EXTERNAL ANALYSIS IN PLS-PATH MODELING FOR THE EVALUATION OF THE PASSANGER SATISFACTION

Anna CRISCI
Second University of Naples

E-mail: crisci.anna@virgilio.it

Antonello D'Ambra
Second University of Naples

E-mail: antonello.dambra@unina2.it

ABSTRACT:

In recent years the need to verify the degree of user satisfaction and the quality of services provided has become a priority for transportation companies. The requirement is not only to move towards higher quality of services provided, but also to strengthen confidence in the companies that provide the service.

This paper shows a study to evaluate the overall passenger satisfaction of public transport service. To explore the collected data, PLS-Path Modeling approach has been adopted, and later the analysis of the data has been further supported by means of the joint use of the PLS-PM and the External Analysis Method, catching the advantages of both.

In particular, the additional element that the present contribution wants to highlight is the use of so-called external information. In fact, passenger satisfaction may be influenced by external factors such as sex, age, educational level, profession, etc.

Therefore, the joint use of the PLS-PM and the External Analysis Method could help the researcher interpret the results objectively.

Key words: *External Analysis, Structural Equation Modeling, PLS- Path Modeling, Passenger Satisfaction.*

1. INTRODUCTION

The activities of Passenger Satisfaction (PS) have a more and more strategic role for the achievement of the objectives of the leader company in the supply of a services [3].

The "total quality" of the services and the full satisfaction of all the users are often unattainable with limited resources. Therefore, it is important to set a goal of possible quality, and an interesting perspective could be "***The satisfaction of users concerning their most important expectations***". The main problem of this approach is that the expectations modify over the time, often while the service is carried out. The PS must be managed along with the service cycle; moreover, in order to estimate the quality of the services supplied, the *so-called "user feedback"* is important [24]. The intangible part in the performance service cannot be subject to a testing or check as in the case of tangible goods. In other words, to measure an effective performance we have to refer to the people involved in it, that is to users who are the only ones that actually test the service, with reflections both on the efficiency of the system and on the quality of the performance.

It is, therefore, essential that the activities of PS:

- are aligned with the more general needs of the company;
- meet the user needs;
- supply elements that allow to improve systematically and competitively organizational structure, directing it to the PS.

In this paper a study of PS evaluation in the transport service in Benevento (Italy) is shown.

In 2011 to estimate the PS a survey was carried out and, a questionnaire, agreed with the responsible of the service, composed by analytical sections of interest (***Quality Factors***) was arranged. The method of extraction of the data is the sample random with a sampling bias and a confidence level, 4.5% and 95%, respectively. The interviewed sample of 400 units is determined by a population of 9,000[1] users daily.

The method of extraction of the data is the sample random with a sampling bias and a confidence level, 4.5% and 95%, respectively. The interviewed sample of 400 units is determined by a population of 9,000[1] users daily.

The five dimensions of the quality are:

- accessibility to the service (A.S);
- condition of the bus (C.B);
- security on board (S.B);
- reliability of the service (R.S);
- overall passenger satisfaction (O.P.S);

To explore the collected data, PLS-Path Modeling approach [4,5,8,21,22] has been adopted, and later the analysis of the data has been further supported by means of the joint use of the PLS-PM and the External Analysis Method [15], catching the advantages of both.

The assumed model is composed of five latent variables and seventeen manifest variables:

[1] The population of 9,000 users was determined by a survey concerning the adoption of the service that the public service company had carried out previously.

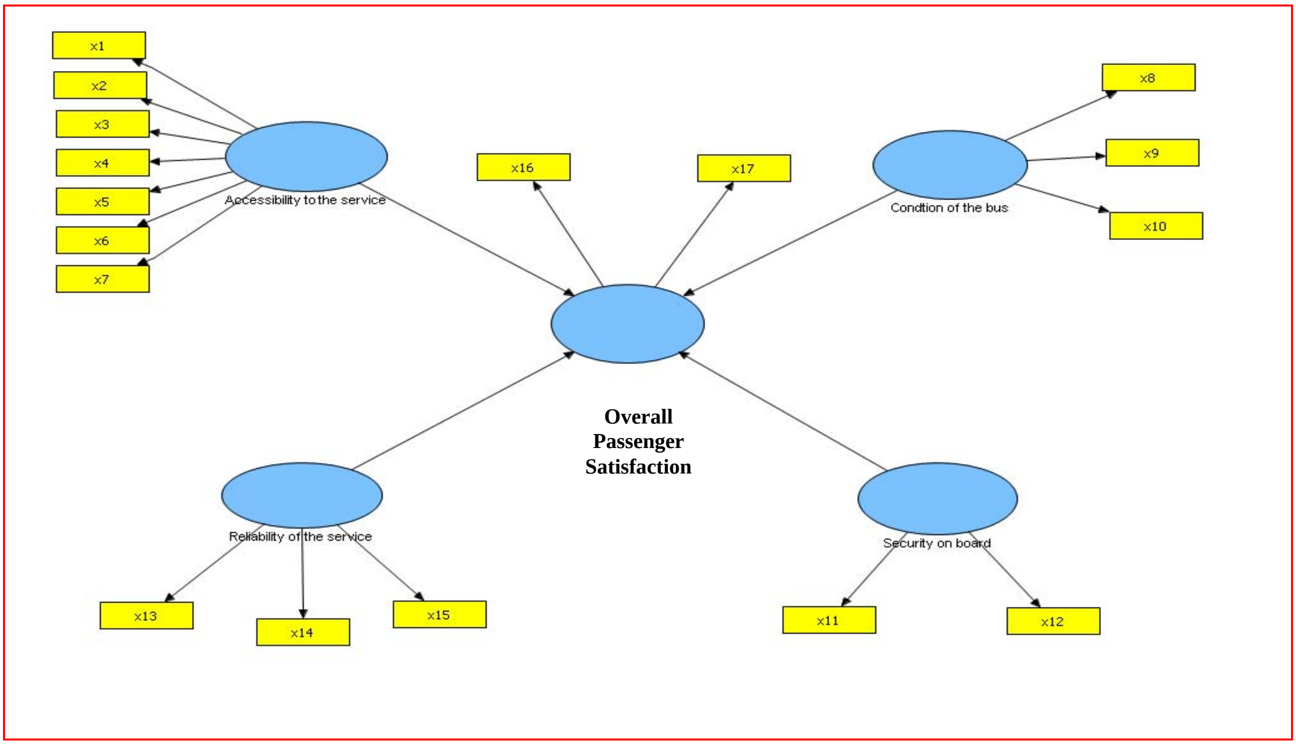

Figure 1: The Theoretical Model for the evaluation of the Passenger Satisfaction of the Public Transport Service

