Some unsolved problems in structural equation models –
Incorporating formative indicators into covariance-based SEM

Dirk Temme
Lutz Hildebrandt

Humboldt University Berlin

Meeting SFB 649 “Economic Risk”
Motzen, May 8-10, 2008
1. Motivation

Discussion on the “correct” specification of measurement models

- Until the end of the 1990s, scale development in marketing research mainly based on the “Churchill paradigm” (Churchill 1979) → reflective measurement model as default → “LISREL” approach (e. g., Anderson/Gerbing 1982).

- “Churchill paradigm” has also been applied to theoretical concepts in management and organizational research (e. g., “strategies”, “market orientation”) as well (Shook et al. 2004).

- Recently, (mis-)specification of measurement models “hot topic” in marketing:
  - Diamantopoulos/Winklhofer (2001)
  - Rossiter (2002)
  - Albers/Hildebrandt (2006)

→ many constructs should better be modelled in a formative manner
→ treating them as reflective can produce grossly misleading results
1. Motivation

Formative measurement model

a) Formative construct operationalized as a composite variable (CV; McDonald 1996):

\[ \eta_j = \gamma_{j1} x_1 + \gamma_{j2} x_2 + \ldots + \gamma_{jQ} x_Q \]

b) Formative construct operationalized as a composite latent variable (CLV; Jarvis et al. 2003):

\[ \eta_j = \gamma_{j1} x_1 + \gamma_{j2} x_2 + \ldots + \gamma_{jQ} x_Q + \zeta_j \]
1. Motivation

Example of a formative construct – Job satisfaction (Churchill et al. 1974)

\[ \zeta_1 \]

\[ \eta_1 \]

\[ x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \quad x_6 \quad x_7 \]

- \( x_1 \): The job
- \( x_2 \): Fellow workers
- \( x_3 \): Supervision
- \( x_4 \): Company policy and support
- \( x_5 \): Pay
- \( x_6 \): Promotion and advancement
- \( x_7 \): Customers
1. Motivation

Estimation of formative measurement models – Alternative approaches

- Discussion on measurement models (as well as new software programs, see Temme/Kreis/Hildebrandt 2008 for an overview) has stimulated renewed interest in partial least squares (PLS, Wold 1966) analysis (e.g., Bliemel et al. 2005; Vinci et al. 2008, Handbook on PLS & Marketing).

- Advantages of partial least squares compared to covariance structure analysis:
  - identification of formative measurement models not an issue
  - scores for (formative) constructs are a core element in parameter estimation
  - no distributional assumptions
  - small samples
1. Motivation

Wide-spread opinions about specifying formative measurement models

• “The PLS method handles both types of indicators [i.e. formative and reflective indicators], whereas other path analytical methods (e.g., LISREL, EQS) can handle only reflective indicators.” (Dellande/Gilly/Graham 2004, JoM, p. 84)

• “[In covariance structure analysis] all indicators must be treated in a reflective manner where they are causally influenced by an underlying construct.” (Chin 1998, p. 301)

• “We chose PLS because of the formative nature of the higher-order value construct.” (Ulaga/Eggert 2006, JoM, p. 129)

• “A PLS analysis is most appropriate when the model incorporates both formative and reflective indicators [...].” (White/Varadarajan/Dacin 2003, JoM, p. 71)

• “Because the latent variable scores are determinate, PLS can be used to model formative indicators [...].“ (Reinartz/Krafft/Hoyer 2004, JoMR, p. 298)
1. Motivation

Estimation of formative measurement models – Alternative approaches (cont.)

- **But:** Covariance structure analysis (“LISREL”) can handle formative constructs too (e.g., MIMIC models, Jöreskog/Goldberger 1975)!

- **Disadvantages** of PLS compared to covariance structure analysis:
  - parameter estimates are known to be biased for all practical purposes (e.g. Dijkstra 1983),
  - does not test for model constraints → **no theory testing**, 
  - no global fit criteria (except GoF, Tenenhaus et al. 2005),
  - assumes that formative constructs can be measured without error → construct operationally defined, no surplus meaning,
  - does not separate between “direct” and “indirect“ effects of explanatory variables on a formative construct.

