

# *Les modèles d'équations structurelles à variables latentes*

Emmanuel Jakobowicz

CNAM – CEDRIC

EDF – R&D

emmanuel.jakobowicz(@)free.fr

*6 décembre 2006*

# *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)**

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*

---

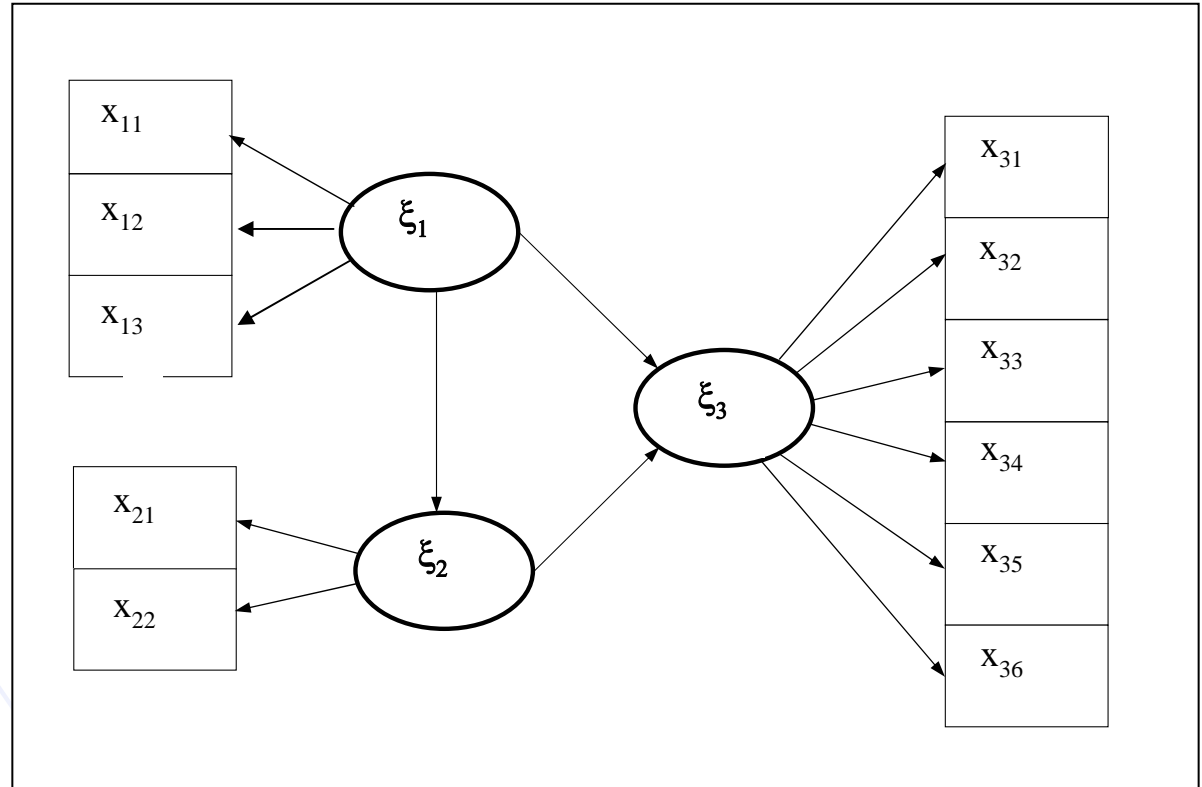
- Méthodes appartenant à la catégorie des modèles structurels à variables latentes (VL)
- $p$  variables observées sur  $n$  individus, réparties en  $J$  blocs de  $k_j$  variables
- L'ensemble des variables sont continues
- Blocs reliés entre eux par un modèle de relations structurelles entre variables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



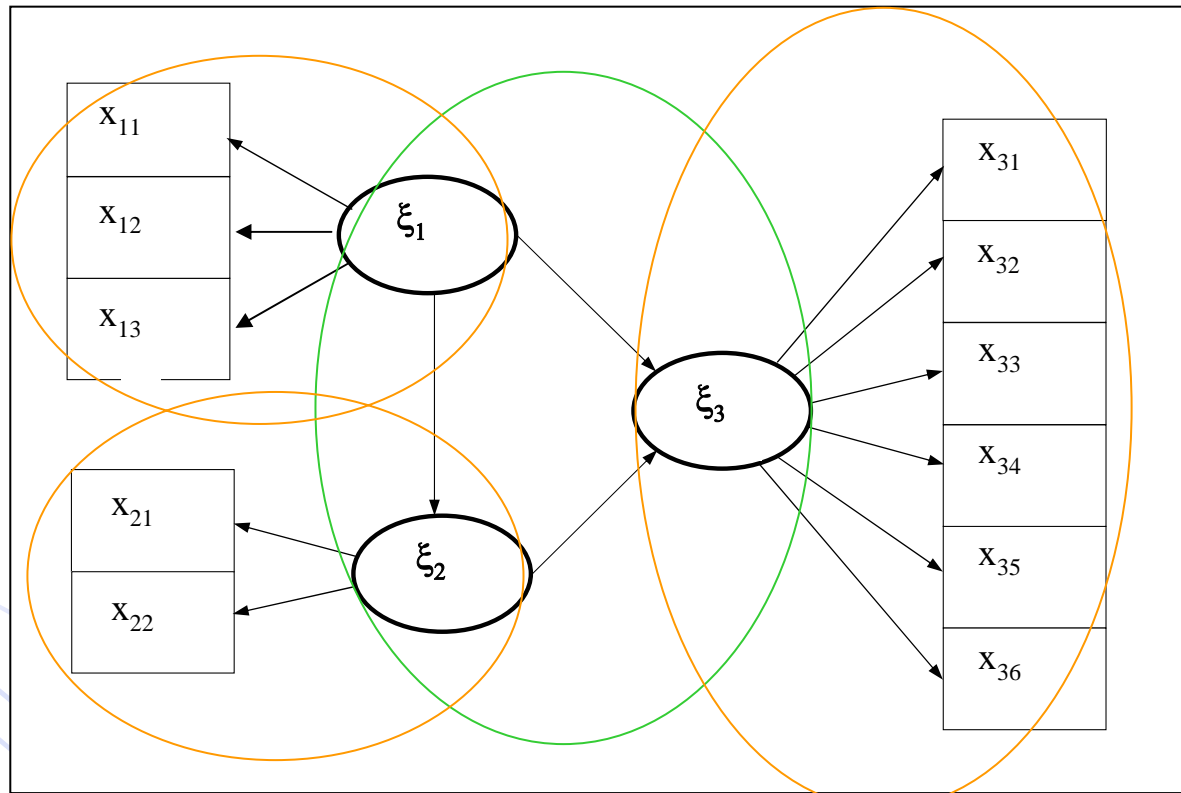
## *Les concepts de base - 3*

---

### **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

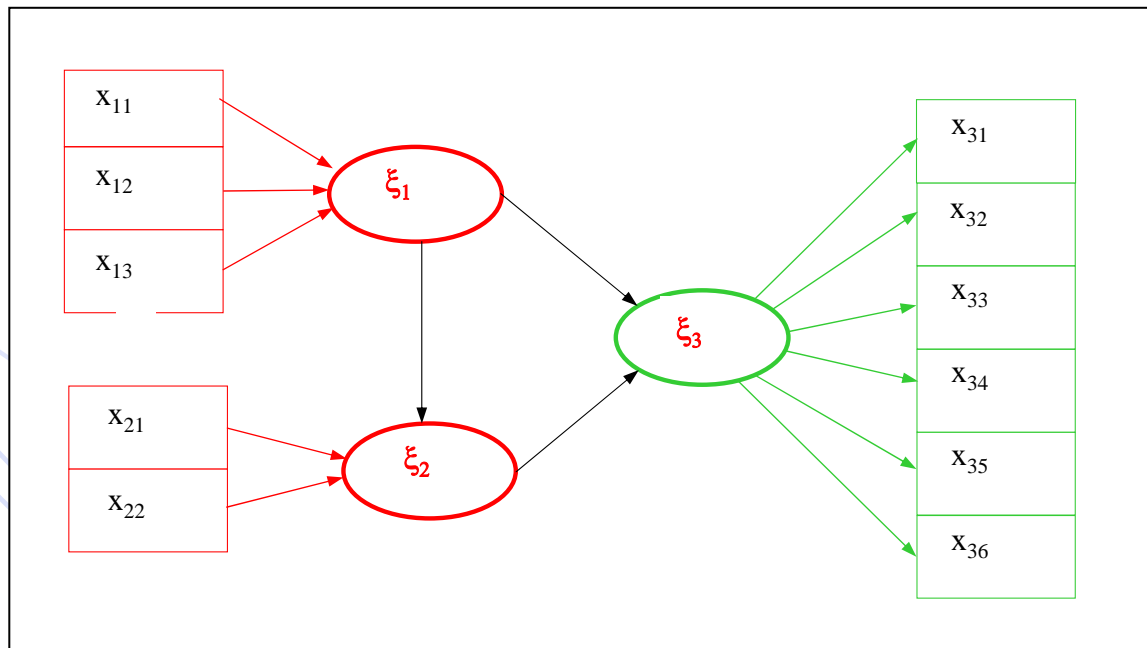
## Le modèle « externe »

