

PLS-VB programme for path modelling with latent variables

Programma PLS-VB per Modelli ad Equazioni Strutturali

Giuseppe Boari, Gabriele Cantaluppi
Istituto di Statistica, Università Cattolica del S. Cuore, Milano
e-mail: giuseppe.boari@unicatt.it, gabriele.cantaluppi@unicatt.it

Riassunto: La stima dei modelli ad equazioni strutturali con variabili latenti viene solitamente effettuata con il metodo denominato PLS (Partial Least Squares). A tale scopo viene presentato il programma PLS-VB, sviluppato in linguaggio Visual Basic Application come modulo aggiuntivo di Microsoft Excel.

Keywords: Structural Equation Models, Latent Variables, Partial Least Squares.

1. Reference model and program description

A structural model may be formulated by the following expressions, which show the relationships among q exogenous and $p-q$ endogenous latent variables (inner model) and the measurement linear process relating each latent variable Y_j to its corresponding manifest indicators $X_{jh}, j=1, \dots, p, h=1, \dots, H_j$ (outer model):

$$[Y_{q+1} \dots Y_p]' = [\Gamma|\mathbf{B}][Y_1 \dots Y_q \ Y_{q+1} \dots Y_p]' + [\zeta_{q+1} \dots \zeta_p]', \quad X_{jh} = \lambda_{jh}Y_j + \varepsilon_{jh}$$

where the elements of the matrix $[\Gamma|\mathbf{B}]$, of dimension $(p-q) \times p$, are the regression coefficients β_{jk} ($j=q+1, \dots, p, k=1, \dots, j-1$). The previous relationships may also be defined with reference to a square matrix $\mathbf{T} = \{t_{jk}\}$ of dimension $p \times p$, with values $t_{jk}=1$ if Y_j depends on Y_k and $t_{jk}=0$ otherwise. We consider models of the recursive type, for which the elements on the diagonal and in the upper triangular part of $\mathbf{T}$ are zeroes, as well as those in the first q rows (corresponding to the exogenous variables). Note that the matrix $[\Gamma|\mathbf{B}]$ corresponds to the last $(p-q)$ rows of $\mathbf{T}$.

For any assigned model structure the PLS-VB programme, according to the algorithm proposed by Lohmöller (1989) – widely used in psychometric and socio-economic research for estimating path model relationships – computes the proper β_{jk} parameters, once the latent scores have been calculated. One may choose, among all competing models, the most parsimonious best describing the linear relationship among exogenous and endogenous variables. To this aim, the r “target” variables (of primary interest with regard to the problem under investigation) are to be specified first.

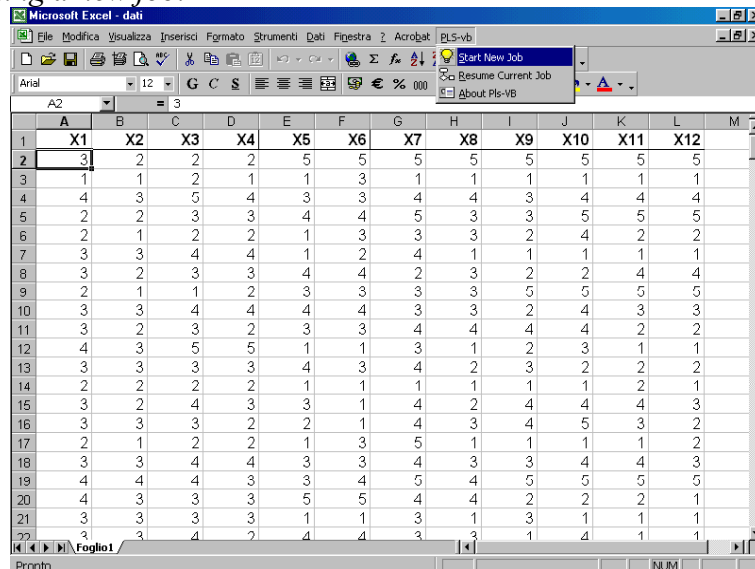
These endogenous variables, the last ones for the sake of simplicity, are considered as basic reference in the criterion adopted to evaluate the model performance. In particular, the criterion chooses, among the models with significant non zero β_{jk} OLS estimates, the one which attains the minimum of the following function:

$$G(g_{p-r+1}, \dots, g_p) = \sum_{j=p-r+1}^p \left(\ln \hat{\sigma}_j^2 + g_j \frac{\ln n}{n} \right) = \sum_{j=p-r+1}^p G_j, \quad (1)$$

where $\hat{\sigma}_j^2$ ($j=p-r+1, \dots, p$) are the residual variances of the regression models relating the r “target” variables Y_j to the corresponding regressors, amounting to $g_j > 0$; note that for $g_j = 0$ we have $\hat{\sigma}_j^2 = \text{Var}(Y_j)$. Observe that (1) recalls the BIC criterion structure, which takes into account a penalty term depending on the model complexity; moreover, it is also equivalent to compare the residual variance proportions.

Once installed the “plsvb.xla” add-in tool, the application may be started with a new job referring to a data set previously loaded or filled into the Excel sheet (see Figure 1). Note that the first line data range should include the manifest variable names, which the program will use further, while the first column is (not necessarily) devoted to contain an individual identification field (case or number name). The current job can be resumed at the point where we decided to hide the dialog box in each of the following steps. Furthermore, whenever a step has been completed, one can press the “Next” Button to advance to the next available step. Step 1 defines the data range, the extent of the scale, which is assumed to be common to all the X observable variables, and the number of Y latent variables, see Figure 2(a). As displayed in Figure 2(b), Step 2 describes the so-called outer (or measurement) model, expressing the reflexive relationships between the latent and the corresponding observable variables. Cronbach’s Alpha Button produces a report (see Figure 3) concerning the reliability analysis of the single scales (the groups of observable variables linked to each latent variable); the “ α if” statistic (alpha if item deleted) is also printed for all the scales containing more than two proxy variables. The next two steps are devoted to the inner (causal or path) model specification. During Step 3 (see Figure 4,a) the latent variables are classified by selecting from the following list: Exogenous, Endogenous, Of Interest and Ignored. The latent variables “Of Interest” are those Endogenous variables with respect to which the G statistic, see (1), is computed (by summing up the G_j statistics computed for the corresponding linear regression sub-models), whereas an “Ignored” latent variable will be excluded (together with all its corresponding observables) from the subsequent analyses. In Step 4 (see Figure 4,b) relationships among the latent variables are defined, by filling the **T** matrix of dependencies: a unitary value means that the corresponding column variable is connected to the corresponding row variable through an oriented arrow, while a null value means no relationship (a double click toggles between these values).

Figure 1 *Starting a new job.*



	A	B	C	D	E	F	G	H	I	J	K	L	M
	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	
1													
2	3	2	2	2	5	5	5	5	5	5	5	5	
3	1	1	2	1	1	3	1	1	1	1	1	1	
4	4	3	5	4	3	3	4	4	3	4	4	4	
5	2	2	3	3	4	4	5	3	3	5	5	5	
6	2	1	2	2	1	3	3	3	2	4	2	2	
7	3	3	4	4	1	2	4	1	1	1	1	1	
8	3	2	3	3	4	4	2	3	2	2	4	4	
9	2	1	1	2	3	3	3	3	5	5	5	5	
10	3	3	4	4	4	4	3	3	2	4	3	3	
11	3	2	3	2	3	3	4	4	4	4	2	2	
12	4	3	5	5	1	1	3	1	2	3	1	1	
13	3	3	3	3	4	3	4	2	3	2	2	2	
14	2	2	2	2	1	1	1	1	1	1	2	1	
15	3	2	4	3	3	1	4	2	4	4	4	3	
16	3	3	3	2	2	1	4	3	4	5	3	2	
17	2	1	2	2	1	3	5	1	1	1	1	2	
18	3	3	4	4	3	3	4	3	3	4	4	3	
19	4	4	4	3	3	4	5	4	5	5	5	5	
20	4	3	3	3	5	5	4	4	2	2	2	1	
21	3	3	3	3	1	1	3	1	3	1	1	1	
22	3	3	4	2	4	4	3	3	1	4	1	1	