- **→ Good reasons to use “LISREL”** provided that the model can be identified.
Outline

1. Motivation

2. Identification of recursive SEM including formative constructs – A new approach

3. Specification of SEM including endogenous formative constructs
   3.1 Exogenous vs. endogenous indicators for endogenous CLV
   3.2 Empirical study

4. Conclusions
Identification of recursive SEM including formative constr.

Identification conditions (e.g., Temme 2006)

Basic identification conditions:

**t-rule**: \[ t \leq \frac{1}{2}(p + q)(p + q + 1) \]

Scaling rule:

1. Fix the path from a formative indicator to the CLV (e.g., to 1).

2. Fix the path from the CLV to one of its reflective indicators or to a latent variable measured by reflective indicators.

3. Fix the variance of the CLV to 1.

Further identification conditions (as stated in the literature):

- specifying at least two additional reflective indicators (MIMIC models, Jöreskog/Goldberger 1975) or, alternatively

- specifying direct effects on at least two unrelated reflective latent variables (MacCallum/Browne 1993; Bollen/Davis 1994; Jarvis/MacKenzie/Podsakoff 2003)
2. Identification of recursive SEM including formative constr.

Smallest identified formative measurement model

\[ \eta_1^F = 1 \cdot x_1 + \gamma_{12} x_2 + \cdots + \gamma_{1Q_1} x_{Q_1} + \zeta_1, \]

\[ \eta_2^R = \beta_{21} \eta_1^F + \zeta_2, \]

\[ \eta_3^R = \beta_{31} \eta_1^F + \zeta_3, \]

where \( \eta_2^R \) and \( \eta_3^R \) are reflective constructs which have identified measurement models (e.g., three reflective indicators each).
2. Identification of recursive SEM including formative constr.

More complex models

Identified models

Non-identified models

$F = \text{formative measurement model}$

$R = \text{reflective measurement model}$
Further identification options?

Does identification of formative constructs always require “at least two unrelated latent constructs with reflective indicators” (Jarvis/MacKenzie/Podsakoff 2003)?
2. Identification of recursive SEM including formative constr.

Example: Model by Law/Wong (1999)
“[…] the error term for the job perception construct was not identified […]. This is because the job perception construct did not have two paths emanating from it that led to independent constructs. It had two paths leading from it, but the two constructs were causally related.” (MacKenzie/Podsakoff/Jarvis 2005, p. 716)

But: Model by Law/Wong (1999) is identified!
2. Identification of recursive SEM including formative constr.

Assessing identification of models with structural relations between directly influenced latent constructs

Three-step graphically-oriented approach (see also Cantaluppi 2002):

1. Transform the original graph into a graph without CLV.
2. Check identification of the transformed graph.
3. Check that the parameters of the original model can be unambiguously derived from the parameters of the transformed model.
Step 1: START (original model)

2. Identification of recursive SEM including formative constr.
2. Identification of recursive SEM including formative constr.

Step 1: END (transformed model)

- Autonomy
- Identity
- Feedback
- Significance
- Variety

Job satisfaction

LMX

Liking

Turnover intention

non-recursive!
2. Identification of recursive SEM including formative constr.

Step 2: Graphic proof of identification (*Rigdon* 1995) → Transformed model is identified

All shared predictors of job satisfaction and turnover intention eliminated
2. Identification of recursive SEM including formative constr.

Step 3: Algebraic proof of identification → Original model is identified

\[ \eta_2 = 1 \cdot \beta_{21} x_1 + \gamma_{12} \beta_{21} x_2 + \gamma_{13} \beta_{21} x_3 + \gamma_{14} \beta_{21} x_4 + \gamma_{15} \beta_{21} x_5 + \gamma_{26} x_6 + \gamma_{27} x_7 + (\beta_{21} \zeta_1 + \zeta_2) \rightarrow \text{All coefficients are identified} \]

\[ \eta_3 = 1 \cdot \beta_{31} x_1 + \gamma_{12} \beta_{31} x_2 + \gamma_{13} \beta_{31} x_3 + \gamma_{14} \beta_{31} x_4 + \gamma_{15} \beta_{31} x_5 + \gamma_{16} \beta_{31} x_6 + \gamma_{37} x_7 + \beta_{32} \eta_2 + (\beta_{31} \zeta_1 + \zeta_3) \rightarrow \text{All coefficients are identified} \]

\[ \psi^*_{23} = \beta_{21} \beta_{31} \psi_{11} \rightarrow \psi_{11} = \frac{\psi^*_{23}}{\beta_{21} \beta_{31}} \]

\[ \psi^*_{22} = \beta_{21}^2 \psi_{11} + \psi_{22} \rightarrow \psi_{22} = \psi^*_{22} - \beta_{21}^2 \psi_{11} \]

\[ \psi^*_{33} = \beta_{31}^2 \psi_{11} + \psi_{33} \rightarrow \psi_{33} = \psi^*_{33} - \beta_{31}^2 \psi_{11} \]

