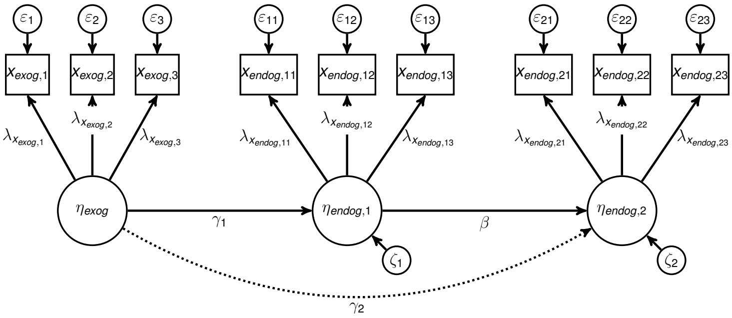
A Prediction Method for Ordinal Consistent Partial Least Squares

SIS 2019
Smart Statistics for Smart Applications,
Conference of the Italian Statistical Society,
Milan, 18-21 June 2019

Gabriele Cantaluppi
Florian Schuberth
Università Cattolica del Sacro Cuore, Milano
University of Twente, The Netherlands

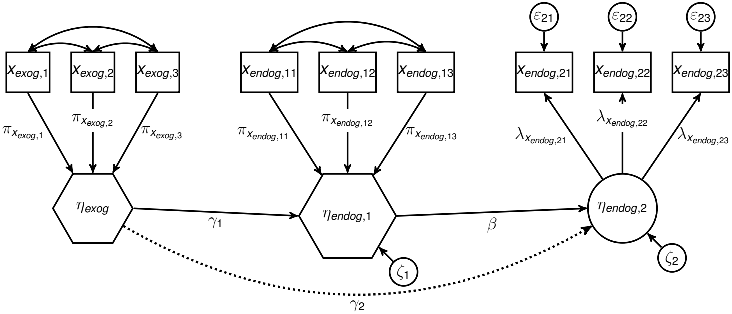
Outline

The PLS path model



$$\eta_{endog} = \mathbf{B}\eta_{endog} + \boldsymbol{\Gamma}\eta_{exog} + \zeta$$
$$\boldsymbol{X} = \boldsymbol{\Lambda}\eta + \varepsilon$$

- $\boldsymbol{X}$ manifest indicators
- η latent variables
- $\mathbf{B}$, $\boldsymbol{\Gamma}$ inner model structural relationships
- $\boldsymbol{\Lambda}$ outer model linear relationships - (common factors)



$$\boldsymbol{\eta}_{endog} = \mathbf{B}\boldsymbol{\eta}_{endog} + \boldsymbol{\Gamma}\boldsymbol{\eta}_{exog} + \boldsymbol{\zeta}$$

$$\boldsymbol{X} = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \quad \text{and} \quad \boldsymbol{\eta} = \boldsymbol{\Pi}\boldsymbol{X}$$

- $\boldsymbol{X}$ manifest indicators
- $\boldsymbol{\eta}$ latent variables
- $\mathbf{B}$, $\boldsymbol{\Gamma}$ inner model structural relationships
- $\boldsymbol{\Lambda}$ outer model linear relationships (common factors)
- $\boldsymbol{\Pi}$ outer model linear relationships (composite)

Partial least squares path modeling (PLS) is a variance-based estimator in structural equation modeling that

1. first creates linear combinations of observable indicators as stand-ins for the theoretical concepts
   (it obtains weights defining these linear combinations)
2. and subsequently estimates the model parameters[1].

[1] Herman Wold. "Path models with latent variables: The NIPALS approach". In: *Quantitative Sociology*. Ed. by H.M. Blalock et al. International Perspectives on Mathematical and Statistical Modeling. Academic Press, 1975. Chap. 11, pp. 307–357.