### Type réflectif

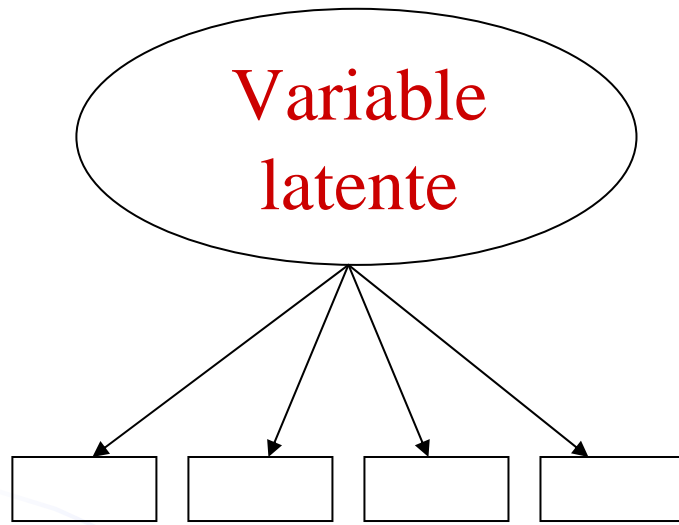
Variables manifestes → reflet de leur variable latente

### Type formatif

Variable latente  $\xi_j$  → reflet des variables manifestes du bloc  $X_j$

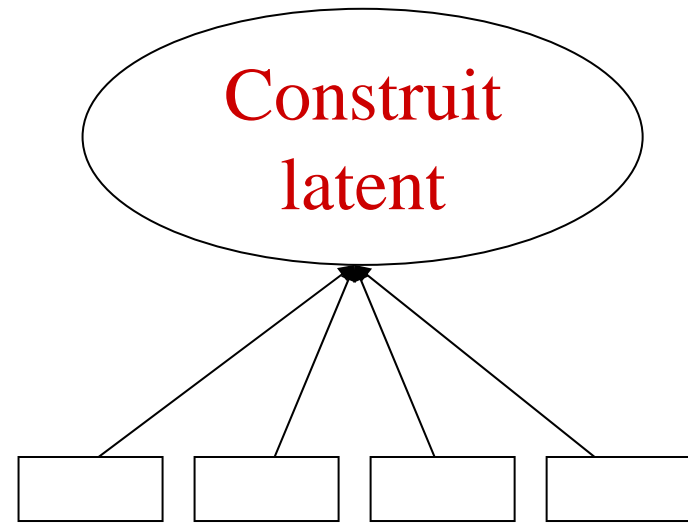


## Le modèle de mesure - 2



Indicateurs réfléchitifs

p.ex. Satisfaction des clients

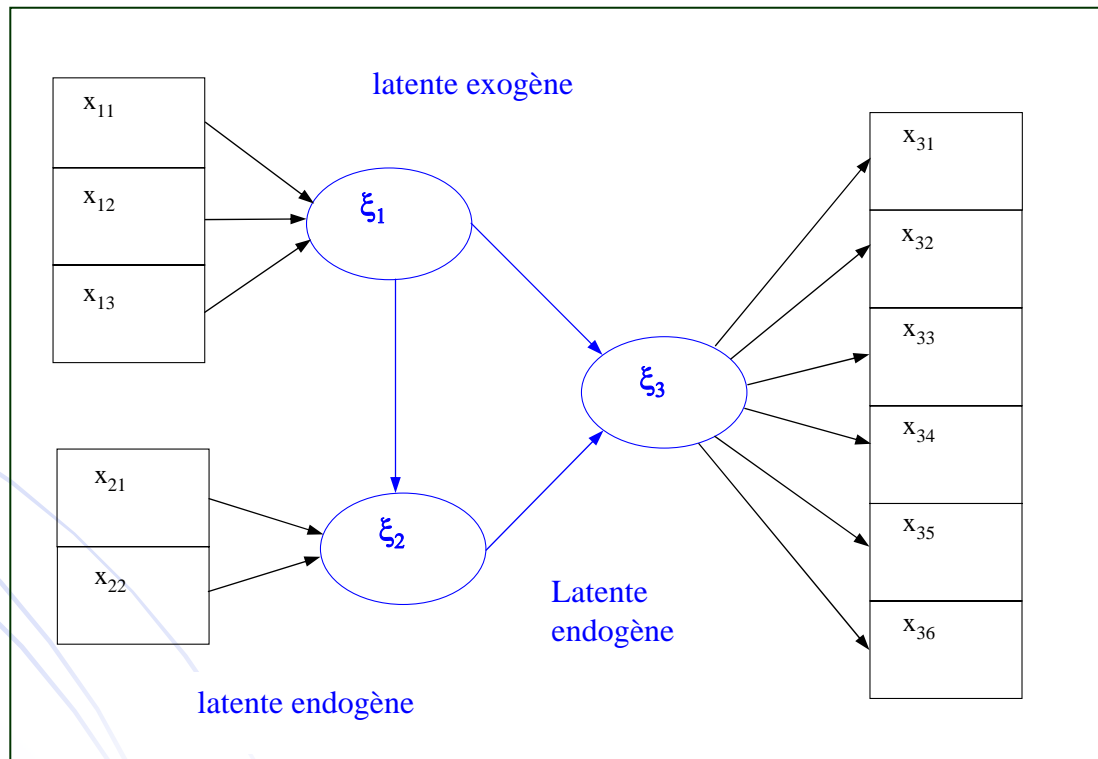


Indicateurs formatifs

p.ex. Indicateurs sociaux-culturels

# Le modèle structurel

Le modèle « interne » → liaisons entre variables latentes



# *L'estimation du modèle*

---

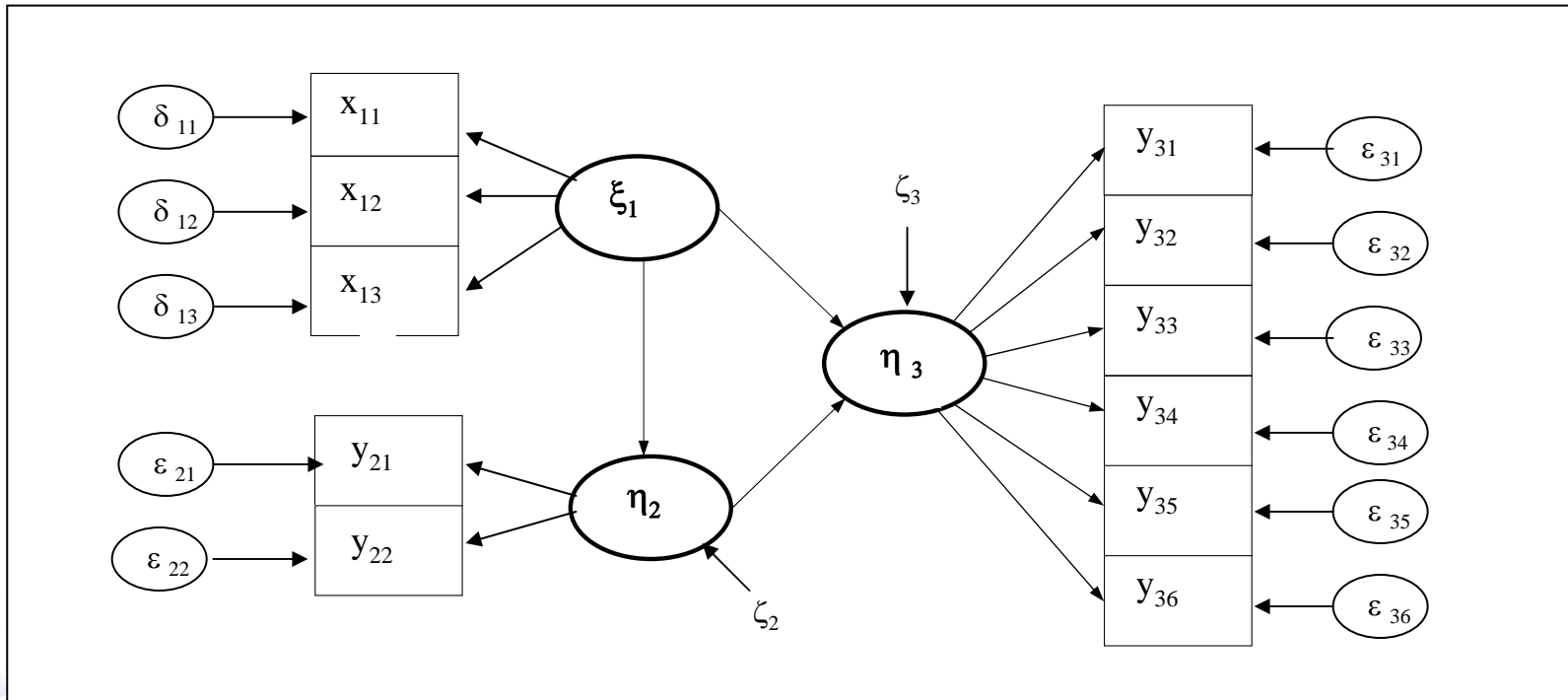
- *L'estimation des paramètres de ce modèles peut se faire soit:*
  - *Par l'approche LISREL (linear structural relationships).*
  - *Par l'approche PLS (partial least squares path modeling)*

# La méthode LISREL

Analyse de la structure de covariance  
Structural Equation Modeling (SEM)  
Covariance Structure Analysis

...

# Le modèle structurel de la covariance (LISREL)



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

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

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

*Eq. du modèle structurel*

*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
  - 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

## **Les moins connus:**

- **EQS** (Bentler, 1995)
- **M-PLUS** (Muthén & Muthén)
- **TETRAD** (Spirtes, Scheines et Glymour)

# *Le principe*

---

- Une méthodologie générale pour spécifier, estimer, comparer et évaluer des modèles de relations entre variables.
- On va chercher à confirmer 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
- $\Sigma$  = Matrice de covariance au niveau de la population
- $S$  = Matrice des covariances observées
- $C$  = Matrice des covariances obtenue grâce au modèle
- $\Phi$  = Matrice de covariance de  $\xi$
- $\Psi$  = Matrice de covariance de  $\zeta$

