Summer School on Capability and Multidimensional Poverty
27 August-8 September, 2009
Lima, Peru
Latent Variable Model with Multiple Indicators: An Introduction

Maria Emma Santos & Gaston Yalonetzky

28 August, 2009
Conceptual Motivation of MIMIC

• To model the causes of a (or more) posited latent variable(s) that has multiple observed indicators.

• It goes a step further in the theoretical explanation of the latent variable.
Diagram

Latent Factor

Causes (Exogenous Variables)

Observed Indicators

\[ x_1 \] \[ x_2 \] \[ x_3 \] \[ x_4 \] \[ x_5 \]

\[ y_1 \] \[ y_2 \] \[ y_3 \] \[ y_4 \]

Causes (Exogenous Variables)
MIMIC Model – One factor
(Joreskog & Goldberg, 1975)

\[ y = \lambda f + \varepsilon \]
\[ f = \beta' x + \nu \]
MIMIC Model – One factor

\[ y = \lambda (\beta' x + \nu) + \varepsilon = \Pi' x + \eta \]

• Where the reduced form coefficient matrix is

\[
\Pi = \beta \lambda'
\]

And the reduced-form disturbance vector

\[ \eta = \lambda \nu + \varepsilon \]

has a covariance matrix:

\[ \Omega = E(\eta \eta') \]

\[ = E[(\lambda \nu + \varepsilon)(\lambda \nu + \varepsilon)'] = \sigma^2 \lambda \lambda' + \Psi^2 \]
Two restrictions:

1. The kxm coefficient matrix $\Pi$ has rank one, the km elements of $\beta$ and $\lambda$ (typical of structural equation modelling)

2. The mxn covariance matrix $\Omega$ is the sum of a rank-one matrix and a diagonal matrix, the $m(m+1)/2$ distinct elements of $\Omega$ being expressed in terms of the $1+2m$ elements of $\sigma^2, \lambda, \theta$. (typical of factor analysis: intercorr of observable variables are assumed to be accounted for by a common latent variable).

Estimation is done through MLE or – alternatively– through minimum distance estimators (moments method)
MIMIC Model – One factor

• The estimator of $f$ is given by:

$$\hat{f} = (1 - \lambda' \Omega^{-1} \lambda)^{-1} (\beta' x + \lambda' \Psi^{-1} y)$$

with

$$V(\epsilon) = \Psi; V(\nu) = \sigma^2; \Omega = \lambda \lambda' + \Psi$$
MIMIC Model – Multiple factors
(Krishnakumar & Nagar, 2005)

\[ y = \Lambda \hat{f} + \varepsilon \]

\[ \hat{f} = Bx + \nu \]
Diagram

**Causes** (Exogenous Variables)

- \( x_1 \)
- \( x_2 \)
- \( x_3 \)
- \( x_4 \)
- \( x_5 \)

**Latent Factors**

- \( f_1 \)
- \( f_2 \)

**Observed Indicators**

- \( y_1 \)
- \( y_2 \)
- \( y_3 \)
- \( y_4 \)
MIMIC Model – Multiple factors

- The estimator of $f$ is given by:

$$\hat{f} = (I - \Lambda'\Omega^{-1}\Lambda)^{-1}(Bx + \Lambda'\Psi^{-1}y)$$

with $V(\epsilon) = \Psi; V(\nu) = \sigma^2 I; \Omega = \Psi + \Lambda\Lambda'$

$$\hat{f} = (I + \Lambda'\Psi^{-1}\Lambda)^{-1}Bx + (I + \Lambda'\Psi^{-1}\Lambda)^{-1}\Lambda\Psi^{-1}y$$

- CAUSES TERM
- INDICATORS TERM
Goodness of fit measures for MIMIC (and SEM)

