PLS SEM Notes

Steven Hurwitt

April 21st, 2020

1 Alternative Estimation Methods

This is pulled from Jason Newsom's class notes for PSY 523 at Portland State University. Additional details were gathered from this thesis by Frances Chumney from the University of Nebraska.

To introduce some of the notation:

1. $\Sigma(\hat{\theta})$ is the covariance structure - a composite of the covariance matrix of the predictor, covariance matrix of the criterion and the covariance matrix of the predictor with the criterion.
2. $\hat{\theta}$ are the estimated parameters.
3. S is the covariance matrix observed in the data.
4. p is the number of observed indicators for the endogeneous latent factors, and q is the number of observed indicators for the exogeneous latent factors.

1.1 Maximum Likelihood (ML)

Maximum likelihood uses derivatives to minimize this fit function:

$$F_{ML} = \log \mid \Sigma(\hat{\theta}) \mid + tr\big(S\Sigma^{-1}(\hat{\theta})\big) - \log \mid S \mid - (p+q)$$

1.2 Generalized Least Squares (GLS)

Generalized least squares minimizes the discrepancy between S and Σ, but uses a weight matrix for the residuals, W (usually the inverse of the covariance S).

$$F_{GLS} = \left(\frac{1}{2}\right) tr\Big(\big[S - \Sigma(\hat{\theta})W^{-1}\big]^2\Big)$$

1.3 Weighted Least Squares (WLS)

Weighted least squares does not assume multivariate normality and instead is based on the variances and kurtosises (covariance of the covariance). This differs from unweighted least squares (ULS) and the above methods in that the choice of W differs.

$$F_{WLS} = \big(s - \sigma\big)^T W^{-1} \big(s - \sigma\big)$$

1.4 Unweighted Least Squares (ULS)

This method essentially seeks to minimize

$$\|S - \Sigma(\hat{\theta})\|^2$$

which can be judged by a Goodness of Fit Index (GFI) or CMIN criterion.

1.5 Diagonally Weighted Least Squares (DWLS)

This differs from WLS in that it uses a modified approach thatis a multiple-step estimation involving polychoric correlations as input to create the asymptotic covariance matrix used for weighting in the WLS estimation. This is usually paired with robust estimation adjustments that improve standard error, chi-square and fit indices.

2 SmartPLS vs. seminr

Thanks to the author's help on Github, I was able to match up the model for the Mobi dataset in both seminr and SmartPLS.

The R code is as follows:

```r
#create measurement model matrix
mobi_mm <- constructs(
  composite("Image",         multi_items("IMAG", 1:5), weights = mode_A),
  composite("Expectation",   multi_items("CUEX", 1:3), weights = mode_A),
  composite("Quality",       multi_items("PERQ", 1:7), weights = mode_A),
  composite("Value",         multi_items("PERV", 1:2), weights = mode_A),
  composite("Satisfaction", multi_items("CUSA", 1:3), weights = mode_A),
  composite("Complaints",   single_item("CUSCO"), weights = mode_A),
  higher_composite("HOC", c("Value", "Satisfaction"), orthogonal, mode_A),
  interaction_term(iv = "Image", moderator = "Expectation", method =  orthogonal, weights = mode_A),
  composite("Loyalty",      multi_items("CUSL", 1:3), weights = mode_A)
)

#create structural model
mobi_sm <- relationships(
  paths(from = "Image",        to = c("Expectation", "Satisfaction", "Loyalty")),
  paths(from = "Expectation",  to = c("Quality", "Value", "Satisfaction")),
  paths(from = "Quality",      to = c("Value", "Satisfaction")),
  paths(from = "Value",        to = c("Satisfaction")),
  paths(from = "Satisfaction", to = c("Complaints", "Loyalty")),
  paths(from = "Complaints",   to = "Loyalty")
)

#model estimation
mobi_pls <- estimate_pls(data = mobi,
                         measurement_model = mobi_mm,
                         structural_model = mobi_sm,
                         inner_weights = path_weighting)

summary(mobi_pls)

mobi_pls$path_coef
mobi_pls$outer_loadings
mobi_pls$outer_weights
mobi_pls$rSquared
```

