The Role of Assumptions in Latent Variable Model Estimation: Elucidation and Tuning

Karen Bandeen-Roche
Department of Biostatistics
Johns Hopkins University

Department of Epidemiology and Biostatistics
Memorial Sloan-Kettering Cancer Center
May 30, 2007

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 <u>latent</u> (underlying) U_i:

  $$F_{Y|U,x}(y|U=u,x;\pi) \qquad \textit{(Measurement)}$$

  > Focus on distribution, regression re U_i:

  $$F_{U|x}(u|x;\beta) \qquad \textit{(Structural)}$$

  > 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***

- Application: Post-traumatic Stress Disorder

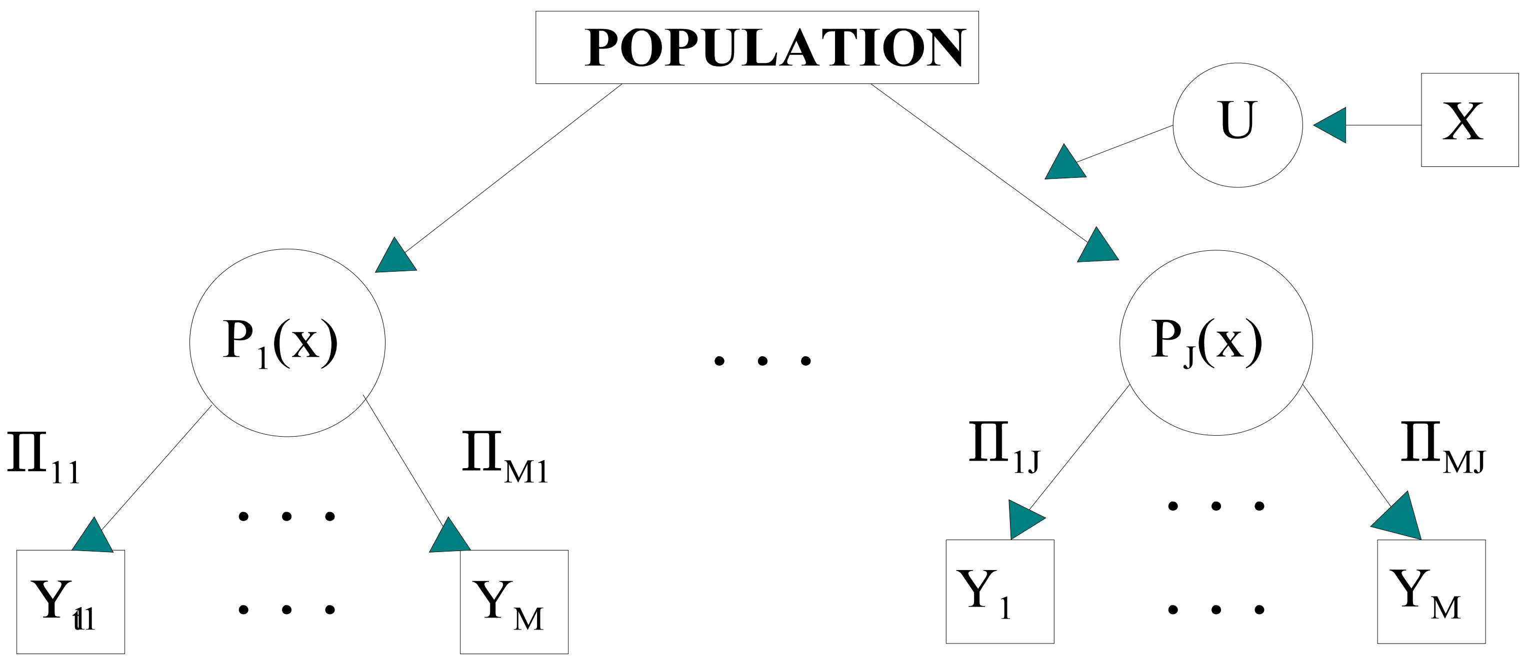
- Refocusing: Methodology to counterbalance competing assumptions

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



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

References: Dayton & Macready 1988, van der Heidjen et al., 1996; Bandeen-Roche et al., 1997

Latent Class Regression (LCR) Model

- **Model**:

$$\boldsymbol{f_{Y|x}(y|x) = \sum_{j=1}^{J} P_j(x,\beta) \prod_{m=1}^{M} \pi_{mj}^{y_m}(1-\pi_{mj})^{1-y_m}}$$

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

- **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:

  — <u>Asymptotically</u>:

  > 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 (β^*,π^*).

