

The Pathmox approach for PLS path modeling segmentation

Abstract

Modeling has very often failed to live up to expectations, mostly because of the difficulty of comprehending relationships within phenomena and expressing them in mathematical models. Reality is frequently too complex to be reflected in a single model. This is often the case of marketing research, where variables relating to socioeconomics, psychographics or consumption constitute potential sources of heterogeneity. In such cases, the assumption of “one model fits all” is unrealistic and may lead to inaccurate decisions. Thus, heterogeneity is a major issue in modeling. Once a model has been fitted to a complete data set that fulfills all validation criteria, it is difficult to establish whether it is valid for the whole population or it is merely an average artifact from several sub-populations. The purpose of this paper is to present the Pathmox approach to deal with heterogeneity in Partial Least Squares Path Modeling (PLS-PM). The idea behind Pathmox is to build a tree of path models that have look-alike structure as a binary decision tree, with different models for each of its nodes. The split criterion consists of an F statistic comparing two structural models, which test the identity of two regression models. In order to ensure the suitability of the split criterion, a simulation study was conducted. Finally, we have applied Pathmox to a survey that measured *Satisfaction* among Spanish mobile phone operators. Results suggest that the Pathmox approach performs adequately in detecting PLS-PM heterogeneity.

Keywords: Heterogeneity, Partial Least Squares Path Modeling, Segmentation, Model Comparison, Segmentation Trees.

1 Introduction

In survey research, it is very often taken for granted that a single model will represent the whole population that is, homogeneity is presumed for all the individuals. However, it is sometimes reasonable to cast doubt on such supposed homogeneity, and it is even logical to assume that groups have different behaviors. This is especially true of market research studies, where different customers/users may have different needs ([1],[2], [3] and [4]). If we do not consider the possible existence of segments in our population, the one-size-fits-all solution might be inappropriate and may generate inaccurate and/or inadequate results. This is particularly the case of Structural Equation Modeling (SEM) studies, where variables of socio-demographics, psychographics or consumption may stem from potential sources of heterogeneity. As [3] pointed out SEM has become a quasi-standard in marketing research. Its main advantages are: it allows to model latent concepts together with observed variables; researchers can test their current knowledge of phenomena using information gathered from the data; and they can validate theories, establish the main drivers of a phenomenon, measure intangibles, and advance new hypotheses. Regarding the problem of customer satisfaction, Partial Least Squares Path Modeling (PLS-PM) has become a reference methodology ([5], [6], [7], [8], [9], [10], [11], [12] and [13]). PLS-PM is based on an iterative algorithm whose objective is to approximate the relationships among the latent constructs and to estimate them as a linear composite of indicators. Its main advantages are: its broad applicability conditions, the lack of any requirement about the probabilistic nature of the data, the clear quantification of the latent constructs and the easy inclusion of formative constructs in the model, making it very suitable for marketing research.

In this paper, we describe a tree methodology to reveal the heterogeneity existing within data. We call it the Pathmox approach. Even though presented within the framework of PLS-PM, its principles can be easily generalized to any kind of modeling. In section 2, we present the case study used to illustrate and justify the proposed approach. In section 3, we introduce the concept of heterogeneity and the solutions for dealing with it. In section 4, we present the Pathmox segmentation tree, with a detailed discussion of the problem concerning the comparison of two PLS path models. In section 5, we show the undertaken simulation with artificial data in order to evaluate the sensitivity of the F statistic split criterion under a set of experimental but realistic conditions. And, in section 6, we present the results obtained from applying the Pathmox approach to the case study. Finally, there are the conclusions on the suitability of the proposed approach (sections 7 and 8).

2 The problem and the data

To illustrate the problem of heterogeneity, we use an example based on the typical customer satisfaction model used in the mobile phone operator sector ([12],[14], and [15]). The mobile phone sector is extremely competitive, therefore knowing the drivers of *Satisfaction* and *Loyalty* is of paramount importance. However, since the mobile phone market is broad and diverse, we may wonder if we need more than one model to accurately represent *Satisfaction* and *Loyalty*.

The objective is to elucidate the presence of heterogeneity and establish its potential causes. In particular, it is our aim to investigate whether there are any differences between the three main competitors mobile operators, that are A, B and C. These represent the cheapest, the most prevalent and the most elite operators, respectively. Here, we present the results obtained from a 2006 pilot study on the Spanish cell phone industry. For this approach, we need to differentiate between two sets of variables: one made up of the manifest variables that will define the PLS-PM model, which we will call *modeling variables*; and the other, representing potential causes of heterogeneity, which we will call *segmentation variables*.

In our case study we use the “csimobile” data on R package [16] which includes 26 manifest variables measured on an 11-point ordinal scale ranging from *very satisfied* (10) to *very dissatisfied* (0) (see Table 1). We also have 8 segmentation variables that will serve the purpose of identifying the source of heterogeneity (see Table 2).

Following the classical Customer Satisfaction Index presented by [14], the model (see Figure 1) includes seven constructs: *Image*, *Expectations*, *Perceived Quality*, *Perceived Value*, *Satisfaction*, *Complaints*, and *Loyalty*. A description of the latent variables is offered in Table 3.

The question is whether the drivers of *Satisfaction* and *Loyalty* are the same for the whole population or if they are different for some groups. In other words, we want to find out if there is a single model for the whole population, or there are different segments, each one with its own specific model.

3 The concept of heterogeneity

According to its classical definition, heterogeneity refers to mixtures of populations that form differentiated clusters. We refer to this situation as *population heterogeneity*. A different situation is *model heterogeneity*, whose objective is to find segments following different models (in our case, PLS path models). Such distinction may seem obscure at first sight, but, in effect, it is not. The cluster analysis does not take into account the hypothesized structural relationships among variables, whereas modeling segmentation does consider the assumed structural relations.

According to the a priori knowledge about heterogeneity, we can distinguish between two situations in model heterogeneity:

1. *Known heterogeneity*: the number and definition of segments is known a priori; for example, we know that heterogeneity is defined by age and education, and segments are then defined by the crossing of these two variables;
2. *Unknown heterogeneity*: neither the number of segments nor their composition are known; in such a case, we can even distinguish between situations where we have observed covariates as segmentation variables from situations without them:
 - (a) *Unknown segments without segmentation variables*: in this case, segmentation is based only on the information provided by the modeling variables. It is assumed that the population is a mixture of several subpopulations, each one characterized by its own model and mixed in different proportions; then, iterative algorithms, such as latent class analysis or adaptive clustering algorithm, allow us to estimate the probabilities of membership for every individual and to obtain class-specific estimates of every model. This approach needs to identify the existing number of segments. Proposals under this option include the Finite Mixture Partial Least Squares (FIMIX-PLS) approach ([17], [18] and [19]), PLS Typological Path Modeling ([20] and [21]) and the Response Based Procedure for detecting Unit Segments (REBUS) in PLS-PM ([13], [22] and [23]);
 - (b) *Unknown segments with segmentation variables*: this is the case in many survey-based studies which include variables such as sociodemographic variables (e.g., age, gender, family size, occupation or education), geographic variables (e.g., country, state, neighborhood, size of metropolitan area or climate), psychographic variables (e.g., lifestyle, values or personality) and behavioral variables (e.g., usage rate, seeking profits, readiness to buy, and user status), that can be used in the process of segment identification ([16], [24], [25], [26], [27] and [28]).

Regarding the type of heterogeneity, path models may differ in various aspects:

1. They may differ at the *topological level*, that is, in the assumed network linking the latent variables;
2. They may also differ at the *structural level*, i.e., changes may occur in the influences of the construct effects; for example, the teenage group may consider Image to be more important for Satisfaction than the adult group does;

3. They may be distinct at the *measurement level*, that is, in how the constructs are defined by their indicators (factor invariance problem).

In this paper, we will address only the second problem, that is, to determine the importance of the effects among constructs in the different detected segments. We also assume that heterogeneity is assignable to the so called segmentation variables.

3.1 Models comparison: dealing with heterogeneity at structural level

There are several ways of dealing with assignable heterogeneity: by introducing moderating variables into the model, by multi-group approach, or by using a global comparison test.

