

Equations structurelles avec PLS-SEM

Véronique Cariou

Statistics, Sensometrics and Chemometrics Unit
Oniris, INRAE, Nantes, France

Outline

1 Syllabus

2 Composite-based

3 Composite models

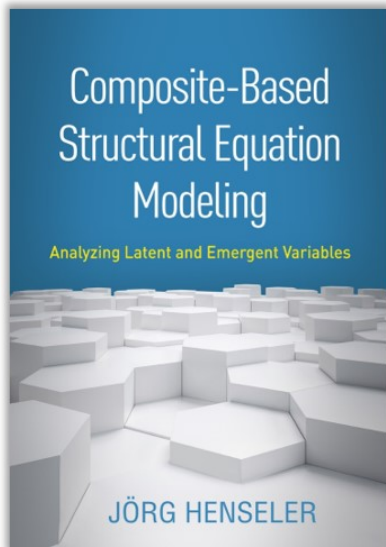
4 Estimation

5 Application

Plan du cours

- CH I : Composite-based methods
 - Définition des méthodes dites variance-based
 - Comment distinguer les deux approches ?
 - PLS-SEM et l'approche de Wold
- CH II : Composite models
 - Quel construit cherche t'on à modéliser ?
 - Variable émergente
- CH III : Estimation du modèle
 - PLS-SEM pour les modèles composites
 - PLSc pour les modèles common-factor
- CH IV : Application
 - RedLosses

Ressource



Ressource : Henseler, J. (2020). Composite-based structural equation modeling : Analyzing latent and emergent variables. Guilford Publications.

Outline

- 1 Syllabus
- 2 Composite-based**
- 3 Composite models
- 4 Estimation
- 5 Application

Méthodes SEM

SEM comme une collection de méthodes statistiques *"that allow a set of relationships between one or more independent variables (IVs), either continuous or discrete, and one or more dependent variables (DVs), either continuous or discrete, to be examined"* Ullman & Bentler (2003)

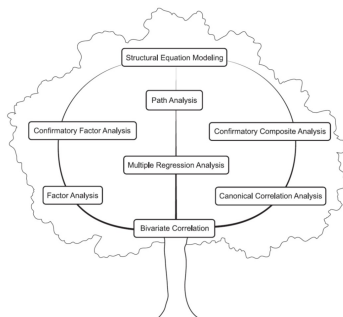


Figure – Famille des méthodes SEM Henseler (2020)

Théorie sous-jacente

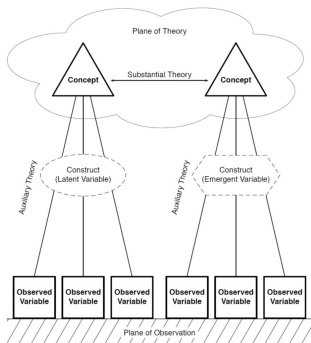


Figure – Construits *Henseler (2020)*

Construits : variables statistiques non directement observables mais pouvant être déduites des variables observables.

- Variables latentes : modélisées comme des facteurs communs sous-jacents à un ensemble de variables observées.
- Variables émergentes : modélisées comme des composites de variables observées.

Méthodes composite-based

"Composite-based SEM are those SEM techniques that involve composites in the estimation phase." (Henseler, 2020, p. 11)

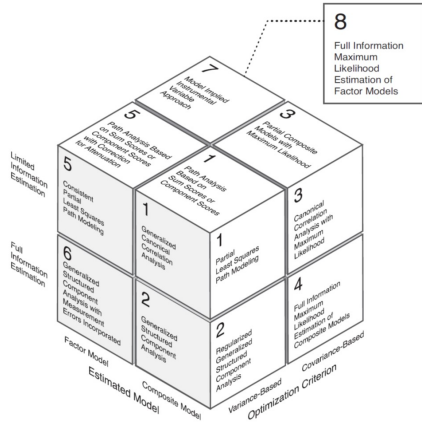


Figure – Typologie des méthodes Henseler (2020)

Méthodes composite-based vs variance-based

Méthodes variance-based (composite-based) vs covariance-based (Jöreskog, 1978 ; Rigdon, 1998) :

- covariance-based : minimise l'écart entre la matrice de variance-covariance empirique et celle impliquée par le modèle
- variance-based : maximise un certain critère d'interdépendance entre les proxies représentant les variables non observées et s'exprimant sous la forme de combinaisons linéaires des variables observées

PLS-SEM

CSA



K. Jöreskog (1973)

GSCA

H. Hwang & Y. Takane
(2004, 2014)

PLS

H. Wold (1966, 1975)
S. Wold & al. (1984)

- En 1975, H. Wold étend l'algorithme itératif NIPALS à une procédure plus générale pour l'estimation de relations entre plusieurs blocs de variables.
- Cette "soft" modélisation évite les hypothèses restrictives qui sous-tendent les techniques du maximum de vraisemblance.

Débats autour de PLS-SEM et développements

Des critiques autour de :

- incohérence des paramètres dans le cas des modèles réflectifs
- manque de mesures d'adéquation
- limite pour la validité discriminante (Rönkkö & Evermann, 2013).

