UNIVERSIDAD SANTO TOMAS
PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA

Comunicaciones en Estadística
August 2016, Vol. 9, No. 2, pp. 157–186

Loyalty Model from a Structural Equation Model Analysis[1]

Modelo de lealtad a partir de un análisis de ecuaciones estructurales

Robert Romero[a]
gilromero@usantotomas.edu.co

Giovanny Babativa M.[b]
josebabativa@usantotomas.edu.co

Abstract

Day by day it becomes more important for companies to own a database of loyal customers, since continuous monitoring and expansion of business with them gives higher returns than the pursuit of new customers. Several methodologies have been used to measure satisfaction and loyalty of customers, most of them are based on Behavioral Psychology from positivist approach and focuses on the description of manifest behaviors measured directly, its main goal is to establish the direct importance that each service driver has on loyalty, this is the case of methodologies based on regression models. In this paper we illustrate an application of a loyalty model that seeks, through an analysis of structural equations with latent variables, to de?ne the size of the effect on the loyalty of phenomena like overall satisfaction with service, repurchase intention and recommendation.

***Keywords*:** direct effect, DWLS, exploratory factor analysis, indirect effect, latent variable, measure model.

Resumen

Cada día toma mayor importancia para las empresas contar con bases de datos de clientes leales, ya que el continuo seguimiento y la ampliación de negocios con ellos otorga mayor rentabilidad que la consecución de nuevos clientes. Son varias las metodologías que se han utilizado con la finalidad de medir la satisfacción y lealtad, la mayoría de ellas están basadas en la psicología conductista propia del positivismo y se apalancan en la descripción de conductas manifiestas medidas directamente y tienen como principal objetivo establecer la importancia directa que cada *driver* del servicio tiene sobre la lealtad; tal es el caso de las metodologías

[1]Romero, R., Babativa, G. Loyalty Model from a Structural Equation Model Analysis. *Comunicaciones en Estadística*, **9**(2), 147-176.

[a]Statistician, Santo Tomás University, Colombia.
[b]Dean of the Faculty of Statistics, Santo Tomás University, Colombia.

basadas en modelos de regresión. En este artículo se ejemplifica la aplicación de un modelo de lealtad que busca, a través de un análisis de ecuaciones estructurales con variables latentes, determinar el tamaño del efecto que tienen sobre la lealtad fenómenos como la satisfacción con el servicio, la intención de recompra y la recomendación.

***Palabras clave*:** análisis factorial exploratorio, DWLS, efecto directo, efecto indirecto, modelo de medida, variable latente.

1 Introduction

The relationship with their customers it's always been important for a company; therefore patterns to describe it have been set out in different works. In this search the use of statistical models that allow establishing the way in which the different variables determine the behavior of a customer, has become relevant.

Regression models are among the used models to describe the relationship between the component variables of satisfaction and the loyalty, factorial analysis and lately the models of structural equations (SEM), which tend to be the more accurate to make the estimation because some of these variables are exogenous and endogenous at the same time and it's not easy to establish a direct relation between them. In this case the model of structural equations allows establishing the direct and indirect effects of variables between them; the said values are used to build an indicator to establish strategies to raise the loyalty level of customers.

This paper explores the use of models of structural equations (SEM) with the purpose of building a model of loyalty from the components of the service offered to the customer. The first part describes the general theory of SEM models; the second one makes the descriptive analysis of data and examines the possible relations between the variables that were used; by last a practical example of the estimation with SEM models and its respective results is introduced and conclusions and future lines of research are established.

2 Fundamental concepts

Nowadays the analysis of SEM models is an important multivariate technique applied in different professional fields; however, statistical techniques supporting it are not of recent appearance and it had to pass quite some time before it could exploit its potential, this thanks to the computational advances that, like in many other fields, have been the artery of the knowledge development; they also allowed a deeper research of the technique, which has consisted mainly in simulation studies.

Roughly, SEM models allow the representation of a series of hypothesis of relations (mainly linear) between a series of measured variables and caused in turn by a diversity of underlying phenomena, which are not directly observable and in this case are called latent (in other fields of research they are know as factors or constructs). Latent variables are of great importance in many disciplines but they lack an accurate measurement mode, as regards their existence or influence in other phenomena.

Possible examples of latent variables may be air quality, happiness or intelligence, phenomena that, given its unobservability, can't be measured directly, therefore researchers define a series of operational tools through which they can be built indirectly. These tools are called manifest variables (observed) and in the methodology the SEM models serve as indicators of the underlying phenomena they represent.

The term SEM is a generalization for several types of models and statistically represents an extension of procedures of general lineal models (MLG), such as ANOVA, analysis of multiple regression and factorial analysis, which have the following features that make them different to the techniques of classic modeling:

They are generally conceived as theoretical constructs of phenomena that are not directly measurable.

They take into account possible measurement errors of variables with which latent factors are built. The variances of the error terms are the parameters to estimate when adjusting the model, therefore it is correct in this case to call it analysis of structure of covariances.

Models are adjusted from matrices of interrelation index (matrix of correlations or covariances), although sometimes the analysis over the median of variables will also be done.

Among the types of SEM models the following may be mentioned:

Path Analysis: This type of analysis seeks, through the support of a path diagram, to decompose the covariance among the model's variables with the purpose of establishing the measure of relation between the causal effect and the measure of covariation. Such relations may be direct, spurious, indirect or combined. It also allows measuring the causal effect (indirect and total) that a variable has on another in the model. According to Long (1983) they are also called models of structure of covariances and they split into two: the model of structural component and the model of measurement component.

Factorial confirmatory model: It allows analyzing the patterns of relation or causation between latent variables (constructs) of the structural model, with the purpose of verifying if this is valid or if its interrelations generate some plausible interpretation.

Growth curve models: The models previously mentioned are based on cross-sectional data obtained through a simple of individuals on a time point t. Growth curve models allow the analysis the dynamic of changes and evolutions of the behavior of processes under study for linear data.

3 Path Diagram

Sometimes the systems of equations are very complex and require the introduction of many relations between the variables; in this case the graphic representation of the model under consideration through a casual diagram or path diagram as in figure 1 is generally preferred; this type of representations equals a set of equations that set up the model. In the graphic representation a special notation is used:

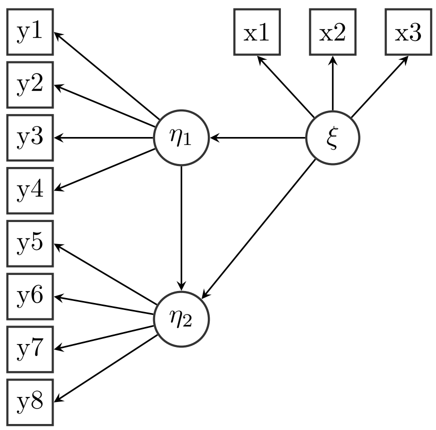
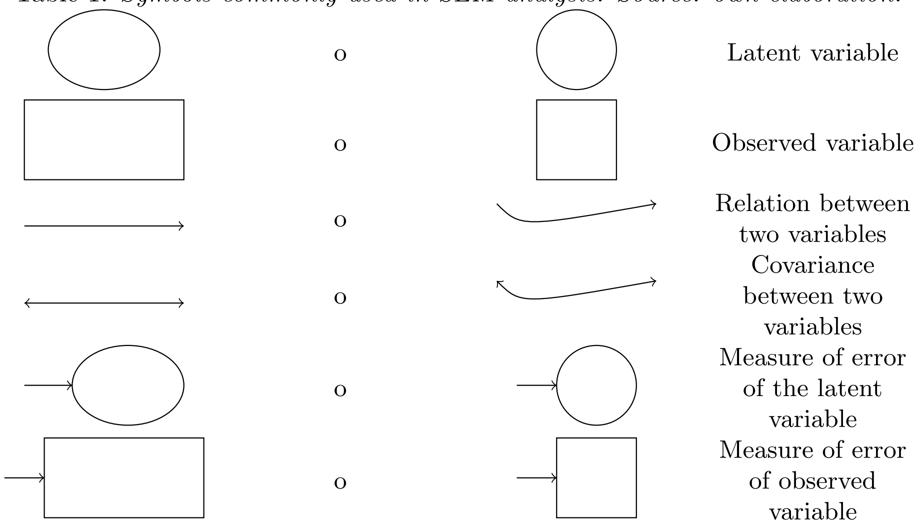


Figure 1: *Model of structural equation represented by a path diagram. Source: own elaboration.*

When model includes both observed and latent variables, first ones are represented with squares or rectangles and the second ones with circle or ellipsis; the arrows coming out of latent variables to the observed variables are called measurement relations. Observed variables are affected by a random term of error, which is represented in the diagram by a directional arrow pointing to the observable variable. A bidirectional arrow joining them represents the covariation between two measurement errors. An arrow from the exogenous variable to the endogenous variable indicates the relation between variables (table 1).