1. **Moderating variables** Heterogeneity is usually tested by including the segmentation variables as terms in the model (called moderating variables) ([29] and [30]). These represent the heterogeneity we want to test:

$$y = f\left(\beta_0 + \beta'_1 x + \beta_{0U} U + \beta'_{1U} Ux\right) + \varepsilon \quad (1)$$

where x represents the vector of the model's explanatory variables, and U the moderating (or interaction) variable we want to test. Heterogeneity is assessed by means of a statistical test on the coefficients of the interaction terms ($\beta_{0U} = 0$ and $\beta'_{1U} = 0$). In practice, this approach has limited possibilities for exploration; it implies that the source(s) of heterogeneity must be known beforehand, with a small number (normally one) of moderating variables (segments are defined by the levels of the moderating variable).

2. **Multi-group approach** Another way of dealing with heterogeneity is to fit the model in all groups, and then to assess the similarity of the obtained coefficients ([31], [32], [33], [34] and [35]):

$$y_A = f\left(\beta_{0A} + \beta'_{1A} x_A\right) + \varepsilon_A \quad (2)$$

$$y_B = f\left(\beta_{0B} + \beta'_{1B} x_B\right) + \varepsilon_B \quad (3)$$

where A and B represent the two segments whose homogeneity we want to test. Here, heterogeneity is assessed variable by variable, establishing whether or not the difference in coefficients is significant ($\beta_{0A} = \beta_{0B}$ and $\beta'_{1A} = \beta'_{1B}$). As stated before, there are limited possibilities for exploration: the source(s) of heterogeneity must be known in advance and we can only deal with a limited number of segments (normally two).

3. **Global comparison test** We test the identity of two models. One represents the homogeneity situation, and the other represents the heterogeneity we want to test:

$$H_0: \begin{bmatrix} y_A \\ y_B \end{bmatrix} = f \left[\begin{pmatrix} 1 & x_A \\ 1 & x_B \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \right] + \begin{pmatrix} \varepsilon_A \\ \varepsilon_B \end{pmatrix} \quad (4)$$

$$H_1: \begin{bmatrix} y_A \\ y_B \end{bmatrix} = f \left[\begin{pmatrix} 1 & x_A & 0 & 0 \\ 0 & 0 & 1 & x_B \end{pmatrix} \begin{pmatrix} \beta_{0A} \\ \beta_{1A} \\ \beta_{0B} \\ \beta_{1B} \end{pmatrix} \right] + \begin{pmatrix} \varepsilon_A \\ \varepsilon_B \end{pmatrix} \quad (5)$$

where the null hypothesis (H_0) represents the homogeneity, and the alternative hypothesis (H_1) represents the heterogeneity that we want to test. Heterogeneity is then assessed by looking at the global statistic of the model comparison (i.e., in regression, it involves comparing the sum of squares of residuals for both models: $SSR_0 = SSR_1$).

4 Path modeling segmentation trees: the Pathmox approach

Path modeling segmentation can be defined as the process through which a set of elements is divided into distinct subsets, the segments, which have their own structural model.

Starting with the PLS-PM model that is fitted to the entire data, the algorithm detects the optimal split between the two groups whose path models are as different as possible. We will call these two groups child nodes, while the global PLS defines the parent node.

We use the segmentation variables to find the optimal split by calculating all possible splits in which the parent node can be divided into two segments. As a split criterion, we make a global comparison of two models, one corresponding to the parent node (the homogeneous model) and the other corresponding to the child nodes (the heterogeneous model).

The final outcome consists of a binary tree of models, where every branch of the tree identifies a segment of the population with a specific PLS-PM model in the terminal node of that branch.

We call it a Pathmox tree (“mox” stands for “divide into two” in the Nahuatl (Aztec) language), and it constitutes the standard for PLS Path Modeling Segmentation.

4.1 Model comparison in Pathmox

Comparison of models is a crucial step in the Pathmox approach. Let us consider a path model (see Figure 2) with two endogenous latent variables, η_1 and η_2 . Its generalization into more complex models is straightforward. The structural equations for both endogenous constructs are:

$$\eta_1 = \beta_1 \xi_1 + \beta_2 \xi_2 + \zeta_1 \quad (6)$$

$$\eta_2 = \beta_3 \xi_1 + \beta_4 \xi_2 + \beta_5 \eta_1 + \zeta_2 \quad (7)$$

The disturbance terms ζ_1 and ζ_2 are assumed to be normally distributed with zero mean and finite variance; that is, $E(\zeta_1) = E(\zeta_2) = 0$ and $Var(\zeta_1) = Var(\zeta_2) = \sigma^2 I$. It is also assumed that

$$\text{Cov}(\zeta_1, \zeta_2) = 0.$$

We can define the following matrices:

$$X_1 = [\xi_1, \xi_2] \quad \text{a column matrix with the explicative latent variables of } \eta_1 \quad (8)$$

$$B_1 = [\beta_1, \beta_2] \quad \text{a vector of path coefficients for the regression of } \eta_1 \quad (9)$$

$$X_2 = [\xi_1, \xi_2, \eta_1] \quad \text{a column matrix with the explicative latent variables of } \eta_2 \quad (10)$$

$$B_2 = [\beta_3, \beta_4, \beta_5] \quad \text{a vector of path coefficients for the regression of } \eta_2 \quad (11)$$

The structural equations in matrix form are expressed as:

$$\eta_1 = X_1 B_1 + \zeta_1 \quad (12)$$

$$\eta_2 = X_2 B_2 + \zeta_2 \quad (13)$$

We assume that the parent node is divided into two child nodes or segments. In each segment, we compute its own structural model:

$$\text{Segment}^A : \eta_1^A = X_1^A B_1^A + \zeta_1^A \quad \text{and} \quad \eta_2^A = X_2^A B_2^A + \zeta_2^A \quad (14)$$

$$\text{Segment}^B : \eta_1^B = X_1^B B_1^B + \zeta_1^B \quad \text{and} \quad \eta_2^B = X_2^B B_2^B + \zeta_2^B \quad (15)$$

with $\zeta_1^A \sim N(0, \sigma^2 I)$ and $\zeta_2^A \sim N(0, \sigma^2 I)$, $\zeta_1^B \sim N(0, \sigma^2 I)$ and $\zeta_2^B \sim N(0, \sigma^2 I)$.

Now, we want to investigate whether the models of the two segments A and B are different. Thus, the test hypothesis will be: *The coefficients of segments A and B are equal.* We can define a null hypothesis and an alternative hypothesis for the structural model as:

$$H_0 : \begin{bmatrix} \eta_1^A \\ \eta_2^A \\ \cdot \\ \eta_1^B \\ \eta_2^B \end{bmatrix}_{[2n,1]} = \begin{bmatrix} X_1^A & 0 \\ 0 & X_2^A \\ \cdot & \cdot \\ X_1^B & 0 \\ 0 & X_2^B \end{bmatrix}_{[2n, p_1+p_2]} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}_{[p_1+p_2, 1]} + \begin{bmatrix} \zeta_1^A \\ \zeta_2^A \\ \cdot \\ \zeta_1^B \\ \zeta_2^B \end{bmatrix}_{[2n,1]} \quad (16)$$

$$H_1 : \begin{bmatrix} \eta_1^A \\ \eta_2^A \\ \cdot \\ \eta_1^B \\ \eta_2^B \end{bmatrix}_{[2n,1]} = \begin{bmatrix} X_1^A & 0 & 0 & 0 \\ 0 & X_2^A & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & X_1^B & 0 \\ 0 & 0 & 0 & X_2^B \end{bmatrix}_{[2n, 2p_1+2p_2]} \begin{bmatrix} \beta_1^A \\ \beta_1^B \\ \cdot \\ \beta_2^A \\ \beta_2^B \end{bmatrix}_{[2p_1+2p_2, 1]} + \begin{bmatrix} \zeta_1^A \\ \zeta_2^A \\ \cdot \\ \zeta_1^B \\ \zeta_2^B \end{bmatrix}_{[2n,1]} \quad (17)$$

Then, assuming that the random perturbations associated with the latent variables are uncorrelated with equal variance, we can apply Lemmas 1 and 2 of [36]. Hence, the F statistic measuring the discrepancy between the two models:

$$F_{Global} = \frac{(SS_{H_0} - SS_{H_1}) / (p_1 + p_2)}{SS_{H_1} / [2n - 2(p_1 + p_2)]} \quad (18)$$

where SS_{H_0} and SS_{H_1} stand for the corresponding sum of squares of residuals in both models. It then follows, under the null hypothesis, that we have an F distribution with $p_1 + p_2$ and $2n - 2(p_1 + p_2)$ degrees of freedom.