Et des développements contribuant à un enrichissement mutuel :

- clarification des théories auxiliaires
- correction pour l'atténuation (PLS cohérent ; voir Dijkstra & Henseler, 2015)
- tests basés sur le bootstrap de ajustement du modèle (voir Dijkstra & Henseler, 2015)

Package cSEM et SEMinR

- cSEM package (Rademaker) : *M. E., Schuberth, F., Schamberger, T., Klesel, M., Dijkstra, T. K., & Henseler, J. (2020). R package cSEM : Composite-based structural equation modeling version 0.3.0.*
- SEMinR package (Ray & Danks) ; *Hair, J. F., Hult, G. T., Ringle, C. M., Sarstedt, M., Danks, N. P., & Ray, S. (2021). Partial least squares structural equation modeling (PLS-SEM) using R : A workbook. Springer Nature.*

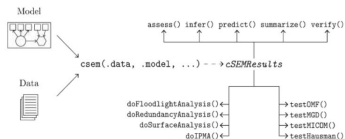


Figure 1: The basic usage of cSEM

Source: <https://m-e-rademaker.github.io/cSEM/>

Chuah, F., Memon, M. A., Ramayah, T., Cheah, J. H., Ting, H., & Cham, T. H. (2021). PLS-SEM using R : An introduction to csem and SEMinR. *Journal of Applied Structural Equation Modeling*, 2(5)

Outline

- 1 Syllabus
- 2 Composite-based
- 3 Composite models**
- 4 Estimation
- 5 Application

Modéliser un construit/concept

Dans le cadre des modèles common-factors, les étapes sont :

- 1 une variable latente associée au concept théorique (réduction de la dimensionnalité),
- 2 le facteur représentant le concept théorique se comporte-t-il comme une variable latente (indépendance conditionnelle) ?
- 3 la variable latente dérivée empiriquement couvre-t-elle suffisamment la signification du concept théorique ?

Modéliser un construit/concept

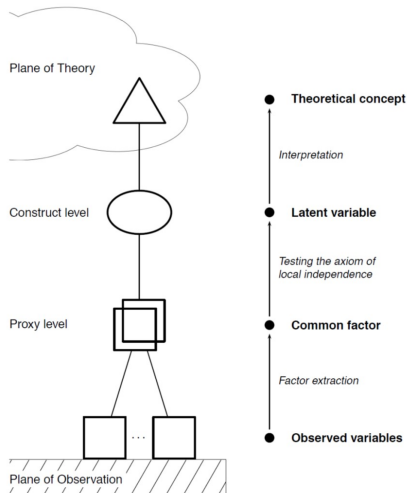
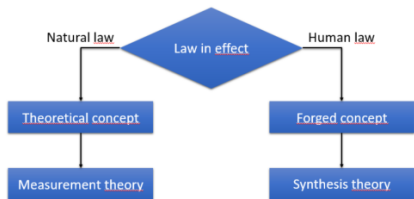


Figure – Théorie de la mesure - Henseler (2020)

Quel type de construit/concept ?



Science of the existing

Human biology
Physics
Psychology
Service research
Human-media interaction
Economics
Consumer behavior
Acoustics
Linguistics
History of arts

Science of the artificial

Medicine
Engineering
Psychotherapy
Service design
Interaction design
Economic policy
Advertising
Sound design/music
Computer linguistics
Arts

Figure – Théorie de la mesure vs synthèse

Un concept émergent nécessite de le développer et de l'évaluer et non de le mesurer.

Modéliser un concept émergent

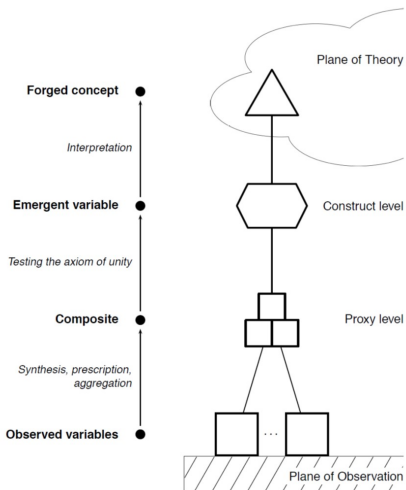


Figure – Variable émergente - Henseler (2020)

Variable émergente et modèle composite

Dans le cas des modèles composites, les indicateurs forment le construit comme combinaison linéaire (avec la prise en compte de poids) sans erreur (Dijkstra, 2017 ; Henseler, 2020). Les relations entre poids et composites sont les suivantes :

$$\xi = \mathbf{XW}_X$$

$$\eta = \mathbf{YW}_Y$$

avec :

$$\mathbf{W}_X = \begin{pmatrix} \mathbf{w}_1^{(X)} & 0 & \dots & 0 \\ 0 & \mathbf{w}_2^{(X)} & & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \mathbf{w}_p^{(X)} \end{pmatrix}, \quad \mathbf{W}_Y = \begin{pmatrix} \mathbf{w}_1^{(Y)} & 0 & \dots & 0 \\ 0 & \mathbf{w}_2^{(Y)} & & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \mathbf{w}_q^{(Y)} \end{pmatrix}.$$

