Les modèles d'équations structurelles à variables latentes

Cours de Statistique Multivariée Approfondie

giorgio.russolillo@cnam.fr

le cnam

Modèles d'équations structurelles à variables latentes

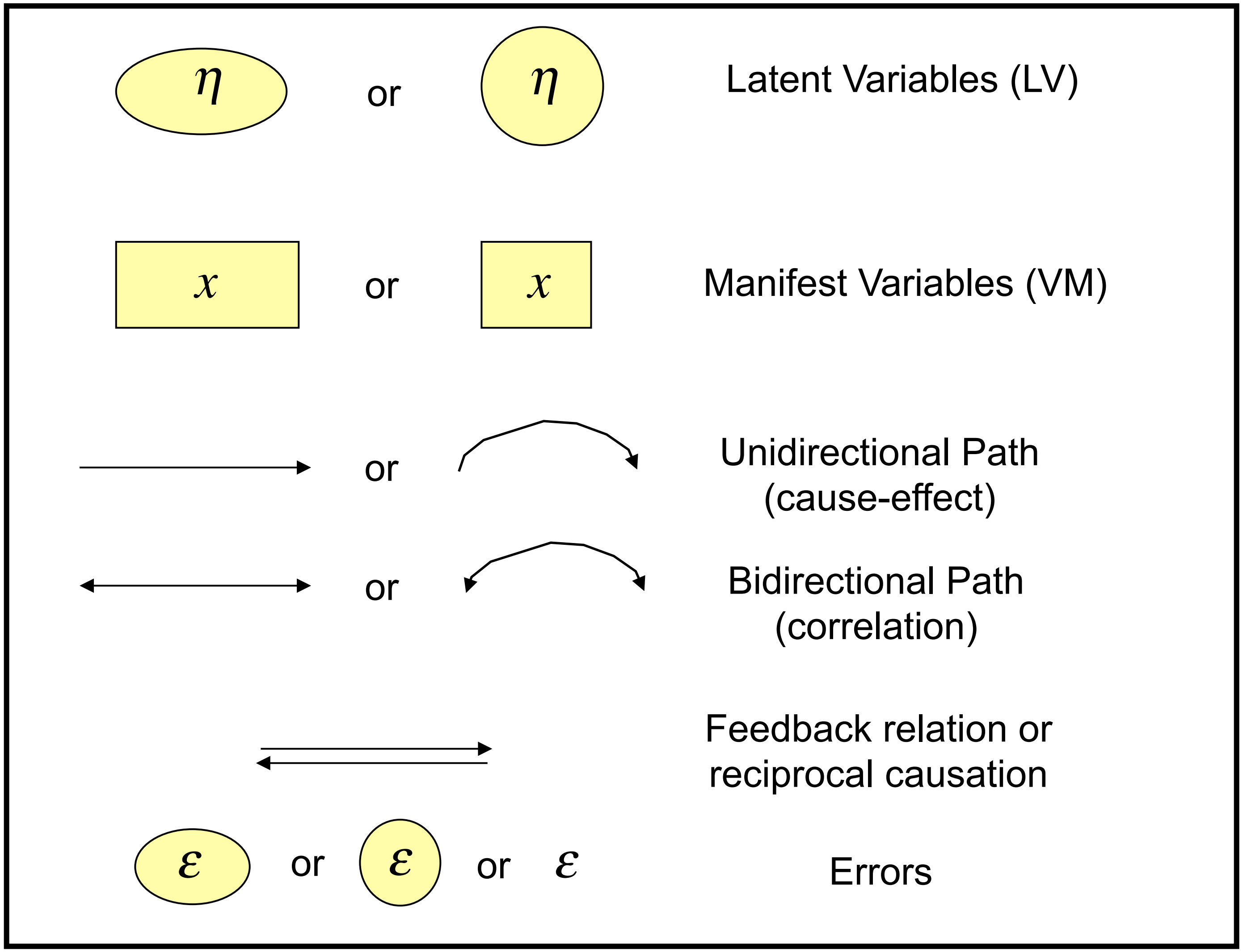
- Les modèles d'équation structurelles ou SEM permettent de quantifier des **relations de cause à effet** décrites par un modèle théorique

- L'intérêt se concentre sur les **variables latentes** (non observables par une voie directe): les variables psychologiques abstraits comme «intelligence», «satisfaction», «statut social», «capacité», «confiance»