## Estimation des paramètres du modèle - 2

$$\begin{aligned}\eta_{(m \times 1)} &= B_{(m \times m)} * \eta_{(m \times 1)} + \Gamma_{(m \times n)} * \xi_{(n \times 1)} + \zeta_{(m \times 1)} \\ y_{(p \times 1)} &= \Lambda_{y(p \times m)} * \eta_{(m \times 1)} + \varepsilon_{(p \times 1)} \\ x_{(q \times 1)} &= \Lambda_{x(q \times n)} * \xi_{(n \times 1)} + \delta_{(q \times 1)}\end{aligned}$$

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

$$\begin{aligned}C &= \begin{pmatrix} \text{cov}(\eta) & \text{cov}(\eta, \xi) \\ \text{cov}(\xi, \eta) & \text{cov}(\xi) \end{pmatrix} \\ &= \begin{pmatrix} (I - B)^{-1} (\Gamma \Phi \Gamma' + \Psi) & (I - B)^{-1} \Gamma \Phi \\ \Phi \Gamma' (I - B)^{-1} & \Phi \end{pmatrix}\end{aligned}$$

## *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).

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 du ML revient à la minimisation de:

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

## Autres fonctions: - Generalised Least Squares

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

## - Unweighted Least Squares

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

## - Asymptotically Distribution Free

# Validation simple du modèle

## Tests de validation globale du modèle:

- Si le modèle étudié est « exact » alors

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

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

- Le modèle est accepté si  $\frac{\chi^2}{DF} \leq 3$  et  $p\text{-valeur} \leq 0.05$   
(seuils généralement utilisés)

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

# Un indice plus « évolué »: RMSEA

## Le RMSEA (Root Mean Square Error of Approximation): (Steiger et Lind)

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

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

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

En pratique, on l'estime avec:

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

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

# *Un exemple: L'engagement amoureux*

---

## **Données:**

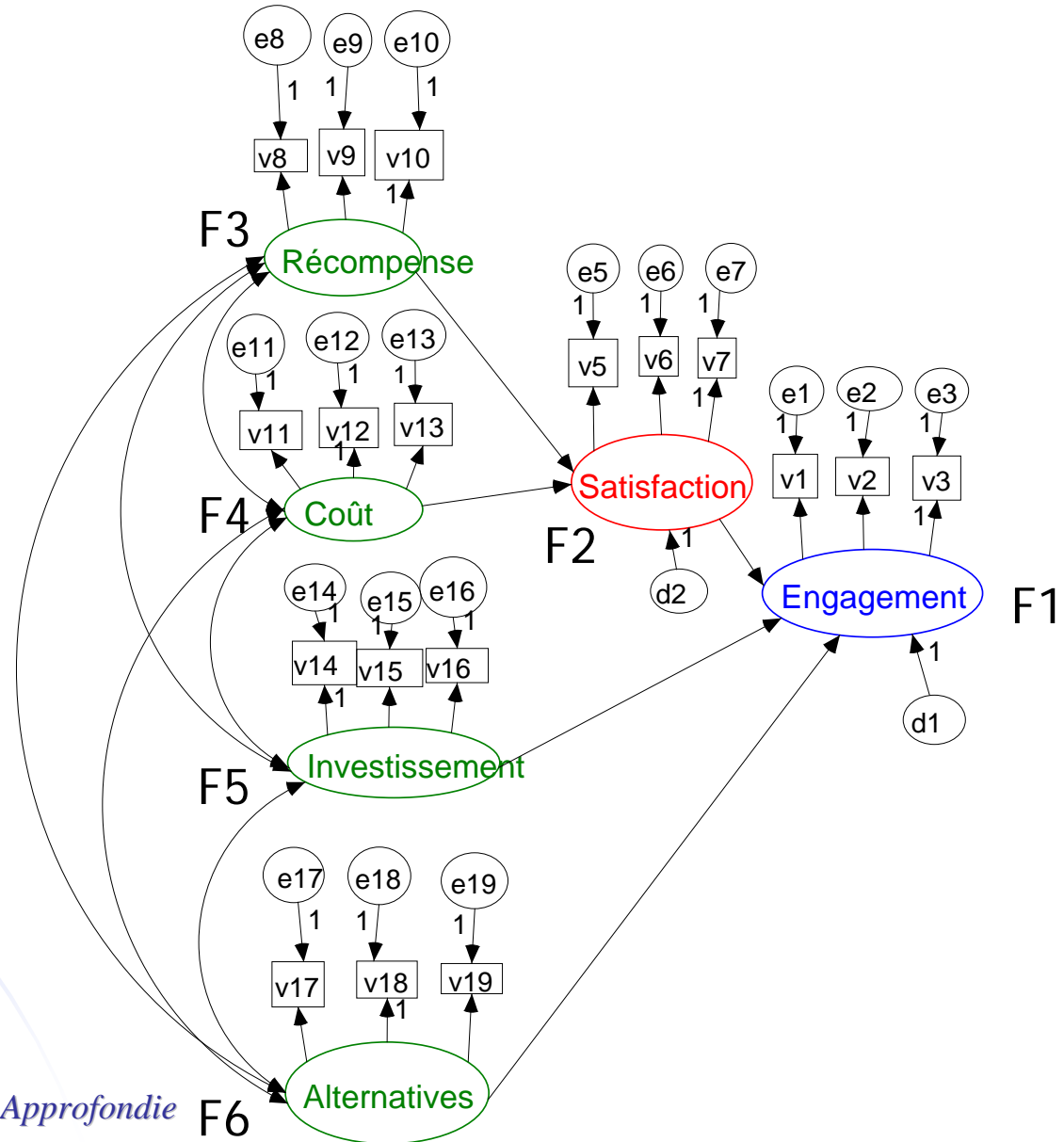
- C.E. Rusbult, « Commitment and satisfaction in romantic associations », *Journal of Experimental Social Psychology*, 1980