The PLS Path Modeling is a statistical method which was developed for the Analysis Structural Models with latent variables. One of the difficulties for a researcher in the economic-social sciences in the specification of a statistical model describing the casual-effect relationships between the variables derives from the fact that the variables which are object of the analysis are not directly observable (*i.e. latent*), for example, the performance, the passenger satisfaction, the social status etc. Although such latent variables (*LVs or latent construct*) cannot be directly observable, the use of proper indicators (*i.e. manifest variables, MVs*) can make the measurement of such constructs easy.

As opposed to the covariance-based approach or LISREL [2,10,11,12,13,14], the aim of the PLS is to obtain the scores of the latent variables for predicted purposes without using the model to explain the covariation of all the indicators [9,23]. For example, a researcher may be interested in what dimensions of the service quality can more influence the PS.

The joint analysis of the original data with external information (e.g. socio-demographic features of the subjects: sex, age, education, job, etc.) can lead the researcher to a more objective interpretation of the obtained results. In other words, the role of the external information may be that of establishing, for example, the existence of a relationship between the assumed links and the demographic information on the subjects.

The paper is organized as follows: in sections 2 and 3, the PLS-PM approach and the External Analysis are shown, respectively. In section 4, the estimations of the parameters, the general results of the Path-model, and the results of the joint analysis of the PLS-PM and External Analysis are shown.

2. The PLS-Path Modeling Approach

PLS –Path Modeling aims to estimate the relationships among J blocks of variables, which are expression of unobservable constructs. Specifically, PLS-PM estimates the network of relations among the manifest variables and their own latent variables, and the latent variables inside the model through a system of interdependent equations based on simple and multiple regression. Formally, let us usually assume K variables observed on N units. The resulting data x_{nkj} are collected in a partitioned table of standardized data $\mathbf{X}$:

$$X = [X_1, \dots, X_j, \dots, X_J],$$

where X_j is the generic j-th block.

The path models in the PLS involve three sets of relations:

1. ***Inner Model or Structural model,*** which refers to the structural model and specifies the relationships between the latent variables LVs. A latent variable can play both predictand role and predictor one; a latent variable which is never predicted is called an exogenous variable, otherwise, it is called endogenous variable.

The structural model can be expressed as:

$$\xi_j = \sum_i \beta_{ji}\xi_i + \zeta_j \qquad (1)$$

where β_{ji} is called the path coefficient (representing the path from i-th latent variable to the j-th latent variable) and we indicate with B the matrix of all the path coefficients in the model. It is a square matrix of 0/1, and its dimensions are equal to the number of LVs. This matrix indicates the structural relationship between LVs. ζ_j is the inner residual term, and the diagonal variance/covariance matrix among inner terms is indicated with Ψ .

2. ***Outer Model or Measurement Model***, which refers to the measurement model and specifies the relationships between the constructs and the associated indicators MVs. Two ways to establish these links can be distinguished as follows [6]:

- ***Reflective way***: in which the indicators (manifest variables) are regarded to be reflections or manifestations of their latent variables: a variation of the construct yields a variation in the measures. As a result, the direction of causality is from the construct to the indicator.

Each manifest variables represents the corresponding latent variable, which is linked to the latent variable by means of a simple regression model.

The reflective indicators of the latent construct should be internally consistent, and, as it is assumed that all the measures are indicators equally valid of a latent construct, they are interchangeable.

The reflective measures are at the basis of the theory of the classical tests, of the reliability estimation, of and factorial analysis, each of them considers the manifest variable x_{jk} being a linear combination of its latent variable ξ_j:

$$x_{jk} = \lambda_{jk}\xi_j + \varepsilon_{jk} \qquad (2)$$

where $\lambda_{jk}$is the generic loading coefficient associated to the k-th manifest variable in the j block, and we indicate with Λ the matrix containing all the loading coefficients in the block.

ε_{jk} represents the generic outer residual term associated to the generic manifest variable and the corresponding diagonal variance/ covariance matrix is indicated with Θ^{ε} .

Predictor specification is adopted,

$$E(x_{jk}|\xi_j) = \lambda_{jk}\xi_j \quad (3)$$

which implies that

$$E(\varepsilon_{jk}) = E(\xi_j \varepsilon_{jk}) = 0 \quad (4)$$

the residuals have zero mean and are uncorrelated with MVs.

- ***Formative way***: in which the indicators are regarded as causes of their latent constructs: a variation of the measures yields a variation in the construct. As a result, the direction of causality is from the indicator to the construct. The elimination of items that have low correlations compared with the overall indicators will compromise the construct validation, narrowing the domain.

This is one of the reasons by which the reliability measures of the internal consistency should not be used to estimate the fitting of the formative models. Moreover, the multi-collinearity between the indicators may be a serious problem for the parameter estimations of the measurement model when the indicators are formative, but it is a good point when the indicators are reflective.

