Free Model for Generalized Path Modeling and Comparison with Bayesian Networks

Christian Derquenne

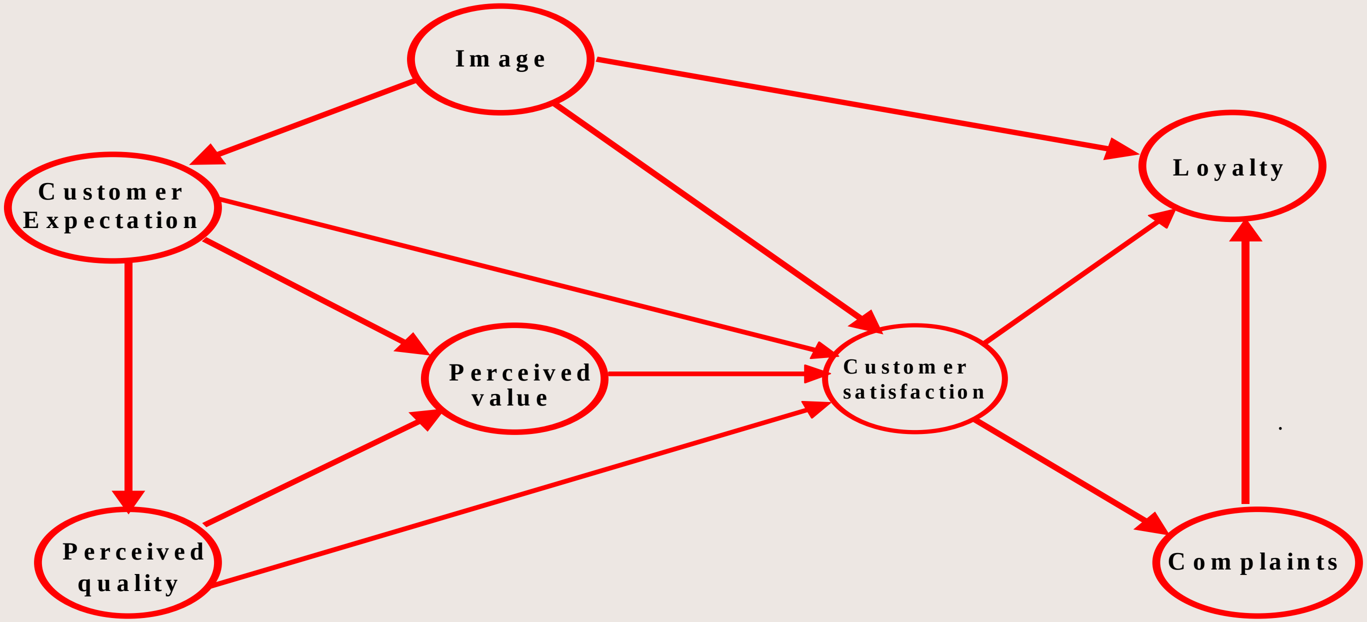
EDF

Research and Development

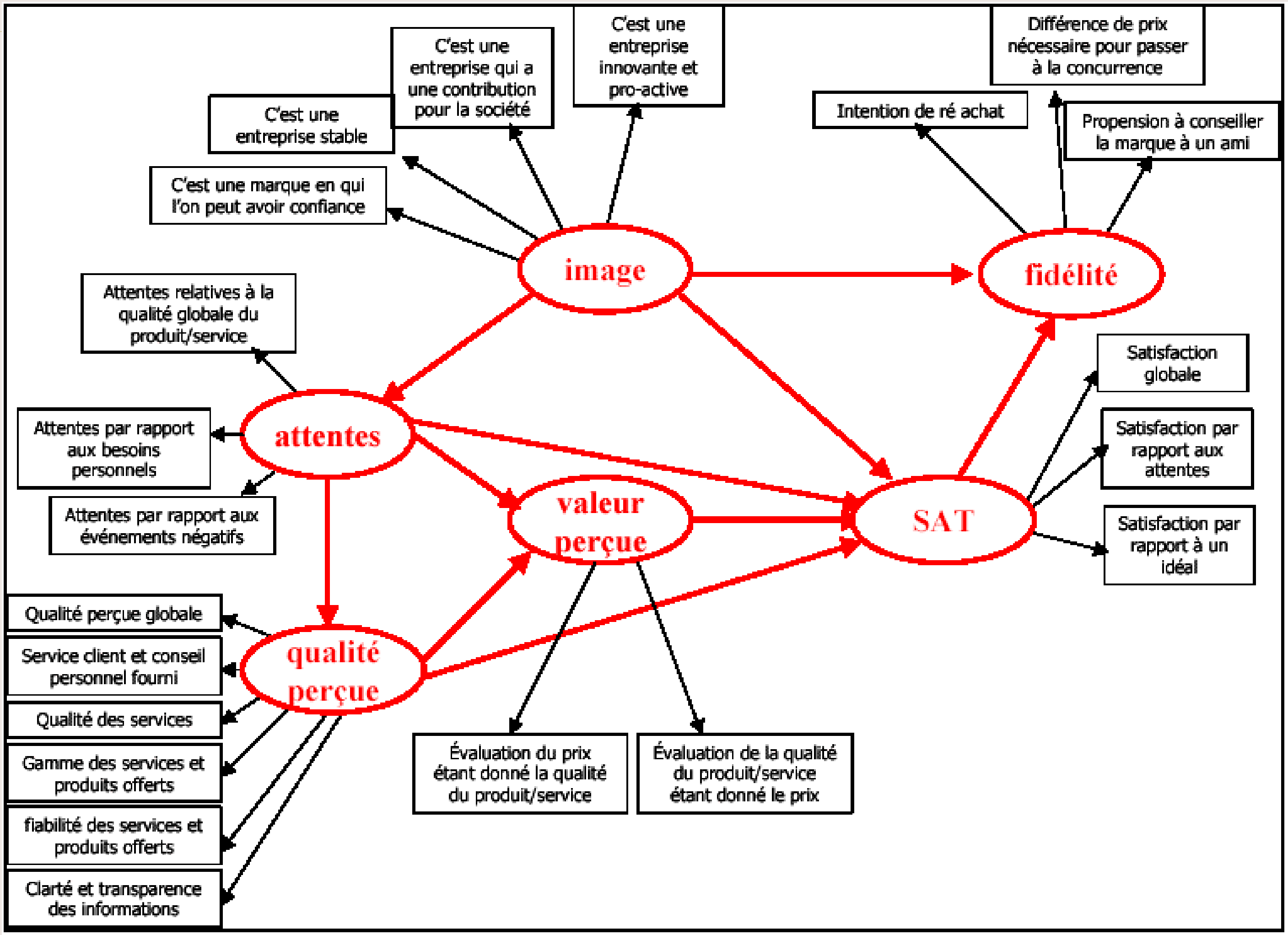
Outline

Issues and motivations

- **ECSI model** (or others) are **defined by the experts** whom are the knowledge of their application domains (marketing, policy, economics, sensiometrics…)



Issues and motivations



Issues and motivations

Un peu de vocabulaire :