  —Y_i asymptotically equivalent in distribution to Y^*, generated as:

  i) Generate U_i^* — distribution determined by (β^*,π^*), $G_{Y|x}(y|x)$;

  ii) Generate Y^*—distribution determined by (β^*,π^*), $G_{Y|x}(y|x)$, 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

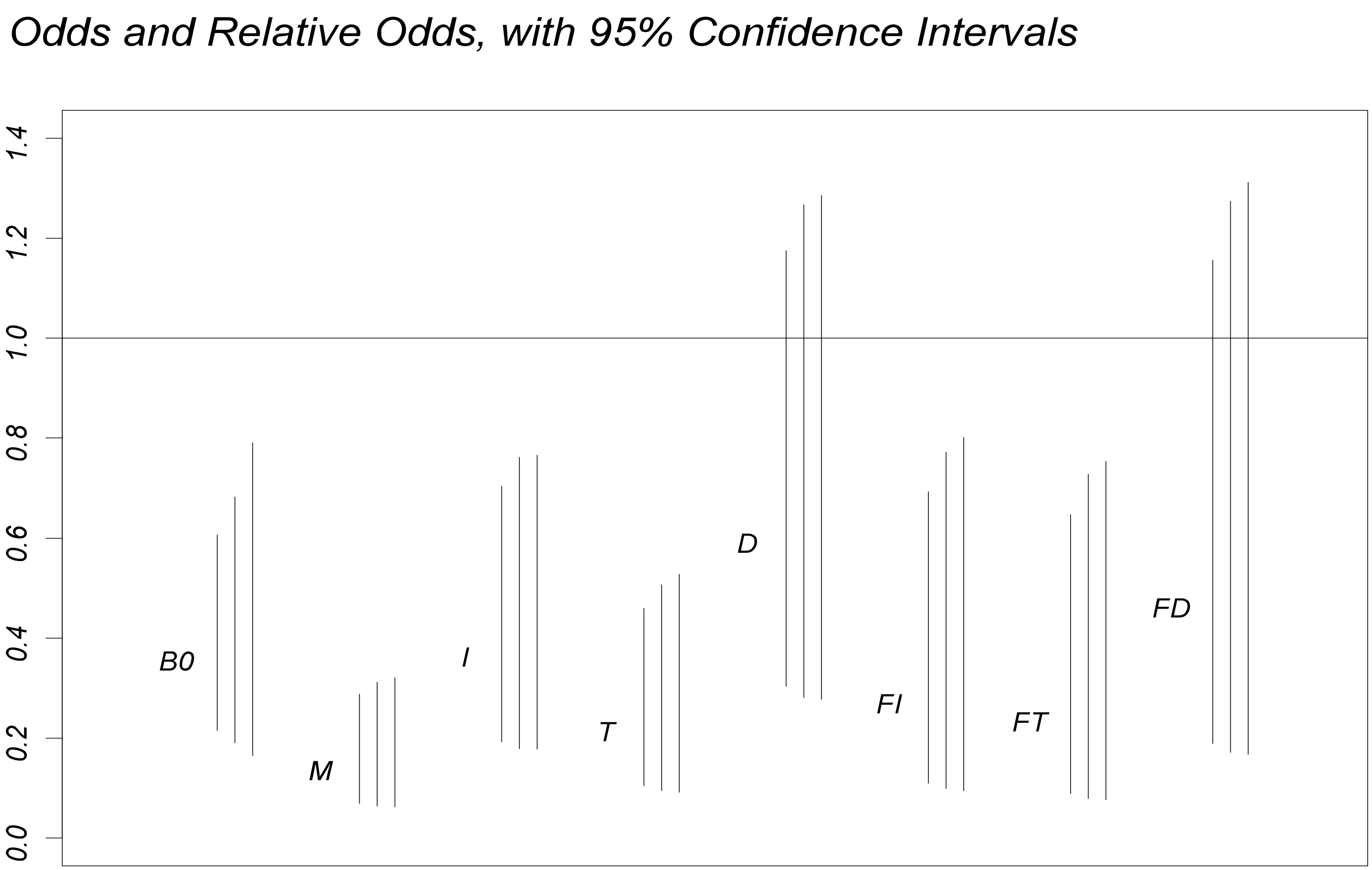
Gender	**Trauma Type**: **percentage distribution**				**n**
	Personal Assault	*Other Injury*	*Trauma to loved one*	*Sudden death*	
Male	14.2	**37.7**	26.9	21.3	964
Female	14.3	26.3	**32.2**	**27.2**	863
Total	14.2	32.3	29.4	24.1	1827