Modèle composite

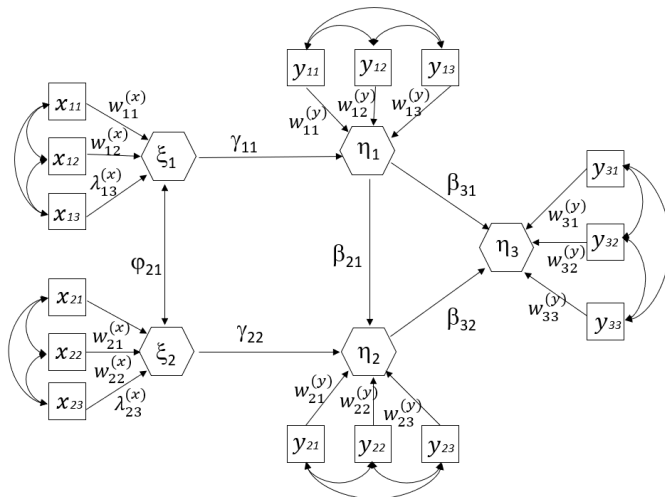


Figure – Exemple de modèle composite

Éléments de SEM

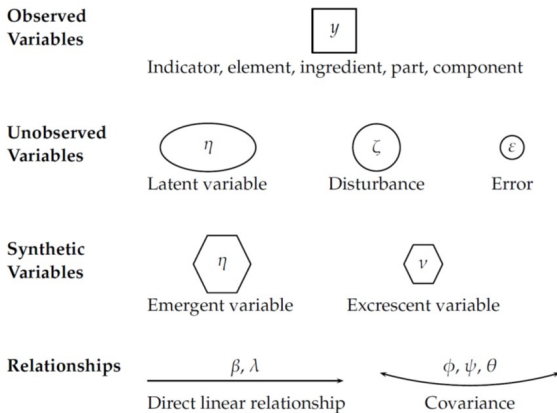


Figure – Variables et paramètres de SEM - Henseler (2020)

Spécificités des trois types de modèle

Characteristic	Component	Effect indicator	Cause indicator
Observed variable's role	Ingredient, part, element	Consequence	Antecedent, cause
Corresponding construct	Emergent variable	Latent variable	Latent variable
Correlations among observed variables	High correlations are common, but not required	High correlations are expected	No reason to expect the measures are correlated
Proneness to measurement error	Can contain measurement error	Contains measurement error	Can contain measurement error
Informative about measurement error	Not informative about measurement error	Jointly informative about measurement error	Not informative about measurement error
Consequences of dropping an indicator	Dropping an indicator alters the construct and may change its meaning	Dropping an indicator does not alter the meaning of the construct	Dropping an indicator increases the error on construct level

Figure – Roles des variables observées - Henseler (2020)

Modèle sur la stabilité politique avec 47 pays et 3 blocs de variables - Russett (1964)

- Indicateurs des inégalités agricoles :
 - gini : indice de concentration Gini
 - farm : pourcentage de propriétaires fonciers qui occupent collectivement la moitié de toutes les terres agricoles
 - rent : pourcentage du nombre total d'exploitations agricoles qui louent toutes leur terre $\ln(x + 1)$
- Indicateurs de développement industriel :
 - gnpr : produit national brut par habitant de 1955 en \$ $\ln(x)$
 - labo : pourcentage de la population active dans l'agriculture $\ln(x)$
- Indicateurs de stabilité politique :
 - inst : instabilité du personnel selon la durée du mandat $\exp(x - 16,3)$
 - ecks : nombre total d'incidents politiques violents $\ln(x + 1)$
 - deat : nombre de personnes tuées à la suite d'attaques internes pour 1 000 000 d'habitants, $\ln(x + 1)$
 - stab : 1 si le pays a une démocratie stable, 0 sinon
 - dict : 1 si dictature, 0 sinon. sinon.

Script R

```
#devtools::install_github("M-E-Rademaker/cSEM")
library(readxl)
library(cSEM) # autres packages SEMinR, cbsem, plspm
setwd(setwd("~/Publication/2024/Chimionométrie/Cours/R"))

Russett <- as.data.frame(readxl::read_excel("Russett.xlsx"))

# Model specification
# =~ specify reflective measurement models; it assigns observed variables to a latent variable.
# <~ specify composite models; it assigns observed variables to an emergent variable.
# ~ specify the inner model; it defines on which independent variables a construct shall be regressed.
# ~ specify correlations among measurement errors within a block of indicators.

model_Russett = ' # Specify the composite models
AgrIneq <~ gini + farm + rent
IndDev <~ gnpr + labo
PolInst <~ inst + ecks + deat + stab + dict
# Specify the relation among the emergent variables
PolInst ~ AgrIneq + IndDev
'

out <- csem(.data = Russett, .model = model_Russett,
            .PLS_weight_scheme_inner = 'factorial', #see estimation
            .tolerance = 1e-06
)

summarize(out)
```

Outputs (1)

----- Overview -----

General information:

```
-----  
Estimation status           = ok  
Number of observations      = 47  
weight estimator           = PLS-PM  
Inner weighting scheme     = "factorial"  
Type of indicator correlation = Pearson  
Path model estimator       = OLS  
Second-order approach      = NA  
Type of path model         = Linear  
Disattenuated              = No
```

Construct details:

```
-----  
Name      Modeled as  Order      Mode  
AgrIneq  Composite    First order "modeB"  
IndDev   Composite    First order "modeB"  
PolInst  Composite    First order "modeB"
```

