Modeling Individual-Level Change in Marketing Phenomena

Jan-Benedict E.M. Steenkamp
Hans Baumgartner
Frenkel ter Hofstede
Modeling Individual-Level Change in Marketing Phenomena

Abstract

Many important marketing issues deal with the study of change in marketing variables based on an analysis of repeated measurements of entities (consumers, salespeople, companies, brands, etc.) observed at different points in time or at different levels of an independent variable. Traditionally, such data have been analyzed using OLS regression pooled across repeated measurements. We describe and illustrate the potential of three powerful new methodologies for studying individual-level change in marketing phenomena, viz., hierarchical linear modeling, latent curve analysis, and hierarchical linear Bayes. These techniques can be used to simultaneously model (1) the average change trajectory at the population level, (2) individual differences in this average trajectory across entities, and (3) the systematic effect of exogenous variables on individual change processes. The three methods are compared on a number of substantive and statistical modeling criteria. An empirical illustration is provided using data on household purchases of private label cola over a five-year period. None of the three techniques clearly dominates the others, and their relative strength is contingent on the specific research situation and the data. However, the three new methods are preferable to OLS regression in terms of statistical assumptions made and substantive findings obtained.
1. Introduction

Many important marketing issues deal with the study of change in marketing variables based on an analysis of repeated measurements of entities observed at different points in time or at different levels of an independent variable.

Consider the following illustrative examples:

1. Companies routinely monitor sales in product categories over time to assess changes in demand and to anticipate future growth opportunities (Urban and Hauser 1993). Although developments in the market as a whole provide useful information, it would be helpful to know how much variability there is in aggregate indices of change and whether this variability is systematically related to other variables such as characteristics of consumers.

2. An understanding of the dynamics of consumer perceptions of and preferences for brands and stores is of great importance in the development of competitive strategies, particularly in positioning and repositioning decisions (cf. Moore and Winer 1987). Companies are interested in how brand and store images change over time and whether these changes are the same for different segments of consumers.

3. For effective supervision and control of salespeople, sales in a salesperson’s territory should be monitored over time and compared with the performance of other salespeople (Churchill, Ford, and Walker 1990). Furthermore, sales managers would like to know which factors (e.g., personal characteristics, differences in sales territories) account for individual differences in average levels of performance and changes in performance over time.

4. Advertisers are frequently interested in consumers’ moment-to-moment emotional reactions to ads. Recent research has indicated that the trajectory of emotions evoked during exposure to an ad is an important determinant of consumers’ overall reactions to ads (Baumgartner, Sujan, and Padgett 1997). Furthermore, work by Hughes (1991, 1992) suggests that there are systematic individual differences in consumers’ continuous responses to ads. A model in which both of these phenomena are studied simultaneously could provide new insights into the effects of advertising and help advertisers design more effective ads.
5. Measurement and tracking of consumers’ brand awareness and brand associations are essential for effective management of brand equity (Keller 1993). Changes in these measures over time provide feedback on the effectiveness of past marketing strategies and they may signal a need for possible future action (Shocker, Srivastava, and Ruekert 1994).

6. Public policy researchers are interested in the development of consumers’ usage of addictive substances such as cigarettes or drugs from early adolescence to adulthood. The goal is to identify key patterns of change in substance use and to investigate the role of advertising and other marketing influences in this process (cf. Pollay et al. 1996).

The common denominator in these scenarios is that repeated measures on some variable of interest (e.g., sales, preferences, consumption) are available for a sample of entities (consumers, salespeople, companies, etc.) and that the researcher would like to specify a model that describes and explains the change process in question. Such a model should have several desirable characteristics. First, the model should capture both the overall pattern of change over time and individual variation in the basic change pattern. The overall, population-level model provides the general picture of the phenomenon, while individual variation in this pattern can point to interesting segments that may warrant special attention. Second, the model should be able to accommodate a variety of different change processes and allow a flexible specification of the error structure. Finally, the model should be able to incorporate possible predictors of individual differences in change trajectories so that one can gain insight into the reasons for individual variation.

The purpose of this paper is to describe and compare three recently developed methodologies that meet these criteria: hierarchical linear modeling (Bryk and Raudenbush 1992), latent curve analysis (Meredith and Tisak 1990), and hierarchical linear Bayes (Blattberg and George 1991; Rossi and Allenby 1993). These models have a hierarchical structure consisting of two levels. The first level describes the change trajectory at the group level and individual differences in this average trajectory, if any. Analyzing the repeated observations of the same individuals (or other entities) over time makes it possible to study intra-individual change that is uncontaminated by inter-individual differences. A wide range of functional trajectories can be tested, and a parsimonious representation of the phenomenon under investigation is achieved by
concentrating on the parameters of the change trajectory rather than on the original repeated measures. At the second level, the parameters of the change trajectory are the focus of the analysis. Hypotheses about systematic influences on the individual change trajectories can be investigated, and the parameters of the latent change curve can be specified as antecedents of other constructs. In this way, relationships are established between the change process of interest and other variables.

The remainder of this paper is organized as follows. In section 2, we will present a general model formulation, which nests the three modeling approaches listed above. In section 3, we will compare these three techniques with the approach traditionally used to analyze data on individual change - OLS regression applied to individual-level data that are pooled across time periods. The methods are evaluated on a number of substantive and statistical modeling criteria. An empirical illustration is provided in section 4, using household panel data of Dutch consumers’ purchases of private label cola over a five-year period. The systematic impact on differences in individual change profiles is investigated for sociodemographics and consumer attitudes. Finally, conclusions and suggestions for future research are provided in section 5.

2. A General Model Describing and Explaining Change

In this section, we will describe a general model for studying change processes that has the desirable properties listed in the previous section. We assume that the researcher has repeated measurements for a number of entities on some continuous outcome (or set of outcomes) of interest. In a marketing context, entities may refer to consumers, salespeople, stores, companies, strategic business units (SBUs), sales territories, countries and the like. Repeated observations could result from multiple measurements that are explicitly related to time (e.g., sales per year), or they could be based on responses to several levels of another variable, such as recall of advertising copy as a function of different numbers of ad exposures. For ease of exposition, we will assume (without loss of generality) that the term entities refers to individuals and that repeated measurements are a function of time.