• There are several of these measures
• The basic statistic (upon which some measures are based) is called “chi-square” because it is asymptotically distributed as chi-square:

\[ c^2 = (N - 1) F_{ML} \left( S, \sum (\theta) \right) \]

\[ F_{ML} \left( S, \sum (\theta) \right) = \ln \left| \sum (\theta) \right| + tr \left[ S \sum^{-1} (\theta) \right] - \ln |S| - p \]
Goodness of fit measures for MIMIC (and SEM)

- where:
  - S: sample covariance matrix of the observed variables
  - Sigma (theta): population covariance matrix implied by the model
  - Theta: tx1 vector of free parameters
  - p: number of observed variables
Intuition: The F function has to be minimized (it is the minus of a likelihood function)

A model with good fit has a low F. The null hypothesis is that the model fits the data well (low c²)

The problem with c² is that with large sample sizes rejection of the good-fit hypothesis is pervasive

To solve this problem alternative statistics are proposed
Goodness of fit measures for MIMIC (and SEM)

- One of such statistics is the Root Mean Error of Approximation (RMEA):

\[
RMSEA = \sqrt{\frac{c^2}{df} - \frac{1}{N} - 1}
\]

\[
df = \frac{1}{2} \left( p + 1 \right) p - t
\]

- Notice that the sample size effect is canceled out
Prior Steps: EFA & CFA + Theory

• Before doing a MIMIC it is recommended to perform EFA and CFA to devise the structure of the indicators and latent factors.
• However, the final decision on the number of factors should always be consistent with theory.
An example of Application with Empowerment

• One aspect of empowerment: autonomy.
• Self-Determination Theory – SDT – (developed by psychologists Ryan and Deci) defines a person to be **autonomous** when his or her behavior is experienced as willingly enacted and when he or she fully endorses the actions in which he or she is engaged and/or the values expressed by them. People are therefore most autonomous when they act in accord with their authentic interests or integrated values or desires.
An example of Application with Empowerment

- Ryan and Deci proposed the **Relative Autonomy Index**.
- The index items are beneficial in shedding light on constraints to agency that may arise from sources outside the household; exploring motivations for choices made and whether they are congruent with the respondent’s values; (Ibrahim and Alkire 2007, p. 25).
- This index has been shown to be robust internationally, as it appears equally applicable to the situation of groups in individual and collective societies and in vertical and horizontal cultures (Chirkov *et al.* 2003, 2005, Ibrahim and Alkire 2005, p. 25).
An example of Application with Empowerment

• OPHI module on empowerment has adopted the autonomy questions developed by Ryan and Deci amongst other questions on empowerment.
• The autonomy questions are context-specific.
• A variant of OPHI’s empowerment module was implemented by UNDP LAC in DF of Mexico, Bs. As. Argentina and Managua, Nicaragua. It is a variant in that it has focused on autonomy in decision-making rather than in behaviour in general.
• So far, only the results from Mexico have been analysed.
Example of the agency questions

• When decisions are made regarding minor household expenditures, who is it that normally takes the decision?
  1. You make decisions alone
  2. You and your partner make decisions together
  3. You and other household members make decisions
  4. You and another person
  5. Your partner/spouse alone
  6. Someone different from you and your partner.

• To what extent do you feel you can make your own personal decisions regarding minor household expenditures if you want to?
  1. To a high extent
  2. To a medium extent
  3. To a small extent
  4. Not at all
Example of the agency questions-SDT

- Statements 1 and 2 correspond to "**External Pressure**": low autonomy
- Statements 3 and 4: to "**Introjected Regulation**": when regulations are taken in, but not accepted as one’s own. It refers to behaviours that are performed to avoid guilt or anxiety or attain ego enhancement such as pride.
- Statement 5: "**Identified or Integrated Regulation**": a conscious valuing of a behavioural goal or regulation, such that the action is accepted or owned as personally important
- **Relative Autonomy Index:**
  \[
  \text{RAI} = (-1)[\text{External Pressure Answer 1}] + (-1)[\text{External Pressure Answer 1}] + (-1)[\text{Introjected Regulation Answer}] + (3)[\text{Identified or Integrated Regulation Answer}].
  \]
- The Index varies between -9 and 9, increasing in autonomy.
Example of the agency questions