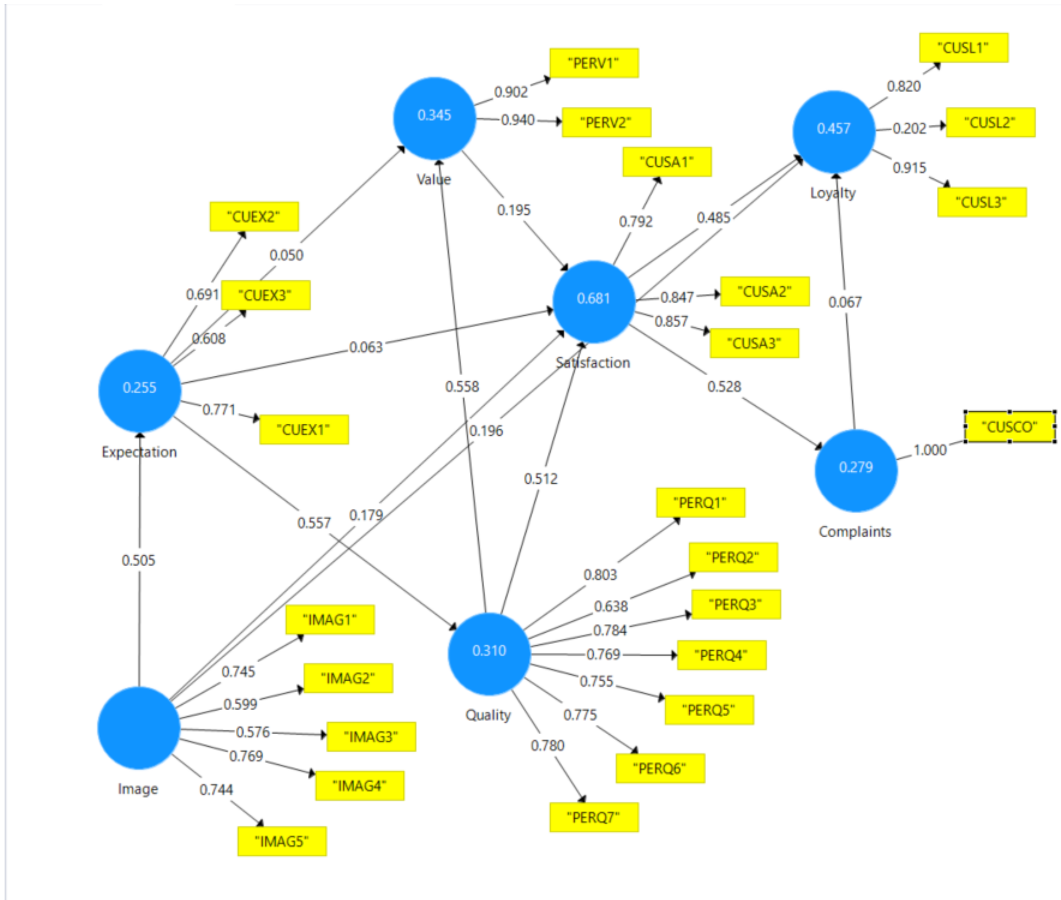
which gives the following path coefficients,

```
> mobi_pls$path_coef
             Image Expectation  Quality      Value Satisfaction Complaints    Loyalty
Image            0   0.5049139 0.000000 0.00000000   0.17873950  0.0000000 0.19575533
Expectation      0   0.0000000 0.556749 0.04998839   0.06252287  0.0000000 0.00000000
Quality          0   0.0000000 0.000000 0.55830438   0.51202394  0.0000000 0.00000000
Value            0   0.0000000 0.000000 0.00000000   0.19476510  0.0000000 0.00000000
Satisfaction     0   0.0000000 0.000000 0.00000000   0.00000000  0.5280662 0.48547762
Complaints       0   0.0000000 0.000000 0.00000000   0.00000000  0.0000000 0.06692607
Loyalty          0   0.0000000 0.000000 0.00000000   0.00000000  0.0000000 0.00000000
> 
```

and outer loadings.

```
> mobi_pls$outer_loadings
                 Image Expectation   Quality     Value Satisfaction Complaints   Loyalty
IMAG1        0.7452081   0.0000000 0.0000000 0.0000000    0.0000000          0 0.0000000
IMAG2        0.5992004   0.0000000 0.0000000 0.0000000    0.0000000          0 0.0000000
IMAG3        0.5763590   0.0000000 0.0000000 0.0000000    0.0000000          0 0.0000000
IMAG4        0.7687617   0.0000000 0.0000000 0.0000000    0.0000000          0 0.0000000
IMAG5        0.7444524   0.0000000 0.0000000 0.0000000    0.0000000          0 0.0000000
CUEX1        0.0000000   0.7707672 0.0000000 0.0000000    0.0000000          0 0.0000000
CUEX2        0.0000000   0.6912455 0.0000000 0.0000000    0.0000000          0 0.0000000
CUEX3        0.0000000   0.6078126 0.0000000 0.0000000    0.0000000          0 0.0000000
PERQ1        0.0000000   0.0000000 0.8031781 0.0000000    0.0000000          0 0.0000000
PERQ2        0.0000000   0.0000000 0.6381464 0.0000000    0.0000000          0 0.0000000
PERQ3        0.0000000   0.0000000 0.7837469 0.0000000    0.0000000          0 0.0000000
PERQ4        0.0000000   0.0000000 0.7694797 0.0000000    0.0000000          0 0.0000000
PERQ5        0.0000000   0.0000000 0.7547214 0.0000000    0.0000000          0 0.0000000
PERQ6        0.0000000   0.0000000 0.7746433 0.0000000    0.0000000          0 0.0000000
PERQ7        0.0000000   0.0000000 0.7798648 0.0000000    0.0000000          0 0.0000000
PERV1        0.0000000   0.0000000 0.0000000 0.9022112    0.0000000          0 0.0000000
PERV2        0.0000000   0.0000000 0.0000000 0.9396248    0.0000000          0 0.0000000
CUSA1        0.0000000   0.0000000 0.0000000 0.0000000    0.7924124          0 0.0000000
CUSA2        0.0000000   0.0000000 0.0000000 0.0000000    0.8470215          0 0.0000000
CUSA3        0.0000000   0.0000000 0.0000000 0.0000000    0.8566927          0 0.0000000
CUSCO        0.0000000   0.0000000 0.0000000 0.0000000    0.0000000          1 0.0000000
Value        0.0000000   0.0000000 0.0000000 0.0000000    0.0000000          0 0.0000000
Satisfaction 0.0000000   0.0000000 0.0000000 0.0000000    0.0000000          0 0.0000000
CUSL1        0.0000000   0.0000000 0.0000000 0.0000000    0.0000000          0 0.8204132
CUSL2        0.0000000   0.0000000 0.0000000 0.0000000    0.0000000          0 0.2020217
CUSL3        0.0000000   0.0000000 0.0000000 0.0000000    0.0000000          0 0.9154363
```

