le cnam Paris

Structural Equation Modeling with factor and composites for the basic design

A. Tenenhaus, M. Tenenhaus & T. Dijkstra

2026/02/05@CNAM

L2S Laboratoire Signaux & Systèmes

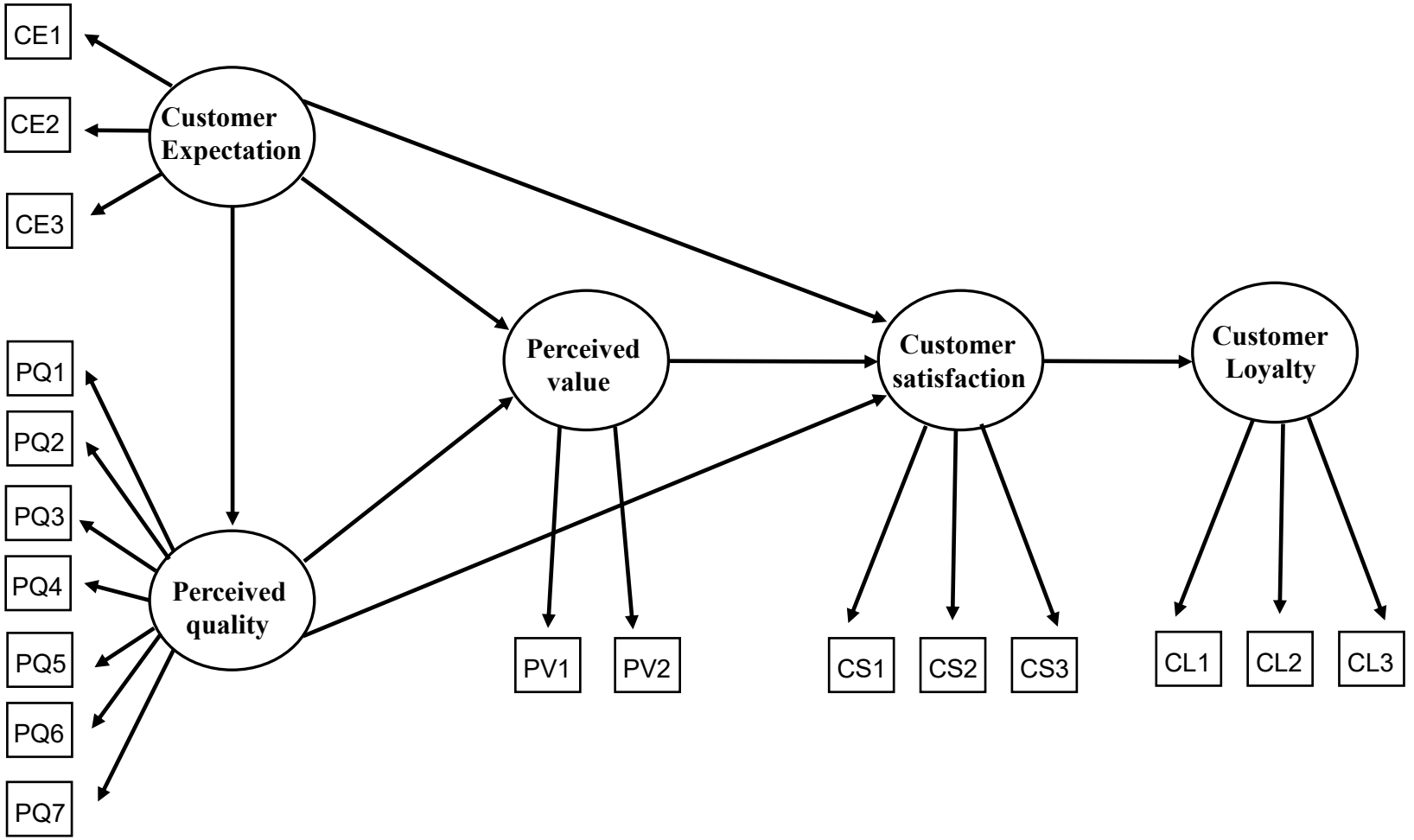
HEC PARIS

university of groningen

Herman Wold and his research group in 1977. From left to right: Theo Dijkstra, Heino Apel, Herman Wold and Baldwin Hui.

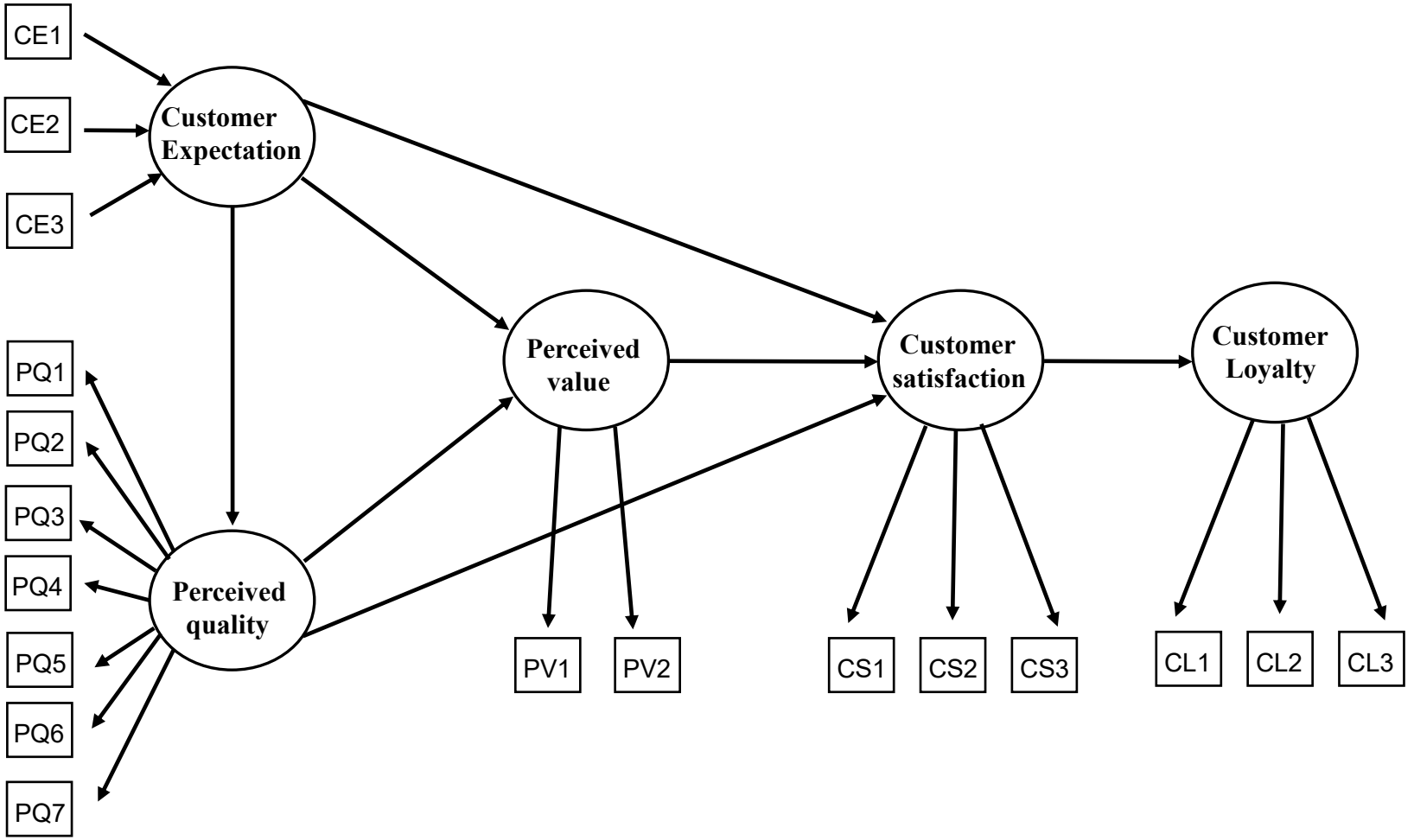
Joint work with Theo Dijkstra who passed away on September 2020.

Path model describing causes and consequences of Customer Satisfaction



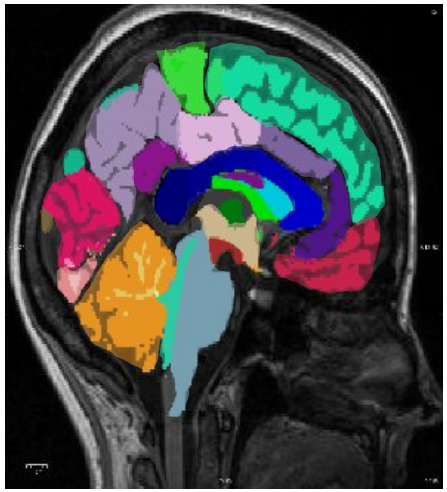
ECSI is an economic indicator that measures customer satisfaction

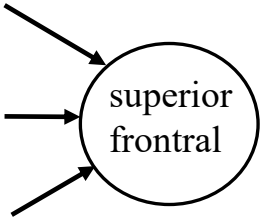
Path model describing causes and consequences of Customer Satisfaction

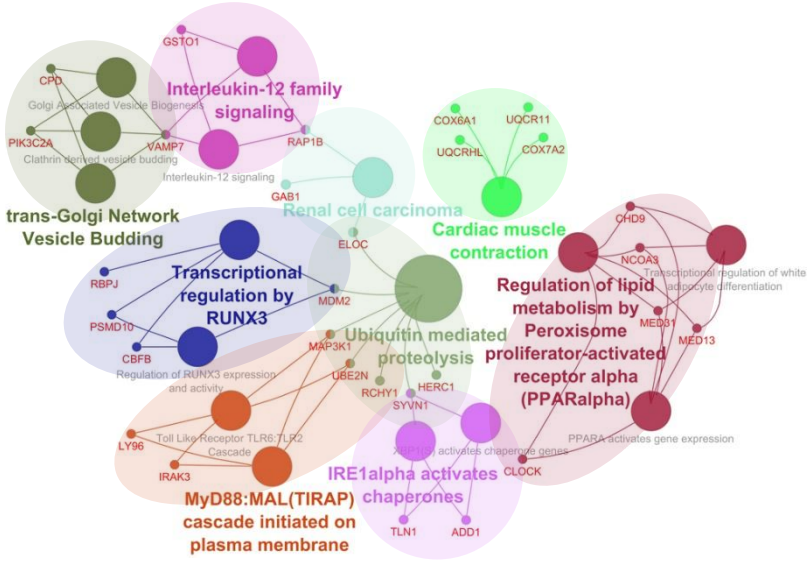


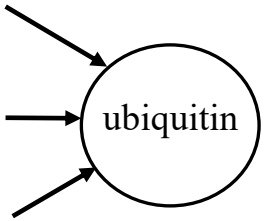
ECSI is an economic indicator that measures customer satisfaction

factor or composite?