----- Estimates -----

Outputs (2)

```

----- Estimates -----
Estimated path coefficients:
=====
Path              Estimate  Std. error  t-stat.  p-value
PoInst ~ AgrIneq  0.3379      NA          NA       NA
PoInst ~ IndDev   0.5926      NA          NA       NA
=====
Estimated loadings:
=====
Loading           Estimate  Std. error  t-stat.  p-value
AgrIneq =~ gini   0.5365      NA          NA       NA
AgrIneq =~ farm   0.6715      NA          NA       NA
AgrIneq =~ rent  -0.2644      NA          NA       NA
IndDev =~ gnpr    -0.9094      NA          NA       NA
IndDev =~ labo    0.9824      NA          NA       NA
PoInst =~ inst    0.1923      NA          NA       NA
PoInst =~ ecks    0.6310      NA          NA       NA
PoInst =~ deat    0.5240      NA          NA       NA
PoInst =~ stab   -0.9685      NA          NA       NA
PoInst =~ dict    0.7395      NA          NA       NA
=====
Estimated weights:
=====
weight           Estimate  Std. error  t-stat.  p-value
AgrIneq <~ gini  -1.0570      NA          NA       NA
AgrIneq <~ farm  2.0241      NA          NA       NA
AgrIneq <~ rent  -0.7859      NA          NA       NA
IndDev <~ gnpr   -0.3228      NA          NA       NA
IndDev <~ labo   0.7191      NA          NA       NA
PoInst <~ inst   -0.1337      NA          NA       NA
PoInst <~ ecks   0.1287      NA          NA       NA
PoInst <~ deat   -0.0854      NA          NA       NA
PoInst <~ stab   -0.8337      NA          NA       NA
PoInst <~ dict   0.2459      NA          NA       NA

```

Outputs (3)

Estimated construct correlations:

Correlation	Estimate	Std. error	t-stat.	p-value
AgrIneq ~ IndDev	0.4875	NA	NA	NA

Estimated indicator correlations:

Correlation	Estimate	Std. error	t-stat.	p-value
gini ~ farm	0.9376	NA	NA	NA
gini ~ rent	0.3873	NA	NA	NA
farm ~ rent	0.4599	NA	NA	NA
gnpr ~ labo	-0.8156	NA	NA	NA
inst ~ ecks	0.3261	NA	NA	NA
inst ~ deat	0.0835	NA	NA	NA
inst ~ stab	-0.3434	NA	NA	NA
inst ~ dict	0.0198	NA	NA	NA
ecks ~ deat	0.6277	NA	NA	NA
ecks ~ stab	-0.6034	NA	NA	NA
ecks ~ dict	0.3920	NA	NA	NA
deat ~ stab	-0.4905	NA	NA	NA
deat ~ dict	0.5321	NA	NA	NA
stab ~ dict	-0.5893	NA	NA	NA

----- Effects -----

Estimated total effects:

Total effect	Estimate	Std. error	t-stat.	p-value
PolInst ~ AgrIneq	0.3379	NA	NA	NA
PolInst ~ IndDev	0.5926	NA	NA	NA

Outline

- 1 Syllabus
- 2 Composite-based
- 3 Composite models
- 4 Estimation**
- 5 Application

PLS-SEM pour modèles composites

Le critère d'optimisation de PLS peut être compris comme une somme pondérée des R^2 du modèle d'équations structurelles.

Les principaux résultats de l'algorithme PLS-SEM sont les proxies z_j , c'est-à-dire les scores obtenus comme combinaisons linéaires des variables observées des blocs respectifs.

La matrice de corrélation impliquée par le modèle interne est alors assimilée à la matrice des corrélations entre les proxies.

L'algorithme PLS peut être vu comme une séquence de régressions en termes de vecteurs de poids (Lohmöller, 1989).

Algorithme PLS-SEM en 4 étapes principales

1 Calcul des outer weights

les poids \mathbf{w}_p sont obtenus par un algorithme itératif alternant le calcul des proxies selon les modèles interne et externe

2 calcul des composites

les composites sont estimées sur la base des proxies

$$\hat{\xi}_p \propto \mathbf{X}_p \mathbf{w}_p$$

$$\hat{\eta}_q \propto \mathbf{Y}_q \mathbf{W}_q$$

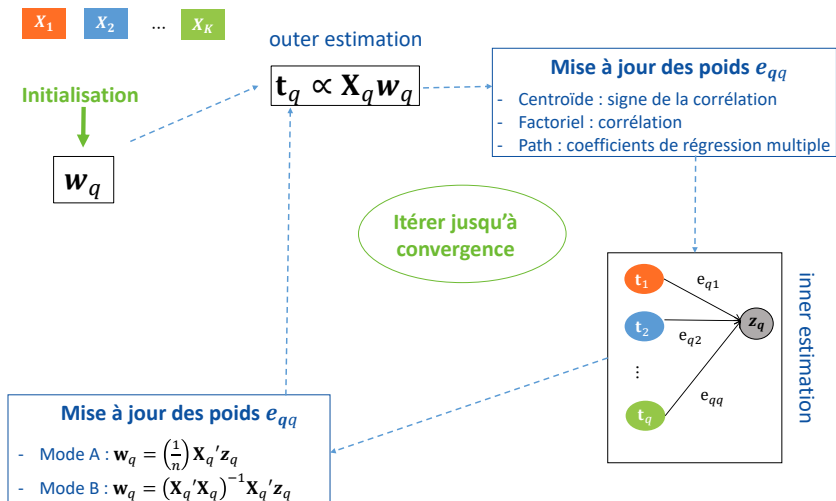
3 estimation des path coefficients $\mathbf{\Gamma}$ et \mathbf{B}

une séquence de régressions est opérée de manière à estimer les coefficients de régression associés au modèle structurel

