

Creating and running a basic XLSTAT-PLSPM project with Excel 2007

[PLSPM_ECSI2007.ppmx](#)

Principles of PLS path modeling

Partial Least Squares Path Modeling (PLS-PM) is a statistical approach for modeling complex multivariable relationships (structural equation models) among observed and latent variables. Since a few years, this approach has been enjoying increasing popularity in several sciences (Esposito Vinzi et al., 2007). Structural Equation Models include a number of statistical methodologies allowing the estimation of a causal theoretical network of relationships linking latent complex concepts, each measured by means of a number of observable indicators.

The first presentation of the finalized PLS approach to path models with latent variables has been published by Wold in 1979 and then the main references on the PLS algorithm are Wold (1982 and 1985).

Herman Wold opposed LISREL (Jöreskog, 1970) "hard modeling" (heavy distribution assumptions, several hundreds of cases necessary) to PLS "soft modeling" (very few distribution assumptions, few cases can suffice). These two approaches to Structural Equation Modeling have been compared in Jöreskog and Wold (1982).

From the standpoint of structural equation modeling, PLS-PM is a component-based approach where the concept of causality is formulated in terms of linear conditional expectation. PLS-PM seeks for optimal linear predictive relationships rather than for causal mechanisms thus privileging a prediction-relevance oriented discovery process to the statistical testing of causal hypotheses. Two very important review papers on PLS approach to Structural Equation Modeling are Chin (1998, more application oriented) and Tenenhaus et al. (2005, more theory oriented).

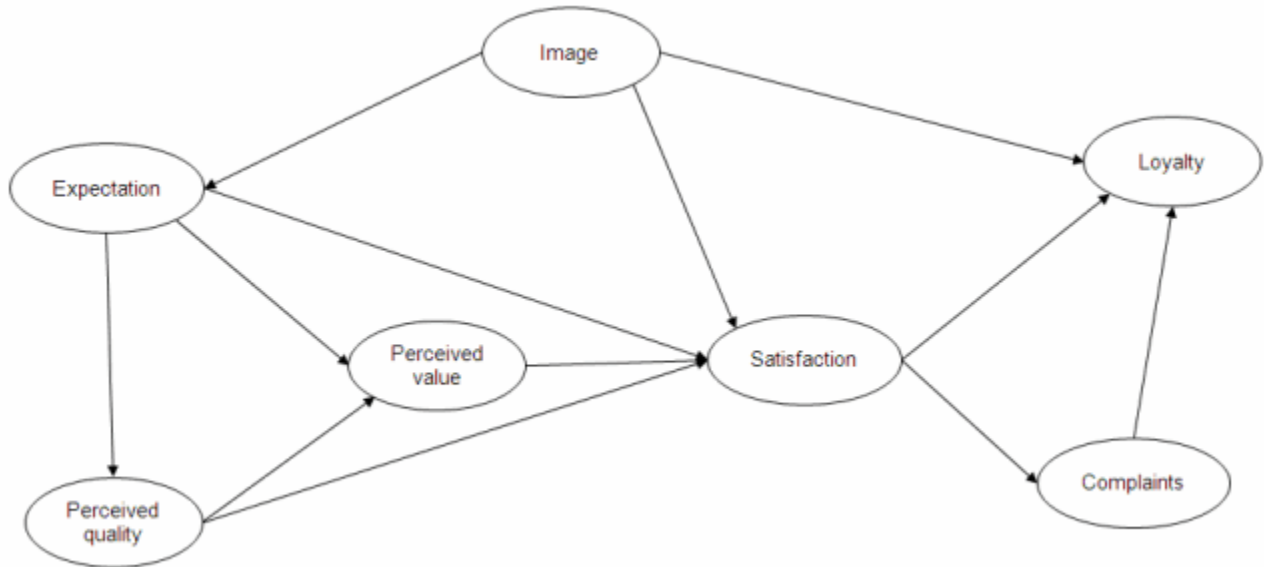
Furthermore, PLS Path Modeling can be used for analyzing multiple tables and it is directly related to more classical data analysis methods used in this field. In fact, PLS-PM may be also viewed as a very flexible approach to multi-block (or multiple table) analysis by means of both the hierarchical PLS path model and the confirmatory PLS path model (Tenenhaus and Hanafi, 2007). This approach clearly shows how the "data-driven" tradition of multiple table analysis can be somehow merged in the "theory-driven" tradition of structural equation modeling so as to allow running the analysis of multi-block data in light of current knowledge on conceptual relationships between tables.

In this tutorial we guide you step by step to show you how to create a project, define a model, estimate the parameters and analyze the results. This tutorial is based on the following paper: [Tenenhaus M., Esposito Vinzi V., Chatelin Y.-M. and Lauro C. (2005). PLS Path Modeling. *Computational Statistics & Data Analysis*, 48(1), 159-205].

PLS path modeling with XLSTAT-PLSPM

Dataset for the PLS path modeling analysis

The application is based on real life data, where 250 customers of mobile phone operators have been asked several questions in order to be able to model their loyalty. The PLSPM model is based on the European Consumer Satisfaction Index (ECSI). In the ECSI model, the latent variables (concepts that cannot be directly measured) are interrelated as displayed below.



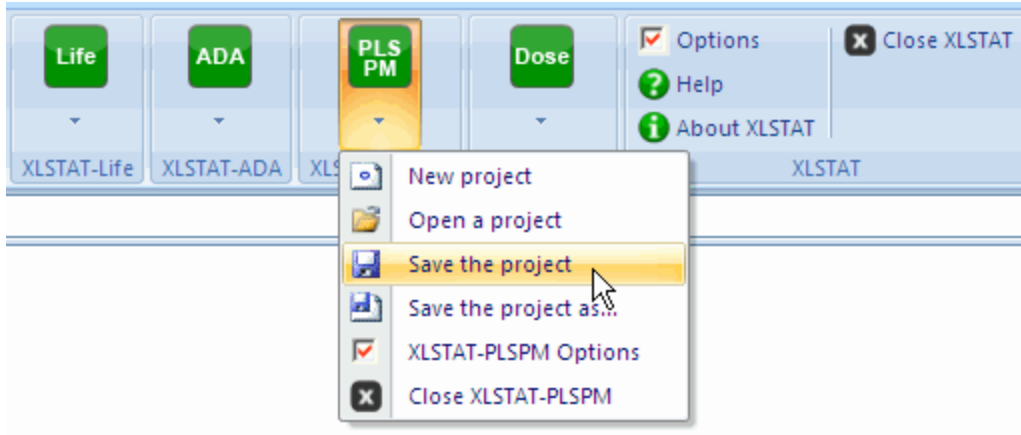
Each latent variable is related to one or more manifest variables that are measured. In this application case, the manifest variables questions are on a 0-100 scale. For example, for the Image latent variable the five manifest variables are:

- It can be trusted in what it says and does
- It is stable and firmly established
- It has a social contribution for the society
- It is concerned with customers
- It is innovative and forward looking

An XLSTAT-PLSPM project sheet containing both the data and the results for use in this tutorial can be downloaded by clicking [here](#).

XLSTAT-PLSPM projects are special Excel workbook templates. When you create a new project, its default name starts with PLSPMBook.

You can then save it to the name you want, but make sure you use the "Save" or "Save as" command of the XLSTAT-PLSPM menu to save it in the folder dedicated to the PLSPM projects using the *.ppmx extension.



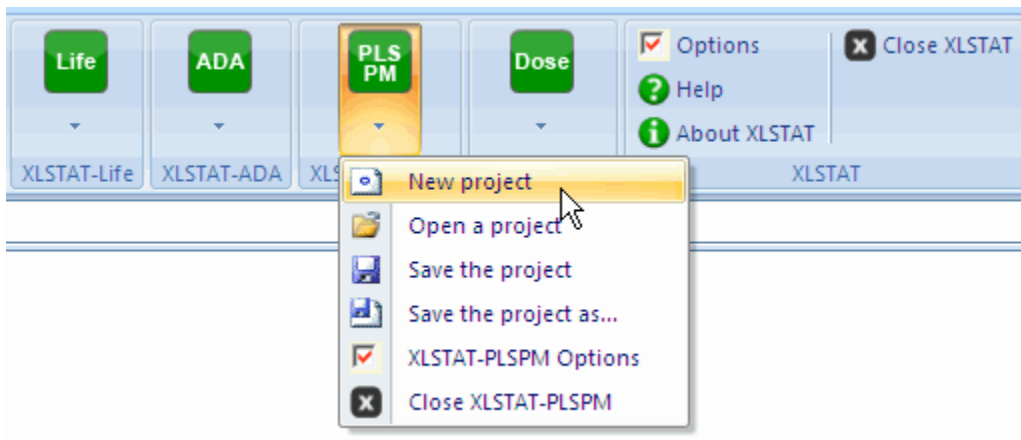
Note: when you open the `PLSPathModeling_EC SI.ppm` file, the graphical representation might look bad. This is due to the fact that the representation depends on your screen settings. To improve the display, click the "Optimize the display" button (see below).

A raw XLSTAT-PLSPM project contains two sheets that cannot be removed:

- D1: This sheet is empty and you need to add all the input data that you want to use into that worksheet.
- PLSPMGraph: This sheet is blank and is used to design the model. When you select this sheet, the "Path modeling" menu is displayed on the upper left part of the page.

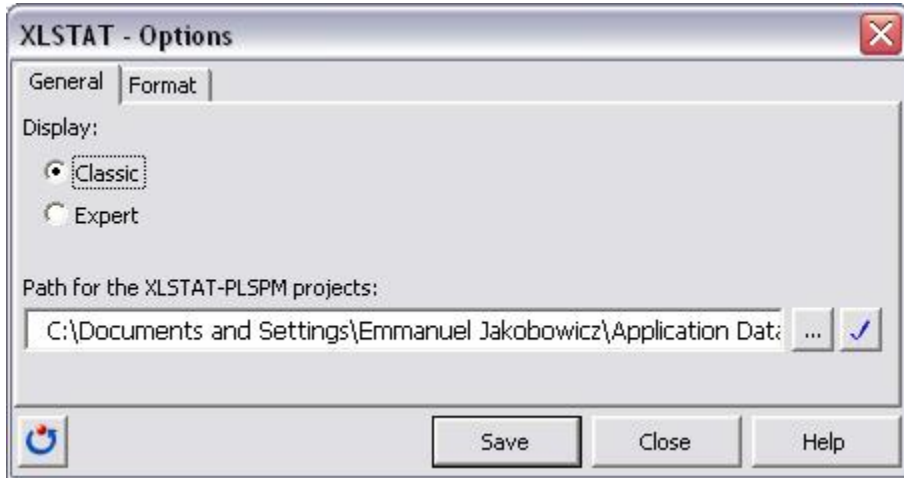
Setting up a PLS path modeling analysis

To create the project used in this tutorial, we first generated a new project using the XLSTAT-PLSPM menu:

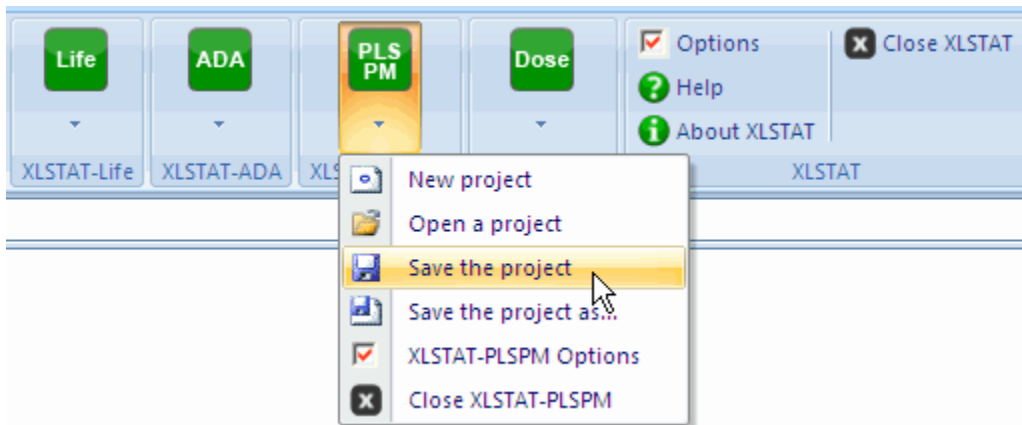


PLS Path Modeling is a complex method and the PLSPM module of XLSTAT has many options and specificities. In order to simplify the application of a simple model, two displays are available.