Table 1: *Symbols commonly used in SEM analysis. Source: own elaboration.*



4 Path Analysis

The main goal of the path analysis was already determined in section 2 ; in order to develop completely the concept the relation types that can lead two variables x and y to covary will be detailed.

x and y may covary if y has any effect on x (or the opposite), as the relation represented by a simple regression model, present in figure 2:

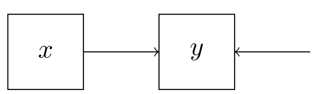


Figure 2: *Path Diagram simple regression. Source: own elaboration.*

These relations are called direct, but can also be reciprocal (?gure 3):

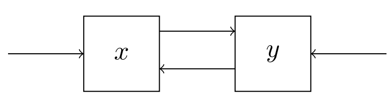


Figure 3: *Path Diagram reciprocal relation. Source: own elaboration.*

x and y covary if they have a common cause z; this type of relation is called spurious (?gure 4):

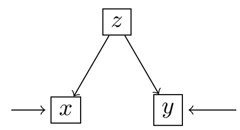


Figure 4: *Path Diagram spurious relation. Source: own elaboration.*

x and y covary also if they are related through a third variable z; this type of relation is called indirect (figure 5):

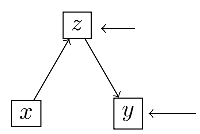


Figure 5: *Path Diagram indirect relation. Source: own elaboration.*

Once the different covariance types between the model's variables are pointed, a series of decomposition rules that allow to establish the relations between the covariances and the parameters of the model are implanted; once these relations are established it's possible to make the calculation of parameter estimation. It's important not to lose sight of the fact that the model's adjustment must be made on the matrices of covariances of the observed variables, which are previously focused to put aside the effect that may have the median of each observed variable.

The variances and covariances of observed variables are themselves initial measures of the model. According to Batista & Coenders (2000) to derive the other parameters it follows that:

> The covariance between two variables is calculated as the sum between the direct, indirect, spurious and combined effects. Each one of them represents in the path diagram a possible way to join the variables. The effect is calculated as the product of variance of the starting variable by all the parameters associated to the arrows plotted until they join the two variables of interest (figures 6 and 7).

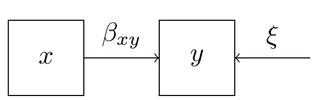
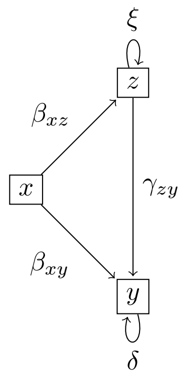


Figure 6: *Total effect of x on y. Source: own elaboration.*



Direct effect:	β_{xy}
Indirect effect:	$\beta_{xz} * \gamma_{zy}$
Total efecto:	$\mathbf{B}_{xy} = \beta_{xy} + \beta_{xz} * \gamma_{zy}$

Figure 7: *Decomposition of the effect of x on y. Source: own elaboration.*

> Variance of an exogenous variable is calculated as the variance of the term of error plus the variable explained by other variables in the model. Also, the explained variance can be expressed as a function of all the exogenous variables with direct effect on the endogenous variable, as the rise in all products between the direct effects and the covariances between endogenous and exogenous variables related by the effects.

Applying the previous rules a system of structural equations is obtained, which expresses the matrix of covariances in function of the model parameters

$$\Sigma = \Sigma(\theta) \tag{1}$$

Where θ is a vector that contains the model parameters (direct and indirect effects, variances, covariances of errors, of perturbations and exogenous variables)

5 Stages of construction of SEM models

As seen, the analysis of SEM models requires a prior knowledge of possible interactions or relations between both variables of the model, exogenous and endogenous; this means that in the path diagram hypothetic relations of phenomena under study are plotted on a graph. From this point a series of steps[1] for the researcher to follow is enumerated before applying the model in the practice; these steps are repeated, since depending on what happens in any of them you have to come back to the first one:

[1]For a detailed illustration of these steps see Romero (2015).

Specification

Identification

Data collect

Estimation

Diagnostic

Evaluate if model is suitable (if it doesn't match: come back to step 1)

Use

6 Application for the loyalty model

Perhaps one of the most common concerns of companies, regardless of the country of origin, is to know how loyal are their customers; for this purpose they have constructed multiple types of models that deal with topics ranging from brand equity to evaluation of the dimensions of the final service offered.

This paper takes as baseline the simulation of an investigation about the quality of the attention received by the user in a service company. It's important to take into account that unlike the assessment of physical products, service evaluation is based on intangible components, which are subject to the subjectivity of the user according to the experience (Saurina 1997).

To this end a measure of satisfaction with the received service was taken through following variables:

intenc: the intention of taking again the services of the company.
recom: the disposition to recommend the company to other people.
schedule: customer service hours.
location: location of the headquarters in the place where it is needed.
comfort: the comfort of the headquarters.
senaliza: security feeling in the headquarters.
security: security feeling in the headquarters.
personal: number of people serving.
time: wait time in collection area.
amabi: friendliness and respect received in collection area.
agili: agility of the person serving in collection area.
manejo: lines arrangement in collection area.
espera: wait time for adviser.
respet: kindness and respect of adviser.
interes: interest shown by adviser.

claro: clarity of information provided by adviser.

soluc: effective solution of needs.

ag: agility in the response received.

cupos: quotas granted.

requisit: requirements to adquire the service.

trami: agility in procedures.

antigue: the recognition of seniority as customer.

comporta: recognition of the good use of the service.

acompa: the service after sale.

nuevo: offering of new products.

These variables are measured on an ordinal scale of one to ten, with 1 being the least qualification and 10 the highest. In this paper I will do the calculation of an indicator of the customer's loyalty from the method of structural equations, which will allow decomposing the direct and indirect effects that different latent dimensions of service have on loyalty. A factorial exploratory analysis is made before the structural analysis with the purpose of finding the underlying relations between observed variables.

The theoretical model that will be introduced contemplates that the above-mentioned variables make dimensions of the service, which have an indirect effect on loyalty and a direct effect on satisfaction, which affects in turn the customer's loyalty.

6.1 Exploratory Analysis of variables

Having listed the variables with which the indicator of loyalty will be calculated through the global model of structural equations and taking into account their ordinal character, in this section the results of the application of the normality tests are shown. The scale for each one of the variables is between 1 and 10, which is enough wide to be treated as numerical variables (though not continuous), for this reason it is valid to make analysis and statistical tests in order to verify the assumptions of models.