4 estimation des loadings

Les loadings sont estimés sur la base des proxies et des poids

Algorithme PLS-SEM



Les variables observées sont supposées standardisées

Etape finale de calcul

- Pour chaque variable observée, les loadings peuvent être déterminés comme des corrélations entre elle et le proxy qui lui est associé :

$$\forall j, q \lambda_{qj} = \text{cor}(\mathbf{z}_q, \mathbf{y}_{qj})$$

- Le vecteur de poids \mathbf{w}_q peut être déterminé sur la base de la matrice de corrélation \mathbf{R}_q des variables observées \mathbf{y}_{qj} et des loadings :

$$\forall j, q \mathbf{w}_{qj} = \mathbf{R}_q^{-1} \lambda_{qj}$$

PLS-SEM pour l'estimation des paramètres d'un modèle common-factor

Correction proposée par Dijkstra dans PLSc (1981) : la corrélation entre deux variables latentes est égale à la corrélation entre leurs scores de construit divisée par la moyenne géométrique de la fiabilité de leurs scores :

$$\text{cor}(\eta_p, \eta_q) = \rho_{pq} = \frac{\text{cor}(\mathbf{z}_p, \mathbf{z}_q)}{\sqrt{\rho_A(\mathbf{z}_p) \times \rho_A(\mathbf{z}_q)}}$$

avec :

$$\rho_A(\mathbf{z}_p) = (\mathbf{w}_p^\top \mathbf{w}_p)^2 \sqrt{\lambda_p^\top \Sigma_{pp} \lambda_p}$$

Modèle sur la stabilité politique avec 47 pays et 3 blocs de variables - Russett (1964)

- Indicateurs des inégalités agricoles :
 - gini : indice de concentration Gini
 - farm : pourcentage de propriétaires fonciers qui occupent collectivement la moitié de toutes les terres agricoles
 - rent : pourcentage du nombre total d'exploitations agricoles qui louent toutes leur terre $\ln(x + 1)$
- Indicateurs de développement industriel :
 - gnpr : produit national brut par habitant de 1955 en \$ $\ln(x)$
 - labo : pourcentage de la population active dans l'agriculture $\ln(x)$
- Indicateurs de stabilité politique :
 - inst : instabilité du personnel selon la durée du mandat $\exp(x - 16,3)$
 - ecks : nombre total d'incidents politiques violents $\ln(x + 1)$
 - deat : nombre de personnes tuées à la suite d'attaques internes pour 1 000 000 d'habitants, $\ln(x + 1)$
 - stab : 1 si le pays a une démocratie stable, 0 sinon
 - dict : 1 si dictature, 0 sinon. sinon.

- les modèles de mesure réfléchifs des variables latentes sont estimés par PLS_c (PLS-SEM mode A puis correction)
- les modèles composites sont estimés par le mode B
- plus généralement, le mode appliqué peut être ajusté avec cinq options pour déterminer les poids : Mode A, Mode B, Mode BNNLS, somme des scores/poids unitaires et poids définis par l'utilisateur.

cSEM sur Russett

- les indicateurs d'inégalités agricoles doivent avoir des poids non négatifs
- les indicateurs de développement industriel sont contraints d'avoir des pondérations égales en valeur absolue mais différentes en signe
- le mode A est appliqué pour l'instabilité politique
- les proxies sont liées sans prise en compte du path diagram
- la qualité du modèle est estimée par bootstrap

```
out <- csem(.data = Russett, .model = model_Russett,  
  .PLS_weight_scheme_inner = 'path',  
  .PLS_ignore_structural_model = TRUE,  
  .iter_max = 100,  
  .tolerance = 1e-06,  
  .PLS_modes = list(AgrIneq = 'modeBNNLS',  
                    IndDev = c(1,-1),  
                    PolInst = 'modeA'),  
  .resample_method='bootstrap'  
)  
summarize(out)
```

Outputs (1)

----- overview -----

General information:

```

-----
Estimation status           = Ok
Number of observations      = 47
weight estimator           = PLS-PM
Inner weighting scheme     = "path"
Type of indicator correlation = Pearson
Path model estimator       = OLS
Second-order approach      = NA
Type of path model         = Linear
Disattenuated              = No
  
```

Resample information:

```

-----
Resample method            = "bootstrap"
Number of resamples        = 499
Number of admissible results = 499
Approach to handle inadmissibles = "drop"
Sign change option         = "none"
Random seed                 = 774016719
  
```

Construct details:

```

-----
Name      Modeled as  Order      Mode
-----
AgrIneq   Composite      First order "modeBNNLS"
IndDev    Composite      First order "fixed"
PolInst   Composite      First order "modeA"
  
```

----- Estimates -----

Estimated path coefficients:

```

-----
Path              Estimate  Std. error  t-stat.  p-value  CI_percentile
PolInst ~ AgrIneq  0.2064     0.1067     1.9355   0.0529   [-0.0096; 0.4355 ]
PolInst ~ IndDev   -0.6960    0.1014    -6.8648   0.0000   [-0.8520; -0.5330 ]
  
```

Outputs (2)

Estimated loadings:

=====

Loading	Estimate	Std. error	t-stat.	p-value	CI_percentile 95%
AgrIneq =~ gini	0.9376	0.0920	10.1899	0.0000	[0.8402; 1.0000]
AgrIneq =~ farm	1.0000	0.0857	11.6674	0.0000	[0.8325; 1.0000]
AgrIneq =~ rent	0.4599	0.1758	2.6156	0.0089	[0.0859; 0.7447]
IndDev =~ gnpr	0.9528	0.0107	89.0989	0.0000	[0.9305; 0.9708]
IndDev =~ labo	-0.9528	0.0107	-89.0989	0.0000	[-0.9708; -0.9305]
PolInst =~ inst	0.3341	0.1546	2.1607	0.0307	[-0.0041; 0.5846]
PolInst =~ ecks	0.8034	0.0961	8.3573	0.0000	[0.6383; 0.8973]
PolInst =~ deat	0.7860	0.0835	9.4173	0.0000	[0.6891; 0.8728]
PolInst =~ stab	-0.8591	0.0897	-9.5749	0.0000	[-0.9234; -0.7313]
PolInst =~ dict	0.7680	0.0856	8.9685	0.0000	[0.6446; 0.8670]