This corresponds with the results from SmartPLS:



3 PLS-PM vs. CB-SEM

Again, the authors provided helpful clarification on github. Essentially, they said:

PLS-PM cannot natively model reflective ”common factor” constructs like covariance-based SEM (CB-SEM) does. The closest that PLS-PM can come to modeling reflective constructs is by doing a post-hoc adjustment to get nearly the same weights/paths as LAVAAN/LISREL (see Dijkstra & Henseler 2015). The reflective() function in SEMinR will try to do these post-hoc adjustments to simulate common factor results.

Dijkstra, T. K., & Henseler, J. (2015). Consistent partial least squares path modeling. MIS quarterly, 39(2). Available at: https://pdfs.semanticscholar.org/7e56/cb95c8996a46c5dff13267a651f382a73567.pdf.

4 Data Simulation Procedure

The results in Section 2 seem promising regarding the abilities of seminr to replicate SmartPLS with a little more flexibility (reflective vs. composite factors and regression vs. classification weights). This helps us answer some of the questions from last week's notes - essentially if we use composite factors with mode A weights, we will arrive at the SmartPLS results.

This brings us to the main work in the deliverable - sequential data generation based on Factor Loadings, Cronbach's Alpha, Composite Reliability and AVE.

4.1 Simulated Raw Datasets

This is the part that I am admittedly most confused about - it seems the data will all be ratings, so we will need a way to generate these algorithmically based on the different factors.

To help further my understanding, I will break down the variables described in the study background:

1. Optimism - level of optimism (three items)
2. Perspective-taking - level of perspective-taking (five items)
3. Information Interpretation - level of information technology implementation (three items)
4. Team Performance - evaluation of team performance (three items)
5. LMX - perception of quality of relationship with their leaders (seven items)
6. Feedback Self-Efficacy - perception of feedback self-efficacy (three items)
7. IT Experience - evaluation of IT experience (three items)
8. Team Performance - evaluation of team performance (three items)

4.2 Data Questions

This brings me to some questions about the data:

1. What will the variables look like, are they indicator variables rated out of the total items?
2. Will they be generated as in the Latent Growth Model (based off of a normally distributed variable multiplied by the factor loadings)?

3. How do we account for Composite Reliability in the generated data? See this reference for a good break down on the formula (where λ_i are loadings and θ_i are item residuals):

$$\rho_c = \frac{\left(\sum \lambda_i\right)^2}{\left(\sum \lambda_i\right)^2 + \sum Var(\theta_i)}$$

4. Same for Cronbach's Alpha (measured for a quantity that's the sum of K components $X = Y_1 + ... + Y_k$):

$$\alpha = \frac{K}{K-1}\left(1 - \frac{\sum \sigma^2_{Y_i}}{\sigma^2_X}\right)$$

5. and Average Variance Extracted (AVE):

$$AVE = \frac{\sum \lambda_i^2}{\sum \lambda_i^2 + \sum Var(e_i)}$$

as AVE and CR depend only on the factor loadings and error variance.

4.3 Technical Report

This can be done with R markdown similar to the Latent Growth notes (Lai 2020) and last week's notes.

4.4 Validation using SmartPLS

This will also be pretty straightforward to verify, as long as we run PLS-PM with composite factors and mode A weights in seminr.

5 JASP: Potential for PLS-PM

I can reach out to the author of the JASP software and try to get some technical requirements on what would be needed to add PLS-PM functionality, possibly through seminr.