To make the normality tests the value p associated to the statistical of Anderson-Darling (table 2) is used, taking into account that his null hypothesis is:

$$H_0 : X \sim N(\mu, \sigma) \tag{2}$$

Table 2: *Test of Anderson Darling to contrast normality. Source: own elaboration.*

Variable	**Statistic**	***p*-value**
schedule	631,5722	0,00
location	733,9707	0,00
comfort	545,6445	0,00
senaliza	609,0171	0,00
security	668,7004	0,00
personal	250,9835	0,00
time	236,6273	0,00
amabi	802,5184	0,00
agili	391,6763	0,00
manejo	391,1865	0,00
espera	338,9154	0,00
respet	876,6862	0,00
interes	630,6738	0,00
claro	684,7826	0,00
soluc	603,0671	0,00
ag	585,1464	0,00
cupos	504,0901	0,00
requisit	453,0338	0,00
trami	459,0459	0,00
antigue	474,4710	0,00
comporta	509,7517	0,00
acompa	432,8783	0,00
nuevo	404,6410	0,00

Table 3: *Multivariate normality test of Mardia. Source: own elaboration.*

g1p:	103,7106
chi.skew:	142809,5
p.value.skew:	0,00
g2p:	1320,195
z.kurtosis:	998,6971
p.value.kurt:	0,00
chi.small.skew:	142865,7
p.value.small:	0,00

In this case the test rejects the hypothesis of normality for each observed variable; the results of the test of Mardia also reject the hypothesis of multivariate normality; however the models of structural equations manage robust methods for no normal variables, therefore its application is viable.

To analyze the data consistency we used the Alpha coefficient of Cronbach, which estimates the information reliability through a set of items that were measured. The general procedure to calculate the Alpha of Cronbach goes from the correlation matrix of Pearson, since the use of a Likert scale with an amount greater than 6 categories stabilizes the coefficient (Gelin et al. 2003) (table 4).

Table 4: *Alpha of Cronbach. Source: own elaboration.*

alpha	**std.alpha**	**Guttman's Lambda 6**
0,96	0,96	0,97

The Alpha coef?cient of Cronbach is of 0.96, that's why it's concluded that variables collect with high reliability the required information (Streiner 2003).

6.2 Exploratory factorial analysis

Before we begin the construction phase of the model, it's necessary to give coherence to the information's diversity that is being measured through some technique that allows finding, from its structure of correlation, underlying relations among vectors of analysis, this with the aim of defining groups of variables that are highly correlated with each other to the k latent factors that explain the greater amount of variance of the original $\mathbf{X}$ matrix. There are different statistical techniques of interdependence; in this case we use the factorial exploratory analysis (EFA).

Although the EFA is based on the assumption of normality, its use is considered in this article, since the target prior to the specification of the structural model is to contextualize the situation and have a measurement model that provides a basis for a causal analysis of relations between latent variables (Loehlin 2004).

6.3 Treatment of ordinal variables

The first step for the realization of an EFA is to make an assessment of the correlation matrix with the aim of establishing whether it justifies its implementation; however, it should be noted that in this case we are dealing with ordinal discrete variables, so the appropriateness of using the correlation matrix of Pearson should be studied, because sometimes it is not appropriate for the analysis; in these cases are the called polychoric correlations the ones that must be used as the starting point (Olsson 1979). This type of correlations are used to relate features that in principle are continuous, but were measured with an ordinal scale; a clear example are the features measured through the Likert scales used in this paper. Coenders et al. (1979) affirm that:

> A typical approach for modelling ordinal variables is to assume that there is an underlying variable y_i^* for each ordinal variable y_i and that each y_i is related to y_i^* through the passing function:

$$y_i = k \text{ when } \tau_{ik-1} < y_i^* \leq \tau_{ik}$$

for $k = 1, ..., m_i$, where $\tau_i = -\infty$, $\tau_{ik} < \tau_{ik+1}, \tau_{imi} = \infty$. The parameters τ_i with $i = 1, ..., m_{i-1}$, are called thresholds of the i-th variable.

However, depending on the estimation method, the matrix of polychoric correlations can be defined as no positive, then the factorial analysis would not be possible; so, taking as computational support the `psych` package of the `R` environment, the correlation matrices of Pearson and polychoric are analyzed by bootstrap tests to determine with which of them it's possible to do the EFA. In this simulation the cases are created through MASR and correlations are calculated, as many times as there are iterations. The median of correlation and its respective standard deviation are calculated based on the Z transformation of correlations of Fisher.

If it's denoted by $\hat{\rho}$ the polychoric correlation of interest, the interval of the Fisher transformation will be defined by:

$$z(\hat{\rho}) \pm z_{\delta/2} * SE(\hat{\rho})/(1 - \hat{\rho}^2) \tag{3}$$

Where

$$z(\hat{\rho}) = 0.5 * ln[(1 + \rho)/(1 - \rho)] \tag{4}$$

And in which $SE(\hat{\rho})$ is the standard error of polychoric correlation (Hoyle 2012). Results are shown in figure 8:

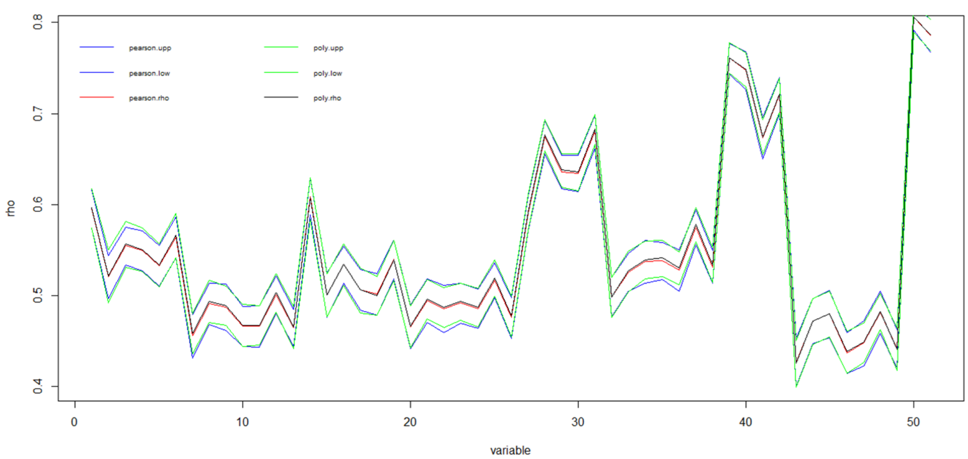


Figure 8: *CI (95%) generated by Bootstrap. Source: own elaboration.*

Figure 8 shows that results generated by the polychoric correlations method and the correlation matrix of Pearson for continuous data are very similar to each other;

in addition to this exists the possibility of obtaining a polychoric correlation matrix that is is not defined as positive, therefore in this paper we opted for using the classic correlation of Pearson to make the exploratory factorial analysis.

6.4 Number of factors to retain

As initial rule, the number of factors to retain must be close to the number of positive eigenvalues of the correlation matrix (Field 2000); however in this case it's possible to obtain a great amount of positive eigenvalues but very close to zero, which sometimes lead to handle with a great amount of factors that contribute with very little information of analysis; in this case the rule of Very Structure Simple is used, according to the cases cited by Romero (2015).

Results obtained by the method VSS with assistance of package `psych` of environment `R` show that the optimum number to retain is four factors, since at this point is the biggest number of factors with which $VSS_{vk} \approx 1$ (figure 9).

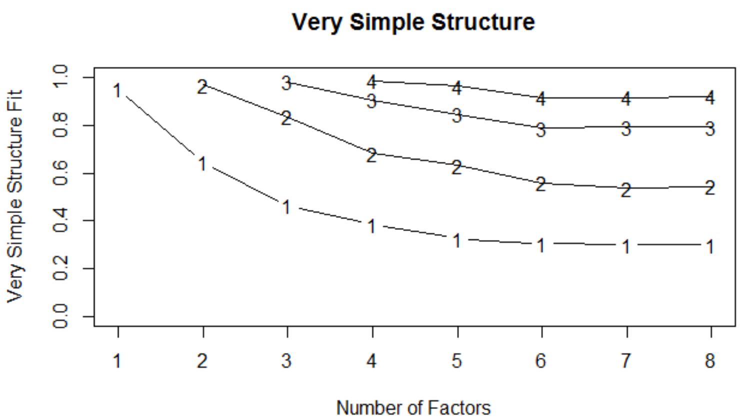


Figure 9: *Factors to retain by the rule VSS. Source: own elaboration.*

The method used for the extraction of factors is the main factor, in which the factorial matrix is extracted with the property that factors explain the maximum variance and are additionally uncorrelated; a common factor underlying to the variables is assumed in this, that's why the extraction of the maximum variance in each factor is pretended, so at the end the k resulting factors explain the greatest amount of common variance, although the charges of extracted factors do not differ substantially from the method of the principal component (Rietveld & Van Hout 1993).