Input: $\mathbf{\Sigma}_{XX}$ (Covariance matrix of manifest indicators), $\mathbf{T}$ (Matrix representation of structural relationships among latent variables)

Initialisation: Choose $n+m$ arbitrary vectors of weights $\mathbf{w}_j^{(0)} = \left[0, \ldots, 0, w_{j1}^{(0)}, \ldots, w_{jp_j}^{(0)}, 0, \ldots, 0\right]'$, $j = 1, 2, \ldots, n+m$, each summing up to 1 and build the matrix $\mathbf{W}_{-1}^{(0)} = \left[\mathbf{w}_1^{(0)}, \ldots, \mathbf{w}_j^{(0)}, \ldots, \mathbf{w}_{n+m}^{(0)}\right]$.

For $s = 0, 1, \ldots$ (until convergence)

For $k = 0, 1, \ldots, m$: Set $\mathbf{W}_{\mathrm{TEMP}} = \mathbf{W}_{-1}^{(s)}$ and compute:

$$
\begin{aligned}
\mathbf{\Sigma}_{\hat{\eta}\hat{\eta},k}^{(s)} &= \mathbf{W}_{k-1}^{(s)\prime}\mathbf{\Sigma}_{XX}\mathbf{W}_{k-1}^{(s)} \\
{}_S\mathbf{W}_{k-1}^{(s)} &= \mathbf{W}_{k-1}^{(s)}\left[\Sigma_{\hat{\eta}\hat{\eta},k-1}^{(s)} * \mathbf{I}\right]^{-1/2} \\
\mathbf{\Sigma}_{\hat{\eta}\hat{\eta},k}^{(s)} &= {}_S\mathbf{W}_{k-1}^{(s)\prime}\mathbf{\Sigma}_{XX}\,{}_S\mathbf{W}_{k-1}^{(s)} \\
\mathbf{\Upsilon}_k &= (\mathbf{T}+\mathbf{T}') * sign\left(\mathbf{\Sigma}_{\hat{\eta}\hat{\eta},k}^{(s)}\right) \\
\mathbf{\Sigma}_{XZ,k}^{(s)} &= \mathbf{\Sigma}_{XX}\,{}_S\mathbf{W}_{k-1}^{(s)}\mathbf{\Upsilon}_k \\
\mathbf{\Sigma}_{X\hat{\eta},k}^{(s)} &= \mathbf{\Sigma}_{XX}\,{}_S\mathbf{W}_{k-1}^{(s)} \\
\mathbf{C}_k^{(s)} &= \varkappa_{{}_S\mathbf{W}_{k-1}^{(s)}} * \mathbf{\Sigma}_{XZ,k}^{(s)} \quad \text{and} \quad \pm = sign\left\{\mathbf{1}_p'\left[sign\left(\varkappa_{{}_S\mathbf{W}_{k-1}^{(s)}} * \mathbf{\Sigma}_{X\hat{\eta},k}^{(s)}\right)\right]\right\} \\
\mathbf{W}_k^{(s+1)} &= \mathbf{C}_k^{(s)}[diag(\mathbf{1}_{p+q}'\mathbf{C}_k^{(s)})]^{-1}diag(\pm) \\
\mathbf{W}_k^{(s+1)} &= \mathbf{W}_{k-1}^{(s)}\begin{bmatrix} \mathbf{O}_{n+k,n+k} & \mathbf{O}_{n+k,m-k} \\ \mathbf{O}_{m-k,n+k} & \mathbf{I}_{m-k,m-k} \end{bmatrix} + \mathbf{W}_k^{(s+1)}\begin{bmatrix} \mathbf{I}_{n+k,n+k} & \mathbf{O}_{n+k,m-k} \\ \mathbf{O}_{m-k,n+k} & \mathbf{O}_{m-k,m-k} \end{bmatrix}
\end{aligned}
$$

$$
\begin{cases} \text{If } k < m \text{ set } \mathbf{W}_k^{(s)} = \mathbf{W}_k^{(s+1)} \\ \text{If } k = m \text{ set } \mathbf{W}_{-1}^{(s+1)} = \mathbf{W}_m^{(s+1)} \text{ and proceed with next } s \end{cases}
$$

End
End

Motivation

- In recent years, PLS's predictive capacities have gained increasingly more attention[2].
- All major developments and extensions of PLS for predictive modeling assume metric observable indicators, but in empirical research often we have ordinal categorical indicators.

