Partial Least Squares Structural Equation Modeling Ia

Introduction to PLS-SEM and model specification

Morten Berg Jensen

Department of Economics and Business

February 22, 2024

Outline

Outcome

This lecture will help you to understand

- The Partial Least Squares approach to structural equation modeling
- The use of reflective and formative indicators
- Specification of the structural model

Who am I

- Associate Professor at Department of Economics and Business Economics
- Research dealing with quantitative analysis of issues related to
  - Questionnaire data
  - Analysis of scanner data
  - Health economics
- Teaching experience includes
  - Customer Analytics
  - Business Data Analysis
  - Econometrics
  - Machine Learning
- Contact details
  - Email: mbj@econ.au.dk
  - Office: 2621(B)-1a

Teaching philosophy

- Teaching material
  - Literature
  - Slides
  - R files
  - Videos
- How to use the material
  - Literature - have a look before the lecture
  - Slides and R code - participate in the lecture, we present the concepts, the challenges, and hands-on exercises
  - Literature - read it carefully after the lecture/when preparing for the exam
  - Slides and R code - rehearse again the applications after the lecture/when preparing for the exam

Structural equation models – recap

- Structural equation models can be viewed as a combination of
  - Path analysis where we try to study the patterns of causation in a network
  - The analysis of latent variables (as in e.g. factor analysis)
- Thus, a structural equation model consists of two parts:
  - Measurement part, which links observed variables to latent variables – using several variables to measure a concept makes it more likely that all aspects of the concept are represented
  - Structural part, which links latent variables to each other via a system of equations
- The equations represent the researcher's hypothesis about causal relationships between variables

PLS and CB structural equation modeling

- Two types of modeling:
  1. Partial Least Squares SEM (PLS-SEM) which is primarily exploratory
  2. Covariance-Based SEM (CB-SEM) which is primarily confirmatory

PLS-SEM and CB-SEM

- CB-SEM (think of the lectures with Ana Alina)
  - Analyze the covariance matrix
  - Focus on covariation between indicators
  - Focus on theory testing and confirmation
  - Can accommodate circular relationships
- PLS-SEM
  - Latent variables are handled as linear combinations of observed variables
  - Latent variable scores act as proxies – can use latent variable scores in follow-up analysis
  - Focus on explaining the variance in the dependent variables
  - Focus on testing a theoretical framework from a prediction perspective
  - Exploratory research
  - Can be used when distributional assumptions are a concern

Examining the data I

Table 1.3 Minimum sample sizes for different levels of minimum path coefficients (p_{min}) and a power of 80%

p_{min}	Significance level		
	1%	5%	10%
0.05–0.1	1004	619	451
0.11–0.2	251	155	113
0.21–0.3	112	69	51
0.31–0.4	63	39	29
0.41–0.5	41	25	19

Source: Hair et al. (2022), Chap. 1; used with permission by Sage

Figure: Table 1.3

- Sample sizes are derived from inverse square root method
- Power: Probability to correctly reject a false null hypothesis

Examining the data II

- Missing values
  - Scarce knowledge on how advanced missing data procedures work for PLS-SEM
  - Remove an observation if it has many missing values for specific constructs
  - Mean replacement often used in literature
- PLS-SEM does not require data to be normally distributed
- Extreme non-normal data might inflate standard errors obtained from bootstrapping (more on bootstrapping in later lectures) which decrease the power of hypothesis testing
- Use metric scaled variables (ratio and interval scaled variables)
- We will not treat how to use binary coded variables

Diagrams and Notation I

- x_1, ..., x_{10} are called indicators, items or manifest variables
- e_7, ..., e_9, z_3, z_4 are errors and represent unexplained variance – are often omitted from figures
- The arrows show predictive relationships – with strong theoretical support, they can be interpreted causally

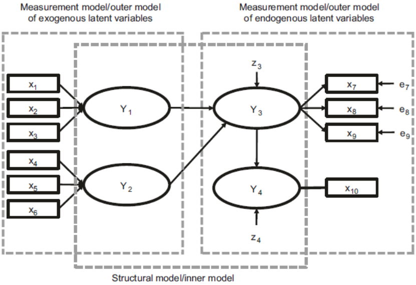

Fig. 1.1 A simple path model. (Source: Hair et al., 2022, Chap. 1; used with permission by Sage)

Figure: Fig 1.1

Diagrams and Notation II

- The arrows point away from $x_1, \ldots, x_6$ – these items have no error terms
- The arrows point toward $x_7, \ldots, x_9$ – these items have error terms
- x_{10} and Y_4 is connected by a line because Y_4 is measured by a single item – this item has no error term