The Varimax rotation was applied to this method, with the aim of maximizing the weights of each factor, and hoping so that each variable is well represented in only one of them and at the same time the maximum number of variables correlated to each factor is minimized. Resulting analysis extracting the four factors through the mentioned method is shown in tables 5, 6 and 7:

Table 5: *Results of the exploratory factorial analysis. Source: own elaboration.*

Ítem	**FAC1**	**FAC2**	**FAC3**	**FAC4**	**Communality**	**Unicity**
antigue		0,78			0,75	0,25
comporta		0,78			0,75	0,25
nuevo		0,73			0,68	0,32
cupos		0,73			0,66	0,34
acompa		0,72			0,72	0,28
requisit		0,67			0,65	0,35
trami		0,67			0,65	0,35
claro			0,79		0,81	0,19
interes			0,76		0,82	0,18
ag			0,76		0,8	0,2
soluc			0,74		0,76	0,24
respet			0,7		0,69	0,31
espera	0,46		0,49		0,64	0,36
tiempo	0,79				0,81	0,19
personal	0,71				0,73	0,27
agili	0,66				0,71	0,29
manejo	0,65				0,7	0,3
senaliza				0,65	0,66	0,34
seguridad				0,6	0,56	0,44
horario				0,57	0,51	0,49
comodidad				0,57	0,63	0,37
ubicacion				0,55	0,44	0,56
amabi				0,45	0,58	0,42

Table 6: *Statistics of summary. Source: own elaboration.*

Item	**FAC1**	**FAC2**	**FAC3**	**FAC4**
SS loadings	4,99	4,26	3,32	3,14
Proportion Var	0,22	0,19	0,14	0,14
Proportion Explained	0,32	0,27	0,21	0,2
Correlation of scores with factors	0,94	0,94	0,9	0,84
Multiple R square of scores with factors	0,88	0,88	0,82	0,7

In the analysis of the exploratory stage a total of four factors that pick up a 68% of variance were founded, such factors generate the underlying dimensions of the service according to the evaluations made to users. These dimensions clearly differentiate each of the moments of attention that occur in service companies.

First factor: The first factor is formed by a total of seven variables with weightings greater tan 0,65; in turn these variables do not

Table 7: *Dimensions and associated variables. Source: own elaboration.*

Dimension	**Variable**
Behavior	antigue: the recognition of seniority as customer comporta: recognition of the good use of the service. nuevo: offering of new products. cupos: quotas granted acompa: the service after sale requisit: requirements to acquire the service trami: agility in procedures
Advisors	claro: clarity of information provided by the advisor. interes: interest shown by adviser ag: agility in the response received soluc: effective solution of needs respet: kindness and respect of adviser espera: wait time for adviser
Cash registers	tiempo: wait time in the line. personal: number of people serving agili: agility of the person serving in cash registers manejo: suitable handling of lines for customers
Offices	senaliza: signaling inside the offices seguridad: security feeling in the offices horario: customer service hours in offices comodidad: the comfort of the offices location: location of offices in the place where it is needed amabi: friendliness and respect of people who serves

reach in other factors importances greater than 0.25, all these variables refer to the satisfaction with the behavior of services offered by the company.

Second factor: It consists of six variables associated to the personal attention given by the advisers of service whose weightings are greater tan 0,49.

Third factor: It consists of four variables with weightings greater tan 0,65 and it is formed by the variables that evaluate the received attention in cash registers.

Fourth factor: It is formed by six variables that have weightings greater than 0,45 and make the underlying dimension of attentions in offices.

6.5 Specification stage

This section aims to formally establish the model; four configurations of interdependence were previously established, in which factors are latent measures of the satisfaction with each one of the moments of the service and in which the existence of a linear relation between the factors and the manifest variables is assumed. So the first factor (Offices) takes the shape presented in figure 10.

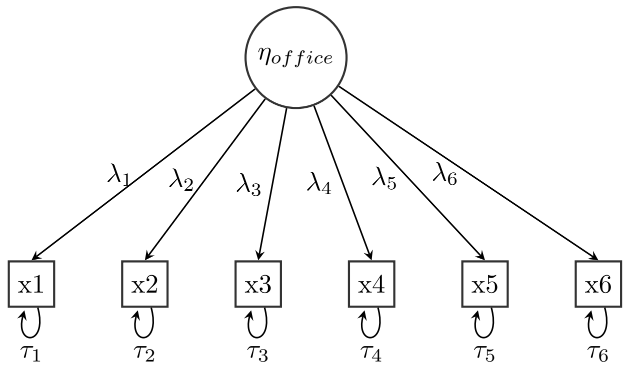


Figure 10: *Path Diagram factor 1 -Offices. Source: own elaboration.*

Path Diagram expresses the direct effect that has the underlying variable of satisfaction with offices on the manifest variables that can be expressed in the measure model:

$$\mathbf{X} = \Lambda\eta + \tau \tag{5}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \end{bmatrix} \begin{bmatrix} \eta \end{bmatrix} + \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \\ \tau_5 \\ \tau_6 \end{bmatrix} \tag{6}$$

Where η is the latent variable Offices, Λ is the matrix of unknown effects to estimate, τ is the matrix of measurement error and finally $\mathbf{X}$ the matrix of observed variables, where x_1 = senaliza, x_2=seguridad, x_3=horario, x_4=comodidad, x_5=ubicacion y x_6=amabi. The specification of the other factors is made similarly. To end the specification stage it's necessary to take into account that previous latent factors determined in exploratory way have effect (direct and indirect) on the satisfaction of service, because at this stage the following structural hypothesis are defined:

In order to finish the specification stage, it should not be forgotten that the previously identified latent factors have an effect (direct and indirect) on the satisfaction of the service, so at this stage the following structural assumptions are defined:

> H_1: Underlying factor offices has a significant effect on the satisfaction with the service, it is, $\eta_{offices} \neq 0$.

H_2: Underlying factor cash registers has a significant effect on the satisfaction with the service, it is, $\eta_{cash\ registers} \neq 0$.

H_3: Underlying factor advisers has a significant effect on the satisfaction with the service, it is, $\eta_{advisers} \neq 0$.

H_4: Underlying factor behavior has a significant effect on the satisfaction with the service, it is, $\eta_{behavior} \neq 0$.

H_5: Overall satisfaction has a significant effect on the loyalty with the brand, it is, $\beta_{sat_{gen}} \neq 0$.

H_6: Recommendation has a significant effect on the loyalty with the brand, it is, $\beta_{recom} \neq 0$.

H_7: The intention of buying again has a significant effect on the brand, it is, $\beta_{intenc} \neq 0$.

