



Une comparaison expérimentale entre les approches PLS et LISREL

Valentina Stan, Gilbert Saporta

► To cite this version:

Valentina Stan, Gilbert Saporta. Une comparaison expérimentale entre les approches PLS et LISREL. 38èmes Journées de Statistique, May 2006, Clamart, France. hal-01125190

HAL Id: hal-01125190

<https://hal.science/hal-01125190>

Submitted on 25 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNE COMPARAISON EXPÉRIMENTALE ENTRE LES APPROCHES PLS ET LISREL

Valentina Stan; Gilbert Saporta

Conservatoire National des Arts et Métiers
292 Rue Saint Martin; F 75141 Paris Cedex 03

valentina_titu@yahoo.fr saporta@cnam.fr

Résumé : Les modèles d'équations structurelles sont utilisés en marketing spécialement dans les études de satisfaction. Nous présentons les principaux concepts des modèles d'équations structurelles. Une comparaison sur le même jeu de données entre PLS (Partial Least Squares) et LISREL (Linear Structural Relationships) est réalisée. Les aspects théoriques sont vérifiés : PLS favorise le modèle externe et LISREL le modèle interne.

Mots clés : marketing, satisfaction, modèles d'équations structurelles, approche PLS, LISREL

Abstract: Structural equation modeling is used in marketing, especially for consumer satisfaction. We present the main concepts of structural equation modeling. A comparison between PLS (Partial Least Squares) and LISREL (Linear Structural Relationships), is realized on the same data. The theoretical aspects are verified: PLS favours the outer model and LISREL inner model.

Keywords: marketing, satisfaction, structural equation modeling, PLS, LISREL

1. Introduction

Les modèles d'équations structurelles sont issus des recherches sur la causalité, les approches les plus connues étant LISREL (Linear Structural Relationships) et PLS. Elles sont utilisées dans plusieurs domaines dont le marketing (spécialement pour des études de satisfaction), la sociologie et la psychologie.

L'approche la plus utilisée est LISREL (Jöreskog & al, 1982), mais PLS (Wold, 1975; 1985), moins connue, présente une série d'avantages par rapport à la première. Ces avantages sont présentés dans la comparaison réalisée au paragraphe 3.

Après une brève présentation des concepts utilisés dans les modèles d'équations structurelles, nous comparons sur le même jeu de données les résultats obtenus en utilisant LISREL et PLS.

2. Généralités

Un modèle structurel utilise 2 types de sous-modèles :

- le modèle « externe » ou modèle de mesure, qui lie les variables manifestes et leurs latentes;
- le modèle « interne » ou structurel qui connecte les variables latentes entre elles.

2.1. Modèle « externe » ou de mesure

Il faut spécifier le type de relations entre les variables latentes et leurs variables manifestes. Il y a 2 possibilités principales : le type réflectif et le type formatif.

a) Type réflectif

Le type réflectif requiert l'unidimensionnalité du bloc. On considère un modèle de régression simple de chaque manifeste sur sa latente (Tenenhaus & al, 2005) :

$$x_{jh} = \pi_{jh}^0 + \pi_{jh} \xi_j + \varepsilon_{jh} \quad \forall h=1, \dots, K_j; \quad \forall j=1, \dots, J$$

π_{jh}^0 = terme constant; π_{jh} = coefficient de régression; ξ_j = variable latente; ε_{jh} = terme résiduel

b) Type formatif

La variable latente est une fonction linéaire des variables manifestes associées (Tenenhaus & al, 2005): $\xi_j = \sum_{h=1}^{k_j} \varpi_{jh} x_{jh} + \delta_j$

$\varpi_{jh} (h=1, \dots, k_j)$ = coefficients de la régression multiple de ξ_j sur les x_{jh} ; δ_j = terme résiduel

2.2. Modèle « interne » ou structurel

Ce modèle concerne l'estimation des relations entre les variables latentes. Les relations entre les variables latentes sont de la forme suivante : $\xi_j = \beta_j^0 + \sum_{i=1, i \neq j}^J \beta_{ji} \xi_i + \zeta_j \quad \forall j=1, \dots, J$

β_j^0 = terme constant; β_{ji} = coefficient de régression; ζ_j = terme résiduel.

