents (79.33% on average) are in the in-active segment $s$=1. Contrarily, if at a specific time-point $t$ a country has many respondents that are actives ($s$=5) the corresponding segmentation structure of the country at time $t$ is more likely to be allocated to for example $h$=9. To enhance interpretation of Part 2 of Table 3, we have ordered the nine higher-level latent classes according to decreasing proportions of the inactive respondent-level, i.e., respondents-level latent class $s$=1 and increasing proportions of the active respondents, i.e., $s$=5. In the next section we discuss the clustering of the 14 countries at the three time-points into the nine higher-level clusters described in part 2 of Table 3.

### *5.3. Clustering of Countries*

As mentioned above, the analyzed dataset contains 42 higher-level units, i.e., 14 countries at three-time-points, 1969, 1990 and 2002. Allocation of the 42 higher-level units in *H* higher-level latent classes is based on the occurrence of the five respondent-level latent classes. The respondent-level latent classes represent the lower level segmentation structure in the multilevel LCA model. For example, respondent level latent class $s$=1 in Part 1 of Table 3 is defined by an average ownership probability of 0.0014 for shares, 0.0004 for investment trusts, 0.0017 for bonds, 0.0278 for the life insurance product and 0.0144 for the savings account. These proportions reflect ownership probabilities of the five analyzed financial products for respondents allocated to latent class $s$=1. If the occurrence of the five respondent level latent classes is highly similar across higher-level units, the higher-level units involved will be allocated to the same country-by-time latent class. Consider for an illustrative hypothetical example that in 1969, we find 25% of the Belgian respondents in lower-level latent class *s*=1, 50% of the respondents in this country is found in *s*=3 and the remaining 25% of the Belgian respondents is in *s*=4. Moreover, in France (1969) we find a highly similar segmentation structure, with 20% of the respondents in *s*=1, 60% of the respondents in this country in *s*=3 and the remaining 20% of the French respondents is in latent class *s*=4. Under such circumstances Belgium (1969) and France (1969) are likely to be allocated to the same higher level latent class *h*. However, if two or more higher-level units are characterized by dissimilar segmentation structures, they will be allocated to different country-by-time clusters. Building on our previous hypothetical example involving Belgium and France, consider that in Great Britain in 2003 50% of the respondents are allocated to latent class *s*=2 and the other 50% to *s*=5. Under these circumstances Great Britain (2003) is unlikely to be allocated to the same higher level latent class *h* as Belgium (1969) and France (1969). Segmentation structures of countries at different time-points can also be allocated to the same country-by-time cluster. For example, the respondent level segmentation structure of Portugal in 2003 is highly similar to the structure characterizing Denmark in 1990. Therefore, Portugal-2003 and Denmark-1990 are placed in the same higher-level latent class *h*=5. This suggests Portugal is lagging in terms of the development of consumer financial product portfolios and Denmark is more advanced.

Table 5 reports the complete allocation of the 14 countries in 1969, 1990 and 2003 to the nine higher-level latent classes. The 42 higher-level units were allocated to the *H* higher level latent classes using the modal classification rule, based on the posterior membership probabilities in the upward-downward implementation of the EM algorithm that is discussed in section 3. A very obvious observation that can be derived from Table 5 is that financial product portfolios have become more divergent across the 14 analyzed countries. In 1969 and

1990 these countries were allocated to only two and three of the higher-level latent classes. This increased to five higher-level latent classes in 2003.

INSERT TABLE 5 HERE

A second finding reported in Table 5 is that some sets of countries have displayed similar developmental patterns in terms of the occurrence of financial product-portfolio based respondent-level latent classes. This is reflected in the same or highly similar allocation of countries to the higher-level latent classes across the three time-points represented in the analyzed data. First of all, in six of the analyzed countries the development, in terms of the allocation to higher-level latent classes, is the same: (1) Austria, Belgium, Luxembourg, Finland, France and Netherlands. These countries were in higher-level latent class $h$=2 in 1969, higher-level latent class $h$=5 in 1990 and higher-level latent class $h$=7 in 2003.