The latent variable ξ_j is assumed to be a combination linear of its manifest variables x_{jk}:

$$\xi_j = \sum_k \pi_{jk} x_{jk} + \delta_j \quad (5)$$

assuming predictor specification:

$$E(\xi_j|x_{jk}) = \sum_k \pi_{jk} x_{jk} \quad (6)$$

which means that residuals have zero mean and are uncorrelated with the MVs

$$E(\delta_j) = E(\xi_j \delta_j) = 0 \quad (7)$$

3. ***Weight relations.*** The specification of the relations between LVs and their set of indicators is carried out at a conceptual level. In other words, the outer relations refer to the indicator and the "true" LV, which is unknown. As a result, the weight relations must be defined for completeness. The estimation of LVs are defined as follows:

$$\hat{\xi}_j = \sum_k \tilde{w}_{jk} \, x_{jk} \quad (8)$$

where $\tilde{w}_{jk}$ are the weights used to estimate the LVs as linear combinations of their observed MVs.

2.1 The PLS Algorithm

By means of this algorithm both the estimation of the latent variables and the estimation of the parameters is obtained. The PLS-PM implies some steps.

The first step consists in an iterative procedure made up of simple and/or multiple regressions by taking into account the relation of the internal model, of the external model, and of weight relations. The result is the estimation of a set of weights which are used to calculate the scores of the latent variables as linear combinations of their manifest variables. Once the estimations are obtained the following steps imply the non-iterative estimation of the structural model and measurement model coefficients.

2.1.1 *PLS Algorithm: stage 1*

Stage 1.1: outer approximation and inner approximation

In order to estimate the parameter, two double approximations for LVs are considered by PLS algorithm [17,20]:

- ***the outer approximation or external estimation***, called Y_j, is used for the measurement model. In this stage we find an initial proxy of each LV, ξ_j, as a linear combination of its MVs x_{jk}. The external estimation is obtained as the product of the block of MVs and the outer weights w_{jk};
- ***the inner approximation or internal estimation***, called Z_j, is used for the structural model. The connections among LVs are taken into account in order to get a proxy of each LV worked out as weighted aggregate of its adjacent LVs. The internal estimation is obtained as the product of the external estimation Y_i (of ξ_i) and the so-called inner weights, e_{ji}.

There are three ways to calculate the internal weights:

- **centroid scheme (Wold)**: Il centroid scheme is the scheme of the original algorithm by Wold. This scheme considers only the direction of the sign among the latent variables $sign\{cor(Y_j, Y_i)\}$, without considering neither the direction nor the power of the paths in the structural model.
- **factorial scheme (Lohmöller)**: this scheme uses the correlation coefficients $cor(Y_j, Y_i)$ as internal weights instead of using only the correlation sign and, therefore, it considers not only the direction of the sign but also the power of link of the paths in the structural model.
- **path weighting scheme**: in this case the latent variables are divided into predictors and followers according to the cause- effect relations between the two latent variables. A variable can be either a follower (if it is yielded by another a latent variable), or a predictor (if it is the cause of another latent variable. If ξ_i is follower of ξ_j, then, the internal weight is equal to the correlation between Y_j e Y_i. On the other hand, for the predictors ξ_i of ξ_j, the internal weights are the regression coefficients of the Y_i in the multiple regression of Y_j on Y_i associated with the predictors ξ_j.

The path weighting scheme has the advantage to consider both the direction and the power of the paths in the structural model.

Even though the path weighting scheme seems the most coherent with direction of the structural relations between latent variables, the centroid scheme is very often used as it adapts well to cases where the manifest variables in a block are strongly correlated to each other. The factorial scheme, instead, is better suited to cases where such correlations are weaker. In spite of the different common practices, it is strongly recommended to use the path weighting scheme. Indeed, this is the only estimation scheme that explicitly considers the direction of relationships as specified in the predictive path model [8].

Stage 1.2: Updating outer weights

The external estimation is conceived as a step in which the information contained in the internal relations are incorporated in estimation process of the latent variables.

Once the internal approximation is carried out, the internal estimation Z_i must be considered with respect to their indicator.

There are two ways to calculate the external weights:

1. Mode A: is preferred when the indicators are linked to their latent variables by means of the reflective way, in which each weight w_{jk} is the coefficient regression of Z_j in the simple regression of x_{jk} on Z_j, that is the simple regression $x_{jk} = w_{jk}Z_j$ in which:

$$w_{jk} = (Z_j^t Z_j)^{-1} Z_j^t x_{jk} = cor(x_{jk}, Z_j) \quad (9)$$

2. ***Mode B***: : is preferred when the indicators are linked to their latent variables by means of the formative way, in which Z_j is regressed on the block of indicators linked to the latent construct ξ_j, and the w_j of weight w_{jk} is the regression coefficients in the multiple regression

$$Z_j = \sum_k w_{jk} x_{jk} \quad (10)$$

and it is defined by means of:

$$w_j = (X_j^t X_j)^{-1} X_j^t Z_j \quad (11)$$

in which X_j is the matrix with columns of manifest variables x_{jk}.

Stage 1.3: Checking convergence

The algorithm is repeated as far as the weight convergence. In each iterative step, for example, S=1,2,3.., the convergence is verified comparing the external weights of the S step (new weights) with the weights of the S-1 step (old weights). For example, Wold (1982) suggested $|w_{jk}^{S-1} - w_{jk}^{S}| < 10^{-5}$ as a criterion of convergence.

2.1.2 algorithm PLS: stage 2

The second step of the algorithm consists in the estimation of the structural coefficients (path coefficients) and of the parameters of the measurement model (loading coefficients).

For the structural model the path coefficients are estimated by ordinary last squares in the multiple regression of Y_j on the Y_i.

For the measurement model the loading coefficients are estimated depending on the corresponding way. In particular, in the reflective way, the loading coefficients are the regression coefficients of the simple linear regression of x_{jk} on Y_j. By contrast, in the formative way, the weight coefficients π's coincide with outer weights as we perform the multiple linear regression Y_j on x_{jk}.