2.3. Approche LISREL

LISREL (Jöreskog & al, 1982) estime à l'aide du maximum de vraisemblance le système d'équations structurelles. Chaque variable manifeste s'écrit en fonction de sa variable latente : $x_{jh} = \lambda_{jh} \xi_j + \varepsilon_{jh}$. Le premier coefficient λ_{j1} est fixé à 1 et les autres sont estimés par maximum de vraisemblance en supposant que les variables manifestes suivent une loi multinormale.

2.4. L'algorithme PLS

L'algorithme utilisé dans l'approche PLS consiste en l'estimation, d'une part, des variables latentes et d'autre part, des équations structurelles. Pour les variables latentes, l'estimation se fait par itération entre l'estimation externe et l'estimation interne. Dans l'estimation externe, les variables latentes standardisées sont estimées comme une combinaison linéaire de leurs variables manifestes centrées. Dans l'estimation interne, on estime les variables latentes en considérant les relations les reliant. La valeur initiale des coefficients de régression est égale à ± 1 , en fonction de la corrélation entre les variables latentes, ou de celle entre les variables latentes et leurs manifestes associées.

Les équations structurelles sont estimées à l'aide de régressions multiples classiques (OLS) dans lesquelles les variables latentes sont remplacées par leurs estimations.

3. Comparaison entre PLS et LISREL

Le tableau suivant, inspiré de Jöreskog et Wold (1982), Chin (2000), Esposito Vinzi (2003) présente une comparaison entre PLS et LISREL selon différents critères :

Critère	PLS	LISREL
Objectif	Orientée vers la réalisation des prévisions	Orientée vers l'estimation des paramètres
Méthodologie	Basée sur variance	Basée sur covariance
Variables latentes (VL)	Combinaison linéaire de ses VM	Combinaison linéaire de toutes les VM
Relations entre VL et VM associées	Type réflectif ou formatif	Type réflectif
Optimalité	Pour la précision des prévisions	Pour la précision des paramètres
Qualité des sous-modèles	Modèle externe meilleur car les VL sont contenues dans l'espace de leurs VM	Modèle interne meilleur car les VL sont estimées dans un espace non restreint
Hypothèses	Unidimensionnalité (réflectif)	Multinormalité des données + unidimensionnalité
Complexité modèle	Grande (ex: 100 VL, 1000 VM)	Réduite ou modéré (<100 VM)
Taille min échantillon	30-100 cas	200-800 cas
Traitement données manquantes	NIPALS	Maximum de vraisemblance
Identification	Dans le cadre du modèle récursif, toujours identifiée	Dépend du modèle; idéal : 4 ou plusieurs VM pour une VL, 3 pour être correctement identifiée

Table 1 : Comparaison PLS – LISREL

4. Validation du modèle

Pour vérifier «la qualité» du modèle obtenu en utilisant l'approche LISREL, il y a plusieurs possibilités parmi lesquelles : le test du χ^2 (Jöreskog, 1967), le test du χ^2 de Satorra-Bentler (1999), la racine carrée de la moyenne des erreurs d'approximations (Steiger et Lind, 1980).

Dans l'approche PLS il n'existe pas d'indice global permettant de juger la qualité du modèle dans son ensemble, car il n'existe pas de fonction à optimiser comme en LISREL.

On définit trois niveaux de validation du modèle : la qualité du modèle externe, la qualité du modèle interne, la qualité de chaque équation structurelle.