4.2 Pathmox algorithm description

The algorithm of Pathmox starts with the estimation of the global PLS Path Model at the root node. Then, using the segmentation variables, all possible binary splits of data are produced, and for each partition local models are calculated. Among all the splits, the best one is selected by means of the F -test comparing the inner models. This process is recursively applied for each child node. The stop criterion is based on the significance level of the p-value associated with the F statistic. Additionally, two stop parameters are also considered: (1) the number of individuals in a node and (2) the growing level of the depth of the tree.

Algorithm 1 PATHMOX Algorithm

Step 1. Start with the global PLS path model at the root node

Step 2. Establish a set of admissible partitions for each segmentation variable in each node of the tree

Step 3. Detect the best partition by:

- 3.1. Forming all binary partitions in all segmentation variables
- 3.2. Applying the F -test, calculating for each partition a p-value
- 3.3. Sorting the p-values in a descending order
- 3.4. Choosing as the best partition the one associated to the lowest p-value

Step 4. *If* (stop criteria¹ = **false**) *then*
repeat **step 3**

1. Possible stop criteria:

- a. The p-values F -test are not significant
 - b. The number of individuals in the group falls below a fixed level
 - c. The maximum level of the tree depth is attained
-

The Pathmox algorithm has been implemented as in [16]. The output of Pathmox is a binary tree, where each split is defined in such a way that the PLS-PM inner models of the child nodes maximize the differences between them.

4.3 Validation of segments

Once a Pathmox tree has been constructed and the final nodes have been obtained, it is necessary to validate the different models obtained in the terminal nodes. We propose to do so by using a distribution-free procedure, in order to skip the conditioning on probabilistic assumptions. This is a natural way to conduct hypothesis testing in the PLS-PM context; following [9], it involves performing a bootstrapping procedure to validate the results of each terminal segment. The bootstrap samples are built by performing a re-sampling operation that uses replacements from the original sample. The samples consist of the same number of units as in the original sample, and the number of re-samples is defined by the user. Hence, bootstrap confidence intervals of the path coefficients can be obtained from the re-sampling procedure, and the confidence intervals allow to detect the significant coefficients for each segment.

5 Simulation with artificial data

A simulation study was conducted in order to obtain some insight into the behavior of the split criterion under different experimental conditions. The idea was to check the ability of the F statistic to detect significant differences between the models, controlling the conditions of the

split. We carried out a Monte Carlo simulation study with a simple path model. We used two exogenous latent variables (ξ_1 and ξ_2) and one endogenous latent variable (η) (see Figure 3). We varied the data generating conditions according the following factors: the skewness of the distributions of the manifest variables, the sample size, the path coefficients for each segment, the disturbance term for the endogenous construct and the measurement error for the indicators. In this way, we tried to reproduce similar conditions to those encountered in real-life applications of customer satisfaction studies.

5.1 Simulated model

The data of the Model, shown in Figure 3, were generated by following a two-step procedure [40]: first, we generated the latent variables by following the relationship specified in the structural model. The inner structure is defined as:

$$\eta = \beta_1 \xi_1 + \beta_2 \xi_2 + \zeta_\eta \quad (19)$$

where β are regression coefficients and ζ is the error term associated with the endogenous latent variables; second, we define the measurement models for ξ_1 , ξ_2 and η reflective:

$$\begin{aligned} x_{11} &= \lambda_{x_{11}} \xi_1 + \varepsilon_{x_{11}} & x_{21} &= \lambda_{x_{21}} \xi_2 + \varepsilon_{x_{21}} & x_{31} &= \lambda_{x_{31}} \eta + \varepsilon_{y_{11}} \\ x_{12} &= \lambda_{x_{12}} \xi_1 + \varepsilon_{x_{12}} & x_{22} &= \lambda_{x_{22}} \xi_2 + \varepsilon_{x_{22}} & x_{32} &= \lambda_{x_{32}} \eta + \varepsilon_{y_{12}} \\ x_{13} &= \lambda_{x_{13}} \xi_1 + \varepsilon_{x_{13}} & x_{23} &= \lambda_{x_{23}} \xi_2 + \varepsilon_{x_{23}} & x_{33} &= \lambda_{x_{33}} \eta + \varepsilon_{x_{13}} \end{aligned} \quad (20)$$

where x denotes the manifest variables, the λ terms are coefficients, and the ε terms are random errors.

5.2 Simulation design

The levels of all factors of the experimental design have been selected according to the following conditions:

1. The exogenous constructs (ξ_1) and (ξ_2) have been realized from a Beta distribution $B \sim (u, v)$ in order to take into account both symmetry and skewness in distributions for the latent variables. Three cases for the beta distribution are considered: 1) B(6,6) symmetric case; 2) B(9,4) moderately right-skewed case; and 3) B(9,1) right-skewed case. In Figure 4 we give a visual display of one instance of such distributions.
2. The disturbance term ζ for the endogenous variable η follows a Uniform distribution $U \sim (a, b)$ with expectation zero and variance with three levels. The levels are chosen such that the variance of ζ accounts for 10%, 30%, and 50% of the total variance of η .

3. We have selected fixed unequal loading coefficients (0.75, 0.80 and 0.85) in order to have similar values to those found in real applications.
4. The error terms ε 's for the manifest variables are also uniformly distributed with expectation zero and variance with three levels (accounting for 10%, 30%, and 50% of the total variance of x_{ji}).
5. We consider four sample sizes as the total number of cases: {100, 200, 500 and 1000}, which implies segments of sizes {50, 100, 250 and 500}.
6. Finally, the path coefficients for each pair of segments (A and B) are determined in the following way. Segment A is kept fixed with values for its path coefficients $\beta_1^A = \beta_2^A = 0.5$. Segment B will vary its path coefficients, starting with identical values to the first segment (i.e., $\beta_1^B = \beta_2^B = 0.5$), and increasing the coefficient β_1^B by 0.05 at each step; conversely, decreasing β_2^B by 0.05 at each step. The final values of the path coefficients in segment B are $\beta_1^B = 0.90$ and $\beta_2^B = 0.10$.

Thus, we have $3 \times 3 \times 3 \times 4 = 108$ scenarios in total, which is the number of possible combinations of sample sizes, Beta distributions, and noise levels. For each possible scenario, we compared nine pairs of inner models that correspond to the nine sets of path coefficients. For each set of path coefficients, we ran 100 repetitions in which we generated two segments or child nodes. In each repetition we computed the corresponding PLS models for each node and applied the F -test to compare the coefficients of both nodes by calculating the associated p-value. We also calculated the mean squared Euclidean distance between the path coefficients of Node A and Node B ($= \frac{1}{r} \sum_{k=1}^r (\beta_{1k} - \beta_{2k})^2$, where r is the total number of path coefficients to have a closeness value of inner models between child nodes A and B). From the 100 repetitions, we calculated the average of the F -tests p-value, and the mean value of the distance of path coefficients between models. For the computational aspects, we employed the statistical software R, using the functions “rbeta” and “runif” to generate the required distributions.

5.3 Simulation results

To assess the influence of each experimental condition on the split criterion, we focused on the evolution of the p-value computed from the F statistic by examining how it is affected by the different levels of the experimental factor. For ease of interpretation, we included in each plot the Lowess regression line¹ of the p-values with respect to the evaluated experimental condition.

Figure 5, shows the influence of the skewness of the distributions of manifest variables upon the p-values of the F statistic split criterion. We can observe that no difference appears, regardless of whether the distributions are symmetrical (normal-like), fairly skewed or clearly skewed.