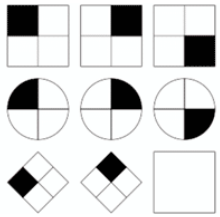


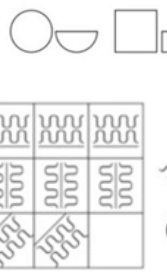


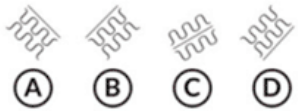


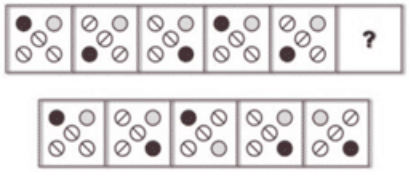


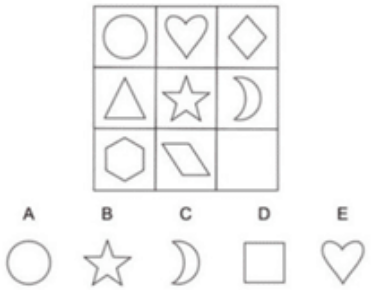
factor or composite?

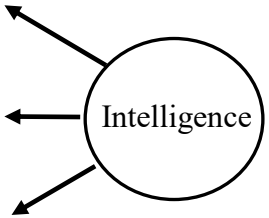






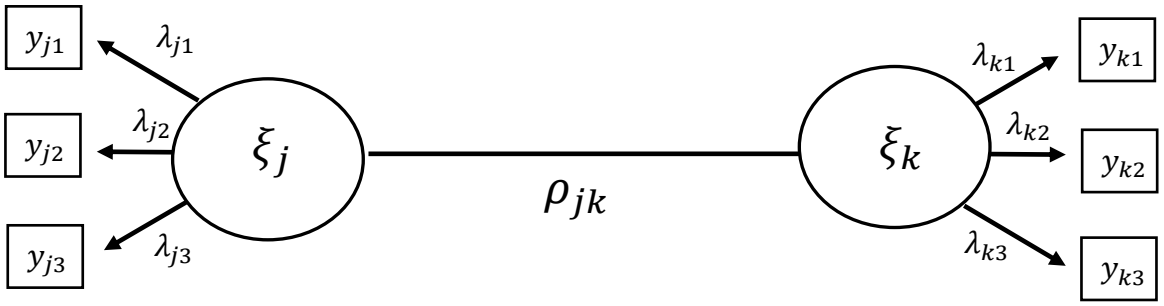






The Wold's principle for <u>factors</u> and composites

Standardized LVs ξ_j



$$\operatorname{cov}(y_{jh}, y_{kl}) = \underbrace{\operatorname{cov}(y_{jh}, \xi_j)}_{\lambda_{jh}} \times \underbrace{\operatorname{cor}(\xi_j, \xi_k)}_{\rho_{jk}} \times \underbrace{\operatorname{cov}(y_{kl}, \xi_k)}_{\lambda_{kl}}$$

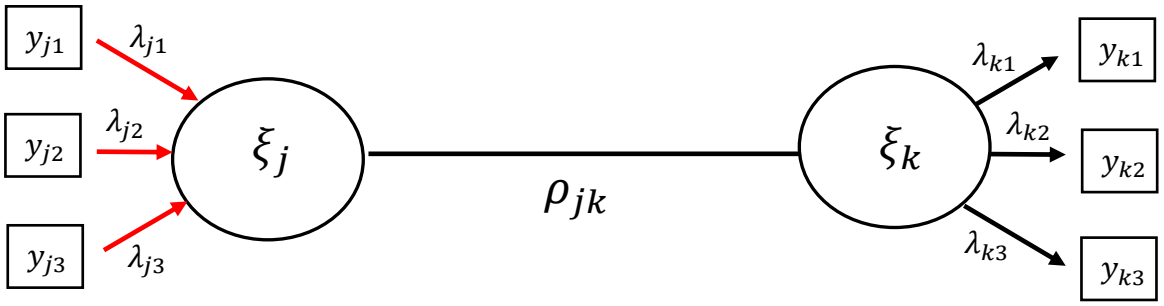
Or, in more compact matrix notations:

$$\boldsymbol{\Sigma}_{jk} = \rho_{jk} \boldsymbol{\lambda}_j \boldsymbol{\lambda}_k^{\top} \qquad \text{where } \boldsymbol{\lambda}_j = \begin{bmatrix} \lambda_{j1} \\ \lambda_{j2} \\ \lambda_{j3} \end{bmatrix} = \operatorname{cov}(\boldsymbol{y}_j, \xi_j)$$

The Wold's principle for factors and <u>composites</u>

Standardized LVs ξ_j

$$\xi_j = \mathbf{w}_j^\top \boldsymbol{y}_j$$



$$\underbrace{\mathrm{cov}(y_{jh}, y_{kl})}_{} = \underbrace{\mathrm{cov}(y_{jh}, \xi_j)}_{\lambda_{jh}} \times \underbrace{\mathrm{cor}(\xi_j, \xi_k)}_{\rho_{jk}} \times \underbrace{\mathrm{cov}(y_{kl}, \xi_k)}_{\lambda_{kl}}$$

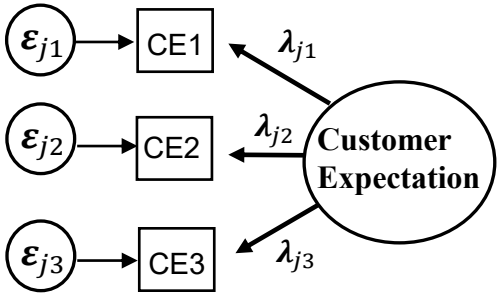
Or, in more compact matrix notations:

$$\boldsymbol{\Sigma}_{jk} = \rho_{jk} \boldsymbol{\lambda}_j \boldsymbol{\lambda}_k^\top$$

where $\boldsymbol{\lambda}_j = \mathrm{cov}(\boldsymbol{y}_j, \xi_j) = \boldsymbol{\Sigma}_{jj} \mathbf{w}_j$

$$\Rightarrow \mathbf{w}_j = \boldsymbol{\Sigma}_{jj}^{-1} \boldsymbol{\lambda}_j$$

The measurement model for factors



$$\begin{bmatrix} \text{CE1} \\ \text{CE2} \\ \text{CE3} \end{bmatrix} = \begin{bmatrix} y_{j1} \\ y_{j2} \\ y_{j3} \end{bmatrix} = \begin{bmatrix} \lambda_{j1} \\ \lambda_{j2} \\ \lambda_{j3} \end{bmatrix} \xi_j + \begin{bmatrix} \varepsilon_{j1} \\ \varepsilon_{j2} \\ \varepsilon_{j3} \end{bmatrix}$$

The basic measurement model is defined for all j **:**

$$\boldsymbol{y}_j = \boldsymbol{\lambda}_j \xi_j + \boldsymbol{\varepsilon}_j$$

Standardized LVs ξ_j

$\varepsilon_{jh} \perp \varepsilon_{jl}$

$\boldsymbol{\varepsilon}_j \perp \xi_j$

- The basic measurement model for factor implies:
  - ► $\boldsymbol{\Sigma}_{jj} = \mathbb{E}\left(\boldsymbol{y}_j \boldsymbol{y}_j^\top\right) = \boldsymbol{\lambda}_j \boldsymbol{\lambda}_j^\top + \boldsymbol{\Theta}_j$

    $\mathbb{E}\left(\boldsymbol{\varepsilon}_j \boldsymbol{\varepsilon}_j^\top\right) = \text{diag}\left(\theta_{j1}, \ldots, \theta_{jh}, \ldots, \theta_{jp_j}\right)$
  - ► $\boldsymbol{\Sigma}_{jk} = \mathbb{E}\left(\boldsymbol{y}_j \boldsymbol{y}_k^\top\right) = \rho_{jk} \boldsymbol{\lambda}_j \boldsymbol{\lambda}_k^\top$

The measurement model for factors

The basic measurement model for factor only implies:

- $\boldsymbol{\Sigma}_{jj} = \boldsymbol{\lambda}_j \boldsymbol{\lambda}_j^{\top} + \boldsymbol{\Theta}_j$
- $\boldsymbol{\Sigma}_{jk} = \rho_{jk} \boldsymbol{\lambda}_j \boldsymbol{\lambda}_k^{\top}$

and reduces to:

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\lambda}_1 \boldsymbol{\lambda}_1^{\top} + \boldsymbol{\Theta}_1 & & & \\ \rho_{21} \boldsymbol{\lambda}_2 \boldsymbol{\lambda}_1^{\top} & \boldsymbol{\lambda}_2 \boldsymbol{\lambda}_2^{\top} + \boldsymbol{\Theta}_2 & & \\ \rho_{31} \boldsymbol{\lambda}_1 \boldsymbol{\lambda}_3^{\top} & \rho_{32} \boldsymbol{\lambda}_3 \boldsymbol{\lambda}_2^{\top} & \boldsymbol{\lambda}_3 \boldsymbol{\lambda}_3^{\top} + \boldsymbol{\Theta}_3 & \\ \vdots & \vdots & \vdots & \ddots \\ \rho_{J1} \boldsymbol{\lambda}_J \boldsymbol{\lambda}_1^{\top} & \rho_{J2} \boldsymbol{\lambda}_J \boldsymbol{\lambda}_2^{\top} & \rho_{J3} \boldsymbol{\lambda}_J \boldsymbol{\lambda}_3^{\top} & \boldsymbol{\lambda}_J \boldsymbol{\lambda}_J^{\top} + \boldsymbol{\Theta}_J \end{pmatrix}$$

The measurement model for factors

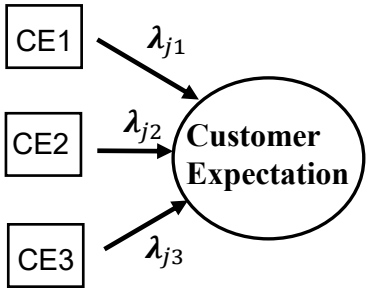
The basic measurement model for factor only implies:

- $\boldsymbol{\Sigma}_{jj} = \boldsymbol{\lambda}_j \boldsymbol{\lambda}_j^{\top} + \boldsymbol{\Theta}_j$
- $\boldsymbol{\Sigma}_{jk} = \rho_{jk} \boldsymbol{\lambda}_j \boldsymbol{\lambda}_k^{\top}$

and reduces to:

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda} \mathbf{R} \boldsymbol{\Lambda}^{\top} + \boldsymbol{\Theta}$$

The measurement model for <u>composites</u>



- $\xi_j = \mathbf{w}_j^\top \boldsymbol{y}_j$ Standardized LVs ξ_j

- The basic measurement model for composites assumes:
  - $\boldsymbol{\Sigma}_{jj}$ invertible with no particular structure $\left(\boldsymbol{\Sigma}_{jj} \neq \boldsymbol{\lambda}_j \boldsymbol{\lambda}_j^\top + \boldsymbol{\Theta}_j\right)$
  - $\boldsymbol{\Sigma}_{jk} = \rho_{jk} \boldsymbol{\lambda}_j \boldsymbol{\lambda}_k^\top$ (Wold's principle)

The measurement model for <u>composites</u>

The basic measurement model for composites only assumes:

- no particular structure for $\boldsymbol{\Sigma}_{jj}$

  (variables are freely correlated)

- $\boldsymbol{\Sigma}_{jk} = \rho_{jk}\boldsymbol{\lambda}_j\boldsymbol{\lambda}_k^{\top}$

  (Wold's principle)

and reduces to:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & & & & \\ \rho_{21}\boldsymbol{\lambda}_2\boldsymbol{\lambda}_1^{\top} & \boldsymbol{\Sigma}_{22} & & & \\ \rho_{31}\boldsymbol{\lambda}_1\boldsymbol{\lambda}_3^{\top} & \rho_{32}\boldsymbol{\lambda}_3\boldsymbol{\lambda}_2^{\top} & \boldsymbol{\Sigma}_{33} & & \\ \vdots & \vdots & \vdots & \ddots & \\ \rho_{J1}\boldsymbol{\lambda}_J\boldsymbol{\lambda}_1^{\top} & \rho_{J2}\boldsymbol{\lambda}_J\boldsymbol{\lambda}_2^{\top} & \rho_{J3}\boldsymbol{\lambda}_J\boldsymbol{\lambda}_3^{\top} & \dots & \boldsymbol{\Sigma}_{JJ} \end{pmatrix}$$

The measurement model for factors and composites

The basic measurement model for **factors and composites** implies:

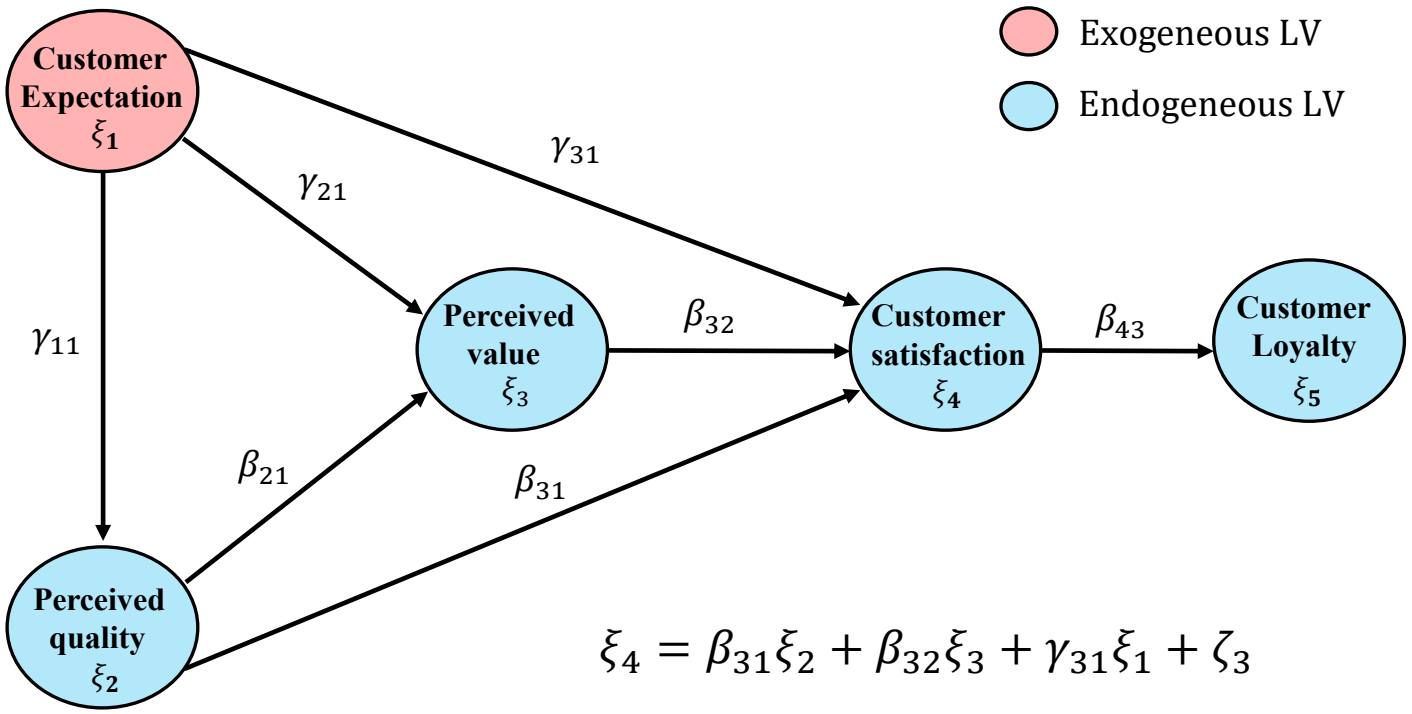
$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{pmatrix} \widetilde{\boldsymbol{\Sigma}}_{11} & & & & \\ \rho_{21}\boldsymbol{\lambda}_2\boldsymbol{\lambda}_1^\top & \widetilde{\boldsymbol{\Sigma}}_{22} & & & \\ \rho_{31}\boldsymbol{\lambda}_1\boldsymbol{\lambda}_3^\top & \rho_{32}\boldsymbol{\lambda}_3\boldsymbol{\lambda}_2^\top & \widetilde{\boldsymbol{\Sigma}}_{33} & & \\ \vdots & \vdots & \vdots & \ddots & \\ \rho_{J1}\boldsymbol{\lambda}_J\boldsymbol{\lambda}_1^\top & \rho_{J2}\boldsymbol{\lambda}_J\boldsymbol{\lambda}_2^\top & \rho_{J3}\boldsymbol{\lambda}_J\boldsymbol{\lambda}_3^\top & \dots & \widetilde{\boldsymbol{\Sigma}}_{JJ} \end{pmatrix}$$

$$\text{where } \widetilde{\boldsymbol{\Sigma}}_{jj} = \begin{cases} \boldsymbol{\lambda}_j\boldsymbol{\lambda}_j^\top + \boldsymbol{\Theta}_j, & \text{for block associated with factors,} \\ \boldsymbol{\Sigma}_{jj}, & \text{for block associated with composites} \end{cases}$$

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}\mathbf{R}\boldsymbol{\Lambda}^\top + \boldsymbol{\Theta} + \text{diag}\left[\boldsymbol{\Sigma}_{11} - \left(\boldsymbol{\lambda}_1\boldsymbol{\lambda}_1^\top + \boldsymbol{\Theta}_1\right), \dots, \boldsymbol{\Sigma}_{JJ} - \left(\boldsymbol{\lambda}_J\boldsymbol{\lambda}_J^\top + \boldsymbol{\Theta}_J\right)\right]$$

The structural model

Recursive model (here!)