La communauté évalue la qualité du modèle externe pour chaque bloc. Pour un bloc j, on a :

$$communauté_j = \frac{1}{p_j} \sum_{h=1}^{p_j} cor^2(x_{jh}, Y_j) \quad \overline{communauté} = \frac{1}{p} \sum_{j=1}^J p_j communauté_j$$

$$Y_j = \text{estimations externes des variables latentes} : \begin{cases} Y_j \propto \pm \left[\sum w_{jh} (x_{jh} - \bar{x}_{jh}) \right] \\ w_{jh} = \text{poids externes (outer weights)} \end{cases}$$

La redondance évalue la qualité du modèle interne pour chaque bloc endogène j :

$$redondance_j = communauté_j \times R^2(Y_j, \{Y_j', \text{explique } Y_j\})$$

R^2 est le coefficient de détermination et mesure la qualité du modèle interne. Il est calculé pour chaque variable endogène, en fonction des variables latentes explicatives.

Amato et al. (2004) ont proposé un critère global qui peut être utilisé pour choisir un modèle parmi

plusieurs : $GoF = \sqrt{communauté \times R^2}$

5. Application pratique

Nous appliquons l'approche PLS (avec le logiciel SPAD 5.6) et l'approche LISREL (avec le logiciel LISREL 8.54) sur des données aimablement fournies par la compagnie PSA (Peugeot Citroën).

5.1. Questionnaire

Les données représentent des notes de satisfaction sur une échelle de 0 à 10, de 32 prestations, recueillies auprès d'environ 3000 clients.

Variable	Libellés	Variable	Libellés
Satisfaction générale	S01	Autoradio - lecteur de cassette-CD	S17
Qualité en général	S02	Chauffage - ventilation	S18
Rapport qualité - prix	S03	Capacité du coffre	S19
Absence de petits défauts irritants	S04	Sécurité	S20
Absence de bruits et vibrations	S05	Freinage	S21
Etat général de la peinture	S06	Accélération	S22
Robustesse des commandes, boutons et commodos	S33	Tenue de route	S23
Solidité et robustesse de la carrosserie	S08	Confort de suspension	S24
Mécanismes des serrures, portes et fenêtres	S09	Silence en roulant	S25
Modularité de l'espace intérieur et des sièges	S34	Maniabilité	S26
Habitabilité intérieure	S11	Direction	S27
Tableau de bord: qualité des matériaux et finition	S12	Passage de vitesses	S28
Intérieur: qualité des matériaux et finition	S13	Fiabilité mécanique	S29
Confort des sièges avant	S14	Consommation de carburant	S30
Position de conduite	S15	Efficacité du garagiste	S31
Visibilité à partir du siège conducteur	S16	Coûts d'entretien et réparations	S32

Table 2 : Variables manifestes

5.2. Schéma de causalité

A l'aide d'experts, les variables ont été regroupées en 8 blocs (Stan et Saporta, 2005), liés par le schéma de causalité suivant :

Blocs	VM	Blocs	VM	Blocs	VM
Satisfaction générale (SG)	S01	Solidité (SOLI)	S06	Qualité conduite (QCON)	S20
	S02		S08		S21
	S03		S09		S22
Qualité de construction (QC)	S04	Confort conduite (CC)	S33		S23
	S05		S12		S24
Maintenance (MEN)	S31		S13		S25
	S32		S14		S26
Design intérieur (DI)	S11		S15		S27
	S34		S16		S28
	S19	Confort intérieur (CI)	S17		S29
			S18		S30

Table 3 : Blocs de variables manifestes

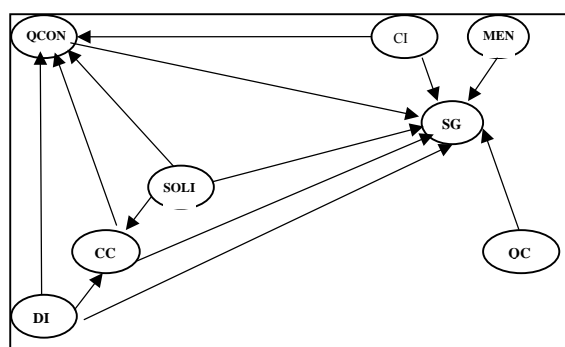


Figure 1 : Schéma de causalité

5.3. Résultats

Modèle externe

A la fin de l'algorithme de l'approche PLS on obtient les poids finaux permettant de relier les variables manifestes aux variables latentes du modèle. Un exemple pour la satisfaction générale :

