Les modèles d'équations structurelles à variables latentes

Emmanuel Jakobowicz
Addinsoft
jakobowicz(@)xlstat.com

23 mars 2011

Plan du cours

Aujourd'hui:

La théorie des modèles structurels à variables latentes

La semaine prochaine:

Des démonstrations et de petits exercices

Plan du cours d'aujourd'hui - 1

1. Introduction

Les concepts de base pour la modélisation par équations structurelles

2. Modélisation d'équations structurelles par le maximum de vraisemblance (LISREL)

Le modèle LISREL
Estimation du modèle
Indices de validation
Un exemple
Indices de modification

Plan du cours d’aujourd’hui - 2

3. Modélisation d’équations structurelles par l’approche PLS (Partial Least Squares Path Modeling)

Le modèle PLS
Le principe
L’algorithme PLS et ses variantes
L’initialisation des poids
Les indices de qualité d’ajustement
Un exemple

4. Comparaisons des 2 approches

Aspects théoriques
Aspects pratiques
Un petit exemple

Les concepts de base - 1

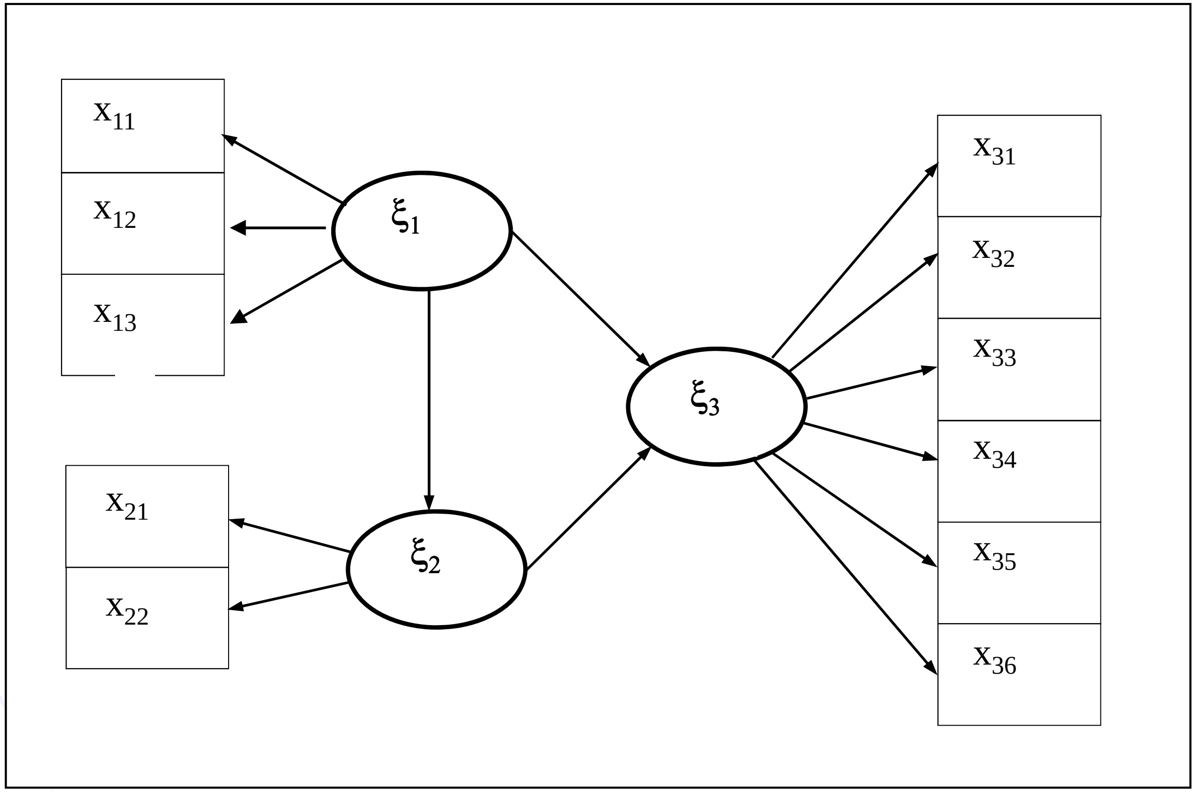
- Méthodes appartenant à la catégorie des modèles structurels à variables latentes (VL)
- *p* variables observées sur *n* individus, réparties en J blocs de k_j variables
- L'ensemble des variables sont continues
- Blocs reliés entre eux par un modèle de relations structurelles entre variables latentes

Les concepts de base - 2

Variables observées = **variables manifestes** (VM)

Les **variables latentes** (VL) non observées, existent au travers des variables manifestes avec lesquelles elles sont en relation

A chaque bloc X_j on associe une seule variable latente ξ_j

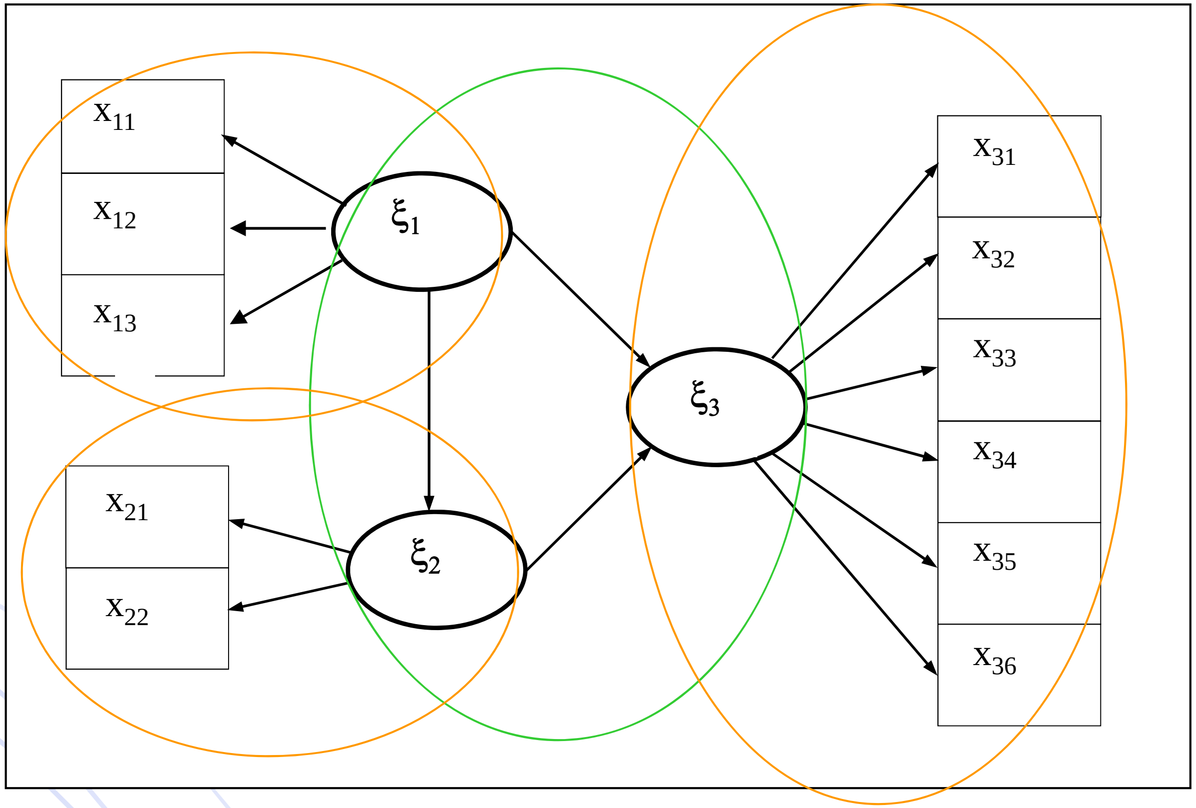


Les concepts de base - 3

2 sous-modèles :

- *Modèle « externe» ou de mesure - lie les VM et leurs VL;*
- *Modèle « interne» ou structurel - connecte les VL.*

Le modèle



Modèle structurel (interne) Modèle de mesure (externe)

Le modèle de mesure - 1

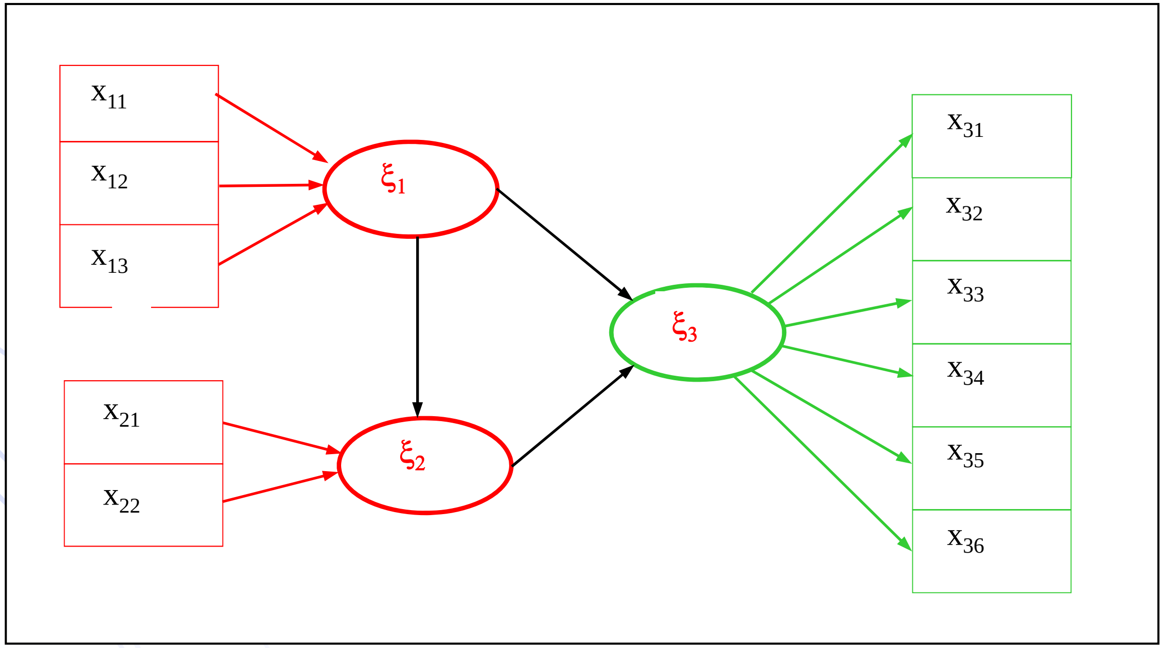
Le modèle « externe »