2.2 **Summary**

The procedure works on centered (or standardized) data, and it starts by choosing arbitrary weights (e.g *1,0..0*)[19]. Chin (1999) suggested starting with equal weights for all indicators (the loadings are set to 1) to get a first approximation of the LVs as a simple sum of its indicators. The inner relations among LVs are considered to estimate the internal approximation by choosing three options: centroid, factoring and path scheme. After obtaining the internal approximation, the algorithm turns around the external relations with the estimate of outer weights obtained by means of mode A (*reflective*) or by mode B (*formative*). The procedure is repeated until convergence of the weights is obtained. Once convergence of the weights is obtained and LVs are estimated, the parameters of the structural and measurement models are calculated by means of the ordinary least squares (OLS).

3. THE EXTERNAL ANALYSIS

A way to add external information to the model on the subjects (e.g, age, sex, education, job, etc..) and/or on the variables is the External Analysis (Takane e Shibayama, 1991). The addition of external information on the subjects and/or variables could help the researcher interpret the results objectively.

This method decomposes the original data into several components (External Analysis), those that can be explained, and those that cannot be explained, by the external information.

We denote a matrix of the data composed of *N*-subject and *n*-variables with *X*. The data may consist in *N* subjects' judgment on *n*-dimensions of the supplied service public quality, or any other multivariate observations. The data may be raw or pre-processed, for example, by standardizations or other transformations.

We denote a matrix of the information on the users with G, and the matrix of the information on the variables with H. For example, G may be an *N*-component vector of ones, a matrix of dummy variables, or a matrix of continuous variables characterizing the subjects.

Similarly, H may be an *N*-component vector of ones, or any other matrix of explanatory variables characterizing the relationships among columns of *X*. Both G and H can be put in any analysis by means of orthogonal projector operators, that is, orthogonal projection operators onto spaces spanned by the column vectors of G and H with $P_G = G(G'G)^{-}G'$ e $P_H = H(H'H)^{-}H'$, respectively. It is well known that both P_G and P_H are unique even if $(G'G)^{-}$ and $(H'H)^{-}$are unique. Moreover, $Q_G = I - P_G$, and $Q_H =$ I- P_H are both orthogonal projector operators that are orthogonal to P_G and P_H, respectively.

The matrix of the measurements *X* can be decomposed according to the following relationship: $P_G X P_H + Q_G X P_H + P_G X Q_H + Q_G X Q_H$. Each term has a precise statistic meaning in it; in particular: $P_G X P_H$ indicates the effect of row and column information; $Q_G X P_H$ is the effect of column information without the effect of the row one; $P_G X Q_H$ is the effect of row information without the column ones and $Q_G X Q_H$ is the part which disregards external information.

Once the matrix *X* is decomposed, it may be interesting to carry out the scheme of analysis introduced a prior on each term to catch (to clean) the influence of the external information on the evaluation of the supplied service quality. The external information has different roles, for example, eliminating the different reference systems of the users/citizens as well as catching the effects of the information both on the components and on the dimensions.

4. RESULTS OF THE ANALYSIS

The objective of the research is to analyse the passenger satisfaction of the public transport service by means of the PLS-PM approach and, later, of the combined use of the PLS approach and External Analysis Method. For this reason, five latent variables (LVs) have been identified, each of them measured by proper indicators (i.e. MVs). The LVs and the corresponding MVs identified are:

Table 1: Latent Variable and Manifest Variables

Latent Variables	**Manifest Variables**
Accessibility to the service (Exogenous)	X1: availability of the company time- tables; X2: information on the time tables and the runs; X3: purchase of the tickets and season-tickets; X4: access to the buses; X5: services for disabled people; X6: bus-shelter condition; X7: behaviour of the company staff (kindness, propriety, presentability)
Condition of the buses (Exogenous)	X8: motor vehicle cleanness; X9: journey comfort; X10: reduction of the environmental impact;
Security on board (Exogenous)	X11: journey Security; X12: safety for the users as regards stealing and attacks;
Reliability of the service (Exogenous)	X13: punctuality, i.e observance of the times. X14: regularity, i.e ability to carry out the predicted runs; X15: company time-tables and their connections with other means of transport.
Overall Passenger Satisfaction (Endogenous)	X16: judgment on the company as a whole ; X17: degree of improvement of the service reached in the latest year.

The LVs- ***accessibility to the service, condition of the buses***, ***security on board***, ***reliability of the service-*** are exogenous LVs, i.e. they are variables which are never predicted and behave just only as predictors, while the ***overall passenger satisfaction*** latent dimension is an endogenous LV (i.e. dependent).

Any analysis of the structural equation modeling implies an explicit specification of the path-model which, in its turn, is composed of the measurement model (Outer Model) and of the structural model (Inner Model).

4.1 The Outer Model and the Inner Model

● **Outer Model**

The outer model establishes the relationship between the block of the manifest variables and their corresponding latent variables. As regards the LVs **accessibility to the service**, **condition of the buses**, **security on board**, **reliability of the service** and **overall passenger/user satisfaction** the MVs are linked to them in a "reflective" way. In other words, the MVs are considered reflections or manifestations of the LVs. For these LVs, in correspondence with their respective MVs, we read the std. loadings, that is standardized regression coefficients (it is about a simple linear regression).

Test of the unidimensionality of the reflective MVs block.

A way to check the quality of the measurement model is the verification of the unidimensionality of the reflective MVs block. The reflective indicators of the LV should be coherent internally, and, as it is supposed, that all the measures are indicators equally valid of a LV, they are interchangeable.