$SG = 0,22S01 + 0,57S02 + 0,48S03$. Dans l'approche LISREL il n'existe pas d'équations de ce type; chaque manifeste s'écrit en fonction de la latente associé, par exemple : $S01 = 0,96SG + \varepsilon$

Le tableau suivant montre que toutes les variables latentes sont bien corrélées avec leurs manifestes, donc les blocs thématiques semblent bien construits.

Corrélations PLS						Corrélations LISREL					
S01	0,6442	S11	0,8801	S21	0,6739	S01	0,608	S11	0,735	S21	0,624
S02	0,8706	S34	0,8286	S22	0,6757	S02	0,872	S34	0,806	S22	0,632
S03	0,7397	S19	0,7015	S23	0,7716	S03	0,424	S19	0,548	S23	0,742
S04	0,8780	S12	0,7711	S24	0,7362	S04	0,742	S12	0,648	S24	0,700
S05	0,8449	S13	0,7903	S25	0,6218	S05	0,656	S13	0,678	S25	0,574
S31	0,8739	S14	0,8122	S26	0,7837	S31	0,775	S14	0,800	S26	0,775
S32	0,8332	S15	0,8283	S27	0,7841	S32	0,592	S15	0,825	S27	0,775
S06	0,6534	S16	0,6595	S28	0,6396	S06	0,520	S16	0,592	S28	0,592
S08	0,7867	S17	0,8110	S29	0,5690	S08	0,574	S17	0,600	S29	0,500
S09	0,7057	S18	0,8665	S30	0,4743	S09	0,663	S18	0,678	S30	0,412
S33	0,7061	S20	0,7315			S33	0,600	S20	0,686		

Table 4 : Corrélations entre les variables manifestes et leurs latentes

Ces résultats montrent également que les corrélations entre les variables manifestes et leurs latentes sont plus importantes en PLS qu'en LISREL, car la première approche favorise le modèle externe.

Modèle interne

Les coefficients de détermination obtenus permettent de vérifier que les équations structurelles sont moins significatives en PLS qu'en LISREL (les R^2 sont plus forts) - la deuxième approche favorise le modèle interne :