FOR THOSE THAT ANSWERED THAT THEY DO NOT MAKE DECISIONS:

Please tell me how true is for you each of the following statements regarding the way in which minor household expenditures are done in the household considering the following scale:


1. I cannot make decisions regarding minor household purchases. I do not have that option.
2. I cannot make decisions regarding minor household purchases because my partner does not allow me to do so.
3. I cannot make decisions regarding minor household purchases because another person, the society, organisations or my community does not allow me to do so.
4. I do not make decisions on minor household purchases because that is what other people expect from me. If I made those decisions they could blame me or I could feel guilty.
5. I do not make decisions on minor household purchases because personally I prefer no to do so.
Example of the agency questions

FOR THOSE WHO ANSWERED THAT THEY MAKE DECISIONS

Please tell me how true is for you each of the following statements regarding the way in which minor household expenditures are done in the household considering the following scale:


1. I make/take part in the decisions regarding minor household purchases because I have no other choice. I have to do it.
2. I make/take part in the decisions regarding minor household purchases according to what my partner obliges me to do.
3. I make/take part in the decisions regarding minor household purchases according to what another person, the society, organisations or my community oblige me to do so.
4. I make/take part in the decisions on minor household purchases because that is what other people expect from me. If I would not do it they could blame me or I could feel guilty.
5. I freely choose to make/take part in the decisions on minor household purchases.
Example of the agency questions

• The same set of autonomy questions was done for decisions regarding:
  – The role in the household: to work or not, and in which type of job.
  – The education of the children
PCA on the SDT autonomy questions

MINOR HOUSEHOLD PURCHASES

Factor analysis/correlation  
Number of obs = 1370  
Method: principal-component factors  
Retained factors = 1  
Rotation: (unrotated)  
Number of params = 4

<table>
<thead>
<tr>
<th>Factor</th>
<th>Eigenvalue</th>
<th>Difference</th>
<th>Proportion</th>
<th>Cumulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor1</td>
<td>2.06770</td>
<td>1.08621</td>
<td>0.5169</td>
<td>0.5169</td>
</tr>
<tr>
<td>Factor2</td>
<td>0.98149</td>
<td>0.43808</td>
<td>0.2454</td>
<td>0.7623</td>
</tr>
<tr>
<td>Factor3</td>
<td>0.54341</td>
<td>0.13601</td>
<td>0.1359</td>
<td>0.8981</td>
</tr>
<tr>
<td>Factor4</td>
<td>0.40740</td>
<td></td>
<td>0.1019</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

LR test: independent vs. saturated: chi2(6) = 1094.39 Prob>chi2 = 0.0000
PCA on the SDT autonomy questions

MINOR HOUSEHOLD PURCHASES
Factor loadings (pattern matrix) and unique variances

<table>
<thead>
<tr>
<th>Variable</th>
<th>Factor1</th>
<th>Uniqueness</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAI_comba</td>
<td>0.8057</td>
<td>0.3509</td>
</tr>
<tr>
<td>RAI_combb</td>
<td>0.8532</td>
<td>0.2721</td>
</tr>
<tr>
<td>RAI_combc</td>
<td>0.7969</td>
<td>0.3650</td>
</tr>
<tr>
<td>RAI_combd</td>
<td>-0.2360</td>
<td>0.9443</td>
</tr>
</tbody>
</table>

Average interitem covariance: 0.14
PCA on the SDT autonomy questions

ROLE IN THE HOUSEHOLD

Factor analysis/correlation  Number of obs  =  1342
Method: principal-component factors  Retained factors =  1
Rotation: (unrotated)  Number of params =  4