[2] Jan-Michael Becker, Arun Rai, and Edward Rigdon. "Predictive validity and formative measurement in structural equation modeling: Embracing practical relevance". In: *Proceedings of the International Conference on Information Systems*. 2013, pp. 1–19
Gabriel Cepeda Carrión et al. "Prediction-oriented modeling in business research by means of PLS path modeling: Introduction to a JBR special section". In: *Journal of Business Research* 69.10 (2016), pp. 4545–4551
Edward E Rigdon. "Rethinking partial least squares path modeling: In praise of simple methods". In: *Long Range Planning* 45.5 (2012), pp. 341–358
Galit Shmueli et al. "The elephant in the room: Predictive performance of PLS models". In: *Journal of Business Research* 69.10 (2016), pp. 4552–4564.

Several approaches have been developed to deal with the non-metric nature of the indicators, such as:

- non-metric partial least squares[3],
- partial maximum likelihood partial least squares[4],
- ordinal partial least squares (OrdPLS)[5], and
- ordinal consistent partial least squares (OrdPLSc)[6].

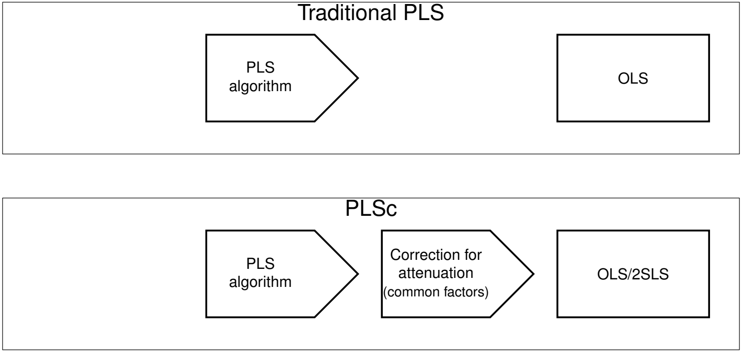
---

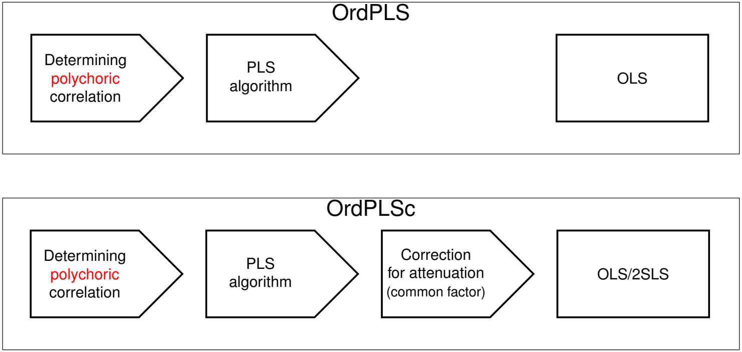
[3] Giorgio Russolillo. "Non-metric partial least squares". In: *Electronic Journal of Statistics* 6 (2012), pp. 1641–1669.

[4] Emmanuel Jakobowicz and Christian Derquenne. "A modified PLS path modeling algorithm handling reflective categorical variables and a new model building strategy". In: *Computational Statistics & Data Analysis* 51.8 (2007), pp. 3666–3678.

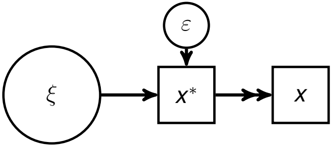
[5] Gabriele Cantaluppi. "A Partial Least Squares Algorithm Handling Ordinal Variables also in Presence of a Small Number of Categories". In: *arXiv preprint arXiv:1212.5049* (2012)
Gabriele Cantaluppi and Giuseppe Boari. "A Partial Least Squares Algorithm Handling Ordinal Variables". In: *The multiple facets of partial least squares and related methods: PLS, Paris, 2014*. Ed. by H. Abdi. Cham: Springer, 2016.

[6] Florian Schuberth and Gabriele Cantaluppi. "Ordinal Consistent Partial Least Squares". In: *Partial Least Squares Path Modeling: Basic Concepts, Methodological Issues and Applications*. Ed. by Hengky Latan and Richard Noonan. Cham: Springer, 2017
Florian Schuberth, Jörg Henseler, and Theo K. Dijkstra. "Partial least squares path modeling using ordinal categorical indicators". In: *Quality & Quantity* 52.1 (2018), pp. 9–35.