Suppose observations are available on some variable \( y_{iit} \), whose change pattern is to be analyzed. The subscripts refer to individual i’s \( i = 1, \ldots, I \) observed score on some characteristic of interest at time t \( t = 1, \ldots, T \). In the application presented below, \( y_{iit} \) represents a household’s
annual consumption of private label cola. We formulate a model that describes the overall pattern of change underlying the repeated observations of individuals in the population and individual change in the basic change pattern. This is done by letting the repeated measurements of individuals follow the same type of functional relation, but allowing individuals to differ in the values of the parameters characterizing the common functional form.

In the current presentation we concentrate on models that are linear in the parameters, where the change trajectories are represented by polynomials. Polynomials are of interest in their own right (e.g., often change processes are assumed to follow a linear, quadratic, or cubic pattern) and they may be used to approximate other models, including models that are nonlinear in the parameters (Burchinal and Applebaum 1991). Ideally, theory will guide the specification of the general functional form of the change trajectory. However, often the marketing mechanisms governing the change process are not well understood and the change function is identified using the data.

We express $y_{it}$ as a weighted sum of basis functions ($\lambda_{tk}$) plus error:

$$y_{it} = \sum_{k=1}^{K} \lambda_{tk} \eta_{ik} + \varepsilon_{it},$$

where the weight $\eta_{ik}$ ($k=1,...,K$) is the unknown change parameter of individual i and $\varepsilon_{it}$ captures the error of measurement. The unknown scores $\eta_{ik}$ ($k=1,...,K$) are individual latent curve parameters which determine the particular shape of the functional form for individual i. Since the change parameters are individual-specific, the formulation allows for individual variation in the change trajectory. A basis function is an attribute of the change trajectory over repeated measurements, and collectively the set of K basis functions defines the general functional form of the change trajectory. For example, if the change trajectory is specified to be quadratic ($K=3$), then $\lambda_{t1}=1$, and $\lambda_{t2}$ and $\lambda_{t3}$ are evaluations of linear and quadratic functions, respectively. In principle, there are two ways to specify $\lambda_{tk}$. One way is to define $\lambda_{tk} = (t-1)^{(k-1)}$. Then $\eta_{i1}$ corresponds to an individual’s expected score at the beginning of the data series, and the resulting model is therefore called the initial-status model. If, on the other hand, $\lambda_{tk}$ is specified using orthogonal polynomial coefficients, $\eta_{i1}$ is equivalent to the error-corrected average of a person’s measurements over time. The latter is denoted as time-averaged coding (Stoolmiller 1995). Ideally, the number of basis functions K should be small so that the repeated measurements can be summarized succinctly with a relatively simple functional relation.
Equation (1) is highly flexible in that it can be extended by adding other effects. Effects that are of particular importance when studying change are structural breaks in the basic change pattern (e.g., due to a major new product introduction; cf. Dekimpe, Hanssens, and Siva-Russo 1999). When a structural break occurs after period \( t^* \), it can be incorporated by adding an additional regressor \( \lambda_{tk} = 1 \) if \( t > t^* \) and \( \lambda_{tk} = 0 \) otherwise. Equation (1) also accommodates piecewise linear and spline change models, where the change curve is represented by a number of connected line segments. Moreover, if a sufficient number of time periods before and after the break is available, one can specify different change processes for the distinct time segments.

Equation (1) can be written in matrix notation by letting \( y_i = (y_{it}) \), \( \eta_i = (\eta_{ik}) \), \( \Lambda = [\lambda_{tk}] \) and \( \varepsilon_i = (\varepsilon_{it}) \):

\[
\begin{align*}
y_i &= \Lambda \eta_i + \varepsilon_i, \\
\end{align*}
\]

where \( \eta_i \) is a \( K \times 1 \) vector of change parameters for individual \( i \) corresponding to a given functional form common to all individuals and \( \Lambda \) is a \( T \times K \) matrix containing the evaluations of the basis functions in different time periods. The error vector \( \varepsilon_i \) is normally distributed with covariance matrix \( \Theta \) (\( T \times T \)):

\[
\varepsilon_i \sim N(0, \Theta).
\]

The covariance matrix \( \Theta \) can be specified in a number of ways. One may assume the residuals to be homoscedastic and distributed independently across time periods, in which case \( \Theta = \theta I_T \), where \( \theta \) is a scalar. A structure with heteroscedastic and uncorrelated errors is obtained by taking \( \Theta = \text{diag}(\theta_1, \ldots, \theta_T) \). Even more flexible error structures are possible by specifying certain off-diagonal entries of \( \Theta \) as unrestricted.

The formulation in Equations (2) and (3) merely describes individual change trajectories. Marketers will usually also be interested in trying to explain differences in individual change trajectories. Because individual characteristics may account for these differences, the change parameters \( \eta_i \) are related to predictor variables that are hypothesized to affect individual change patterns. Let \( \xi_i = (\xi_{ip}) \) be a vector of predictor variables reflecting \( P \) observed characteristics of individual \( i \). Then the vector of individual change parameters \( \eta_i \) can be expressed as a linear combination of the elements in \( \xi_i \) and a parameter matrix \( \Gamma \), an intercept \( \gamma_0 \) and an error term \( \delta_i \):

\[
\eta_i = \gamma_0 + \Gamma \xi_i + \delta_i,
\]
where $\Gamma$ is a $K \times P$ matrix whose entries $\gamma_{kp}$ express the structural effect of variable $p$ on the $k$-th change parameter. The intercept $\gamma_0 = (\gamma_{0k})$ describes the change pattern when all predictor variables are equal to zero ($\xi_i = 0$). If the predictors are mean-centered, then $\gamma_0$ will describe the average change curve across subjects.

The remaining variation in the change parameters that is not explained by the predictors $\xi_i$ is captured by $\delta_i$ and is assumed to be normally distributed with zero mean and covariance $\Psi$:

$$
\delta_i \sim N(0, \Psi).
$$

The diagonal entries $\psi_{kk}$ of $\Psi$ reflect the amount of unexplained variation in the change parameters and the off-diagonal elements $\psi_{kk'}$ reflect their covariances.