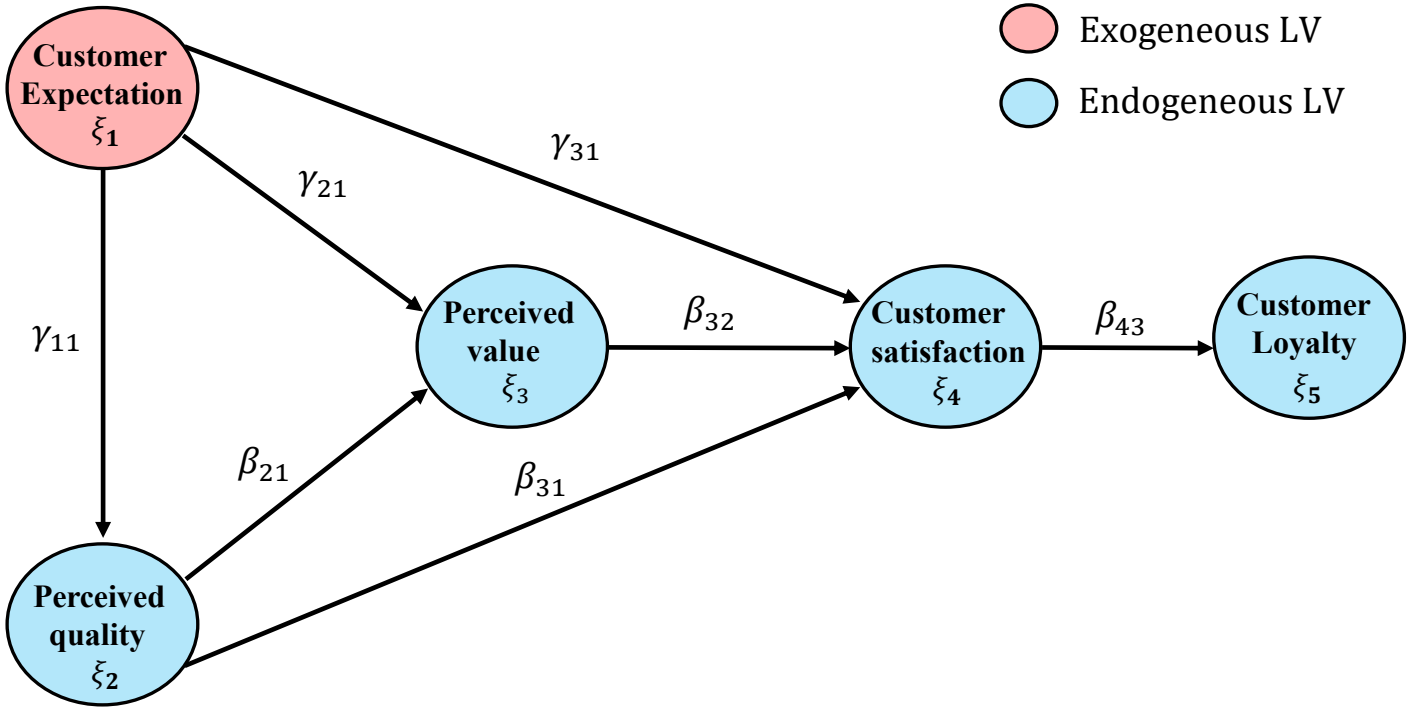
$$\xi_4 = \beta_{31}\xi_2 + \beta_{32}\xi_3 + \gamma_{31}\xi_1 + \zeta_3$$

The ith structural equation can be expressed as follows:

$$\eta_{n+i} = \boldsymbol{\beta}_i^{\top}\boldsymbol{\xi}_{endo_i} + \boldsymbol{\gamma}_i^{\top}\boldsymbol{\xi}_{exo_i} + \zeta_i$$

The structural model

Recursive model



In matrix notations:

$$\begin{pmatrix}\xi_2\\ \xi_3\\ \xi_4\\ \xi_5\end{pmatrix} = \begin{pmatrix}0 & 0 & 0 & 0\\ \beta_{21} & 0 & 0 & 0\\ \beta_{31} & \beta_{32} & 0 & 0\\ 0 & 0 & \beta_{43} & 0\end{pmatrix}\begin{pmatrix}\xi_2\\ \xi_3\\ \xi_4\\ \xi_5\end{pmatrix} + \begin{pmatrix}\gamma_{11}\\ \gamma_{21}\\ \gamma_{31}\\ 0\end{pmatrix}\eta_1 + \begin{pmatrix}\zeta_1\\ \zeta_2\\ \zeta_3\\ \zeta_4\end{pmatrix}$$

$$\boldsymbol{\xi}_{endo} = \mathbf{B}\boldsymbol{\xi}_{endo} + \boldsymbol{\Gamma}\boldsymbol{\xi}_{exo} + \boldsymbol{\zeta}$$

The structural model

All the structural equations can be expressed compactly:

$$\boldsymbol{\xi}_{endo} = \mathbf{B}\boldsymbol{\xi}_{endo} + \boldsymbol{\Gamma}\boldsymbol{\xi}_{exo} + \boldsymbol{\zeta}$$

Denoting $\boldsymbol{\xi} = (\boldsymbol{\xi}_{exo}; \boldsymbol{\xi}_{endo})$, the correlation matrix between LVs $\mathbf{R}(\boldsymbol{\theta}) = \mathbb{E}(\boldsymbol{\xi}\boldsymbol{\xi}^{\top})$ can be expressed in terms of $\mathbf{B}, \boldsymbol{\Gamma}, \boldsymbol{\Phi} = \mathbb{E}(\boldsymbol{\xi}_{exo}\boldsymbol{\xi}_{exo}^{\top})$ and $\boldsymbol{\Psi} = \mathbb{E}(\boldsymbol{\zeta}\boldsymbol{\zeta}^{\top})$:

$$\mathbf{R}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbb{E}(\boldsymbol{\xi}_{exo}\boldsymbol{\xi}_{exo}^{\top}) & \mathbb{E}(\boldsymbol{\xi}_{exo}\boldsymbol{\xi}_{endo}^{\top}) \\ \mathbb{E}(\boldsymbol{\xi}_{endo}\boldsymbol{\xi}_{exo}^{\top}) & \mathbb{E}(\boldsymbol{\xi}_{endo}\boldsymbol{\xi}_{endo}^{\top}) \end{pmatrix}$$

(restricted) MLE for SEM with factors and composites

The SEM for factors and composites can be summarized by:

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}\mathbf{R}(\boldsymbol{\theta})\boldsymbol{\Lambda}^{\top} + \boldsymbol{\Theta} + \mathrm{diag}\left[\boldsymbol{\Sigma}_{11} - \left(\boldsymbol{\lambda}_1\boldsymbol{\lambda}_1^{\top} + \boldsymbol{\Theta}_1\right), \ldots, \boldsymbol{\Sigma}_{JJ} - \left(\boldsymbol{\lambda}_J\boldsymbol{\lambda}_J^{\top} + \boldsymbol{\Theta}_J\right)\right]$$

Assuming that the vector $\boldsymbol{y}$ of observed variables has a multinormal distribution, we consider the minimizing the maximum likelihood fitting function

$$F(\boldsymbol{\theta}) = \log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| + \mathrm{tr}\left(\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{S}\right) - \log|\mathbf{S}| - p$$

subject to $\boldsymbol{\lambda}_j^{\top}\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\lambda}_j$=1 for block associated with composites

SOLNP (Ye, 1987) for minimizing $F(\boldsymbol{\theta})$.

Ye (1987), *Interior algorithms for linear, quadratic, and linearly constrained non linear programming*, PhD Thesis, Department of EES Stanford University, Stanford CA.

SEM for factors and composites

The SEM for factors and composites can be summarized by:

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}\mathbf{R}(\boldsymbol{\theta})\boldsymbol{\Lambda}^{\top} + \boldsymbol{\Theta} + \mathrm{diag}\left[\boldsymbol{\Sigma}_{11} - \left(\boldsymbol{\lambda}_1\boldsymbol{\lambda}_1^{\top} + \boldsymbol{\Theta}_1\right), \ldots, \boldsymbol{\Sigma}_{JJ} - \left(\boldsymbol{\lambda}_J\boldsymbol{\lambda}_J^{\top} + \boldsymbol{\Theta}_J\right)\right]$$

Measurement model parameters	$\boldsymbol{\Lambda}$
	$\boldsymbol{\Theta}$
Latent variable model parameters	$\mathbf{B}$
	$\boldsymbol{\Gamma}$
	$\boldsymbol{\Phi}$
	$\boldsymbol{\Psi}$

IMPORTANT RESULTS

All these parameters can be expressed in terms of $\boldsymbol{\Lambda}$ and $\boldsymbol{\Sigma}$!