In the plot of Figure 6, we can observe the influence of the difference between the inner PLS-PM models of nodes A and B . The more distant the path coefficients, the more significant

¹The Lowess regression is computed using the “Lowess” function of the R software with its default implemented parameters.

the p-values of the F statistic criterion become. The p-values clearly become significant when increasing the difference between the path coefficients of child nodes, as expected.

In Figure 7 the influence of the sample size on the significance of results can be observed, and it is found, as expected, that p-values decrease (i.e., they become more significant) as the sample size increases.

Finally, Figure 8 shows the influence of the level of noise in the endogenous latent variable and in the disturbance terms of manifest variables. In both cases, the plot looks similar: the greater the error, the less significant the F statistic. In all cases, we have represented the Lowess curves corresponding to the three situations of skewness (B(6,6), B(9,4) and B(9,1)) in order to emphasize the non-difference among the different distributions of data.

6 Application

We have used the Spanish mobile phone data as a benchmark for the Pathmox approach. The analysis has been performed by using the R packages “plspm” and “pathmox” [16] and [39]. The inner model, specified in Figure 1, has been designed to measure the cause-effect relationships of *Satisfaction* and *Loyalty* and their predictors. Both constructs are the two main endogenous intangibles that company management wants to explain and monitor. The antecedents of *Satisfaction* are *Image*, *Perceived Quality*, and *Perceived Value*, while *Loyalty* is driven by *Image*, *Satisfaction* and *Complaints*².

The measurement model of the latent variables is taken in a reflective way in all cases except the *Perceived Quality* block, since it is indeed of a formative nature and represents a composition of different facets of quality. Since all manifest variables use the same scale, no standardization is used. This allows us to measure the latent concepts on the same original scale as the manifest variables.

6.1 PLS-PM main results

Here, we present the main results of the fitted PLS-PM model. In order to simplify the interpretation, we merely present the results of *Satisfaction* and *Loyalty*. As indicated by [32] we report:

1. The *unidimensionality* of the reflective latent constructs (Table 5). All Cronbach’s α are higher than 0.86. For *Perceived Quality* its value has not been provided since it is formative, nevertheless, the sizes of the first and second eigenvalues confirm that even this construct is unidimensional in that case.
2. *Reliability*, as measured by the Average Variance Extracted for each construct in respect to its indicators (Table 6). In all cases, it is greater than 0.70, meaning indicator coherence in

²It is worth noticing that we have not taken into consideration the direct effect of *Expectations* on *Satisfaction* and *Perceived Value*, since they usually result to be non-significant in satisfaction studies.

measuring the construct.

3. The *estimated inner model* (Figure 9). We can see that the main (direct) drivers for *Satisfaction* are *Perceived Value* and *Image*, whereas *Satisfaction* is the main antecedent for *Loyalty*.
4. The *significance of the path coefficients*. We assess this by looking at the bootstrap confidence intervals of path coefficients (Table 7), of which all are significant except for two: the effect of *Perceived Quality* on *Satisfaction* (which is likely to indicate that the quality offered by different operators is not a distinctive driver for *Satisfaction*); and the effect of *Complaints* on *Loyalty* (indicating that the level of *Complaints* is not an important driver for staying with an operator).
5. *Predictability of the model* (Table 6), with an R^2 of 0.789 and 0.655 for the *Satisfaction* and *Loyalty* constructs, respectively (which are normal values in customer satisfaction studies).

6.2 Looking for heterogeneity

Despite the reduced sample size of the pilot study, it seems that the obtained model is good enough to draw conclusions about how *Satisfaction* and *Loyalty* are formed among mobile phone customers. However, the question remains about whether the effects we have detected are valid for the whole population or if they are in fact artificial averages of underlying subpopulations. To answer that question and discover potential sources of heterogeneity, we have performed a Pathmax analysis using the segmentation variables in Table 2. All variables are binary, with two exceptions: *Education*, which is declared ordinal; and *Operator*, which is taken as nominal. The tree we are interested in has to be necessarily small since the sample size is not at all big, and also because model segmentation does not involve looking for big trees with a large number of leaves, as this could become subject to overfitting. Instead, we are interested in a moderate final number of segments, which can be interpreted and made operational for marketing purposes. In this case, we have limited the tree by imposing a maximum depth of two levels (limiting the final number of segments to 4, at most); a minimum admissible size of 12% for nodes (= 10 individuals), which are really few but enough to allow a split in the *Operator* variable, since we are interested in investigating the differences among customers according to their mobile provider; and a threshold significance that is set at 0.05 for the split criterion. The obtained Pathmax tree is shown in Figure 10. It is a tree with a total number of three terminal nodes. The root node corresponds to the previous global model that was calculated over the entire sample of 87 customers.

The first split results from the segmentation variable *Operator*, which has a corresponding F statistic p-value of 0.0001. This split divides the 87 customers of the root node into two subsets. One subset is made up of 12 customers from operator A (node 2). The other subset consists of 75 customers from Operators B and C (node 3). For Node 2, it is not possible to find any further significant splits that verify the size condition, but Node 3 is divided by *Education* on

nodes 6 and 7, with an F statistic p-value of 0.0155. Node 6 represents the path model of the 19 customers with a basic education, whereas Node 7 represents the path model of the segment with 56 customers who have a high-school or university education.

All the splits performed are highly significant and define a meaningful tree of models. The path coefficients of the structural model for each final segment (as well as the path coefficients of the global model) are listed in Table 8. Focusing on the *Satisfaction* and *Loyalty* model, we can see that, for customers of operator A, the important driver for *Satisfaction* is the *Image* of the operator, whereas for customers of B and C, the important driver for *Satisfaction* is the *Perceived Value* obtained from the mobile phone. In relation to the role of *Perceived Quality* in the three segments (which turned out to be non-significant in the global model), it has a positive effect for customers of operator A and customers of B and C operators who have a basic education, whereas it is irrelevant for customers of operators B and C with higher education. Looking at the path coefficients of the *Loyalty* construct in the three segments, it happens that *Image* is again the important driver for customers of operator A, whereas *Satisfaction* is the important driver for customers of B and C. The latter customers differ on the role of *Image*, which is more important to less educated people. It is also worth noticing the different role of *Complaints* in the three segments: it is irrelevant for customers of A, as well for those of B and C who have a basic education; but it is more important to customers of B and C with higher education.

In Figure 11, we can see the differential models in each leaf in relation to *Satisfaction* and *Loyalty*. For both latent constructs, we represent their global model and the difference between the path coefficients of the local model respect to the global one.

6.3 Bootstrap validation

We may wonder about the significance of the local models obtained so far. In order to evaluate the significance of the estimates obtained in each final segment, we have applied bootstrapping performing 100 re-samples. The results of the bootstrap confidence intervals in the 5th and 95th percentile are listed in Table 9. As we can see, many path coefficients are non-significant, since the sizes of the terminal nodes are scarce. This is the case for Node 2 (clients of Operator A), which has just 12 individuals. Here, the confidence intervals are very large and only the effect of *Image* on *Loyalty* appears to be significant. For customers of operators B and C who have a basic education, only the effects of *Perceived Value* on *Satisfaction* and *Satisfaction* on *Loyalty* are significant. For customers of operators B and C with higher education, the effects of *Image* and *Perceived Value* on *Satisfaction* appear to be significant, as well as *Satisfaction* on *Loyalty*. Even if very few intervals are significant, we can appreciate differences between the estimated local and global models.

7 Discussion of the findings

Modeling is not an easy task. It involves capturing the signals of phenomena and their relationships. It is also true that reality is (almost) always heterogeneous and failure to consider this will lead to inaccurate results. Checking the heterogeneity of models must be standard practice when working with survey data.

Here, we have seen the suitability of the Pathmox approach in order to detect heterogeneity. This methodology can be applied regardless of whether the latent variables are reflective or formative, since we are interested only in detecting differences in the path coefficients. Moreover, Pathmox allows the researcher to find out which are the factors that define the heterogeneity in the data, if they are unknown. The results are easy to interpret by using a visual display. Pathmox is computationally inexpensive and allows the management of a large number of segmentation variables with complex models. Finally, the validity of the obtained local models is assessed by bootstrapping.