Equations (2) to (5) describe a general model that subsumes a number of specific approaches that have been proposed to study change processes: OLS regression, hierarchical linear modeling, latent curve analysis, and the hierarchical linear Bayes model. Below we will show how these approaches are special cases of this general specification.

An approach traditionally used to analyze data on individual change is OLS regression applied to individual-level data pooled across time periods. The pooled regression model is obtained by restricting $\delta_i$ and $\Psi$ in Equations (4) and (5) to zero and by substituting the expression for $\eta_i$ in (4) into Equation (2), which results in the following standard regression equation:

$$
y_i = \Lambda \gamma_0 + (\xi_i' \otimes \Lambda) \cdot \text{vec}(\Gamma) + \epsilon_i.
$$

This specification combines Equations (2) to (5), where the errors of measurement and the unexplained variance in the change parameters are lumped together and are assumed to be homoscedastic and uncorrelated ($\Theta = \theta \mathbf{I}_T$, $\delta_i = 0$). This approach poses conceptual problems because the unit of analysis for testing all effects is the individual observation at time $t$ ($y_{it}$), whereas the unit of analysis for testing hypotheses concerning the effects of individual characteristics should be the individual. In addition, pooled OLS regression presents statistical problems because it ignores the within-subject dependence of the observations, which violates basic OSL assumptions and leads to incorrect estimates of standard errors (Goldstein 1995).

A more efficient approach is hierarchical linear modeling (HLM) developed by Bryk and Raudenbush (1987, 1992). HLM is a statistical method developed for data that typically contain
an inherent hierarchical structure. Change data considered in this paper have such a hierarchical structure: occasions of measurement are nested within individuals. HLM simultaneously estimates Equations (2) to (5) under the assumption of homoscedastic and uncorrelated errors in Equation (3): \( \Theta = \Theta I_T \). In Equation (5) HLM does not impose restrictions on \( \Psi \), the covariance matrix of the change parameters. The well-known random-coefficient regression model is a special case of HLM, viz., when \( P=0 \), and in this case \( \gamma_{k0} \) represents the aggregate change parameters. Unbiased estimates of the model parameters and standard errors are obtained by restricted maximum likelihood (Bryk and Raudenbush 1992).

Latent curve analysis (LCA; Meredith and Tisak 1990) originates from the psychometric literature. A latent curve model is a special case of a confirmatory factor model with mean structures, where Equation (2) describes a measurement model for \( y_{it} \) and Equation (4) is the structural model (Meredith and Tisak 1990). The variables \( \eta_{ik} \) and \( \xi_{ip} \) represent latent factors, the \( \lambda_{tk} \) are factor loadings, and the \( \gamma_{kp} \) are structural parameters. The factor loadings \( \lambda_{tk} \) in Equation (2) are assumed to be completely or at least partially known since they contain the specification of the hypothesized common change trajectory. As in HLM, LCA allows for a full specification of the entries of \( \Psi \) in Equation (5). Moreover, as long as the number of degrees of freedom is sufficient, LCA permits a flexible specification of the structure of \( \Theta \). In general, parameters of the latent curve models are obtained by applying maximum likelihood.

The previous approaches view the free model parameters as fixed quantities that are estimated by the model. In the hierarchical linear Bayes model (HLB; Blattberg and George 1991; Lindley and Smith 1972), these parameters are regarded as stochastic quantities that follow a particular probability distribution. The general HLB model is obtained from Equations (2) to (5) by imposing “hyper prior distributions” on the parameters \( \Gamma \), \( \Theta \), and \( \Psi \). These distributions are usually taken from so-called conjugate families, which facilitates estimation using the Gibbs sampler (Gelfand and Smith, 1990). A flexible model is obtained when the following specification is used: \( (\gamma_0, \Gamma) \sim \text{Normal}(g_0, G_0) \), \( \Theta^{-1} \sim \text{Wishart}(a_0, A_0) \), and \( \Psi^{-1} \sim \text{Wishart}(b_0, B_0) \), where \( g_0 \), \( G_0 \), \( a_0 \), \( A_0 \), \( b_0 \), and \( B_0 \) are specified by the researcher (cf. Lenk et al. 1996). In this specification, estimates of the full error covariance matrix \( \Theta \) can be obtained.
Alternative choices for the distribution of \( \Theta \) are possible as well, describing simpler error structures, such as heteroscedasticity, autocorrelation or a combination of the two.

3. Comparison of Techniques

The pooled OLS regression model specified in Equation (6) is convenient to use, but it is stringent in its modeling assumptions and accommodates only very simple change models. By pooling observations across time-periods, the approach assumes exchangeability of observations across subjects and does not account for the inherently hierarchical structure of the change process. OLS provides unbiased estimates of the structural parameters in \( \Gamma \) if the HLM specification is correct, but the estimates are less efficient and the standard errors will be biased (Raudenbush 1995). Therefore, the estimates of the structural parameters for pooled OLS and HLM are very similar, but OLS will provide incorrect variance estimates. This will result in incorrect inferences about the relations of the curve parameters to their predictor values, among other things (Goldstein 1995). These problems may lead to erroneous conclusions about the basic change pattern and the structural relations between change parameters and predictor variables.

In contrast to the pooled OLS model, HLM, LCA, and HLB model individual differences in patterns of change in a more appropriate and flexible way. HLM, LCA, and HLB are powerful and versatile approaches to modeling individual differences in change in marketing phenomena. Yet, all three techniques have their specific strengths and weaknesses. Below, we will systematically compare these techniques on three “substantive modeling” criteria and five “statistical” criteria.

3.1 Substantive modeling criteria

*Flexibility in estimating different change processes.* All three techniques can accommodate a wide range of models that are linear in the parameters. Moreover, some models that are nonlinear in the parameters, such as the Gompertz curve, can also be estimated (Browne 1993). LCA can also estimate elements of \( \Lambda \), which makes it possible to estimate relative change models, also called linear spline models (Meredith and Tisak 1990). To identify such models, the intercept and two loadings on the slope factor have to be fixed, but the remaining elements can be freely estimated. This allows approximating the change function by a piecewise curve that consists of...
linear segments connecting all successive time periods. Such relative change models cannot be estimated with HLM or HLB since the elements of $\Lambda$ have to be fixed a priori. For HLM and HLB certain coding schemes exist for $\Lambda$ so that the curve consists of a limited number of line segments, but a piecewise curve that connects curves in all successive time periods will not be identified without imposing restrictions on the error variances (cf. Bryk and Raudenbush 1992, pp. 148-149). This is an important strength of LCA as relative change models have proven to be very useful in change research (e.g., Duncan and Duncan 1996).

