On the Targets of Latent Variable Model Estimation

Karen Bandeen-Roche
Department of Biostatistics
Johns Hopkins University

Department of Mathematics and Statistics
Miami University
December 8, 2005

With thanks to: Howard Chilcoat, Naomi Breslau, Linda Fried
Introduction: Statistical Problem

• Observed variables (i=1,...,n):  \( Y_i = M \)-variate;  \( x_i = P \)-variate

• Focus:  response (Y) distribution = \( G_{Y|x}(y|x) \);  x-dependence

• Modeling issue: flexible or theory-based?
  — Flexible:  \( g_m(E[Y_{im}|x_i]) = f_m(x_i), m=1,...,M \)
  — Theory-based:
    > \( Y_i \) generated from latent (underlying) \( U_i \):
      \[
      F_{Y|U,x}(y|U=u,x;\pi) \quad (Measurement)
      \]
    > Focus on distribution, regression re \( U_i \):
      \[
      F_{U|x}(u|x;\beta) \quad (Structural)
      \]
    > Overall, hierarchical, model:
      \[
      F_{Y|x}(y|x) = \int F_{Y|U,x}(y|U=u,x)dF_{U|x}(u|x)
      \]
Motivation

The Debate over Mixture and Latent Variable Models

• **In favor**: they
  — acknowledge **measurement problems**: errors, differential reporting
  — **summarize** multiple measures **parsimoniously**
  — operationalize **theory**
  — describe population **heterogeneity**

• **Against**: their
  — **modeling assumptions** may determine scientific conclusions

  — **interpretation** may be ambiguous
    > nature of latent variables?
    > comparable fit of very different models
    > seeing is believing
Possible Approaches to the Debate

- Argue advantages of favorite method

- Hybrid approaches:
  - Parallel analyses (e.g. *Bandeen-Roche et al. AJE 1999*)
  - Marginal mean + LV-based association
    (e.g. *Heagerty, Biometrics, 2001*)

- Sensitivity analyses

- “Popperian”
  - Pose parsimonious model
  - Learn how it fails to describe the world
Outline

- Modeling and estimation framework

- Specifying the target of estimation
  - *Supposing that the target uniquely exists* ...
    - Strategy for delineating it
    - Validity of the strategy
  - *Unique existence of the target*

- Applications
  - Post-traumatic Stress Disorder
  - Basic task disability in older women
Application: Post-traumatic Stress Disorder Ascertainment

- PTSD
  - Follows a qualifying traumatic event
    > This study: personal assault, other personal injury/trauma, trauma to loved one, sudden death of loved one = “x”, along with gender
  - Criterion endorsement of symptoms related to the event ⇒ diagnosis
    > Binary report on 17 symptoms = “Y”

- A recent study (Chilcoat & Breslau, Arch Gen Psych, 1998)
  - Telephone interview in metropolitan Detroit
  - n=1827 with a qualifying event

  - Analytic issues
    > Nosology
    > Does diagnosis differ by trauma type or gender?
    > Are female assault victims particularly at risk?
Model 1
Latent Class Regression

\[ P_1(x) \]
\[ \ldots \]
\[ \Pi_{11} \]
\[ \ldots \]
\[ Y_{d1} \]
\[ \ldots \]
\[ Y_M \]

\[ \ldots \]
\[ \Pi_{M1} \]

\[ \ldots \]
\[ \Pi_{1J} \]
\[ \ldots \]
\[ Y_1 \]
\[ \ldots \]
\[ Y_M \]

\[ P_j(x) = \Pr\{U = j|x\} \]
\[ \pi_{mj} = \Pr\{Y_m = 1|U = j\} \]

References: Dayton & Macready 1988, van der Heijden et al., 1996; Bandeen-Roche et al., 1997
Latent Class Regression (LCR) Model

- Model:
  \[ f_{y|x}(y|x) = \sum_{j=1}^{J} P_j(x, \beta) \prod_{m=1}^{M} \pi_{jm} y^m (1-\pi_{jm})^{1-y^m} \]