The default one, called “classic”, displays the main functions necessary to apply PLS Path Modeling and a more sophisticated one, called “expert”, displays lots of new options like multigroup testing, moderating effect estimation, superbloc procedure... To modify this option, click on the XLSTAT-PLPM options **button on the XLSTAT-PLSPM toolbar**.



We then saved it as **PLSPM_ECSEI.ppmx** using the **Save as** command of the same menu.



Then, we copied the data that were available in an Excel file, and pasted them into the D1 sheet of the Project. Once this is done, you are ready to start creating the model. Move to the PLSPMGraph sheet. The toolbar is displayed on the upper left corner of that sheet. You can find details on the function of each button in the help.



To create several latent variables in a row, double click on the circle button so that it stays pressed while you add variables:



You can then add the arrows that indicate how the latent variables are related.

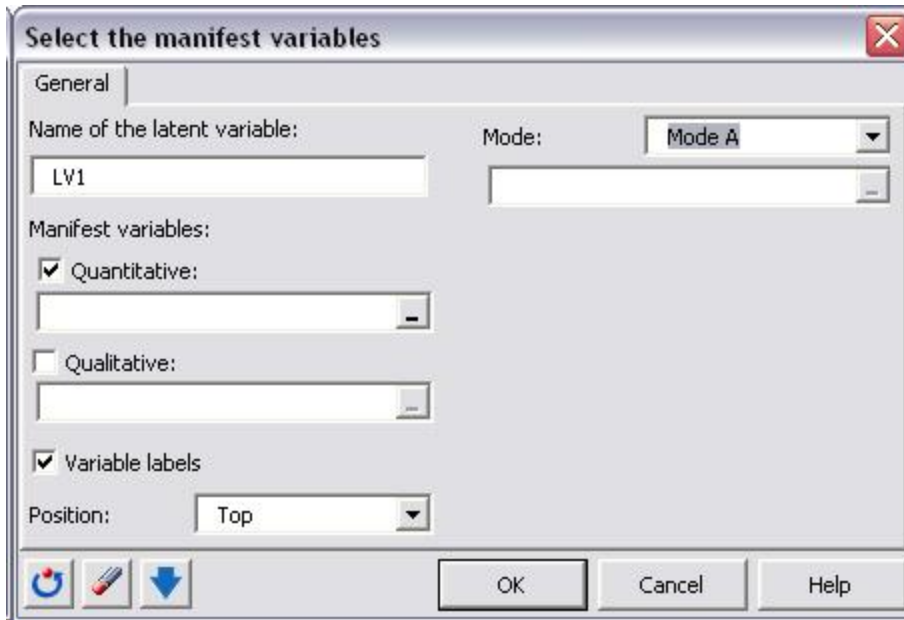
To add an arrow, click on the latent variable from which it should start, then press Ctrl and click on the latent variable where the arrow should end. Then click on the arrow button or use the following keyboard shortcut: Ctrl+L.



Once all the arrows have been added, you can define the manifest variables that relate to each latent variable (this can also be done after adding the latent variables). To add manifest variables to a latent variable, select the latent variable and click on the MV button in the toolbar:

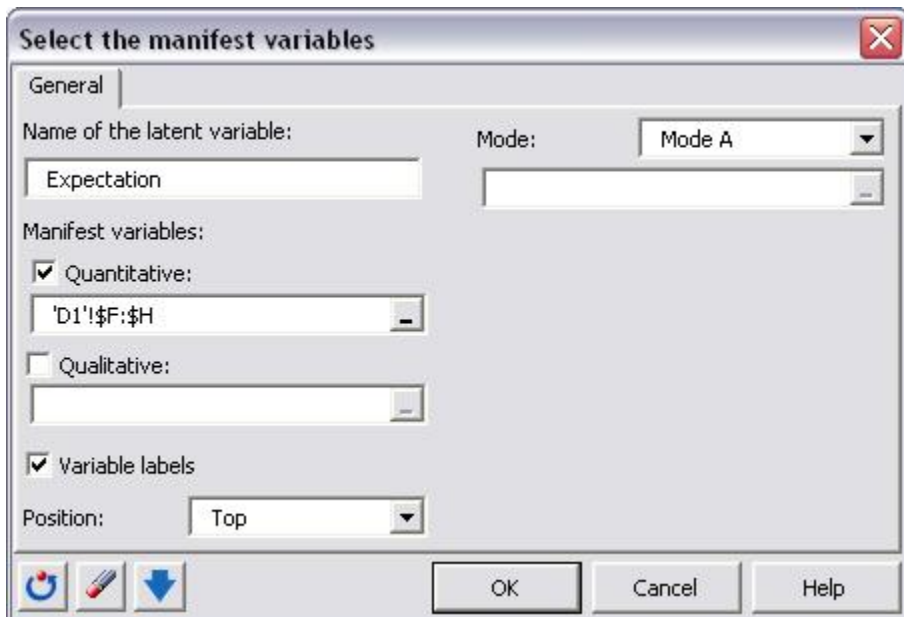


This activates the D1 sheet and displays a dialog box where you give a proper name to the latent variable, select the manifest variables on D1 and define a few settings.