**Modeling the antecedents and consequences of change-related parameters.** LCA is very flexible in modeling the relations between change-related parameters and their antecedents and consequences. Antecedents and consequences can either be time-invariant (e.g., gender, socioeconomic status) or change over time (attitudes, lifestyle). They can also be fallible or perfectly measured characteristics, latent or observed variables. Given the prevalence of random measurement error in marketing measures, it is especially important to account for measure unreliability since it can bias parameter estimates to a significant degree. While time-invariant covariates can be incorporated into all three approaches, in the current framework HLM and HLB can only accommodate perfectly measured, observed predictor variables, and consequences of curve parameters cannot be specified.

**Modeling multivariate change.** Sometimes, one is interested in studying similarities and differences in processes of change for multiple variables. For example, differences in the process of advertising wear-out for informational and transformational (image) ads can be examined by comparing attitude change trajectories as a function of the number of ad exposures. LCA allows the researcher to model (1) individual change in multiple domains simultaneously, (2) the interrelations between individual change processes in these domains, and (3) the effect of predictor variables on the simultaneous and joint associations of these changes (Duncan and Duncan 1996). Multivariate change models can also be estimated using HLM or HLB by adding an additional layer to the hierarchical structure of equations. However, these models are more stringent in their assumptions about the simultaneous associations. In HLM and HLB the domain-specific change parameters are assumed to follow a normal distribution across domains and the simultaneous change pattern is just represented by their means. LCA is more flexible in
that one can specify separate latent curve models for different domains (e.g., submarkets) and then specify second-order factors of intercept, slope, etc. to account for the covariation among the first-order factors (i.e., curve parameters) (Duncan and Duncan 1996).

3.2 Statistical criteria

*Flexibility of the error covariance structure of the time-varying measures.* HLM is restrictive in this respect. The errors in the time-varying measures are assumed to be homoscedastic and uncorrelated. LCA is much more flexible with respect to the specification of the matrix $\Theta$. The baseline case of homoscedasticity and zero autocorrelation of the errors in $y_{it}$ can be easily relaxed. Three particularly important extensions are heteroscedasticity, autocorrelation among temporally adjacent pairs of error terms, and first-order autoregressive error structures. Other structures are also possible, provided the time series is sufficiently long, although complex error specifications may lead to estimation problems (e.g., negative error variances).

HLB is also very flexible in the specification of the error covariance structure, and it does not yield negative error variances. The full covariance matrix of the error terms can be estimated, which is not identified in LCA. In this way, correlations between errors can be fully accounted for, although one does not gain much insight into the nature of the process. However, HLB is restrictive in allowing specifications of particular forms of autocorrelation. HLB does accommodate autoregressive and heteroscedastic error structures, but in most cases it is very difficult, if not impossible, to impose restrictions on the error variances and covariances. This is in contrast to LCA, where, as long as the number of degrees of freedom is sufficiently large, a wide range of error specifications is possible by imposing equality restrictions on the elements of $\Theta$.

*Distributional assumptions.* All three approaches make assumptions about the distribution of the data and parameters. The restricted maximum likelihood procedure employed by HLM assumes that the time-varying measure $y_{it}$ and change parameters are normally distributed. HLB makes additional assumptions about hyper prior distributions of the model parameters. For both HLM and HLB, specifications of the distribution of the time-varying measure $y_{it}$ have been extended to accommodate distributions other than the normal distribution (e.g., Poisson distributions for count processes). In this extended framework, the measure whose change
pattern is to be analyzed is unobserved (McCullagh and Nelder 1989; Zeger and Karim 1991). The maximum likelihood estimation procedure employed by LCA assumes that the data are distributed according to a multivariate normal distribution. Distribution-free estimation techniques have been developed for LCA, but they require a large number of observations.

In HLB, exact confidence intervals (or credible sets in Bayesian terminology) of the parameter estimates are obtained, which do not rely on asymptotic theory, while the maximum likelihood estimation approach in HLM and LCA assumes asymptotic normality of the parameters. It should be noted, however, that the choice of a hyper prior distribution for the parameters in HLB will become more influential when sample sizes are very small. With smaller sample sizes, the information in the sample will decrease and HLB estimates will rely more on the prior information specified by the researcher. However, when very diffuse and weakly informative prior distributions are used, this effect will be moderate.

Estimation when data are not time-structured. For ease of exposition, it is assumed in this paper that the data are “time-structured.” This means that all individuals are measured at the same points in time and there are no missing data. However, all three techniques can handle missing data and cases where different individuals are measured on different occasions. Non-time structured data pose no particular problem for HLB and HLM. For LCA, the likelihood function cannot be defined in terms of the moments of the observed data. Instead, the model has to be fitted to the raw, individual-level data and to whatever data are available for each individual (Arbuckle 1996). This new estimation procedure is available in the latest version of LISREL (LISREL 8.50) as well as in other programs (e.g., Amos), but relatively little is known about model behavior in relation to missing data patterns. Moreover, when individual observations are obtained at completely different points in time, so that the occasion of measurement becomes a continuous quantity, the missing data procedure is inappropriate.

Sample size requirements. With large samples, the three techniques will provide similar results. Small samples pose no particular problem for HLB as the confidence intervals of its parameters do not rely on asymptotic theory. The statistical properties of the maximum likelihood estimates obtained by HLM and LCA only hold asymptotically. HLB is especially powerful vis-à-
vis the other techniques when the sample size is small relative to the number of parameters estimated.