- 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

SYMPTOM CLASS	**SYMPTOM (prevalence)**	**SYMPTOM PROBABILITY (π)**		
		Class 1 - NO PTSD	Class 2 - SOME SYMPTOMS	Class 3 - PTSD
RE-EXPERIENCE	Recurrent thoughts (.49)	.20	**.74**	.96
	Distress to event cues (.42)	.12	**.68**	.88
	Reactivity to cues (.31)	.05	**.51**	.77
AVOIDANCE/ NUMBING	Avoid related thoughts (.28)	.08	.37	.75
	Avoid activities (.24)	.05	.34	.66
	Detachment (.15)	**.01**	**.14**	**.64**
INCREASED AROUSAL	Difficulty sleeping (.19)	.02	.18	.78
	Irritability (.21)	.02	.22	.83
	Difficulty concentrating (.25)	.03	.30	.89
MEAN PREVALENCE-BASELINE		.52	.33	.14

[Omitted: nightmares, flashback; **amnesia, ↓interest, ↓affect, short future;** hypervigilance, startle]

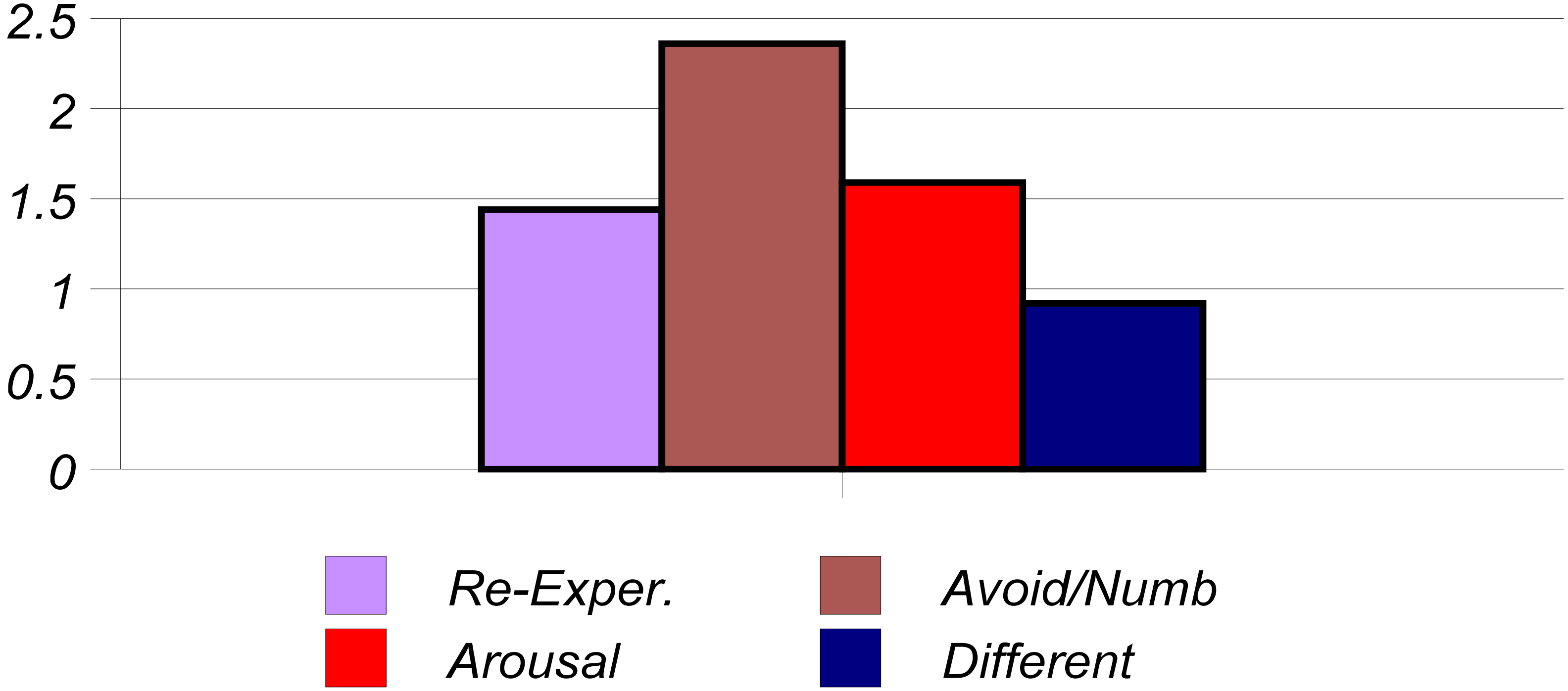

Odds and Relative Odds, with 95% Confidence Intervals
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
B0
M
I
T
D
FI
FT
FD

PTSD: DIAGNOSIS, LCR MEASUREMENT MODEL

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

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

Diagnosis: Conditional Independence (Pairwise ORs)



Re-Exper.	*1.44*
Avoid/Numb	*2.36*
Arousal	*1.59*
Different	*0.92*

Summary
PTSD Analysis

- The analysis hypothesizes that PTSD is
  - — a syndrome comprising <u>**unaffected**</u>, <u>**subclinically affected**</u>, and <u>**diseased**</u> subpopulations of those suffering traumas
  - — reported homogeneously within subpopulations
- The hypotheses are consistent with current diagnostic criteria
- <u>Gender x type interactions</u>: 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