Next to this Table 5 shows that the three Mediterranean countries represented in the dataset, Italy, Portugal and Spain, also underwent similar developments. These three countries were allocated to higher-level latent class $h$=1 in 1969. They proceeded to higher-level latent class $h$=4 in 1990 and then Portugal and Spain entered higher-level latent class $h$=5 in 2003, which also includes some of the more advanced North-European markets in a previous time-point, 1990. This lagging of two of the three Mediterranean countries suggests that successful marketing activities used in the more advanced markets at a previous time-point can be used later in less advanced markets, which is obviously relevant for applied marketers in the financial services sector. Italy was allocated to higher-level latent class $h$=6 in 2003, a higher-level latent class that doesn't include any other countries at any time-point, implying that in 2003 Italy had a relatively unique segmentation structure.

A third set of countries, GB, Ireland and Denmark, started in higher-level latent class $h$=7 in 1969 and proceeded to higher-level latent class $h$=8 in 2003, via different higher-level latent classes in 1990. That is GB and Denmark were in higher-level latent class $h$=5 in 1990 and Ireland was in higher-level latent class $h$=4, with the Mediterranean countries. The economic boom of the 1990's may explain why Ireland has developed rapidly between 1990 and 2003 in terms of consumer financial product portfolios, catching up with more developed North European countries.

Two countries display divergent developments, i.e., West-Germany and Sweden. Perhaps conditions specific for these countries have resulted in their unique developmental pattern. For example, Sweden is allocated to the higher-level cluster $h$=9 which represents an advanced segmentation structure, see Table 3. No other countries have been allocated to $h$=9 at any of the three measurement occasions. This finding may be due to the Swedish pension scheme in which all citizens in paid employment are provided with the opportunity to allocate part of their savings to equity funds and other risk-bearing securities [16].

In sum, although some countries display similar patterns in development, countries do not all follow the same developmental process in terms of the financial product portfolios that that consumers have. Segmentation structures develop in different ways and at different speeds. These differences have resulted in an increased diversity in segmentation structures across the 14 analyzed countries in 2003.

## 6. Discussion

In this paper we have illustrated the application of the multilevel LCA model for deriving insights into cross-national longitudinal developments by analyzing multiple cross sectional datasets. In the absence of cross-national longitudinal data our approach can provide relevant insight into such developments at the country level. The application of the multilevel LCA model reported in the current paper has led to some interesting findings. These results will be discussed below and then we conclude with a discussion on methodological implications of the application of the multilevel LCA model that was presented in this paper.

The reported empirical results are relevant at the consumer level and at the higher level of countries at different time-points, i.e., 1969, 1990 and 2003. Regarding the former, our findings suggest that consumers in the 14 analyzed EU countries generally acquire products according to increasing risk levels, i.e., less risky products are acquired before more risky products, in section 5.2. The suggested pattern does not seem to be disturbed by contextual differences between countries; otherwise we would have found segments in which higher-order products have higher penetrations than the more basic products. Although countries are developing along the same pattern across time-points up to 34 years apart (1969 and 2003), different rates of development seem to apply. For example, the Mediterranean countries in our dataset, Italy, Portugal and Spain, seem to be lagging in the development of consumer financial product portfolios. The differences in the rate of development may have led to a larger number of segments in the 14 analyzed EU countries in 2003 (see Table 5).

The reported results have managerial implications for financial firms in the EU. First, given the similarity in the developmental order of consumer financial product portfolios across countries and time is relevant for financial firms. For example, in all countries respondents owning only a savings account can be offered a life insurance product first, while those respondents that already own a life insurance product can be offered an investment product. Moreover, at the country level marketers can consider the application of market strategies that have proven to be successful in the more advanced countries, such as Sweden, in other countries.

Evidently, our findings call for further research in other services or product markets. Other avenues for future research involve two important limitations of our study. First, we only studied 14 EU countries. Amongst these similar countries we already found substantial differences. Investigating a larger variation of countries would be interesting, particularly if developing countries or the new EU members (joining in 2004 or later) and also Greece would be included. Second, further research could aim to investigate the effects of country-level variables on developments in respondent financial product portfolios.

From a methodological perspective, we have illustrated the application of the multilevel LCA model for analyzing multiple cross-sectional datasets to assess differences between countries and across time-points, which can be used to assess developments in countries and differences in developments across different countries. In the commonly occurring absence of truly longitudinal cross-national datasets this is a relevant application of the multilevel LCA model that can also be used in future research on respondent financial product portfolio development and also for other substantive applications such as developments in ownership combinations of consumer durable products, developments in consumer's personal values across nations or social-demographics developments across nations. Our paper has shown that the application of multilevel LCA to cross-sectional datasets derived from different countries at different time-points can lead to interesting conceptual insights relevant for developing academic theory.