Type réflectif

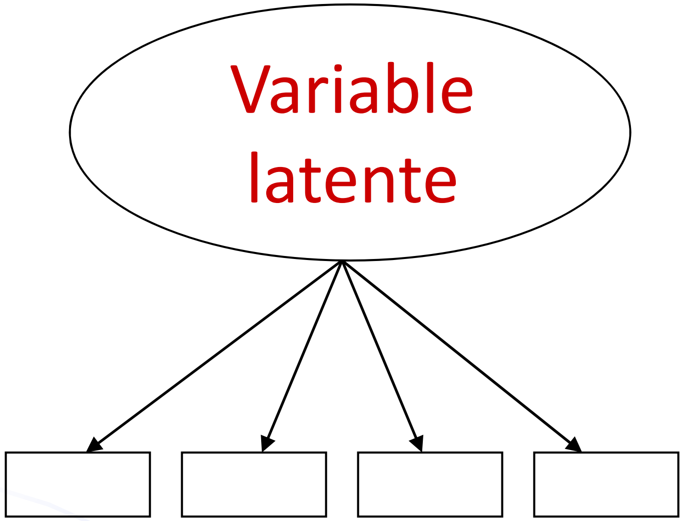
Les variables manifestes sont le reflet de leur variable latente

Type formatif

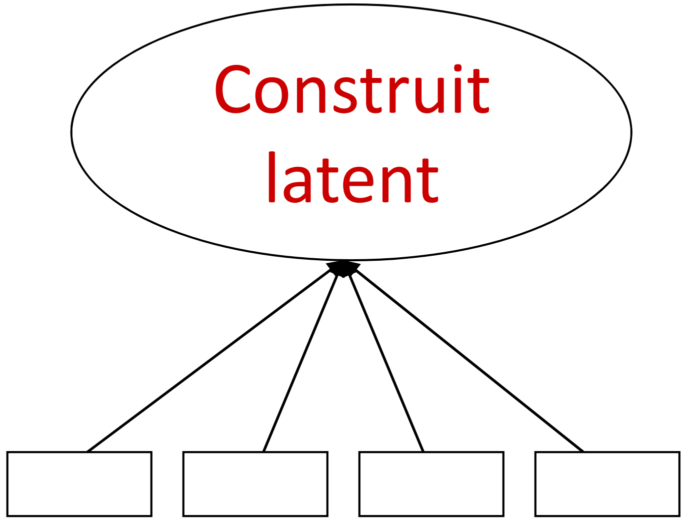
La variable latente ξ_j est le reflet des variables manifestes du bloc X_j



Le modèle de mesure - 2



Indicateurs réflectifs



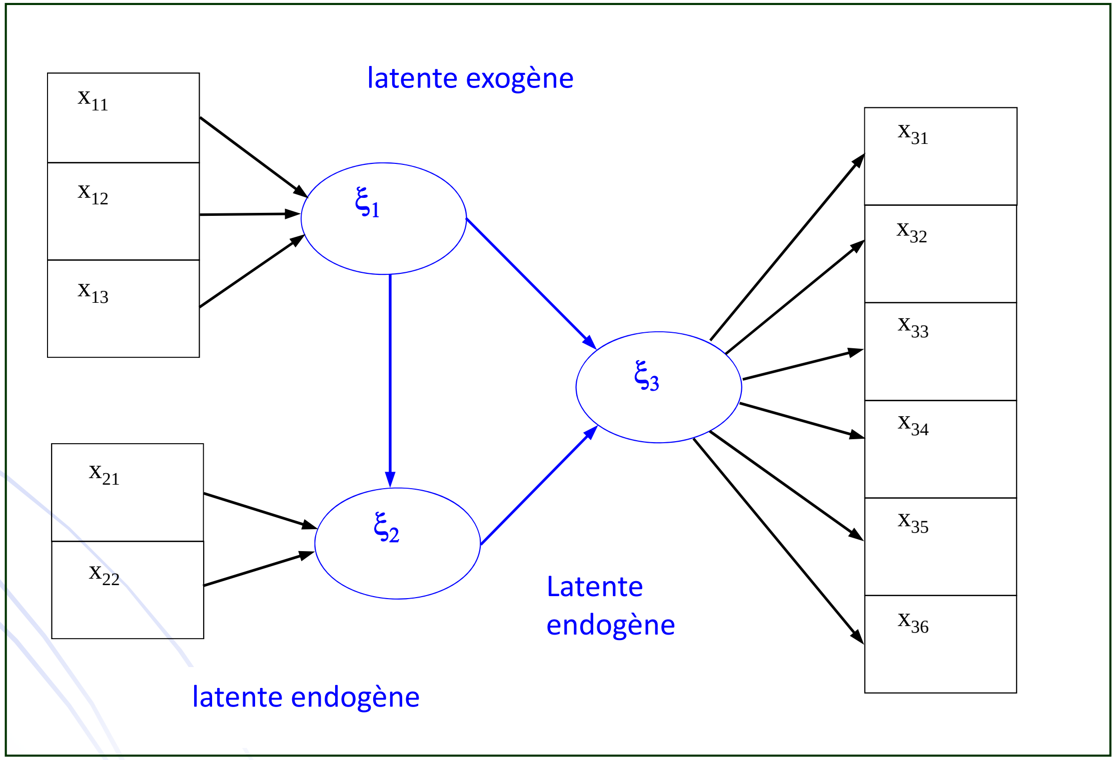
Indicateurs formatifs

p.ex. Satisfaction des clients

p.ex. Indicateurs socioculturels

Le modèle structurel

Le modèle « interne » ➔ liaisons entre variables latentes



L'estimation du modèle

- *L'estimation des paramètres de ce modèle peut se faire soit :*

  - *Par **l'approche LISREL** (linear structural relationships).*

  - *Par **l'approche PLS** (partial least squares path modeling)*

La méthode LISREL

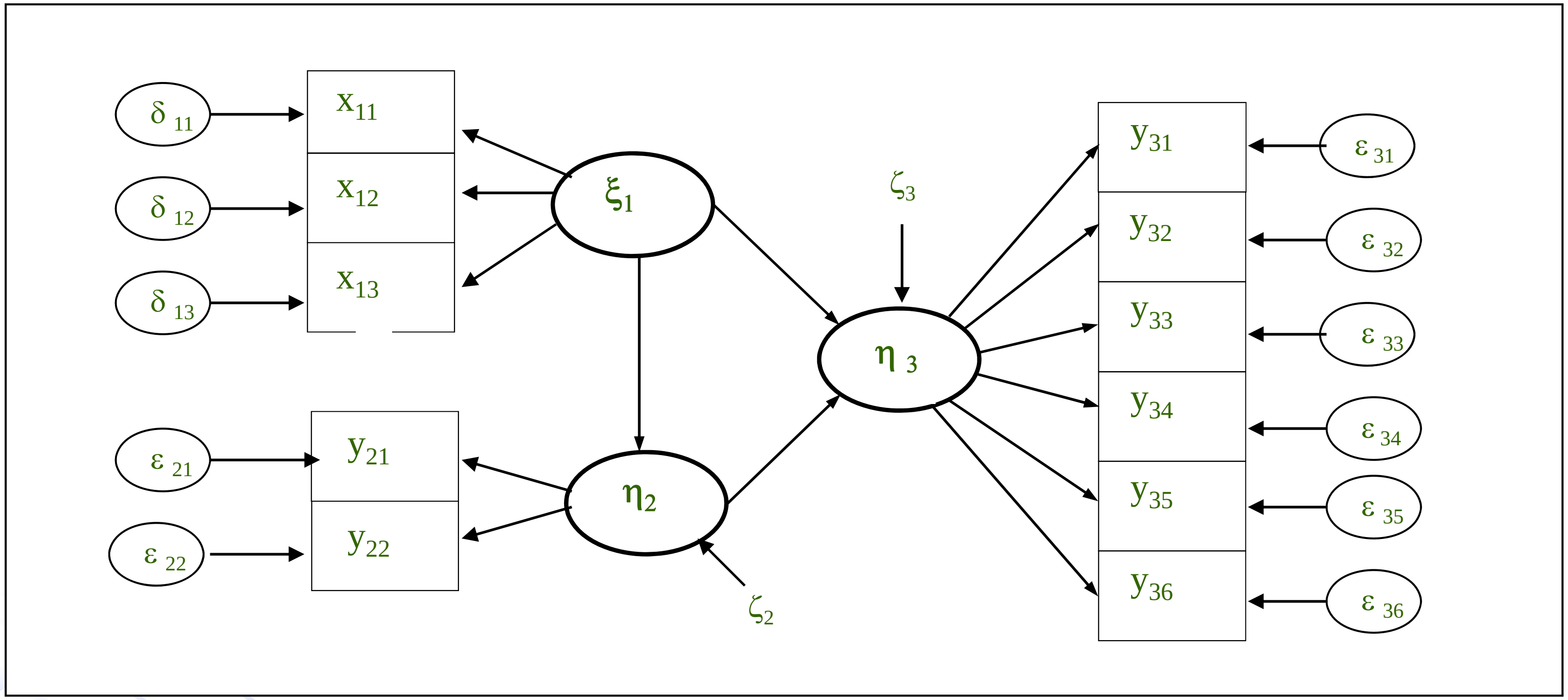
Analyse de la structure de covariance

Structural Equation Modeling (SEM)

Covariance Structure Analysis

...