- Les variables latentes sont mesurées par des **indicateurs observables** appelés **variables manifestes.**

- Les modéles d'équations tructurelles font le pont entre l'analyse de **régression**, l'analyse utilisant des graphes **(path analysis) et l'analyse factorielle**

SEM: conventions de dessin

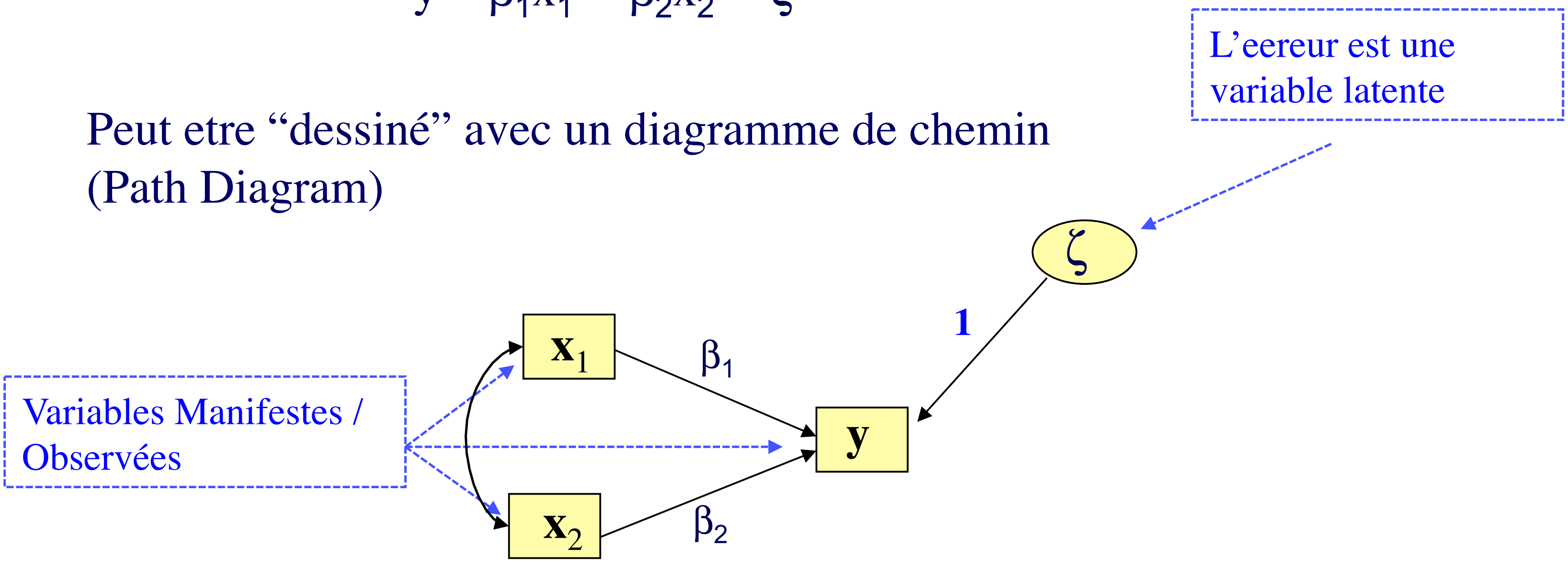


"Dessin" d'un modèle de régression

Ce modèle de régression multiple

$$y = \beta_1 x_1 + \beta_2 x_2 + \zeta$$

Peut etre "dessiné" avec un diagramme de chemin (Path Diagram)

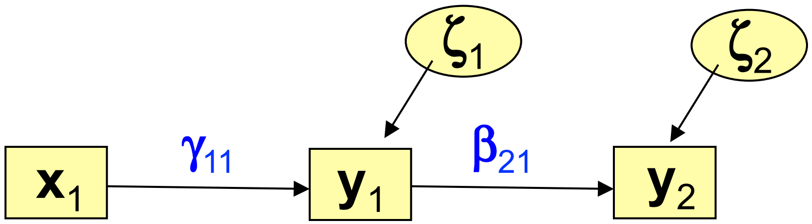


Modéles de chemin avec des variables manifestes

Le modèle de régression multiple peut être généralisée à des chemins où les variables endogènes sont à leur fois causes d'autres variables endogènes.

