Penalized Latent Class Regression: Incorporating Scientific Knowledge into Measurement Models

Jeannie-Marie (Sheppard) Leoutsakos, Karen Bandeen-Roche, PhD, Peter Zandi, PhD, William Eaton, PhD

Johns Hopkins Bloomberg School of Public Health

JSM Session 320: 8/8/2006

The Taxonomy Problem

- A major roadblock in psychiatric epidemiology
- Existing definitions often based on presentation.
- Iterative process: disorder definitions aid identification of risk factors and mechanisms which would in turn lead to better disorder definitions.

Depression

- Depression is heterogeneous in presentation
- Depression may be heterogeneous in its etiopathogenesis
- 5-HTT genotype may play a role in one of these mechanisms
- 5-HTT genotype may therefore be useful in elucidating subtypes of depression

Latent Class Regression Model

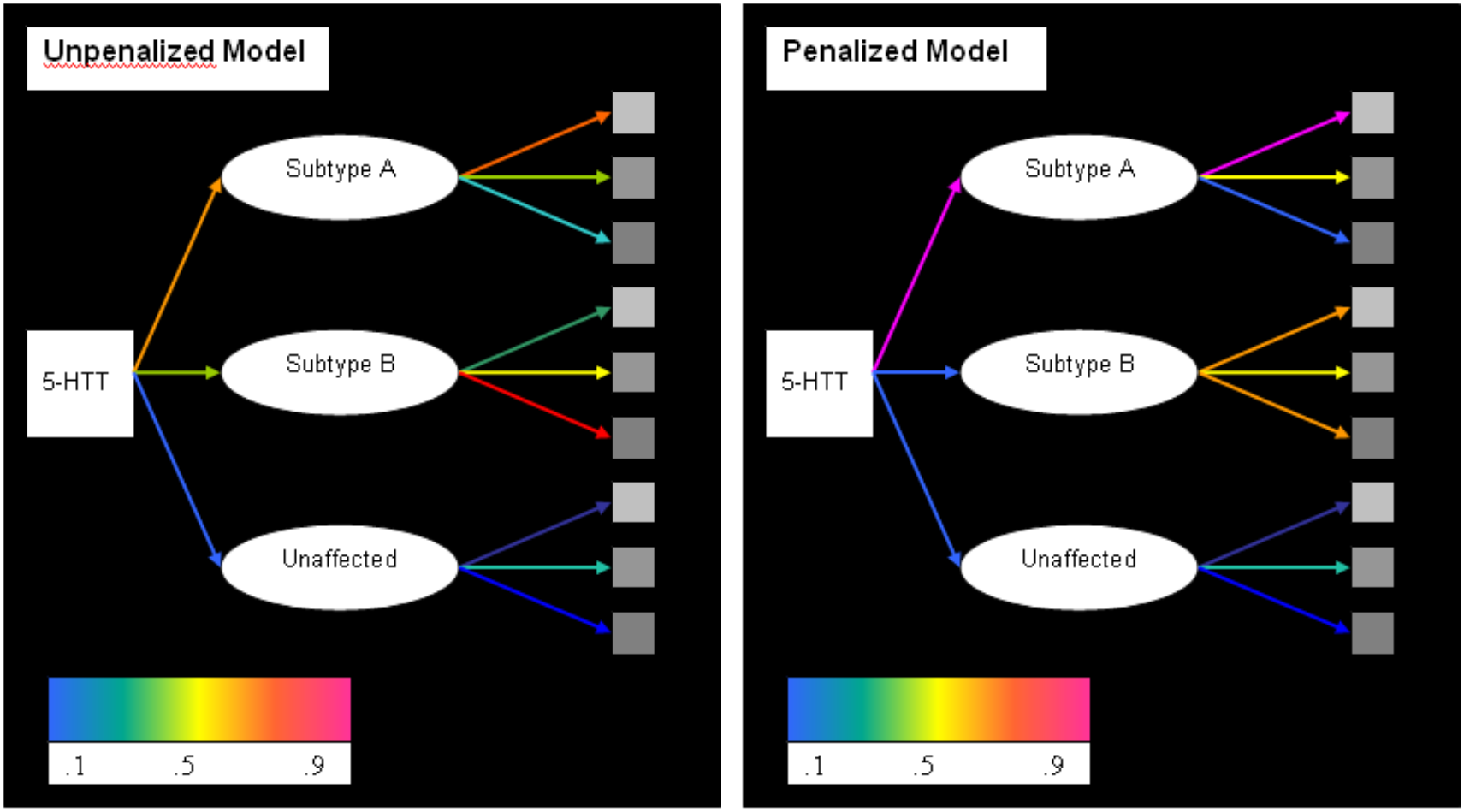
$$L(\pi,\beta)=\sum_{i=1}^{N}\left[\sum_{j=1}^{J}\eta_j(x_i'\beta_j)\prod_{m=1}^{M}\pi_{jm}^{y_{im}}(1-\pi_{jm})^{1-y_{im}}\right]$$

- i=1,…,N : individuals
- j=1,..,J : latent classes
- y_{im}=1,…M : latent class indicators
- X_i=0 or 1 : presence of genotype

Model Assumptions

- Conditional Independence – given class membership, latent class indicators are independent.
- Non-Differential Measurement – given class membership, latent class indicators and covariates are independent.