(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
4. Je me sens bien à propos de 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 bloc est unidimensionnel si  $> 0,7$ )

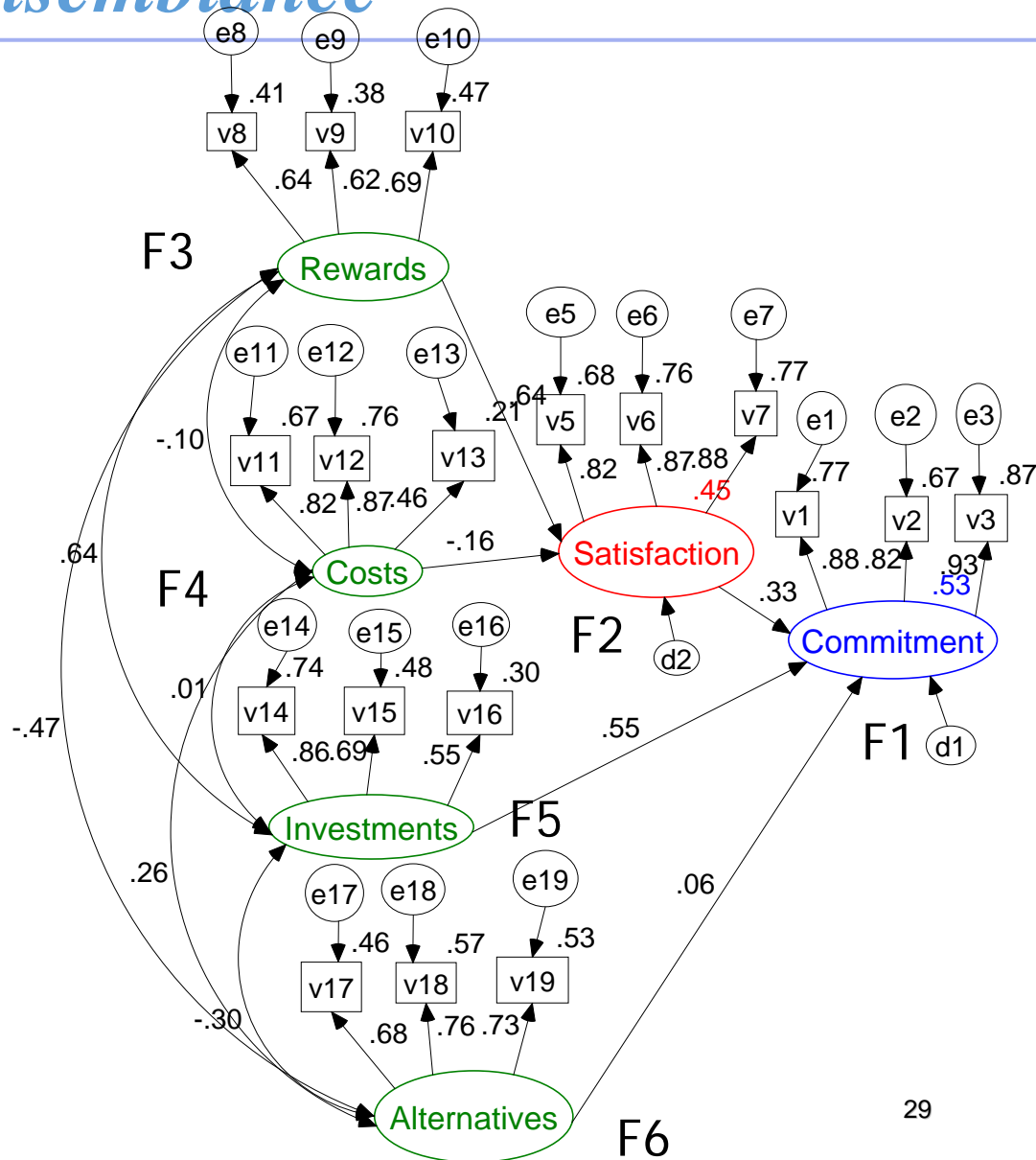
$$\alpha_j = \frac{\frac{1}{p_j(p_j-1)} \sum_{i \neq k} \text{cor}(\mathbf{x}_{ij}, \mathbf{x}_{kj})}{\frac{1}{p_j-1} + \frac{1}{p_j(p_j-1)} \sum_{i \neq k} \text{cor}(\mathbf{x}_{ij}, \mathbf{x}_{kj})} \times \frac{p_j}{p_j-1}$$

$$\rho_j = \frac{\left(\sum_i \lambda_{ij}\right)^2 \times \text{var}(\xi_j)}{\left(\sum_i \lambda_{ij}\right)^2 \times \text{var}(\xi_j) + \sum_i \text{var}(\epsilon_{ij})}$$

# Un exemple: Estimation par maximum de vraisemblance

Chi-square = 216.75  
 DF = 124  
 p-value = .000  
 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		<u>0.9136</u>		

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  $\chi^2$  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**

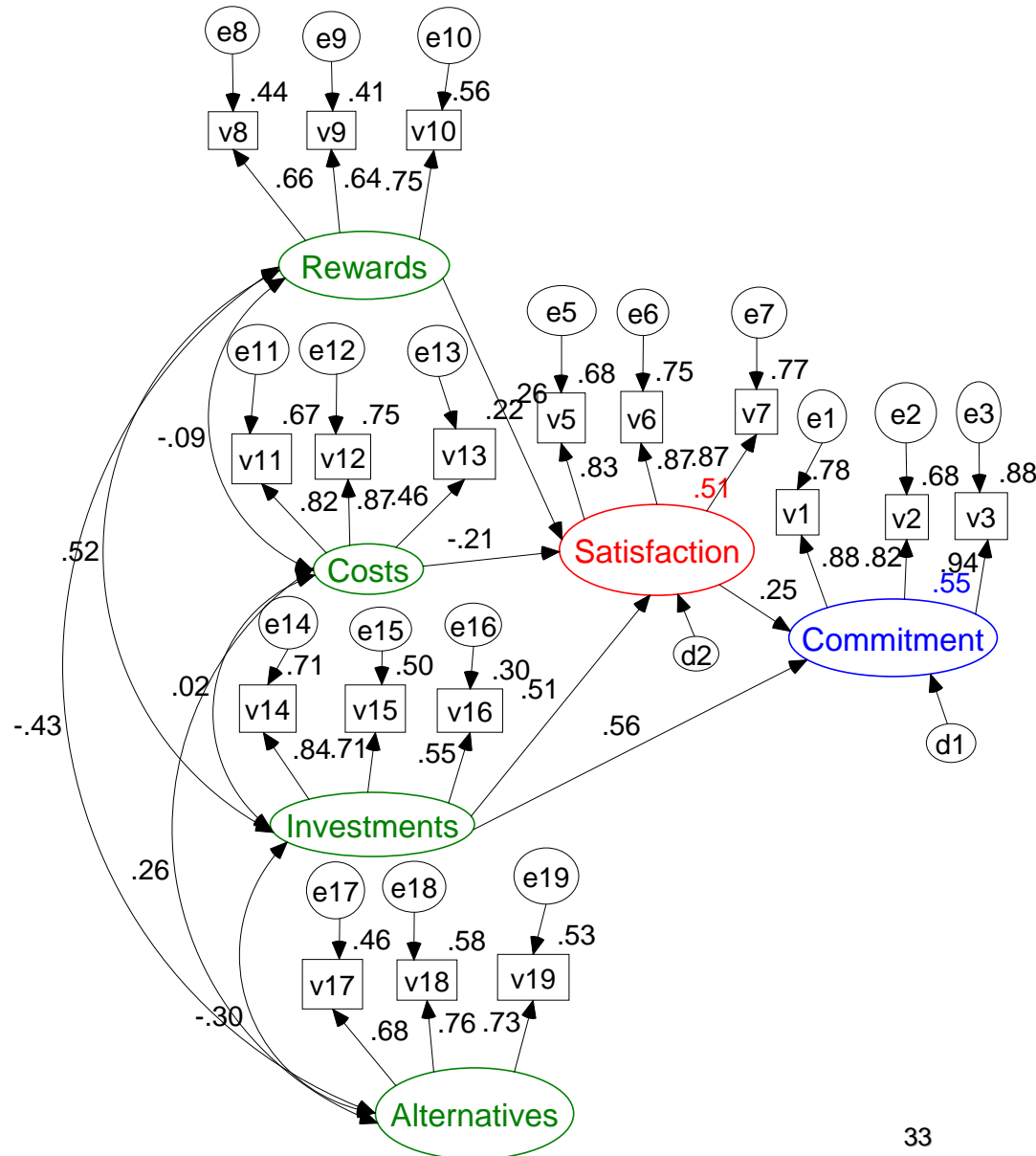
Il existe aussi des indices (Univariate Wald Test) qui estime l'augmentation du  $\chi^2$  quand on retire un lien au modèle