$$y_1 = \gamma_{11}\, x_1 + \zeta_1$$

$$y_2 = \beta_{21}\, y_1 + \zeta_2$$

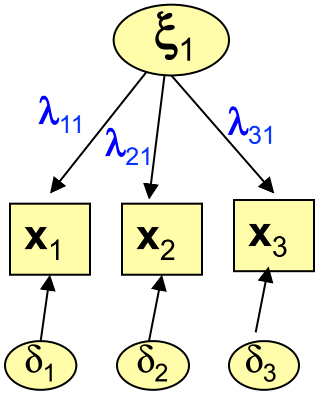


$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \quad \Gamma = \begin{bmatrix} \gamma_{11} \\ 0 \end{bmatrix} \quad \mathbf{x} = [x_1] \quad \mathrm{B} = \begin{bmatrix} 0 & 0 \\ \beta_{21} & 0 \end{bmatrix} \quad \zeta = \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}$$

$$\mathbf{y} = \Gamma\mathbf{x} + \mathbf{B}\mathbf{y} + \zeta$$

Modèle factoriel

$$x_1 = \lambda_{11}\,\xi_1 + \delta_1$$
$$x_2 = \lambda_{21}\,\xi_1 + \delta_2$$
$$x_3 = \lambda_{31}\,\xi_1 + \delta_3$$



$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad \Lambda_x = \begin{bmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \end{bmatrix} \quad \xi = [\xi_1] \quad \delta = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix}$$

$$\mathbf{x} = \Lambda_x \xi + \delta$$

Modèle factoriel multiple

$x_1 = \lambda_{11}\,\xi_1 + \delta_1$,

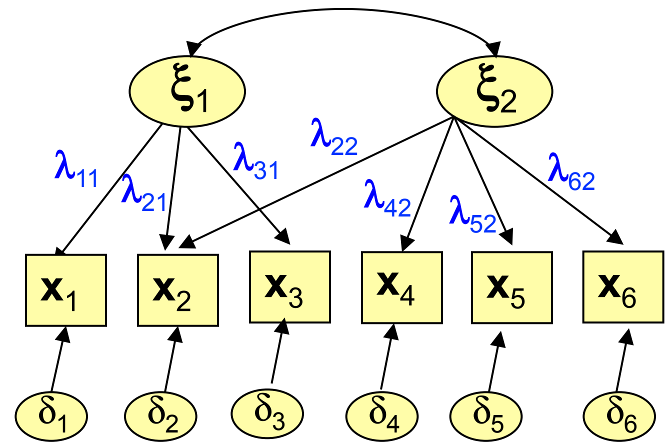
$x_2 = \lambda_{21}\,\xi_1 + \lambda_{22}\,\xi_2 + \delta_2$