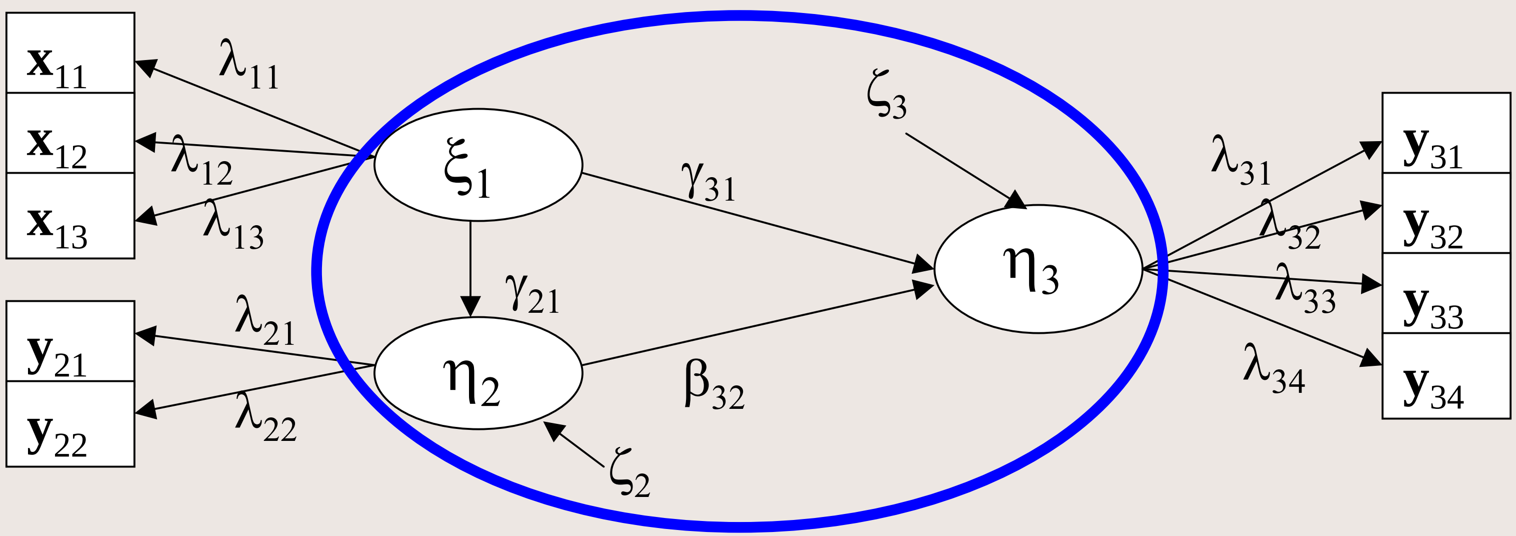
- Variable latente **exogène** = {Image}
- Variables latentes **endogènes** = {Qualité perçue ; Attentes ; Valeur perçue ; Satisfaction ; Fidélité}
- Variable latente **cible** = {Fidélité}
- Variables manifestes liées au bloc « **Attentes** »

{Attentes relatives à la qualité globale du produit/service ;
Attentes par rapport aux besoins personnels ;
Attentes par rapport aux événements négatifs }

- Modèle de mesure (externe) : lien entre les variables manifestes et les variables latentes
- Modèle de structure (interne) : lien entre les variables latentes

Issues and motivations

Structural model (internal)



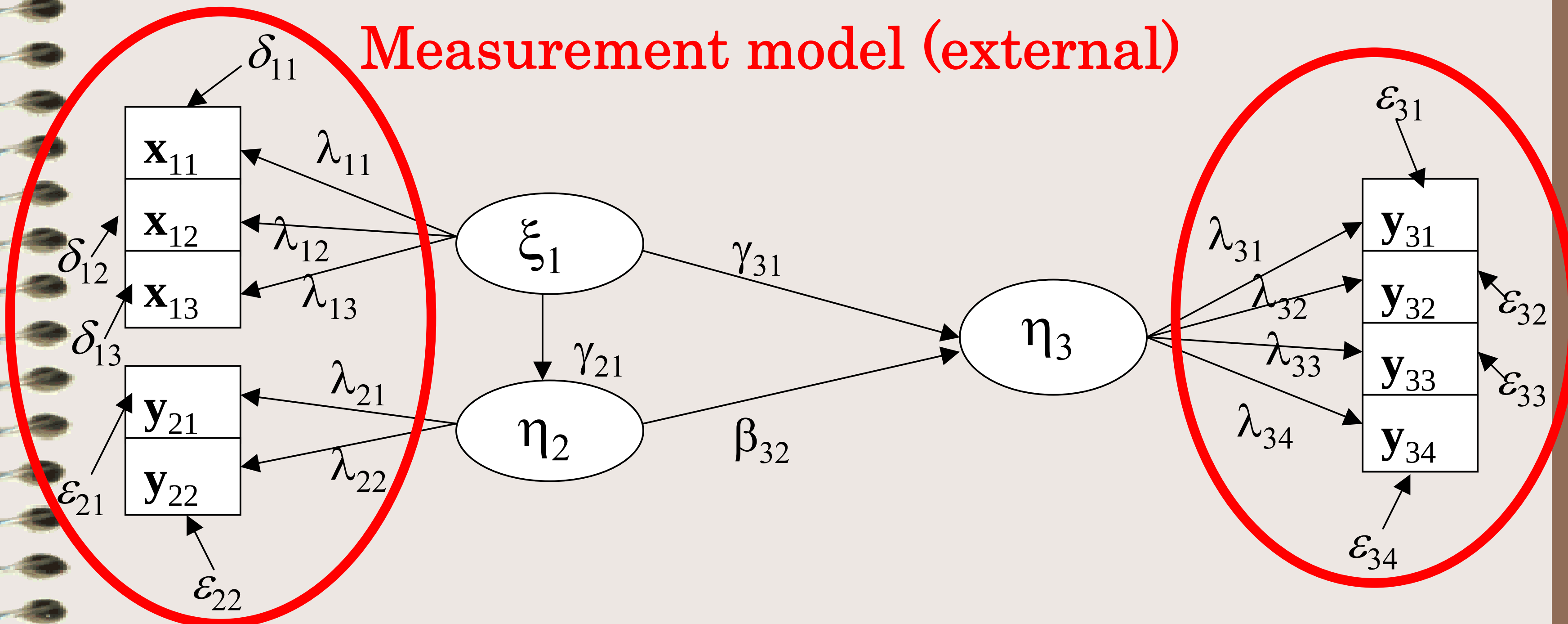
ξ_i, η_j : latent variables

β_{23}, γ_{21} et γ_{31} : path coefficients

$$\eta_2 = \gamma_{21}\xi_1 + \zeta_2$$

$$\eta_3 = \gamma_{31}\xi_1 + \beta_{32}\eta_2 + \zeta_3$$

Issues and motivations



$\mathbf{x}_{jh,}$: manifest var. (observed) of exogenous latent variables

$$x_{1h} = \lambda_{1h}\xi_1 + \delta_{1h},\ h = 1,2,3$$

λ_{1h} : external weight

$\mathbf{y}_{jh}$: manifest var. (observed) of endogenous latent variables

$$y_{jh} = \lambda_{jh}\eta_j + \varepsilon_{jh}\ \forall j,h$$

λ_{jh} : external weight

Méthodes d'estimation de ces modèles complexes

Deux types de structure des données

- Matrice de covariances (corrélations) entre les variables manifestes : **Méthode LISREL** (**LI**near **S**tructural **REL**ationships) : Minimise l'écart entre les matrices de covariances théorique (issues du modèle) et observée. Données Gaussiennes – EMV. [Jöreskog, 1979]

- Matrice individus×variables : **Approche PLS** (**P**artial **L**east **S**quares) [Wold H., 1982] et **approche RFPC** (**R**egression on the **F**irst **P**rincipal **C**omponents) [Derquenne Ch., 2001] : Minimise la variance résiduelle de toutes les variables à expliquer du modèle. Pas d'hypothèse de normalité, moindres carrés ordinaires.