- ***Dillon –Goldestein's ρ (o Jöreskog):*** a block is unidimensional if this index is >0.7.

$$\rho = \frac{(\sum_{k=1}^{p} \lambda_{jk})^2 \operatorname{var}(\xi_j)}{\sum_{k=1}^{p} \operatorname{var}(\xi_j) + \sum_{k=1}^{p} \operatorname{var}(\varepsilon_{jk})} \tag{12}$$

As, in practice, we do not know the real values of λ_{jk} and ξ_j, an estimate of Dillon Goldstein's ρ is needed. The approximation of the latent variable is achieved by using the first principal component t_{j1} of the *j-th* block of indicators; the approximation of the loading coefficient is taken as the correlation between t_j and the observed variable x_k, cor(t_1,x_k); the term var (ε_k) is approximated by 1-cor^2(t_1,x_k). Then, estimate of Dillon-Goldstein's is given by

$$\hat{\rho} = \frac{[\sum_{k=1}^{p} cor(x_{jk}, t_{j1})]^2}{\sum_{k=1}^{p} [cor(x_{jk}, t_{j1})]^2 + \sum_{k=1}^{p} [1 - cor^2(x_{jk}, t_{j1})]} \tag{13}$$

The value of the index is > 0.7 for all the observed VMs blocks (Table 2) .

Table 2: Block Unidimensionality

LVs	Type Measure	MVs	DG.rho
A.S	Reflective	7	0,8373
C.B	Reflective	3	0,8025
S.B	Reflective	2	0,7258
R.S	Reflective	3	0,8706
O.P.S	Reflective	2	0,8180

● **Inner Model**

The inner model considers only the LVs, which are assumed to be linearly interconnected according to a casual-effect relationship model. The present study aims at verifying, from an explorative and non confirmative view point, the existence of meaningful relations between the following LVs:

1. accessibility to the service and overall passenger satisfaction
2. condition of the buses and overall passenger satisfaction
3. security on board and overall passenger satisfaction
4. reliability of the service and overall passenger satisfaction

Like the measurement model, the structural model also requires to be validated. In particular, in correspondence with the endogenous LV, we read the coefficient of determination R^2. For each regression in the structural model we have an R^2 that is interpreted similarly as any multiple regression analysis. R^2 indicates the amount of variance in the endogenous latent variable explained by its independent latent variables. A satisfying R^2 (0,57) is obtained for the latent variable Overall Passenger Satisfaction (Table 3). Moreover, if we consider the ***Average Redundance*** (note that the redundancy index represents the power of the set of independent latent variables to explain the variation in the dependent latent variable), a satisfying prediction endogenous LV -**Overall Passenger Satisfaction** (0.38)- from LVs exogenous LVs is obtained.

It is important to note in table 3 the values of AVE (Average Variance Extracted) that tries to measure the amount of variance that a LV captures from its indicators in relation to the amount of variance due to measurement error. AVE is, in most cases, > 0.50, which means that 50% or more variance of the indicators should be accounted for.

Table 3: Summary Inner Model

	LV.Type	**Measure**	**MVs**	**R-square**	**Av.Redun**	**AVE**
A.S	Exogen	Rflct	7	0	0	0.43
C.B	Exogen	Rflct	3	0	0	0.58
S.B	Exogen	Rflct	2	0	0	0.53
R.S	Exogen	Rflct	3	0	0	0.69
O.P.S	Endogen	Rflct	2	0,57	0,38	0.68

GoF (Goodness of fit):

The Goodness of Fit (GoF) is a global criterion developed by Amato, S., Esposito Vinzi, V., and Tenenhaus, M. [1], and it is represents a compromise between the quality of the measurement model and the quality of the structural model. A satisfying GoF is obtained both for the outer model (0,98) and for the inner model (0,62).

4.2 Parameter estimation and validation by re-sampling methods

To estimate the parameter of the model, we have used the module R-package. To calculate the inner estimates of the latent variables, we have used the path weighting scheme.

The non-parametric bootstrap [7,16] procedure can be used in PLS-PM to provide confidence intervals for all parameter estimation, building the basis for statistical inference. In general, the bootstrap technique provides an estimation the shape, spread, and bias of the sampling distribution of a specific statistic.

Bootstrapping treats the observed sample as if it represented the population. Bootstrap samples are created by randomly drawing cases with replacement from the original sample.

In particular, bootstrap has been based on 100 samples, and 95% confidence intervals have been asked for. The confidence intervals indicate the regression coefficients which are significant [18].

If a confidence interval for an estimated path coefficient does not include zero, the hypothesis that the parameter equal zero is rejected.

As shown in the table 4, all the parameters of the measurement model (loading coefficients) are significant (the confidence intervals never include zero).

Moreover, the p-value associated to the t- statistic bootstrap, in this specific case, is, in most cases, > 0.05 and this implies the acceptation of the hypothesis according to which the distribution sampling is centered at the true value of the parameter.

Moreover, the MVs that best reflect their corresponding LVs are:

- ***accessibility to the service***: behaviour of the company staff (loading coefficient: 0,72), availability of the company time- tables (loading coefficient: 0,67), access to the buses (loading coefficient: 0,67).
- ***condition of the bus***: motor vehicle cleanness (loading coefficient: 0.86); journey comfort (loading coefficient: 0,85);
- ***security of board***: journey security (loading coefficient: 0,98) ;
- ***reliability of the service:*** regularity, i.e ability to carry out the predicted runs (loading coefficient: 0,84), company time-tables and their connections with other means of transport (loading coefficient:0,83), punctuality, i.e observance of the times (loading coefficient:0,82);
- ***overall passenger satisfaction:*** judgment on the company as a whole (loading coefficient:0,92);

Table 4: Bootstrap validation for loading coefficients (** significant 5%)