# Un exemple: les indices de modification

Chi-square = 183.191  
 DF = 124  
 p-value = .000  
 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

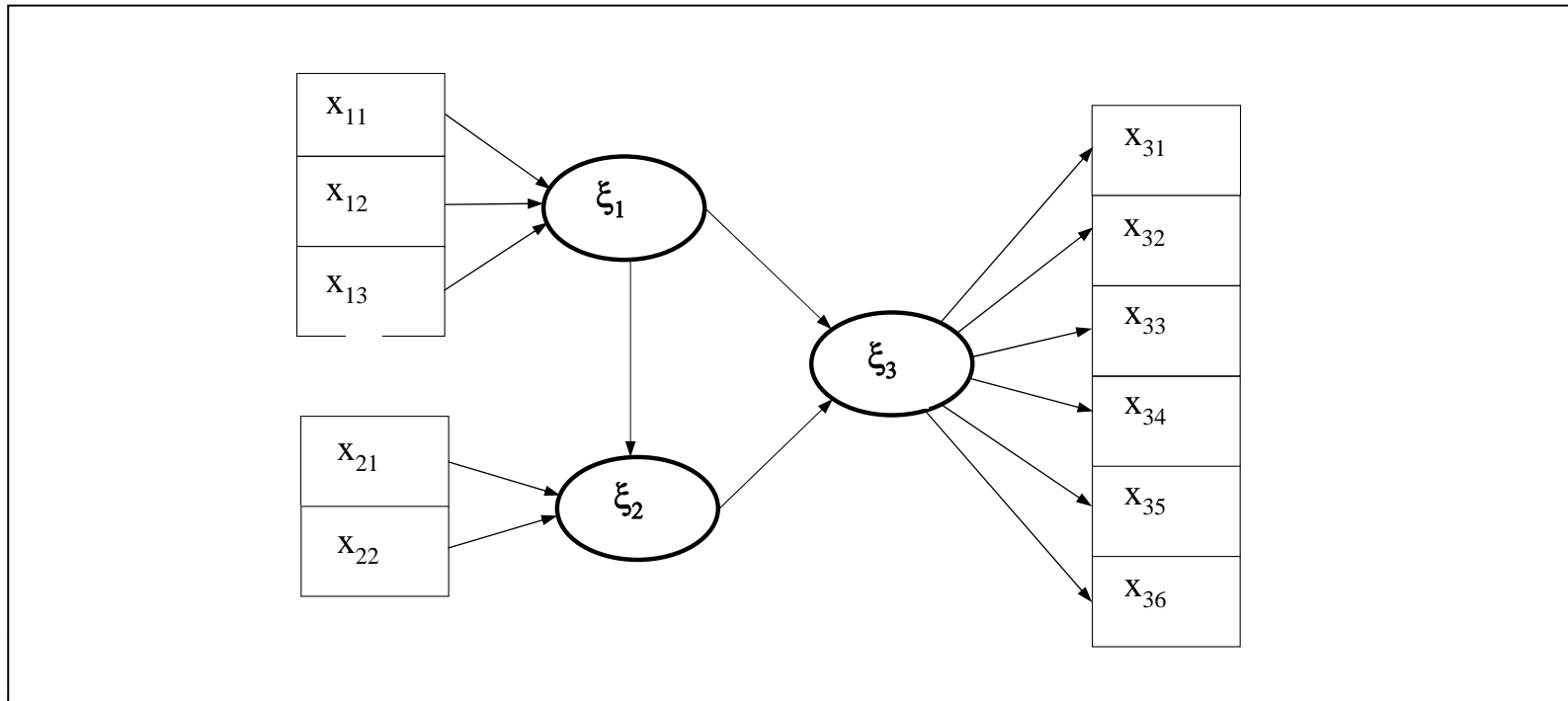
➔ Les indices de qualité d'ajustement sont améliorés



# L'approche PLS

Partial Least Squares Path Modeling

# Le « path diagram » (PLS)



$$\mathbf{X}_{jh} = \pi_{jh} \xi_j + \boldsymbol{\varepsilon}_{jh}$$

$$\xi_j = \sum_i \beta_{ji} \xi_i + \mathbf{v}_j$$

# 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.

# Les logiciels

## Les principaux:

- LVPLS (Löhmoller, 1989): Premier logiciel, très ancien
- PLS-Graph (W. Chin, 1996): très complet et convivial, usage gratuit pour le milieu académique
- SPAD-PLS (SPAD, 2005): Seul logiciel commercial, complet
- SmartPLS (Ringue, 2004): Logiciel gratuit et complet ([www.smartpls.de](http://www.smartpls.de))

## *Approche PLS – Le principe*

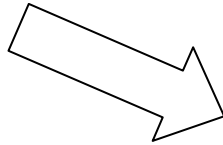
### **Principe :**

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.

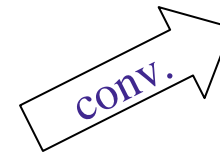
# Algorithme PLS

Poids externes  
initiaux  $w_{jh}$



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

Estimation externe de  $\xi_j$



Estimation des  
paramètres par  
régression

Calcul de poids  
externes  $w_{ij}$



Calcul des poids  
internes  $e_{ji}$



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

Estimation interne de  $\xi_j$

## 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 estime les variables latentes en se basant sur le modèle externe (chaque variable latente est estimée en fonction des variables manifestes de son bloc)

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

3. On estime les variables latentes en se basant sur le modèle interne (chaque variable latente est estimée en fonction des autres variables latentes qui lui sont liées)

$$\mathbf{z}_j = \sum_{y_i \rightarrow y_j} e_{ji} \mathbf{y}_i$$

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): 
$$w_{jh} = \text{COV} \left( x_{jh}, Z_j \right)$$

→ 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:  $e_{ji} = \text{coef. de régression multiple de } \mathbf{y}_j \text{ sur } \mathbf{y}_i \text{ si elles sont reliées}$

## *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 basé 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): Communauté et redondance

Communauté → évalue la qualité du modèle externe :

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

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) :

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

*Autre solutions:* la validation croisée (cv-communauté, cv-redondance, Q<sup>2</sup> 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*