<table>
<thead>
<tr>
<th>Factor</th>
<th>Eigenvalue</th>
<th>Difference</th>
<th>Proportion</th>
<th>Cumulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor1</td>
<td>2.05167</td>
<td>1.07102</td>
<td>0.5129</td>
<td>0.5129</td>
</tr>
<tr>
<td>Factor2</td>
<td>0.98065</td>
<td>0.38876</td>
<td>0.2452</td>
<td>0.7581</td>
</tr>
<tr>
<td>Factor3</td>
<td>0.59189</td>
<td>0.21609</td>
<td>0.1480</td>
<td>0.9061</td>
</tr>
<tr>
<td>Factor4</td>
<td>0.37580</td>
<td>.</td>
<td>0.0939</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

LR test: independent vs. saturated: chi2(6) = 1077.28 Prob>chi2 = 0.0000
PCA on the SDT autonomy questions

**ROLE IN THE HOUSEHOLD**
Factor loadings (pattern matrix) and unique variances

<table>
<thead>
<tr>
<th>Variable</th>
<th>Factor1</th>
<th>Uniqueness</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAI_comba1</td>
<td>0.7933</td>
<td>0.3706</td>
</tr>
<tr>
<td>RAI_combb2</td>
<td>0.8683</td>
<td>0.2460</td>
</tr>
<tr>
<td>RAI_combc3</td>
<td>0.7753</td>
<td>0.3989</td>
</tr>
<tr>
<td>RAI_combd4</td>
<td>-0.2591</td>
<td>0.9329</td>
</tr>
</tbody>
</table>

Average interitem covariance: 0.11
PCA on the SDT autonomy questions

**EDUCATION OF THE CHILDREN**

<table>
<thead>
<tr>
<th>Factor</th>
<th>Eigenvalue</th>
<th>Difference</th>
<th>Proportion</th>
<th>Cumulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor 1</td>
<td>2.17150</td>
<td>1.18896</td>
<td>0.5429</td>
<td>0.5429</td>
</tr>
<tr>
<td>Factor 2</td>
<td>0.98255</td>
<td>0.47572</td>
<td>0.2456</td>
<td>0.7885</td>
</tr>
<tr>
<td>Factor 3</td>
<td>0.50683</td>
<td>0.16771</td>
<td>0.1267</td>
<td>0.9152</td>
</tr>
<tr>
<td>Factor 4</td>
<td>0.33912</td>
<td>.</td>
<td>0.0848</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

LR test: independent vs. saturated: chi2(6) = 1366.15 Prob>chi2 = 0.0000
PCA on the SDT autonomy questions

**EDUCATION OF THE CHILDREN**

Factor loadings (pattern matrix) and unique variances

<table>
<thead>
<tr>
<th>Variable</th>
<th>Factor1</th>
<th>Uniqueness</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAI_combaa</td>
<td>0.8277</td>
<td>0.3149</td>
</tr>
<tr>
<td>RAI_combbb</td>
<td>0.8807</td>
<td>0.2243</td>
</tr>
<tr>
<td>RAI_combcc</td>
<td>0.8178</td>
<td>0.3313</td>
</tr>
<tr>
<td>RAI_combdd</td>
<td>-0.2048</td>
<td>0.9580</td>
</tr>
</tbody>
</table>

*External/Introjected*

*Identified*
PCA on the SDT autonomy questions

- Previous analysis suggests that it is valid to combine the 4 items of each question into one index.
- Q/: is autonomy in one context related to autonomy in another?
- Using the RAI in each context + other empowerment qs we performed EFA.
EFA for empowerment: 7 indicators

EXPLORATORY FACTOR ANALYSIS WITH 1 FACTOR(S): Not good results

TESTS OF MODEL FIT

Chi-Square Test of Model Fit
Value 115.898*
Degrees of Freedom 5**
P-Value 0.0000

Chi-Square Test of Model Fit for the Baseline Model

RMSEA (Root Mean Square Error Of Approximation)
Estimate 0.127
EFA for empowerment: 7 indicators