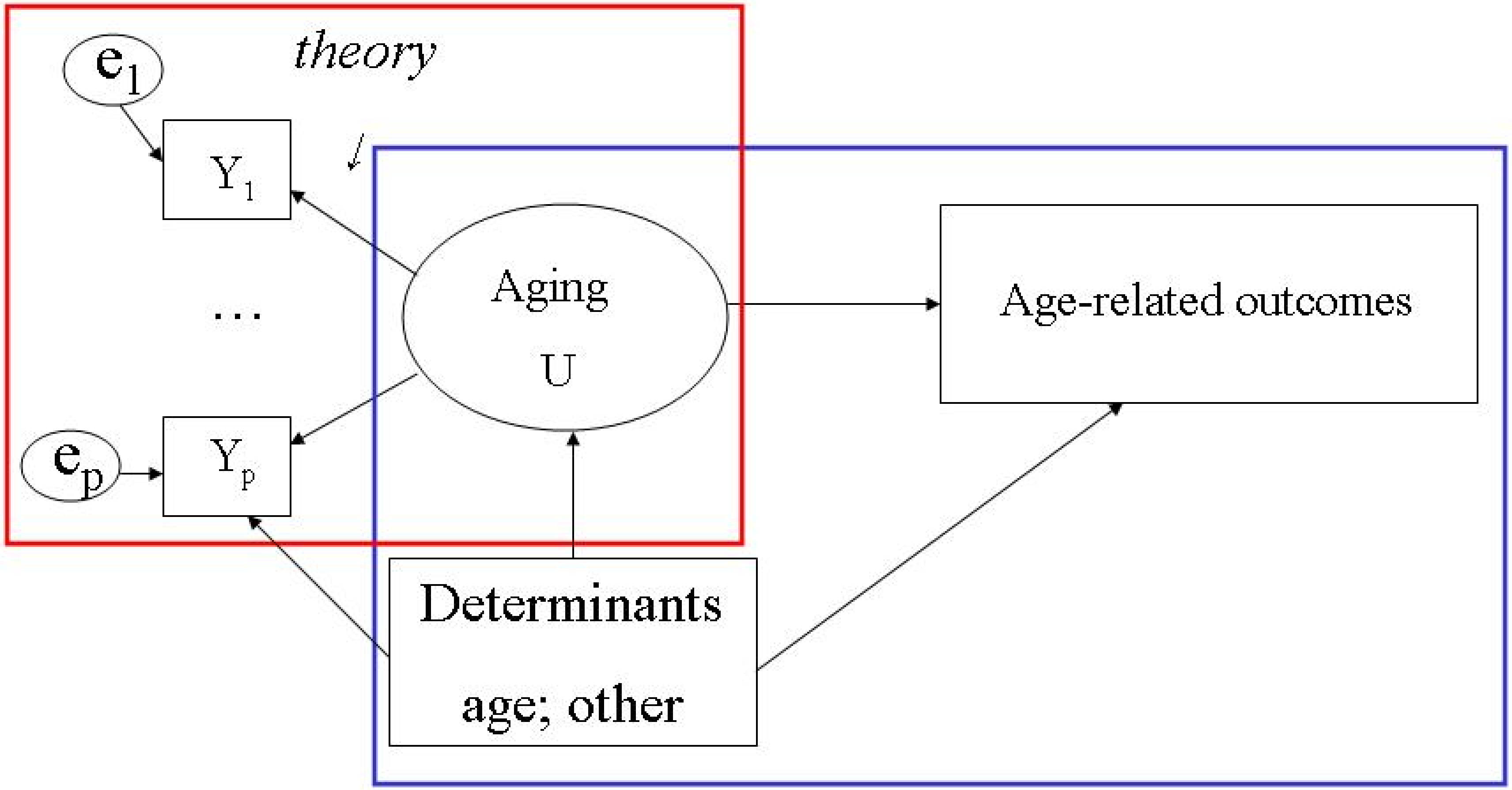
Summary

- 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
  - New work: On compromise between potentially competing validation criteria
- Strengths / benefits
  - Improved use / usefulness of latent variable models
  - Attention to estimability
  - Allows some distrust of the data

Refocusing of the aim: Measurement

- A well defined target; a less-well-defined operationalization
- We seek measures that have validity for representing their targets
  - LV assumptions encode validity criteria,...
  - ... some better than others
- Objective: Method to unify multiple validation aspects in 1 analysis
  - Balancing potentially conflicting aspects
  - **<u>Today's focus</u>**: "scale" weighing the balance

Measurement
Latent Variable Paradigm



Rationale of the New Work

- Which deserves pre-eminence?
  - Internally validating assumptions?
  - Externally validating assumptions?
  - Some compromise?

- Proposal: Allow compromise via "penalized" fitting

Implementing penalization

- On LCR kernel: Houseman, Coull & Betensky, *BMCS, 2006*

- On LCR mixing distribution: Sheppard Ph.D. thesis

- Key questions
  - Form of the penalty
  - Different purpose than usual?
  - What is the objective function?