## References

1. Andrews, R.L., Currim, I.S.: Recovering and Profiling the True Segmentation Structure in Markets: An Empirical Investigation. International Journal of Research in Marketing **20**, 177-192 (2003)
2. Asparouhov, T. & Muthen, B.: Multilevel modeling of complex survey data. Proceedings of the Joint Statistical Meeting in Seattle, August 2006. ASA section on Survey Research Methods, 2718-2726 (2006)
3. Bartolucci, F., Farcomeni, A., Pennoni, F.: Latent Markov Models for Longitudinal Data. Chapman and Hall / CRC Press, Boca Raton (2013)
4. Bartolucci, F., Bacci, S., Gnaldi, M.: MultiCIRT: An R package for multidimensional latent class item response models. Computational Statistics and Data Analysis **71**, 971-985 (2014)
5. Baum, L.E., Petrie, T., Soules, G., Weiss, N.: A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. Annals of Mathematical Statistics **41**, 164-171 (1970)
6. Bijmolt, T.H.A., Paas, L.J., Vermunt, J.K..: Country and respondent segmentation: Multilevel latent class analysis of financial product ownership. International Journal of Research in Marketing **21**, 323-340 (2004)
7. Browning, M., Lusardi, A.: Respondent Savings: Micro Theories and Micro Facts. Journal of Economic Literature **34**, 1797-1855 (1996)
8. Craig, C.S., Douglas, S.P.: Cross-country Marketing Research, 3rd Ed. Wiley, New York (2005)
9. Dayton, C.M., Macready, G.B.: Concomitant-variable latent-class models. Journal of the American Statistical Association **83**: 173-178 (1988).
10. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Society Series B **39**, 1-38 (1977)
11. De Mooij, M., Hofstede, G.: Convergence and divergence in respondent behavior: Implications for international retailing. Journal of Retailing **78**, 61-69 (2002)

12. Formann, A.K.: Mixture analysis of multivariate categorical data with covariates and missing entries. Computational Statistics and Data Analysis **51**, 5236-5246 (2007)
13. Gnaldi, M., Bartolucci, F., Bassi, F.: A multilevel finite mixture item response model to cluster examinees and schools. Technical Reprot: arXiv: 1408.2319 (Paper accepted for publication in Advances on Data Analysis and Clssification) (2014)
14. Goodman, L.A.: Exploratory latent structure analysis using identifiable an unidentifiable models. Biometrika **61**, 215-231 (1974)
15. Gunnarson, J., Wahlund, R.: Respondent financial strategies in Sweden: An exploratory study. Journal of Economic Psychology **18**, 211-233 (1997)
16. Hedesström, T.M., Svedsäter, H., Gärling, T.: Indentifying heuristic choice rules in the Swedish premium pension scheme. Journal of Behavioral Finance **5**, 32-42 (2004)
17. Hofstede, G. (2001). Cultures consequences: Comparing values, behaviors, institutions and organizations across nations (second edition). Sage Publications, Thousand Oaks (2001).
18. Kamakura, W.A., Ramaswami, S.N., Srivastava, R.K.: Applying latent trait analysis in the evaluation of prospect for cross-selling financial products. International Journal of Research in Marketing **8**, 329-349 (1991)
19. Lazarfeld, P.F., Henry, N.W.: Latent Structure Analysis. Houghton Mifflin, Boston (1968)
20. Lindqvist, A.: A note on the determinants of respondent saving behaviour. Journal of Economic Psychology **1**, 39-57 (1981)
21. Lukociene, O., Varriale, R., Vermunt, J.K.: The simultaneous decision(s) about the number of lower- and higher-level classes in multilevel latent class analysis. Sociological Methodology **40**, 247-283 (2010)