EXPLORATORY FACTOR ANALYSIS WITH 2 FACTOR(S): GOOD RESULTS

EIGENVALUES FOR SAMPLE CORRELATION MATRIX

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.224</td>
<td>1.718</td>
<td>1.197</td>
<td>0.790</td>
<td>0.491</td>
</tr>
</tbody>
</table>

EIGENVALUES FOR SAMPLE CORRELATION MATRIX

<table>
<thead>
<tr>
<th></th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.402</td>
<td>0.178</td>
</tr>
</tbody>
</table>

TESTS OF MODEL FIT

Chi-Square Test of Model Fit

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>8.707*</td>
</tr>
<tr>
<td>Degrees of Freedom</td>
<td>3**</td>
</tr>
<tr>
<td>P-Value</td>
<td>0.0334</td>
</tr>
</tbody>
</table>

RMSEA (Root Mean Square Error Of Approximation)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>0.037</td>
</tr>
</tbody>
</table>

SRMR (Standardized Root Mean Square Residual)

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>0.046</td>
</tr>
</tbody>
</table>

MINIMUM ROTATION FUNCTION VALUE 0.16017
## EFA for empowerment: 7 indicators

### FACTOR STRUCTURE

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAI PURCHASES</td>
<td>0.721</td>
<td>0.031</td>
</tr>
<tr>
<td>RAI HH ROLE</td>
<td>0.801</td>
<td>-0.047</td>
</tr>
<tr>
<td>RAI CHI EDUC</td>
<td>0.692</td>
<td>-0.135</td>
</tr>
<tr>
<td>GLOBAL DEC MAK</td>
<td>-0.133</td>
<td>0.041</td>
</tr>
<tr>
<td>CHANGE 1</td>
<td>-0.027</td>
<td>0.748</td>
</tr>
<tr>
<td>CHANGE 2</td>
<td>-0.070</td>
<td>1.140</td>
</tr>
<tr>
<td>LADDER</td>
<td>-0.018</td>
<td>0.067</td>
</tr>
</tbody>
</table>
Results of MIMIC for autonomy

Estimator WLSMV
Maximum number of iterations 1000
Convergence criterion 0.500D-04
Maximum number of steepest descent iterations 20
Maximum number of iterations for H1 2000
Convergence criterion for H1 0.100D-03
Parameterization DELTA

TESTS OF MODEL FIT
Chi-Square Test of Model Fit
  Value 12.238*
  Degrees of Freedom 10**
  P-Value 0.2695

RMSEA (Root Mean Square Error Of Approximation)
  Estimate 0.015
### MIMIC Results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>S.E.</th>
<th>Est./S.E.</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAI Purchases*</td>
<td>1.000</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>RAI Role in HH</td>
<td>1.127</td>
<td>0.116</td>
<td>9.739</td>
<td>0.000</td>
</tr>
<tr>
<td>RAI Chil. Educ.</td>
<td>0.871</td>
<td>0.133</td>
<td>6.536</td>
<td>0.000</td>
</tr>
</tbody>
</table>
## Autonomy ON