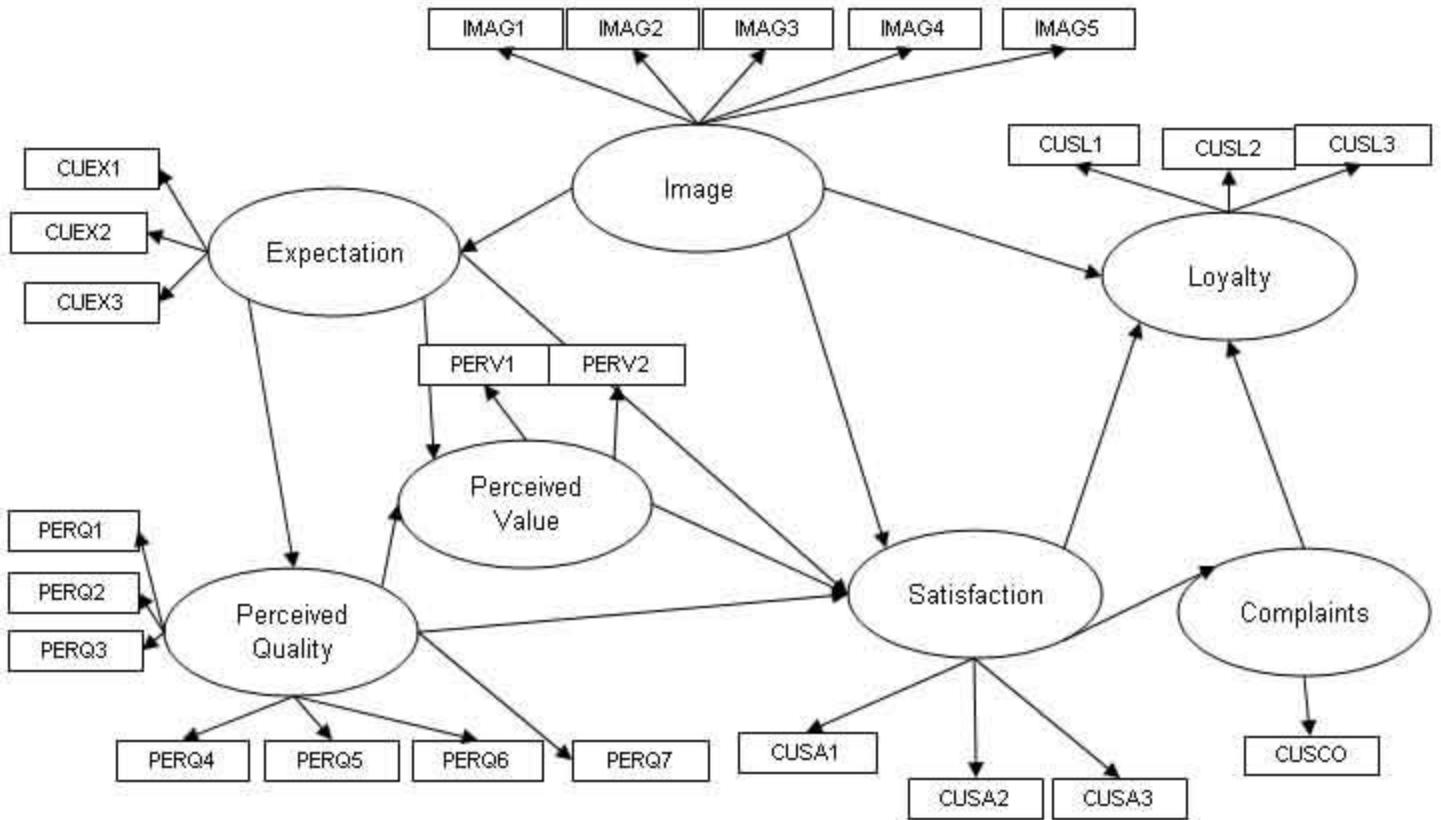


The mode has to be defined. In Mode A (reflective mode) the latent variable is responsible for what is measured for the manifest variables, and in Mode B (formative mode), the manifest variables construct the latent variable.

For example, this is how the dialog box looked like once filled in for the latent variable Expectation:



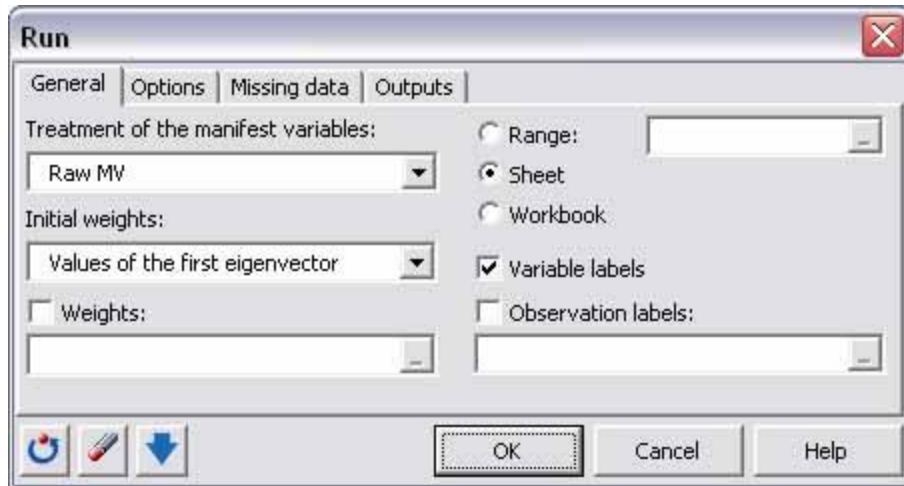
The obtained model has the following form:



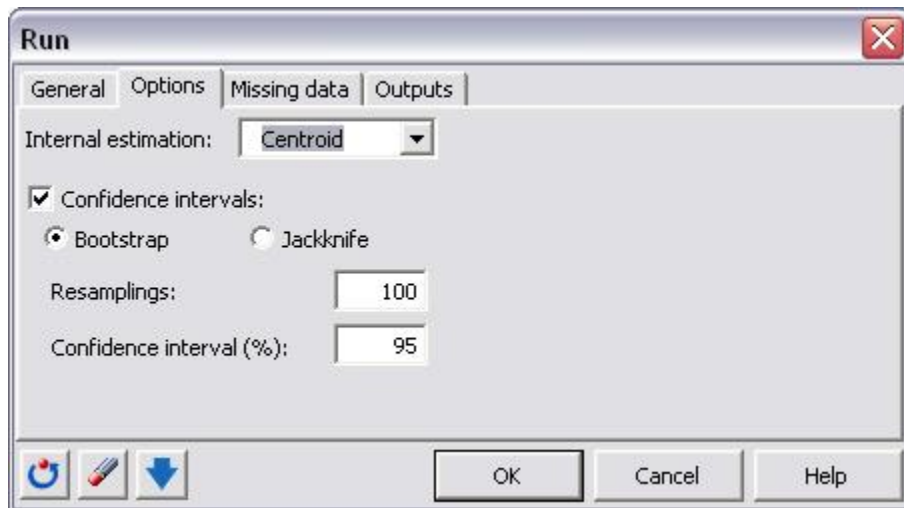
Once the manifest variables have been defined for each latent variable and latent variables are linked, you can start computing the model. To run the model, click the:



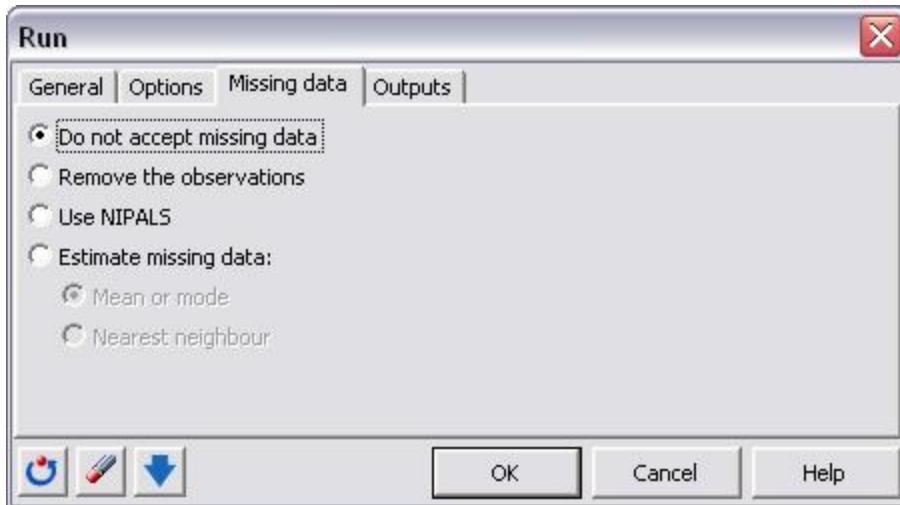
This displays the Run dialog box, where many options are available. For this tutorial the following options have been used:



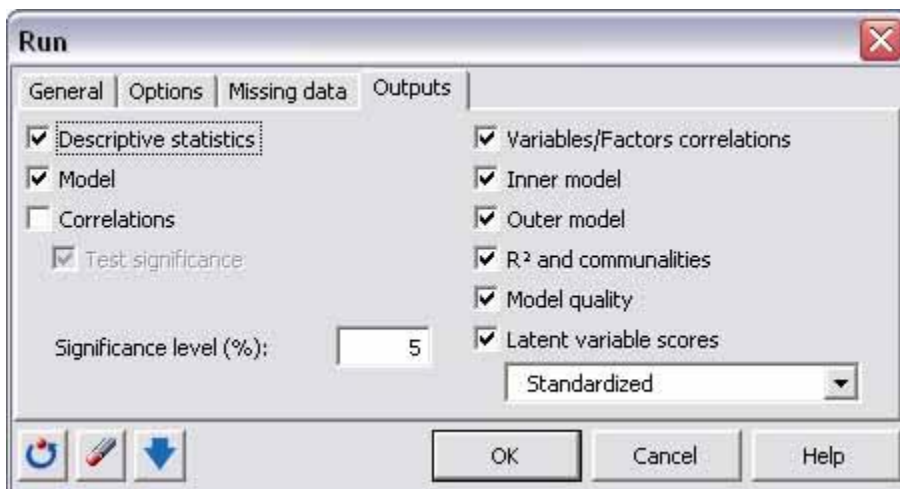
PLS path modeling is based on an iterative algorithm and thus should be initialized. For this application, manifest variables (observed variables) are treated with no prior transformations (4 different settings are available) because all variables are on the same scale. The initial values for the outer weights are the values of the first eigenvector when performing a principal component analysis on the manifest variables associated to a latent variable (2 different settings are available).



We use the centroid scheme for inner weights estimation. Confidence intervals are obtained using bootstrap resample.



In our simple example, there are no missing data in the dataset. We, thus, do not accept missing data.



Finally, for the output, all boxes are checked (except correlations) and we will study each output in the following part.

Results and interpretation of a PLS-PM project

In the results, information related to the manifest variables, the measurement model and the structural model are first summarized.

Numerical results for a PLS path modeling analysis

The first important elements are the composite reliability indexes:

Composite reliability:						
Latent variable	Dimensions	Cronbach's alpha	Dillon-Goldstein's rho (PC)	Condition number	Critical value	Eigenvalues
Image	5	0,714	0,815	2,302	397,512	954,371
						410,830
						238,913
						203,315
						180,130
Expectation	3	0,433	0,717	1,631	422,482	616,460
						419,159
						231,827
Perceived	7	0,872	0,903	3,642	349,451	1400,119
						321,027
						213,002
						159,806
						127,371
						119,276
Perceived	2	0,817	0,921	2,430	503,840	861,752
						145,928
Satisfaction	3	0,770	0,879	2,494	316,491	683,025
						156,609
						109,840
Complaints	1					
Loyalty	3	0,442	0,723	1,923	825,127	1190,304
						963,315
						321,763

In this application, latent variable are reflective. The blocks have to be one-dimensional. We can see that Dillon-Goldstein's rho is higher than 0.7 and that the first eigenvalue is always far greater than the second one. Expectation and loyalty have bad values for the Cronbach's alpha and a second dimension could be significant. In this tutorial, we will focus on the case of one dimension.

If you are interested in further dimensions, you can study the correlations between manifest variables and factors in a principal component analysis applied on each block of manifest variables. We will not focus on that point and consider only one dimension.