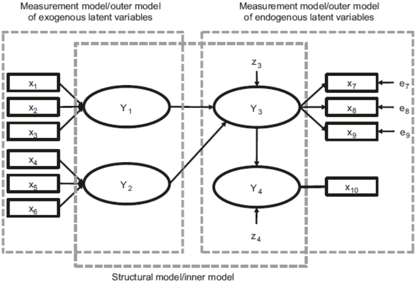

Fig. 1.1 A simple path model. (Source: Hair et al., 2022, Chap. 1; used with permission by Sage)

Figure: Fig 1.1

A systematic procedure for applying PLS-SEM

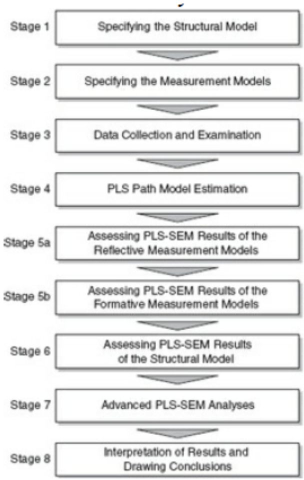

Figure: Procedure for applying PLS-SEM. (Source: Hair et al. 2022)

Specifying the structural model

- Specified from left (exogenous constructs) to right (endogenous constructs)
- Theory and logic should always be applied to specify the relations
- If theory is unclear, use best judgement
- Can specify competing models
- "*A parsimonious approach to theoretical specification is far more powerful than the broad application of a shotgun*" (Falk and Miller, 1992, p. 24; A Primer for Soft Modeling, University of Akron Press)
- No causal loops

Specifying the structural model – mediating effects

- Mediating effects are used to reflect the ”true” relationship between variables

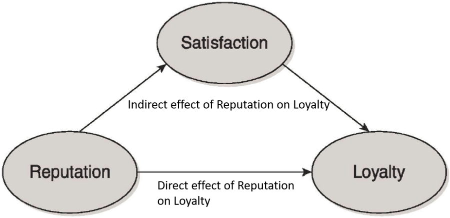

Figure: Structural model with mediating effects. (Source: Hair et al. 2022)

Specifying the measurement model I

- Hypothesis testing among structural relationships will only be valid and reliable if the measurement model accurately can explain the construct
- Often rely on prior research on how the construct should be measured

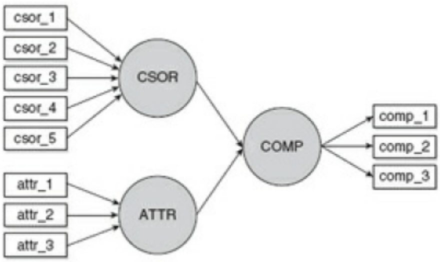

Figure: Model with reflective (right) and formative (left) measurements. (Source: Hair et al. 2022)

Specifying the measurement model II

- The latent variable scores in PLS-SEM are approximations to the constructs we try to measure
- E.g. let Y be a latent variable constructed as a weighted average of k indicators (a composite) – $Y = \sum_{i=1}^{k} w_i x_i$ where w_i are estimated weights
- How we estimate the weights, w_i, that depends on the type of measurement model we specify (reflective or formative)

Specifying the measurement model - reflective measurement models

- Representative sample of all the possible indicators available within the conceptual domain of the construct
- The construct causes the indicators
- Indicators are highly correlated
- Any single indicator can be left out as long as the construct has sufficient reliability
- A set of reflective measures is called a scale
- Often referred to as “Mode A” in PLS-SEM

Specifying the measurement model - formative measurement models

- Based on the assumption that causal indicators forms the construct by a linear combination
- Cannot remove or add indicators without changing the meaning of the construct
- Indicators need not be highly correlated, because they explain different aspects of the construct
- Often called a formative index
- Often referred to as "Mode B" in PLS-SEM

Specifying the measurement model - choosing between formative and reflective measurements I

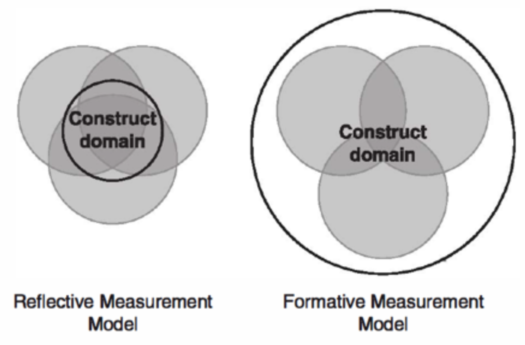

Figure: The conceptual difference between formative and reflective measures. (Source: Hair et al. 2022)

