Limited-Information Testing for Structural Models with Categorical Data

Scott Monroe and Li Cai

IMPS, 2013
1. A Motivating Example

2. Goodness-of-Fit Testing

3. Simulation Study

4. Empirical Application

5. Conclusion
A Motivating Example: PISA Student Questionnaire

Example PISA (2003) Items Measuring Self-Related Cognition in Mathematics

- *How much do you disagree or agree with the following statements?*
  - I learn mathematics quickly.
  - I get very nervous doing mathematics problems.

- *How confident do you feel about having to do the following calculations?*
  - Using a <train timetable>, how long it would take to get from Zedville to Zedtown?
A Proposed Ordinal Structural Model

Latent Mediation Model for PISA Questionnaire Data

- **PSC**: Positive self-concept as a mathematics student
- **ANX**: Mathematics anxiety
- **TASK**: Task-specific confidence
This research considers the \textit{multistage} estimator, which estimates:

1. thresholds by ML
2. polychoric correlations by ML
   - stages 1 and 2 yield a sample polychoric correlation matrix
3. structural parameters by some form of least squares
First type:
statistic based on minimized fit-function value

- Let $F$ be the minimum fit function value from estimation
- Then, $T = (N - 1)F$ is used to construct a test statistic
- Typically, $T$ is adjusted to approximate a chi-square variate using moment-matching (e.g., Satorra and Bentler, 1994)
  - define $T_U$ and $T_D$ as mean- and variance-adjusted stats based on ULS and DWLS, respectively
Second type: statistic based on contingency table residuals (Maydeu-Olivares, 2001)

- theoretical appeal of accounting for all levels of uncertainty
- Maydeu-Olivares (2001) derived 3 test statistics:
  1. distributional
  2. structural
  3. overall
- like $T_U$ and $T_D$, all 3 statistics formed by matching moments
Maydeu-Olivares and Joe (2005, 2006) proposed $M_2$

- quadratic form based on first- and second-order marginal residuals
- *limited-information* statistic
- $M_2^*$, a version of $M_2$ for polytomous responses (Joe and Maydeu-Olivares, 2010, Cai and Hansen, 2012)
- chi-square distributed
$M_2$ has been successfully applied to many IRT models, estimated by ML.

But, $M_2$ is not limited to IRT or ML (Maydeu-Olivares and Joe, 2006).

The current research uses $M_2$ and $M^*_2$ as an overall test for ordinal structural models, estimated by the multistage estimator.
Simulation Study

• Purpose:
  1. show $M_2$ is chi-squared
  2. compare $M_2$ to $T_U$ and $T_D$ in terms of calibration and power

• Conditions:
  • 500 replications attempted
  • model identical to PISA example (latent mediation)
  • $N = 100, 200, 500, 1000$
  • $K = 2$ or 4 categories per item
  • model misspecification via Tucker, Koopman, and Linn (TKL, 1969)
Calibration of Test Statistics

QQ Plot for N=1000, K=4, Null Condition

Statistics
- $M_2$
- $T_U$
- $T_D$

KS p-value
- $M_2 = 0.36$
- $T_U = 0.25$
- $T_D = 0.32$
Calibration of Test Statistics

QQ Plot for N=200, K=4, Null Condition

Statistics

- $M_2$
- $T_U$
- $T_D$

KS p-value

- $M_2 = 0.57$
- $T_U = 0.05$
- $T_D < 0.01$
Calibration of Test Statistics

QQ Plot for N=100, K=4, Null Condition

Statistics

- $M_2$
- $T_U$
- $T_D$

KS p-value

- $M_2 = 0.08$
- $T_U < 0.01$
- $T_D < 0.01$
Calibration of Test Statistics

QQ Plot for N=1000, K=2, Null Condition

Statistics
- \( M_2 \)
- \( T_U \)
- \( T_D \)

KS p-value
- \( M_2 = 0.48 \)
- \( T_U = 0.2 \)
- \( T_D = 0.08 \)
Calibration of Test Statistics

