Partial Least Squares Structural Equation Modeling II

Evaluation of reflective measurement models

Morten Berg Jensen

Department of Economics and Business

February 26, 2023

Outline

Outcome

This lecture will help you to understand

- The necessary steps for an assessment of the reflective measurement model
  - Indicator reliability
  - Internal consistency reliability
  - Convergent validity
  - Discriminant validity

Evaluation of measurement models

- Reliability and validity of the measurement model must be established before we can evaluate the structural model

$$x_k = t_k + \overbrace{\varepsilon_{k,r} + \varepsilon_{k,s}}^{\text{Measurement error}}$$

Measured value

True value

Random error

Systematic error

Threat to reliability

Threat to validity

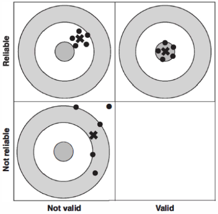

Figure: Comparing validity and reliability. (Source: Sarstedt and Mooi 2019)

Evaluation of the reflective measurement model - overview

- Indicator reliability
  - Size of correlation weights/loadings
- Internal consistency reliability
  - Cronbach's alpha
  - Composite reliability
- Convergent validity
  - Average variance extracted
- Discriminant validity
  - Fornell-Larcker criterion
  - Cross-loadings
  - Heterotrait-monotrait (HTMT) ratio of correlations

Reliability

- Seeks to answer the question: Do we have problems with a high level of random error?
- There are different types of reliability, but we will focus on
  - *Indicator reliability* – where we assess if the construct explains more variance of the indicator compared to what is left in the error term
  - *Internal consistency reliability* – where we assess the consistency of results across items for the same construct

Indicator reliability I

- High loadings indicate that different measures of a construct have much in common
- The squared loading of a standardized indicator tells us how much of the variance in an indicator is explained by the construct
- As a rule of thumb, we would like that a construct explains more than 50% of an indicators variance: loading > 0.708
- See next slide for (updated) recommendations

Indicator reliability II

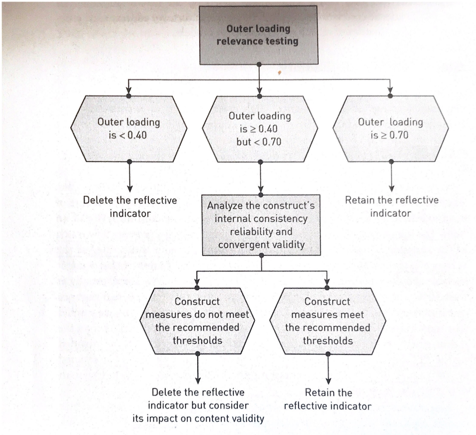

Figure: Outer loading relevance testing. (Source: Hair et al. 2022)

Internal consistency reliability – Cronbachs alpha I

- Cronbach's alpha: Based on the intercorrelations of the observed indicator variables

$$\text{Cronbach's } \alpha = \left(\frac{K}{K-1}\right)\left(1 - \frac{\sum_{k=1}^{K} s_k^2}{s_t^2}\right)$$

where K is the number of indicators for a specific construct, s_k^2 is the variance of indicator k and s_t^2 is the variance of the sum of all K indicators

Internal consistency reliability – Cronbachs alpha II

- Is equivalent to the average correlation between all indicator-split-half construct scores
- Assumes that all indicators have equal loadings on the construct
- Tends to underestimate the internal consistency reliability (conservative measure)

Internal consistency reliability – Composite reliability I

- Let us assume that we only have random error in our measurement of an indicator. For each indicator k and observation i we have

$$x_{i,k} = t_{i,k} + \varepsilon_{i,k,r}$$

  where $t_{i,k}$ is the true value, $\varepsilon_{i,k,r}$ is the random error and where we assume that $cov(t_{i,k}, \varepsilon_{i,k,r}) = 0$
- Let the true value of the indicator be related to our construct as

$$t_{i,k} = \mu_k + l_k Y_i$$

  where l_k is the loading of indicator k and μ_k is its mean.
- We can then write

$$x_{i,k} = \mu_k + l_k Y_i + \varepsilon_{i,k,r}$$

Internal consistency reliability – Composite reliability II

- Using our assumption from the previous lecture that $\mathbf{V}(Y_i) = 1$ we can decompose the variance of indicator k as

$$\mathbf{V}(x_{i,k}) = l_k^2 + \theta_k$$

where $\theta_k = \mathbf{V}(\varepsilon_{i,k,r})$

- Note, when we work with standardized indicators, we have $\theta_k = 1 - l_k^2$
- Composite reliability of indicator k is given as

$$\rho_k = \frac{l_k^2}{l_k^2 + \theta_k}$$

Internal consistency reliability – Composite reliability III

- The composite reliability for a construct measured via K indicators is given as

$$\rho_c = \frac{\left(\sum_{k=1}^{K} l_k\right)^2}{\left(\sum_{k=1}^{K} l_k\right)^2 + \sum_{k=1}^{K} \theta_k}$$

- In contrast to Cronbach's α, we do not assume equal loadings when using the composite reliability
- However, the composite reliability tends to overestimate the internal consistency reliability

Internal consistency reliability– wrap up

- Satisfactory ranges
  - Exploratory research: 0.6-0.7
  - Advanced stages of research: 0.7-0.9
- Values above 0.9 (definitely above 0.95) $\rightarrow$ likely that the indicators measure the same aspect of the construct. Hence, we only measure a small part of the construct domain, which threatens validity (e.g. using semantically redundant questions)
- Values below 0.6 $\rightarrow$ lack of internal consistency reliability
- The true reliability usually lies between Cronbach's α (lower bound) and the composite reliability (upper bound)

Validity

- Validity refers to the extent a measurement instrument (the indicators in our setting) captures what it is intended to measure (the construct in our setting)
- As with measures of reliability, there are many different ways to measure validity (and many types of validity)
- We will be concerned with convergent and discriminant validity
- However, you should keep content or face validity in the back of your mind when you empirically investigate your constructs
  - Does it make sense to use this set of indicators to measure this construct?