---

## **Instabilité politique**

**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)

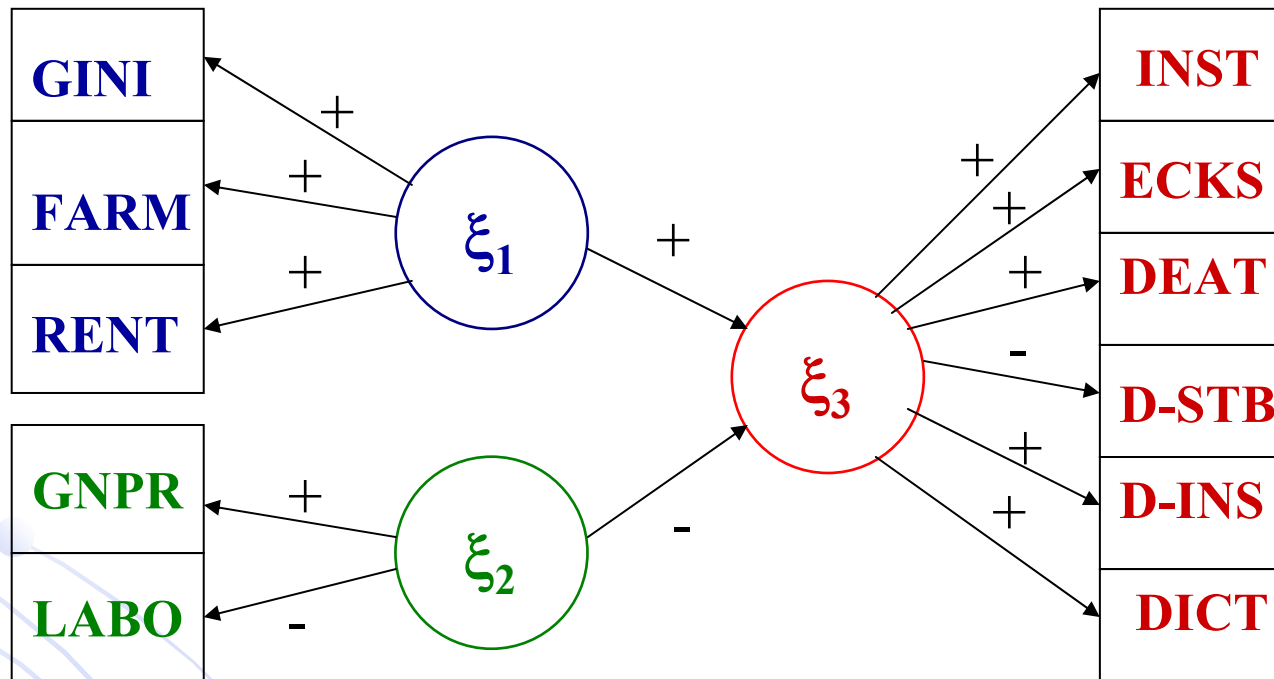
**D-STAB** : Démocratie stable

**D-INS** : Démocratie instable

**DICT** : Dictature

# Un exemple: Le modèle

## Inégalité agricole ( $\xi_1$ )

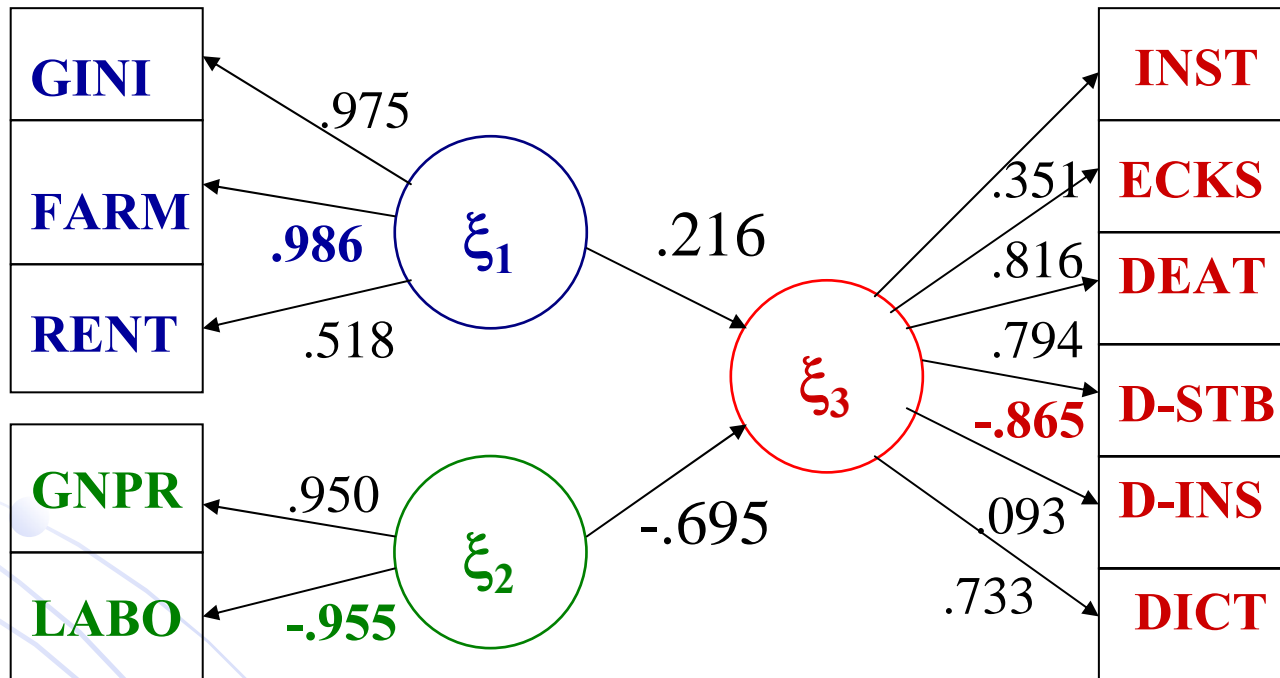


## Développement industriel ( $\xi_2$ )

## Instabilité politique ( $\xi_3$ )

# Un exemple: Estimation avec PLS

## Inégalité agricole ( $\xi_1$ )



## Développement industriel ( $\xi_2$ )

Les coefficients sur les arcs sont des corrélations

## Instabilité politique ( $\xi_3$ )

# *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 externe 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

```
Outer Model
=====
Variable          Weight          Corr.
-----
      ineg_agr    outward
gini                .4567          .9745
farm                .5125          .9857
rent                .1018          .5156
-----
      dev_indu    outward
gnpr                .5113          .9501
labo               -.5384         -.9551
-----
      inst_pol    outward
inst                .1187          .3676
ecks                .2855          .8241
deat                .2977          .7910
demostab           -.3271         -.8635
demoinst           .0370          .1037
dictatur           .2758          .7227
=====
```

# Un exemple: Le modèle interne

Estimation des variables latentes (scores obtenus après convergence de l'algorithme)