- Structural model assumption: \([U_i | x_i] = Pr\{U_i = j|x_i\} = P_j(x_i, \beta)\)
  - \(RPR_j = Pr\{U_i = j|x_i\}/Pr\{U_i = J|x_i\}; j=1,...,J\)

- Measurement assumptions: \([Y_i | U_i]\)
  - conditional independence
  - nondifferential measurement
  > reporting heterogeneity unrelated to measured, unmeasured characteristics

- Fitting: ML w EM; robust variance (e.g. Muthén & Muthén 1998, M-Plus)

- Posterior latent outcome info: \(Pr\{U_i = j|Y_i, x_i; \theta = (\pi, \beta)\}\)
Methodology
Delineating the Target of Measurement

- **Fit an initial model**: ML, Bayes, etc.

- **Obtain posterior latent outcome** info — e.g. \( f_{u|Y,x}(u|Y,x;\theta) \)
  — This talk: empirical Bayes

- RANDOMLY generate “empirical LVs,” \( V_i \), according to \( f_{u|Y,x}(u|Y,x;\hat{\theta}) \)

- Analyze \( V_i \) AS \( U_i \) (accounting for variability in first-stage estimation)

- Estimate measurement structure through empirical analysis of \( Y_i|V_i,x_i \)
Methodology
Properties “whatever” the True Distribution

- Under Huber (1967)-like conditions:

  — **Asymptotically**:

    > Randomization imposes limiting hierarchical model, except $[Y|V,x]$ arbitrary (and specifiable)

    i.e. *underlying variable distribution has an estimable interpretation even if assumptions are violated*

    > No bias in substituting $V_i$ for $U_i$.

    i.e. *regression of $V_i$ on $x_i$ and model-based LV regression eventually equivalent*
Methodology
More formal statement

- Under Huber (1967)-like conditions:

  - \((\hat{\beta}, \hat{\pi})\) converge in probability to limits \((\beta^*, \pi^*)\).

  - \(Y_i\) asymptotically equivalent in distribution to \(Y^*\), generated as:
    
    i) Generate \(U_i^*\) — distribution determined by \((\beta^*, \pi^*)\), \(G_{Y|x}(y|x)\);
    
    ii) Generate \(Y^*\) — distribution determined by \((\beta^*, \pi^*)\), \(G_{Y|x}(y|x)\), \(U_i^*\)

  - \{Pr[\(Y_i \leq y|V_i, x_i\)], i=1,2,...\} converges in distribution to
    \{Pr[\(Y_i^* \leq y|U_i^*, x_i\)], i=1,2,...\}, for each supported \(y\).

  - \(V_i\) converges in distribution to \(U_i^*\).
## PTSD Study: Descriptive Statistics

<table>
<thead>
<tr>
<th>Gender</th>
<th>Trauma Type: percentage distribution</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Personal Assault</td>
<td>Other Injury</td>
</tr>
<tr>
<td>Male</td>
<td>14.2</td>
<td>37.7</td>
</tr>
<tr>
<td>Female</td>
<td>14.3</td>
<td>26.3</td>
</tr>
<tr>
<td>Total</td>
<td>14.2</td>
<td>32.3</td>
</tr>
</tbody>
</table>

- PTSD symptom criteria met: 11.8% (n=215)
  - By gender: 8.3% of men, 15.6% of women
  - By trauma: assault (26.9%), sudden death (14.8%), other injury (8.1%), trauma to loved one (6.0%)
  - Interactions: female x assault (↑), female x other (↓)
  - Criterion issue? 60% reported symptoms short of diagnosis
Latent Class Model for PTSD: 9 items