In this case the overall satisfaction is given by the present diagram in figure 11:

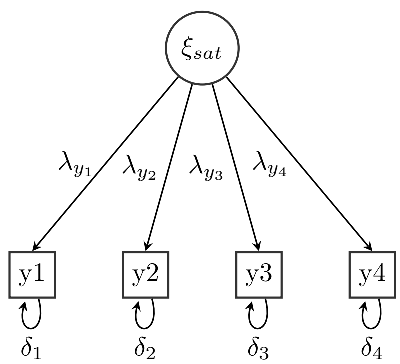


Figure 11: *Path Diagram factor Satisfaction. Source: own elaboration.*

And the equation representing its model of measure expresses as it follows:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} \lambda_{y_1} \\ \lambda_{y_2} \\ \lambda_{y_3} \\ \lambda_{y_4} \end{bmatrix} [\xi] + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{bmatrix} \tag{7}$$

Where ξ is the latent variable of overall satisfaction, Λ is the matrix of unknown effects to estimate, δ is the matrix of measurement error and finally $\mathbf{Y}$ is the matrix of unknown effects to estimate, where y_1= physical plant, y_2= financial area, y_3= personal attention, y_4= variety of products. In figure 12 we can see the path of the whole model's specification:

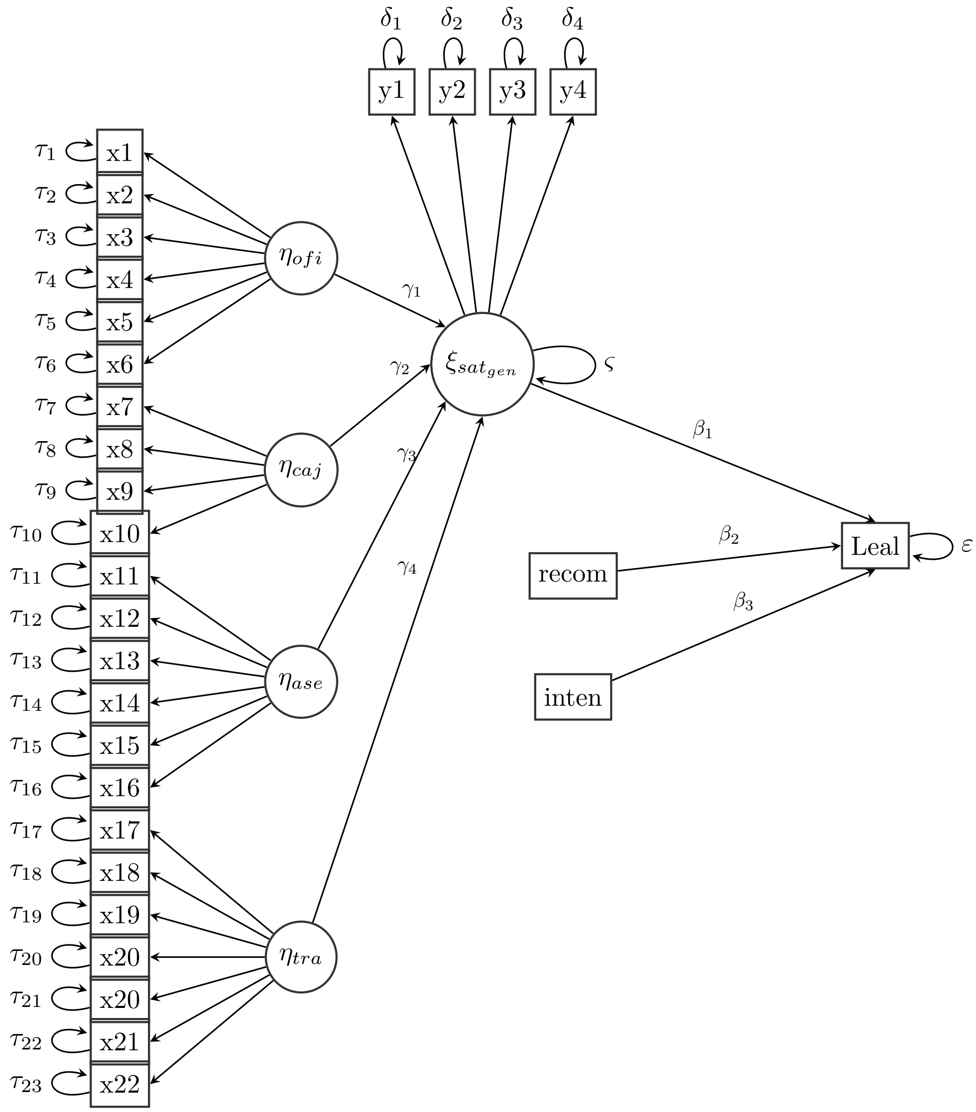


Figure 12: *Path Diagram specification stage. Source: own elaboration.*

And so the structural model is given by the expression:

$$\left[\xi\right] = \begin{bmatrix}\gamma_1 & \gamma_2 & \gamma_3 & \gamma_4\end{bmatrix}\begin{bmatrix}\eta_1\\ \eta_2\\ \eta_3\\ \eta_4\end{bmatrix} + \left[\varsigma\right] \tag{8}$$

And finally:

$$Leal = \begin{bmatrix}\beta_1 & \beta_2 & \beta_3\end{bmatrix} \begin{bmatrix}\xi \\ recom \\ inten\end{bmatrix} + \begin{bmatrix}\varepsilon\end{bmatrix} \tag{9}$$

At first glance analyzing the degrees of freedom of the model, this is overidentified:

$$dof = m(m+1)/2 - 2m - \xi(\xi - 1)/2 = 400 > 0$$

It also meets the rule t-rule with which we can verify that we are not facing a non-identified model:

$$t \leq \left(\frac{1}{2}\right)(p+q)(p+q+1) \tag{10}$$

6.6 Estimation stage

Given the non-normality of variables under study and its ordinal character, it is chosen for estimation a technique that does not make any assumption about the distribution of variables, so that the kurtosis presented does not affect the standard errors of estimation.

This method minimizes the function:

$$F_{DWLS} = [\hat{\rho} - \rho(\theta)]' \, diag(W_{\rho\rho})^{-1}[\hat{\rho} - \rho(\theta)] \tag{11}$$

Since these methods seek to reduce the distance between the observed variances and the modelled design through the matrix $\mathbf{W}$, calculated by the weighted least squares (WLS), this will be precisely the estimation method chosen for the model, particularly the robust method ($DWLS$) based on polychoric correlations of observed variables and give the results observed in Figure 13 and table 8:

Results in figure 13 show that recommendation has a bigger effect on loyalty than the produced by the buyback or the satisfaction with the service; however there is a problem of model's specification, since the variance of the factor overall satisfaction is negative, therefore it must be specified again to avoid this problem. It can be observed in table 8 that in general the variables entered in the model are statistically significant in the calculation of each factor.

The overall effect of each variable is calculated multiplying its estimations, following in the diagram the route that connects the variable in question with the loyalty. Calculated effects are shown in table 9.

Initially it's observed that among those variables that affect loyalty indirectly, is the relationship with the advisers and the factor offices the ones that have the greatest indirect effect, while in the level of direct effects is the recommendation the one that has greater significance in the measure of loyalty.

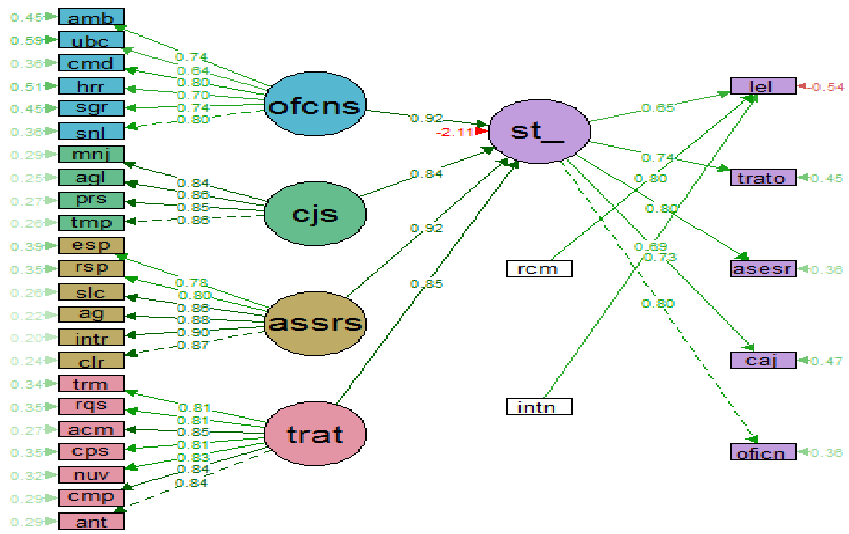


Figure 13: *Initial estimation of model. Source: own elaboration.*

Table 9: *Effect on loyalty. Source: own elaboration.*

Variable	Estimate	Indirect effec.	Direct effec.	Total
offices	0,924	0,603372	0	0,603372
cash registers	0,836	0,545908	0	0,545908
advisers	0,919	0,600107	0	0,600107
trat	0,848	0,553744	0	0,553744
sat_{gen}	0,653	0	0,653	0,653
intenc	0,688	0	0,688	0,688
recom	0,803	0	0,803	0,803

6.7 Diagnosis of fit

A condition of the models of structural equations is that the quality of their adjustment is measured through the capacity that have the estimations to reproduce the sample matrix of covariances, or the matrix of correlations in the case of standardized estimations, which is derived through $\Sigma(\theta) = \Sigma(\hat{\theta})$ where $\hat{\theta} = (\Lambda, \eta, \xi)$ is the vector of parameters that relates the covariances between the variables and the parameters. If the model is well adjusted, the residual matrix must be the closest to the null matrix; to verify the quality of this adjustment it's necessary to do the relevant diagnosis.