&

$\hat{\boldsymbol{\Lambda}}$ can be obtained by SVD

$\boldsymbol{\lambda}_j$ is obtained by SVD

- The normalized loading vector $\boldsymbol{\lambda}_j^* = \boldsymbol{\lambda}_j / \|\boldsymbol{\lambda}_j\|$ is the unit norm eigenvector of the rank-1 matrix $\boldsymbol{\Sigma}_{jk}\boldsymbol{\Sigma}_{kj} = \rho_{jk}^2 \|\boldsymbol{\lambda}_k\|^2 \boldsymbol{\lambda}_j \boldsymbol{\lambda}_j^\top$:

$$\Rightarrow \quad \boldsymbol{\lambda}_j = \|\boldsymbol{\lambda}_j\| \boldsymbol{\lambda}_j^*$$

To avoid choosing a specific pair, k, j we can choose $\boldsymbol{\lambda}_j^*$ as unit norm eigenvector of the rank-1 matrix:

$$\sum_{k=1:k\neq j}^{J} \boldsymbol{\Sigma}_{jk}\boldsymbol{\Sigma}_{kj} = \sum_{k=1:k\neq j}^{J} \rho_{jk}^2 \|\boldsymbol{\lambda}_k\|^2 \boldsymbol{\lambda}_j \boldsymbol{\lambda}_j^\top$$

Parameters of the model in terms of Λ^* and Σ

$\boldsymbol{\lambda}_j^*$ is the unit norm eigenvector of $\sum_{k=1:k\neq j}^{J} \boldsymbol{\Sigma}_{jk}\boldsymbol{\Sigma}_{kj}$

Correlations between LVs

$$\boldsymbol{\Sigma}_{jk} = \rho_{jk}\boldsymbol{\lambda}_j\boldsymbol{\lambda}_k^\top \Leftrightarrow \boldsymbol{\lambda}_j^\top\boldsymbol{\Sigma}_{jk}\boldsymbol{\lambda}_k = \rho_{jk}\boldsymbol{\lambda}_j^\top\boldsymbol{\lambda}_j\boldsymbol{\lambda}_k^\top\boldsymbol{\lambda}_k$$

$$\Leftrightarrow \rho_{jk} = \frac{\boldsymbol{\lambda}_j^\top\boldsymbol{\Sigma}_{jk}\boldsymbol{\lambda}_k}{\|\boldsymbol{\lambda}_j\|^2\|\boldsymbol{\lambda}_k\|^2} = \frac{\boldsymbol{\lambda}_j^{*\top}\boldsymbol{\Sigma}_{jk}\boldsymbol{\lambda}_k^*}{\|\boldsymbol{\lambda}_j\|\|\boldsymbol{\lambda}_k\|}$$

Norm of the loading vector $\boldsymbol{\lambda}_j$

$$\|\boldsymbol{\lambda}_j\|^2 = \begin{cases} \dfrac{\boldsymbol{\lambda}_j^{*\top}\left(\boldsymbol{\Sigma}_{jj} - \mathrm{diag}(\boldsymbol{\Sigma}_{jj})\right)\boldsymbol{\lambda}_j^*}{1 - \boldsymbol{\lambda}_j^{*\top}\mathrm{diag}\left(\boldsymbol{\lambda}_j^*\boldsymbol{\lambda}_j^{*\top}\right)\boldsymbol{\lambda}_j^*}, & \text{for factor} \\ \dfrac{1}{\boldsymbol{\lambda}_j^{*\top}\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\lambda}_j^*}, & \text{for composite} \end{cases}$$

The measurement model

$$\boldsymbol{y}_j = \boldsymbol{\lambda}_j\,\xi_j + \boldsymbol{\varepsilon}_j \qquad \varepsilon_{jh} \perp \varepsilon_{jl},\ \boldsymbol{\varepsilon}_j \perp \xi_j$$

Standardized ξ_j

$$\boldsymbol{\Theta}_j = \mathbb{E}\left(\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j^\top\right) = \mathrm{diag}\left(\theta_{j1}, \ldots, \theta_{jh}, \ldots, \theta_{jp_j}\right)$$

Variance of the residuals θ_{jh}

$$\begin{aligned} \theta_{jh} &= \mathrm{var}(\varepsilon_{jh}) = \mathrm{var}(y_{jh} - \lambda_{jh}\xi_j) \\ &= \mathrm{var}(y_{jh}) - 2\mathrm{cov}(y_{jh}, \lambda_{jh}\xi_j) + \mathrm{var}(\lambda_{jh}\xi_j) \\ &= \mathrm{var}(y_{jh}) - 2\mathrm{cov}(\lambda_{jh}\xi_j + \varepsilon_{jh}, \lambda_{jh}\xi_j) + \lambda_{jh}^2 \\ &= \mathrm{var}(y_{jh}) - \lambda_{jh}^2 \\ &= \mathrm{var}(y_{jh}) - \lambda_{jh}^{2*}\|\lambda_j\|^2 \end{aligned}$$

The same stands for $\mathbf{B}, \boldsymbol{\Gamma}, \boldsymbol{\Phi}, \boldsymbol{\Psi}$

All the parameters can be expressed in terms of λ_j^* and Σ_{jj} !!

SVD-SEM

Eigen decomposition

of $\sum_{k=1:k\neq j}^{J} \mathbf{S}_{jk}\mathbf{S}_{kj}$ $\rightarrow$ $\hat{\boldsymbol{\lambda}}_j^*$ $\rightarrow$ $\widehat{\|\boldsymbol{\lambda}_j\|}$ $\rightarrow$ $\hat{\boldsymbol{\lambda}}_j = \widehat{\|\boldsymbol{\lambda}_j\|}\hat{\boldsymbol{\lambda}}_j^*$

$\hat{\boldsymbol{\Theta}}$

$$\hat{\mathbf{w}}_j = \frac{\mathbf{S}_{jj}^{-1}\hat{\boldsymbol{\lambda}}_j^*}{\sqrt{\hat{\boldsymbol{\lambda}}_j^{*\top}\mathbf{S}_{jj}^{-1}\hat{\boldsymbol{\lambda}}_j^*}}$$

$$\tilde{\rho}_{jk} = \frac{\hat{\boldsymbol{\lambda}}_j^{*\top}\mathbf{S}_{jk}\hat{\boldsymbol{\lambda}}_k^*}{\widehat{\|\boldsymbol{\lambda}_j\|}\widehat{\|\boldsymbol{\lambda}_k\|}}$$

$$\tilde{\mathbf{R}} = \{\tilde{\rho}_{jk}\}$$

$\hat{\mathbf{R}}$

$\hat{\mathbf{B}}, \hat{\boldsymbol{\Gamma}}, \hat{\boldsymbol{\Psi}}, \hat{\boldsymbol{\Phi}}$

All these estimators are Consistent and Asymptotically Normal (CAN)

The model implied covariance matrix $\hat{\boldsymbol{\Sigma}} = \{\hat{\boldsymbol{\Sigma}}_{jk}\}$ is deduced:

$$\hat{\boldsymbol{\Sigma}}_{jk} = \hat{\rho}_{jk}\hat{\boldsymbol{\lambda}}_j\hat{\boldsymbol{\lambda}}_k^{\top}, \text{for j} \neq k$$

$$\hat{\boldsymbol{\Sigma}}_{jj} = \begin{cases} \hat{\boldsymbol{\lambda}}_j\hat{\boldsymbol{\lambda}}_j^{\top} + \hat{\boldsymbol{\Theta}}_j, & \text{for reflective block} \\ \mathbf{S}_{jj}, & \text{for formative block} \end{cases}$$

SEM with factors and composites
Monte-Carlo Simulation

We generate 1000 samples of n =300 observations by random drawings from $\mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$.

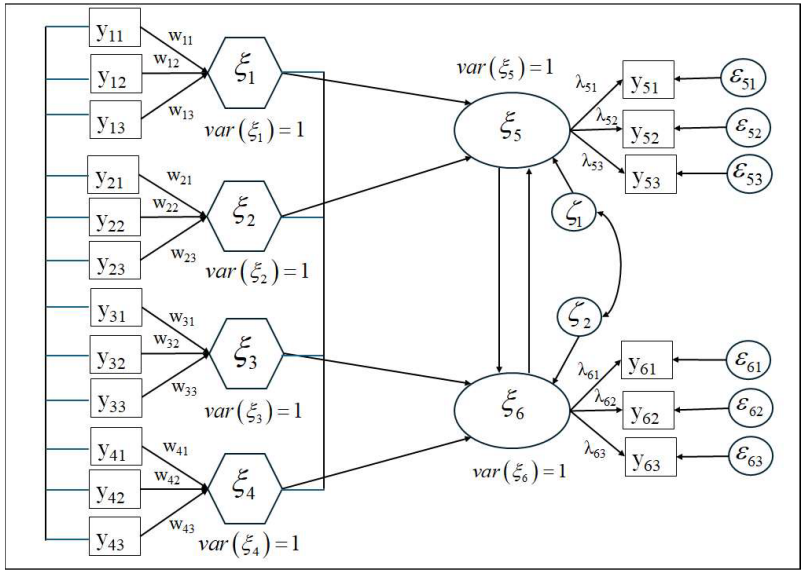


TABLE 2: Loading estimates.