	Original (O)	Mean Boot.	Std.Err	Bootstrap t-statistic	p-value	Conf. Intervals (95%)
X1	0,6296	0,6285	0,0420	-0,2628	0,7932	0,5459;0,7111
X2	0,6711	0,6684	0,0384	-0,6949	0,4888	0,593;0,7439
X3	0,5983	0,6004	0,0497	0,4342	0,6651	0,5026;0,6982
X4	0,6753	0,6717	0,0409	-0,8917	0,3747	0,5913;0,7520
X5	0,6331	0,6359	0,0378	0,7294	0,4675	0,5616;0,7102
X6	0,6153	0,6167	0,0513	0,2722	0,7860	0,5159;0,7175
X7	0,7261	0,7269	0,0306	0,2532	0,8006	0,6667;0,7871
X8	0,8665	0,8687	0,0198	1,1066	0,2712	0,8298;0,9075
X9	0,8544	0,8546	0,0213	0,1125	0,9106	0,8128;0,8965
X10	0,5116	0,5082	0,06	-0,5752	0,5665	0,3902;0,6261
X11	0,9830	0,978	0,0207	-2,4172	0,0175	0,9374;1,0187
X12	0,3185	0,3272	0,1207	0,7171	0,4750	0,09;0,5643
X13	0,8226	0,8229	0,0258	-0,2140	0,8310	0,7713;0,8728
X14	0,8438	0,8434	0,0246	-0,1678	0,8671	0,7951;0,8917
X15	0,8271	0,8286	0,0235	0,6284	0,5312	0,7825;0,8747
X16	0,9267	0,9246	0,0163	-1,2653	0,2052	0,8926;0,9566
X17	0,7029	0,7089	0,0457	1,3232	0,1888	0,6191;0,7987

If we consider the analysis of the path- coefficient we can infer a significance of all the links.

In particular, the exogenous LVs that mostly affect the endogenous LV (overall passenger satisfaction) are: **accessibility to the service** (path coefficient: 0,41) and **reliability of the service** (path coefficient: 0,22).

The value of the link between the **accessibility to the service** and the **overall passenger satisfaction** indicates that when the accessibility to the service rises by one unit (in particular, the behavior of the company staff, the access to the buses, the information on time table and runs) the ***overall users satisfaction*** rises by 0,41. Similarly, the value of the link between the **reliability of the service** and the **overall passenger satisfaction** indicates that when the **reliability of the service** rises by one unit (in particular, regularity, company time- table and their connection with other means of transport, punctuality) the ***overall users satisfaction*** rises by 0.22.

Table 5: Bootstrap Validation for path coefficients (**significant 5%)

	Original (O)	Mean Boot	Std.Err	Bootstrap t-statistic	p-value	Conf. Intervals (95%)
A.S→ O.P.S	0,4105	0,4121	0,0612	0,2578	0,7971	0,2918;0,5324
C.B→ O.P.S	0,1364	0,1337	0,0537	-0,5038	0,6158	0,0281;0,2392
S.B→ O.P.S	0,1423	0,1445	0,0464	0,4640	0,6437	0,0532;0,2358
R.S→ O.P.S	0,2228	0,2227	0,0473	-0,0186	0,9852	0,1297;0,3158

The full specification of the path diagram (Structural model and Measurement model), along with the indication of the loading coefficients and path coefficients, is the following:

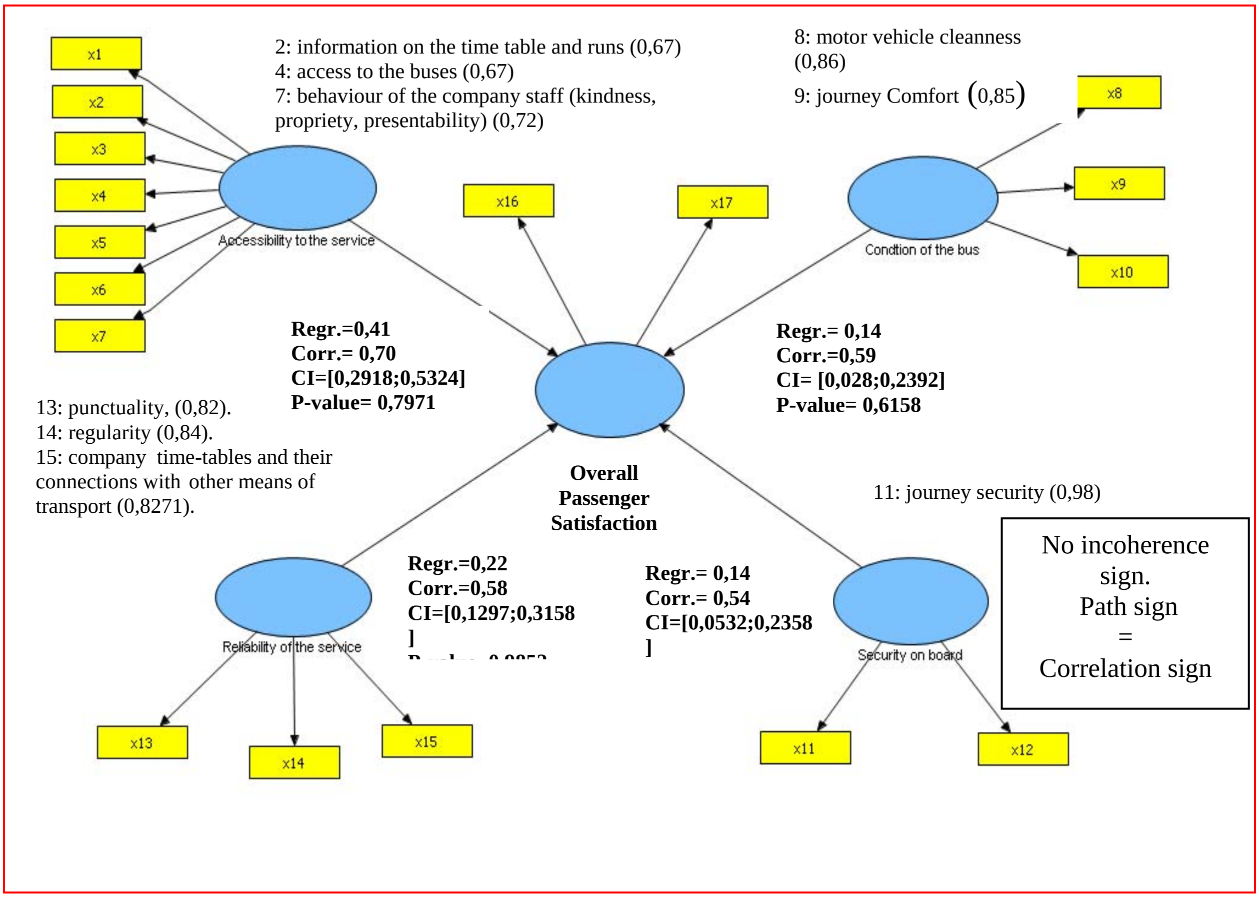

Figure 2: The specification of the path diagram (Structural Model and Measurement Model): path coefficients, correlations, loading coefficients and bootstrap results.