Table 8: *Initial estimations. Source: own elaboration.*

Variable/factor	**Estimate**	**Std.err**	**Z-value**	**P($>$ $\lvert z\rvert$)**	**Std.lv**	**Std.all**
Regressions:						
sat_gene						
oficinas	1,042	0,017	60,355	0	0,924	0,924
cajas	0,59	0,009	63,886	0	0,836	0,836
asesores	0,885	0,014	65,284	0	0,919	0,919
trat	0,545	0,007	83,501	0	0,848	0,848
lealtad						
intenc	0,585	0,02	29,922	0	0,585	0,688
recom	0,761	0,028	27,443	0	0,761	0,803
sat_gene	0,958	0,009	105,588	0	1,377	0,653
trat =						
antigue	1				2,237	0,842
comporta	0,956	0,011	85,695	0	2,138	0,845
nuevo	0,922	0,011	85,2	0	2,062	0,828
cupos	0,845	0,01	82,581	0	1,89	0,807
acompa	0,968	0,011	86,102	0	2,165	0,855
requisit	0,753	0,009	82,184	0	1,683	0,807
trami	0,744	0,009	81,573	0	1,665	0,812
asesores =						
claro	1				1,492	0,872
interes	1,133	0,017	65,851	0	1,69	0,895
ag	1,104	0,017	65,858	0	1,647	0,883
soluc	1,148	0,017	65,895	0	1,712	0,863
respet	0,747	0,012	61,478	0	1,115	0,804
espera	1,148	0,017	66,513	0	1,713	0,78
cajas =						
tiempo	1				2,036	0,863
personal	0,953	0,014	68,827	0	1,94	0,852
agili	0,824	0,012	66,645	0	1,678	0,864
manejo	0,871	0,013	67,399	0	1,773	0,843
oficinas =						
senaliza	1				1,275	0,802
seguridad	0,899	0,015	59,536	0	1,146	0,741
horario	0,92	0,015	59,539	0	1,172	0,699
comodidad	1,123	0,018	61,802	0	1,431	0,802
ubicacion	0,82	0,014	58,543	0	1,045	0,64
amabi	0,823	0,014	58,096	0	1,049	0,741
sat_gene =						
planta_fisica	1				1,437	0,801
financiera	0,864	0,008	109,268	0	1,243	0,73
atencion	0,968	0,009	112,323	0	1,391	0,802
variedad	1,004	0,009	112,506	0	1,444	0,745

Several tests have been designed for this, they obey to different motivations, some of them contrast the fit of $\Sigma(\theta)$, some others verify the parsimony model and other compare the adjusted model with the saturated based model.

However, there are different situations that can alter its result, therefore not all tests can always be used; in this paper 1000 simulations by Bootstrap at different sample levels (100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500, 7000, 7500, 8000) were done to prove its stability (figure14) before choosing some test to do the diagnosis of model's adjustment.

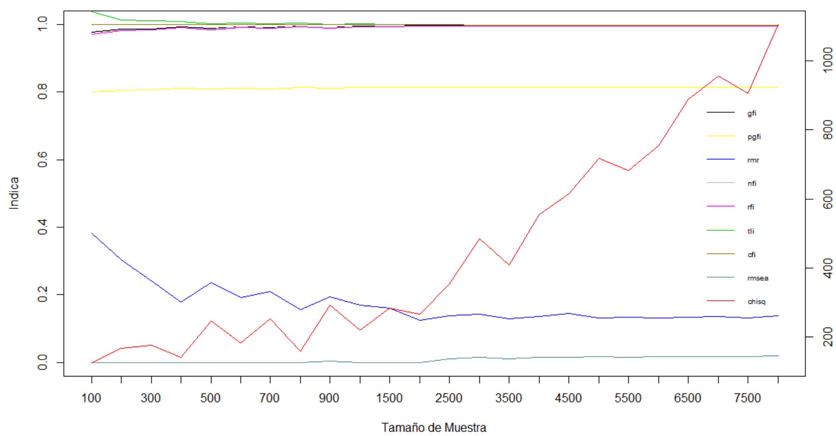


Figure 14: *Stability of adjustment at different simple levels. Source: own elaboration.*

Figure 14 shows that the χ^2 test in particular increases its statistical as the sample size grows, leading always to rejection for big n; this result is in line with the researches that concluded that for big samples small variations between the matrices of sampling and estimated covariances are detected as significant (Bentler & Bonett 1980).

It happens the opposite with the test RMR, since as the sample increases, its estimations decay, finding its stability for sample sizes above 2.000 cases; test TLI, RFI and GFI meanwhile reach stability in sample sizes of 500 or 1000 elements. Tests CFI, RMSEA and PGFI are the most stable and contrast the fit of $\Sigma(\theta)$, of model's residues and the contrast against the saturated model[2].

Table 10: *Indices of goodness of fits. Source: own elaboration.*

Test	**Statistic**	**Optimal fit value**	**Result**
χ^2	228551,8671	$\approx 0,0$	Bad fit
cfi	0,4568	≈ 1	Bad fit
rmsea	0,2624	$\leq 0,1$	Bad fit
pgfi	0,4522	≈ 1	Bad fit

[2]For a detailed illustrations of tests see Romero (2015).

When analyzing the most stable tests of model adjustment presented in table 10, we find that the initial model does not show an accurate adjustment of the sampling matrix of covariances.

6.8 Modification of the model

The results of this model show a bad adjustment, modifications will be done in the search of a better one; such modifications pretend to set free parameters that were set at zero (for example, the covariance between two variables). We used the test of multipliers of Lagrange (modification indices) in order to maximize the adjustment of the model; this test allows establishing the modifications that will contribute with a bigger change in the statistics of χ^2 resulting of the liberation of these parameters restricted to zero. However, to establish these modifications depends basically on the implications that such change has on the theory of the model, so not all changes are viable, or even, some changes may report variations in the theory of the model.

These modifications are included one by one and after each new change the adjustment is reviewed again to find the right one. The contrast between each model will be done through the statistical LR test, which allows showing the difference between the log- verisimilitudes of the models (initial and modified) through the statistics:

$$LR = -2(l(\pi^0|y) - l(\pi^1|y)) \sim \chi^2_1 \tag{12}$$

Modifications are stopped at the time that each liberation of parameters made does not generate big changes of χ^2 between nested models (seventh model in this case); in figure 15 it's observed that from the model number seven, the new nested models do not generate greater loss of χ^2 compared to the previous model, therefore these seven changes, which consist of the liberation of the parameters of covariance between latent factors indirectly affecting the overall satisfaction, will be introduced to the initial model.

6.9 Final model

Verifying the test of goodness of fit 11, we see that it finally reached a good model, both in total explained variability as at the level of parsimony and minimal residues (table 11).