Structure de données individuelles : l'approche PLS

Caractéristiques principales :

→ Analyse de la structure du tableau des **individus**

→ **Estimations séparées** des modèles de mesure et de structure (pas de critère global à optimiser)

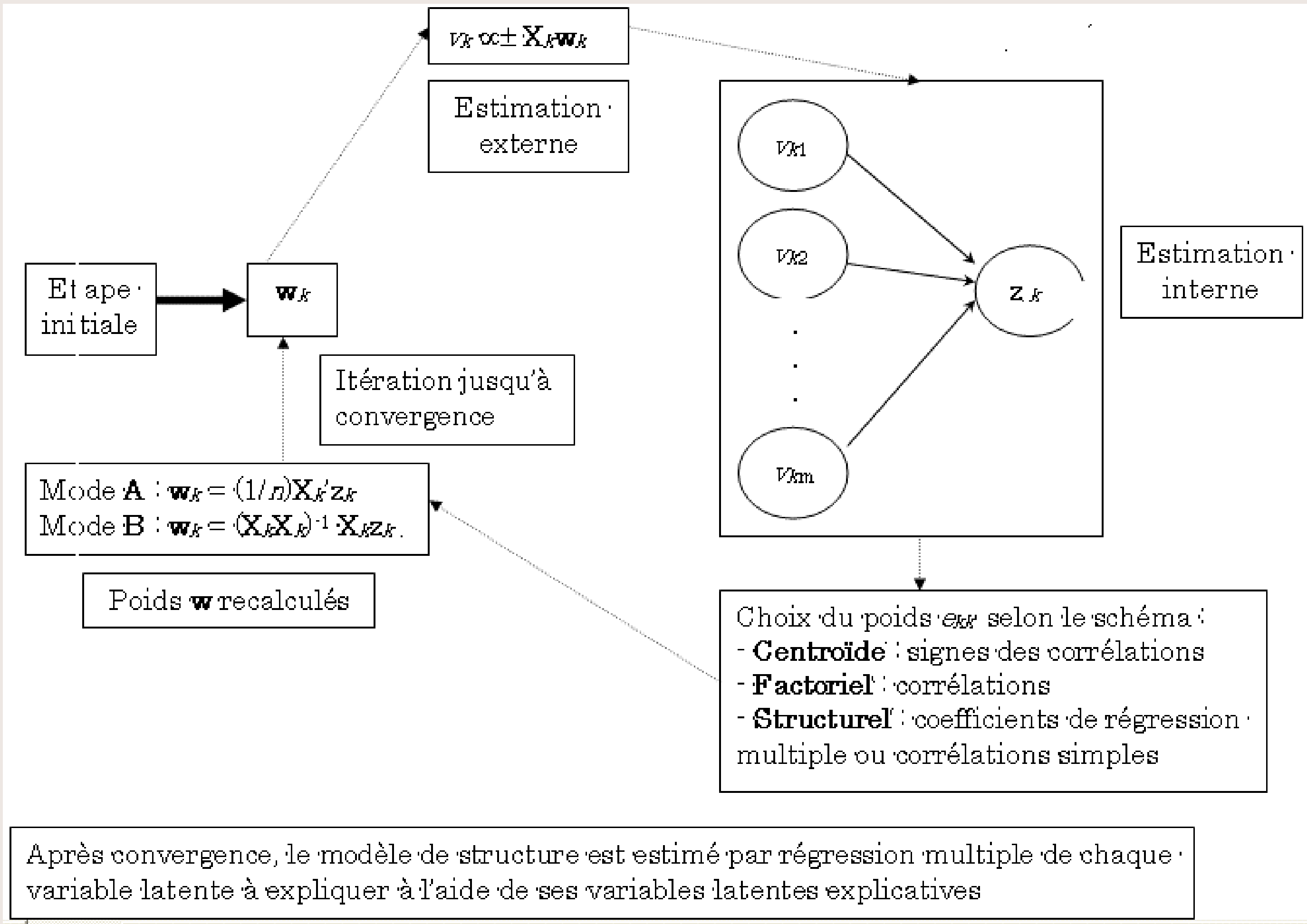
→ Estimation par la méthode **des moindres carrés ordinaires (PLS = Partial Least Squares)**

→ **Mode réflectif** et **formatif** pour le modèle externe

→ Méthode développée par **H. Wold** (1982, 1985)

Structure de données individuelles : l’approche PLS

Synoptique de l’estimation du modèle



L’approche RFPC

Caractéristiques principales :

→ Analyse de la structure du tableau des **individus**

→ **Estimations séparées** des modèles de mesure et de structure (pas de critère global à optimiser)

→ Utilisation de l’ACP pour estimer les variables latentes

⇒ Pas de système de calculs alternés entre le modèle interne et externe

⇒ Pas de problème de biais des paramètres

→ **Mode réflectif** seulement pour le modèle externe

Issues and motivations

- ECSI model (or others) are defined by the experts whom are the knowledge of their application domains (marketing, policy, economics, sensiometrics…)

⇒ **but these models are based on strong working hypothesis**

⇒ **fixed models**

Based on SEM (Strutural Equations Modelling) :

- Covariance (LISREL) [Jöreskog, 1979]
- PLS Approach [Wold H, 1982]
- RFPC Approach [Derquenne Ch., 2001]

Issues and motivations

These models are fixed at four main levels :

(i) **Choice of the manifest variables** corresponding to the questions of survey, for instance