*Estimation of individual coefficients.* All three techniques allow the estimation of the coefficients of individual change functions, based on the overall solution. Estimation of these coefficients is least attractive for LCA because the individual curve parameters (intercept, slope, etc.) are actually factor scores, which are indeterminate. This problem does not emerge for the other two techniques since no latent factors are involved. With both HLB and HLM one can estimate the coefficients of individual curves as well as the confidence intervals of these coefficients. The techniques increase the reliability of the individual estimates using Bayesian shrinkage estimation. An individual estimate is *shrunk* toward the population estimates, where the amount of shrinkage is inversely related to the reliability of the individual estimate. HLM uses empirical Bayes estimates that use sample variance and parameter variance to determine the amount of shrinkage. HLB provides full Bayes estimates that are more realistic because they take the unreliability of the other parameters into account. HLB has an additional advantage over the other two techniques. It not only allows one to predict values of individual change curves at unobserved points in time (e.g., t=2.5), but it also makes it possible to compute exact confidence bounds of the curves.

Table 1 summarizes the results of the comparison between HLM, LCA, and HLB.

--- Table 1 about here ---

4. Empirical Illustration

4.1 Data

We will use household data for a sample of 708 Dutch consumers concerning their purchases of private label cola during the period 1994-1998 to illustrate and compare HLM, LCA, HLB, and pooled OLS. We only included consumers who made at least one purchase of private label cola in the period considered. We also had data for four sociodemographics, age of the primary shopper in the household (AGE), household size (HHSIZE), socioeconomic status (SES), and region, and two psychographic variables, innovativeness and importance attached to healthy
eating (HEALTH). Socioeconomic status was based on the official classification scheme of the Dutch Association of Market Research Agencies, which distinguishes between 5 categories. Region was coded as a dummy variable with 0 when the person lived in the highly urbanized area in Holland (“The Randstad”) and 1 when he or she lived in other parts of the Netherlands. The primary shopper’s degree of consumer innovativeness was measured by a ten-item, five-point Likert scale (see Steenkamp, ter Hofstede, and Wedel 1999). The summated score ($\alpha = .82$) was used in the analyses. Finally, the importance attached to healthy eating was measured with a five-point scale. The data were provided by a large market research agency in the Netherlands. The purchase data (in liters) were aggregated into yearly intervals so that we had five repeated measures for each household in the panel.

Substantively, private labels are becoming ever more important in the Western world. They have been described as “a formidable force in an already highly competitive grocery environment” (Hoch 1996, p. 89). Due to increased concentration in retailing, retail chains can develop their own brands. Consumers have also become much more positive toward private labels because of consistent quality improvement over the last 10 to 15 years (Bronnenberg and Wathieu 1996; Kapferer 1997).

4.2 Analytical approach

The aggregate data revealed a structural break between 1995 and 1996. Private label purchases increased considerably in 1996, which was due to the introduction of a new, high-quality, relatively high-priced private label called “First Choice” by one of the largest retail organizations in the Netherlands (Superunie). This private label is produced by the Canadian company Cott. In 1997, it got the coveted designation “best value for money” from the Dutch equivalent of Consumer Reports and was rated higher on taste than Coca Cola (Consumentengids 1997). Thus, we added a structural break to our model to capture this effect.

Once we include the structural break, a linear model adequately summarized the data. We estimated the effect of all six predictor variables on the intercept, structural break, and slope. We used HLM version 4 (Bryk, Raudenbush, and Congdon 1996) to estimate the HLM model and LISREL 8 (Jöreskog and Sörbom 1993) to estimate the LCA model. The HLB programs were written by the authors using Gauss (Aptech 1995).
intercept and the slope, and the predictors were mean-centered. The matrix Λ was assumed to be fixed for all methods, and was coded as: \( \lambda_{t1} = 1 \); \( \lambda_{t2} = 1 \) if \( t \geq 3 \), \( \lambda_{t2} = 0 \) if \( t < 3 \), \( \lambda_{t3} = t - 3 \), where \( t = 1, \ldots, 5 \). Except for consumer innovativeness, all predictors were assumed to be measured without error. LCA allows us to take measurement error of consumer innovativeness into account by setting its error variance to \( (1 - \alpha) \) times the variance of innovativeness (Jöreskog and Sörbom 1993). For the other methods, the error variance of consumer innovativeness was constrained to zero.

For pooled regression and HLM we assumed the errors of measurement \( (\varepsilon_{it}) \) to have constant variance and to be uncorrelated over time. In the case of pooled regression the level-2 errors \( (\delta_i) \) were set to zero. For LCA, we chose a heteroscedastic, first-order autocorrelated (tri-diagonal) specification for the errors of measurement (a more extensive error structure is not identified). For HLB, we estimated a model with a full covariance matrix \( \Theta \). In terms of error specification, these models represent the most comprehensive structures that can be estimated by these techniques for these data.4

Initial analyses for both HLM and LCA revealed that, when the predictor variables were included, the residual variation in the structural break was not significantly different from zero. Hence, this error variance was constrained to zero for these two techniques. Constraining non-significant error variances in change parameters to zero also increases the stability of the parameter estimates in these models (e.g., Bryk and Raudenbush 1992).

4.3 Results

Comparison of methods. Table 2 presents the estimates of the structural parameters.5 The similarity of the structural parameter estimates (\( \gamma \)'s) and their standard errors across techniques is examined using the coefficient of alienation and the mean absolute difference (MAD). The coefficient of alienation \( \kappa \) (Borg and Leutner 1985) is a measure of similarity between two vectors of parameter estimates, based on Tucker’s congruence coefficient \( c \) (\( \kappa = (1 - c^2)^{-0.5} \)). Perfect similarity is given by \( \kappa = 0 \) and total lack of similarity is indicated by \( \kappa = 1 \). MAD is the mean of the absolute differences between corresponding parameter estimates. It has a more straightforward interpretation than \( \kappa \) but has no upper bound. The results are reported in Table 3.
As expected, pooled OLS and HLM yield the same estimates of the structural parameters. These estimates are slightly more similar to the LCA estimates than to the HLB estimates. The structural parameters of LCA and HLB are more similar to each other than they are to the other two techniques. This effect is particularly apparent for MAD. When we turn to the estimates of the standard errors, we find that the differences between HLM, LCA, and HLB are modest, when compared to the differences of these three techniques with pooled OLS. Both $\kappa$ and MAD are much higher when the comparison involves pooled OLS.\(^6\)