From a practical point of view, we recommend limiting the over-fitting tendency of trees by obtaining small trees in order to better guarantee that the obtained local models are robust and valid. However, the obtained results are based on a classical parametric statistic: the F statistic, which supposes the normality assumption of the perturbation terms with equal variance in all endogenous constructs; assumptions are rarely met in practice. Although violations of these assumptions may lead to an invalid F statistic, the simulations undertaken show the robustness of the test.

It is worth noticing that Pathmox just focuses on differences in the structural model, whereas the outer model does not intervene in the split criterion; and it is recalculated in every child node. This implies that we cannot assure factor invariance (i.e. all constructs are equally defined in all segments). In case that we need to compare the constructs across segments, we advise to employ a common model, usually the global one, in all terminal nodes.

The objective of Pathmox is to discover different segments by following different models without taking into account the predictability of such models. Gaining insight into how endogenous constructs are formed in the detected segments can help adopt specific policies (i.e., whether we need to invest more in *Image* to improve *Satisfaction* in one segment or invest more in *Perceived Quality* to improve *Satisfaction* in another segment).

However, when using Pathmox, assessing the best model segmentation often implies improving the predictability in the child nodes as well. In the case of the undertaken application, the R^2 values for the three detected segments³ are: 0.838, 0.921 and 0.733 for *Satisfaction*, whereas it was 0.789 for the global model, and 0.910, 0.794, and 0.544, being 0.666 the R^2 of *Loyalty* in the global model. Furthermore, the usual statistics are available for all models in segments regarding the validity, significance of the path coefficients and predictability to assess their practical value.

Finally, the Pathmox approach follows a data mining strategy that retrieves the information contained in the segmentation variables we are dealing with, providing a meaningful ordering of segmentation variables per node. Hence Pathmox should not be taken as a device for testing a formal hypothesis.

³Model 1: the A operator customers; Model 2: the B, C operator customers with basic education; and Model 3: the B, C operator customers with higher education.

8 Conclusions

The Pathmox approach is a realistic solution to unveil the existing heterogeneity in a survey, assuming that we have observed a set of segmentation variables, which is almost the case in marketing research. The undertaken simulations have shown the ability of the F statistic to detect differences in the path coefficients of PLS-PM models in accordance with the experimental conditions: the sensitivity of the F statistic is guaranteed by larger sample size, lower levels of random perturbations and clearer differences in the segments. The distribution of manifest variables, does not have any effect on the results, whether they are symmetrical or skewed.

Applying the Pathmox approach to the primary Spanish mobile phone providers clearly shows its effectiveness. Although the global model suits all the standard validation criteria of PLS-PM models, the use of Pathmox highlighted two sources of heterogeneity: the *phone operator* (which, as expected, behaves differently) and *level of education*. Therefore, looking at the main drivers of *Satisfaction* and *Loyalty* reveals that *Image* is an important driver of *Loyalty* for customers of the least expensive operator. For customers of the most prevalent and elitist operators, *Perceived Value* influences their *Satisfaction* and so does *Loyalty*. And among the higher educated customers in the latter segment, *Image* is also an important driver for *Satisfaction*.

Finally, although the Pathmox approach is presented within the PLS-PM framework, the idea of comparing two models (i.e., the homogeneity and heterogeneity hypotheses) can easily be generalized into a broader kind of modeling. However, this would require knowing the discrepancy distribution of the two measures of fit in the null and alternative models.

References

- [1] Muthén B.O. Latent Variable Modeling in Heterogeneous Populations. *Psychometrika* 1989; 54: 557-585. DOI 10.1007/BF02296397
- [2] Jedidi K., Jagpal H.S., DeSarbo W.S. Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science* 1997; 16: 39-59. DOI : 10.1287/mksc.16.1.39
- [3] Hair J.F., Sarstedt M., Ringle C.M. An assessment of the use of partial least squares structural equation modeling in marketing research. *Journal of the Academy of Marketing Science* 2012; 40: 414-430. DOI: 10.1007/s11747-011-0261-6
- [4] Bagozzi R.P., Yi Y. Specification, evaluation, and interpretation of structural equation models. *Journal of the Academy of Marketing Science* 2012; 40: 8-34. DOI: 10.1007/s11747-011-0278-x
- [5] Wold H. Soft modeling: The Basic Design and Some Extensions. In: K.G. Jöreskog and H. Wold. (Eds.), *Systems under indirect observation: causality, structure, prediction*, Vol. 2, North Holland, Amsterdam, 1982; 1-54.
- [6] Fornell C., Bookstein F.L. Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory. *Journal of Marketing Research* XIX 1982; 440-452.
- [7] Wold H. Partial Least Squares. In: S. Kotz and Johnson, N.L. (Eds.), *Encyclopedia of Statistical Sciences*, Vol. 6, Wiley, New York, 1985; 581-591.
- [8] Lohmöller J.B. Latent variable path modeling with partial least squares. *Physica Heidelberg*, 1989.

- [9] Chin W.W. The partial least squares approach for structural equation modeling. In G.A. Marcoulides (eds.). *Modern methods for business research*. Lawrence Erlbaum. New Jersey, 1998; 295-336.
- [10] Anderson E., Fornell C. Foundations of the American Customer Satisfaction Index. *Total Quality Management* 2000; 11(7): 869-882. DOI:10.1080/09544120050135425
- [11] Hackl P., Westlund A.H. On structural equation modeling for customer satisfaction measurement. *Total Quality Management* 2000; 11(4,5,6): 820-825. DOI: 10.1080/09544120050008264
- [12] Tenenhaus M., Esposito Vinzi V., Chatelin Y.M., Lauro C. PLS Path Modeling, *Computational Statistics and Data Analysis* 2005; 48: 159-205. DOI: 10.1016/j.csda.2004.03.005
- [13] Esposito Vinzi V., Trinchera L., Amato S. PLS Path Modeling: From Foundations to Recent Developments and Open Issues for Model Assessment and Improvement. In: V. Esposito Vinzi, W.W. Chin, W.W., J. Henseler and H. Wang (Eds.), *Handbook of Partial Least Squares. Concepts, Methods and Applications*. Springer, 2010.
- [14] Fornell C.G., Johnson M.D., Anderson E.W., Cha J., Everitt B. The American Customer Satisfaction Index: Nature, Purpose and Findings. *Journal of Marketing* 1996; 60(4): 7-18.
- [15] Westlund A.H., Cassel C.M., Eklöf J., Hackl P. Structural analysis and measurement of customer perceptions, assuming measurement and specifications errors. *Total Quality Management* 2001; 12(7,8): 873-881. DOI:10.1080/09544120100000010
- [16] Sanchez G., Aluja T. Pathmox: Segmentation trees in partial least squares path modeling (R package Version 0.4-1) [Computer software manual]. Available from <http://CRAN.R-project.org/package=pathmox/> [2012].
- [17] Hahn C., Johnson M.D., Herrmann A., Huber A. Capturing Customer Heterogeneity Using a Finite Mixture PLS Approach. *Schmalenbach Business Review*, 2002; 54: 243-269.
- [18] Ringle C.M., Wende S., Will A. Customer Segmentation with FIMIX-PLS. In: T. Aluja, J. Casanovas, V. Esposito Vinzi, A. Morineau, M. Tenenhaus (Eds.), *Proceedings of the PLS05 International Symposium, SPAD Test&go*, Paris, France, 2005; 507-514.
- [19] Ringle C.M. Segmentation for Path Models and Unobserved Heterogeneity: The Finite Mixture Partial Least Squares Approach. *Research Papers on Marketing and Retailing*. University of Hamburg, 2006; No. 35.
- [20] Squillacciotti S. Prediction oriented classification in PLS Path Modelling. In: T. Aluja, J. Casanovas, V. Esposito Vinzi, A. Morineau, M. Tenenhaus (Eds.), *Proceedings of the PLS05 International Symposium, SPAD Test&go*, Paris, France, 2005; 499-506.
- [21] Squillacciotti S. Prediction Oriented Classification in PLS Path Modeling. In: V. Esposito Vinzi, W.W. Chin, W.W., J. Henseler and H. Wang (Eds.), *Handbook of Partial Least Squares. Concepts, Methods and Applications*. Springer, 2010.