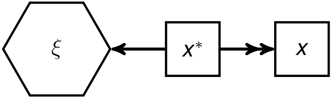

Traditional PLS
PLS algorithm
OLS
PLSc
PLS algorithm
Correction for attenuation (common factors)
OLS/2SLS



OrdPLS
Determining polychoric correlation
PLS algorithm
OLS
OrdPLSc
Determining polychoric correlation
PLS algorithm
Correction for attenuation (common factor)
OLS/2SLS



Exogenous common factor: ordinal categorical indicators

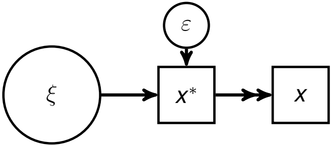


Exogenous composite model: ordinal categorical indicators

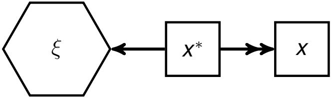
An ordinal categorical indicator x is assumed to be the outcome of a polytomized standard normally distributed latent random variable x^*:

$$x = m \quad \text{if} \quad \tau_{m-1} \leq x^* < \tau_m \quad m = 1, \ldots, M, \tag{1}$$

threshold parameters $\tau_0, ..., \tau_M$ determine observed categories.



Exogenous common factor: ordinal categorical indicators



Exogenous composite model: ordinal categorical indicators

An ordinal categorical indicator x is assumed to be the outcome of a polytomized standard normally distributed latent random variable x^*:

$$x = m \quad \text{if} \quad \tau_{m-1} \leq x^* < \tau_m \quad m = 1, \ldots, M, \tag{1}$$

threshold parameters $\tau_0, ..., \tau_M$ determine observed categories. Notation

$$\boldsymbol{X}_j^* \quad \rightarrow \quad \boldsymbol{X}_j. \tag{2}$$

will be used for the transformation corresponding to the block of indicators related to construct j.

Prediction Method for PLS and PLSc

Let $\boldsymbol{X}^{o}_{j,exog}$, $j = 1, ..., J_{exog}$ be a new set of observations of the metric indicators $\boldsymbol{x}_{j,exog}$:

Prediction Method for PLS and PLSc

Let $\boldsymbol{X}^{o}_{j,exog}$, $j = 1, ..., J_{exog}$ be a new set of observations of the metric indicators $\boldsymbol{x}_{j,exog}$:

- **Step 1** Predict scores of exogenous constructs

$$\hat{\eta}^{o}_{j,exog} = \boldsymbol{X}^{o}_{j,exog} \hat{\boldsymbol{w}}_j, \quad j = 1, \ldots, J_{exog}$$

Prediction Method for PLS and PLSc

Let $\boldsymbol{X}^o_{j,exog}$, $j = 1, ..., J_{exog}$ be a new set of observations of the metric indicators $\boldsymbol{x}_{j,exog}$:

- **Step 1** Predict scores of exogenous constructs

$$\hat{\eta}^o_{j,exog} = \boldsymbol{X}^o_{j,exog}\hat{\boldsymbol{w}}_j, \quad j = 1, \ldots, J_{exog}$$

- **Step 2** Predict the scores of the endogenous constructs in accordance with the structural model:

$$\hat{\boldsymbol{\eta}}^o_{endog} = (\mathbf{I} - \hat{\boldsymbol{B}})^{-1}\hat{\Gamma}\hat{\boldsymbol{\eta}}^o_{exog}$$

where $\hat{\boldsymbol{B}}$ and $\hat{\Gamma}$ contain path coefficient estimates based on the original dataset.

Prediction Method for PLS and PLSc

Let $\boldsymbol{X}^o_{j,exog}$, $j = 1, ..., J_{exog}$ be a new set of observations of the metric indicators $\boldsymbol{x}_{j,exog}$:

- **Step 1** Predict scores of exogenous constructs

$$\hat{\eta}^o_{j,exog} = \boldsymbol{X}^o_{j,exog}\hat{\boldsymbol{w}}_j, \quad j = 1, \ldots, J_{exog}$$

- **Step 2** Predict the scores of the endogenous constructs in accordance with the structural model:

$$\hat{\eta}^o_{endog} = (\mathbf{I} - \hat{\boldsymbol{B}})^{-1}\hat{\Gamma}\hat{\eta}^o_{exog}$$

  where $\hat{\boldsymbol{B}}$ and $\hat{\Gamma}$ contain path coefficient estimates based on the original dataset.