The relatively large difference in structural parameter estimates between LCA and HLM is mostly due to the difference in error specification assumed by these models (the LCA model with the simpler error structure did not provide an adequate fit to the data and the more complicated error structure cannot be estimated with HLM). When we compare HLM to the LCA model with homoscedastic and uncorrelated errors, we find that $\kappa=.024$ (vs. .190 for the full LCA model) and MAD=.027 (vs. .281). The difference between HLM and the HLB model with homoscedastic and uncorrelated errors was reduced much less ($\kappa=.175$ vs. .218 for the correlated errors model; MAD=.274 vs. .312). Differences in error specification between models also affected the similarity estimates for the standard errors, but the effect was smaller as the standard errors were more alike in the first place. When we compare HLM to the homoscedastic, uncorrelated LCA model, $\kappa=.006$ (vs. .022 for the full model) and MAD=.005 (vs. .082). For the comparison with the homoscedastic, uncorrelated errors HLB model, the results were: $\kappa=.116$ (vs. .056) and MAD=.087 (vs. .129).

On average, the standard errors of the structural parameter estimates were smallest for HLB. The standard errors of the LCA (HLM) parameters were on average 8.5% (20.0%) larger. As expected, the major difference is with pooled OLS. Its standard errors were on average 70.0%, 41.9%, and 58.6% larger than the HLB, HLM, and LCA standard errors, respectively.

A final and important consideration is the identification of significant parameters, as substantive conclusions and interpretation of marketing phenomena are largely, if not exclusively, based on significant parameters. LCA and HLB yield the same set of significant parameters, where we indicate HLB parameters as significant when zero was not contained in the 95% credible sets. HLM identifies the same set of significant parameters but in addition also finds a significant effect of health on the structural break. Pooled OLS yields fewer significant
results, which is due to the larger standard errors of the structural parameters. Next, we turn to the substantive interpretation of the results.

**Substantive findings.** The parameters $\gamma_{10}$, $\gamma_{20}$, and $\gamma_{30}$ specify the expected change curve of consumers with an average score on the predictor variables. The other $\gamma$’s in Table 2 indicate systematic influences of the predictors on the individual change trajectories. We will discuss the results using the significant estimates provided by the HLB model. The substantive conclusions using LCA and (with the exception of one effect) HLM were the same. The estimated average annual private label purchases of cola over the period considered, corrected for the increase in purchases after 1995 due to the structural break, was 7.3 liters ($\gamma_{10}=7.343$). Purchase of private label cola is higher among larger households, which is consistent with Hoch (1996) who reports that larger households are more prone to buy private labels across a set of 14 product categories. The estimated effect is about 2 liters per additional household member ($\gamma_{13}=2.063$). Private label cola purchase is lower among consumers attaching greater importance to healthy eating ($\gamma_{12}=-4.598$), which is consistent with the unhealthy image of cola in the Netherlands (Consumentengids 1997).

The introduction of “First Choice” increased estimated annual private label purchases by about 3.5 liters ($\gamma_{20}=3.518$). The magnitude of this effect compared to the baseline purchase level is in line with the U.K. experience after Sainsbury’s Classic Cola was introduced (Wileman and Jary 1997). The new brand appealed especially to innovative consumers ($\gamma_{21}=.399$), who have a greater tendency to try out new products and brands (Steenkamp, ter Hofstede, and Wedel 1999). Among consumers in the 90th percentile of innovativeness, the introduction of “First Choice” led to an increase of private label purchases by 7.0 liters per household, whereas purchases increased by only 0.2 liter among consumers in the 10th percentile.

Apart from the structural break due to new product introduction, there is no general long-term upward trend in private label purchases ($\gamma_{30}=.186$, n.s.). However, the major high-quality new product introduction had a significant and persistent effect. This supports other findings indicating that quality improvement is a key factor underlying private label success (Dhar and Hoch 1997; Hoch 1996). However, the lack of an aggregate trend masks differential trends among segments in the market, related to SES ($\gamma_{35}=.800$) and consumer innovativeness ($\gamma_{31}=-.180$). Private label purchases increased by 1.7 liters per year among the highest SES
consumers, while they declined by 1.5 liters per year among the lowest SES consumers. Furthermore, there was an increase among less innovative consumers (1.7 liters per year for the 10\textsuperscript{th} percentile) and a negative trend among the innovative consumers (1.4 liters per year for the 90\textsuperscript{th} percentile). Thus, on the one hand innovative consumers are attracted to the new brand, but on the other hand have decreased their consumption of private label colas over time and are more prone to seek new experiences, either with manufacturer brands or with other drinks providing more stimulation (cf. Zuckerman 1994).

As mentioned above, LCA yields the same substantive conclusions. Although LCA takes the error in measuring consumer innovativeness into account, the substantive conclusions remain the same. However, this need not always be the case and depends on the reliability of the predictor variables and their intercorrelations (Bollen 1989). Furthermore, a comparison of the LCA results with and without correction for measurement error in innovativeness revealed that failure to account for measurement error led to an underestimation of the effects of consumer innovativeness by 15 to 20 percent.

In general, HLM yielded the same substantive conclusions as HLB and LCA. However, HLM also identified a significant effect of the importance of healthy eating on the structural break ($\gamma_{22} = 2.517, p < .05$). This indicates that the introduction of First Choice increased cola consumption more for health-conscious people. Since First Choice makes no health claim, the rationale for this effect is not apparent.

Pooled OLS failed to identify two important effects. The structural break is not significant and the effect of SES on the trend is not significant. This is due to the inappropriate assumptions of pooled OLS leading to incorrect standard errors.

--- Tables 2 and 3 about here ---

5. Conclusions

We describe three new techniques that can be used for modeling individual-level change in marketing phenomena, viz. hierarchical linear modeling, latent curve analysis, and hierarchical linear Bayes. These techniques can be used to simultaneously model (1) the average change trajectory at the population level, (2) individual differences in this average trajectory across
entities, and (3) the systematic effect of exogenous variables on individual change processes. These techniques present a major advance over traditional pooled regression analysis.