$x_3 = \lambda_{31}\,\xi_1 + \delta_3$

$x_4 = \lambda_{42}\,\xi_2 + \delta_4$

$x_5 = \lambda_{52}\,\xi_2 + \delta_5$

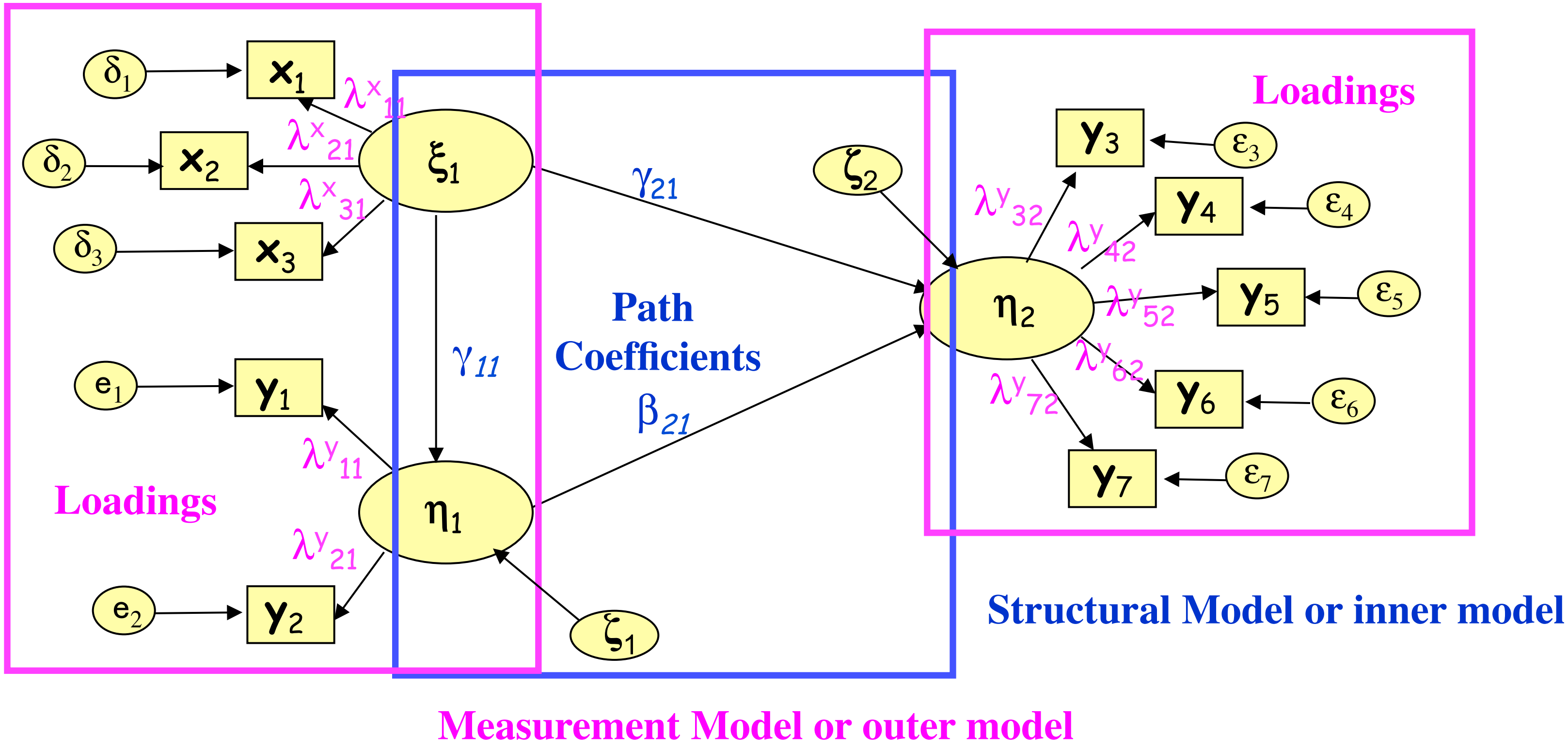
$x_6 = \lambda_{62}\,\xi_2 + \delta_6$



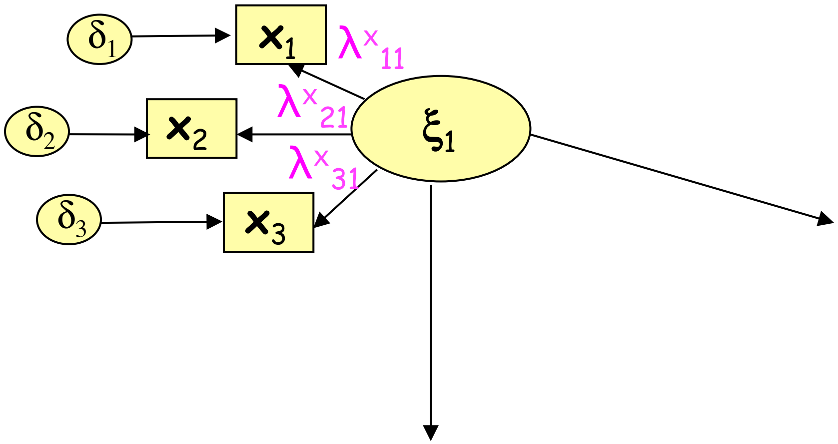
$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_6 \end{bmatrix} \qquad \Lambda_x = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix} \qquad \xi = \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} \qquad \delta = \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_6 \end{bmatrix}$$