Estimated weights:

=====

weight	Estimate	Std. error	t-stat.	p-value	CI_percentile 95%
AgrIneq <= gini	0.0000	0.4869	0.0000	1.0000	[0.0000; 1.0000]
AgrIneq <= farm	1.0000	0.4930	2.0283	0.0425	[0.0000; 1.0000]
AgrIneq <= rent	0.0000	0.1640	0.0000	1.0000	[0.0000; 0.4895]
IndDev <= gnpr	0.5248	0.0059	88.5890	0.0000	[0.5150; 0.5374]
IndDev <= labo	-0.5248	0.0059	-88.5890	0.0000	[-0.5374; -0.5150]
PolInst <= inst	0.0989	0.0627	1.5789	0.1144	[-0.0398; 0.2062]
PolInst <= ecks	0.2690	0.0461	5.8289	0.0000	[0.1874; 0.3489]
PolInst <= deat	0.2739	0.0426	6.4352	0.0000	[0.2131; 0.3533]
PolInst <= stab	-0.3448	0.0490	-7.0419	0.0000	[-0.4123; -0.2713]
PolInst <= dict	0.3116	0.0431	7.2263	0.0000	[0.2492; 0.3774]

Outputs (3)

Estimated construct correlations:

Correlation	Estimate	Std. error	t-stat.	p-value	CI_percentile 95%
AgrIneq ~ IndDev	-0.3544	0.1460	-2.4278	0.0152	[-0.5879; -0.0262]

Estimated indicator correlations:

Correlation	Estimate	Std. error	t-stat.	p-value	CI_percentile 95%
gini ~ farm	0.9376	0.0160	58.6652	0.0000	[0.9001; 0.9604]
gini ~ rent	0.3873	0.1570	2.4670	0.0136	[0.0280; 0.6192]
farm ~ rent	0.4599	0.1601	2.8719	0.0041	[0.0924; 0.7059]
gnpr ~ labo	-0.8156	0.0407	-20.0644	0.0000	[-0.8850; -0.7315]
inst ~ ecks	0.3261	0.0984	3.3133	0.0009	[0.1029; 0.5000]
inst ~ deat	0.0835	0.1423	0.5868	0.5573	[-0.1782; 0.3884]
inst ~ stab	-0.3434	0.0952	-3.6070	0.0003	[-0.5148; -0.1479]
inst ~ dict	0.0198	0.1459	0.1354	0.8923	[-0.2723; 0.3021]
ecks ~ deat	0.6277	0.0629	9.9874	0.0000	[0.5033; 0.7461]
ecks ~ stab	-0.6034	0.1133	-5.3271	0.0000	[-0.7859; -0.3454]
ecks ~ dict	0.3920	0.1100	3.5648	0.0004	[0.1662; 0.5911]
deat ~ stab	-0.4905	0.0761	-6.4445	0.0000	[-0.6322; -0.3358]
deat ~ dict	0.5321	0.1093	4.8671	0.0000	[0.3086; 0.7551]
stab ~ dict	-0.5893	0.0846	-6.9654	0.0000	[-0.7657; -0.4310]

Effects

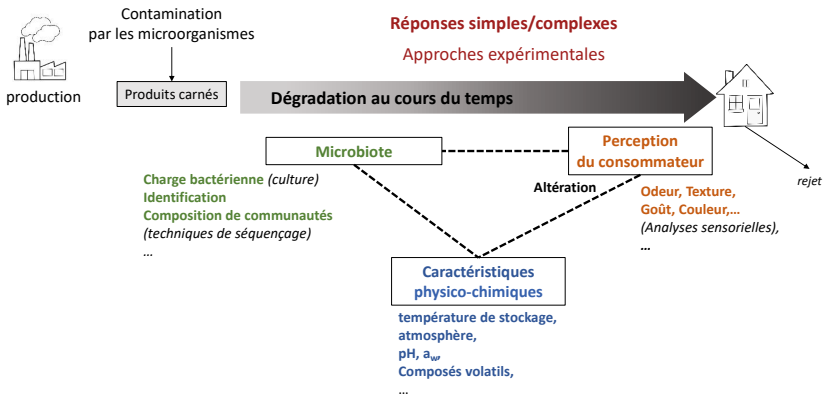
Estimated total effects:

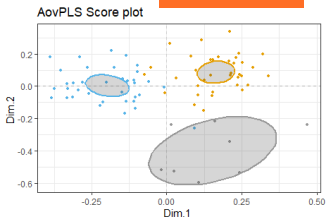
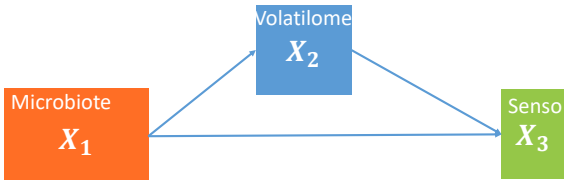
Total effect	Estimate	Std. error	t-stat.	p-value	CI_percentile 95%
PoInst ~ AgrIneq	0.2064	0.1067	1.9355	0.0529	[-0.0096; 0.4355]
PoInst ~ IndDev	-0.6960	0.1014	-6.8648	0.0000	[-0.8520; -0.5330]