- **Step 3** Predict the the values of the observable metric indicators belonging to the endogenous constructs as

$$\hat{\boldsymbol{X}}^o_{j,endog} = \hat{\Lambda}\hat{\eta}^o_{endog}$$

Prediction Method for OrdPLS and OrdPLSc

We assume that the new set of observations $\boldsymbol{X}^{o}_{j,exog}$, $j = 1, ..., J_{exog}$ of the ordinal categorical indicators $\boldsymbol{x}_{j,exog}$ are generated according to:

$$\boldsymbol{X}^{*}_{j,exogTRUNC}(continuous) \quad \rightarrow \quad \boldsymbol{X}^{o}_{j,exog}. \qquad (3)$$

- The dataset $\boldsymbol{X}^{*}_{j,exogTRUNC}$ has the same correlation matrix as the polychoric correlation of exogenous manifest variables of the original dataset.
- The domain of $\boldsymbol{X}^{*}_{j,exogTRUNC}$ is defined by the same thresholds (τ_{j-1}, τ_{j}) obtained when computing the polychoric correlation matrix on the original dataset.

Step 1
Calculate the construct scores of the exogenous constructs

- PLS(c): construct scores are linear combination of observable indicators $\boldsymbol{x}_j$
- OrdPLS(c): construct scores are linear combinations of unobservable random variables $\boldsymbol{x}^*_{j,exogTRUNC}$

Step 1

Calculate the construct scores of the exogenous constructs

- PLS(c): construct scores are linear combination of observable indicators $\boldsymbol{x}_j$
- OrdPLS(c): construct scores are linear combinations of unobservable random variables $\boldsymbol{x}^*_{j,exogTRUNC}$

The distribution of the construct scores can be approximated by simulation.
A sufficient number of drawings ($n_{pred} = 100$) from each truncated multivariate normal distribution $\boldsymbol{X}^*_{j,exogTRUNC}$, corresponding to $\boldsymbol{x}^o_{j,exog}$, is used to calculate the linear combinations.

Step 1

Calculate the construct scores of the exogenous constructs

- PLS(c): construct scores are linear combination of observable indicators $\boldsymbol{x}_j$
- OrdPLS(c): construct scores are linear combinations of unobservable random variables $\boldsymbol{x}^*_{j,exogTRUNC}$

The distribution of the construct scores can be approximated by simulation.
A sufficient number of drawings ($n_{pred} = 100$) from each truncated multivariate normal distribution $\boldsymbol{X}^*_{j,exogTRUNC}$, corresponding to $\boldsymbol{x}^o_{j,exog}$, is used to calculate the linear combinations.
We obtain n_{pred} predictions of possible sets of construct scores for each subject, consistent with the observed categories:

$$\hat{\eta}^{o,p}_{j,exog} = \boldsymbol{X}^*_{j,exogTRUNC}\hat{\boldsymbol{w}}_j, \quad j = 1, \ldots, J_{exog}, \; p = 1, \ldots, n_{pred}. \tag{4}$$

Step 2

Predict the scores of the endogenous constructs in accordance with the structural model:

$$\hat{\eta}^{o,p}_{endog} = (\mathbf{I} - \hat{\boldsymbol{B}})^{-1} \hat{\Gamma} \hat{\eta}^{o,p}_{exog},$$

where $\hat{\boldsymbol{B}}$ and $\hat{\Gamma}$ contain path coefficient estimates based on the original dataset.
As an outcome, we obtain n_{pred} predicted values for the endogenous constructs.

Step 3

Predict the continuous latent variables underlying the categorical indicators belonging to the endogenous constructs.