Le modèle structurel de la covariance (LISREL)



$$\eta_{(m \times 1)} = B_{(m \times m)} * \eta_{(m \times 1)} + \Gamma_{(m \times n)} * \xi_{(n \times 1)} + \zeta_{(m \times 1)}$$

Eq. du modèle structurel

$$y_{(p \times 1)} = \Lambda_{y(p \times m)} * \eta_{(m \times 1)} + \varepsilon_{(p \times 1)}$$

$$x_{(q \times 1)} = \Lambda_{x(q \times n)} * \xi_{(n \times 1)} + \delta_{(q \times 1)}$$

Eq. du modèle de mesure

Caractéristiques

- Une approche statistique pour tester des hypothèses sur les relations entre variables observées et latentes (Hoyle, 1995)
- Le fondement statistique de la méthode LISREL est **la covariance**

Les pré-requis:

- Modèles statistiques linéaires
- Valide seulement sous certaines conditions:
  - Indépendance des observations (multiniveaux possible)
  - Normalité multivariée des données
  - Unidimensionnalité des blocs de variables

- La méthode LISREL est une méthode **a priori** et nécessite que les chercheurs pensent en termes de modèles et d'hypothèses

Les logiciels

Les principaux:

- LISREL (Jöreskog & Sörbom, 1996) : mis au point par le créateur de la méthode, très complet
- AMOS (Arbuckle, 1999, SPSS) : très complet et convivial
- CALIS (SAS): procédure intégrée à SAS

Le principe

- Une méthodologie générale pour <u>spécifier</u>, <u>estimer</u>, <u>comparer</u> et <u>évaluer</u> des modèles de relations entre variables.
- On va chercher à <u>confirmer</u> une théorie

- ***Procédure:***
  - Construction d'un modèle
  - Collecter les données pour tester le modèle
  - Le modèle est comparé aux données et évalué
  - Si nécessaire, le modèle est modifié et testé avec de nouvelles données

Estimation des paramètres du modèle - 1

Notations

- $p + q$ = Nombre de variables manifestes
- n = Nombre d'observations
- Σ = Matrice de covariance au niveau de la population
- S = Matrice des covariances observées
- C = Matrice des covariances obtenue grâce au modèle
- Φ = Matrice de covariance de ξ
- Ψ = Matrice de covariance de ζ

Estimation des paramètres du modèle - 2

$$\eta_{(m \times 1)} = B_{(m \times m)} * \eta_{(m \times 1)} + \Gamma_{(m \times n)} * \xi_{(n \times 1)} + \zeta_{(m \times 1)}$$

$$y_{(p \times 1)} = \Lambda_{y(p \times m)} * \eta_{(m \times 1)} + \varepsilon_{(p \times 1)}$$

$$x_{(q \times 1)} = \Lambda_{X(q \times n)} * \xi_{(n \times 1)} + \delta_{(q \times 1)}$$

On va tenter d'obtenir une matrice de covariance à partir de ce modèle. Cette matrice aura donc la forme:

$$C = \begin{pmatrix} \mathrm{cov}(x) & \mathrm{cov}(x,y) \\ \mathrm{cov}(y,x) & \mathrm{cov}(y) \end{pmatrix}$$

$$= \begin{pmatrix} (I-B)^{-1}(\Gamma\Phi\Gamma' + \Psi) & (I-B)^{-1}\Gamma\Phi \\ \Phi\Gamma'(I-B)^{-1} & \Phi \end{pmatrix}$$

Estimation des paramètres du modèle - 3

La méthode LISREL consiste en l'utilisation d'un estimateur afin de rendre la matrice de covariance calculée à partir du modèle (C) le plus proche possible de la matrice de covariance observée (S) en terme de maximum de vraisemblance.

On va donc estimer les paramètres du modèle (les éléments de la matrice C) de façon à minimiser cette différence.

Estimation des paramètres du modèle - 4

Utilisation du maximum de vraisemblance(MLE):

Si on suppose que les données sont **normales multivariées**,

l'estimateur ML revient à la minimisation de :

$$F_{ML} = \log\left|\mathbf{C}\right| + tr\left(\mathbf{S}\mathbf{C}^{-1}\right) - \log\left|\mathbf{S}\right| - \left(p+q\right)$$

Autres fonctions:

- Generalised Least Squares

$$F_{GLS} = \frac{1}{2} tr\left[\mathbf{S}^{-1}\left(\mathbf{S}\text{-}\mathbf{C}\right)^2\right]$$

- Unweighted Least Squares

$$F_{ULS} = \frac{1}{2} tr\left[\left(\mathbf{S}\text{-}\mathbf{C}\right)^2\right]$$

- Asymptotically Distribution Free

$$F_{ADF/WLS} = \left(\underline{\mathbf{s}} - \underline{\mathbf{c}}\right)^{\mathrm{T}} \mathbf{W}^{-1}\left(\underline{\mathbf{s}} - \underline{\mathbf{c}}\right)$$

Validation simple du modèle

Tests de validation globale du modèle:

- Si le modèle étudié est « exact » alors

$$(n-1)F = \chi^2(DF)$$

- Les degrés de liberté (DF) = nb de covariances – nb de paramètres

- Le modèle est accepté si $\dfrac{\chi^2}{DF} \leq 3$ ou p-valeur > 0,05
(seuils généralement utilisés)

Il existe d'autres indices de validation qui seront plus performants

Un indice plus « évolué »: RMSEA

<u>Le RMSEA (Root Mean Square Error of Approximation):</u> (Steiger et Lind)

- Cet indice calcule la différence entre la matrice de covariance obtenue et celle de la population globale :

$$\boxed{RMSEA = \sqrt{\frac{F_0}{DF}}}$$

Où $F_{ML} = \log|\mathbf{C}| + tr\left(\Sigma \mathbf{C}^{-1}\right) - \log|\Sigma| - (p+q)$

En pratique, on l'estime avec :

$$\boxed{RMSEA_{estimated} = \sqrt{\frac{F}{DF} - \frac{1}{n-1}}}$$

On l'accepte en dessous de 0.08 en général, un intervalle de confiance peut être obtenu.

Un exemple: L'engagement amoureux

Données:

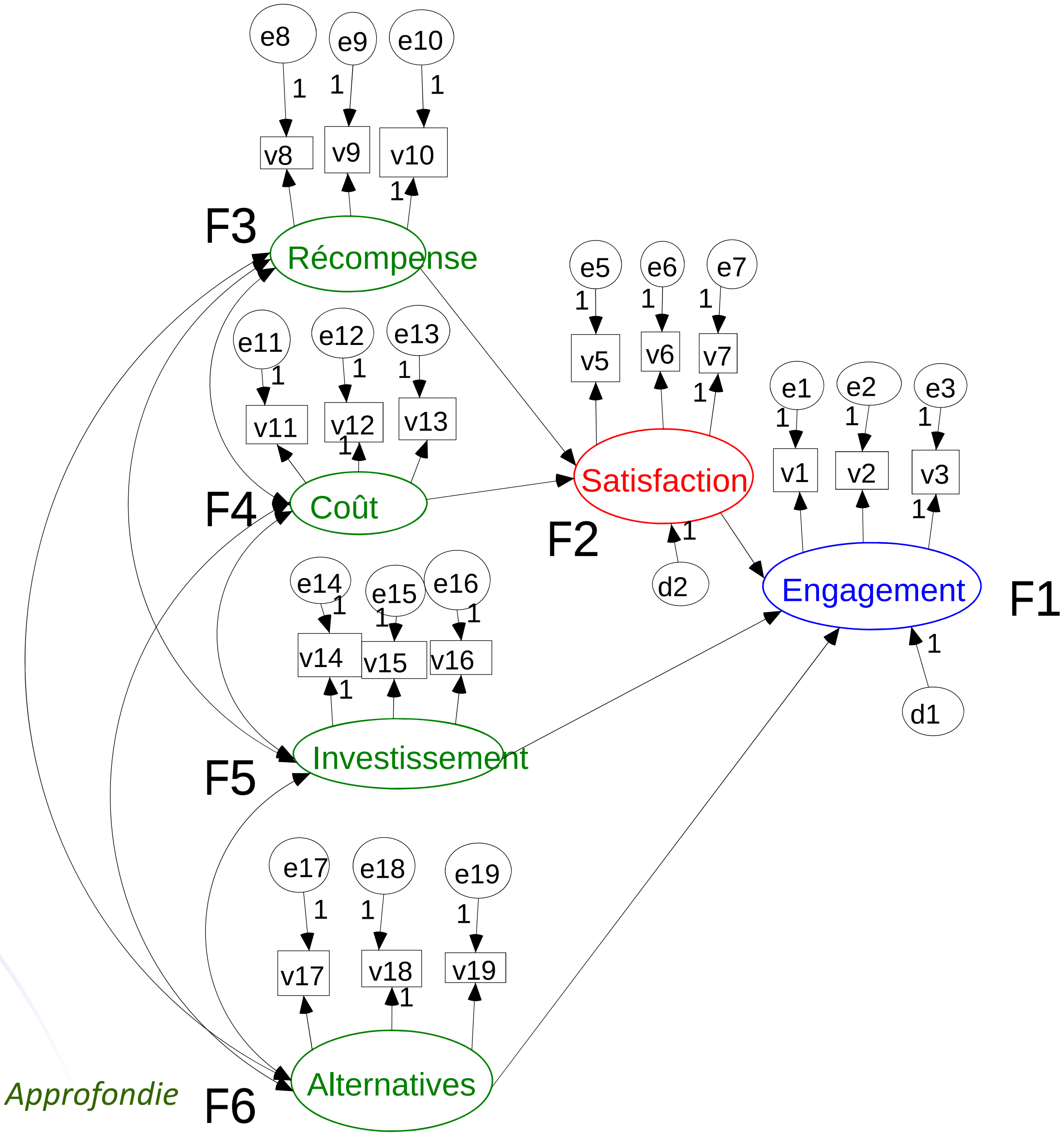
- C.E. Rusbult, « Commitment and satisfaction in romantic associations », *Journal of Experimental Social Psychology,* 1980

(Exemple tiré de la présentation «SEM assesment » par V. Esposito Vinzi)

6 blocs de variables manifestes:

- L'engagement
- La satisfaction
- Les récompenses
- Le coût
- La taille de l'investissement
- Les alternatives

Un exemple: Le modèle



Un exemple: Les blocs

Engagement intention de poursuivre une relation (F1)

Satisfaction réponse émotionnelle à une relation (F2)

Taille de l'investissement temps et effort nécessaire au maintien de la relation (F5)

Solutions alternatives (F6)

Récompenses quantité de bonnes choses associées à cette relation (F3)

Coût quantité de mauvaises choses associées à cette relation (F4)

Un exemple: Les questions

Donner une note de 1 à 7 en fonction de votre degré d’agrément avec l’affirmation

Satisfaction :

1. Je suis satisfait de ma relation

2. Ma relation actuelle est proche de la relation idéale

3. Je suis plus satisfait que la moyenne par ma relation actuelle

Taille de l’investissement :

1. J’ai investi beaucoup de temps dans ma relation actuelle
2. J’ai investi beaucoup d’énergie dans ma relation actuelle
3. J’ai investi beaucoup de ressources afin de développer ma relation actuelle

Une propriété importante: l'unidimensionnalité

- Dans le cadre des techniques de modèles structurels à variables latentes une propriété importante est souvent exigée: l'unidimensionnalité des blocs de variables manifestes.