- [22] Trinchera L., Squillacciotti S., Esposito Vinzi V., Tenenhaus M. PLS path modeling in presence of a group structure: REBUS-PLS, a new response-based approach. In: H. Martens and T. Naes. (Eds.), Proceedings of the PLS07 International Symposium, Matforsk, As, Norway, 2007; 79-82.
- [23] Esposito Vinzi V., Trinchera L., Squillacciotti S., Tenenhaus M. REBUS-PLS: A Response-Based Procedure for detecting Unit Segments in PLS Path Modeling. Applied Stochastic Models in Business and Industry, 2008; 24: 439-458. DOI: 10.1002/asmb.728
- [24] Sanchez G. Pathmox approach: Segmentation trees in partial least squares path modeling (Unpublished doctoral dissertation). Universitat Politècnica de Catalunya, Catalonia, Spain. <http://gastonsanchez.com/thesis/pathmox-approach-thesis-gaston-sanchez.pdf> [2012]
- [25] Sanchez G., Aluja T. Pathmox: A PLS segmentation algorithm. In Electronic Proceedings of the workshop on Knowledge Extraction and Modeling (KNEMO). V. Esposito-Vinzi, C. Lauro, A. Braveman, H.A.L. Kiers, M.G. Schimeck (eds.) Anacapri (Italy), 2006.
- [26] Sanchez G., Aluja T. A simulation study of Pathmox sensitivity. In Proceedings of the PLS07 International Symposium. H.Martens & T. Naes (eds.) Matforsk, As (Norway), 2007; 33-36.
- [27] Sanchez G., Aluja T. A simulation study of Pathmox with non-normal data. In Proceedings of the PLS'09 International Symposium. V. Esposito-Vinzi, M. Tenenhaus, R. Guan (eds.) Publishing House of Electronics Industry. Beijing, 2009; 54-59.
- [28] Brandmaier A.M., Oertzen T., McArdle J.J., Lindenberger U. Structural Equation Model Trees. Psychological Methods. American Psychological Association, 2013; Vol. 18, No. 1, 71-86. DOI: 10.1037/a0030001
- [29] Baron R.M., Kenny D.A. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology 1986; 5(6): 1173-1182. DOI: 10.1037/0022-3514.51.6.1173
- [30] Tenenhaus M., Mauger E., Guinot C. Use of ULS-SEM and PLS-SEM to Measure a Group Effect in a Regression Model Relating Two Blocks of Binary Variables. In: V. Esposito Vinzi, W.W. Chin, W.W., J. Henseler and H. Wang (Eds.), Handbook of Partial Least Squares. Concepts, Methods and Applications. Springer, 2010.
- [31] Chin W.W, Marcolin B.L., Newsted P.R. A Partial Least Squares Latent Variable Modeling Approach for Measuring Interaction Effects: Results from a Monte Carlo Simulation Study and an Electronic-Mail Emotion/Adoption Study. Information Systems Research 2003; 14(2): 189-217.
- [32] Chin, W.W. How to Write Up and Report PLS Analyses. In Chin, Wynne W., Henseler, J., Wang, H. (Eds.), Book title: Handbook of Partial Least Squares. Springer, 2010.
- [33] Chin W.W. A permutation based procedure for multi-group comparison of PLS models. In Proceedings of the PLS03 International Symposium, M. Vilares, M. Tenenhaus, P. Coleho, V. Esposito-Vinzi (eds.). Decisia, 2003b; 33-43.

- [34] Henseler J., Fassot G. Testing Moderating Effects in PLS Path Models. In: T. Aluja, J. Casanovas, V. Esposito Vinzi, A. Morineau, M. Tenenhaus (Eds.), Proceedings of the PLS05 International Symposium, SPAD Test&go, Paris, France, 2005; 371-377.
- [35] Henseler J., Chin W.W. A Comparison of Approaches for the Analysis of Interaction Effects Between Latent Variables Using Partial Least Squares Modeling. *Structural Equation Modeling: A Multidisciplinary Journal* 2010; 17: 1, 82-109. DOI:10.1080/10705510903439003
- [36] Lebart L., Morineau A., Fénelon J.P. *Traitement des données statistiques*. Dunod, 1979.
- [37] Chow G.C. Test of equality between sets of coefficients in two linear regressions. *Econometrica* 1960; 28(3): 591-605.
- [38] Cassel C., Hackl P., Anders H.W. Robustness of partial least-squares method for estimating latent variable quality structures. *Journal of Applied Statistics* 1999; 26(4): 435-446. DOI:10.1080/09544120050135443
- [39] R package plsmp. <http://cran.r-project.org/web/packages/plsmp/index.html/> [2013].
- [40] Reinartz W.J., Echambadi R., Chin W.W. Generating Non-normal Data for Simulation of Structural Equation Models Using Mattson's Method. *Multivariate Behavioral Research* 2002; 37(2): 227-244. DOI:10.1207/S15327906MBR3702_3
- [41] Sarstedt, M., Ringle C.M., Christian M., Henseler, J., Hair J.F. On the emancipation of PLS-SEM: a commentary on Rigdon. *Long Range Planning* 2012; 47 (3), 154-160.

Table 1: Description of the manifest variables for each of the latent constructs

Latent variables	Description of Indicators
Image	1) Trustworthiness 2) Dynamic 3) Company's solidness 4) Innovation and forward looking 5) Caring about the customer's needs
Expectations	1) Providing customer information 2) Providing clarity in payments 3) Providing customer service 4) Providing adapted services
Perceived Quality	1) Quality connection 2) Degree to which customer feels well informed 3) Clarity and transparency of transactions 4) Range of products and services 5) Customer service
Perceived Value	1) Beneficial services and products 2) Quality relative to price 3) Price relative to quality
Customer Satisfaction	1) Overall rating of satisfaction 2) Fulfillment of expectations 3) Rating the performance relative to customer's ideal operator
Complaints	1) Adequate response 2) Quick response 3) Effectiveness
Loyalty	1) Propensity to choose the same operator if the customer would had to choose again 2) Customer's intention to recommend the operator to friends or colleagues

Table 2: Codification of segmentation variables according to their type of scale and levels

Name	Scale	N. levels	Levels description
Gender	binary	2	female, male
Age	binary	2	under 25, 25 or over
Occupation	binary	2	employee, student
Education	ordinal	3	basic, high-school, university
Contract's typology	binary	2	contract, prepay
Operator	nominal	3	A, B, C
Switch of operator	binary	2	YES, NO

Table 3: Description of latent variables of Mobile dataset

LV	Description
Image	Includes variables such as trustworthiness, dynamic, solidness, innovation, and caring about customer's needs.
Expectation	Includes variables such as products and services provided and expectations for the overall quality.
Quality	Includes variables such as reliable products and services, range of products and services, and overall perceived quality.
Complaints	Includes one variable defining how well or poorly customers' complaints were handled.
Value	Includes variables such as service and products, quality relative to price, and price relative to quality.
Satisfaction	Includes variables such as overall rating of satisfaction, fulfillment of expectations, satisfaction relative to other phone providers.
Loyalty	Includes variables such as propensity to choose the same phone provider again, intention to recommend the phone provider to friends.