Penalization
Very brief background

- Fitting: minimize

$$-2 \ln L(\theta;Y,x) + \lambda g(\theta)$$

- Examples
  - "Ridge": $g(\theta) = \Sigma_j \theta_j^2$
  - "Lasso": $g(\theta) = \Sigma_j |\theta_j|$

Green, Int Stat Rev, 1987; Tibshirani, JRSS-B, 1996

Penalization
Very brief background

- A useful equivalence: penalized fit obtains via formulating parameters as crossed random effects
  - "Ridge": $\theta_j \sim N(0,\sigma\{\lambda\}^2)$
  - "Lasso": $\theta_j \sim$ double $\exp(0,h\{\lambda\})$

Wahba, JRSS-B, 1978; Ngo & Wand, J Stat Software, 2004

Form of the penalty
Current case

- Usual purpose: regularization
- Here: secondary validation
- <u>Discriminant hypothesis</u>: Genotypes predispose individuals to only one “subtype” of depression

Form of the penalty
Genetic subtypes example

- Say, LCR with one normal class (1) and two disordered classes (2, 3):

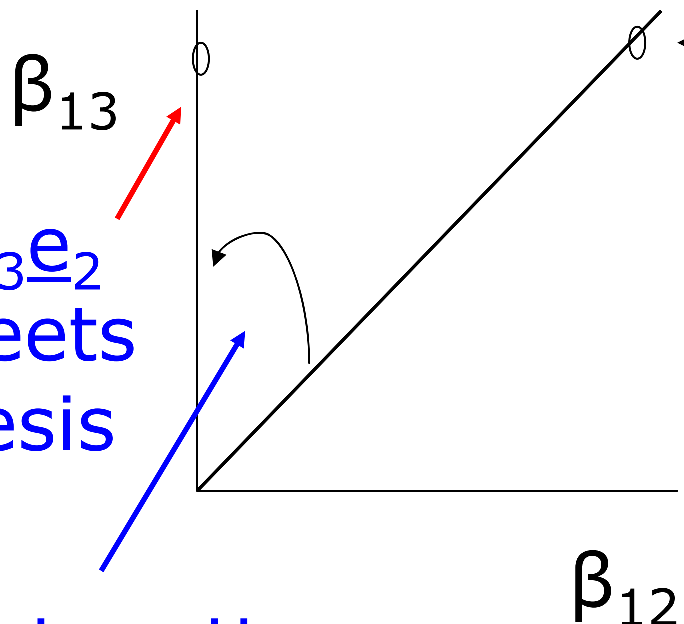
- Hypothesis: β_{1j} negligible, and $\beta_{1j'}$ appreciable, in

$$\log\left(\frac{p_k}{p_1}\right) = \beta_{0k} + \beta_{1k}x$$

with p_k = pr(class k); x=genotype indicator;
k=2,3; j, j′ ϵ {2,3}; j ≠ j′

Form of the penalty
Genetic subtypes example

- Ridge, lasso not quite right



$\beta_{12}\underline{e}_1 + \beta_{13}\underline{e}_2$ here contradicts hypothesis

$\beta_{12}\underline{e}_1 + \beta_{13}\underline{e}_2$ here meets hypothesis

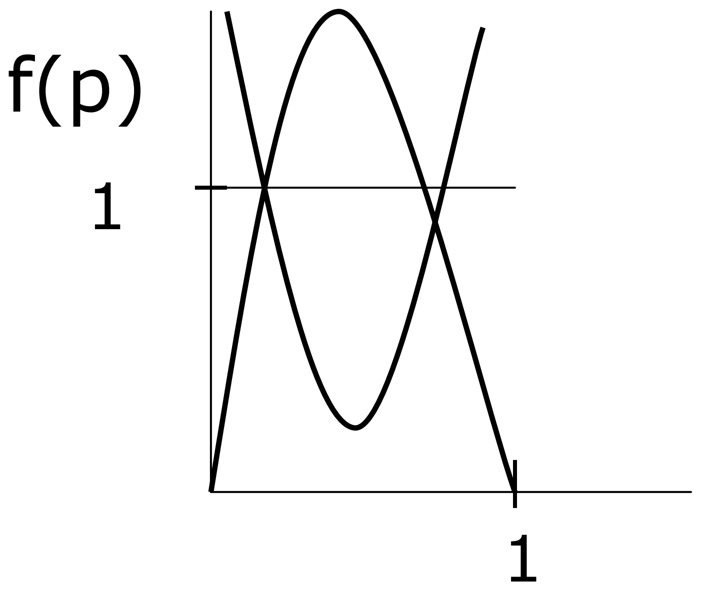
What matters: angle α

Form of the penalty
Genetic subtypes example

- Approach 1
  - Consider α ϵ [0,90] degrees
  - Desired orientations are cos(α)=1, sin(α)=1
  - i.e., goal: minimize cos(α)+sin(α)
  - i.e. minimize $$\frac{|\beta_{12}|+|\beta_{13}|}{\sqrt{\beta_{12}^2+\beta_{13}^2}}$$