$$\mathbf{x} = \Lambda_x \xi + \delta$$

Modéle de chemin avec des variables latentes



Le modéle de mesure (externe) pour les VM exogènes



$x_1 = \lambda^x{}_{11}\ \xi_1 + \delta_1$

$x_2 = \lambda^x{}_{21}\ \xi_1 + \delta_2$

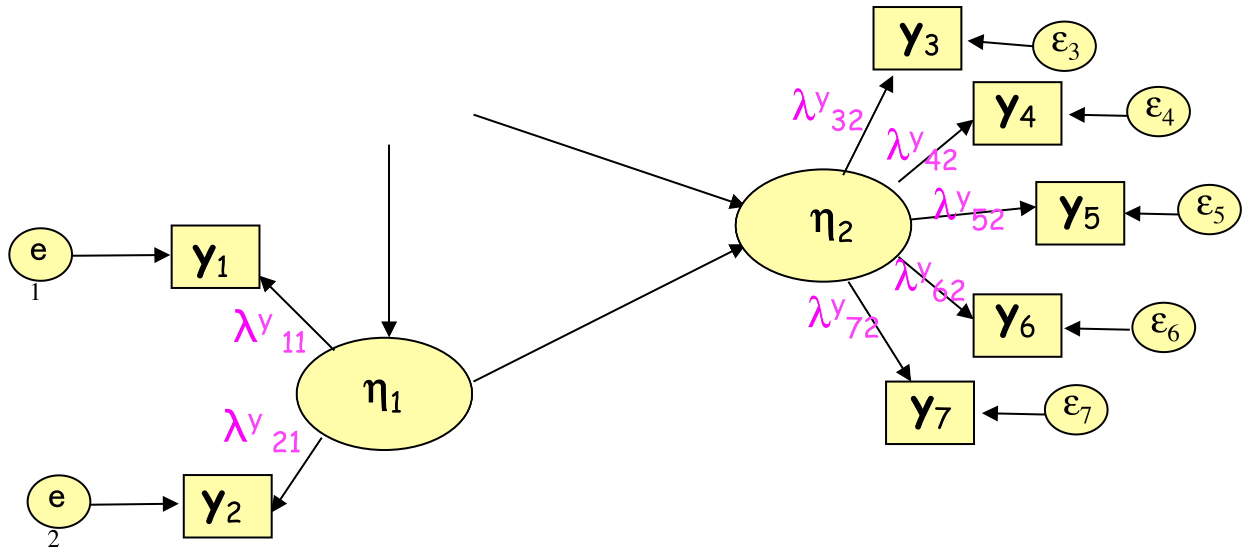
$x_3 = \lambda^x{}_{31}\ \xi_1 + \delta_3$

$$\underset{(P\times 1)}{\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}} \quad \underset{(P\times M)}{\Lambda_x = \begin{bmatrix} \lambda^x_{11} \\ \lambda^x_{21} \\ \lambda^x_{31} \end{bmatrix}} \quad \underset{(M\times 1)}{\xi = [\xi_1]} \quad \underset{(P\times 1)}{\delta = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix}}$$

$$\mathbf{x} = \Lambda_x \xi + \delta$$

Le modéle de mesure (externe) pour les VM endogènes

$$y_1 = \lambda^{y_{11}}\eta_1 + \varepsilon_1$$
$$y_2 = \lambda^{y_{21}}\eta_1 + \varepsilon_2$$
$$y_3 = \lambda^{y_{32}}\eta_2 + \varepsilon_3$$
$$y_4 = \lambda^{y_{42}}\eta_2 + \varepsilon_4$$
$$y_5 = \lambda^{y_{52}}\eta_2 + \varepsilon_5$$
$$y_6 = \lambda^{y_{62}}\eta_2 + \varepsilon_6$$
$$y_7 = \lambda^{y_{72}}\eta_2 + \varepsilon_7$$