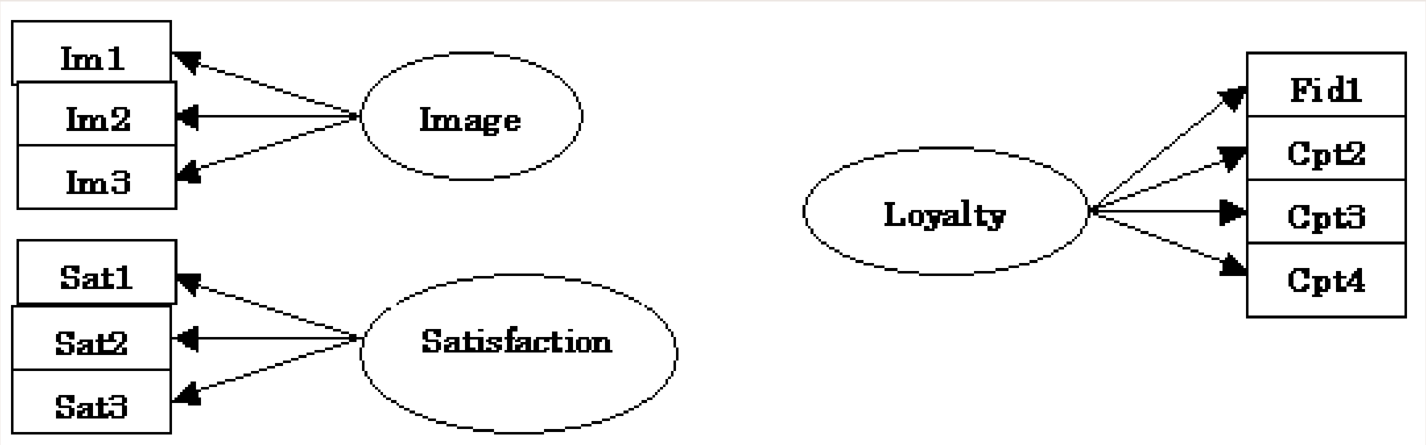
Im1	Sat2	Cpt4
Im2	Sat3	Fid1
Im3	Cpt2	
Sat1	Cpt3	

Issues and motivations

These models are fixed at four main levels :

(i) Choice of the manifest variables corresponding to the questions of survey, for instance

(ii) **Choice of blocks of manifest variables** constituting the latent variables

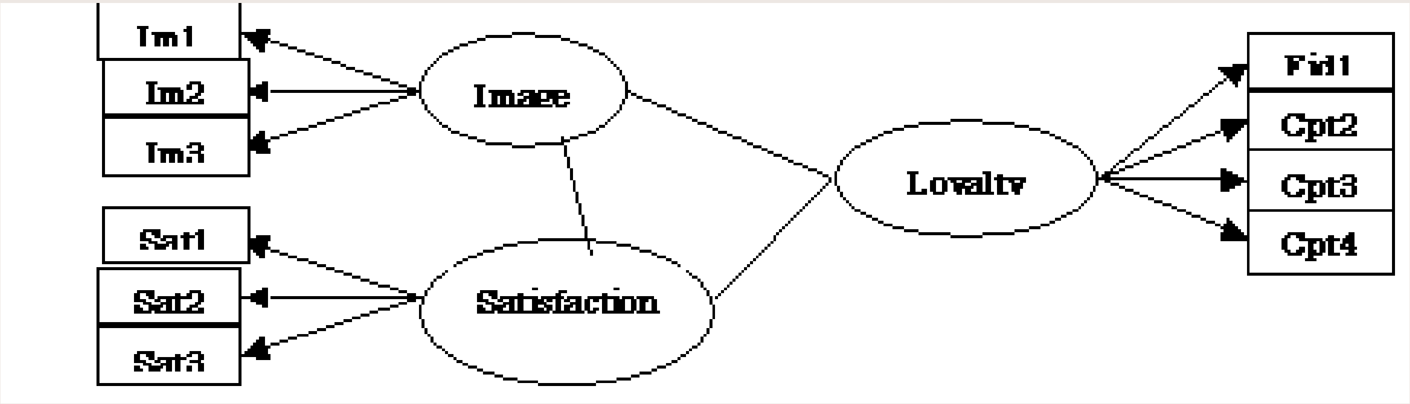


Issues and motivations

These models are fixed at four main levels :

(ii) Choice of blocks of manifest variables constituting the latent variables

(iii) Determination of presence or absence of **links between blocks**

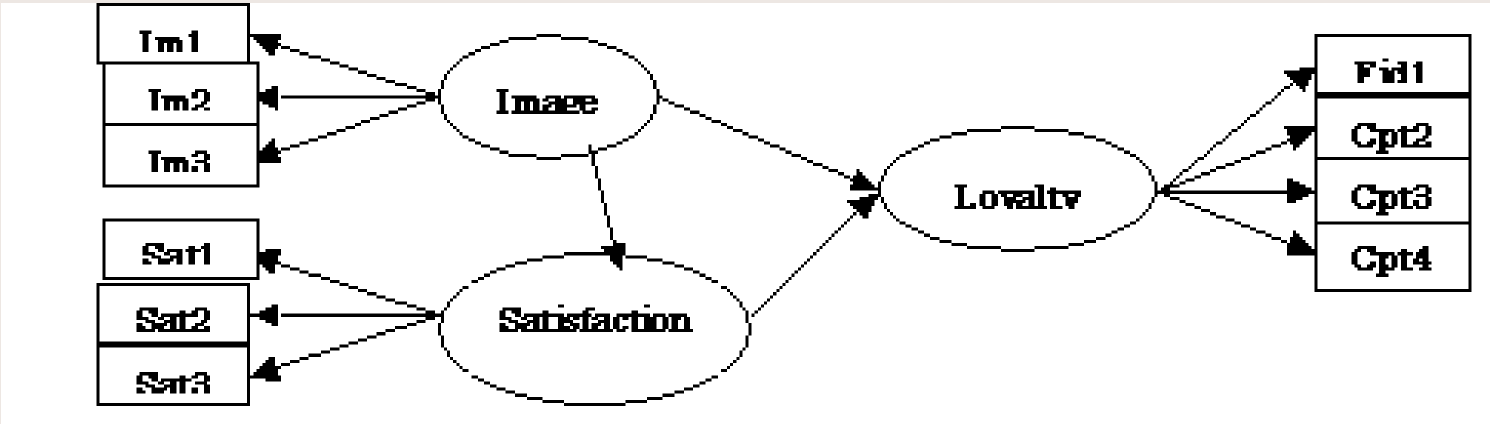


Issues and motivations

These models are fixed at four main levels :

(iii) Determination of presence or absence of links between blocks

(iv) Orientation of the links



Issues and motivations

The last three levels can be thrown back into question

⇒ We can leave “to talk the data”

⇒ Using of exploratory analysis

Feasible solution : A free model approach

Issues and motivations

The last three levels can be thrown back into question

⇒ We can leave "to talk the data"
⇒ Using of exploratory analysis

Feasible solution : A free model approach

⇒ A graphical model based on "causalities" (correlations) explored by direct/indirect observations

Outline

1 Issues and motivations

2 How to build a free model ?

3 Generalized Free Model approach

4 Bayesian Networks approach

5 Application to marketing data

6 Concluding remarks, applications and perspectives

How to build a free model ?

The free model in five steps (numeric scale)

Step 1 : Constituting of block of manifest variables
→ Clustering of manifest variables (Factor Analysis - rotation)

Step 2 : Building the latent variables
→ First component of Principal Components Analysis

Step 3 : Searching for links between latent variables
→ Using of partial correlation (Pearson or Spearman)

Step 4 : Orientation of links between latent variables
→ Usually made by expert, but possibility to use criteria: global R^2, GoF [Amato et al., 2004]

Step 5 : Estimation of the free model
→ Regression on the First Principal Components (RFPC) [Derquenne, 2000]

How to build a free model ?