A set of n_{pred} predictions for the indicators belonging to the j-th endogenous constructs can be obtained as

$$\hat{\boldsymbol{X}}^{*}_{j,endog} = \hat{\boldsymbol{\Lambda}}_j \hat{\boldsymbol{\eta}}^{o,p}_{j,endog}$$

Step 4

Predict the the values of the observable ordinal categorical indicators belonging to the endogenous constructs

To obtain predictions on ordinal scales, the n_{pred} (continuous) predictions of the components of $\hat{\boldsymbol{x}}^*_{j,endog}$ can be summarized by their mean or their median and subsequently transformed according to Equation 2

$$\boldsymbol{X}^*_j \quad \rightarrow \quad \boldsymbol{X}_j$$

by using the threshold parameter estimates based on the original dataset.

Moreover, a *mode* estimate of the category can be obtained from the n_{pred} estimates of the components of $\boldsymbol{x}^*_{j,endog}$ by considering the maximum empirical density on the intervals defined by the thresholds in Equation 2.

Monte Carlo Simulation

- Two population models
  - a model with three common factors and
  - a model with one common factor and two composites
- Different numbers of consecutive categories (2, 3, 4, 5, 7)
- Different ordinal categorical indicator distributions: symmetric, (alternating) moderate asymmetric, and (alternating) extreme asymmetric
- Different sample sizes (250 and 500 observations)
- 1,000 replications and evaluation on corresponding test datasets of size 1,000
- Each condition was estimated by OrdPLSc, PLSc, OrdPLS, and PLS, by removing estimates inadmissible with consistent methods
- PLS and PLSc predictions were rounded, to obtain values on the ordinal scale

Predictive performance evaluation

Mean squared error on the test dataset

$$MSE_k = \frac{1}{1000}\sum_{i=1}^{1000}\left(\hat{x}_{ij} - x_{ij}\right)^2, k = 1, 2, \ldots, 1000,$$

where x_{ij} represents the actual observed category of the j-th indicator for the i-th observation in the test data set and $\hat{x}_{ij}$ is the corresponding predicted category.

Predictive performance evaluation

Mean squared error on the test dataset

$$MSE_k = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{x}_{ij} - x_{ij})^2 \, , k = 1, 2, \ldots, 1000,$$

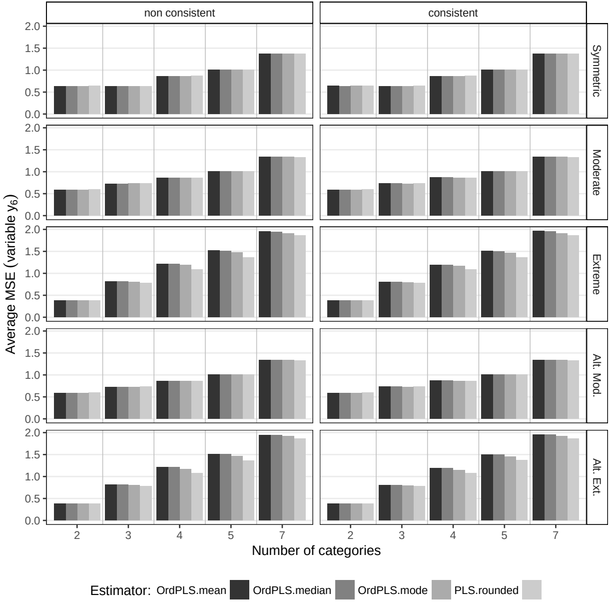
where x_{ij} represents the actual observed category of the j-th indicator for the i-th observation in the test data set and $\hat{x}_{ij}$ is the corresponding predicted category.