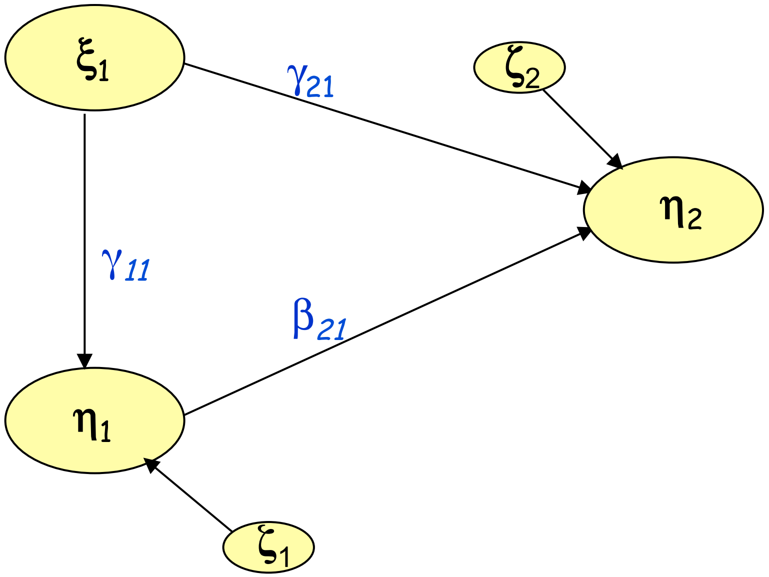
$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \end{bmatrix} \qquad \Lambda_y = \begin{pmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \\ 0 & \lambda_{72} \end{pmatrix} \qquad \eta = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \end{bmatrix}$$

$$\mathbf{y} = \Lambda_y \eta + \varepsilon$$

Le modèle structurel (interne)

$$\eta_1 = \gamma_{11}\,\xi_1 + \zeta_1$$

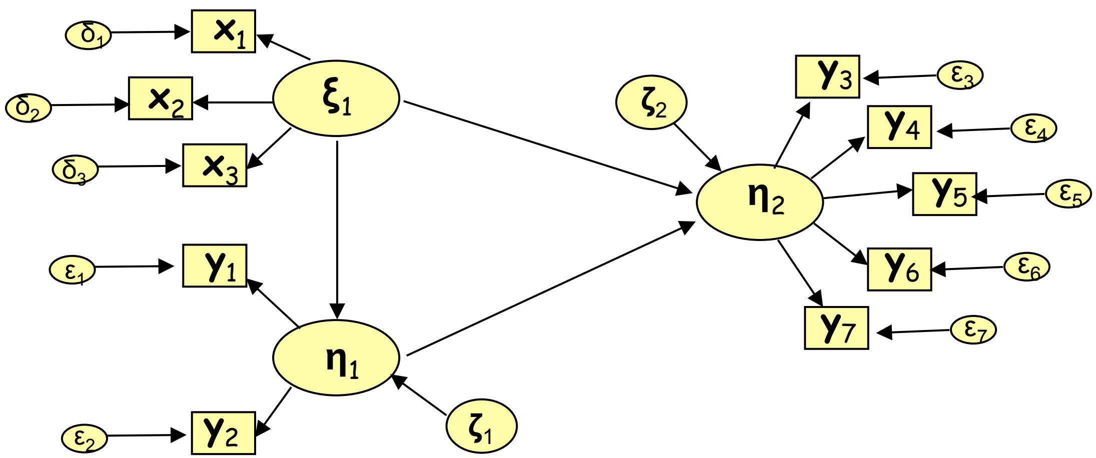
$$\eta_2 = \gamma_{21}\,\xi_1 + \beta_{21}\eta_1 + \zeta_2$$



$$\eta = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} \quad \xi = \begin{bmatrix} \xi_1 \end{bmatrix} \quad \Gamma = \begin{bmatrix} \gamma_{11} \\ \gamma_{21} \end{bmatrix} \quad \mathrm{B} = \begin{bmatrix} 0 & 0 \\ \beta_{21} & 0 \end{bmatrix} \quad \zeta = \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}$$

$$\eta = \Gamma\xi + \mathrm{B}\eta + \zeta$$

Le modèle global



$$\mathbf{x} = \Lambda_x \xi + \delta$$

$$\mathbf{y} = \Lambda_y \eta + \varepsilon$$

Modèles de mesure

Modéle structurel

$$\eta = \Gamma\xi + \mathrm{B}\eta + \zeta \Leftrightarrow \eta = (\mathbf{I} - \mathbf{B})^{-1}(\Gamma\xi + \zeta)$$

Two families of methods for estimating paths models with latent variables

Méthodes basées sur la covariance

Le but est de reproduire de matrice de covariance de l'échantillon des variables manifestes à l'aide des paramètres du modèle:

- c'est une approche confirmative visant à valider un modèle

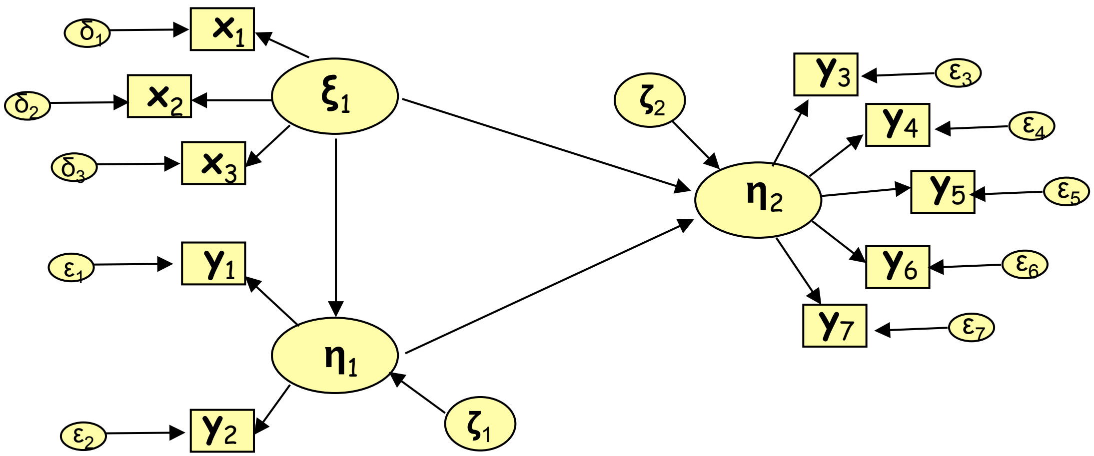
SEM

Component-based Methods

Le but est de fournir une estimation des scores de variables latentes

- il est une approche plus exploratoire (stratégie opérationnel)

Les hypothèses du modèle



$$\mathbf{x} = \Lambda_x \xi + \delta$$

$$\mathbf{y} = \Lambda_y \eta + \varepsilon$$

$$\eta = (\mathbf{I} - \mathbf{B})^{-1}(\Gamma \xi + \zeta)$$

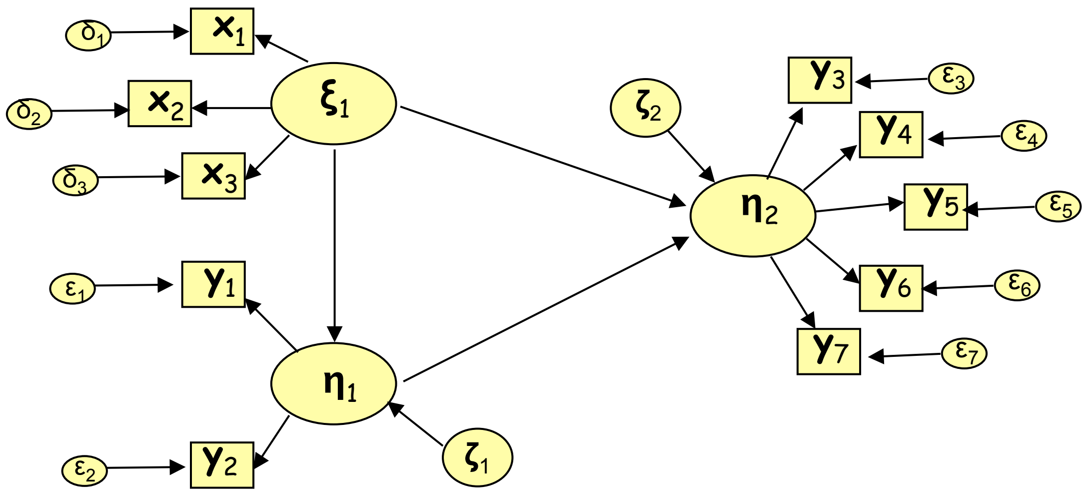
$$E(\mathbf{x}) = 0,\ E(\mathbf{y}) = 0,\ E(\xi) = 0,\ E(\eta) = 0,\ E(\delta) = 0,\ E(\varepsilon) = 0,\ E(\zeta) = 0$$

$$E(\delta\varepsilon') = 0,\ E\big((\zeta\delta')\big) = 0,\ E(\zeta\varepsilon') = 0$$

$$E(\delta\xi') = 0,\ E(\delta\eta') = 0,\ E(\varepsilon\xi') = 0,\ E(\varepsilon\eta') = 0$$

$$E(\zeta\xi') = 0$$

Covariance Structure Analysis (CSA, aka LISREL)



$$\Sigma = \begin{bmatrix} \Sigma_{xx} & \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}$$