Specifying the measurement model - choosing between formative and reflective measurements II

- It is not fully resolved which measurement model is appropriate in certain situations – however use the following questions as guidelines
  - Theory testing or managerial recommendation?
  - Which way does the causality go (from indicator to construct or opposite)?
  - Are items mutually interchangeable?
  - Is the construct a trait that explain the indicators, or is the construct defined by the indicators?
  - We can use empirical means to determine the measurement perspective: Confirmatory tetrad analysis for PLS-SEM (CTA-PLS)

Specifying the measurement model - choosing between formative and reflective measurements III

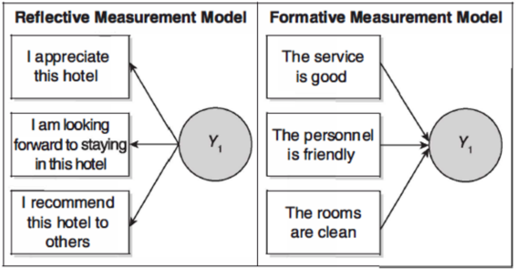

Figure: Satisfaction as a formatively and reflectively measured construct. (Source: Hair et al. 2022)

Specifying the measurement model – single-item measures

- Benefits
  - Ease of application
  - Promotes higher response rate (mitigating mental fatigue for respondents)
- Drawbacks
  - Harder to impute missing values
  - Does not remove measurement error
  - Can create validity problems for the construct measured by the single indicator
  - PLS-SEM is consistent at large, and one indicator is certainly not large
- Is appropriate when measuring observable characteristics – e.g. sales and profits

Customer satisfaction model I

- To familiarize you with the specification and estimation of models in R we will use a modified version of the European Customer Satisfaction Index (ECSI) model
- See chapter 2 in Hair et al. 2021 for references providing further descriptions
- The answers from 250 customers from a mobile telephone provider are provided in `satisfaction.csv`
- The reference model is shown on the next slide

Customer satisfaction model II

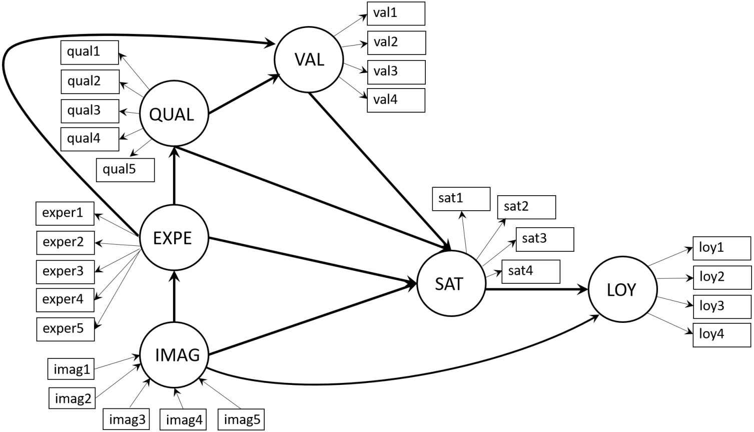

Figure: The model specified in the R script on Brightspace

Exercises

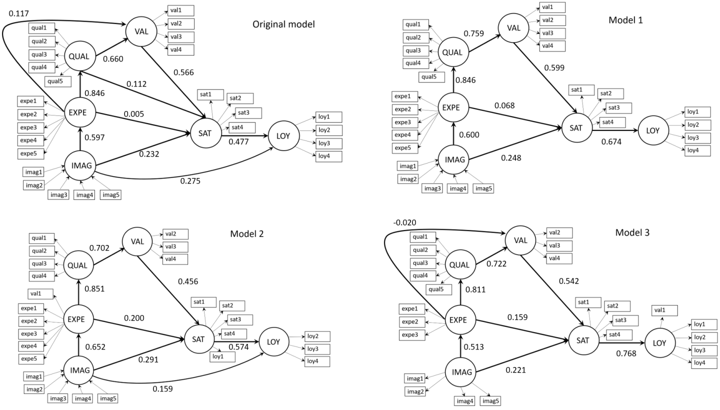

Figure: The models you should specify in R

- Modify the R script from Brightspace such that you estimate Models 1-3 above