		Mean of $\hat{\lambda}_{ij}(\ell)$, $\ell = 1,\ldots,1000$		Standard Deviation of $\hat{\lambda}_{ij}(\ell)$, $\ell = 1,\ldots,1000$		Mean of $\hat{\sigma}(\hat{\lambda}_{ij}^{\mathrm{RML}})[\ell]$, $\ell = 1,\ldots,1000$
Parameter	Population value	(1) SVD-SEM	(2) RML-SEM	(3) SVD-SEM	(4) RML-SEM	(5) RML-SEM
λ_{11}	.7316	.726	.728	.075	.069	.068
λ_{12}	.7746	.771	.771	.071	.066	.064
λ_{13}	.8176	.808	.810	.064	.059	.061
λ_{21}	.7316	.728	.728	.066	.066	.063
λ_{22}	.7746	.774	.774	.063	.063	.061
λ_{23}	.8176	.815	.814	.061	.060	.058
λ_{31}	.5763	.574	.574	.073	.071	.068
λ_{32}	.7822	.779	.779	.061	.058	.058
λ_{33}	.9057	.901	.902	.052	.051	.050
λ_{41}	.5764	.575	.574	.080	.078	.075
λ_{42}	.7822	.777	.777	.066	.064	.063
λ_{43}	.9057	.899	.900	.054	.053	.053
λ_{51}	.7	.699	.700	.071	.057	.057
λ_{52}	.7	.694	.695	.071	.058	.057
λ_{53}	.7	.699	.700	.069	.056	.057
λ_{61}	.7	.697	.698	.056	.054	.053
λ_{62}	.7	.700	.700	.058	.056	.053
λ_{63}	.7	.699	.699	.055	.055	.053
			MEAN	.065	.061	.060

SEM with factors and composites
Monte-Carlo Simulation

We generate 1000 samples of n =300 observations by random drawings from $\mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$.

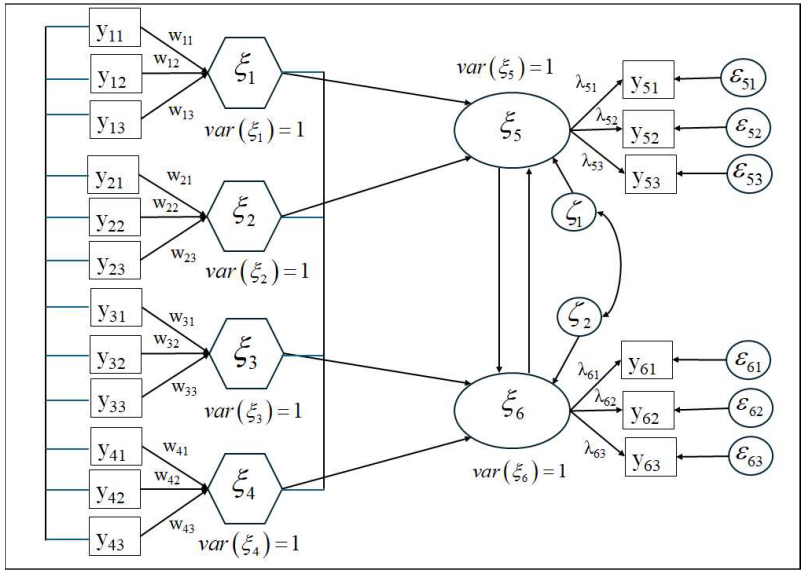


TABLE 3: Estimate of measurement error variances.

		Mean of estimate		Standard deviation of estimate		Mean of standard error of estimate
Parameter	Population value	SVD-SEM	RML-SEM	SVD-SEM	RML-SEM	RML-SEM
θ_{51}	.51	.505	.506	.078	.058	.056
θ_{52}	.51	.511	.510	.081	.056	.056
θ_{53}	.51	.508	.509	.081	.056	.056
θ_{61}	.51	.510	.508	.053	.050	.048
θ_{62}	.51	.506	.507	.050	.049	.048
θ_{63}	.51	.506	.507	.050	.047	.048
			MEAN	.066	.053	.052

SEM with factors and composites
Monte-Carlo Simulation

We generate 1000 samples of n =300 observations by random drawings from $\mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$.

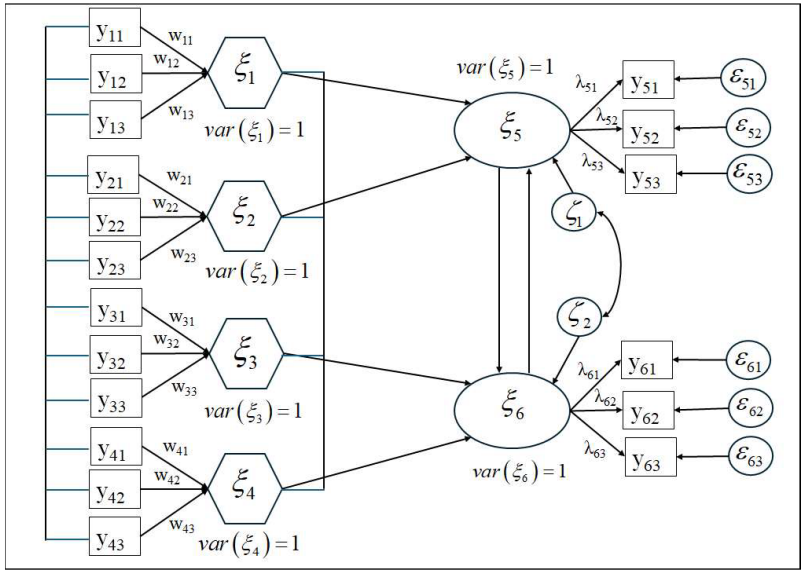


TABLE 4: Weight estimates for formative blocks.

		Mean of estimate		Standard deviation of estimate	
Parameter	Population value	SVD-SEM	RML-SEM	SVD-SEM	RML-SEM
w_{11}	.4303	.426	.427	.084	.072
w_{12}	.4303	.433	.432	.087	.076
w_{13}	.4303	.428	.430	.092	.079
w_{21}	.4303	.428	.429	.066	.065
w_{22}	.4303	.430	.431	.072	.071
w_{23}	.4303	.429	.428	.075	.074
w_{31}	.2058	.207	.207	.065	.059
w_{32}	.4117	.413	.413	.065	.059
w_{33}	.6175	.615	.616	.066	.060
w_{41}	.2058	.211	.208	.082	.078
w_{42}	.4117	.410	.410	.078	.075
w_{43}	.6175	.614	.616	.079	.077
			MEAN	.076	.070

SEM with factors and composites
Monte-Carlo Simulation

We generate 1000 samples of n =300 observations by random drawings from $\mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$.

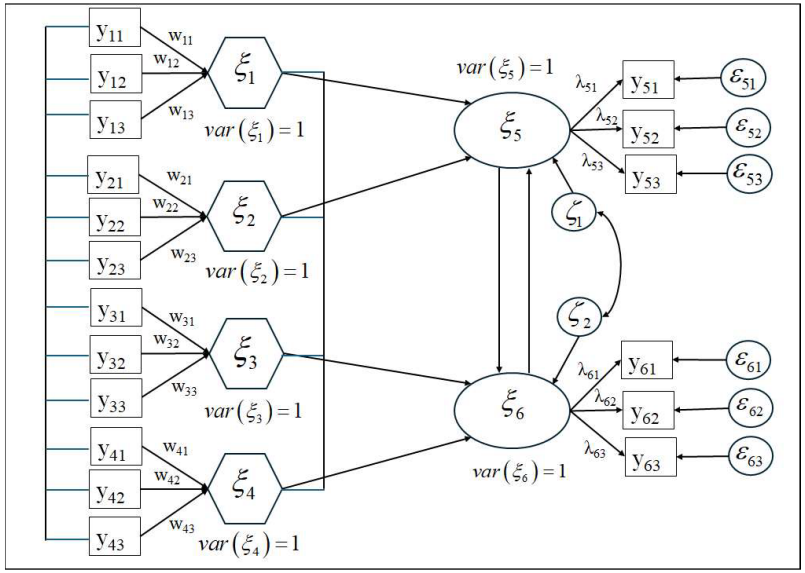


TABLE 6: Estimate of factor/composite correlations.

		Mean of estimate		Standard deviation of estimate		Mean of standard error of estimate
Parameter	Population value	SVD-SEM	RML-SEM	SVD-SEM	RML-SEM	RML-SEM
ρ_{12}	.5000	.504	.503	.043	.043	.043
ρ_{13}	.5000	.501	.500	.044	.044	.043
ρ_{14}	.5000	.502	.502	.043	.043	.043
ρ_{15}	.0500	.050	.047	.072	.069	NA
ρ_{16}	.4000	.401	.400	.054	.054	NA
ρ_{23}	.5000	.502	.501	.043	.043	.043
ρ_{24}	.5000	.503	.503	.043	.044	.043
ρ_{25}	.5071	.506	.507	.054	.053	NA
ρ_{26}	.6286	.630	.630	.043	.043	NA
ρ_{34}	.5000	.501	.500	.044	.043	.043
ρ_{35}	.2929	.294	.291	.060	.061	NA
ρ_{36}	.7714	.772	.772	.036	.035	NA
ρ_{45}	.2571	.259	.256	.052	.052	NA
ρ_{46}	.6286	.629	.628	.047	.045	NA
ρ_{56}	.7071	.706	.706	.050	.051	.052
			MEAN	.049	.048	
			MEAN in Φ_{endo} and Φ_{exo}	.044	.044	.044

SEM with factors and composites
Monte-Carlo Simulation

We generate 1000 samples of n =300 observations by random drawings from $\mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$.