<table>
<thead>
<tr>
<th>SYMPTOM CLASS</th>
<th>SYMPTOM (prevalence)</th>
<th>SYMPTOM PROBABILITY ($\pi$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Class 1 - NO PTSD</td>
</tr>
<tr>
<td>RE-EXPERIENCE</td>
<td>Recurrent thoughts (.49)</td>
<td>.20</td>
</tr>
<tr>
<td></td>
<td>Distress to event cues (.42)</td>
<td>.12</td>
</tr>
<tr>
<td></td>
<td>Reactivity to cues (.31)</td>
<td>.05</td>
</tr>
<tr>
<td>AVOIDANCE/NUMBING</td>
<td>Avoid related thoughts (.28)</td>
<td>.08</td>
</tr>
<tr>
<td></td>
<td>Avoid activities (.24)</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>Detachement (.15)</td>
<td><strong>.01</strong></td>
</tr>
<tr>
<td>INCREASED AROUSAL</td>
<td>Difficulty sleeping (.19)</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td>Irritability (.21)</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td>Difficulty concentrating (.25)</td>
<td>.03</td>
</tr>
<tr>
<td>MEAN PREVALENCE-BASELINE</td>
<td></td>
<td>.52</td>
</tr>
</tbody>
</table>

[Omitted: nightmares, flashback; amnesia, ↓interest, ↓affect, short future; hypervigilance, startle]
Odds and Relative Odds, with 95% Confidence Intervals
PTSD: DIAGNOSIS, LCR MEASUREMENT MODEL

- **Method:** Regress item responses on covariates “controlling” for class — For simplicity: non-assaultive traumas merged into “other trauma”

<table>
<thead>
<tr>
<th>Variable</th>
<th>Odds Ratio or Interaction Ratio (CI)</th>
<th>By-item Odds Ratio MODEL 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>1.07 (0.93,1.22)</td>
<td>1.07 (0.93,1.22)</td>
</tr>
<tr>
<td>Trauma = other than assault (recur.)</td>
<td>3.19 (1.89,5.40)</td>
<td>3.19 (1.89,5.40)</td>
</tr>
<tr>
<td>Cue distress x other trauma</td>
<td>0.18 (0.09,0.38)</td>
<td>0.58 (0.36,0.92)</td>
</tr>
<tr>
<td>Cue reactivity x other trauma</td>
<td>0.14 (0.07,0.28)</td>
<td>0.44 (0.27,0.72)</td>
</tr>
<tr>
<td>Avoid thoughts x other trauma</td>
<td>0.21 (0.11,0.41)</td>
<td>0.68 (0.44,1.05)</td>
</tr>
<tr>
<td>Avoid activities x other trauma</td>
<td>0.11 (0.05,0.22)</td>
<td>0.35 (0.21,0.58)</td>
</tr>
<tr>
<td>Detachment x other trauma</td>
<td>0.27 (0.13,0.58)</td>
<td>0.88 (0.51,1.49)</td>
</tr>
<tr>
<td>Difficulty sleep x other trauma</td>
<td>0.43 (0.21,0.90)</td>
<td>1.37 (0.78,2.42)</td>
</tr>
<tr>
<td>Irritability x other trauma</td>
<td>0.28 (0.13,0.61)</td>
<td>0.91 (0.52,1.59)</td>
</tr>
<tr>
<td>Concentration x other trauma</td>
<td>0.73 (0.36,1.47)</td>
<td>2.33 (1.35,4.03)</td>
</tr>
</tbody>
</table>
Summary
PTSD Analysis

- The analysis hypothesizes that PTSD is
  - a syndrome comprising unaffected, subclinically affected, and diseased subpopulations of those suffering traumas
  - reported homogeneously within subpopulations
- The hypotheses are consistent with current diagnostic criteria
- Gender x type interactions: are strongly indicated
  - Female assault victims at particular risk
  - ... given the subpopulations defined by the model
Summary
PTSD Analysis

- Symptoms appeared differentially sensitive to different traumas

  Within classes: those who had a non-assaultive trauma were

  — less prone to report distress to cues, reactivity to cues, avoiding thoughts, & avoiding activities

  — more prone to report recurrent thoughts & difficulty concentrating

- Concern: Current criteria may better detect psychiatric sequelae to assault than to traumas other than assault
Characterization of the Target Parameters ($\beta^*, \pi^*$)