The free model in five steps (numeric scale)

Step 1 : Constituting of block of manifest variables

→ Clustering of manifest variables (Factor Analysis - rotation)

Step 2 : Building the latent variables

→ First component of PCA or NIPALS [Wold H., 1973]

Step 3 : Searching for links between latent variables

→ Using of partial correlation (Pearson or Spearman)

Step 4 : Orientation of links between latent variables

→ Usually made by expert, but possibility to use criteria: global R^2, GoF [Amato et al., 2004]

Step 5 : Estimation of the free model

→ Regression on the First Principal Components (RFPC) [Derquenne, 2000]

How to build a free model ?

The free model in five steps (numeric scale)

Step 1 : Constituting of block of manifest variables

→ Clustering of manifest variables (Factor Analysis - rotation)

Step 2 : Building the latent variables

→ First component of PCA or NIPALS [Wold H., 1973]

Step 3 : Searching for links between latent variables

→ Using of partial correlation (Pearson or Spearman)

Step 4 : Orientation of links between latent variables

→ Usually made by expert, but possibility to use criteria: global R^2, GoF [Amato et al., 2004]

Step 5 : Estimation of the free model

→ Regression on the First Principal Components (RFPC) [Derquenne, 2000] (or LISREL ; PLS Approach)

How to build a free model ?

The free model in five steps (numeric scale)

Step 1 : Constituting of block of manifest variables

→ Clustering of manifest variables (Factor Analysis - rotation)

Step 2 : Building the latent variables

→ First component of PCA or NIPALS [Wold H., 1973]

Step 3 : Searching for links between latent variables

→ Using of partial correlation (Pearson or Spearman)

Step 4 : Orientation of links between latent variables

→ Usually made by expert, but possibility to use criteria: global R^2, GoF [Amato et al., 2004]

Step 5 : Estimation of the free model

→ Regression on the First Principal Components (RFPC) [Derquenne, 2000] (or LISREL ; PLS Approach)

How to build a free model ?

The free model in five steps (numeric scale)

Step 1 : Constituting of block of manifest variables

→ Clustering of manifest variables (Factor Analysis - rotation)

Step 2 : Building the latent variables

→ First component of PCA or NIPALS [Wold H., 1973]

Step 3 : Searching for links between latent variables

→ Using of partial correlation (Pearson or Spearman)

Step 4 : Orientation of links between latent variables

→ Usually made by expert, but possibility to use criteria: global R^2, GoF [Amato et al., 2004]

Step 5 : Estimation of the free model

→ Regression on the First Principal Components (RFPC) [Derquenne et al, 2003] (or LISREL ; PLS Approach)

How to build a free model ?

Context : In the applications of path modelling, data can be measured on different scales (survey questionnaires, databases, metadata, …)

→ *pseudo-numeric* (score from 1 to 10)

→ *semantic* ({yes ; no} or {very satisfy ; nearly satisfy ; nearly not satisfy ; not satisfy}, {electricity ; gas ; fuel ; other},…)

Problem : Step 1 and step 2 work only with manifest variables on a numeric scale

Solution : To generalize the free model approach to other scales of manifest variables (binary, ordinal, nominal, …)

How to build a free model ?

Context : In the applications of path modeling, data can be measured on different scales (survey questionnaires, databases, metadata, …)

→ *pseudo-numeric* (score from 1 to 10)

→ *semantic* ({yes ; no} or {very satisfy ; nearly satisfy ; nearly not satisfy ; not satisfy}, {electricity ; gas ; fuel ; other},…)

Problem : Step 1 and step 2 are only available for manifest variables on a numeric scale

Solution : To generalize the free model approach to other scales of manifest variables (binary, ordinal, nominal, …)

How to build a free model ?

Context : In the applications of path modeling, data can be measured on different scales (survey questionnaires, databases, metadata, ...)

→ *pseudo-numeric* (score from 1 to 10)

→ *semantic* ({yes ; no} or {very satisfy ; nearly satisfy ; nearly not satisfy ; not satisfy}, {electricity ; gas ; fuel ; other},...)

Problem : Step 1 and step 2 are only available for manifest variables on a numeric scale

Solution : To **generalize the free model approach** to other scales of manifest variables (binary, ordinal, nominal, ...)

Outline

1 Issues and motivations

2 How to build a free model ?

3 Generalized Free Model approach

4 Bayesian Networks approach

5 Application to marketing data

6 Concluding remarks, applications and perspectives

Generalized Free Model

The Generalized Free Model (GFM) in five steps

Step 1 : Constituting of block of manifest variables

→ Clustering of manifest variables (Exponential Family + Generalized Linear Models)

Step 2 : Building the latent variables

→ First component of the GPCA (Generalized NIPALS) [Derquenne, 2005]

Step 3 : Searching for links between latent variables

→ Using of partial correlation (Pearson or Spearman)

Step 4 : Orientation of links between latent variables

→ Usually made by expert, but possibility to use criteria: global R^2, GoF [Amato et al., 2004]

Step 5 : Estimation of the free model

→ Regression on the First Generalized Principal Components (RFGPC) [Derquenne, 2005]

Generalized Free Model

The Generalized Free Model (GFM) in five steps

Step 1 : Constituting of block of manifest variables

→ Clustering of manifest variables (Exponential Family + Generalized Linear Models)

Step 2 : Building the latent variables

→ First component of the GPCA (Generalized NIPALS) [Derquenne, 2005]

Step 3 : Searching for links between latent variables

→ Using of partial correlation (Pearson or Spearman)

Step 4 : Orientation of links between latent variables

→ Usually made by expert, but possibility to use criteria: global R^2, GoF [Amato et al., 2004]

Step 5 : Estimation of the free model

→ Regression on the First Generalized Principal Components (RFGPC) [Derquenne, 2005]

Generalized Free Model

The Generalized Free Model (GFM) in five steps

Step 1 : Constituting of block of manifest variables

→ Clustering of manifest variables (Exponential Family + Generalized Linear Models)

Step 2 : Building the latent variables

→ First component of the GPCA (Generalized NIPALS) [Derquenne, 2005]

Step 3 : Searching for links between latent variables

→ Using of partial correlation (Pearson or Spearman)

Step 4 : Orientation of links between latent variables

→ Usually made by expert, but possibility to use criteria: global R^2, GoF [Amato et al., 2004]

Step 5 : Estimation of the free model

→ Regression on the First Generalized Principal Components (RFGPC) [Derquenne, 2005]

Generalized Free Model

The Generalized Free Model (GFM) in five steps

Step 1 : Constituting of block of manifest variables

→ Clustering of manifest variables (Exponential Family + Generalized Linear Models)

Step 2 : Building the latent variables

→ First component of the GPCA (Generalized NIPALS) [Derquenne, 2005]

Step 3 : Searching for links between latent variables

→ Using of partial correlation (Pearson or Spearman)