Outline

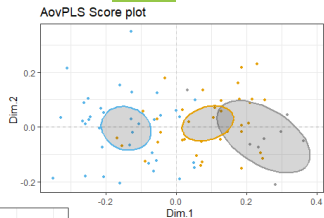
- 1 Syllabus
- 2 Composite-based
- 3 Composite models
- 4 Estimation
- 5 Application**

Suivi des caractéristiques de la viande

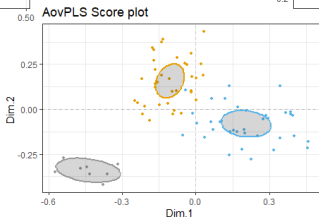




Microbiote



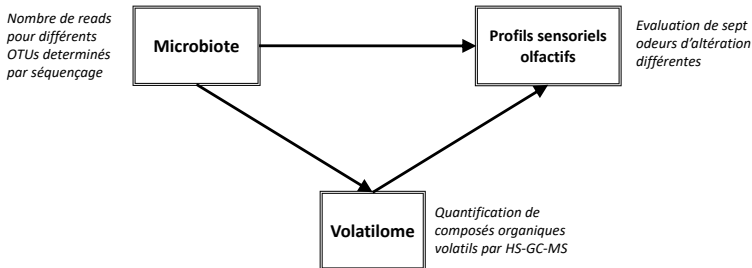
Volatilome



Senso

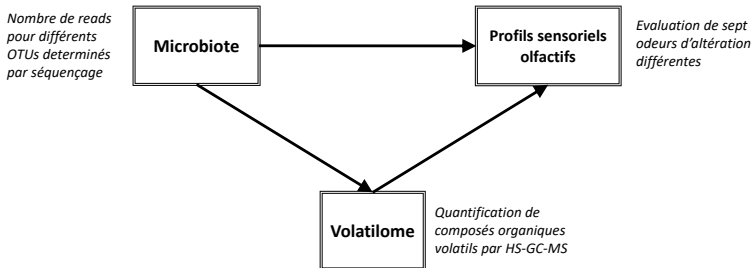
niveau
 ● 2
 ● 8
 ● 15
 Facteur Temps

Intégration de données : données utilisées et hypothèse



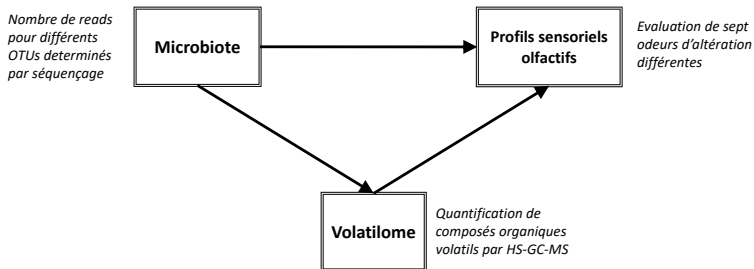
- Données de grande dimension et hétérogènes,
- Liens orientés *a priori* à tester,
- Prétraitements de données nécessaires dans chaque bloc,
- Intégration de données par l'approche Path-ComDim (Cariou et al. 2018)

Intégration de données : données utilisées et hypothèse

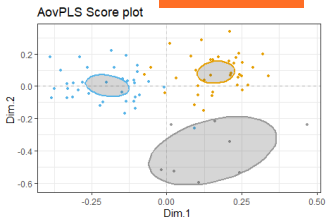
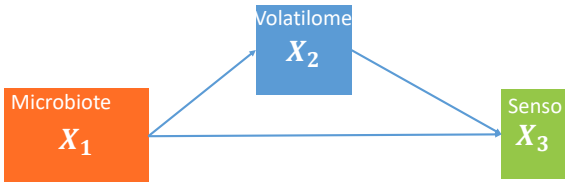


- Données de grande dimension et hétérogènes,
- Liens orientés *a priori* à tester,
- Prétraitements de données nécessaires dans chaque bloc,
- Intégration de données par l'approche Path-ComDim (Cariou et al. 2018)

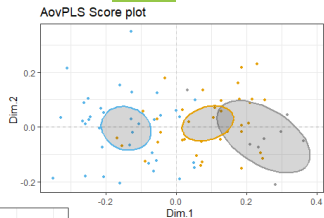
Intégration de données : données utilisées et hypothèse



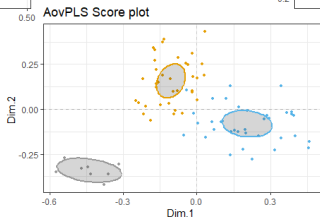
- Données de grande dimension et hétérogènes,
- Liens orientés *a priori* à tester,
- Prétraitements de données nécessaires dans chaque bloc,
- Intégration de données par l'approche Path-ComDim (Cariou et al. 2018)