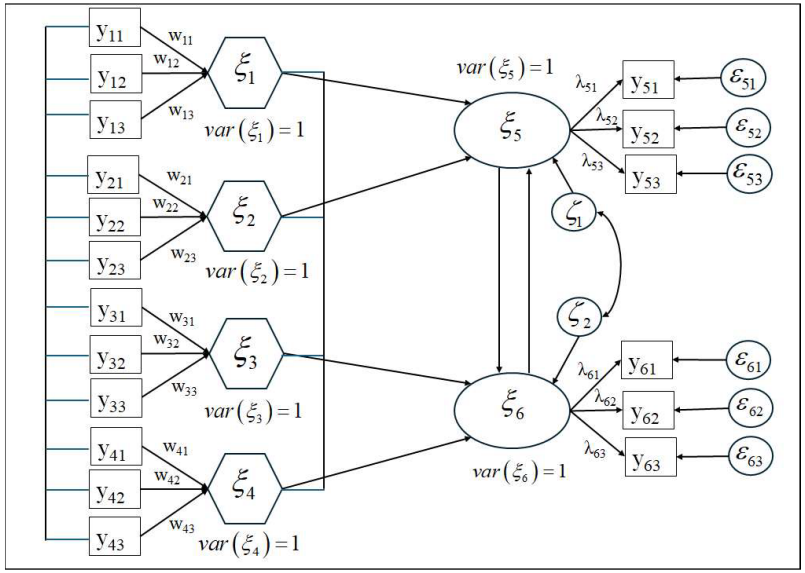
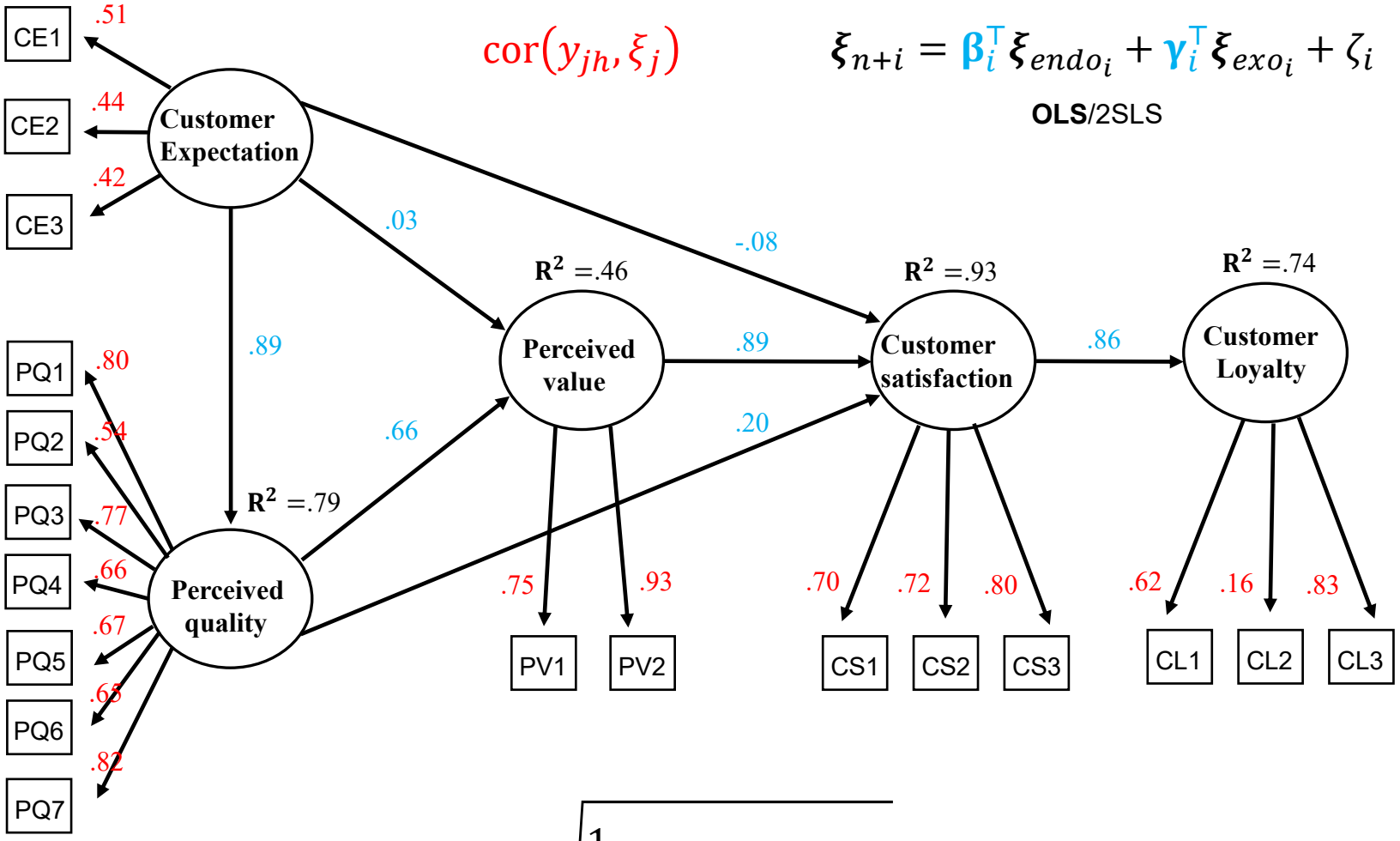


TABLE 5: Estimate of structural equations.

		Mean of estimate		Standard deviation of estimate		Mean of standard error of estimate
Parameter	Population value	SVD-SEM	RML-SEM	SVD-SEM	RML-SEM	RML-SEM
γ_{11}	-.3000	-.302	-.305	.065	.062	.060
γ_{12}	.5000	.501	.505	.078	.078	.078
γ_{23}	.5000	.500	.501	.053	.050	.050
γ_{24}	.2500	.249	.249	.054	.049	.047
β_{12}	.2500	.251	.246	.096	.094	.093
β_{21}	.5000	.502	.503	.086	.089	.087
			MEAN	.072	.070	.069
ψ_{11}	.5189	.515	.516	.077	.077	NA
ψ_{22}	.1054	.107	.107	.048	.048	NA
ψ_{12}	-.0295	-.031	-.030	.072	.074	NA
R_1^2	.4811	.485	.484	.077	.077	NA
R_2^2	.8946	.893	.893	.048	.048	NA

svdSEM for Customer Satisfaction

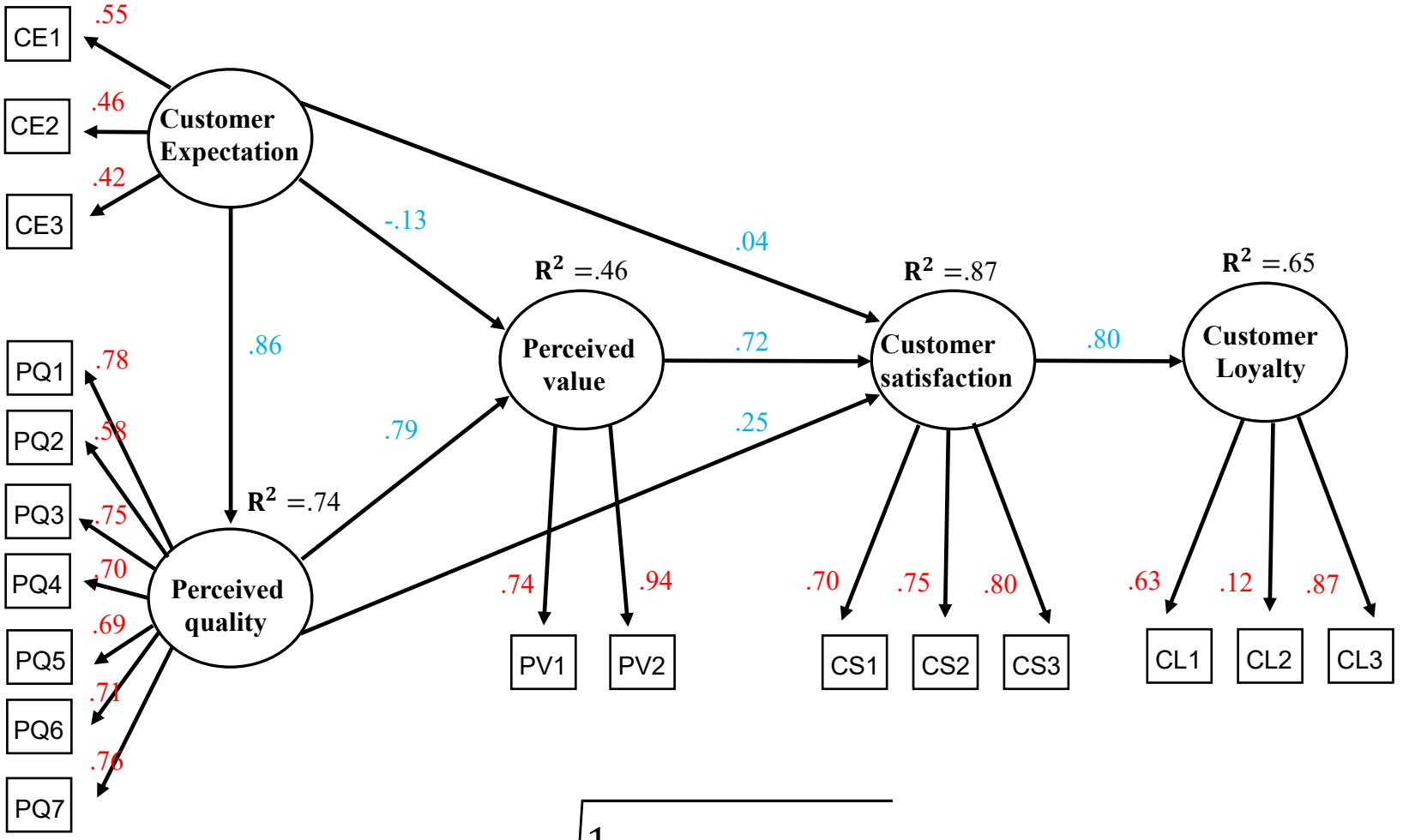
standardized results



$$d_{LS}(\hat{\boldsymbol{\Sigma}}, \mathbf{S}) = \sqrt{\frac{1}{2}\mathrm{trace}\left[(\hat{\boldsymbol{\Sigma}} - \mathbf{S})^2\right]} = 0.52$$