Huber (1967), Proc. 5th Berkeley Symposium

- **Notation**
  - True distribution: \{Y_1, ..., Y_n\} i.i.d. with $Y_i \sim f^*_Y(y)$
  - Model: $Y_i \sim f_Y(y; \beta, \pi)$ ~ an **LCA** mass function
  - Derivative operator: $D_\phi = \text{gradient wrt } (\beta, \pi)$

- **If** ($\beta^*, \pi^*$) **exist**: they minimize Kullback-Leibler distance between $f$ & $f^*$

- **When do** ($\beta^*, \pi^*$) **exist**?
  - Regularity conditions
  - **Key**: $E_{f^*}[D_\phi \ln f_Y(y; \beta, \pi)]$ has a unique 0
Existence of the Target Parameters ($\beta^*, \pi^*$)

Verification

- Two strategies
  - Theory
    - Geometry (e.g. Lindsay, Ann Stat., 1983)
    - Global identifiability
  - Direct examination
    \[ \sum_{l=1}^{2^M} [D_\phi \ln f_Y(y_i;u,v)]_{\beta,\pi} \Pr\{Y=y_i\} \text{ as a function of } (\beta, \pi) \text{ (grid)} \]
    \[ \Pr\{Y=y\} = f^*_Y(y) \text{ unknown; estimate by empirical } \hat{\Pr}\{Y=y\} \]

- A key aid: substantive conceptual framework
  - Reduction of parameter space
Example: Self-reported Disability among Older Adults

- **Import**: Medicare funding weighs prevalence of self-reported disability
  - Recent report: disability decreasing (Manton et al., 1998)

- **This talk**: Basic functioning in The Women’s Health and Aging Study
  - “Basic function” via “Do you have difficulty …”
    -洗澡, 准备餐, 穿衣服, 用马桶 (M=4)
  - 7 rounds every 6 months: n=1002 at baseline
  - Aims: disability prevalence trend + role of covariates, x_{it}
  - Potential failure to measure as intended: trust effect

- Why not a “harder” outcome than self-report?
  - Distinct dimension of health (e.g., Jette, 1980)
  - Increasingly a focus of interventions
Example
Women’s Health and Aging Study

- **Conceptual framework:** Task hierarchy (*Fried et al., J Clin Epi, 1999*)
  - Difficulty ordering according to physiological demand
  - Basic functioning: bathing = most difficult; others = “parallel”

- **Idealized** conditional probabilities (πs):

<table>
<thead>
<tr>
<th>Task</th>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bathing</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Preparing Meals</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Dressing</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Use toilet</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Example
Women’s Health and Aging Study

- Conceptual framework: Task hierarchy *(Fried et al., J Clin Epi, 1999)*
  — Difficulty ordering according to physiological demand

  — Basic functioning: bathing = most difficult; others = “parallel”

- Modeled conditional probabilities (πs):

<table>
<thead>
<tr>
<th>Task</th>
<th>Class 1</th>
<th>Class 2</th>
<th>Class 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bathing</td>
<td>0</td>
<td>1</td>
<td>π₃</td>
</tr>
<tr>
<td>Preparing Meals</td>
<td>π₁</td>
<td>π₂</td>
<td>π₃</td>
</tr>
<tr>
<td>Dressing</td>
<td>π₁</td>
<td>π₂</td>
<td>π₃</td>
</tr>
<tr>
<td>Use toilet</td>
<td>π₁</td>
<td>π₂</td>
<td>π₃</td>
</tr>
</tbody>
</table>

> Constrained parameter space: π₁ < .5, π₂ < .5, π₃ > .5
Example: Uniqueness of Target Parameters

- First, a test case: \( P_1 = P_2 = 1/3; \) \( (\pi_1, \pi_2, \pi_3) = (.1,.1,.9) \)
- Measure of closeness to 0: Euclidean norm
  > 5-number summary: 0.00, 0.20, 0.33, 0.52, 0.75

- Constrained Grid \((P,\pi)\) with 10 expected gradients closest to 0:

<table>
<thead>
<tr>
<th>( P_1 )</th>
<th>( P_2 )</th>
<th>( \pi_1 )</th>
<th>( \pi_2 )</th>
<th>( \pi_3 )</th>
<th>Norm</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3333</td>
<td>0.3333</td>
<td>0.1000</td>
<td>0.1000</td>
<td>0.9000</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.3333</td>
<td>0.3333</td>
<td>0.0500</td>
<td>0.1000</td>
<td>0.9000</td>
<td>0.0146</td>
</tr>
<tr>
<td>0.3333</td>
<td>0.3333</td>
<td>0.1000</td>
<td>0.0500</td>
<td>0.9000</td>
<td>0.0166</td>
</tr>
<tr>
<td>0.3333</td>
<td>0.3333</td>
<td>0.1000</td>
<td>0.1000</td>
<td>0.9500</td>
<td>0.0169</td>
</tr>
<tr>
<td>0.3333</td>
<td>0.3333</td>
<td>0.1000</td>
<td>0.0500</td>
<td>0.9500</td>
<td>0.0205</td>
</tr>
<tr>
<td>0.3333</td>
<td>0.3333</td>
<td>0.0500</td>
<td>0.0500</td>
<td>0.9000</td>
<td>0.0225</td>
</tr>
<tr>
<td>0.3333</td>
<td>0.3333</td>
<td>0.0500</td>
<td>0.1000</td>
<td>0.9500</td>
<td>0.0238</td>
</tr>
<tr>
<td>0.3333</td>
<td>0.3333</td>
<td>0.0500</td>
<td>0.0500</td>
<td>0.9500</td>
<td>0.0268</td>
</tr>
<tr>
<td>0.2654</td>
<td>0.0132</td>
<td>0.1000</td>
<td>0.4000</td>
<td>0.6000</td>
<td>0.0275</td>
</tr>
<tr>
<td>0.2654</td>
<td>0.0132</td>
<td>0.0500</td>
<td>0.4000</td>
<td>0.6000</td>
<td>0.0293</td>
</tr>
</tbody>
</table>
Example: Uniqueness of Target Parameters

- Test case: $P_1 = P_2 = 1/3; \ (\pi_1, \pi_2, \pi_3) = (.1, .1, .9)$

- **Unconstrained** (wider) Grid with 10 expected gradients closest to 0:

<table>
<thead>
<tr>
<th>$P_1$</th>
<th>$P_2$</th>
<th>$\pi_1$</th>
<th>$\pi_2$</th>
<th>$\pi_3$</th>
<th>Norm</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3333</td>
<td>0.3333</td>
<td>0.1000</td>
<td>0.1000</td>
<td>0.9000</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.2119</td>
<td>0.5761</td>
<td>0.1000</td>
<td>0.5000</td>
<td>0.3000</td>
<td>0.0268  (#)</td>
</tr>
<tr>
<td>0.2119</td>
<td>0.5761</td>
<td>0.1000</td>
<td>0.5000</td>
<td>0.1000</td>
<td>0.0305</td>
</tr>
<tr>
<td>0.0132</td>
<td>0.2654</td>
<td>0.1000</td>
<td>0.9000</td>
<td>0.3000</td>
<td>0.0361</td>
</tr>
<tr>
<td>0.2119</td>
<td>0.5761</td>
<td>0.3000</td>
<td>0.5000</td>
<td>0.1000</td>
<td>0.0378</td>
</tr>
<tr>
<td>0.0132</td>
<td>0.2654</td>
<td>0.5000</td>
<td>0.1000</td>
<td>0.5000</td>
<td>0.0382</td>
</tr>
<tr>
<td>0.0132</td>
<td>0.2654</td>
<td>0.3000</td>
<td>0.9000</td>
<td>0.3000</td>
<td>0.0385</td>
</tr>
<tr>
<td>0.0132</td>
<td>0.2654</td>
<td>0.3000</td>
<td>0.1000</td>
<td>0.5000</td>
<td>0.0387</td>
</tr>
<tr>
<td>0.1554</td>
<td>0.4223</td>
<td>0.1000</td>
<td>0.7000</td>
<td>0.3000</td>
<td>0.0395</td>
</tr>
<tr>
<td>0.0132</td>
<td>0.2654</td>
<td>0.7000</td>
<td>0.1000</td>
<td>0.5000</td>
<td>0.0403</td>
</tr>
</tbody>
</table>