Step 4 : Orientation of links between latent variables

→ Usually made by expert, but possibility to use criteria: global R^2, GoF [Amato et al., 2004]

Step 5 : Estimation of the free model

→ Regression on the First Generalized Principal Components (RFGPC) [Derquenne, 2005]

Generalized Free Model

The Generalized Free Model (GFM) in five steps

Step 1 : Constituting of block of manifest variables

→ Clustering of manifest variables (Exponential Family + Generalized Linear Models)

Step 2 : Building the latent variables

→ First component of the GPCA (Generalized NIPALS) [Derquenne, 2005]

Step 3 : Searching for links between latent variables

→ Using of partial correlation (Pearson or Spearman)

Step 4 : Orientation of links between latent variables

→ Usually made by expert, but possibility to use criteria: global R^2, GoF [Amato et al., 2004]

Step 5 : Estimation of the generalized free model

→ Regression on the First Generalized Principal Components (RFGPC) (or PML Approach) [Derquenne, 2005]

Generalized Free Model

Step 1 : Clustering of manifest variables (1)

Let $(X_1, \ldots, X_p)$ be, p manifest variables measured on different scales

Principle : Each variable (response variable) is modeled by the p-1 others (explanatory variables), through an adapted model

Example : if X_j is ordinal dependent variable with R_j categories

$\Rightarrow$ Ordinal Logit Model : Cumulated probability to category r_j :

$$\Pr\left[X_j \le r_j / X_k, k=1,\ldots p, k \ne j\right] = \exp\left(\alpha_{r_j} + \sum_{k \ne j} \mathrm{x}_k \beta_k\right) \Big/ \left(1 + \exp\left(\alpha_{r_j} + \sum_{k \ne j} \mathrm{x}_k \beta_k\right)\right) \qquad (1)$$

Example : if X_j is nominal dependent variable with R_j categories

$\Rightarrow$ Multinomial Logit Model : Individual probability for category r_j :

$$\Pr\left[X_j = r_j / X_k, k=1,\ldots p, k \ne j\right] = \exp\left(\alpha_{r_j} + \sum_{k \ne j} \mathrm{x}_k \beta_k^{(r_j)}\right) \Big/ \left(1 + \sum_{r_j=1}^{R_j-1} \exp\left(\alpha_{r_j} + \sum_{k \ne j} \mathrm{x}_k \beta_k^{(r_j)}\right)\right) \qquad (2)$$

Generalized Free Model

Step 1：Clustering of manifest variables (2)

Using of **marginal test of likelihood ratio** to calculate the intensity of link between X_j and X_k：

$$\Lambda\left(X_j / X_k\right) = -2\left[l(\hat{\beta}_{(k)}, X_j) - l(\hat{\beta}, X_j)\right] \qquad (3)$$

where $\hat{\beta}$ is a vector of the parameters associated to the full model and $\hat{\beta}_{(k)}$ is a vector of the parameters without X_k

$$p\text{-value}(X_j/X_k) = \Pr\left[\chi^2_{v_k} > \Lambda_{obs}(X_j / X_k)\right] \qquad (4)$$

Calculation of the dissimilarity between X_j and X_k：

$$d(X_j, X_k) = \min[p\text{-value}(X_j/X_k)\ ;\ p\text{-value}(X_k/X_j)] \qquad (5)$$

Generalized Free Model

Step 1: Clustering of manifest variables (3)

We obtain a **symmetric matrix of dissimilarities** (p,p)

The number of clusters G is chosen with the **Ward's hierarchical ascending clustering** on this matrix :

(i) The simplest rule cuts the clustering tree where there is a big loss of inter-cluster inertia

(ii) The improvement of this criterion :

$$R_W^2(G) > 1 - \frac{\alpha^2(p-G)}{T} \qquad (6)$$

where $R_W^2(G)$ is the proportion of within-cluster inertia and T is the total inertia

⇒ **The final result consists in G clusters, containing $(q_1, \ldots, q_G)$ manifest variables**

Generalized Free Model

Step 2 : Building the latent variables (1)

Each cluster contains q_g variables which can be measured on different scales, with $\sum_{g=1}^{G} q_g = p$ where it can have missing data

For each cluster, the **GPCA** is applied with aid of the **Generalized NIPALS algorithm**. This technique deals with too missing data according to the same principle as the NIPALS algorithm

Only the first principal component is used as in RFPC
$\Rightarrow$ **The final result consists in G latent variables, $(Z_1, \ldots, Z_G)$**

Particular case : if a block of variables are only on ordinal scale, without missing data we obtain the first principal component of the Multiple Correspondence Analysis

Generalized Free Model

Step 2 : Building the latent variables (2)

Methods to apply in accordance with the type of variables and missing data on not to obtain a latent variable

	Numeric	**Binary or ordinal**	**Nominal**	**Count**	**Mixed**
Not missing	PCA	MCA	MCA	Generalized NIPALS	Generalized NIPALS
Missing	NIPALS	Categorical NIPALS	Categorical NIPALS	Generalized NIPALS	Generalized NIPALS

Generalized Free Model

Step 3 : Searching for links between latent variables

All the G **latent variables** are on a numeric scale

The **partial correlation** between each couple (Z_s, Z_q) giving the G-2 other Z_g : $\rho(Z_s, Z_q / \forall\ Z_g,\ g \neq s, q)$ is calculated

The associated statistic of test follows a **Student's distribution** with n-G-1 deg. of freedom under H_0 (non-correlation between Z_s and Z_q)

The link between two latent variables is considered significant if the p-value is lower than a fixed threshold (for instance : 0.01)

This procedure leads to $K \in [0, G(G-1)/2]$ links

⇒ **The final result is a non-directed graph**

Generalized Free Model

Step 4 : Orientation of links between latent variables

It is usually preferable to use **an expert decision** to avoid a non efficient model

However, different **statistic solutions** can be proposed

If there are K significant links, then the number of different models is equal to 2^K

Then, each model can be judged by a score, (GoF, global R^2, …), and an expert can choose among the ten highest scores, for instance (these models must not have cycle in the directed graph)

⇒ **The final result is a directed graph**

Generalized Free Model

Step 5: Estimation of the generalized free model

The free model is estimated through the **RFGPC approach** (Regression on the First Generalized Principal Components) [Derquenne, 2005] or by the **PML approach** (Partial Maximum Likelihood) [Derquenne, 2005]

The **internal model** is estimated by means of multiple linear regression between latent variables

The **external model** is estimated by means regression between each latent variable and its own manifest variables

⇒ **The final result is the generalized free model**

Outline

1 Issues and motivations

2 How to build a free model ?

3 Generalized Free Model approach

4 **Bayesian Networks approach**

5 Application to marketing data

6 Concluding remarks, applications and perspectives

Bayesian Networks approach

Bayesian Networks (BN) = graphical form

→ **Nodes** = model's variables

→ **Edges** = direct probability relations

→ Powerful tools for **non-supervised learning** [Jensen, 2001]

→ Used to **find links between observed variables** in order to create a robust outer model

Why to use BN for customer satisfaction surveys ?