Applying PLS path modeling gives the table with GoF indexes:

Goodness of fit index (1):											
	GoF	GoF (Bootstrap)	Standard error	Critical ratio	Lower bound	Upper bound	95% Minimum	1st Quartile	Median	3rd Quartile	Maximum
Absolute	0,465	0,468	0,024	19,337	0,416	0,519	0,414	0,454	0,468	0,484	0,527
Relative	0,937	0,912	0,016	57,648	0,873	0,945	0,851	0,901	0,913	0,921	0,945
Outer model	0,996	0,994	0,002	513,091	0,988	0,997	0,988	0,993	0,994	0,996	0,997
Inner model	0,941	0,917	0,016	58,396	0,879	0,949	0,858	0,906	0,917	0,927	0,950

We can see the absolute GoF is 0.465, very close to the bootstrap estimate. This value is hard to interpret; it could be useful when comparing the global quality of two groups of observations or two different models. The relative GoF is very high. So are inner and outer models GoF.

Then, you should check the cross-loadings:

Cross-loadings (Monofactorial manifest variables / 1):							
	Image	Expectation	received Quar	received Va	Satisfactor	Loyalty	Complaints
IMAG1	0,717	0,347	0,571	0,393	0,540	0,338	0,423
IMAG2	0,566	0,387	0,492	0,269	0,398	0,293	0,188
IMAG3	0,658	0,272	0,367	0,332	0,339	0,309	0,207
IMAG4	0,792	0,374	0,571	0,459	0,542	0,461	0,440
IMAG5	0,698	0,340	0,544	0,260	0,501	0,485	0,337
CUEX1	0,349	0,687	0,437	0,293	0,362	0,268	0,183
CUEX2	0,404	0,644	0,343	0,175	0,345	0,320	0,225
CUEX3	0,285	0,726	0,357	0,273	0,300	0,190	0,126
PERQ1	0,622	0,534	0,778	0,454	0,661	0,461	0,380
PERQ2	0,405	0,308	0,651	0,295	0,474	0,319	0,300
PERQ3	0,621	0,423	0,801	0,467	0,651	0,461	0,472
PERQ4	0,480	0,388	0,760	0,390	0,587	0,353	0,379
PERQ5	0,598	0,406	0,732	0,455	0,517	0,373	0,389
PERQ6	0,551	0,447	0,766	0,405	0,539	0,333	0,418
PERQ7	0,596	0,411	0,803	0,547	0,707	0,446	0,465
PERV1	0,405	0,314	0,477	0,933	0,495	0,435	0,287
PERV2	0,542	0,354	0,594	0,911	0,629	0,525	0,360
CUSA1	0,558	0,495	0,637	0,403	0,711	0,484	0,334
CUSA2	0,524	0,395	0,672	0,480	0,872	0,484	0,416
CUSA3	0,612	0,382	0,684	0,588	0,885	0,609	0,547
CUSL1	0,430	0,281	0,393	0,407	0,455	0,855	0,237
CUSL2	0,109	0,095	0,065	0,148	0,115	0,273	0,122
CUSL3	0,528	0,351	0,537	0,481	0,658	0,869	0,448
CUSCO	0,469	0,250	0,537	0,348	0,540	0,401	1,000

In the case of our dataset, loadings between manifest variables and their own latent variable are the highest.

Then, outer weights and correlations are gathered in two large tables. If we study the correlations between manifest variables and latent variables:

Correlations (Dimension 1):									
Latent variable	Manifest variable	Standardized loadings	Communalities	Redundancies	Standardized loadings (Boc)	Standard error	Critical ratio (CF)	Upper bound (Ser bound (95%))	Lower bound (Ser bound (95%))
Image	IMAG1	0,717	0,514		0,710	0,047	15,312	0,613	0,798
	IMAG2	0,566	0,320		0,564	0,068	8,280	0,395	0,678
	IMAG3	0,658	0,433		0,661	0,065	10,193	0,507	0,785
	IMAG4	0,792	0,627		0,790	0,038	20,593	0,697	0,864
	IMAG5	0,698	0,487		0,693	0,039	18,111	0,622	0,775
Expectation	CUEX1	0,687	0,471	0,115	0,677	0,065	10,498	0,536	0,785
	CUEX2	0,644	0,415	0,101	0,629	0,115	5,595	0,314	0,830
	CUEX3	0,726	0,527	0,128	0,724	0,084	8,645	0,509	0,869
Perceived Quality	PERQ1	0,778	0,605	0,180	0,780	0,028	28,138	0,721	0,834
	PERQ2	0,651	0,423	0,126	0,652	0,051	12,874	0,551	0,746
	PERQ3	0,801	0,641	0,191	0,801	0,029	27,359	0,742	0,860
	PERQ4	0,760	0,578	0,172	0,754	0,043	17,553	0,655	0,843
	PERQ5	0,732	0,536	0,159	0,731	0,039	18,685	0,639	0,813
	PERQ6	0,766	0,587	0,174	0,767	0,055	13,874	0,625	0,864
	PERQ7	0,803	0,644	0,191	0,802	0,038	21,210	0,712	0,861
Perceived Value	PERV1	0,933	0,870	0,291	0,934	0,014	64,756	0,895	0,955
	PERV2	0,911	0,829	0,278	0,912	0,014	66,332	0,869	0,937
Satisfaction	CUSA1	0,711	0,505	0,339	0,712	0,038	18,909	0,624	0,791
	CUSA2	0,872	0,760	0,510	0,868	0,026	33,182	0,792	0,914
	CUSA3	0,885	0,783	0,526	0,881	0,023	38,158	0,824	0,922
Loyalty	CUSL1	0,855	0,731	0,316	0,848	0,039	21,717	0,748	0,913
	CUSL2	0,273	0,075	0,032	0,285	0,111	2,460	0,070	0,530
	CUSL3	0,869	0,755	0,326	0,863	0,026	32,899	0,793	0,909
Complaints	CUSCO	1,000		0,292	1,000	0,000		1,000	1,000