SEM-ML for Customer Satisfaction

Standardized results



$$d_{LS}(\hat{\boldsymbol{\Sigma}}, \mathbf{S}) = \sqrt{\frac{1}{2}\text{trace}\left[(\hat{\boldsymbol{\Sigma}} - \mathbf{S})^2\right]} = 0.46$$

Comparison between SVD-SEM and ML-SEM

LVs	Standardized loadings	svdSEM		ML-SEM	
		value	p-value	value	p-value
Customer expectation	λ_{11}	.512	.000	.549	.000
	λ_{12}	.438	.000	.460	.000
	λ_{13}	.416	.000	.423	.000
Perceived quality	λ_{21}	.804	.000	.775	.000
	λ_{21}	.542	.000	.575	.000
	λ_{23}	.774	.000	.750	.000
	λ_{24}	.660	.000	.707	.000
	λ_{25}	.670	.000	.691	.000
	λ_{26}	.652	.000	.706	.000
	λ_{27}	.819	.000	.756	.000
Perceived value	λ_{31}	.754	.000	.742	.000
	λ_{32}	.929	.000	.943	.000
Customer satisfaction	λ_{41}	.697	.000	.696	.000
	λ_{42}	.722	.000	.752	.000
	λ_{43}	.799	.000	.801	.000
Customer loyalty	λ_{51}	.623	.000	.625	.000
	λ_{52}	.157	.050	.118	.090
	λ_{53}	.830	.000	.865	.000

Comparison between svdSEM and ML-SEM (structural model)

Dependent variable	Explanatory variable	svdSEM		ML	
		coefficient	p-value	coefficient	p-value
Perceived quality	**Customer expectation**	.889	.000	.857	.000
Perceived value	**Customer expectation**	.027	.955	-.129	.619
	Perceived quality	.655	.161	.785	.001
Customer satisfaction	**Customer expectation**	-.081	.760	.038	.838
	Perceived quality	.893	.000	.715	.000
	Perceived value	.196	.030	.245	.000
Customer loyalty	**Customer satisfaction**	.862	.000	.804	.000

R-squared	svdSEM	ML-SEM
Perceived quality	.79	.74
Perceived value	.46	.46
Customer satisfaction	.93	.87
Customer loyalty	.74	.65

Conclusion

1. In addition to RML-SEM, we proposed a non-iterative SVD-based approach for obtaining consistent and asymptotically normal estimators of the parameters of SEM with factors and composites.

2. svdSEM and RML-SEM gives very close results.

3. svdSEM opens interesting avenues to be explored :
    - CAN estimators ⇒ test of significance can be derived,
    - SEM in high dimensional settings,
    - large number of blocks,
    - Several LVs per blocks,
    - Sparsity.

Conclusion

Advances in Data Analysis and Classification
https://doi.org/10.1007/s11634-025-00647-4

ORIGINAL ARTICLE

Structural equation modeling with factors and composites within the framework of the basic design

Arthur Tenenhaus[1] · Michel Tenenhaus[2] · Theo K. Dijkstra[3]

Received: 20 July 2024 / Revised: 4 December 2024 / Accepted: 19 March 2025

Abstract

In this paper, we study structural equation modeling with factors and composites within the framework of the basic design. The PLS approach (i.e. PLS-SEM based on PLS-PM, PLSc and the composite model) used to estimate this model has many limitations, including lack of theoretical foundations and the use of iterative algorithms with no guaranteed convergence. To address these shortcomings, we introduce a novel method, SVD-SEM, which corrects these issues. Indeed, SVD-SEM relies on a non-iterative SVD-based algorithm for parameter estimation, producing consistent and asymptotically normal estimators, offering a statistically and computationally sound alternative to PLS-SEM. Additionally, we present the restricted maximum-likelihood approach (RML-SEM) for the basic design with factors and composites. SVD-SEM can serve as an initial solution for RML-SEM. To illustrate the performance of these methods, we discuss a Monte Carlo simulation on a nonrecursive model with factors and composites.

R package: https://github.com/Tenenhaus/SEMFC

B & Γ in terms of R

- **OLS for recursive model**

$$\begin{pmatrix}\boldsymbol{\beta}_i \\ \boldsymbol{\gamma}_i\end{pmatrix} = \mathbb{E}\left[\boldsymbol{\xi}_i \boldsymbol{\xi}_i^{\top}\right]^{-1} \mathbb{E}\left[\boldsymbol{\xi}_i \xi_{n+i}\right]$$

- **2SLS for nonrecursive model**

$$\begin{pmatrix}\boldsymbol{\beta}_i \\ \boldsymbol{\gamma}_i\end{pmatrix} = \begin{pmatrix} \mathbb{E}\left[\boldsymbol{\xi}_{endo,i} \boldsymbol{\xi}_{exo}^{\top}\right] \mathbb{E}\left[\boldsymbol{\xi}_{exo} \boldsymbol{\xi}_{exo}^{\top}\right]^{-1} \mathbb{E}\left[\boldsymbol{\xi}_{exo} \boldsymbol{\xi}_{endo,i}^{\top}\right] & \mathbb{E}\left[\boldsymbol{\xi}_{endo,i} \boldsymbol{\xi}_{exo,i}^{\top}\right] \\ \mathbb{E}\left[\boldsymbol{\xi}_{exo,i} \boldsymbol{\xi}_{endo,i}^{\top}\right] & \mathbb{E}\left[\boldsymbol{\xi}_{exo,i} \boldsymbol{\xi}_{exo,i}^{\top}\right] \end{pmatrix}^{-1}$$

$$\times \begin{pmatrix} \mathbb{E}\left[\boldsymbol{\xi}_{endo,i} \boldsymbol{\xi}_{exo}^{\top}\right] \mathbb{E}\left[\boldsymbol{\xi}_{exo} \boldsymbol{\xi}_{exo}^{\top}\right]^{-1} \mathbb{E}\left[\boldsymbol{\xi}_{exo} \xi_{n+i}\right] \\ \mathbb{E}\left[\boldsymbol{\xi}_{exo,i} \xi_{n+i}\right] \end{pmatrix}$$

$\|\boldsymbol{\lambda}_j\|$ for reflective block in term of $\boldsymbol{\lambda}_j^*$ and $\boldsymbol{\Sigma}$

- We assume standardized LVs:

$$\text{var}(\xi_j) = \text{var}\left(\mathbf{w}_j^\top \boldsymbol{y}_j\right) = \mathbf{w}_j^\top \boldsymbol{\Sigma}_{jj} \mathbf{w}_j = 1$$

- From $\boldsymbol{\lambda}_j = cov\left(\boldsymbol{y}_j, \mathbf{w}_j^\top \boldsymbol{y}_j\right) = \boldsymbol{\Sigma}_{jj} \mathbf{w}_j \Rightarrow \boldsymbol{\lambda}_j^\top \boldsymbol{\Sigma}_{jj}^{-1} \boldsymbol{\lambda}_j = 1$

From $\boldsymbol{\lambda}_j^* = \dfrac{\boldsymbol{\lambda}_j}{\|\boldsymbol{\lambda}_j\|}$, it comes: $\|\boldsymbol{\lambda}_j\|^2 \boldsymbol{\lambda}_j^{*\top} \boldsymbol{\Sigma}_{jj}^{-1} \boldsymbol{\lambda}_j^* = 1$

$$\Rightarrow \|\boldsymbol{\lambda}_j\|^2 = \frac{1}{\boldsymbol{\lambda}_j^{*\top} \boldsymbol{\Sigma}_{jj}^{-1} \boldsymbol{\lambda}_j^*}, \qquad \text{for formative block.}$$

$\|\boldsymbol{\lambda}_j\|$ for formative block in term of $\boldsymbol{\lambda}_j^*$ and $\boldsymbol{\Sigma}$

The basic measurement model implies: $\boldsymbol{\Sigma}_{jj} = \boldsymbol{\lambda}_j \boldsymbol{\lambda}_j^{\top} + \boldsymbol{\Theta}_j$. Therefore, any of diagonal term $\sigma_{jj,hl}$ of $\boldsymbol{\Sigma}_{jj}$ can be decomposed as:

$$\sigma_{jj,hl} = \lambda_{jh}\lambda_{jl} = \|\boldsymbol{\lambda}_j\|^2 \boldsymbol{\lambda}_{jh}^* \boldsymbol{\lambda}_{jl}^* \quad \Rightarrow \quad \|\boldsymbol{\lambda}_j\|^2 = \frac{\sigma_{jj,hl}}{\boldsymbol{\lambda}_{jh}^* \boldsymbol{\lambda}_{jl}^*}$$

To avoid choosing a specific pair $h \neq l$, the slope of the least squares regression line can be used:

$$\|\boldsymbol{\lambda}_j\|^2 = \frac{\sum_{h<l}\left(\boldsymbol{\lambda}_{jh}^* \boldsymbol{\lambda}_{jl}^*\right)^2 \sigma_{jj,hl}}{\sum_{h<l}\left(\boldsymbol{\lambda}_{jh}^* \boldsymbol{\lambda}_{jl}^*\right)^2}, \qquad \text{for reflective block.}$$