Form of the penalty
Genetic subtypes example

- Approach 2
  - Write $\beta_{12} = p\beta$; $\beta_{13} = (1\text{-}p)\beta$
  - Fit with beta random effect on p



  - Generalization: $\underline{\beta} = \underline{p}\beta$, $\underline{p} \sim$ Dirichlet

Fitting
Approach 2

- E-M algorithm: quite straightforward
- E-step: Computes posterior class membership probabilities given current parameter iterates
- M-step: minimize (e.g. Nelder-Mead)

$$-\sum_{i=1}^{n}\sum_{j=1}^{J} h(j|data)\ln[f_{U|x}(u|x,p,\beta)] + \boxed{(1-\frac{\Delta}{2})\ln[p(1-p)]}$$

Simulation study
Three-class model

- 100 reps; single x~Unif(-.5,.5); n=1000
- Poly Log Reg: $\beta_{02}=\beta_{03}=0$; $\beta_{12}=-1.4$;
  $\beta_{13}\in\{0,-1.4\}$
- Measurement:

Class 1	Class 2	Class 3
.15	.85	.85
.15	.85	.85
.15	.85	.85
.15	.13	.85
.15	.13	.85

Simulation study
Three-class model

- Two assumption scenarios
  - Frank LCR
  - Differential measurement: First three items have increased log(odds =1) per unit x of 1.4 in class 3

Simulation Study
Diff. Meas.; β_{13}=0; β_{12}=-1.4

Param.	Penalized		LCR	
	Estimate	SE	Estimate	SE
β_{13}	-0.04	0.14	-0.54	0.31
β_{12}	-0.79	0.30	-1.01	0.34