It is important to observe that the figure 2 also shows the correlations between LVs as there is no incoherence between the path coefficient sign and the correlations sign; for this reason, none of the original relationships between LVs has been suppressed.

4.3 PLS-PM with external information on subjects

The analysis of the data has been further supported by means of the joint use of two methodologies: the PLS-PM and the External Analysis, catching the advantages of both.

For this purpose, the external information on subjects that has been taken into account is the *job of the interviewed subjects*: 1. Student, 2. Employee, 3. Housewife, 4. Retired, 5. Other".

The choice of working only with the external information on the subjects has been determined by the interest in testing how these preference data are related to the subjects' demographic information, that is their job.

As previously shown, the aim of the External Information is the decomposition of the matrix of the original data (X) in different components, those which can be explained by the external information: "job", and those that cannot be explained by external information.

This external information can be put in the analysis by means of orthogonal projector operators; in this specific case, the orthogonal projector is:

$$P_G = G_{(N,B)} (G'_{(B,N)} G_{(N,B)})^{-} G'_{(B,N)} \tag{14}$$

As it is known, the decomposition of the matrix X is the following:

$$X = P_G X P_H + Q_G X P_H + P_G X Q_H + Q_G X Q_H \tag{15}$$

The PLS-PM has been applied to the first and the last matrix X component, and both the predictive power of the model and the differences in terms of path coefficients between the two PLS- models (i.e: PLS with information and PLS without information) have been evaluated in order to verify whether a relationship between the information taken into account and the theoretical links specified in the path diagram (figure 2) exists.

The addition of the "job" external information improves the quality of the structural model. The coefficient of determination R^2 changes from 0.56 (for the PLS-model without information) into 0.90 (for the PLS-model with information).

Table 6: Determination Coefficient (R^2)

PLS-Model with information	PLS-Model without information
0.90	0.56

The redundancy index also improves by changing from 0,39 into 0,65.

Table 7: Redundance Index

PLS-Model with information	PLS-Model without information
0.65	0.39

The values of AVE (Average Variance Extracted) for PLS-Model with information improved in most cases.

Table 8: Average Variance Extracted

LVs	AVE PLS-Model with information	AVE PLS-Model without information
A.C	0.43	0.42
C.B	0.78	0.58
S.B	0.93	0.53
R.B	0.51	0.69
O.P.S	0.70	0.67

Finally, the Gof (Goodness of fit) for the Inner Model also improves, (from 0,61 for the PLS-model without information to 0,66 for the PLS-model with information) .
By comparing path coefficients of the PLS-PM with information and the PLS-PM without information it follows that, if the external information is absent (e.g: equal jobs), all the links are positive and significant. The addition of the external information further strengthens

some of these links, in particular, the relationship between the access to the service and the overall passenger satisfaction, and between the bus condition and the overall passenger satisfaction. Probably, there are specific categories of subjects whose judgment of preference may be actually linked to the external information taken in consideration, for example, students and employees, workers whose use of the means of the transport should be more frequent. As a result, the company, in order to better the quality of the service, should direct its efforts to improve this aspect of the service.

Table 9: Comparison between the path coefficients of the PLS-PM with external information and of the PLS-PM without external information

	PLS-PM with information	**PLS-PM without information**	**abs. diff**
A.S->O.P.S	**0,5943**	**0,3965**	**0,1978**
C.B->O.P.S	**0,3638**	**0,1416**	**0,2222**
S.B->O.P.S	0,0913	0,1243	0,0330
R.S->O.P.S	0,2860	0,2478	0,0382

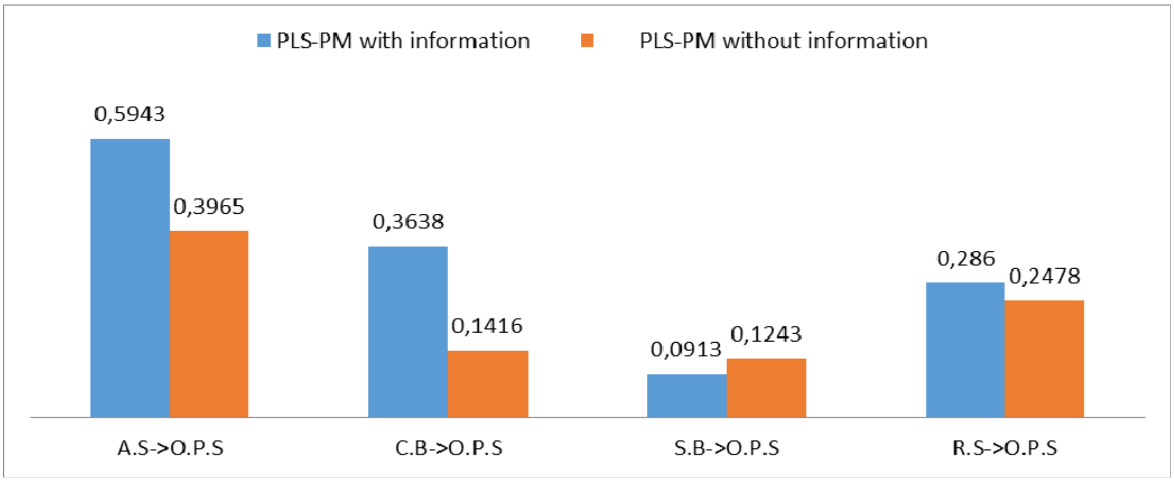

Figure 3: Comparison between path coefficients of the PLS-PM with information and the PLS-PM without information

5. CONCLUSION

The objective of the research has been to analyse the passenger satisfaction of the public transport service by means of the PLS-PM approach and, later, of the combined use of the PLS approach and External Analysis Method. In particular, the External Analysis Method represents the additional element that this contribution has sought to highlight.