Figure 2 *Input step and outer model definition.*

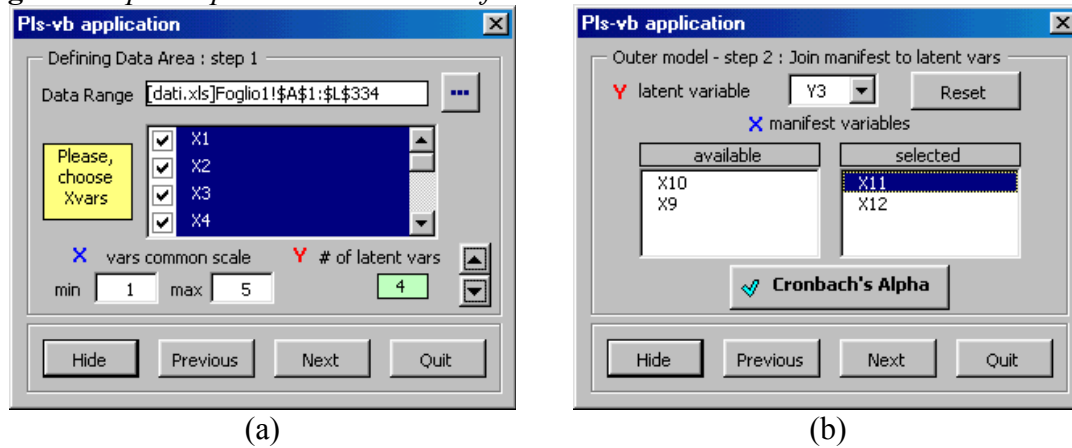
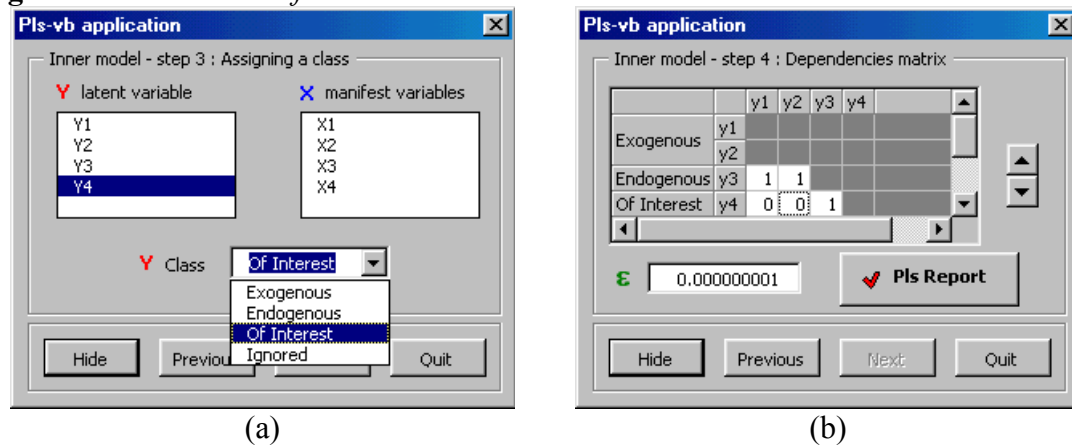


Figure 3 *Reliability analysis report.*

α if deleted						Cronbach α			
Yvar	Xvar	Num	Avg	Var	α if	Yvar	Num	Count	α
1	X7	2	3.1862	1.1460		1	2	333	70.79%
1	X8	2	2.5315	1.3943		2	2	333	78.37%
...	...	...	...	...	...	3	4	333	82.89%
4	X1	4	2.9850	0.4425	58.43%	4	4	333	70.80%
4	X2	4	2.4234	0.6063	62.98%				
4	X3	4	3.3273	0.5160	66.97%				
4	X4	4	2.9730	0.8095	69.88%				

Figure 4 *Inner model definition.*



The fixed constant ϵ , used by the program to terminate the iteration procedure of the PLS algorithm, can also be redefined.

The “PLS Report” Button produces the final reports (see Figure 5 for a comprehensive view) displaying model parameter estimates (β_{ij} s are printed on the right most position) as well as all statistics needed to check model fitting (see Bayol *et al.*, 2000 and Lohmöller, 1989, for an in-depth description).

The various reports will be stored in proper worksheets automatically created by the programme. The last one produced (as displayed in Figure 6) contains the latent score estimates, expressed in the original scale, as well as in centesimal and standard units.

The following technical requirements are at least necessary for PLS-VB programme properly run: Win95/98 or Windows NT and Microsoft Excel 2000.

Figure 5 A comprehensive view of PLS output.