PLS

$$R^2(CC; SOLI, DI) = 0,42$$

$$R^2(QCON; SOLI, DI, CC, CI) = 0,5$$

$$R^2(SG; QC, MEN, SOLI, DI, CC, CI, QCON) = 0,27$$

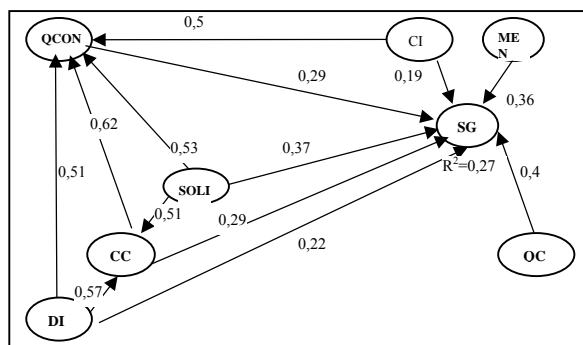


Figure 2 : Schéma de causalité PLS avec les corrélations

LISREL

$$R^2(CC; SOLI, DI) = 0,57$$

$$R^2(QCON; SOLI, DI, CC, CI) = 0,56$$

$$R^2(SG; QC, MEN, SOLI, DI, CC, CI, QCON) = 0,42$$

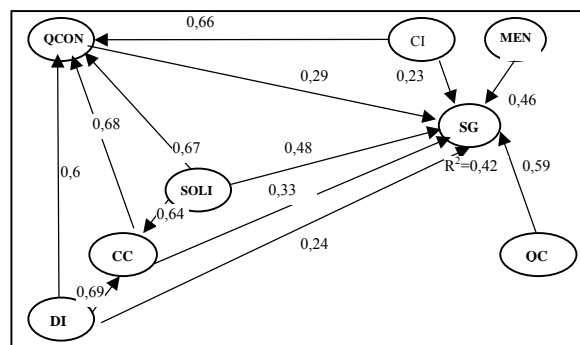


Figure 3 : Schéma de causalité LISREL avec les corrélations

Le tableau suivant présente les corrélations entre les variables latentes :

	PLS									LISREL							
	SG	QC	MEN	SOLI	DI	CC	CI	QCON		SG	QC	MEN	SOLI	DI	CC	CI	QCON
SG	1								1								
QC	0,40	1							0,59	1							
MEN	0,36	0,35	1						0,46	0,56	1						
SOLI	0,37	0,34	0,29	1					0,48	0,47	0,42	1					
DI	0,22	0,10	0,20	0,42	1				0,24	0,12	0,23	0,58	1,00				
CC	0,29	0,15	0,23	0,51	0,57	1			0,33	0,23	0,26	0,64	0,69	1			
CI	0,19	0,12	0,23	0,40	0,38	0,45	1		0,23	0,19	0,36	0,65	0,57	0,51	1		
QCON	0,29	0,20	0,3	0,53	0,51	0,62	0,50	1	0,29	0,24	0,3	0,67	0,6	0,68	0,66	1	

Table 5 : Corrélations entre les variables latentes

Les résultats obtenus montrent que les corrélations entre les variables latentes sont plus fortes en LISREL qu'en PLS, car LISREL favorise le modèle interne.

On peut constater que la «Qualité de construction» est la variable la plus importante pour la «Satisfaction générale» (coefficient de corrélation = 0,4 en PLS et 0,59 en LISREL) et que la moins importante est la «Confort intérieur» (coefficient de corrélation = 0,19 en PLS et 0,23 en LISREL). En conséquence, pour augmenter la «Satisfaction générale», le producteur devrait agir d'abord sur la variable «Qualité construction», puis sur la «Solidité», «Maintenance», «Confort conduite», «Qualité conduite», «Design intérieur», «Confort intérieur».

En conséquence, étant donné le schéma structurel, la détermination de la «Satisfaction générale» est un processus complexe où participent pratiquement toutes les variables latentes, l'équation étant la suivante :

$$PLS : SG = 0,26QC + 0,19MEN + 0,15SOLI + 0,03DI + 0,10CC - 0,03CI + 0,04QCON$$

$$LISREL : SG = 0,4QC + 0,15MEN + 0,24SOLI + 0,028DI + 0,097CC - 0,11CI - 0,28QCON$$

Le critère d'Amato favorise LISREL par rapport à PLS : $GoF_{PLS} = 0,48$

$$GoF_{LISREL} = 0,69$$

Mais les deux approches se rejoignent au niveau du calcul des scores.

Si PLS donne les scores des variables latentes, LISREL ne les fournit pas directement, mais ils peuvent être calculés à l'aide des équations de régression entre les variables latentes et variables

manifestes. Le tableau suivant présente les corrélations entre les scores des variables latentes obtenus en PLS et en LISREL. Toutes les corrélations entre les scores donnés par PLS et ceux donnés par LISREL sont supérieures à 0,94 :