Penalizing the Model



Penalized Latent Class Regression

- Building on work by Houseman, *et al**, two types of bridge penalties, lasso and ridge were examined.

$$L_c(\beta,\pi;\lambda) = L(\beta,\pi) - C(\beta,\lambda)$$

$$C(\beta,\Lambda) = \sum_{j=1}^{J}\sum_{i=1}^{N}\lambda_{ji}\left|\beta_{ji}\right|^{\gamma}$$

* http://www.bepress.com/cgi/viewcontent.cgi?article=1022&context=harvardbiostat

Cross Validated Log-Likelihood Loss

$$\hat{R} = 2L(\pi, \beta) - 2tr(H_C^{-1} nV)$$

Ridge: $H_c^{-1} = H + 2\Lambda$

Lasso: $H_c^{-1} = H + W$

$$W = \begin{bmatrix} |\beta_{pj}|^{-1} & 0 & 0 \\ 0 & |\beta_{pj}|^{-1} & 0 \\ 0 & 0 & |\beta_{PJ}|^{-1} \end{bmatrix}$$

Simulation Studies – Plan I

All scenarios run with both ridge and lasso

- Model is "correct"
  - 3,4,5 class models; 1-3 covariates
    - All non-intercept β's penalized (constant lambda).
    - Only β_1 penalized.

Simulation Studies - Plan II

Model is “Incorrect”

- Conditional dependence
  - Exchangeable correlations between indicators
  - Two indicators correlated within class
- Differential Measurement
  - Association only between covariate(s) and 1- 2 indicators
  - Associations between covariate(s) and 1-2 indicators AND between covariates and class assignment

Simulation Results

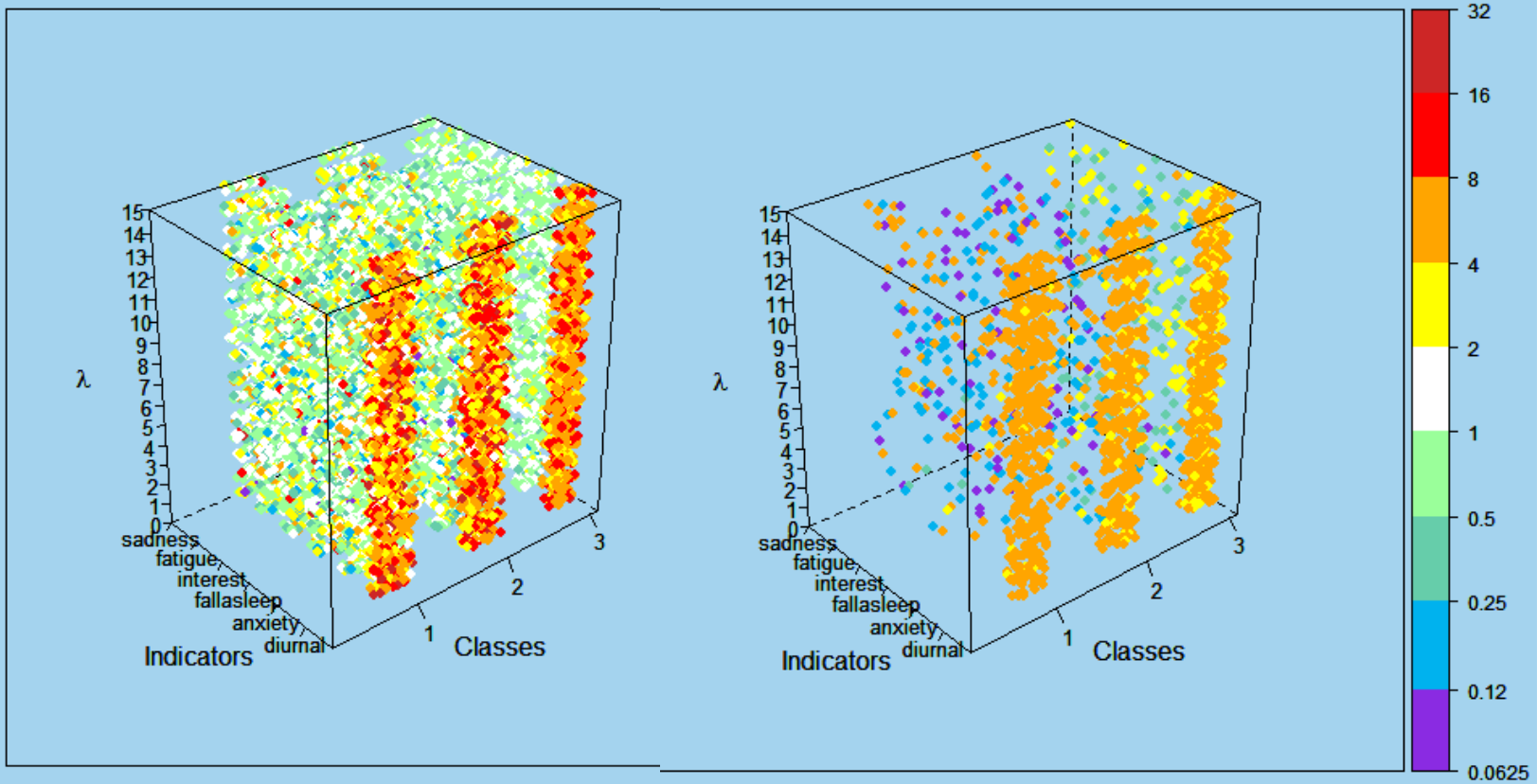
- Except in cases of differential measurement, optimal penalty was almost always 0.
- LASSO penalty seemed to flatten all beta's