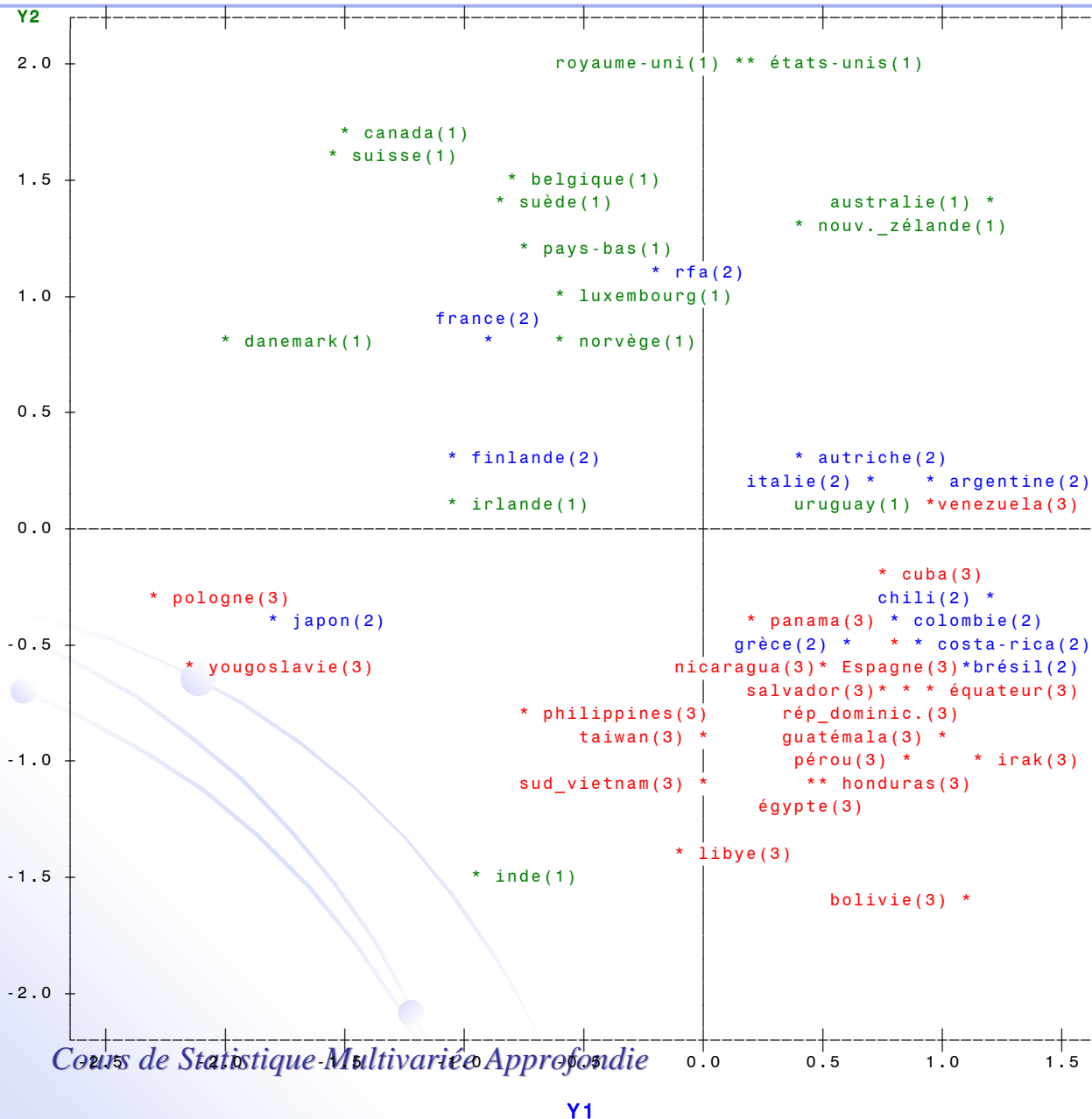
	$Y_1$	$Y_2$	$Y_3$
Argentine	0.96	0.24	0.75
Australie	1.20	1.37	-1.62
Autriche	0.39	0.25	-0.48
:			
France	-0.88	0.80	0.56
:			
Yougoslavie	-2.19	-0.65	0.13

Équation structurelle reliant les variables latentes endogènes aux variables latentes exogènes

$$\text{Instabilité politique} = 0.217 \times \text{Inégalité agricole} - 0.692 \times \text{Développement industriel}$$
$$R^2 = 0.618 \quad (2.24) \quad (-7.22)$$

*Les t de Student sont obtenus grâce à la régression multiple (OLS)*

# Un exemple: Répartition des pays



**Répartition des pays:  
 $Y_1$  = inégalité agricole,  
 $Y_2$  = développement  
industriel**

# Un exemple: Les indices globaux

Outer Model

Variable	Weight	Corr.	Communality	Redundancy	
-----					
ineg_agr	outward				
gini	.4567	.9745	.9496	.0000	
farm	.5125	.9857	.9716	.0000	
rent	.1018	.5156	.2659	.0000	
-----					
dev_indu	outward				
gnpr	.5113	.9501	.9027	.0000	
labo	-.5384	-.9551	.9123	.0000	
-----					
inst_pol	outward				
inst	.1187	.3676	.1352	.0835	} Moyenne=0.28
ecks	.2855	.8241	.6792	.4197	
deat	.2977	.7910	.6257	.3867	
demostab	-.3271	-.8635	.7457	.4608	
demoinst	.0370	.1037	.0107	.0066	
dictatur	.2758	.7227	.5223	.3228	
=====					

$$GoF = \sqrt{\frac{1}{3} \sum Communauté_i \times R^2(Inst.Pol.)} = \sqrt{0.611 \times 0.618} = 0.614$$

# Comparaison des approches LISREL et PLS

# Différences théoriques - 1

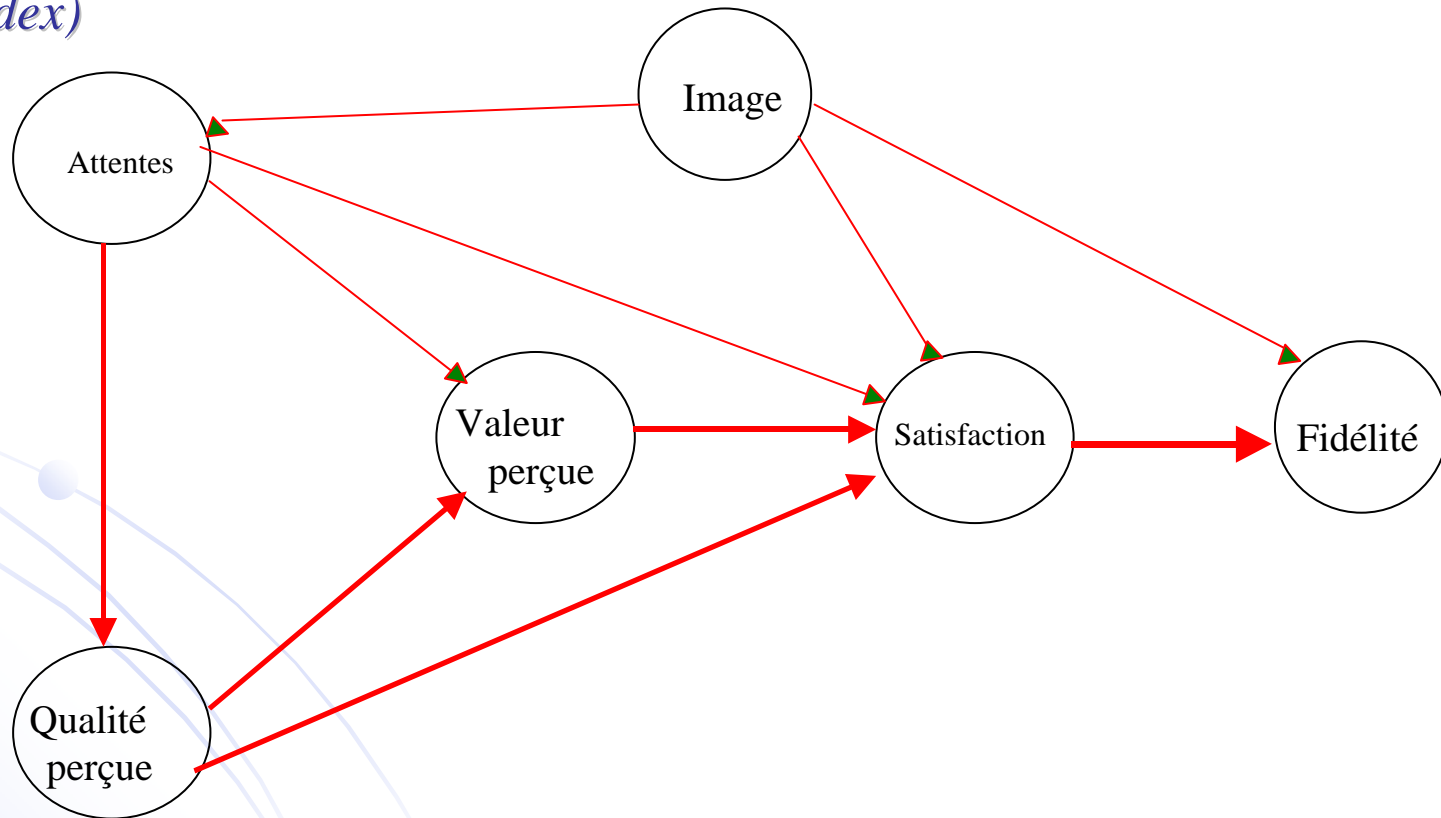
<i>Critère</i>	<i>Approche PLS</i>	<i>LISREL (Linear Structural Relationship)</i>
<i>Objectif</i>	Orientée vers la réalisation des prévisions	Orientée vers l'estimation des paramètres
<i>Méthodologie</i>	Basée sur variance	Basée sur covariance
<i>Variables latentes (VL)</i>	Combinaison linéaire de ses VM	Combinaison linéaire de toutes les VM
<i>Relations entre les VL et les VM</i>	Type réflectif ou formatif	Type réflectif
<i>Optimalité</i>	Pour la précision des prévisions	Pour la précision des paramètres
<i>Qualité des sous-modèles</i>	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