Table 4: Nine pairs of path coefficients for node B

	Pair 1	Pair 2	Pair 3	Pair 4	Pair 5	Pair 6	Pair 7	Pair 8	Pair 9
β_1	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90
β_2	0.50	0.45	0.40	0.35	0.30	0.25	0.20	0.15	0.10

Table 5: Different measures to assess unidimensionality of blocks of indicators

Constructs	Type	N.Indicators	Crombach's α	Dillon ρ	1-Eigenvalue	2-Eigenvalue
Image	Reflective	5	0.894	0.922	3.52	0.538
Expectations	Reflective	4	0.863	0.907	2.84	0.475
Quality	Formative	5	0.000	0.000	3.09	0.602
Value	Reflective	3	0.874	0.923	2.40	0.323
Satisfaction	Reflective	3	0.883	0.928	2.43	0.309
Complaints	Reflective	3	0.932	0.957	2.64	0.246
Loyalty	Reflective	2	0.896	0.951	1.81	0.188

Table 6: Summary results per blocks

Constructs	Type	Measur. type	Number Indicators	R.square	Av. Communalilty	Av. Redundance	AVE
Image	Exogen.	Rfct	5	0.000	0.703	0.000	0.703
Expectations	Endogen.	Rfct	4	0.197	0.705	0.139	0.705
Quality	Endogen.	Frmtv	5	0.257	0.586	0.151	0.000
Value	Endogen.	Rfct	3	0.483	0.798	0.385	0.798
Satisfaction	Endogen.	Rfct	3	0.789	0.809	0.638	0.809
Complaints	Endogen.	Rfct	3	0.255	0.880	0.224	0.880
Loyalty	Endogen.	Rfct	2	0.666	0.906	0.604	0.906

Table 7: Bootstrap confidence intervals for path coefficients (0.025 and 0.975 percentiles)

Paths	Original path value	Bootstrap mean	Bootstrap stand. error	perc. .025	perc. .975	Significance
Image -> Satisfaction	0.427	0.422	0.0934	0.216	0.599	YES
Quality -> Satisfaction	0.101	0.110	0.1223	-0.141	0.330	NO
Value -> Satisfaction	0.472	0.468	0.0813	0.291	0.618	YES
Image -> Loyalty	0.294	0.275	0.1154	0.073	0.458	YES
Satisfaction -> Loyalty	0.485	0.484	0.1200	0.251	0.704	YES
Complaints -> Loyalty	0.131	0.151	0.0876	-0.032	0.308	NO

Table 8: Path coefficients of the root node and the terminal segments

Paths	Root node	A Segment	B,C-basic-ed. Segment	B,C-high-ed. Segment
Image -> Satisfaction	0.427	0.419	0.305	0.423
Quality -> Satisfaction	0.101	0.210	0.206	-0.012
Value -> Satisfaction	0.472	0.393	0.531	0.538
R^2 Satisfaction	0.789	0.838	0.921	0.733
Image -> Loyalty	0.294	0.573	0.368	0.162
Satisfaction -> Loyalty	0.485	0.361	0.575	0.531
Complaints -> Loyalty	0.131	0.079	-0.029	0.163
R^2 Loyalty	0.666	0.910	0.794	0.544

Table 9: Path coefficients of the root node and the terminal segments

Paths	<i>Node₂</i>		<i>Node₆</i>		<i>Node₇</i>	
	<i>A</i>		<i>B-C basic ed.</i>		<i>B-C high ed.</i>	
Image -> Satisfaction	-0.396	0.734	-0.044	0.538	0.183	0.562
Quality -> Satisfaction	-0.391	1.771	-0.133	0.745	-0.220	0.310
Value -> Satisfaction	-0.745	0.902	0.246	2.039	0.307	0.675
Image -> Loyalty	0.181	0.996	-0.211	0.812	-0.124	0.398
Satisfaction -> Loyalty	-0.634	0.597	0.147	1.101	0.266	0.790
Complaints -> Loyalty	-0.154	0.679	-0.331	0.270	-0.010	0.426

Figure 1: Path diagram of the Customer Satisfaction mobile phone model

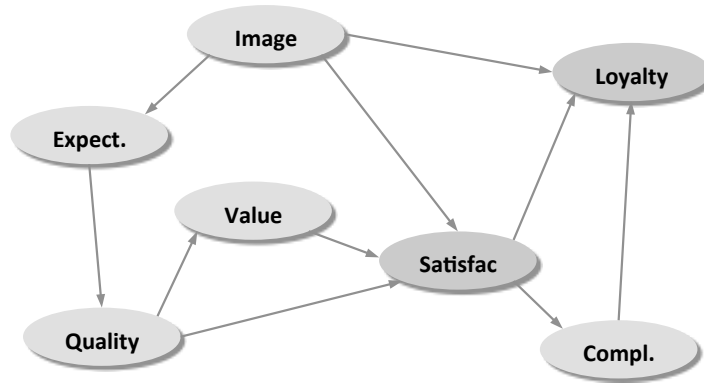


Figure 2: Path diagram of a path model with two endogenous latent variables

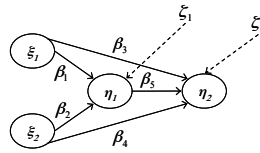


Figure 3: Path Diagram of the structural model in the simulation study

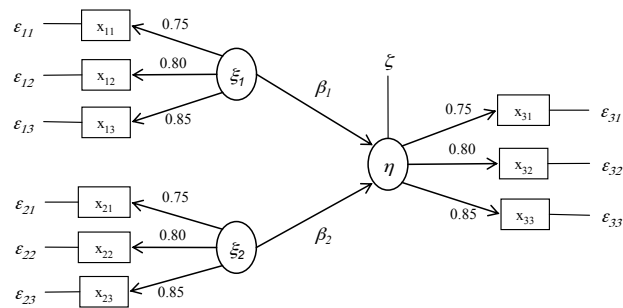


Figure 4: Examples of beta distributions for the manifest variables

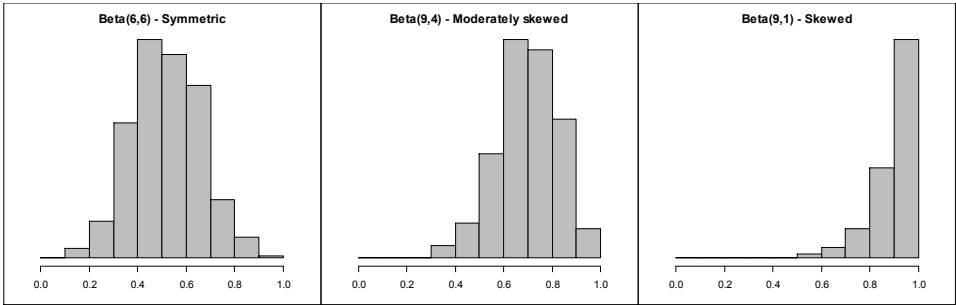


Figure 5: Influence of the skewness of the distributions on the p-value of the test