QQ Plot for N=200, K=2, Null Condition

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Observed</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_2$</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>$T_U$</td>
<td>&lt; 0.01</td>
<td></td>
</tr>
<tr>
<td>$T_D$</td>
<td>&lt; 0.01</td>
<td></td>
</tr>
</tbody>
</table>
Calibration of Test Statistics

QQ Plot for N=100, K=2, Null Condition

Statistics

- $M_2$
- $T_U$
- $T_D$

KS p-value

- $M_2 = 0.23$
- $T_U < 0.01$
- $T_D < 0.01$
Power of Test Statistics at $\alpha = .05$

Misspecification: TKL 10

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Power</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0</td>
<td>$M_2$</td>
</tr>
<tr>
<td>500</td>
<td>0.5</td>
<td>$T_U$</td>
</tr>
<tr>
<td>1000</td>
<td>1</td>
<td>$T_D$</td>
</tr>
</tbody>
</table>

For $K=2$:

- $M_2$
- $T_U$
- $T_D$

For $K=4$:

- $M_2$
- $T_U$
- $T_D$
Power of Test Statistics at $\alpha = .05$

Misspecification: TKL 30

<table>
<thead>
<tr>
<th></th>
<th>K=2</th>
<th></th>
<th>K=4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>500</td>
<td>1000</td>
<td></td>
</tr>
<tr>
<td>Power</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Statistics
- $M_2$
- $T_U$
- $T_D$
An Aside: RMSEA for Discretized Latent Variable

For TKL10, the *population* RMSEA is .033

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>K = 2</th>
<th>K = 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>.017 (.023)</td>
<td>.027 (.028)</td>
</tr>
<tr>
<td>200</td>
<td>.016 (.018)</td>
<td>.022 (.022)</td>
</tr>
<tr>
<td>500</td>
<td>.011 (.011)</td>
<td>.022 (.014)</td>
</tr>
<tr>
<td>1000</td>
<td>.011 (.008)</td>
<td>.025 (.010)</td>
</tr>
</tbody>
</table>
For TKL30, the *population* RMSEA is .070

<table>
<thead>
<tr>
<th>$K$</th>
<th>Sample Size</th>
<th>100</th>
<th>200</th>
<th>500</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>.021 (.023)</td>
<td>.023 (.017)</td>
<td>.026 (.011)</td>
<td>.027 (.006)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.045 (.032)</td>
<td>.046 (.023)</td>
<td>.050 (.011)</td>
<td>.051 (.008)</td>
<td></td>
</tr>
</tbody>
</table>
## Empirical Application

### Results for PISA data example (US sample, $N = 5,086$)

<table>
<thead>
<tr>
<th>Stat</th>
<th>Value</th>
<th>$df$</th>
<th>$p$</th>
<th>TLI</th>
<th>RMSEA</th>
<th>90% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_U$</td>
<td>330.16</td>
<td>30**</td>
<td>&lt; .001</td>
<td>0.995</td>
<td>0.044</td>
<td>(0.040, 0.048)</td>
</tr>
<tr>
<td>$T_D$</td>
<td>571.50</td>
<td>33**</td>
<td>&lt; .001</td>
<td>0.995</td>
<td>0.057</td>
<td>(0.053, 0.061)</td>
</tr>
<tr>
<td>$M_2$</td>
<td>108.62</td>
<td>27</td>
<td>&lt; .001</td>
<td>0.997</td>
<td>0.024</td>
<td>(0.020, 0.029)</td>
</tr>
</tbody>
</table>

*note: ** indicates an approximation to $df$*
Conclusion

$M_2$ can be applied to structural equation models when the data are categorical.

Advantages of $M_2$:
- better calibration than $T_U$ & $T_D$, particularly with small samples
- more powerful

Disadvantages of $M_2$:
- computationally demanding
- not as versatile as traditional stats

Questions:
- how do $M_2$-based fit indices perform?
- does $M_2$ have power against distributional misspecifications?


This research is supported by grants from the Institute of Education Sciences (R305B080016 and R305D100039) and the National Institute on Drug Abuse (R01DA026943 and R01DA030466).