Table 11: *Indices of goodness of final fit. Source: own elaboration.*

Test	**Statistic**	**Optimal fit value**	**Result**
chisq	46935,5		
cfi	0,9	≈ 1	Good fit
rmsea	0,1	$\leq 0,1$	Good fit
pgfi	0,8	≈ 1	Good fit

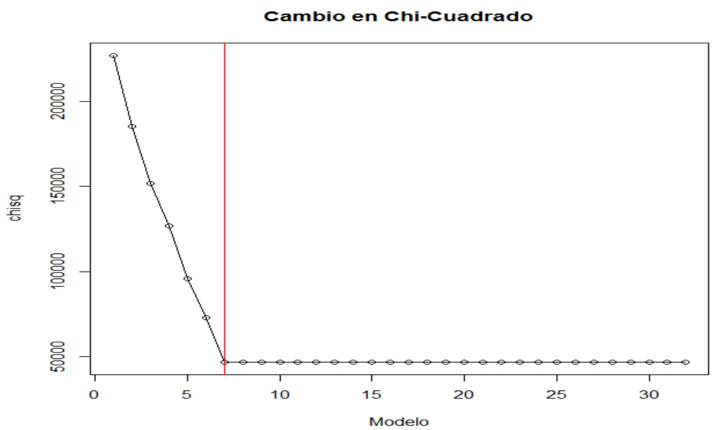


Figure 15: *Modification indices-Changes in* χ^2*. Source: own elaboration.*

Estimations of parameters in figure 16 allow observing that the dimensions offices and advisers are the ones that have the greatest effect on the overall satisfaction of customers. At the same time that satisfaction affects more the loyalty, which is reflected in the coefficient that this variable has on the variable response. It is also observed that the biggest covariance is among the factors offices and cash registers, so a possible adjustment of the model that takes into account these two factors as a single one could also be considered.

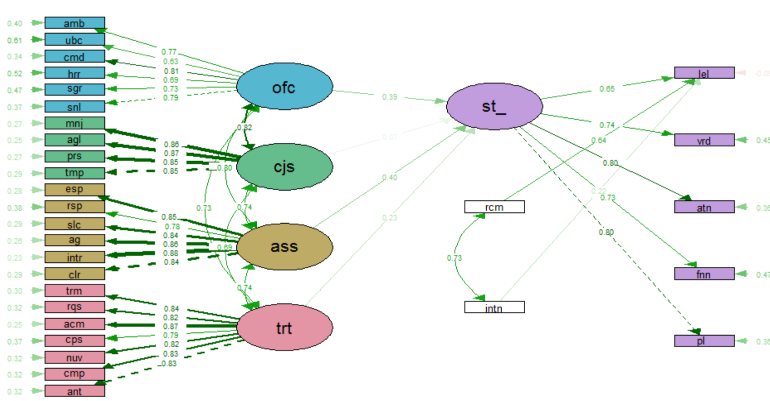


Figure 16: *Path Diagram fitted model. Source: own elaboration.*

Results in 12 and 13 allow proving the hypothesis system, concluding that underlying factors specified in the model have a significant effect on the satisfactions, as also each one of the variables affecting the loyalty of a customer is significant.

It is also observed that the relation of covariation between the component factors of the service is significant, which means that its study should not be done in an isolated way as it would be done in any model of multiple regression, but the existing correlations between the different dimensions of the service must be taken into account, since some affect others.

The expression of the initial model of measure is the following:

$$\mathbf{X} = \Lambda\eta + \tau \tag{13}$$

$$\begin{bmatrix} antigue \\ comporta \\ nuevo \\ cupos \\ acompa \\ requisit \\ trami \\ claro \\ interes \\ ag \\ soluc \\ respet \\ espera \\ time \\ personal \\ agili \\ manejo \\ senaliza \\ seguridad \\ horario \\ comodidad \\ ubicacion \\ amabi \end{bmatrix} = \begin{bmatrix} 0,826 & 0 & 0 & 0 \\ 0,827 & 0 & 0 & 0 \\ 0,822 & 0 & 0 & 0 \\ 0,792 & 0 & 0 & 0 \\ 0,868 & 0 & 0 & 0 \\ 0,825 & 0 & 0 & 0 \\ 0,835 & 0 & 0 & 0 \\ 0 & 0,841 & 0 & 0 \\ 0 & 0,877 & 0 & 0 \\ 0 & 0,862 & 0 & 0 \\ 0 & 0,841 & 0 & 0 \\ 0 & 0,785 & 0 & 0 \\ 0 & 0,85 & 0 & 0 \\ 0 & 0 & 0,845 & 0 \\ 0 & 0 & 0,854 & 0 \\ 0 & 0 & 0,866 & 0 \\ 0 & 0 & 0,857 & 0 \\ 0 & 0 & 0 & 0,793 \\ 0 & 0 & 0 & 0,729 \\ 0 & 0 & 0 & 0,692 \\ 0 & 0 & 0 & 0,812 \\ 0 & 0 & 0 & 0,625 \\ 0 & 0 & 0 & 0,773 \end{bmatrix} \begin{bmatrix} \eta_{trato} \\ \eta_{asesores} \\ \eta_{cajas} \\ \eta_{oficinas} \end{bmatrix} + \begin{bmatrix} 0,317 \\ 0,317 \\ 0,325 \\ 0,373 \\ 0,247 \\ 0,32 \\ 0,302 \\ 0,294 \\ 0,232 \\ 0,258 \\ 0,292 \\ 0,384 \\ 0,277 \\ 0,285 \\ 0,27 \\ 0,25 \\ 0,265 \\ 0,372 \\ 0,469 \\ 0,521 \\ 0,341 \\ 0,609 \\ 0,402 \end{bmatrix}$$

While the structural model is:

$$\left[\xi_{satisfacción}\right] = \begin{bmatrix} 0,369 & 0,292 & -0,098 & 0,231 \end{bmatrix} \begin{bmatrix} \eta_{trato} \\ \eta_{advisers} \\ \eta_{cash\ registers} \\ \eta_{offices} \end{bmatrix} + \left[0,025\right] \tag{14}$$

$$Leal = 0,65 * satisfaction + 0,64 * recomendation + 0,22 * intention - 0.093 \tag{15}$$

Table 12: *Estimations of fitted model. Source: own elaboration.*

Variable/factor	**Estimate**	**Std.err**	**Z-value**	**P($>$ $\|z\|$)**	**Std.lv**	**Std.all**
Regressions:						
sat_gene ∼						
offices	0,442	0,041	10,908	0	0,388	0,388
cash register	0,05	0,019	2,586	0,01	0,07	0,07
advisers	0,395	0,025	15,796	0	0,395	0,395
trat	0,149	0,01	14,335	0	0,227	0,227
leal ∼						
intenc	0,185	0,053	3,516	0	0,185	0,217
recom	0,611	0,062	9,822	0	0,611	0,645
sat_gene	0,957	0,009	105,589	0	1,376	0,652
Latent variables:						
asesores =∼						
claro	1				1,438	0,841
interes	1,151	0,011	108,845	0	1,655	0,877
ag	1,118	0,01	109,03	0	1,607	0,862
soluc	1,161	0,011	108,702	0	1,67	0,841
respet	0,757	0,007	101,08	0	1,088	0,785
espera	1,3	0,012	112,387	0	1,869	0,85
trat =∼						
antigue	1				2,195	0,826
comporta	0,953	0,008	122,828	0	2,092	0,827
nuevo	0,933	0,008	123,158	0	2,048	0,822
cupos	0,845	0,007	118,678	0	1,855	0,792
acompa	1,002	0,008	125,176	0	2,199	0,868
requisit	0,783	0,007	119,869	0	1,72	0,825
trami	0,78	0,007	119,038	0	1,713	0,835
cajas =∼						
tiempo	1				1,995	0,845
personal	0,975	0,008	123,324	0	1,944	0,854
agili	0,843	0,007	118,633	0	1,682	0,866
manejo	0,904	0,008	120,148	0	1,803	0,857
oficinas =∼						
senaliza	1				1,26	0,793
seguridad	0,895	0,009	102,826	0	1,127	0,729
horario	0,921	0,009	102,431	0	1,16	0,692
comodidad	1,15	0,011	107,947	0	1,449	0,812
ubicacion	0,81	0,008	99,617	0	1,021	0,625
amabi	0,869	0,009	101,832	0	1,095	0,773
sat_gene =∼						
planta_fisica	1				1,438	0,801
financiera	0,865	0,008	109,323	0	1,244	0,731
atencion	0,967	0,009	112,364	0	1,39	0,801
variedad	1,004	0,009	112,53	0	1,444	0,745