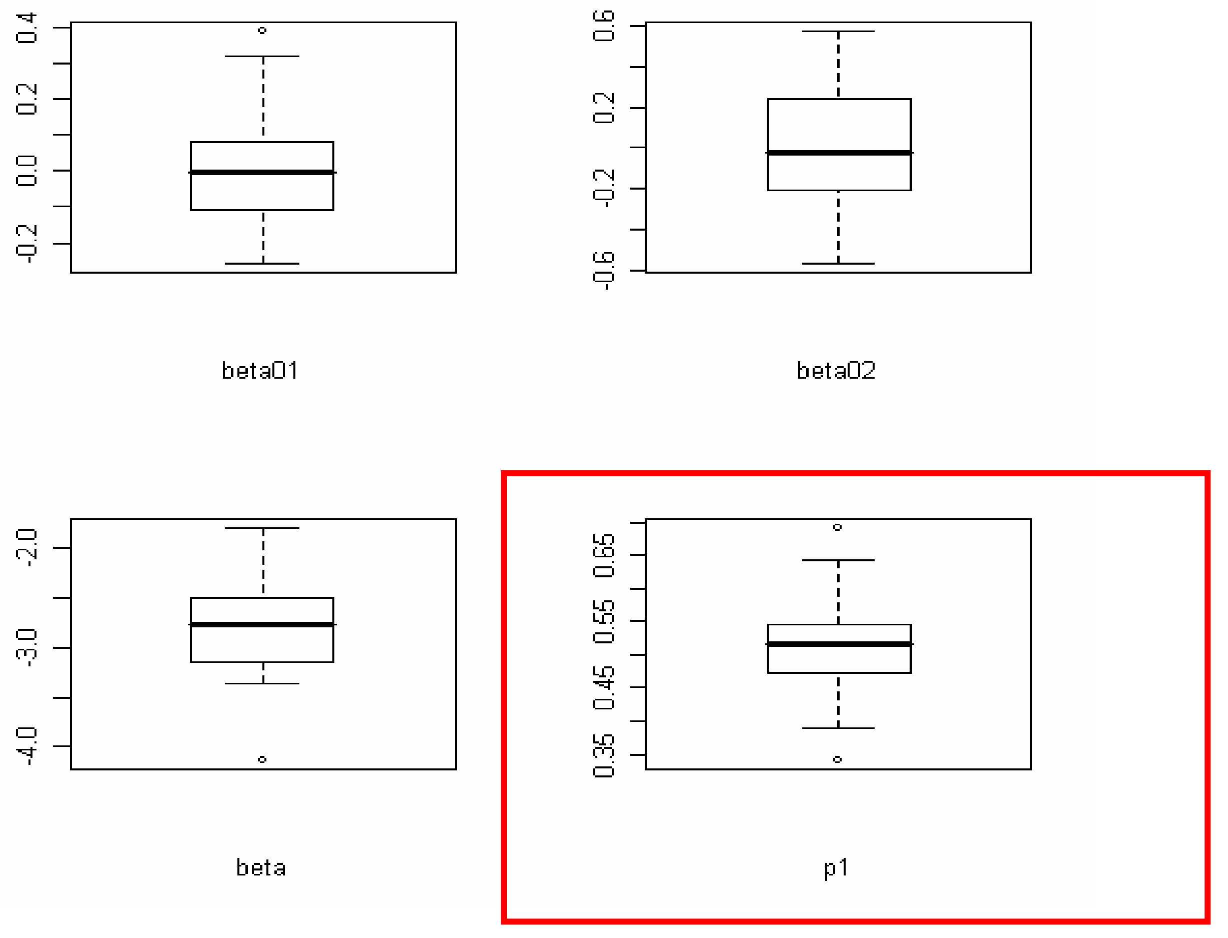
Simulation Study
Non-diff meas; β_{13}=0; β_{12}=-1.4

Param.	Penalized		LCR	
	Estimate	SE	Estimate	SE
β_{13}	0	0	0.04	0.32
β_{12}	-1.42	0.35	-1.41	0.38

Simulation Study
Non-diff meas; $\beta_{12}=\beta_{13}=-1.4$

Param.	Penalized		LCR	
	Estimate	SE	Estimate	SE
β_{13}	-1.45	0.34	-1.45	0.30
β_{12}	-1.38	0.31	-1.38	0.31

Simulation Study
Non-diff meas; $\beta_{12}=\beta_{13}=-1.4$



Discussion

- A primary issue: Why a hierarchical model at all?

  — PTSD: Why not DSM *Y*, delineate 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

Discussion

- Beyond delineation of assumptions….

- Further Work: Uniqueness of target
  - Delineation of plausible models
  - Displays, complicated models
  - ***Implication***: Guidance on parsimony versus complexity

- Further work: Big picture for validation compromise
  - How does measurement conform?
  - How should one determine the magnitude of the compromise?
    - Empirical adjudication
    - Clinical / scientific utility
    - Ultimate: Gold standard (aging: telomeric shortening)

- Why not be Bayesian?

Implications

- Refined understanding of health states and their measurement
  - Integrating biology
  - Increasing sensitivity, specificity
- Heightened accuracy, precision for
  - Delineating etiology
  - Developing and targeting interventions

Acknowledgments

- Hopkins Colleagues

  Linda Fried, Ron Brookmeyer, Yi Huang, Jeremy Walston, Qian-Li Xue, Scott Zeger

- Colleagues outside of Hopkins

  Naomi Breslau, Howard Chilcoat, Luigi Ferrucci, Jack Guralnik, Don Ingram, Richard Miller

- Funding / Institutional Support

  Johns Hopkins Older Americans Independence Center, Alliance for Aging Research, National Institute on Aging, National Institute of Mental Health