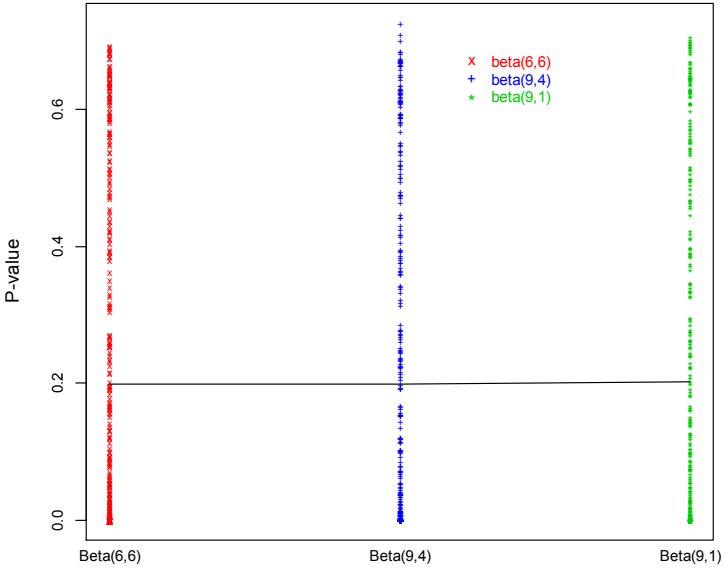


Figure 6: Influence of the distance between path coefficients of nodes A and B

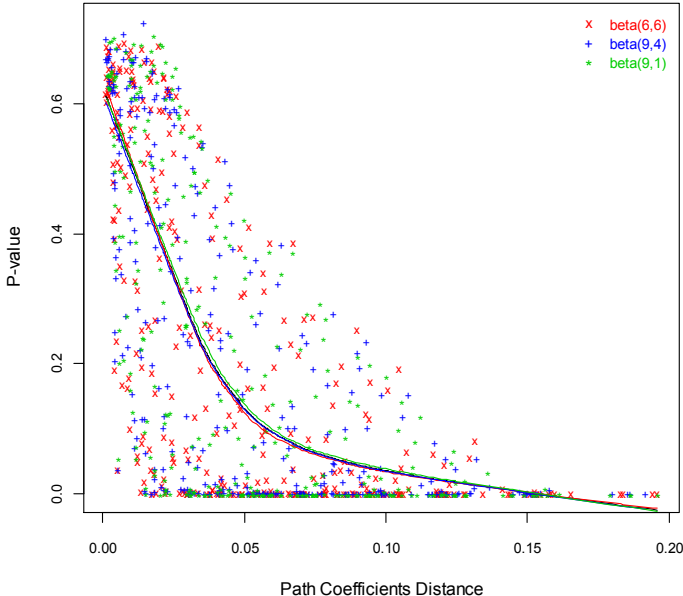


Figure 7: Influence of the sample size of nodes on the p-value of the test

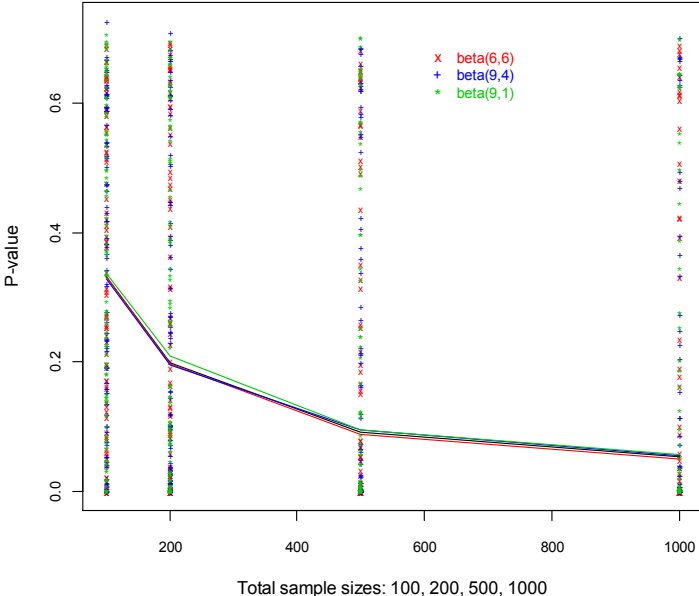


Figure 8: Influence of the level of noise of nodes on the p-value of the test

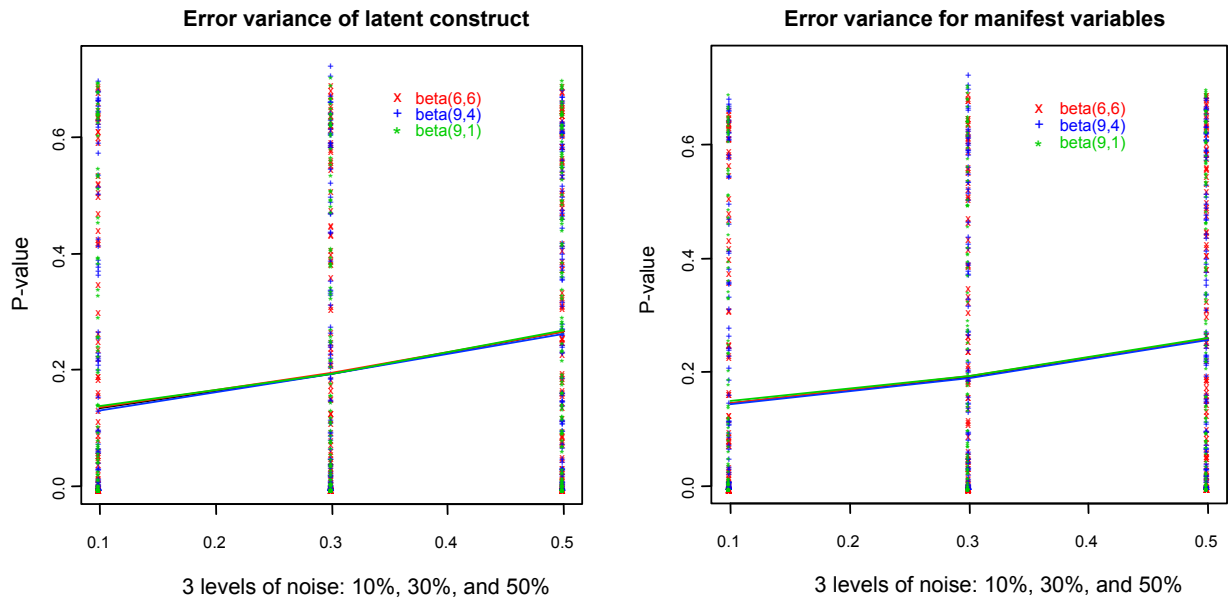


Figure 9: Estimated inner model of mobile data

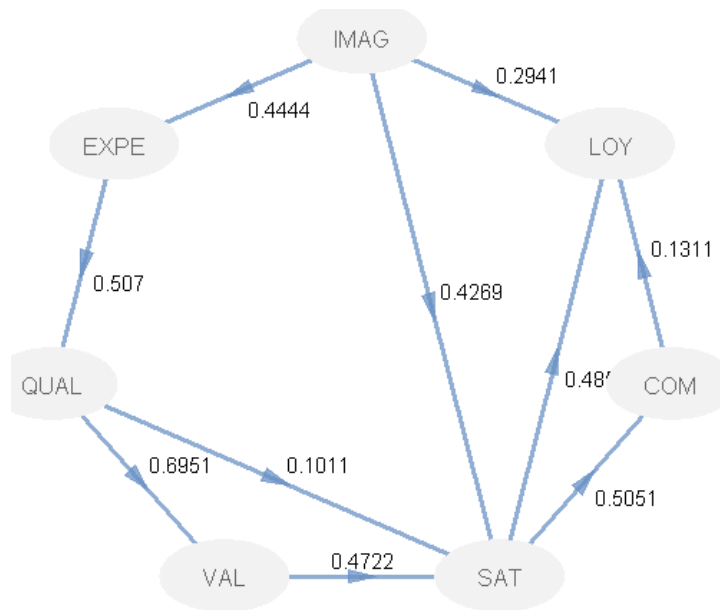


Figure 10: Pathmox tree of mobile data (from R pathmox)

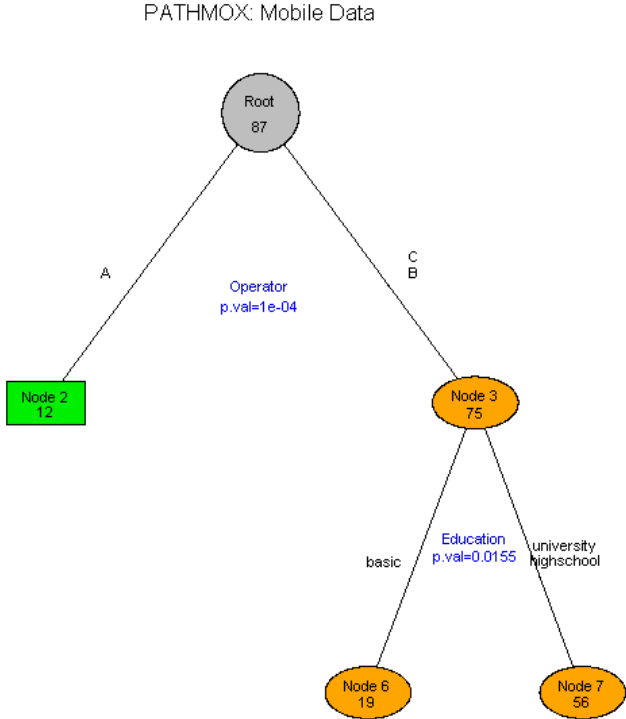


Figure 11: Differential paths for Satisfaction and Loyalty in segments compared to the global model