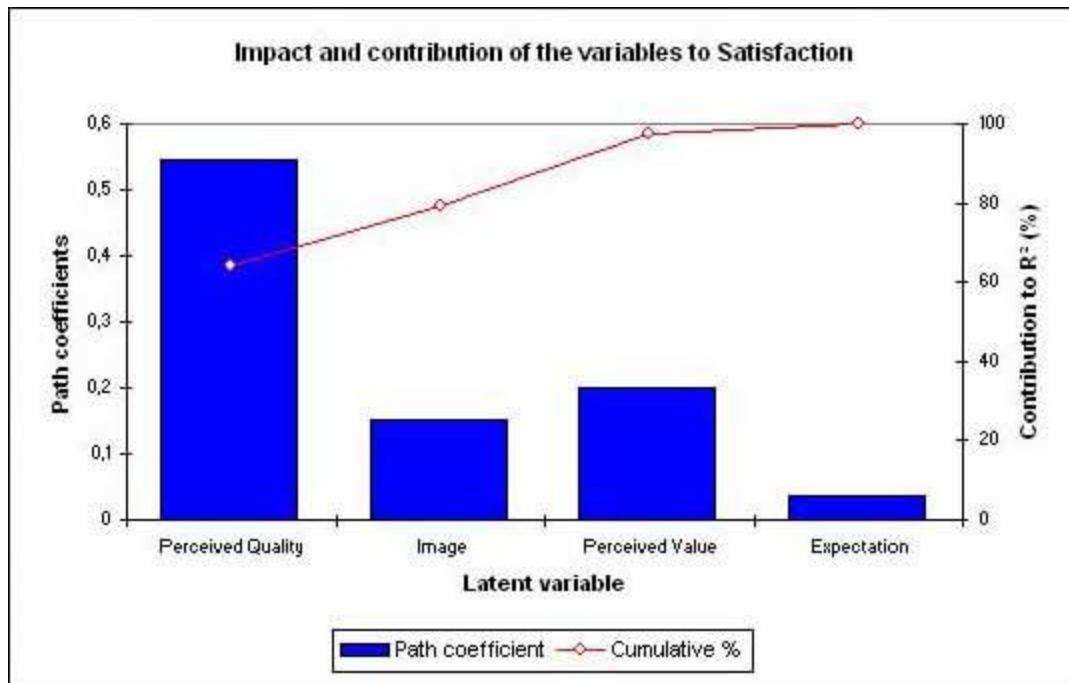
We can see that, for example, the manifest variables CUSA3 and CUSA2 have a greater effect on satisfaction than CUSA1. These tables allow you to see the impact of each manifest variables on its associated latent variable.

The results associated to the structural model follow. For each latent variable, information on the structural model is gathered. In the case of satisfaction, we have:

R ² (Satisfaction / 1):									
R ²	R ² (Bootstrap)	Standard error	Critical ratio (CR)	Lower bound (95%)	Upper bound (95%)				
0,672	0,678	0,040	16,721	0,597	0,752				
Path coefficients (Satisfaction / 1):									
Latent variable	Value	Standard error	t	Pr > t	Value(Bootstrap)	Standard error(Bootstrap)	Critical ratio (CR)	Lower bound (95%)	Upper bound (95%)
Intercept	3,102	6,021	0,515	0,000	1,019	5,423	0,572	-12,642	10,265
Image	0,170	0,061	2,776	0,006	0,181	0,060	2,851	0,073	0,327
Expectation	0,040	0,048	0,837	0,403	0,050	0,053	0,754	-0,063	0,139
Perceived Quality	0,586	0,064	9,128	0,000	0,555	0,085	6,934	0,339	0,724
Perceived Value	0,149	0,034	4,401	0,000	0,154	0,041	3,649	0,049	0,236
Equation of the model:									
Satisfaction = 3.10239801725722+0.170186817393756*Image+4.02146365022616E-02*Expectation+0.585996518809068*Perceived Quality+0.148779893519804*Perceived Value									
Standardized coefficients (Satisfaction / 1):									
Latent variable	Value	Standard error	t	Pr > t					
Image	0,153	0,055	2,776	0,006					
Expectation	0,037	0,044	0,837	0,403					
Perceived Quality	0,544	0,060	9,128	0,000					
Perceived Value	0,200	0,045	4,401	0,000					
Impact and contribution of the variables to Satisfaction (Dimension 1):									
	Perceived Quality	Image	Perceived Value	Expectation					
Correlation	0,791	0,671	0,604	0,481					
Path coefficient	0,544	0,153	0,200	0,037					
Correlation *	0,431	0,102	0,121	0,018					
Contribution	64,133	15,241	17,970	2,657					
Cumulative %	64,133	79,374	97,343	100,000					

A R² of 0.672 can be considered as a good result. We can see that perceived quality has the greatest effect on satisfaction and that the impact of expectation is not significant. The last table summarizes the results and shows that perceived value contributes to 64% of the R² of satisfaction.

The chart illustrates these results:



The next table shows different predictive quality indexes associated to both outer and inner models for each latent variable. Mean values of these indices give a global quality value. We see that the mean of all R² is 0.378 and the R² of satisfaction is the highest one. Communalities are always greater than redundancies because PLS-PM favours the measurement model in its estimation procedure.

Model assessment (Dimension 1):											
Latent variable	Type	Mean	R ²	Adjusted R ²	in Communalities (Mean Redundancy)	D.G. rho	Mean(Bootstrap	Critical ratio (CF)	er bound (Ser bound (95%)		
Image	Exogenous	72,685			0,476	0,818	72,530	76,410	70,711	74,587	
Expectation	Endogenous	72,325	0,243	0,243	0,471	0,115	72,212	73,433	70,464	74,704	
Perceived Quality	Endogenous	74,584	0,297	0,297	0,574	0,170	74,474	81,170	72,540	76,164	
Perceived Value	Endogenous	61,584	0,335	0,332	0,850	0,285	61,580	48,625	58,501	64,013	
Satisfaction	Endogenous	71,250	0,672	0,668	0,683	0,458	71,270	73,336	68,720	73,432	
Complaints	Endogenous	67,423	0,292	0,292	1,000	0,292	1,000	67,643	38,656	63,734	71,201
Loyalty	Endogenous	69,292	0,432	0,427	0,520	0,225	69,329	39,022	66,042	73,221	
Mean			0,378		0,570	0,257					

One of the greatest advantages of PLS-PM is the latent variable scores. They are given and can be used for other statistical treatments with XLSTAT.