→ **Handled using questions with nominal or ordinal scales**

Bayesian Networks approach

Overall principle (1) :

→ Using structure learning heuristics based on network scores by **Minimum Description Length** (MDL)

→ Search based on a global characterization of the data and on the exploitation of the properties of equivalent Bayesian networks with **SopLEQ** algorithm [Jouffe, 2004]

→ Building a **segmentation of variables** with aid of this graph is applied to obtain clusters of nodes (corresponding to clustering variables = **step 1** of **GFM**)

→ For each cluster, a **categorical latent variable is built** by clustering the individuals, as a new node (**step 2**) and the external model is created

Bayesian Networks approach

Overall principle (2) :

→ **Search relations between these new nodes** with SopLEQ again on these new nodes (**step 3**)

→ **Build the structural model**, the position of links (**step 4**) is directly given by SopLEQ (possibly to choose a different position of links among equivalent Bayesian networks)

→ **Estimation of « Free model »** is given by making inference on full model (external and internal models) in selecting a target node among latent variables (**step 5**)

Bayesian Networks approach

Comparison of the both approaches

Steps	***Bayesian Networks***	***Generalized Free Models***
Searching links between manifest variables	SopLEQ Algorithm	Building dissimilarities matrix (p values) between each couple of variables by Generalized Linear Models
Clustering of variables	Segmentation of variables	Ward's distance on the dissimilarities matrix
Building latent variables	Segmentation of data	1st generalized principal component
Searching links between latent variables	SopLEQ Algorithm	Partial correlation test
Position of links between latent variables		By expert ou by selection of the "best" statistical model
Estimation of the Generalized Free Model		RFGPC approach or PML approach

Outline

Application to marketing data

Satisfaction and loyalty survey concerning electric central heating

Context : The marketing Department would like to evaluate the leverage of loyalty and satisfaction with respect to central heating. But, it doesn't exist a specific marketing conceptual model. However a satisfaction and loyalty survey has been conduct, but the questionnaire was not explicitly built to be used for a path modeling. Then this department asks to R&D to construct a first statistical model as an initial marketing model

Application to marketing data

Form of the questionnaire and results of survey :

Several parts :

- socio-economics questions (age, type of household, …)
- specific questions (satisfaction, expectations, loyalty, …)

Results : The answers are distributed on semantic scales

- 4 binary variables
- 5 nominal variables
- 19 ordinal variables

- 6899 questionnaires
- Some contain missing data

Application to marketing data

Results on the Generalized Free Model (1)

Step 1 : Clustering of manifest variables (8 clusters)

Expectation #1 (EXPEC_1: $q_1 = 5$)

Expectation #2 (EXPEC_2: $q_2 = 4$)

Socio-demographics (SOC_DEM: $q_3 = 4$)

Housing (HOUSING: $q_4 = 4$)

Customer's contract (CONTRA: $q_5 = 2$)

Loyalty (LOYALTY: $q_6 =2$)

Satisfaction/recommendation (SAT-REC: $q_7 = 3$)

Perceived value (PER-VAL: $q_8 =7$)

Step 2 : Building the 8 latent variables :

EXPEC_1 ; EXPEC_2 ; SOC_DEM ; HOUSING ; CONTRA ;LOYALTY ; SAT-REC; PER-VAL

Application to marketing data

Results on the Generalized Free Model (2)

Step 3 : Searching for links between latent variables

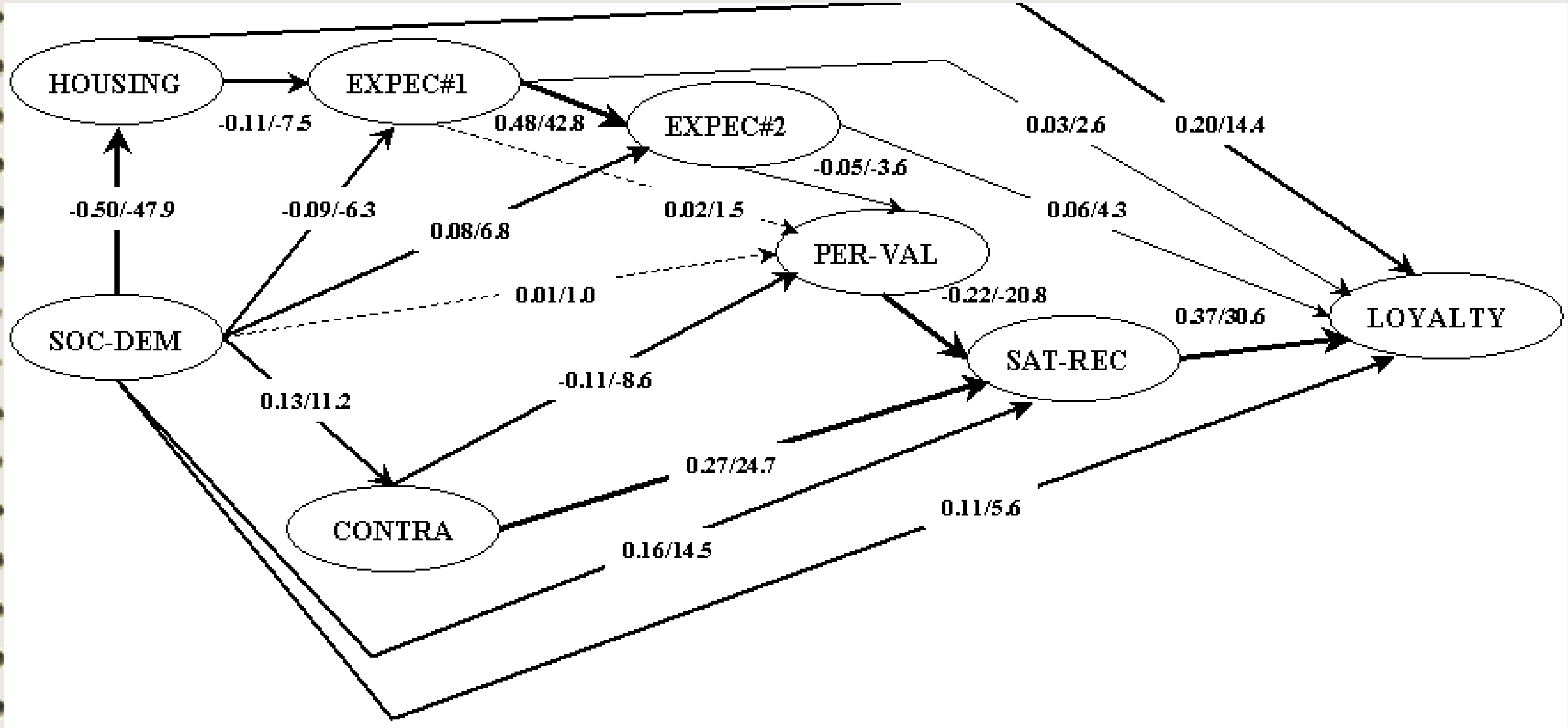
Step 4 : Orientation of links between latent variables