Une fois que le modèle est spécifié, la matrice de covariance des MV est exprimée en termes de paramètres du modèle

$$\Sigma(\Omega) = \Sigma(\Gamma, \mathbf{B}, \Lambda_x, \Lambda_y, \Phi, \Psi, \Theta_\delta, \Theta_\varepsilon)$$

Path Coefficients

Loadings

Exog. LV Covariance

Structural Error Covariance

Measurement Error Covariance

Analysing the covariance

Estimation minimise une certaine fonction de la divergence entre la matrice de covariance (C) impliquée par le modèle et la matrice de covariance (S).

$$\Sigma = \begin{bmatrix} \Sigma_{xx} & \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}$$

Population Covariance matrix

$$S = \begin{bmatrix} s_{xx} & \\ s_{yx} & s_{yy} \end{bmatrix}$$

Empirical covariance matrix

$$\Sigma(\Omega) = \left[\begin{array}{c:c} \Lambda_x \Phi \Lambda'_x + \Theta_\delta & \\ \hdashline \Lambda_y (\mathbf{I}-\mathbf{B})^{-1} \Gamma \Phi' \Lambda'_x & \Lambda_y \left[(\mathbf{I}-\mathbf{B})^{-1} (\Gamma \Phi \Gamma' + \Psi) (\mathbf{I}-\mathbf{B})^{-1'} \right] \Lambda'_y + \Theta_\varepsilon \end{array}\right]$$

"Implied" covariance matrix

$$\Sigma(\hat{\Omega}) = \mathbf{C} = \left[\begin{array}{c:c} \hat{\Lambda}_x \hat{\Phi} \hat{\Lambda}'_x + \hat{\Theta}_\delta & \\ \hdashline \hat{\Lambda}_y (\mathbf{I}-\hat{\mathbf{B}})^{-1} \hat{\Gamma} \hat{\Phi}' \hat{\Lambda}'_x & \hat{\Lambda}_y \left[(\mathbf{I}-\hat{\mathbf{B}})^{-1} (\hat{\Gamma} \hat{\Phi} \hat{\Gamma}' + \hat{\Psi}) (\mathbf{I}-\hat{\mathbf{B}})^{-1'} \right] \hat{\Lambda}'_y + \hat{\Theta}_\varepsilon \end{array}\right]$$

Estimated covariance matrix

$$\boxed{S \approx C}$$

Fonctions de divergence

Maximum Likelihood

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

Unweighted Least Squares

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

Generalised Least Squares

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

Considerations sur l'approache CSA

- Estimateurs avec de bonnes propriétés statistiques
- Les estimations sont obtenues au moyen de ML, ULS, WLS, GLS;
- Taille de l'échantillon: 200-400 cas pour 10-15 variables observées;
- L'identification du modèle (estimation unique des paramètres) peut être un problème (qui peut être résolu en imposant un ensemble de contraintes sur les paramètres)
- Aucune estimation directe de variables latentes -> ce n'est pas son objectif!!

Références

- K.A. Bollen : Structural Equations with Latent Variables, John Wiley & Sons., 1989
- R. Cudeck, S. Du Toit, D. Sörbom (Editors) : Structural Equation Modeling: Present and Future - A Festschrift in honor of Karl G. Jöreskog, Scientific Software International Chicago, 2001
- R. Hoyle : Structural equation modeling: concepts, issues and applications, SAGE Publications, 1995
- K.G. Jöreskog & D. Sörbom : Advances in Factor Analysis and Structural Equation Models, Abt Books, 1979
- K.G. Jöreskog, D. Sörbom, S. Du Toit, M. Du Toit : LISREL8: New Statistical Features, Second Printing with Revisions, Scientific Software International Chicago, 2000
- D. Kaplan : Structural Equation Modeling: Foundations and Extensions, SAGE Publications, 2000
- Rex B. Kline : Principles and Practice of Structural Equation Modeling, The Guilford Press, New York, 2005

- D. Kaplan : Structural Equation Modeling: Foundations and Extensions, SAGE Publications, 2000