- Si la première valeur propre obtenue par ACP est la seule >1 alors le bloc est unidimensionnel
- On peut utiliser l'alpha de Cronbach ou le rho de Dillon-Goldstein (on considère qu'un blocs est unidimensionnel si > 0,7)

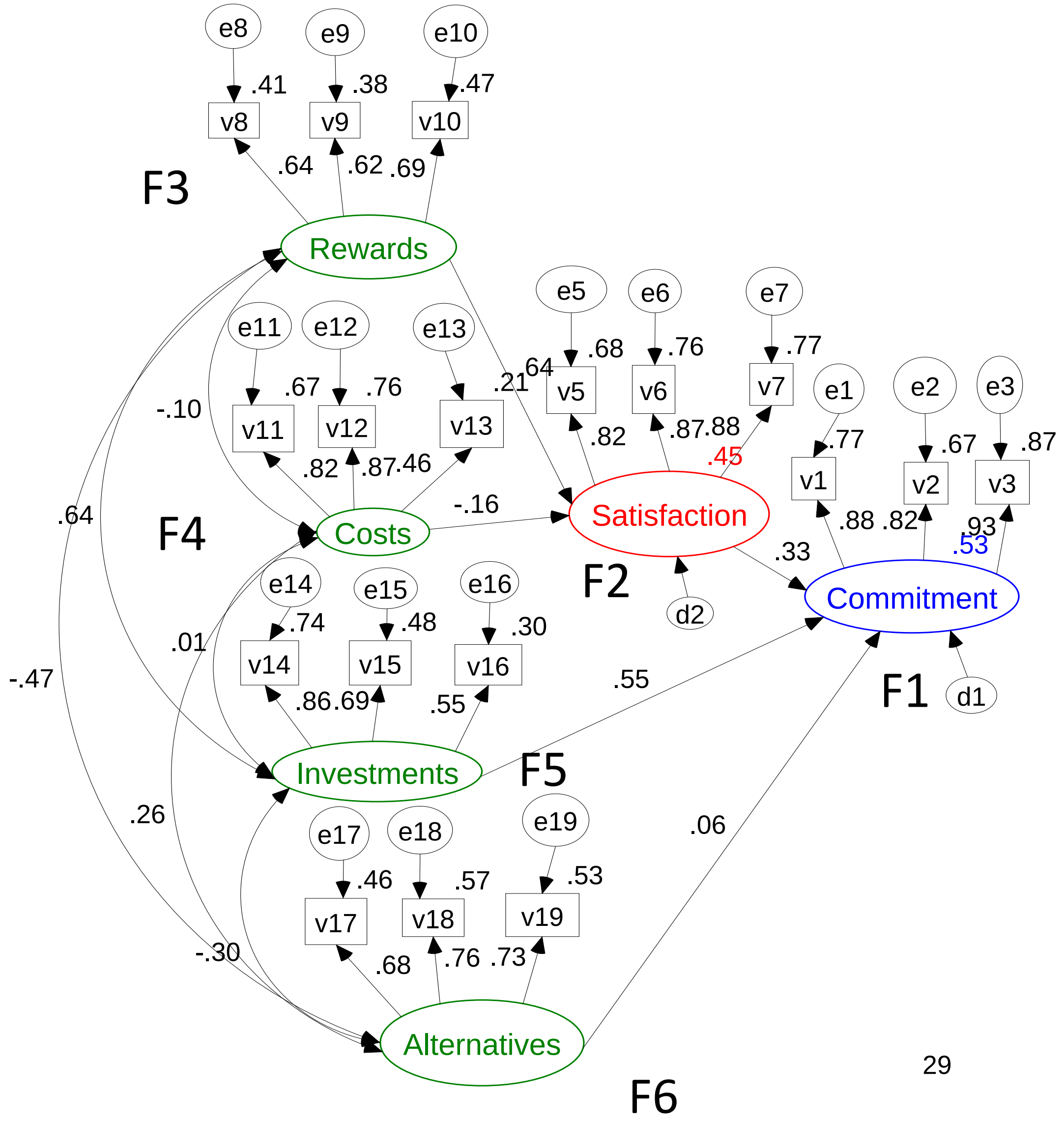
$$\boxed{\alpha_j = \frac{\frac{1}{p_j\left(p_j-1\right)}\sum_{i\neq k}\operatorname{cor}\left(\mathbf{x}_{ij},\mathbf{x}_{kj}\right)}{\frac{1}{p_j-1}+\frac{1}{p_j\left(p_j-1\right)}\sum_{i\neq k}\operatorname{cor}\left(\mathbf{x}_{ij},\mathbf{x}_{kj}\right)}\times\frac{p_j}{p_j-1}}$$

$$\boxed{\rho_j = \frac{\left(\sum_i \lambda_{ij}\right)^2\times\operatorname{var}\left(\boldsymbol{\xi}_j\right)}{\left(\sum_i \lambda_{ij}\right)^2\times\operatorname{var}\left(\boldsymbol{\xi}_j\right)+\sum_i \operatorname{var}\left(\boldsymbol{\varepsilon}_{ij}\right)}}$$

Un exemple: Estimation par maximum de vraisemblance

Chi-square = 216.75
DF = 124
Chi-square/DF = 1.748
RMSEA = .056

Les coefficients sur les arcs sont des corrélations



Un exemple: Les équations structurelles

Les coefficients des équations structurelles sont obtenus lors de l'estimation par maximum de vraisemblance

```
Latent Variable Equations with Estimates

 f1        =    0.4608*f2        +  0.7580*f5        +  0.1000*f6          +  1.0000 d1
 Std Err        0.0910 pf1f2        0.1037 pf1f5        0.1094 pf1f6
 t Value        5.0618              7.3127              0.9136


 f2        =    0.9737*f3        + -0.1213*f4        +  1.0000 d2
 Std Err        0.1321 pf2f3        0.0510 pf2f4
 t Value        7.3690             -2.3777
```

L'engagement ne dépend donc pas significativement des alternatives

Validation croisée du modèle

Le modèle obtenu s'adapte bien aux données, mais ceci ne prouve pas que ce modèle est « le meilleur », nous pouvons dire uniquement que :

Le modèle s'ajuste bien aux données traitées. Pour aller plus loin dans les conclusions, il faut valider le modèle en utilisant de la validation croisée

→L'indice de validation croisée (CVI):

Il mesure la distance entre la matrice de covariance estimée sur l'échantillon d'apprentissage et la matrice de covariance calculée sur les données de validation. Le modèle avec le CVI le plus petit est le plus stable.

Les indices de modification du modèle

Un indice de modification mesure la baisse du Chi² lorsqu'un lien est ajouté au modèle (Univariate Lagrange Multiplier Test)

```
Rank Order of the 5 Largest Modification Indices
```

Row	Column	Chi-Square	Pr > ChiSq
f2	f5	34.34669	<.0001
v2	f5	7.97159	0.0048
v1	f3	7.65396	0.0057
v10	f5	5.64619	0.0175
v18	f3	4.69157	0.0303

La satisfaction dépend aussi de l'investissement

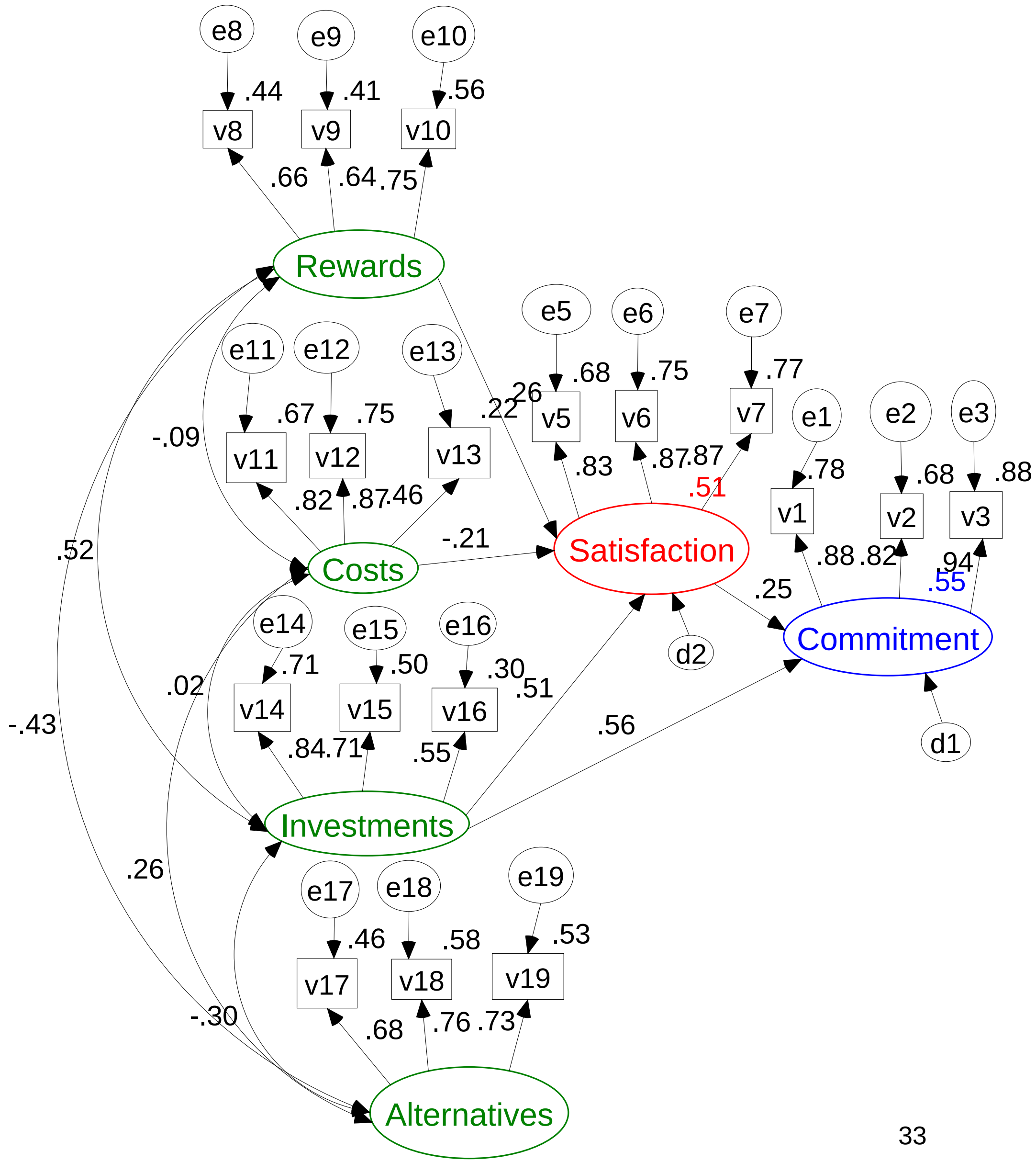
Il existe aussi des indices (Univariate Wald Test) qui estime l'augmentation du chi² quand on retire un lien au modèle