The three methodologies are discussed and compared on three substantive modeling criteria and five statistical criteria. LCA appears to be especially powerful on the three substantive modeling criteria, dominating both HLB and HLM. On the other hand, LCA performs less well on the five statistical criteria. HLM is more powerful in terms of distributional assumptions, estimation with non-time structured data, and estimation of individual coefficients, and HLB is equal or superior to LCA on all statistical criteria considered.

In our empirical illustration the three methods yielded largely the same conclusions. The standard errors of HLB tended to be somewhat smaller than those of HLM and LCA, which may indicate smaller (posterior) variance of the estimates. However, while the prior distributions were chosen to be very diffuse and weakly informative, it may have slightly reduced the standard errors by passing some information to the parameters. For several key parameters, pooled OLS yielded different conclusions. First, it failed to identify the structural break in the market, caused by a major new product introduction. As such, analyzing these data with pooled OLS would lead to the erroneous conclusion that new product development is not a viable way to increase consumption of private label cola. The results based on the new techniques indicate that this is in fact the major route to success in the cola market for store brands. Second, it did not identify a key growth segment in this market, viz. the higher SES households. On both issues, marketing opportunities such as new product development and targeted marketing efforts would not be identified if the researcher had not used more advanced models. Thus, there is a definite substantive advantage to using the more advanced models. Apart from this, the fact that the estimates of the standard errors are incorrect is an important consideration in its own right.

In sum, our conceptual and empirical comparisons indicate that HLM, LCA, and HLB are preferable to pooled OLS. The decision among the three new techniques is less clear-cut and depends on the research context. When the sample is large, the data are interval-scaled and time-structured, there is no interest in individual coefficients, predictor variables are measured with error and/or are time-variant, and the researcher is interested in modeling relative change models, LCA is clearly preferred. On the other hand, when the sample is relatively small, many parameters are to be estimated, the data are not time structured, attention is focused on individual change coefficients, the predictors are time-invariant, and errors of measurement are correlated, HLB is the
preferred technique. Table 1 can help the researcher in deciding which technique offers the greatest strengths in his/her specific research situation.

One final consideration in choosing one of the techniques is their ease of use. Marketing researchers are likely to be most familiar with LCA software, as the other two techniques are less widely used in marketing and have attracted less attention in doctoral programs so far. LCA programs such as LISREL also offer well-known and extensively researched overall fit indices that allow the researcher to assess whether the model is acceptable. On the other hand, a software package estimating HLM is also commercially available (Bryk, Raudenbush, and Congdon 1996), and BUGS (Best et al. 1996) is a promising and flexible Bayesian program. Software is developing rapidly in this area, accommodate more and more flexible models. Newer versions of software estimating HLM accommodate richer error structures and accommodate models of multivariate change.

Future research could extend the techniques to alleviate their weaknesses. All three techniques are still under development and we have described the current state of the art. Both HLM and HLB would benefit from the development of stand-alone goodness-of-fit test statistics. HLB is intensive in terms of computation time and data storage and involves many decisions, including the choice of prior distributions and convergence diagnostics (Zellner and Min 1995). New user-friendly software could help here, just as the increasingly user-friendly versions of LISREL and EQS have done much to make structural equation modeling more accessible to large groups of (academic) marketing researchers.
REFERENCES


<table>
<thead>
<tr>
<th>Criteria</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Substantive modeling criteria</strong></td>
<td></td>
</tr>
<tr>
<td>Flexibility in estimating different models</td>
<td>LCA &gt; HLB ≈ HLM</td>
</tr>
<tr>
<td>Modeling antecedents and consequences</td>
<td>LCA &gt; HLB ≈ HLM</td>
</tr>
<tr>
<td>Modeling multivariate change</td>
<td>LCA ≥ HLB ≈ HLM</td>
</tr>
<tr>
<td><strong>Statistical criteria</strong></td>
<td></td>
</tr>
<tr>
<td>Flexibility in specifying the error covariance structure</td>
<td>HLB ≈ LCA &gt; HLM</td>
</tr>
<tr>
<td>Distributional assumptions</td>
<td>HLB &gt; HLM ≥ LCA</td>
</tr>
<tr>
<td>Estimation with non time-structured data</td>
<td>HLM ≈ HLB ≥ LCA</td>
</tr>
<tr>
<td>Sample size requirements</td>
<td>HLB &gt; HLM ≈ LCA</td>
</tr>
<tr>
<td>Estimation of individual coefficients</td>
<td>HLB &gt; HLM &gt; LCA</td>
</tr>
</tbody>
</table>

