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

1. The PLS path model
2. Prediction Method for PLS and PLSc
3. Prediction Method for OrdPLS and OrdPLSc
4. Monte Carlo Simulation
5. Results and Conclusion
The PLS path model

\[ \eta_{\text{endog}} = B \eta_{\text{endog}} + \Gamma \eta_{\text{exog}} + \zeta \]

\[ \mathbf{X} = \Lambda \eta + \epsilon \]

- \( \mathbf{X} \) manifest indicators
- \( \eta \) latent variables
- \( B, \Gamma \) inner model structural relationships
- \( \Lambda \) outer model linear relationships - (common factors)
A Prediction Method for Ordinal Consistent Partial Least Squares

Gabriele Cantaluppi
Florian Schuberth

The PLS path model
Prediction Method for PLS and PLSc
Prediction Method for OrdPLS and OrdPLSc
Monte Carlo Simulation
Results and Conclusion

\[ \eta_{\text{endog}} = \mathbf{B} \eta_{\text{endog}} + \Gamma \eta_{\text{exog}} + \zeta \]

\[ \mathbf{X} = \Lambda \eta + \varepsilon \quad \text{and} \quad \eta = \Pi \mathbf{X} \]

- \( \mathbf{X} \) manifest indicators
- \( \eta \) latent variables
- \( \mathbf{B}, \Gamma \) inner model structural relationships
- \( \Lambda \) outer model linear relationships (common factors)
- \( \Pi \) 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\(^1\).

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

**Initialisation:** Choose $n + m$ arbitrary vectors of weights $w_j^{(0)} = [0, \ldots, 0, w_j^{(0)}, \ldots, w_j^{(0)}, 0, \ldots, 0]'$, $j = 1, 2, \ldots, n + m$, each summing up to 1 and build the matrix $W_0^{(0)} = [w_1^{(0)}, \ldots, w_j^{(0)}, \ldots, w_{n+m}^{(0)}]$.

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

For $k = 0, 1, \ldots, m$: Set $W_{\text{TEMP}} = W_{-1}^{(s)}$ and compute:

\[
\Sigma_{\hat{\eta}\hat{\eta},k} = W_{k-1}^{(s)'} \Sigma_{XX} W_{k-1}^{(s)}
\]

\[S W_{k-1}^{(s)} = W_{k-1}^{(s)} \left[ \Sigma_{\hat{\eta}\hat{\eta},k-1}^{-1} I \right]^{-1/2}
\]

\[\Sigma_{\hat{\eta}\hat{\eta},k} = S W_{k-1}^{(s)'} \Sigma_{XX} S W_{k-1}^{(s)}
\]

\[\Upsilon_k = (T + T') \ast \text{sign} \left( \Sigma_{\hat{\eta}\hat{\eta},k}^{(s)} \right)
\]

\[\Sigma_{XZ,k} = \Sigma_{XX} S W_{k-1}^{(s)} \Upsilon_k
\]

\[\Sigma_{X\hat{\eta},k} = \Sigma_{XX} S W_{k-1}^{(s)}
\]

\[C_k^{(s)} = \Upsilon_k S W_{k-1}^{(s)} \ast \Sigma_{XZ,k}^{(s)} \quad \text{and} \quad \pm = \text{sign} \left\{ 1_p' \left[ \text{sign} \left( \Upsilon_k S W_{k-1}^{(s)} \ast \Sigma_{X\hat{\eta},k}^{(s)} \right) \right] \right\}
\]

\[W_{k}^{(s+1)} = C_k^{(s)} \left[ \text{diag}(1_p' + q C_k^{(s)}) \right]^{-1} \text{diag}(\pm)
\]

\[W_{k}^{(s+1)} = W_{k-1}^{(s)} \left[ \begin{array}{cc} O_{n+k,n+k} & O_{n+k,m-k} \\ O_{m-k,n+k} & I_{m-k,m-k} \end{array} \right] + W_{k}^{(s+1)} \left[ \begin{array}{cc} I_{n+k,n+k} & O_{n+k,m-k} \\ O_{m-k,n+k} & O_{m-k,m-k} \end{array} \right]
\]

\[\begin{cases} \text{If } k < m \text{ set } W_k^{(s)} = W_k^{(s+1)} \\ \text{If } k = m \text{ set } W_{-1}^{(s+1)} = W_m^{(s+1)} \text{ and proceed with next } s \end{cases}
\]
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.


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)\(^6\).