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

# Un exemple: Un questionnaire de satisfaction

*Nous devons 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:

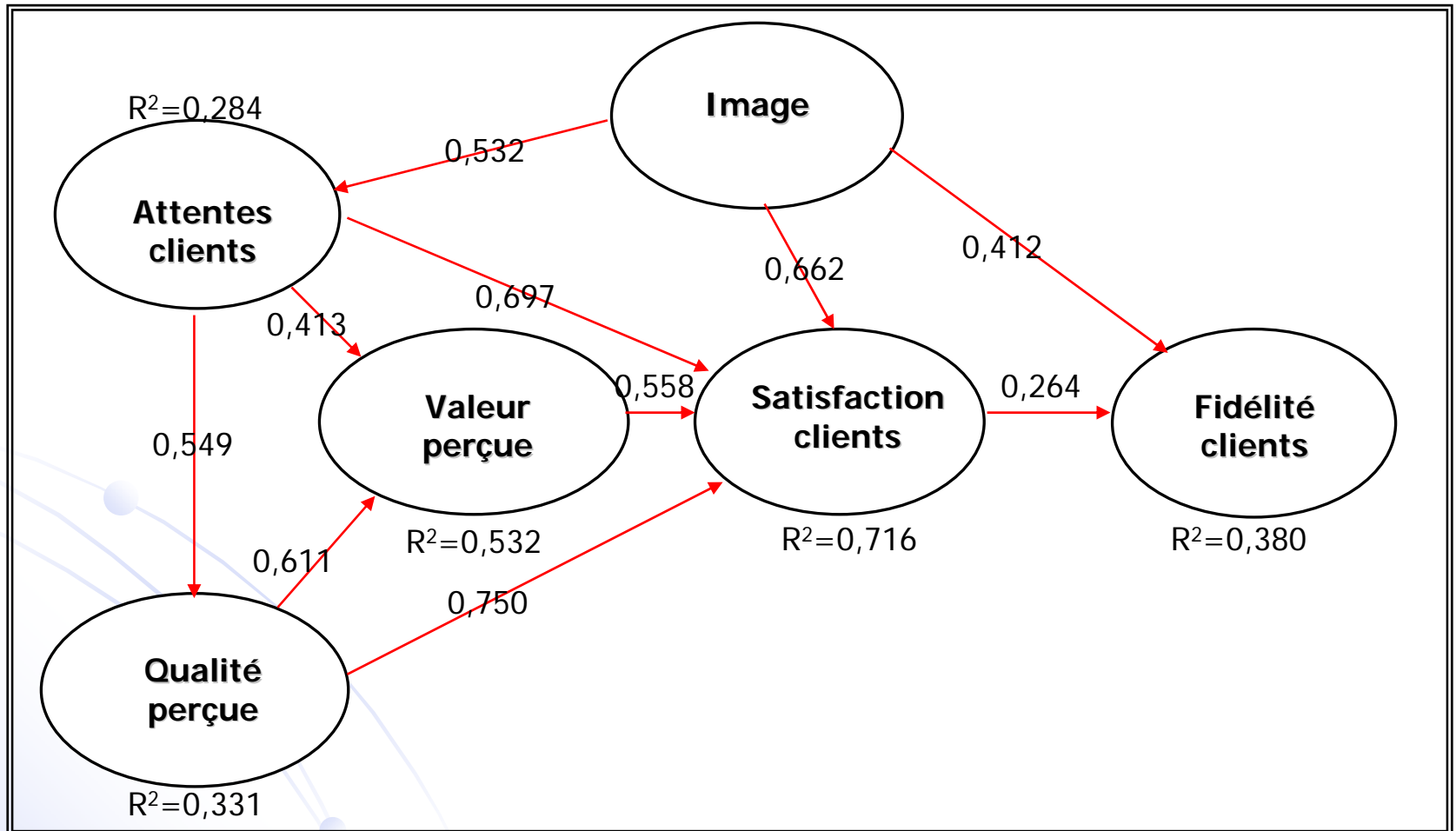
$$Satg = 0,170 \times Image + 0,290 \times Attente + 0,470 \times VP + 0,019 \times QP$$

Pour LISREL:

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

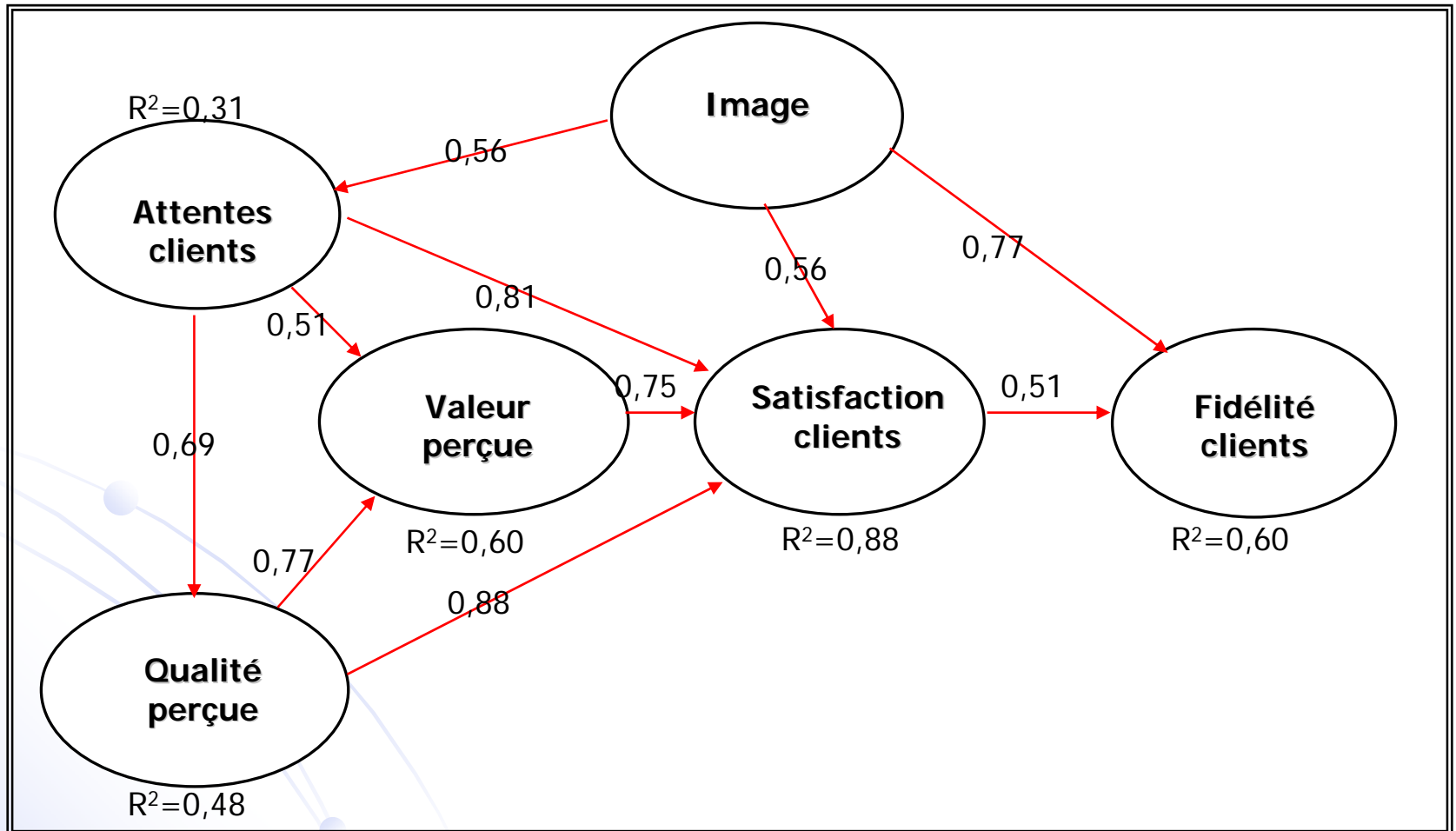
# Application - Modèle interne pour l'approche PLS

## 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:

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

où  $VM_i$  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.

# Bibliographies

- 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.