Un exemple: les indices de modification

Chi-square = 183.191
DF = 124
Chi-square/DF = 1.472
RMSEA = .045

Le modèle a été modifié avec la suppression de l'arc allant des alternatives à l'engagement et l'ajout d'un arc entre investissement et satisfaction

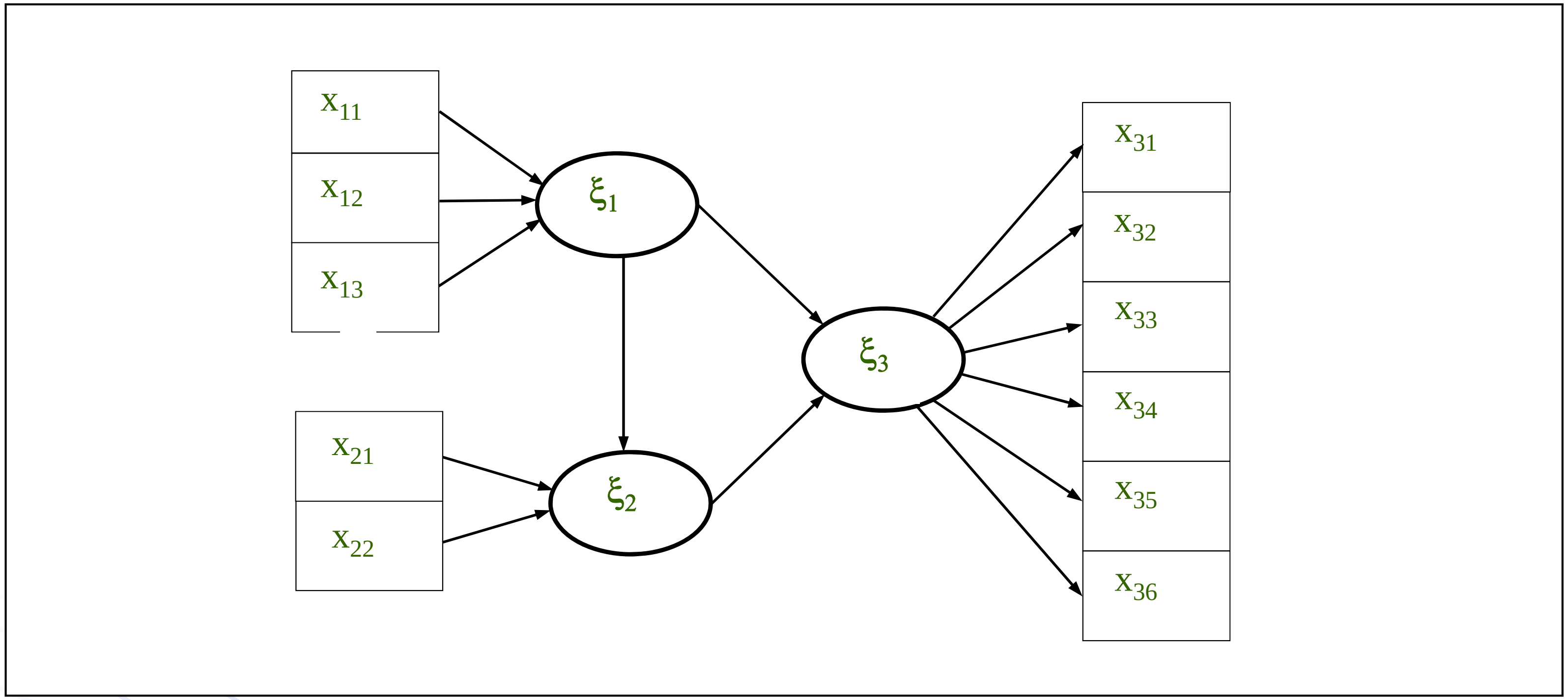
⇨ Les indices de qualité d'ajustement sont améliorés



L'approche PLS

Partial Least Squares Path Modeling

Le « path diagram » (PLS)



$$\mathbf{x}_{jh} = \pi_{jh}\boldsymbol{\xi}_j + \boldsymbol{\varepsilon}_{jh}$$

$$\boldsymbol{\xi}_j = \sum_i \beta_{ji}\boldsymbol{\xi}_i + \mathbf{v}_j$$

Caractéristiques

- Le fondement statistique de l’approche PLS est **la variance**

Les pré-requis:

- Modèles de régressions simples et multiples
- Valide seulement sous certaines conditions:
  - Indépendance des observations (multiniveaux possible)
  - Unidimensionnalité des blocs (dans le cas réflectif)

- L’approche PLS est une méthode **a priori** et nécessite que les chercheurs pensent en termes de modèles et d’hypothèses. Cependant, un aspect prédictif existe aussi.

Les logiciels

Les principaux:

- XLSTAT (PLS-PM, 2009) : Logiciel complet et convivial adapté à Excel (www.xlstat.com)
- LVPLS (Löhmoller, 1989): Premier logiciel, très ancien
- PLS-Graph (W. Chin, 1996): Le plus connu dans le milieu académique

Approche PLS – Le principe

Principe :

1. Méthode basée sur des régressions simples et multiples
2. L'estimation du modèle passe par l'estimation des scores des variables latentes
3. Cette estimation se fait à l'aide d'un algorithme itératif
4. Une fois les scores obtenus, on estime les coefficients du modèle interne par régressions multiples classiques (OLS)
5. Les « loadings » peuvent être retrouvés avec les scores des variables latentes.

Cette approche converge dans la pratique mais cette convergence n'est pas prouvée au-delà de 2 blocs.

Algorithme PLS

Poids externes initiaux w_{jh}

Estimation des paramètres par régression

conv.

$$\mathbf{y}_j = \sum_{h=1}^{p_j} w_{jh}\mathbf{x}_{jh}$$

Estimation externe de ξ_j

Calcul de poids externes w_{ij}

Calcul des poids internes e_{ji}

$$\mathbf{z}_j = \sum_{\xi_i \to \xi_j} e_{ji}\mathbf{y}_i$$

Estimation interne de ξ_j

Approche PLS – Le principe

Principe de l'approche PLS (Wold, 1982) Algorithme itératif avec estimation alternée des variables latentes en fonction de chaque sous-modèle.

1. On fixe les poids externes initiaux w_0
2. On calcule les scores des variables latentes en se basant sur le modèle externe (chaque score associé à une variable latente est calculé en fonction des variables manifestes de son bloc)
$$\mathbf{y}_j = \sum_{h=1}^{p_j} w_{jh} \mathbf{x}_{jh}$$
3. On calcule les scores des variables latentes en se basant sur le modèle interne (chaque score associé à une variable latente est calculé en fonction des autres variables latentes qui lui sont liées)
$$\mathbf{z}_j = \sum_{\xi_i \leftrightarrow \xi_j} e_{ji} \mathbf{y}_i$$

Répéter les points 2 et 3 jusqu'à convergence

Estimation des équations structurelles par régressions multiples (OLS)

Approche PLS – Les poids externes

Initialisation: En général, les poids externes sont fixés à 1 pour toutes les variables manifestes exceptée la dernière de chaque bloc qui est fixé à –1.

Modes d'estimation:

Mode A (Cas réflectif): $$w_{jh} = \operatorname{cov}\left(x_{jh}, Z_j\right)$$

→ Régressions simples

Mode B (Cas formatif): $$w_j = (X_j' X_j)^{-1} X_j' Z_j$$

→ Régressions multiples (OLS)

Approche PLS – Les poids internes

Schémas d'estimation:

Centroïde: $e_{ji} = \text{sign}[\text{cor}(\mathbf{y}_j,\mathbf{y}_i)]$

→ Problèmes avec les corrélations proches de 0

Factoriel: $e_{ji} = \text{cor}(\mathbf{y}_j,\mathbf{y}_i)$

Structurel: e_{ji} = coef. de régression multiple de y_j sur y_i si elles sont reliées

Validation du modèle

La validation se fait en terme de qualité prédictive et non en terme de qualité d'ajustement du modèle aux données.

Les indices de validation sont basés sur la qualité de prédiction des variables manifestes en se basant sur le modèle et sur les scores des variables latentes.

Le R^2 est couramment utilisé, d'autres indices spécifiques existent.

Validation du modèle

Validation (Tenenhaus et al., 2005): **<u>Communauté et redondance</u>**

Communauté ➔ évalue la qualité du modèle externe :

$$communality_j = \frac{1}{p_j} \sum_{h=1}^{p_j} cor^2 \left(x_{jh}, Y_j \right)$$

Redondance ➔ évalue la qualité du modèle interne :

$$redundancy_j = communality_j \times R^2 \left(Y_j, \left\{ Y'_{j'} \text{ explique } Y_j \right\} \right)$$

Critère global utilisé pour choisir le meilleur modèle (Amato § al., 2004) :

$$G \circ F = \sqrt{\overline{communality} \times \overline{R^2}}$$

Autre solutions: la validation croisée (cv-communauté, cv-redondance, Q² de Stone-Geisser)

Remarques

- On obtient donc des scores pour les variables latentes au niveau de chaque individu
- L'utilisation de régressions n'entraîne pas d'hypothèses de normalité
- La convergence de cet algorithme n'est pas prouvée pour plus de deux blocs mais elle est constatée dans la pratique

Cette méthode est plus prédictive que confirmatoire (à l'inverse de la méthode LISREL)

Un exemple : L'instabilité politique

Données: C.E. Russett, *GIFI,* 1964 (Exemple tiré de « PLS Path Modeling » par M. Tenenhaus)

Inégalité agricole

GINI : Inégalité dans la répartition des terres

FARM : % fermiers possédant la moitié des terres (> 50%)

RENT : % fermiers locataires

Développement industriel

GNPR : PNB par habitant ($ 1955)

LABO : % d 'actifs dans l'agriculture

Un exemple : L'instabilité politique

<u>Instabilité politique</u>

INST : Instabilité de l 'exécutif (45-61)