> indicates superior  
≥ indicates probably superior  
≈ indicates no clear superiority
<table>
<thead>
<tr>
<th>Table 2</th>
<th>ESTIMATES OF THE STRUCTURAL PARAMETERS AND THEIR STANDARD ERRORS OF POOLED OLS, HLM, LCA AND HLB MODELS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>POOLED OLS</td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td></td>
</tr>
<tr>
<td>Intercept ((\gamma_{10}))</td>
<td>7.632*</td>
</tr>
<tr>
<td></td>
<td>(1.034)</td>
</tr>
<tr>
<td>Innovativeness ((\gamma_{11}))</td>
<td>-0.024</td>
</tr>
<tr>
<td></td>
<td>(0.165)</td>
</tr>
<tr>
<td>Health ((\gamma_{12}))</td>
<td>-5.044*</td>
</tr>
<tr>
<td></td>
<td>(1.305)</td>
</tr>
<tr>
<td>HHSize ((\gamma_{13}))</td>
<td>2.154*</td>
</tr>
<tr>
<td></td>
<td>(0.825)</td>
</tr>
<tr>
<td>Age ((\gamma_{14}))</td>
<td>-0.045</td>
</tr>
<tr>
<td></td>
<td>(0.069)</td>
</tr>
<tr>
<td>SES ((\gamma_{15}))</td>
<td>0.656</td>
</tr>
<tr>
<td></td>
<td>(1.051)</td>
</tr>
<tr>
<td>Region ((\gamma_{16}))</td>
<td>0.970</td>
</tr>
<tr>
<td></td>
<td>(3.488)</td>
</tr>
<tr>
<td><strong>Structural break</strong></td>
<td></td>
</tr>
<tr>
<td>Intercept ((\gamma_{20}))</td>
<td>2.887</td>
</tr>
<tr>
<td></td>
<td>(1.595)</td>
</tr>
<tr>
<td>Innovativeness ((\gamma_{21}))</td>
<td>0.511*</td>
</tr>
<tr>
<td></td>
<td>(0.254)</td>
</tr>
<tr>
<td>Health ((\gamma_{22}))</td>
<td>2.517</td>
</tr>
<tr>
<td></td>
<td>(2.013)</td>
</tr>
<tr>
<td>HHSize ((\gamma_{23}))</td>
<td>0.683</td>
</tr>
<tr>
<td></td>
<td>(1.272)</td>
</tr>
<tr>
<td>Age ((\gamma_{24}))</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>(0.106)</td>
</tr>
<tr>
<td>SES ((\gamma_{25}))</td>
<td>-1.537</td>
</tr>
<tr>
<td></td>
<td>(1.621)</td>
</tr>
<tr>
<td>Region ((\gamma_{26}))</td>
<td>0.313</td>
</tr>
<tr>
<td></td>
<td>(5.382)</td>
</tr>
<tr>
<td><strong>Trend</strong></td>
<td></td>
</tr>
<tr>
<td>Intercept ((\gamma_{30}))</td>
<td>0.359</td>
</tr>
<tr>
<td></td>
<td>(0.553)</td>
</tr>
<tr>
<td>Innovativeness ((\gamma_{31}))</td>
<td>-0.212*</td>
</tr>
<tr>
<td></td>
<td>(0.088)</td>
</tr>
<tr>
<td>Health ((\gamma_{32}))</td>
<td>-0.851</td>
</tr>
<tr>
<td></td>
<td>(0.697)</td>
</tr>
<tr>
<td>HHSize ((\gamma_{33}))</td>
<td>-0.008</td>
</tr>
<tr>
<td></td>
<td>(0.441)</td>
</tr>
<tr>
<td>Age ((\gamma_{34}))</td>
<td>-0.029</td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
</tr>
<tr>
<td>SES ((\gamma_{35}))</td>
<td>0.901</td>
</tr>
<tr>
<td></td>
<td>(0.562)</td>
</tr>
<tr>
<td>Region ($\gamma_{36}$)</td>
<td>-0.383</td>
</tr>
<tr>
<td>------------------------</td>
<td>--------</td>
</tr>
<tr>
<td></td>
<td>(1.864)</td>
</tr>
</tbody>
</table>
Table 3
OVERVIEW OF THE RESULTS OF THE EMPIRICAL COMPARISON BETWEEN POOLED OLS, HLM, LCA AND HLB

<table>
<thead>
<tr>
<th>Coefficient of Alienation (κ)</th>
<th>Mean Absolute Difference (MAD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pooled OLS</td>
</tr>
<tr>
<td>Structural Parameters</td>
<td></td>
</tr>
<tr>
<td>HLM</td>
<td>.000</td>
</tr>
<tr>
<td>LCA</td>
<td>.190</td>
</tr>
<tr>
<td>HLB</td>
<td>.218</td>
</tr>
<tr>
<td>Standard Errors of Structural Parameters</td>
<td></td>
</tr>
<tr>
<td>HLM</td>
<td>.165</td>
</tr>
<tr>
<td>LCA</td>
<td>.145</td>
</tr>
<tr>
<td>HLB</td>
<td>.169</td>
</tr>
</tbody>
</table>
**FOOTNOTES**

1) Interestingly, a similar thing happened in 1994 in the U.K where Sainbury's Classic Cola (also produced by Cott) caused a large increase in the consumption of private label cola while the share of Coca Cola declined substantially (Kapferer 1997).

2) Following common practice (e.g., Willett and Sayer 1994) and to keep the analysis manageable, we recommend to first fit Equations (2) and (3) to identify the appropriate model of change before adding predictors and estimating the full model in Equations (2)-(5).

3) Diffuse, weakly informative prior distributions were used in order for the prior distributions not to influence the HLB estimates. The prior distributions of a basic HLB model with homoscedastic and uncorrelated errors ($\Theta = \Theta I_3$) were specified as follows: $\gamma_{kp} \sim \mathcal{N}(0,10^6)$, $\theta^{-1} \sim\text{Gamma}(10^{-6},10^{-6})$, and $\Psi^{-1} \sim \text{Wishart}(I_3,1)$. For a fully specified model with heteroscedastic and autocorrelated errors, the prior distributions were: $\gamma_{kp} \sim \mathcal{N}(0,10^6)$, $\Theta^{-1} \sim \text{Wishart}(I_5,1)$, and $\Psi^{-1} \sim \text{Wishart}(I_3,1)$. For this specification, and given the reasonably large sample size, the influence of the prior distribution on the HLB estimates is limited. For each model run, the Gibbs sampler is applied for 15,000 iterations. The sampled parameter values in the first 5,000 iterations were discarded and the remaining values were used to compute the sampling quantities of interest (i.e., the means and standard deviations).

4) The fit indices indicated that the heteroscedastic LCA/HLB model with correlated errors achieved a significantly and substantially better fit than the simple homoscedastic model without correlated errors. For example, the fit indices for the full LCA model were: $\chi^2(17)=37.76$ (p=.003), CFI=.989, TLI=.965, RMSEA=.042. In contrast, the corresponding fit indices for the homoscedastic, uncorrelated errors model were: $\chi^2(25)=356.41$ (p<.001), CFI=.827, TLI=.619, RMSEA=.137.

5) Due to space constraints, the estimates for the error covariance matrices $\Theta$ and $\Psi$ are not reported. A complete overview can be obtained by writing to the authors.

6) One should not compare $\kappa$ and MAD between structural parameters and standard errors. MAD is scale dependent and standard errors were typically smaller than parameter estimates. $\kappa$ is lower for standard errors but this is due to the fact that the ratio of between-method to within-method variation of standard errors is rather large as compared to those of estimates.
This is inherent to computing a summary measure across a large number of different coefficients.

7) Note that cola consumption in the Netherlands is much lower than in the USA.