6.10 Impact of the components of service on the loyalty

For the estimation of the impact that each one of the dimensions of the service has on the customer's loyalty, it's used the total effect of the variables, which is

Table 13: *Estimations of fitted model (cont.). Source: own elaboration.*

Variable/factor	Estimate	Std.err	Z-value	P(> $\|z\|$)	Std.lv	Std.all
Covariances:						
trat ∼∼						
asesores	2,329	0,022	106,305	0	0,738	0,738
oficinas	2,012	0,019	103,253	0	0,728	0,728
cajas	3,008	0,028	108,676	0	0,687	0,687
asesores ∼∼						
oficinas	1,442	0,015	94,379	0	0,796	0,796
cajas	2,109	0,021	99,596	0	0,736	0,736
cajas ∼∼						
oficinas	2,07	0,021	98,007	0	0,824	0,824
intenc ∼∼						
recom	4,03	0,103	39,127	0	4,03	0,729

calculated as the sum of the direct and indirect effects.

Table 14: *Final effect on the loyalty of the customer. Source: own elaboration.*

Variable	Estimate	Indirect effec.	Direct effec.	Total
oficinas	0,388	0,252976	0	0,252976
cajas	0,07	0,04564	0	0,04564
asesores	0,395	0,25754	0	0,25754
trat	0,227	0,148004	0	0,148004
sat$_{gen}$	0,652	0	0,652	0,652
intenc	0,217	0	0,217	0,217
recom	0,645	0	0,645	0,645

Table 14 shows that in the dimensions of satisfaction, the ones that have the biggest effect are advisers and offices, while cash registers have the lower impact on the satisfaction. Globally the recommendation has the biggest effect on loyalty.

6.11 Calculation of loyalty indicator

Different approaches can be used to calculate the indicator, however, the purpose of this work is to extract all the information derived as base for calculation. It is proposed to make the calculation of the loyalty indicator using the total effect of each one of the components of the service as weights that weigh up the indicator, such that:

$$Lealtad = \sum_{i=1}^{3} w_i v \tag{16}$$

Where w_i is a vector of weightings built with the total effects of the variables *recom*, *intenc* and the factor sat_{gen}, such that

$$w_i = \frac{\beta_i}{\sum_{i=1}^{3} \beta_i} \qquad \sum_{i=1}^{3} w_i = 1 \tag{17}$$

and υ is a matrix formed by the variables of overall satisfaction (sat_{gen}), recommendation ($recom$) and repurchase intention ($intenc$).

Then the mathematical expression of the loyalty indicator would be:

$$I_{Loyalty} = w_1 * sat_{gen} + w_2 * recom + w_3 * intenc \tag{18}$$

$$I_{Loyalty} = 0,431 * sat_{gen} + 0,143 * recom + 0,426 * intenc \tag{19}$$

7 Conclusions

The conclusions are a summary of what was already mentioned along this paper; however synthesizing the following can be concluded:

> Regarding the best method of estimation it is found that the function of adjustment by diagonalized weighted least squares (DWLS) is the one that fits the best to data, since it is a function free of distribution and it additionally uses the matrix of polychoric correlations for ordinal variables in the estimation stage.
>
> The overall satisfaction with the service and the recommendation of this are the variables that affect the most on customer's loyalty. It can also be noted that are the constructs of service related to the advisors and the offices the ones that have the greatest effect on the overall satisfaction with the service.
>
> The matrix of action shows that the attributes forming the factors advisers and offices are the best valued and also the ones that have the biggest effect on the loyalty; however the attribute wait is rated below the average, so, in order to improve the loyalty we should aim to raise its average qualification. The attributes forming the factor cash registers generally present a satisfaction below the average and are the ones that least effect have on loyalty.

In addition, during the development of this work other concerns arose such as the behavior of the indicators of the model's adjustment and how to improve the adjustment of this, in this respect we can conclude that:

> In these situations the matrix of polychoric correlations should be used; however, most commercial packages has not yet widespread its use, therefore the realization of a factorial exploratory analysis could be questioned. In this work it's concluded that for scales type Likert of 10 points the matrix of polychoric correlations is very closet to the matrix of correlation of Pearson, both in the value of correlation as in its behavior, therefore in these cases it is valid to use the matrix of Pearson to make the factorial analysis in ordinal variables.

As statistics of diagnosis for the model's adjustment it was found that despite being the most popular, measure χ^2 is not robust because it is directly affected by the sample size, increasing its value as n increases. In relation to RMR, GFI, NFI, RFI and TLI measures, they find stability in samples over 1000 cases, so in many cases it will not be permissive to use them, given the costs of raising large samples. It should be clarified that this analysis was performed on the estimation method DWLS, so in other estimation methods this behavior is unknown.

On the practical part one should be rigorous with the descriptive analysis and validation of data assumptions, because although it is not directly shown in this work, when analyzing the adjustment functions can be easily understood that the values of the parameters change depending on the selected function; but you should also take into account the size of the sample as free distribution methods require a large amount of sample to make the estimates.

Received: 3 November 2016
Accepted: 5 April 2016

References

Batista, J. & Coenders, G. (2000), *Modelos de Ecuaciones Estructurales*, 2 edn, Editorial La Muralla, Madrid, Spain.

Bentler, P. M. & Bonett, D. G. (1980), 'Significance test and goodness of fit in the analysis of covariance structures', *Psychological Bulletin* **88**(3), 588–606.

Coenders, G., Satorra, A. & Saris, W. E. (1979), 'Alternative approaches to structural modeling of ordinal data: A Monte Carlo study', *Structural Equation Modeling: A Multidisciplinary Journal* **4**(4), 261–282.

Field, A. (2000), *Discovering Statistics using SPSS for Windows*, 1 edn, Sage publications, Thousand Oaks, CA.

Gelin, M., Beasley, T. & Zumbo, B. (2003), 'What is the impact on scale reliability and exploratory factor analysis of a Pearson correlation matrix when some respondents are not able to follow the rating scale?'.
*http://faculty.educ.ubc.ca/zumbo/aera/papers/GelinBeasleyZumbo-7Apr.pdf

Hoyle, R. H. (2012), *Handbook of Structural Equation Modeling*, 1 edn, The Guilford Press, New York.

Loehlin, J. C. (2004), *Latent Variable Models*, 4 edn, Laurence Erlbaum Associates Publishers, New Jersey.

Olsson, U. H. (1979), 'Maximum likelihood estimation of the polychoric correlation coefficient', *Psychometrika* **44**(4), 443–460.

Rietveld, T. & Van Hout, R. (1993), *Statistical Techniques for the Study of Language and Language Behaviour*, 1 edn, Mouton de Gruyter, Berlin.

Romero, G. R. (2015), 'Modelo de Lealtad a partir de un análisis de Ecuaciones Estructurales', *Tesis de Pregrado. Universidad Santo Tomás* .

Saurina, C. (1997), 'Evaluación de un modelo de medida de la calidad en el sector servicios. Las entidades financieras de la comarca de Girona', *Annual meeting of the American Educational Research Association (AERA)* **39**, 219–248.

Streiner, D. (2003), 'Starting at the beginning: an introduction to coefficient alpha and internal consistency', *Journal of personality assessment* **8**(1), 99–103.