Proportion of correct predictions over all observations in each test data set

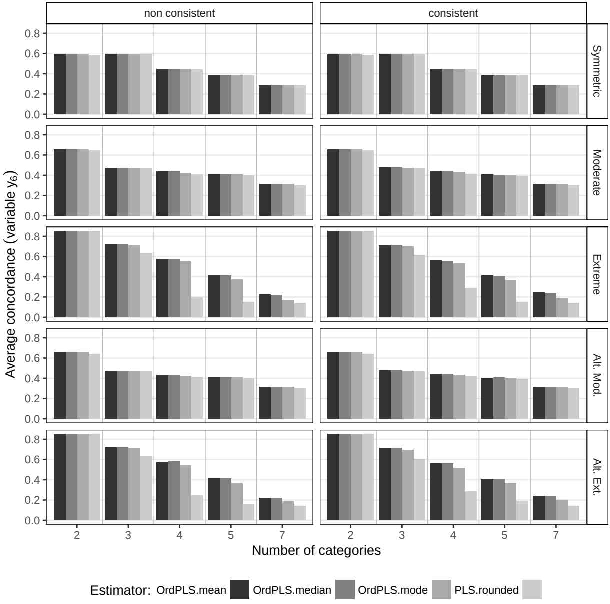
$$conc_k = \frac{1}{1000} \sum_{i=1}^{1000} I\left(\hat{x}_{ij}, x_{ij}\right) , \quad k = 1, 2, \ldots, 1000,$$

where $I(\cdot)$ is the indicator function that takes value 1 when $\hat{x}_{ij} = x_{ij}$ and 0 otherwise.

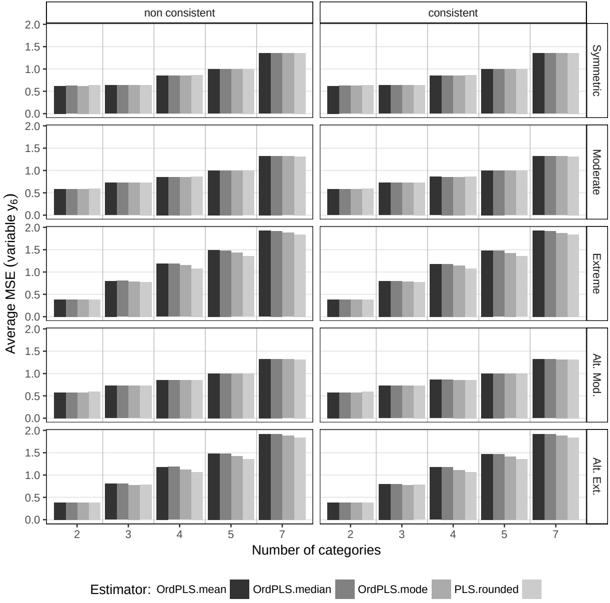
3 common factors model: *MSE* average values of predictions of indicator y_6.



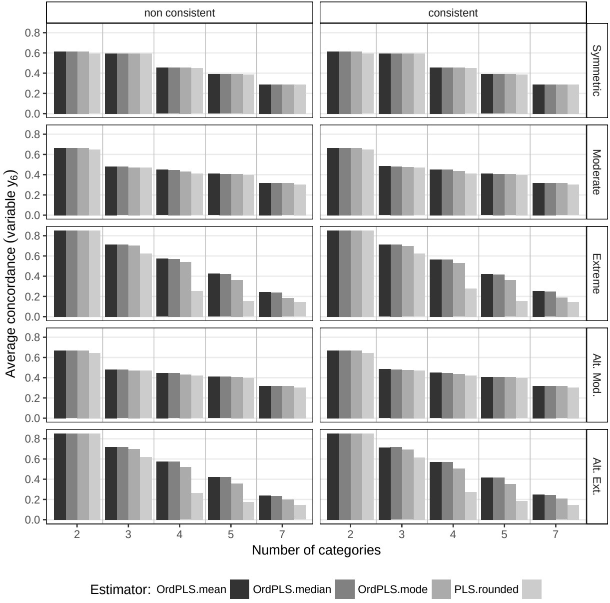
3 common factors model: *conc* average values of predictions of indicator y_6.



2 composites 1 common factor model: *MSE* average values of predictions of endogenous indicator y_6.



2 composites 1 common factor model: *concordance* average values of predictions of endogenous indicator y_6.



Conclusion

- The largest difference in *concordance* measures between rounded PLS(c) and OrdPLS(c) is present for (alternating) extreme distributions with 4-point scales.
- 4-point scales are relevant for empirical research, as they are encountered in questionnaires, e.g., when answers are coded as

very bad   bad   good   very good

- We can conclude that the prediction method for OrdPLS(c) appears to produce more reliable predictions than rounded PLS(c) methods in the case of (alternating) extreme distribution setting.