Couple of latent variables	Partial correlation	p-value	Position of links
(EXPEC#1,EXPEC#2)	+0.6311	< 0.0001	EXPEC#1→EXPEC#2
(LOYALTY,SAT-REC)	+0.5024	< 0.0001	SAT-REC→LOYALTY
(SOC-DEM,HOUSING)	+0.4004	< 0.0001	SOC-DEM→HOUSING
(SAT-REC,PER-VAL)	-0.2560	< 0.0001	PER-VAL→SAT-REC
(SAT-REC,HOUSING)	+0.1555	< 0.0001	HOUSING→SAT-REC
(EXPEC#1,SOC-DEM)	-0.1449	< 0.0001	SOC-DEM→EXPEC#1
(LOYALTY,HOUSING)	-0.1290	< 0.0001	HOUSING→LOYALTY
(SOC-DEM,LOYALTY)	+0.1217	< 0.0001	SOC-DEM→LOYALTY
(CONTRA,SAT-REC)	+0.1022	< 0.0001	CONTRA→SAT-REC
(EXPEC#2,SOC-DEM)	+0.1005	< 0.0001	SOC-DEM→EXPEC#2
(SOC-DEM,PER-VAL)	+0.0893	< 0.0001	SOC-DEM→OPINION
(SOC-DEM,SAT-REC)	+0.0828	< 0.0001	SOC-DEM→SAT-REC
(EXPEC#1,LOYALTY)	-0.0735	< 0.0001	EXPEC#1→LOYALTY
(EXPEC#1,HOUSING)	+0.0697	< 0.0001	EXPEC#1→HOU-EDF
(EXPEC#2,OPINION)	+0.0649	< 0.0001	EXPEC#2→OPINION
(EXPEC#2,LOYALTY)	+0.0471	0.0001	EXPEC#2→LOYALTY
(CONTRA,SOC-DEM)	+0.0469	0.0001	SOC-DEM→CONTRA
(CONTRA,PER-VAL)	-0.0394	0.0011	CONTRA→PER-VAL
(EXPEC#1,PER-VAL)	-0.0353	0.0034	EXPEC#1→PER-VAL

Application to marketing data

Results on the Generalized Free Model (3)

Step 5 : Estimation of the generalized free model by RFGPC



Application to marketing data

Results on the Generalized Free Model (4)

Evaluation of the Structural Model

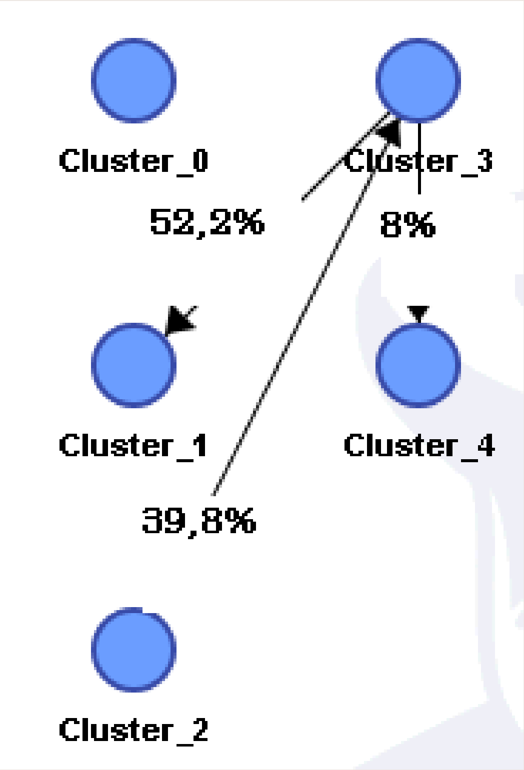
Latent Variables	R2	Adjusted R2	Communalities	Redundancies
EXPEC#1	0.0105	0.0102	0.5915	0.0062
EXPEC#2	0.2266	0.2264	0.5158	0.1169
CONTRA	0.0177	0.0176	0.7005	0.0124
LOYALTY	0.1506	0.1500	0.5632	0.0848
HOUSING	0.2495	0.2494	0.4958	0.1237
PER-VAL	0.0132	0.0126	0.6141	0.0081
SAT-REC	0.1773	0.1770	0.5010	0.0888
SOC_DEM	.	.	0.4820	.

GoF = 0.2621

Application to marketing data

Results on the Bayesian Network

Cluster_0 = {EXPEC#1 ; EXPEC#2}
Cluster_1 = {HOUSING ; SOC_DEM}
Cluster_2 = {PER-VAL}
Cluster_3 = {LOYALTY ; SAT-REC ; (-X_j) }
Cluster_4 = {CONTRA ; (+ X_j)}



Outline

Concluding remarks, applications and perspectives

The proposed method allows to build a **free model with variables measured on mixed scales** and can easily be applied in marketing, chemiometrics, policy, human capital, …

Other methods for creating models exist: **stepwise structural equation modelling** [Hui, 1982] or **bayesian networks** [Jensen, 2001]

Remaining works :

→ To compare with the other methods

→ Future researches concern the improvement of the clustering step

Bibliography

Amato S., Esposito Vinzi V. and Tenenhaus M., (2004), A Global-of-Fit Index for PLS Structural Equation Modelling, *Oral Communication to PLS Club*, HEC School of Management, France.

Derquenne Ch. et Hallais C., (2003), Une méthode alternative à l'approche PLS : comparaison et application aux modèles conceptuels marketing, *Revue de Statistique Appliquée*, **LII**, 37-72.

Derquenne Ch., (2005), Generalized Path Modelling based on the Partial Maximum Likelihood Approach, *PLS'05, Fourth International Symposium on PLS and Related Methods*, Barcelona – Spain.

Derquenne Ch., (2007), Free Model for Generalized Path Modelling, *PLS'07, Fifth International Symposium on PLS and Related Methods*, Barcelona – Norge.

Derquenne Ch., (2008), Free Model for Generalized Path Modelling and Comparison with Bayesian Networks, *Compstat'2008, International Conference on Computational Statistics*, Springer-Verlag, Porto – Portugal.

Bibliography

Hui B.S., (1982), On Building Partial Least Squares Models with Interdependent Inner Relations, pp. 249-271 in K.G. Jöreskog & H. Wold (eds.), *Systems under Indirect Observations: Causality, Structure, Prediction*, Vol 2, Amsterdam, North-Holland.

Jensen F.V., (2001), *Bayesian Networks and Decision Graphs*, Springer, New-York.

Jouffe L. (2004), Bayesialab: the decision support and data mining tools, *ISBA Bulletin*, Vol 11(4), 7-10.

Jöreskog K.G. and Sörbum D. (1979), *Advance in Factor Analysis and Structural Equation Models*, Abt Books, Cambridge.

Wold H., (1973), Nonlinear Iterative Partial Least Squares (NIPALS) Modelling some Current Developments, in Krishnaiah P.R. Editor, *Multivariate Analysis*, III, 391-420, Academic Press, New-York.

Wold, H. (1982): Soft modelling: the basic design and some extensions, In: K.G. Jöreskog & H. Wold (eds.), *Systems under Indirect Observations: Causality, Structure, Prediction,* Vol 2, 1-54, Amsterdam, North-Holland.