A Prediction Method for Ordinal Consistent Partial Least Squares

Gabriele Cantaluppi
Florian Schuberth

The PLS path model
Prediction Method for PLS and PLSc
Prediction Method for OrdPLS and OrdPLSc
Monte Carlo Simulation
Results and Conclusion

1.9

Traditional PLS

PLS algorithm → OLS

PLSc

PLS algorithm → Correction for attenuation (common factors) → OLS/2SLS
A Prediction Method for Ordinal Consistent Partial Least Squares

Gabriele Cantaluppi
Florian Schuberth

The PLS path model
Prediction Method for PLS and PLSc
Prediction Method for OrdPLS and OrdPLSc
Monte Carlo Simulation
Results and Conclusion

1. OrdPLS
- Determining polychoric correlation
- PLS algorithm
- OLS

2. 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 $x^*$:

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

(1)

Threshold parameters $\tau_0, \ldots, \tau_M$ determine observed categories.
An ordinal categorical indicator $x$ is assumed to be the outcome of a polytomized standard normally distributed latent random variable $x^*$:

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

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

$$X^*_j \rightarrow X_j.$$ (2)

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

Prediction Method for PLS and PLSc

Let $X^o_{j,\text{exog}}, j = 1, \ldots, J_{\text{exog}}$ be a new set of observations of the metric indicators $x_{j,\text{exog}}$: 
Prediction Method for PLS and PLSc

Let $X_{j,exog}^o, j = 1, \ldots, J_{exog}$ be a new set of observations of the metric indicators $x_{j,exog}$:

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

$$\hat{\eta}_{j,exog}^o = X_{j,exog}^o \hat{w}_j, \quad j = 1, \ldots, J_{exog}$$
Prediction Method for PLS and PLSc

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

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

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

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

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

where $\hat{B}$ and $\hat{\Gamma}$ contain path coefficient estimates based on the original dataset.
Prediction Method for PLS and PLSc

Let $X_{j,\text{exog}}, j = 1, \ldots, J_{\text{exog}}$ be a new set of observations of the metric indicators $x_{j,\text{exog}}$:

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

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

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

  $$\hat{\eta}_{\text{endog}}^o = (I - \hat{B})^{-1} \hat{\Gamma} \hat{\eta}_{\text{exog}}^o$$

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

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

  $$\hat{X}_{j,\text{endog}}^o = \hat{\Lambda} \hat{\eta}_{\text{endog}}^o$$
Prediction Method for OrdPLS and OrdPLSc

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

$$X_{j,exog}^\ast_{TRUNC}(\text{continuous}) \rightarrow X_{j,exog}^o. \tag{3}$$

- The dataset $X_{j,exog}^\ast_{TRUNC}$ has the same correlation matrix as the polychoric correlation of exogenous manifest variables of the original dataset.
- The domain of $X_{j,exog}^\ast_{TRUNC}$ is defined by the same thresholds $(\tau_{j-1}, \tau_j)$ 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 $x_j$
- OrdPLS(c): construct scores are linear combinations of unobservable random variables $x^*_j, \text{exogTRUNC}$
Step 1
Calculate the construct scores of the exogenous constructs

- PLS(c): construct scores are linear combination of observable indicators $x_j$
- OrdPLS(c): construct scores are linear combinations of unobservable random variables $x_j^{*,\text{exogTRUNC}}$

The distribution of the construct scores can be approximated by simulation. A sufficient number of drawings ($n_{\text{pred}} = 100$) from each truncated multivariate normal distribution $X_j^{*,\text{exogTRUNC}}$, corresponding to $x_j^{o,\text{exog}}$, 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 $x_j$
- OrdPLS(c): construct scores are linear combinations of unobservable random variables $x^*_j, exogTRUNC$

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 $X^*_j, exogTRUNC$, corresponding to $x^o_j, exog$, 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:

$$\hat{\eta}^{o,p}_{j,exog} = X^*_j, exogTRUNC \hat{w}_j, \quad j = 1, \ldots, J_{exog}, \quad p = 1, \ldots, n_{pred}. \quad (4)$$
Step 2

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

\[ \hat{\eta}_{\text{endog}}^{o,p} = (I - \hat{B})^{-1} \hat{\Gamma} \hat{\eta}_{\text{exog}}^{o,p}, \]

where \( \hat{B} \) and \( \hat{\Gamma} \) contain path coefficient estimates based on the original dataset. As an outcome, we obtain \( n_{\text{pred}} \) 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 $n_{\text{pred}}$ predictions for the indicators belonging to the $j$-th endogenous constructs can be obtained as

$$\hat{X}_{j,\text{endog}}^* = \hat{\Lambda}_j \hat{\eta}_{j,\text{endog}}^{o,p}$$
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 $\hat{x}^*_j, endog$ can be summarized by their mean or their median and subsequently transformed according to Equation 2

$$x^*_j \rightarrow x_j$$

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 $x^*_j, endog$ 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

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

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

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

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

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

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 $y_6$.

![Graph showing average concordance for different estimators and number of categories](image_url)
2 composites 1 common factor model: $MSE$ average values of predictions of endogenous indicator $y_6$. 

![Bar chart](image-url)
2 composites 1 common factor model: \textit{concordance average values of predictions of endogenous indicator} $y_6$. 

![Bar charts showing 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.