22. Magidson, J., Vermunt, J.K.: Latent class models. In: D. Kaplan (ed.), The sage handbook of quantitative methodology for the social sciences, pp. 175-198. Sage Publications, Thousand Oakes (2004)
23. Ngwenya, M., Paas, L.J.: Lifecycle effects on consumer financial product portfolios in South Africa: An analysis of four ethnic groups. Journal of Economic Psychology **33**, 8-18 (2012)
24. Paas, L.J.: Acquisition patterns of products facilitating financial transactions: A Cross-country investigation. International Journal of Bank Marketing **19**, 266-275 (2001)
25. Paas, L.J.: Acquisition pattern analysis for evolutionary database marketing. The Service Industries Journal **29**, 805-812 (2009)
26. Paas, L.J.: Comment on: Latent Markov models: a review of a general framework for the analysis of longitudinal data with covariates. TEST **23**, 473-477 (2014)
27. Paas, L.J., Vermunt, J.K., Bijmolt, T.H.A.: Discrete time, discrete state latent Markov modelling for assessing and predicting respondent acquisitions of financial products. Journal of the Royal Statistical Society **A170**, 955-974 (2007)
28. Paccagnella, O, Varriale, R.: Asset ownership of the elderly across Europe: A multilevel latent class analysis to segment country and respondent. Advances in Theoretical and Applied Statistics, 383-393 (2013)
29. Prinzie, A., Poel, D. van den: Investigating purchasing patterns for financial services using Markov, MTD and MTDg models. European Journal of Operations Research **170**, 710-734 (2006)
30. Rabe-Hesketh, S. and Skrondal, A.: Multilevel modeling of complex survey data. *Journal of the Royal Statistical Society, Series A* **169** (4), 805-827 (2006)
31. Schwarz, G.: Estimating the dimensions of a model. Annals of Statistics **6**, 461-464 (1978)

32. Steenkamp, J.-B.E.M.: Moving out of the US silo: A call to arms for conducting cross-country marketing research. Journal of Marketing **69**, 6-8 (2005)

33. Ter Hofstede, F., Steenkamp, J.-B.E.M., Wedel, M.: Cross-country Market Segmentation Based on Respondent-Product Relations. Journal of Marketing Research **36**, 1-17 (1999)

34. Vermunt, J.K.: Multilevel Latent Class Models. Sociological Methodology **33**, 213-239 (2003)

35. Vermunt, J.K.: A hierarchical mixture model for clustering three-way datasets. Computational Statistics and Data Analysis, 51, 5368-5376 (2007)

36. Vermunt, J.K.: Latent class and finite mixture models for multilevel datasets. *Statistical Methods in Medical Research*, 17, 33-51 (2008)

37. Vermunt, J.K., Magidson, J.: Technical Guide for Latent GOLD 5.0: Basic, Advanced, and Syntax. Statistical Innovations Inc., Belmont (2013)

38. Wedel, M., Kamakura, W.A.: Market Segmentation: Conceptual and Methodological Foundations, 2nd Ed. Kluwer, Dordrecht (2000)

*39.* Wärneryd, K.-E.: The Psychology of Saving: A Study of Economic Psychology. Edward Elgar Publishing, Northampton (1999)

*TABLE 1:*
*Product penetrations over time*

| Product | 1969 | 1990 | 2003 |
|---|---|---|---|
| *Investments* | | | |
| (1) Shares | 0.05 | 0.09 | 0.17 |
| (2) Investment trusts | 0.04 | 0.03 | 0.15 |
| (3) Bonds | 0.12 | 0.06 | 0.08 |
| *Long-term contractual saving* | | | |
| (4) Life Insurance | 0.25 | 0.21 | 0.43 |
| *Regular Saving* | | | |
| (5) Savings account | 0.40 | 0.81 | 0.90 |

*TABLE 2:*
*Model selection in the three steps of the Lukociene et al (2010) procedure:*
*Log-likehood values (LL), number of parameters (Npar), and BIC values*

| Step | Higher level (H) | Lower level (S) | LL | Npar | BIC N=4200 | BIC N=42 |
|---|---|---|---|---|---|---|
| 1 | 1 | 1 | -8773 | 5 | 17588 | |
| 1 | 1 | 2 | -7858 | 16 | 15849 | |
| 1 | 1 | 3 | -7733 | 27 | 15692 | |
| 1 | 1 | 4 | -7678 | 38 | 15672 | |
| 1 | 1 | 5 | -7629 | 49 | 15666 | |
| 1 | 1 | 6 | -7602 | 60 | 15704 | |
| 2 | 1 | 5 | -7629 | 49 | | 15440 |
| 2 | 2 | 5 | -7084 | 54 | | 14369 |
| 2 | 3 | 5 | -6980 | 59 | | 14181 |
| 2 | 4 | 5 | -6926 | 64 | | 14092 |
| 2 | 5 | 5 | -6881 | 69 | | 14021 |
| 2 | 6 | 5 | -6851 | 74 | | 13979 |
| 2 | 7 | 5 | -6831 | 79 | | 13957 |
| 2 | 8 | 5 | -6817 | 84 | | 13948 |
| 2 | 9 | 5 | -6807 | 89 | | 13946 |
| 2 | 10 | 5 | -6798 | 94 | | 13948 |
| 3 | 9 | 4 | -6911 | 70 | 14406 | |
| 3 | 9 | 5 | -6807 | 89 | 14356 | |
| 3 | 9 | 6 | -6777 | 108 | 14455 | |