All error variances are identified.
Outline

1. Motivation

2. Identification of recursive SEM including formative constructs – A new approach

3. Specification of SEM including endogenous formative constructs
   3.1 Exogenous vs. endogenous indicators for endogenous CLV
   3.2 Empirical study

4. Conclusions

Types of formative CLV and their specification

1. **Quasi exogenous CLV**: Only influenced by their formative indicators.
   → Covariances between formative indicators and other exogenous variables are set free.

2. **Endogenous CLV**: Influenced also by other variables (e.g., exogenous latent variables representing reflective constructs).
   → Those few methodological articles discussing endogenous CLV (e.g., MacCallum/Browne 1993) recommend specifying formative indicators as exogenous variables.
3.1 Exogenous vs. endogenous indicators for endogenous CLV

Issues resulting from specifying exogenous indicators for endogenous CLV

• **Conceptual issue:** How can explanatory variables influence an endogenous CLV without influencing its formative indicators (e.g., how can a salary increase positively impact on an employee’s job satisfaction without improving his/her satisfaction with pay)?

• **Substantial issue:** Specifying the formative indicators of an endogenous CLV as exogenous variables under- or overestimates the influence of explanatory variables on the formative construct.

→ Formative indicators need to be specified a being endogenous
3.1 Exogenous vs. endogenous indicators for endogenous CLV

Specification including endogenous formative indicators (model A1)

Total effect of $\zeta_1$ on $\eta_3$: $\gamma_{31} + \gamma_{11}\beta_{31} + \gamma_{21}\beta_{32}$

$F = \text{formative measurement model}$

$R = \text{reflective measurement model}$

* = exogenous indicators
3.1 Exogenous vs. endogenous indicators for endogenous CLV

Specification including exogenous formative indicators (model A2)

Total effect of $\zeta_1$ on $\eta_3: \gamma_{31}$

$F$ = formative measurement model
$R$ = reflective measurement model
* = exogenous indicators
3.2 Empirical study

Empirical study on the causes and effects of brand competence

- **Brand competence**: “A competent brand is one that has the ability to solve a consumer’s problem and to meet his or her need.” *(Lau/Lee 2000)*

- **Data**: Survey (GfK Market Research Nuremberg) on brands for *fixatives* and *denture cleansers*, *N* = 261 consumers in Germany

- **Operationalization of brand competence** by four formative indicators (e.g., “stands for hygiene and cleanliness”, “stands for attractiveness and beauty”)

- **Causes of brand competence**:
  - *Advertising* (“really advertises a lot”)
  - *Recommendation* (“is recommended by dentists”)

- **Effect of brand competence**: *brand strength (BPI)*
3.2 Empirical study

Structural equation model of brand competence – causes and effects

- **EXOG1**
- **EXOG2**
- **BC1**
- **BC2**
- **BC3**
- **BC4**

**Brand competence**

- **Brand strength (BPI)**

**Exogenous variables** (advertising, recommendation)

**Formative indicators**
3.2 Empirical study

Results – Fit measures

- **Model 1**: Formative indicators of *brand competence* specified as endogenous variables.

- **Model 2**: Formative indicators of *brand competence* specified as exogenous variables.

- **Model 3**: Modification of model 1 (constraining two coefficients to zero).

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$p$</th>
<th>RMSEA (90 % K. I.)</th>
<th>$p$</th>
<th>CFI</th>
<th>TLI</th>
<th>RMRs</th>
<th>$R^2$ Brand competence</th>
<th>$R^2$ Brand strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>158.05</td>
<td>87</td>
<td>0.000</td>
<td>0.053 (0.038 – 0.067)</td>
<td>0.350</td>
<td>0.987</td>
<td>0.982</td>
<td>0.045</td>
<td>0.61</td>
<td>0.39</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>identical to model 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>158.66</td>
<td>89</td>
<td>0.000</td>
<td>0.052 (0.037 – 0.066)</td>
<td>0.399</td>
<td>0.987</td>
<td>0.983</td>
<td>0.044</td>
<td>0.61</td>
<td>0.39</td>
</tr>
</tbody>
</table>
### 3.2 Empirical study