ECKS : Nb de conflits violents entre communautés (46-61)

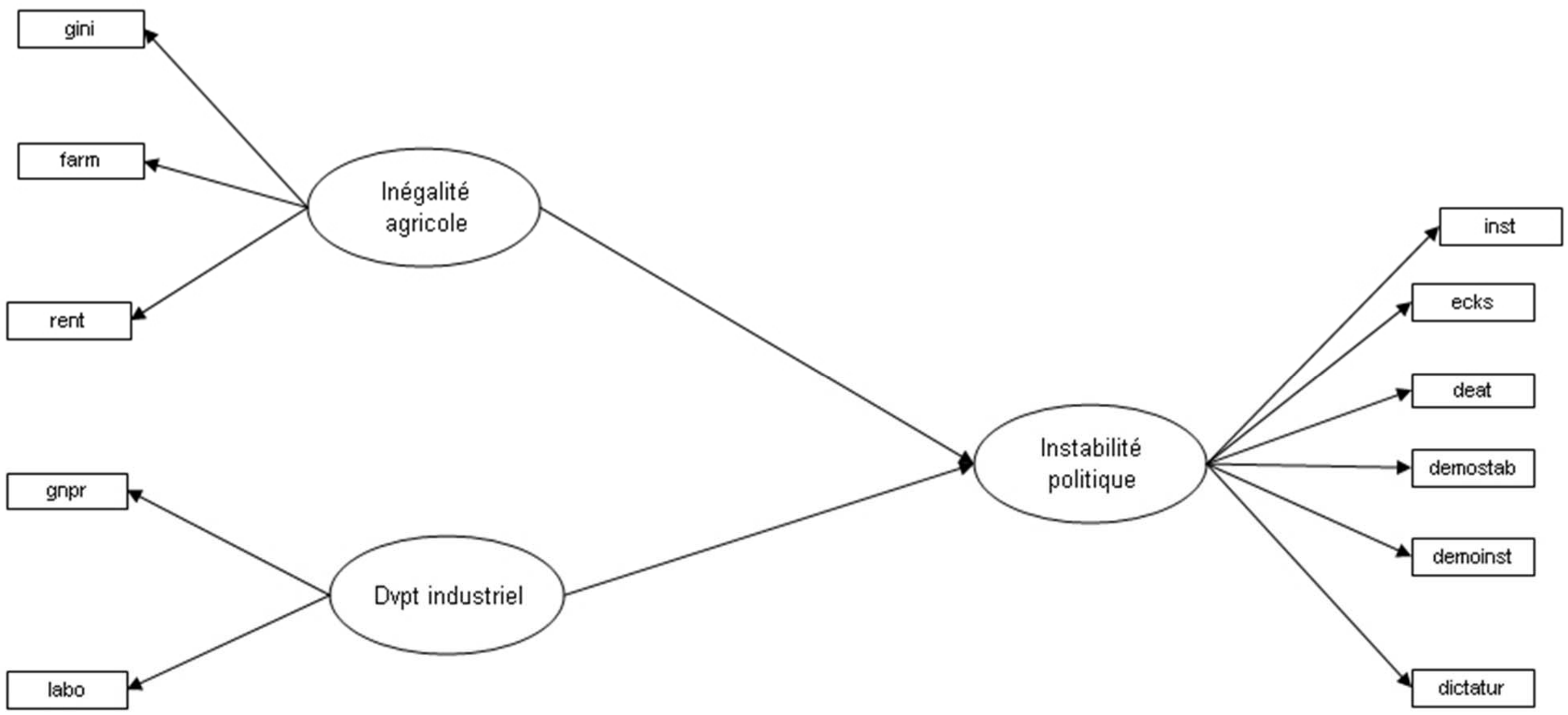
DEAT : Nb de morts dans des manifestations (50-62)

D-STAB : Démocratie stable

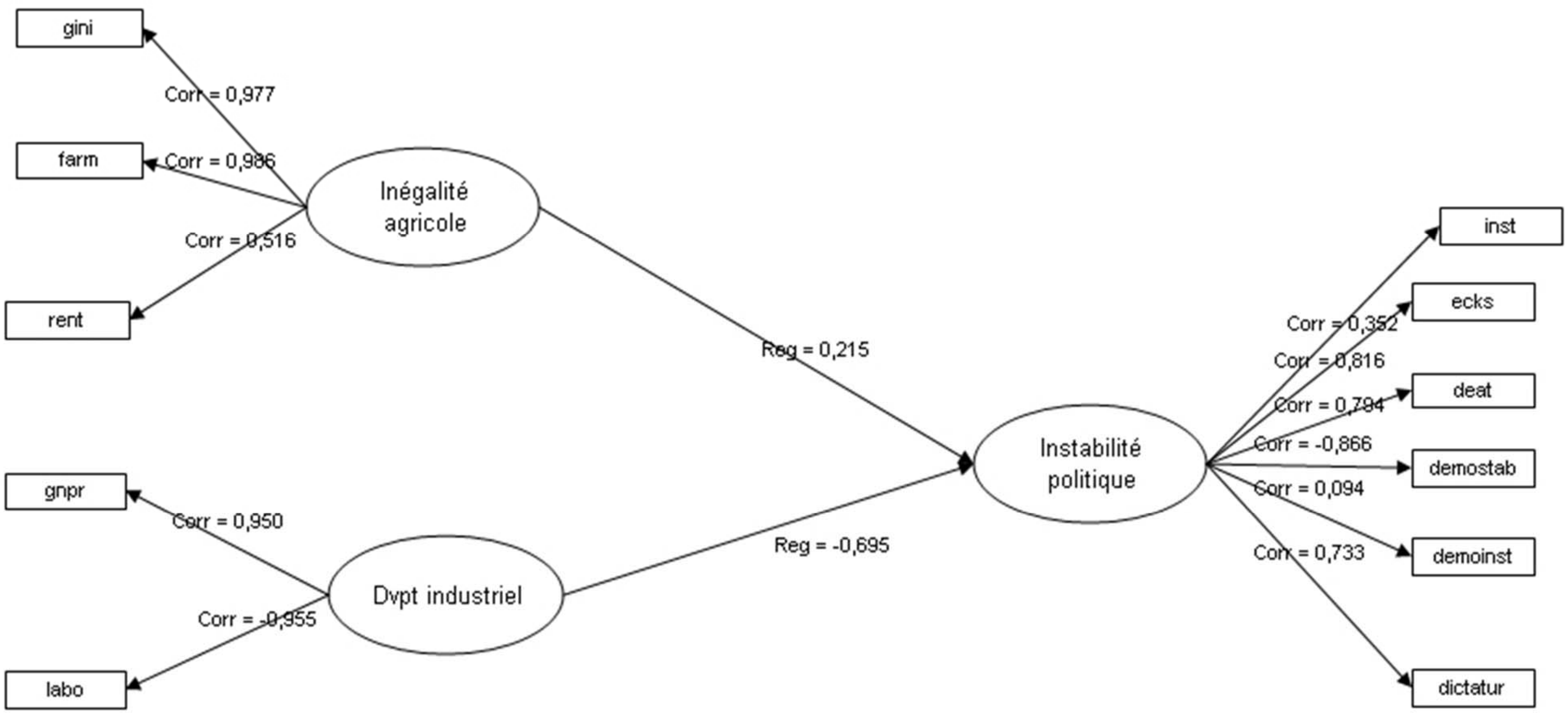
D-INS : Démocratie instable

DICT : Dictature

Un exemple : Le modèle



Un exemple : Estimation avec PLS



Un exemple : Les estimations

(1) Estimation externe de Y_i :

$$Y_1 = w_{11}\text{Gini} + w_{12}\text{Farm} + w_{13}\text{Rent}$$

$$Y_2 = X_2 w_2$$

$$Y_3 = X_3 w_3$$

(2) Estimation interne de Z_i :

$$Z_1 = \text{sign}(\text{cor}(\xi_1, \xi_3)Y_3 = (+1)Y_3$$

$$Z_2 = \text{sign}(\text{cor}(\xi_2, \xi_3)Y_3 = (-1)Y_3$$

$$Z_3 = \text{sign}(\text{cor}(\xi_3, \xi_1)Y_1 + \text{sign}(\text{cor}(\xi_3, \xi_2)Y_2 = (+1)Y_1 + (-1)Y_2$$

Un exemple : Le modèle externe

Variable latente	Variables manifestes	Poids externe	Corrélations
Inégalité agricole	gini	0,032	0,977
	farm	0,077	0,986
	rent	0,085	0,516
Dvpt industriel	gnpr	0,573	0,950
	labo	-0,766	-0,955
Instabilité politique	inst	0,424	0,352
	ecks	0,198	0,816
	deat	0,130	0,794
	demostab	-0,714	-0,866
	demoinst	0,084	0,094
	dictatur	0,569	0,733

Un exemple : Le modèle interne

R^2 (Instabilité politique / 1) :

R^2	R^2(Bootstrap)	Ecart-type	Ratio critique (CR)
0,622	0,657	0,076	8,167

Path coefficients (Instabilité politique / 1) :

Variable latente	Valeur	Ecart-type	t	Pr > |t|
Inégalité agricole	0,215	0,097	2,206	0,033
Dvpt industriel	-0,695	0,097	-7,128	0,000