This study has shown how to use XLSTAT-PLS-PM module in the case of real data. Once the model has been drawn, the procedure is simple. Once the model has been validated, interpretation of the result can be done by reading the tables with path coefficients and correlations.

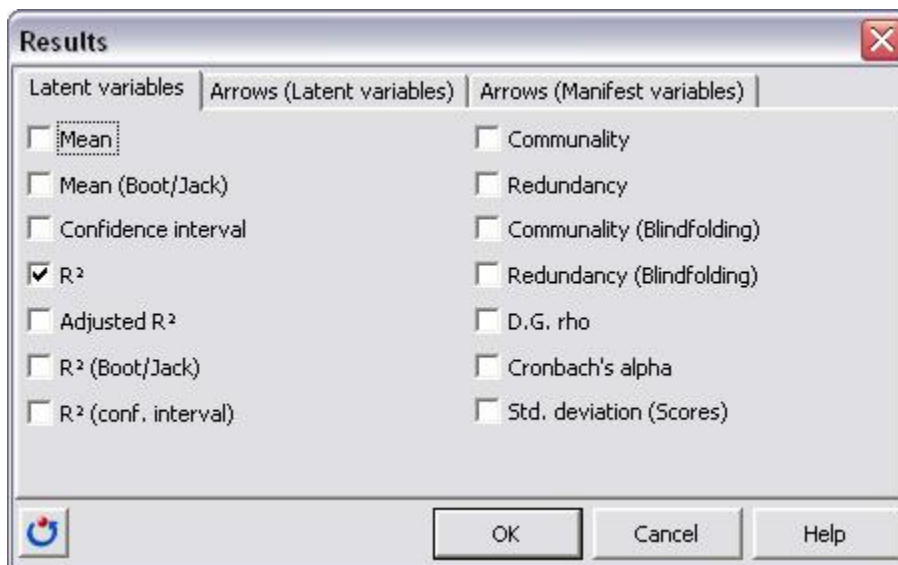
Graphical output for PLS path modelling analysis

You can display many type of results on the path model with XLSTAT-PLSPM. Choose between al the possible indexes obtained when clicking on the button “Choose the results to display”:

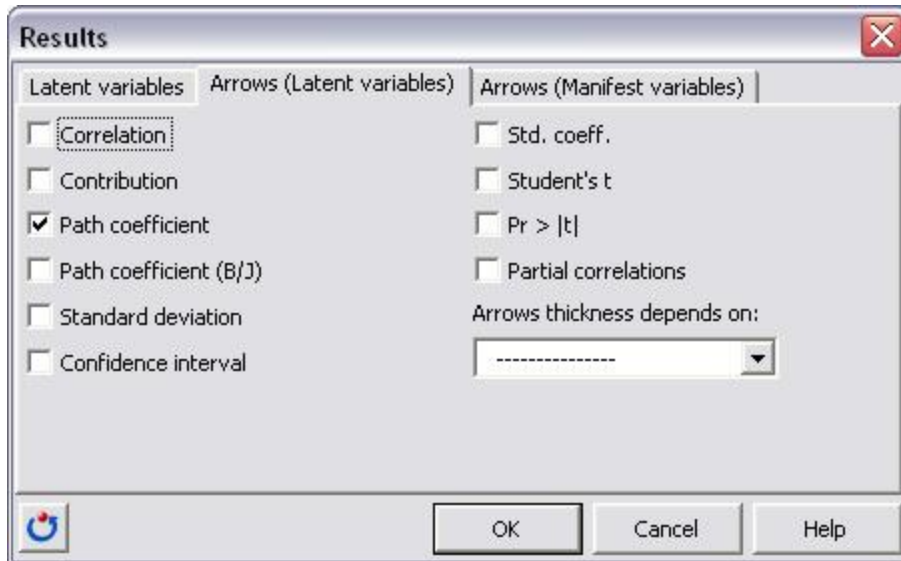


The “results” dialog box appears. It has three pages:

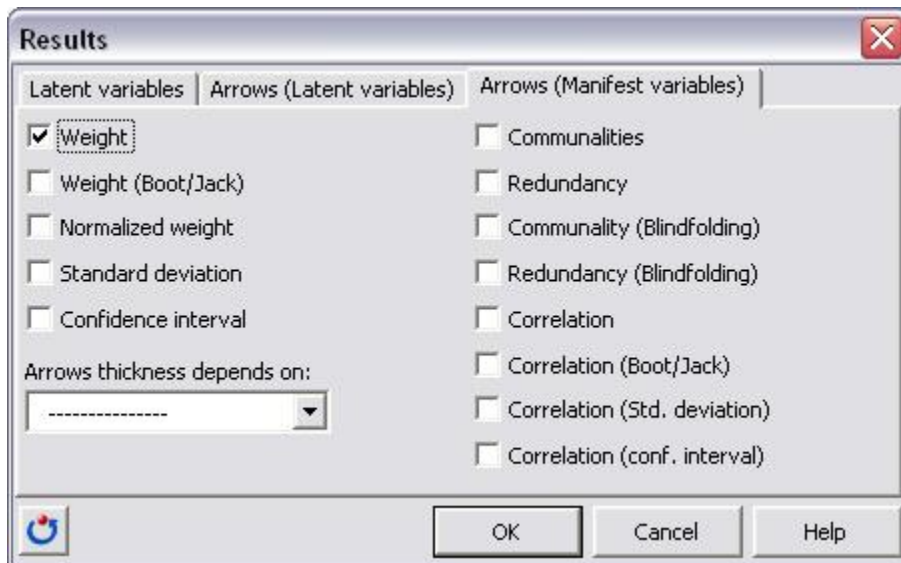
1. the first one concerns the display of indexes on the latent variables:



2. the second one concerns the display of coefficients and indexes on the arrows between latent variables



- the third one concerns the display of coefficients and indexes on the arrows between manifest variables and latent variables:



Results appear on the path model in the worksheet PLSPMgraph when pushing the button “display result”.



You can select the entire diagram and copy it to any other document.