$r(SG_{PLS}; SG_{LISREL})$	0,94
$r(QC_{PLS}; QC_{LISREL})$	0,99
$r(MEN_{PLS}; MEN_{LISREL})$	0,99
$r(DI_{PLS}; DI_{LISREL})$	0,98
$r(SOLI_{PLS}; SOLI_{LISREL})$	0,99
$r(CC_{PLS}; CC_{LISREL})$	0,99
$r(CI_{PLS}; CI_{LISREL})$	0,99
$r(QCON_{PLS}; QCON_{LISREL})$	0,99

Table 6 : Corrélations scores PLS et LISREL

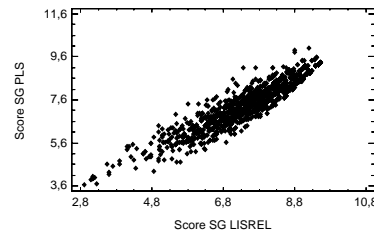


Figure 4 : Comparaison scores PLS et LISREL

Conclusions

Dans l'application pratique réalisée, les deux approches ont montré leur efficacité. Les aspects théoriques sont vérifiés : PLS favorise le modèle externe (corrélations plus fortes entre les variables manifestes et leurs latentes) et LISREL le modèle interne (R^2 plus forts). Pour la «Satisfaction générale», les coefficients de détermination obtenus par les deux approches représentent un bon résultat pour un échantillon de 2922 individus.

Selon le critère d'Amato, les performances de LISREL seraient meilleures que celles de PLS, mais les corrélations entre les scores donnés par PLS et ceux de LISREL, pour toutes les variables latentes sont supérieures à 0,94.

En réalité, l'objectif des deux approches n'est pas le même : pour réaliser des prévisions, PLS est meilleur que LISREL, qui est plus orienté vers l'explication du modèle. D'autre part, LISREL demande une hypothèse de multinormalité des données qui est rarement remplie en marketing.

Bibliographie

- Amato, S. & al. (2004) *A global goodness – of – fit index for PLS structural equation modelling*, Oral Communication to PLS Club, HEC School of Management, France
- Chin, W. W. (2000) *Partial Least Square for researchers: a overview and presentation of recent advances using the PLS approach*, <http://disc-nt.cba.uh.edu/chin/indx.html>
- Esposito Vinzi, V. (2003) *The PLS Approach to Path Modelling*, IASC – IFCS Summer School, Lisbonne
- Jöreskog, K. G. (1967) *Some contributions to maximum likelihood factor analysis*, Psychometrika, 32 (4), pages 443 - 482
- Jöreskog, K. G. et Wold, H. (1982) *The ML and PLS techniques for modelling with latent variables: historical and competitive aspect*, en Jöreskog, K. G. et Wold, H. (Editors), *Systems under indirect observation*, Part 1, pages 263 – 270, North – Holland, Amsterdam
- Steiger, J. H. et Lind, J. C. (1980) *Statistically based tests for the number of common factors*, Proceedings of the annual meeting of Psychometric Society
- Satorra, A. et Bentler, P. M. (1999) *Corrections to test statistics and standard errors in covariance structure analysis*, in Von Eye, A. et Clogg, C. (Editors), *Latent Variables Analysis, Applications to Developmental Research*, pages 399 – 419, Sage Thousand Oaks, Ca
- Stan, V. et Saporta, G. (2005) *Conjoint use of variables clustering and PLS structural equations modelling*, PLS'05, 4th International Symposium on PLS and Related Methods; Barcelone
- Tenenhaus, M. & al. (2005) *PLS path modelling*, Computational Statistics & Data Analysis, volume 48, pages 159-205
- Wold, H. (1975) *Soft modeling by latent variables: The Non – Linear Iterative Partial Least Squares (NIPALS) approach*, in Gani, J. (Editor), *Perspectives in probability and statistics*, pages 117 – 142, Londres
- Wold, H. (1985) *Partial Least Squares*, Kotz, S. et Johnson, N. L. (Editors), *Encyclopedia of Statistical Sciences*, Vol. 6, pages 581-591, John Wiley & Sons, New York