*TABLE 3: The segmentation solution*

*Part 1: Segment-specific product penetrations, $P(Y_{ijtk}|X_{ijt}=s)$*

| | $s$=1 | $s$=2 | $s$=3 | $s$=4 | $s$=5 |
|---|---|---|---|---|---|
| Segment label | In-actives | Savers | Life-savers | Life-savers with bonds | Actives |
| Cluster size | 0.2526 | 0.4104 | 0.1669 | 0.0726 | 0.0974 |
| Shares | 0.0014 | 0.0315 | 0.0989 | 0.1337 | 0.6547 |
| Invest. trusts | 0.0004 | 0.0127 | 0.0463 | 0.0337 | 0.6530 |
| Bonds | 0.0017 | 0.0150 | 0.0107 | 0.5042 | 0.3703 |
| Life insurance | 0.0278 | 0.0387 | 0.9845 | 0.6718 | 0.6584 |
| Savings account | 0.0144 | 0.9642 | 0.9629 | 0.9196 | 0.9860 |

*Part 2: Definition of the higher-level latent classes, $P(X_{ijt}=s|Z_{it}=h)$*

| Higher-level class number | $s$=1 | $s$=2 | $s$=3 | $s$=4 | $s$=5 |
|---|---|---|---|---|---|
| $h$=1 | 0.7933 | 0.1368 | 0.0437 | 0.0103 | 0.0159 |
| $h$=2 | 0.5965 | 0.1881 | 0.1193 | 0.0744 | 0.0217 |
| $h$=3 | 0.5630 | 0.0007 | 0.0134 | 0.4038 | 0.0190 |
| $h$=4 | 0.3265 | 0.5815 | 0.0310 | 0.0394 | 0.0218 |
| $h$=5 | 0.0884 | 0.6290 | 0.1849 | 0.0517 | 0.0459 |
| $h$=6 | 0.1824 | 0.4507 | 0.0947 | 0.0598 | 0.2124 |
| $h$=7 | 0.0239 | 0.5036 | 0.3040 | 0.0040 | 0.1644 |
| $h$=8 | 0.0792 | 0.3314 | 0.3417 | 0.0763 | 0.1714 |
| $h$=9 | 0.0261 | 0.1216 | 0.1463 | 0.0059 | 0.7001 |

*TABLE 4*
*Covariate effects on respondent level segment membership*

| Segment label | In-actives (*s*=1) | Savers (*s*=2) | Life-savers (*s*=3) | Life-savers with bonds (*s*=4) | Actives (*s*=5) | Sign. |
|---|---|---|---|---|---|---|
| Partner | -0.18 | -0.02 | 0.26 | -0.23 | 0.16 | df=4, Wald=22.02, p<0.01 |
| Income > median | -0.24 | -0.21 | -0.02 | -0.14 | 0.61 | df=4, Wald=68.13, p<0.01 |
| Age < 35 | 0.53 | 0.31 | 0.05 | -0.28 | -0.62 | df=12, Wald=102.55, p<0.01 |
| Age 36-50 | -0.20 | -0.05 | 0.39 | -0.17 | 0.04 | |
| Age 51-64 | -0.29 | -0.16 | 0.06 | 0.03 | 0.36 | |
| Age 65+ | -0.04 | -0.10 | -0.49 | 0.42 | 0.22 | |

*TABLE 5*
*Time-specific allocation of countries to the higher-level segments*

| Country | 1969 | 1990 | 2003 |
|---|---|---|---|
| Austria, Belgium, Luxembourg, Finland, France, Netherlands | 2 | 5 | 7 |
| Portugal, Spain | 1 | 4 | 5 |
| Italy | 1 | 4 | 6 |
| Ireland | 3 | 4 | 8 |
| GB, Denmark | 3 | 5 | 8 |
| Sweden | 3 | 5 | 9 |
| West Germany | 3 | 5 | 7 |