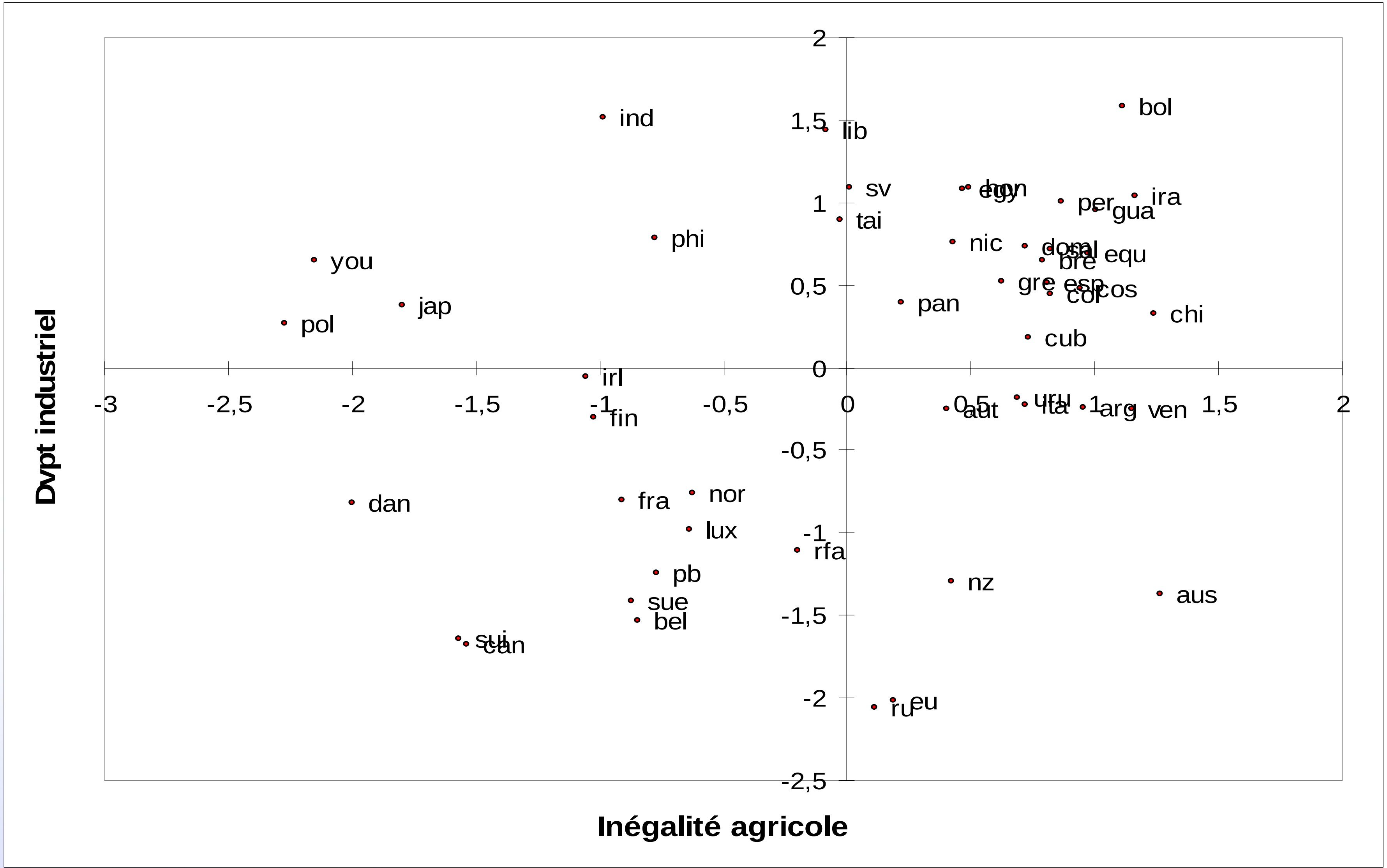
Equation du modèle :

Instabilité politique = 0,215*Inégalité agricole-0,695*Dvpt industriel

Un exemple : Les variables latentes

	Inégalité agricole	Dvpt industriel	Instabilité politique
arg	0,953	-0,238	-0,751
aus	1,265	-1,371	1,601
aut	0,404	-0,253	0,464
bel	-0,848	-1,530	0,881
bol	1,115	1,584	-1,503
bre	0,789	0,654	-0,268
can	-1,539	-1,680	0,972
chi	1,239	0,324	-0,016
..	…	…	…
sv	0,013	1,094	-1,386
esp	0,811	0,516	-0,411
sue	-0,870	-1,410	1,605
sui	-1,568	-1,640	1,605
tai	-0,030	0,898	0,030
ru	0,112	-2,059	1,063
eu	0,187	-2,016	0,964
uru	0,685	-0,179	1,299
ven	1,149	-0,252	-1,142
rfa	-0,199	-1,104	0,477
you	-2,153	0,654	-0,152

Un exemple : Répartition des pays



Un exemple : Les indices globaux

	GoF	GoF (Bootstrap)	Ecart-type
Absolu	0,610	0,624	0,043

Variable latente	Type	Moyenne	R²	R² ajusté	Moyenne Communalités (AVE)	Moyenne Redondances
Inégalité agricole	Exogène	0,000			0,731	
Dvpt industriel	Exogène	0,000			0,907	
Instabilité politique	Endogène	0,000	0,622	0,614	0,452	0,282
Moyenne			0,622		0,611	0,282

Variable latente	ariables manifeste	Corrélations	Communalités	Redondances
Inégalité agricole	gini	0,977	0,955	
	farm	0,986	0,972	
	rent	0,516	0,266	
Dvpt industriel	gnpr	0,950	0,903	
	labo	-0,955	0,912	
Instabilité politique	inst	0,352	0,124	0,077
	ecks	0,816	0,665	0,414
	deat	0,794	0,630	0,392
	demostab	-0,866	0,749	0,466
	demoinst	0,094	0,009	0,006
	dictatur	0,733	0,537	0,334

$$GoF = \sqrt{\frac{1}{3}\sum Communaut\acute{e}_i \times R^2(Inst.Pol.)} = \sqrt{0.611 \times 0.622} = 0.610$$

Comparaison des approches LISREL et PLS

Différences théoriques - 1

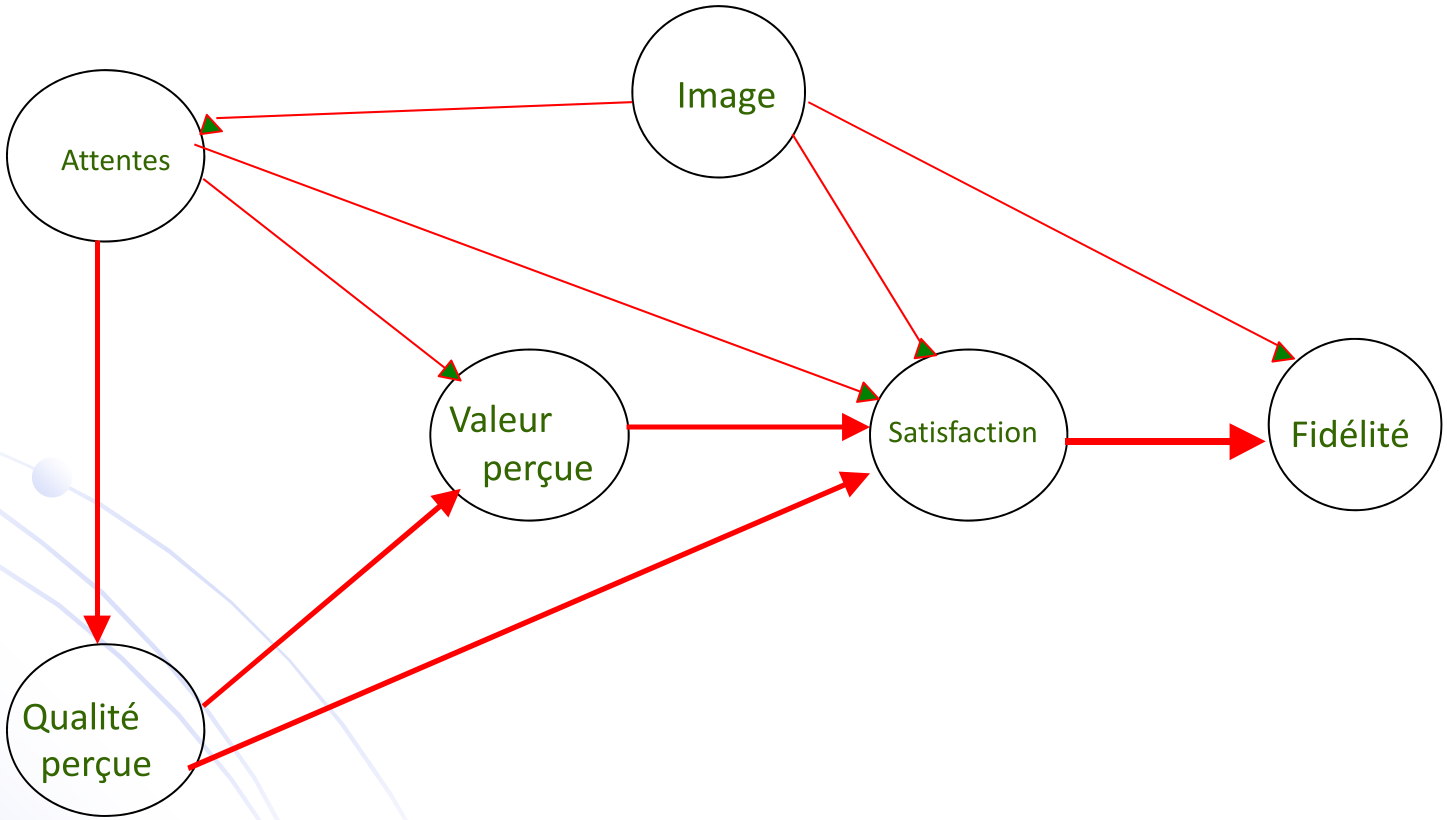
Critère	Approche PLS	LISREL (Linear Structural Relationship)
Objectif	Orientée vers la réalisation des prévisions	Orientée vers l'estimation des paramètres
Méthodologie	Basée sur variance	Basée sur covariance
Variables latentes (VL)	Combinaison linéaire de ses VM	Combinaison linéaire de toutes les VM
Relations entre les VL et les VM	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 estimée dans un espace non restreint

Différences théoriques - 2

Critère	**Approche PLS**	**LISREL (Linear Structural Relationship)**
Hypothèses	Unidimensionnalité (réflectif)	Multinormalité des données + unidimensionnalité
Complexité modèle	Grande (ex: 100 VL, 1000 VM)	Réduite ou modéré (<100 VM)
Taille échantillon	30-100 cas	200-800 cas
Données manquantes	« NIPALS »	Prétraitement
Identification	Dans le cadre du modèle récursif, toujours identifiée	Dépend du modèle: au moins 3 VM par VL pour être correctement identifiée
Consistance	Consistance « au sens large »	Consistance des estimations
Domaines	Marketing, analyse sensorielle…	Sociologie, psychologie…

Un exemple: Un questionnaire de satisfaction

Nous devrons utiliser un modèle simplifié car LISREL ne converge pas sur un modèle complexe (modèle basé sur l'ECSI, European Customer Satisfaction Index)



Application – Modèle interne

Modèle interne

La satisfaction est un processus complexe auquel participent toutes les variables latentes.