> Frame of reference: norm(0.3333, 0.3333, 0.05, 0.05, 0.95) = 0.0268 (#)
Example: Uniqueness of Target Parameters

- **Actual basic functioning data**

- **Constrained Grid** \((P, \pi)\) with 10 expected gradients closest to 0:

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(P_1)</td>
<td>(P_2)</td>
<td>(\pi_1)</td>
<td>(\pi_2)</td>
<td>(\pi_3)</td>
<td>Norm</td>
</tr>
<tr>
<td>0.4223</td>
<td>0.4223</td>
<td>0.0500</td>
<td>0.3000</td>
<td>0.9000</td>
<td>0.0170</td>
</tr>
<tr>
<td><strong>0.4223</strong></td>
<td><strong>0.2900</strong></td>
<td><strong>0.0500</strong></td>
<td><strong>0.1000</strong></td>
<td><strong>0.8000</strong></td>
<td><strong>0.0177</strong></td>
</tr>
<tr>
<td>0.4223</td>
<td>0.4223</td>
<td>0.0500</td>
<td>0.3000</td>
<td>0.8000</td>
<td>0.0183</td>
</tr>
<tr>
<td>0.4223</td>
<td>0.4223</td>
<td>0.1000</td>
<td>0.3000</td>
<td>0.9000</td>
<td>0.0198</td>
</tr>
<tr>
<td>0.4223</td>
<td>0.4223</td>
<td>0.0500</td>
<td>0.3000</td>
<td>0.7000</td>
<td>0.0209</td>
</tr>
<tr>
<td><strong>0.4223</strong></td>
<td><strong>0.2900</strong></td>
<td><strong>0.1000</strong></td>
<td><strong>0.1000</strong></td>
<td><strong>0.8000</strong></td>
<td><strong>0.0209</strong></td>
</tr>
<tr>
<td>0.4223</td>
<td>0.4223</td>
<td>0.1000</td>
<td>0.3000</td>
<td>0.8000</td>
<td>0.0228</td>
</tr>
<tr>
<td>0.4223</td>
<td>0.4223</td>
<td>0.1000</td>
<td>0.3000</td>
<td>0.9500</td>
<td>0.0235</td>
</tr>
<tr>
<td>0.4223</td>
<td>0.4223</td>
<td>0.0500</td>
<td>0.3000</td>
<td>0.9500</td>
<td>0.0246</td>
</tr>
<tr>
<td>0.4223</td>
<td>0.4223</td>
<td>0.1000</td>
<td>0.3000</td>
<td>0.7000</td>
<td>0.0248</td>
</tr>
</tbody>
</table>
DISCUSSION

• What I delineated
  — A philosophy
    > Fit an ideal model
    > Determine the nature of measurement achieved in fact
  — Theory: On the nature of measurement
  — Methodology: To implement the philosophy

• What’s next?
  — Uniqueness of target: Displays, complicated models
  — Implications: Delineation of plausible models
DISCUSSION

• A primary issue: Why a hierarchical model at all?
  — e.g. PTSD: Why not DSM $Y$, delineate its measurement properties

1) Nosology
  a. Central role of cond. independence, non-diff. measurement.
  b. Guidance in creating, say, three rather than two groups.

2) The quest for the “ideal”
  a. Could have turned out that LCR much less subject to NDM, than DSM: i.e. issue with diagnostic criteria rather than items.
  b. In fact: LCR and DSM about equally subject to NDM
  c. Ultimate recommendation: DSM

• Some other issues
  — A seduction: Accuracy property re $V_i$ only for model fit in first stage
  — Why not be Bayesian?
  — Should one be parsimonious or complex?