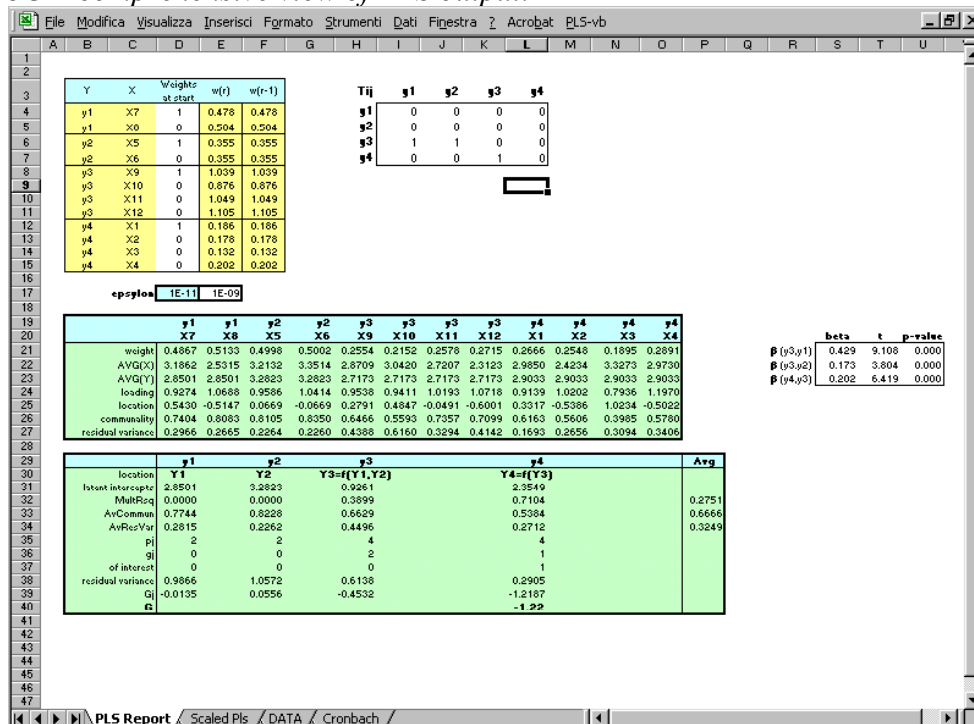
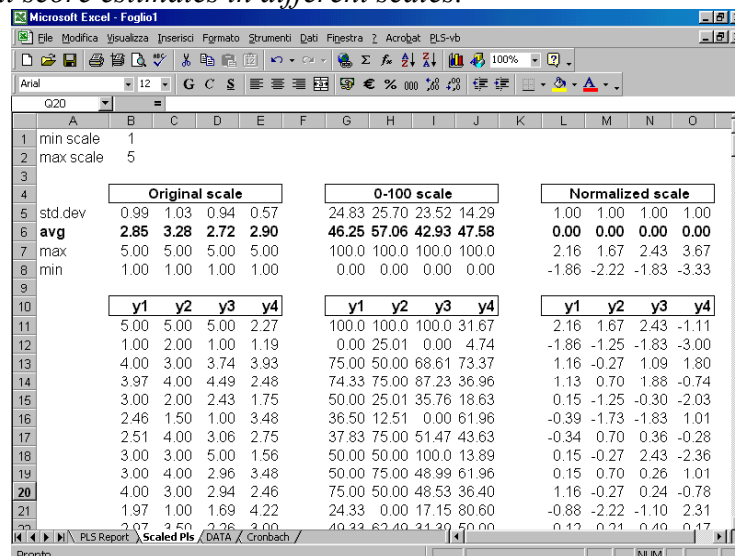


Figure 6 Latent score estimates in different scales.



References

- Bayol M.P., de la Foye A., Tellier C., Tehenhaus M. (2000) Use of the PLS Path Modelling to estimate the European Customer Satisfaction Index (ECSI) model, *Statistica Applicata*, vol. 12(3), 361-375.
- Boari G., Cantaluppi G. (2003) Reperimento di modelli ad equazioni strutturali parsimoniosi con il programma PLS-VB, Serie E.P., N° 115, Istituto di Statistica, Università Cattolica del S. Cuore, Milano.
- Lohmöller J.-B. (1989) *Latent Variable Path Modeling with Partial Least Squares*, Physica-Verlag, Heidelberg.