Convergent validity

- Convergent validity is the extent to which a measure correlates positively with alternative measures of the same construct
- Average variance extracted (AVE), is the average amount of variance in the indicators that the construct can explain
- $AVE = \frac{\sum_{k=1}^{K} l_i^2}{K}$
- $AVE > 0.5$, means that the construct explains more of the variance in its indicators, than the variance that is left unexplained in the error terms

Discriminant validity I

- If we have discriminant validity, we can say from an empirical standpoint, that the construct separates itself from other reflectively measured/single indicator constructs in the model
- Fornell-Larcker Criterion: $\sqrt{AVE}$ of a construct must be higher than the correlation with any other reflectively measured/single indicator construct
- However, Fornell-Larcker criterion does not detect problems with discriminant validity very well

Discriminant validity II

- Heterotrait-monotrait ratio (HTMT): Estimate of what the true correlation would be if the constructs where perfectly measured (i.e. perfectly reliable)
- If construct are conceptually very similar: HTMT< 0.9
- If constructs are conceptually more distinct: HTMT< 0.85
- This should be assessed for any combination of reflectively measured/single item constructs
- Can test HTMT ratio using bootstrap confidence interval. If 1 is inside the confidence interval $\rightarrow$ indicates lack of discriminant validity (more on bootstrap in PLS-SEM later)

Discriminant validity III

$$HTMT = 0.632 = \frac{\frac{1}{9}(0.426 + 0.339 + 0.393 + 0.423 + 0.345 + 0.385 + 0.274 + 0.235 + 0.25)}{\sqrt{\frac{1}{3}(0.77 + 0.701 + 0.665)\,\frac{1}{3}(0.574 + 0.318 + 0.335)}}$$

Average(heterotrait-heteromethod correlations)

Average(monotrait-heteromethod correlation of Y_1)

Average(monotrait-heteromethod correlation of Y_2)

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	1					
x_2	0.77	1				
x_3	0.701	0.665	1			
x_4	0.426	0.339	0.393	1		
x_5	0.423	0.345	0.385	0.574	1	
x_6	0.274	0.235	0.250	0.318	0.335	1

Figure: Visual representation of the HTMT approach. (Source: Hair et al. 2022)

Discriminant validity IV

- Ways to decrease HTMT
  - Increase average monotrait-heteromethod correlation by removing indicators which have low correlation with other indicators of the same construct
  - Decrease average heteromethod-heterotrait correlation by either:
    1. Remove indicators which are highly correlated with indicators from the opposing construct
    2. Reassign indicators to other constructs
- If theory supports it, it is also possible to merge problematic constructs
- When removing/changing indicators or merging/splitting constructs, it is **very important** that this make sense theoretically or conceptually so we preserve content validity

Discriminant validity V

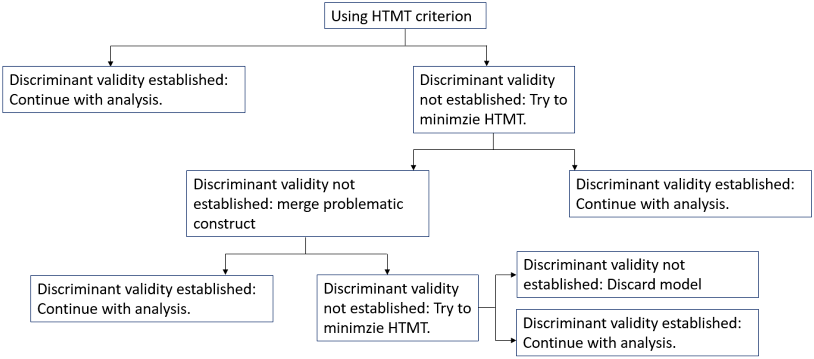

Figure: Handling discriminant validity problems. (Source: Hair et al. 2022)

Corporate reputation model – reflective measurement evaluation I

- Indicator reliability
  - All indicator loadings are above the threshold of 0.7
  - comp_2 has the lowest indicator reliability with a value of 0.638 ($= 0.798^2$) well above 0.5
- Internal consistency reliability
  - Values are within the satisfactory range of 0.7-0.95 and close to the recommended range of 0.8-0.9
  - As expected, composite reliability estimates are higher than Cronbach's alpha
- Convergent validity
  - All reflective constructs have values of AVE above 0.5

Corporate reputation model – reflective measurement evaluation II

- Discriminant validity – Fornell-Larcker
  - Each constructs square root of AVE is higher than the constructs highest correlation with the other reflectively measured/single indicator constructs
- Discriminant validity – HTMT
  - All HTMT ratios are below the conservative treshold of 0.85
  - Using a 95% bootstrap confidence interval, we find that no intervals contain 1 – in fact the upper limit of all intervals are below 0.85

Exercises

- Complete exercises 1 and 2 on page 89 in Hair et al. 2021