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>S.E.</th>
<th>Est./S.E.</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women</td>
<td>0.022</td>
<td>0.112</td>
<td>0.195</td>
<td>0.845</td>
</tr>
<tr>
<td>Btw 19-29</td>
<td>-0.313</td>
<td>0.295</td>
<td>-1.061</td>
<td>0.289</td>
</tr>
<tr>
<td>Btw 45 and 59</td>
<td>-0.078</td>
<td>0.125</td>
<td>-0.626</td>
<td>0.531</td>
</tr>
<tr>
<td><strong>60 or more</strong></td>
<td><strong>-0.515</strong></td>
<td><strong>0.260</strong></td>
<td><strong>-1.984</strong></td>
<td><strong>0.047</strong></td>
</tr>
<tr>
<td>Father Employer</td>
<td>-0.633</td>
<td>0.179</td>
<td>-3.545</td>
<td>0.000</td>
</tr>
<tr>
<td>Father Self-Employed</td>
<td>-0.380</td>
<td>0.151</td>
<td>-2.512</td>
<td>0.012</td>
</tr>
<tr>
<td>Father Private Sect. Empl.</td>
<td>-0.335</td>
<td>0.149</td>
<td>-2.251</td>
<td>0.024</td>
</tr>
<tr>
<td>Mother worked</td>
<td>-0.436</td>
<td>0.183</td>
<td>-2.383</td>
<td>0.017</td>
</tr>
<tr>
<td>Parents with Sec. Ed.</td>
<td>0.004</td>
<td>0.142</td>
<td>0.031</td>
<td>0.975</td>
</tr>
<tr>
<td><strong>Parents with Tert Ed.</strong></td>
<td><strong>0.696</strong></td>
<td><strong>0.235</strong></td>
<td><strong>2.968</strong></td>
<td><strong>0.003</strong></td>
</tr>
<tr>
<td>Father’s age at adult’s birth</td>
<td>0.006</td>
<td>0.018</td>
<td>0.313</td>
<td>0.755</td>
</tr>
<tr>
<td>Mother’s age at adult’s birth</td>
<td>-0.002</td>
<td>0.018</td>
<td>-0.114</td>
<td>0.909</td>
</tr>
<tr>
<td>Lived in DF at age 12</td>
<td>0.055</td>
<td>0.160</td>
<td>0.343</td>
<td>0.732</td>
</tr>
<tr>
<td>Lived in Mex state at age 12</td>
<td>-0.016</td>
<td>0.170</td>
<td>-0.096</td>
<td>0.924</td>
</tr>
</tbody>
</table>

**Residual Variances**

| AUTON          | 0.548 | 0.072 | 7.636 | 0.000 |
Latent variables: myths and realities (Everitt and Dunn, 2001)

- MIMIC and SEM are very attractive
- A whole industry of it developed in social sciences (e.g. psychology, sociology) way before their use for capability-approach studies
- But beware: we need to be mindful all the time of the purpose of these models when we apply them
Latent variables: myths and realities

• Causality:
  – “Correlation does not imply causation” (try reversing roles of variables in a model)
  – These models (as well as many others) capture correlations between variables
  – But proper causal inference requires “active control of variables” (e.g. randomized trials, natural experiments, or statistical patching)
  – So these models are good to provide a parsimonious description of a set of correlated variables with high explanatory power
  – Prior knowledge (or theory) about the relationship between the variables is important for interpretation
  – No surprise that two researchers make come up with different preferred models for the same system
Latent variables: myths and realities

- Latency:
  - In practice latent variables are whatever is contained in the observed variables used to “detect” them
  - In theory latent variables are hypothetical constructs made to understand a phenomenon but for which there is no method for direct measurement
  - Can we make progress with hypothetical constructs?
  - Sometimes!
Latent variables: myths and realities

• Latency:
  – Examples of “sometimes” progress with latent variables: atom particles, gravity, “g” intelligence, agency/empowerment
  – The importance of latent variables is not their “reality” but the ability of models which depend on them to describe and predict phenomena
  – Latent variables also serve to synthesize and summarize the properties of observed variables
  – Latent variables are “as real as their predictive consequences are valid”
Latent variables: myths and realities

• Latent variables in the light of the Friedman-Samuelson debate:
  – Friedman: models should be judged on the grounds of their predictive power regardless of assumptions
  – Samuelson: assumptions are important

• The justification for the use of latent variables is safest under the position of Friedman
Cronbach Alpha Coefficient

• It is a reliability coefficient used in factor analysis: it indicates how well a set of indicators measure a latent variable.

\[ \alpha = \frac{N\overline{c}}{\overline{\nu} + (N - 1)\overline{c}} \]

where \( N \) is the number of indicators, \( \overline{c} \) denotes the average covariances between pairs of indicators and \( \overline{\nu} \) is the average of the indicators’ covariances.

The higher average correlation between indicators, the higher the coefficient, and higher reliability that the indicators are measuring well the latent variable.