The use of the PLS Path modeling has shown that the exogenous LVs that mostly affect the endogenous LV are the **accessibility to the service** and **reliability of the service**. A satisfying GoF is obtained both for the outer model and for the inner model.

Later, the analysis of the data has been further supported by means of the joint use of two methodologies: the PLS-PM and the External Analysis. The external information on subjects that has been taken into account is the Job of the interviewed subjects. The addition of the *job external information* improves the quality of the structural model, of both R^2 and GoF.

In particular, the relationship between the access to the service and the overall passenger satisfaction, and between the bus condition and the overall passenger satisfaction further strengthens.

REFERENCES

1. Amato, S., Esposito Vinzi, V., and Tenenhaus, M., **A global Goodness-of-Fit for PLS structural equation modelling**. Oral communication to PLS Club, HEC School of Management, 2004.
2. Bagozzi, R.P, **"Structural Equation Models in Experimental Research"**, Journal of Marketing Resarch,14, 1977, pp 209-226.
3. Busacca, B., **Le risorse di fiducia dell'impresa. Soddisfazione del cliente, creazione di valore, strategie di accrescimento**, UTET, Libreria, 1994.
4. Chin, W.W. and Newsted, P.R., **Structural Equation Modeling Analysis with Small Samples using Partial Least Squares**. In: *Statistical Strategies for Small Sample Research*, 1999, 307-341. Hoyle R. (Ed). London: Sage.
5. Chin, W.W, **Partial Least Squares for research: An overview and presentation of recent advance using the PLS approach**, 2000, p.68.
6. Edwards, Jeffrey R. and Bagozzi, P.R, **"On the Nature and Direction of Relationships between Construct an Measures"**, *Psycological Method,*5(2), 2000, 155-74.
7. Efron,B, **The jackknife, the bootstrap and other resampling plans**, SIAM,USA: Philadelphia, 1982.
8. Esposito Vinzi, V., Chin, W.W., Henseler,J., and Wang. H, **Handbook of Partial Least Squares, Concepts, Methods and Application**, Springer Handbooks of Computational Statistics, New York, 2010.
9. Fornell, C., and Bookstein, F.L., **Two Structural equation model: LISREL and PLS applied to consumer exit-voice theory**. Journal of Marketing Research,19, 1982, 440-452.
10. Jöreskog, K.G., **Factor Analysis and Its Extensions. Factor Analysis at 100: Historical Developments and Future Decisions**. Department of Psychology, University of North Carolina, 2004, May 13-15.
11. Jöreskog, K.G., and Wold, H., **The ML and PLS Techniques for Modeling with Latent Variables: Historical and Comparative Aspects.** In: Systems under indirect observation: Causality, structure, prediction. Part I, 263-270. K.G. Jöreskog and H. Wold (Eds). Amsterdam: North Holland. 149, 1982.
12. Jöreskog, K.G., **"Testing Structural Equation Models"**. In: *Testing Structural Equation Models.* (Ed) Bollen, K.A and Lang, J.S., Beverly Hills, Sage, 1993, pp. 294-316.
13. Jöreskog, K.G and Sorbon, D., **LISREL 7: A Guide to the Program and Application***,* Chicago, IL, SPSS, Inc, 1989.
14. Jöreskog, K.G and Sorbon, D., ***LISREL 7: User's Reference Guide***. In: Scientific Software, Inc., Mooresville, 1989.

15. Takane, Y. and Shibayama T., **Principal Component Analysis with externalInformation on both subject and variables**, Psicometrica-vol.56, NO. 1, 1991, 97-120.
16. Temme, D., Kreis, H., and Hildebrandt, L., **PLS Path Modeling: A Software Review**. SFB 649 Discussion Paper 2006-084, Economic Risk.
17. Tenenhaus, M. (1999), **L'Approche PLS**, *Statistquèe Applique* XLVII(2), 1999, 5-40.
18. Tenenhaus, M. (2008) Structural Equation Modeling for small samples. HEC school of management (GRECHEC).
19. Tenenhaus, M., Esposito Vinzi, V., Chatelin,Y.M, and Lauro,C., **PLS Path Modeling**, Computational Statistics & Data Analysis 48, 2005, pp.159-205.
20. Wold, H., **Soft modeling: The Basic Design and Some Extensions.** In: *Systems under indirect observation: Causality, structure, prediction*. Part II, 1-54. K.G. Jöreskog and H. Wold (Eds). Amsterdam: North Holland, 1982.
21. Wold, H., **Non Linear estimation by iterative least squares procedures**. In: Researche Paper Statistcs, Festchrft for J. Neyman, 411-444. F.David (Ed).New York: Wiley, 1966.
22. Wold, H., **Estimation of Principal Components an Related Models by Iterative Least Squares**, In: Multivariate Analysis, Proceedings o fan International Symposium, 391-430, P.R. Krishmaiah (Ed). New York: Academic Press, 1966.
23. Sanchez, G., **Partial Least Squares**, Phd thesis, 2009
24. Zeithaml V.A., Parasuraman A., Berry L.L., **Servire la qualità**, McGraw-Hill, 1991.