Microbiote

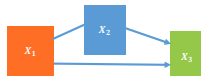


Senso



Volatilome

niveau
 ● 2
 ● 8
 ● 15
 Facteur Temps



```

model_RedLosses = ' # Specify the composite models
Met <- Brochothrix_thermosphacta + Latilactobacillus_curvatus + Latilactobacillus_fuchuensis
      + Paucilactobacillus_oligofermentans + Secundilactobacillus_malefermentans + Psychrobacter_cibarius
      + Psychrobacter_submarinus + Photobacterium_iliopiscarium + Photobacterium_phosphoreum
Volat <- VETHanol + V1_propanol + VAcetoin + VETHyl_acetate + VDIacetyl + V2_butanone
Senso <- Global_oumour_of_rotten_meat + Pungunt_sour + Ethereal_Fermented_fruit + Fermented_meat_0ld_dry_sausage_like_oumour
# Specify the relation among the emergent variables
Volat ~ Met
Senso ~ Met + Volat
'

out <- csem(.data = df, .model = model_RedLosses,
            .PLS_weight_scheme_inner = 'factorial', #see estimation
            .tolerance = 1e-06,
            .resample_method='bootstrap'

)

```

----- Overview -----

General information:

```

-----
Estimation status           = ok
Number of observations      = 80
Weight estimator           = PLS-PM
Inner weighting scheme     = "Factorial"
Type of indicator correlation = Pearson
Path model estimator       = OLS
Second-order approach      = NA
Type of path model         = Linear
Disattenuated              = No

```

Resample information:

```

-----
Resample method            = "bootstrap"
Number of resamples       = 499
Number of admissible results = 465
Approach to handle inadmissibles = "drop"
Sign change option        = "none"
Random seed                = 134659653

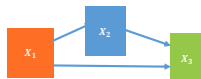
```

Construct details:

```

-----
Name   Modeled as   Order   Mode
Met    Composite    First order  "modeB"
Volat  Composite    First order  "modeB"
Senso  Composite    First order  "modeB"

```



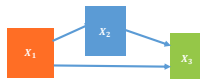
----- Estimates -----

Estimated path coefficients:

Path	Estimate	Std. error	t-stat.	p-value	CI_percentile 95%
volat ~ Met	0.7324	0.2513	2.9141	0.0036	[0.0492; 0.8519]
Senso ~ Met	0.2505	0.1548	1.6181	0.1056	[-0.0917; 0.5313]
Senso ~ Volat	0.5551	0.1166	4.7618	0.0000	[0.3013; 0.7514]

Estimated loadings:

Loading	Estimate	Std. error	t-stat.	p-value	CI_percentile 95%
Met ~ Brochothrix_thermosphacta	-0.1803	0.1842	-0.9792	0.3275	[-0.4903; 0.2311]
Met ~ Latilactobacillus_curvatus	-0.1447	0.1963	-0.7370	0.4611	[-0.4883; 0.2839]
Met ~ Latilactobacillus_fuchuensis	-0.5027	0.2538	-1.9806	0.0476	[-0.7871; 0.2628]
Met ~ Paucilactobacillus_oligofermentans	-0.5831	0.2329	-2.5040	0.0123	[-0.7607; 0.0972]
Met ~ Secundilactobacillus_malefermentans	0.5198	0.1838	2.8277	0.0047	[-0.1092; 0.6700]
Met ~ Psychrobacter_cibarius	0.6774	0.2093	3.2361	0.0012	[0.0093; 0.8066]
Met ~ Psychrobacter_submarinus	0.5431	0.1809	3.0031	0.0027	[-0.0330; 0.7131]
Met ~ Photobacterium_illipiscarium	0.2302	0.1471	1.5645	0.1177	[-0.1062; 0.5029]
Met ~ Photobacterium_phosphoreum	0.2853	0.1891	1.5086	0.1314	[-0.2166; 0.5804]
Volat ~ VETHanol	0.5881	0.1967	2.9901	0.0028	[0.0926; 0.8627]
Volat ~ V1_propanol	0.7809	0.0963	8.1093	0.0000	[0.5356; 0.9040]
Volat ~ VAcetoin	0.3664	0.1887	1.9415	0.0522	[-0.1011; 0.6781]
Volat ~ VETHyl_acetate	0.5417	0.1538	3.5225	0.0004	[0.1884; 0.7689]
Volat ~ VDiacetyl	0.2921	0.1935	1.5089	0.1313	[-0.1794; 0.6192]
Volat ~ V2_butanone	0.7507	0.1116	6.7295	0.0000	[0.4291; 0.8832]
Senso ~ Global_oudeur_of_rotten_meat	0.9847	0.0489	20.1197	0.0000	[0.8164; 0.9973]
Senso ~ Pungunt_Sour	0.8924	0.0898	9.9322	0.0000	[0.6096; 0.9768]
Senso ~ Ethereal_Fermented_fruit	0.8041	0.0723	11.1142	0.0000	[0.6204; 0.9110]
Senso ~ Fermented_meat_Old_dry_sausage_like_oudeur	0.7358	0.1340	5.4911	0.0000	[0.4227; 0.9111]



----- Effects -----

Estimated total effects:

Total effect	Estimate	Std. error	t-stat.	p-value	CI_percentile 95%
Volat ~ Met	0.7324	0.2513	2.9141	0.0036	[0.0492; 0.8519]
Senso ~ Met	0.6570	0.2360	2.7844	0.0054	[0.0455; 0.8183]
Senso ~ Volat	0.5551	0.1166	4.7618	0.0000	[0.3013; 0.7514]

Estimated indirect effects:

Indirect effect	Estimate	Std. error	t-stat.	p-value	CI_percentile 95%
Senso ~ Met	0.4065	0.1627	2.4985	0.0125	[-0.0080; 0.5973]

R2 estimé :
Volat 53%
Senso 57%