Pour PLS (XLSTAT) :

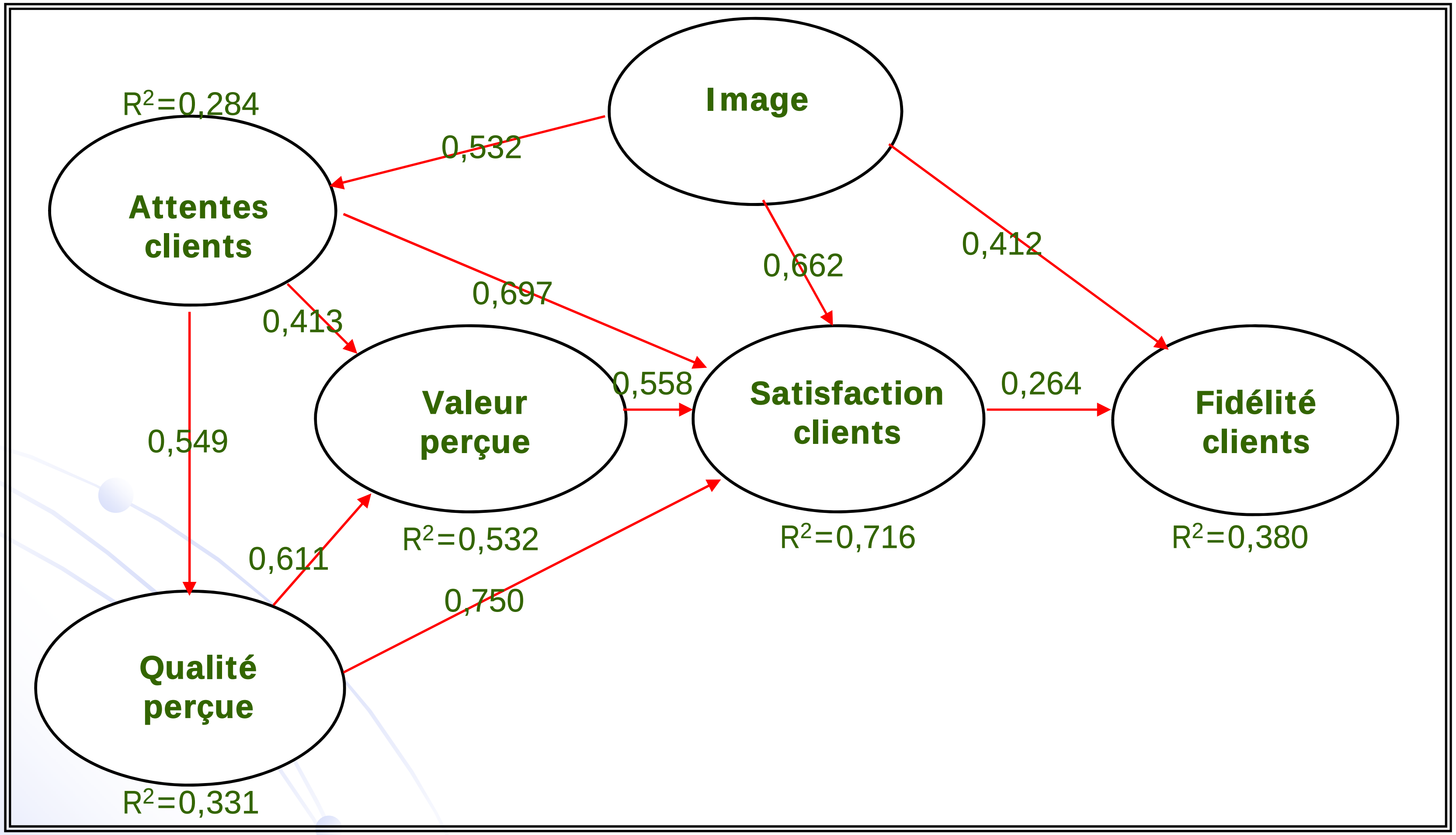
$$Satg = 0{,}170 \times Image + 0{,}290 \times Attente + 0{,}470 \times VP + 0{,}019 \times QP$$

Pour LISREL (AMOS) :

$$Satg = 0{,}228 \times Image + 0{,}329 \times Attente + 0{,}408 \times VP + 0{,}005 \times QP$$

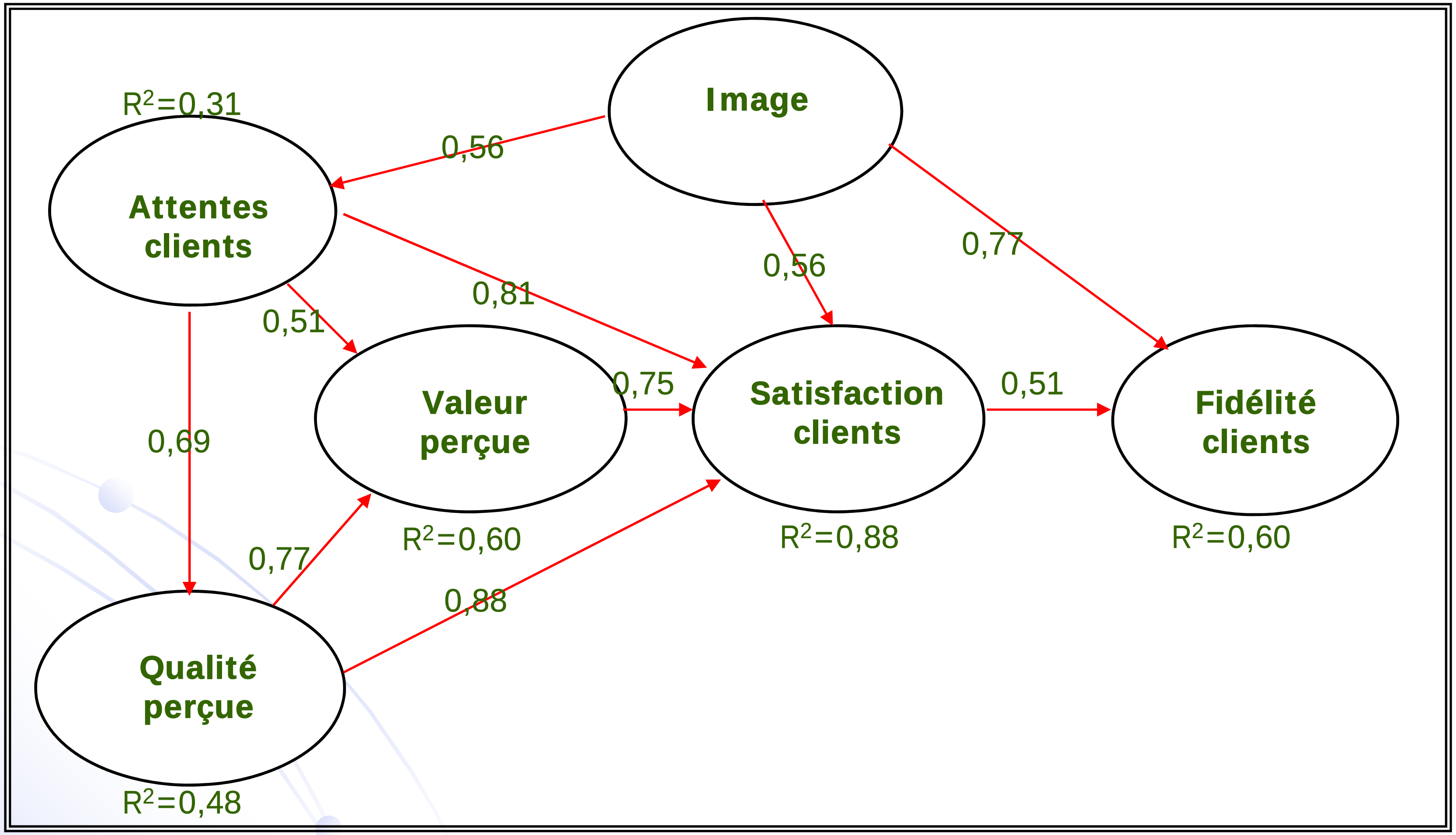
Application - Modèle interne pour l'approche PLS

Schéma de causalité (R^2 et corrélations)



Application - Modèle interne pour la méthode LISREL

Schéma de causalité (R^2 et corrélations)



Application – Le modèle externe

La variable latente satisfaction

Approche PLS:

$$Sat = 0,116 \times VM1 + 0,106 \times VM2 + 0,109 \times VM3 + ... + 0,086 \times VM12$$

où *VMi* représente la variable manifeste *i* associée à la variable latente satisfaction

Méthode LISREL:

Pas d'équations de ce type

Conclusions

- Les deux modèles obtenus sont assez différents
- PLS sous-estime les paramètres du modèle interne
- Les indices de validation ne sont pas comparables:
  - Pour la méthode LISREL, ils sont basés sur le chi^2
  - Pour l'approche PLS, ils sont basés sur les estimations des variables latentes
- Pour une application réelle et complexe, l'approche PLS est plus adaptée, de plus, dans le milieu opérationnel, on cherche plus des prévisions que des estimations précises des paramètres.
- La méthode LISREL est intéressante du point de vue de sa validité théorique et de l'optimalité des estimations des paramètres. Elle s'applique dans le cadre de la validation d'une théorie.

Bibliographie

• Hoyle R., 1995, *Structural equation modeling: concepts, issues and applications*, SAGE Publications

• Jöreskog K.& Sörbom D., 1989, *LISREL7 - A guide to the program and applications*, Second Edition, SPSS Publications.

• Lohmöller J.B., 1987, *LVPLS 1.8 Program Manual*, Universitaet zu Koehn, Zentralarchiv fuer Empirische Sozialforschung, Köln.

• Roussel P., Durrieu F., Campoy E., El Akremi A., 2002, *Méthodes d'Équations Structurelles: Recherche et Applications en Gestion*, Economica, Paris, 2002

• Tenenhaus M., 1998, *La Régression PLS*, Editions Technip, Paris.

• Tenenhaus M., 1999, « L'approche PLS », Revue de Statistique Appliquée, 47 (2), 5-40.

• Tenenhaus M., Esposito Vinzi V., Chatelin Y.-M., and Lauro C. (2005). "PLS Path Modeling", *Computational Statistics & Data Analysis*, 48:159-205.

• Wold H. (1982), "Soft Modelling: The basic design and some extensions" *in* Jöreskog K. and Wold H. (Eds.), *System under indirect observation, vol.2,* North-Holland, Amsterdam, 1-54.