#### Direct effects of the exogenous variables

<table>
<thead>
<tr>
<th>Explanatory variable</th>
<th>Parameter</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advertising (EXOG1)</td>
<td>$\gamma_{11}$</td>
<td>0.077</td>
<td>−</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{21}$</td>
<td>0.102</td>
<td>−</td>
<td>0.101</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{31}$</td>
<td>0.133***</td>
<td>−</td>
<td>0.131***</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{41}$</td>
<td>0.160</td>
<td>−</td>
<td>0.158</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{51}$</td>
<td>0.081</td>
<td>−</td>
<td>0.080</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{22}$</td>
<td>0.099</td>
<td>−</td>
<td>0.098</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{32}$</td>
<td>0.005</td>
<td>−</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{42}$</td>
<td>0.005</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{52}$</td>
<td>0.155***</td>
<td>0.155***</td>
<td>0.148***</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{53}$</td>
<td>0.261</td>
<td>0.261</td>
<td>0.250</td>
</tr>
<tr>
<td>Recommendation (EXOG2)</td>
<td>$\gamma_{12}$</td>
<td>0.097*</td>
<td>−</td>
<td>0.097*</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{22}$</td>
<td>0.120</td>
<td>−</td>
<td>0.121</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{32}$</td>
<td>0.194***</td>
<td>−</td>
<td>0.194***</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{42}$</td>
<td>0.220</td>
<td>−</td>
<td>0.220</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{52}$</td>
<td>0.088</td>
<td>−</td>
<td>0.089</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{43}$</td>
<td>0.102</td>
<td>−</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{53}$</td>
<td>0.160**</td>
<td>−</td>
<td>0.162***</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{54}$</td>
<td>0.161</td>
<td>−</td>
<td>0.163</td>
</tr>
</tbody>
</table>

- **direct effects on formative indicators**
- **direct effect on brand competence**
- **direct effects on formative indicators**

---

* *** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$; standardised parameter estimates in italic.*
## 3.2 Empirical study

### Total and indirect effects

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Indirect effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>advertising → brand competence</td>
<td>0.056**</td>
<td>−</td>
<td>0.054**</td>
</tr>
<tr>
<td>recommendation → brand competence</td>
<td>0.073**</td>
<td>−</td>
<td>0.072**</td>
</tr>
<tr>
<td><strong>Total effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>advertising → brand competence</td>
<td>0.211***</td>
<td>0.155***</td>
<td>0.203***</td>
</tr>
<tr>
<td></td>
<td>0.356</td>
<td>0.261</td>
<td>0.341</td>
</tr>
<tr>
<td>brand strength</td>
<td>0.193***</td>
<td>0.142***</td>
<td>0.186***</td>
</tr>
<tr>
<td></td>
<td>0.222</td>
<td>0.163</td>
<td>0.214</td>
</tr>
<tr>
<td>recommendation → brand competence</td>
<td>0.043</td>
<td>−0.030</td>
<td>0.072**</td>
</tr>
<tr>
<td></td>
<td>0.068</td>
<td>−0.048</td>
<td>0.114</td>
</tr>
<tr>
<td>brand strength</td>
<td>0.039</td>
<td>−0.028</td>
<td>0.066**</td>
</tr>
<tr>
<td></td>
<td>0.043</td>
<td>−0.030</td>
<td>0.071</td>
</tr>
</tbody>
</table>

*** $p \leq 0.01$, ** $p \leq 0.05$, * $p \leq 0.10$; standardised parameter estimates in italic.
Outline

1. Motivation

2. Identification of recursive SEM including formative constructs – A new approach

3. Specification of SEM including endogenous formative constructs
   3.1 Exogenous vs. endogenous indicators for endogenous CLV
   3.2 Empirical study

4. Conclusions
4. Conclusions

Conclusions

• Many theoretical constructs in marketing or management research may be specified as being formative (instead of reflective).

• “LISREL” approach often mistaken for not being able to handle formative constructs → “LISREL” can handle formative constructs given model identification.

• PLS “solves” the identification problem but suffers from various shortcomings compared to the “LISREL” approach.

• Identification rules for formative measurement models in “LISREL” have been incomplete: New types of models formerly considered as nonidentified can now be identified.

• For endogenous formative constructs, indicators (or components) should be specified as being endogenous → using exogenous indicators under- or overestimates the effect of explanatory variables on the formative construct; in PLS, only (biased) total effects of the explanatory variables can be estimated.