Simulation Results – Parameter Estimates

β_{jp}	True Value	λ=0	λ=5	λ=10	λ=15
β_{10}	-1.38	-1.08 (.16)	-1.11 (.13)	-1.06 (.12)	-1.11 (.60)
β_{11}	0	0.02 (.42)	-0.01 (.56)	0.05 (.27)	0.01 (.21)
β_{20}	-1.38	-1.04 (.44)	-1.17 (.29)	-0.97 (.11)	-1.14 (2.09)
β_{21}	0	-0.14 (.47)	-0.06 (.47)	-0.00 (.29)	-0.02 (.66)

λ value	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Count of optimal λ	29	1	0	3	11	16	3	1	1	3	4	1	2	1	0	24

Penalty and Differential Measurement



Application - Background

- Epidemiologic Catchment Area Study
- Began as 5-site study in 1981
- Intended to ascertain incidence and prevalence of major mental disorders.
- Most recent wave (4) included a blood draw.
  - 1071 interviews, 885 samples collected, 673 5-HTT genotypes available. Effective sample size = 652

Application – Study Sample

- Affective symptom in the 30 days prior

	%		
Female	62.7		
Caucasian	74.8		
5-HTT Genotype	SS - 12.2	SL – 47.4	LL – 40.3
Sadness	6.1		
Anxiety	9.8		
Trouble falling asleep	13.5		
Loss of interest	15.3		
Fatigue	9.0		
Diurnal variation	8.4		

Class on S Allele

	Unpenalized			Ridge Penalty		
				Best Penalty (lamda = 5)		
	Class 1	Class 2	Class 3	Class 1	Class 2	Class 3
Sadness	.73 (.26)	.14 (.07)	.02 (.01)	.63 (.17)	.12 (.08)	.02 (.01)
Anxiety	.92 (.26)	.11 (.11)	.06 (.01)	.79 (.18)	.07 (.12)	.06 (.01)
Falling Asleep	.66 (.13)	.44 (.09)	.07 (.01)	.63 (.12)	.44 (.10)	.07 (.01)
Interest	.55 (.12)	.40 (.09)	.10 (.02)	.53 (.11)	.39 (.10)	.10 (.01)
Fatigue	.32 (.11)	.54 (.15)	.02 (.01)	.31 (.10)	.62 (.14)	.02 (.01)
AM/PM	.30 (.12)	.54 (.14)	.02 (.01)	.32 (.11)	.58 (.14)	.02 (.01)
Intercept	-3.18 (.87)	-1.95 (.51)		-2.63 (.33)	-2.52 (.22)	
>=1 S Allele	0.28 (.59)	-0.21 (.34)		-0.11 (.48)	0.32 (.40)	

Class on S allele and Race

	Unpenalized			Ridge Penalty Best Penalty (lambda = 6)		
	Class 1	Class 2	Class 3	Class 1	Class 2	Class 3
Sadness	.65 (.19)	.13 (.07)	.02 (.01)	.62 (.16)	.13 (.07)	.02 (.01)
Anxiety	.86 (.18)	.08 (.12)	.05 (.01)	.82 (.16)	.06 (.12)	.05 (.01)
Falling Asleep	.64 (.12)	.47 (.09)	.07 (.01)	.63 (.12)	.47 (.10)	.07 (.01)
Interest	.53 (.11)	.42 (.15)	.10 (.02)	.52 (.10)	.41 (.14)	.11 (.02)
Fatigue	.29 (.10)	.62 (.15)	.03 (.01)	.29 (.10)	.64 (.14)	.03 (.01)
AM/PM	.30 (.11)	.59 (.15)	.02 (.01)	.30 (.11)	.61 (.01)	.02 (.01)
Intercept	-3.68 (.81)	-2.75 (.72)		-3.37 (.62)	-2.98 (.51)	
>=1 S Allele	0.20 (.48)	-0.10 (.38)		-0.13 (	0.09	
Race (Caucasian)	1.05 (.56)	0.74 (.49)		1.03 (.55)	0.78 (.50)	

Future Directions

- Faster software
- Explore other types penalties
- Explore wider range of candidate lambdas
- Effect of sample size
- Other genes/study samples

Acknowledgements

- NIMH Training Program in Psychiatric Epidemiology T32-MH-14592
- Adam Kaplin, MD, PhD, Johns Hopkins Psychiatry
- Epidemiological Catchment Area Study Staff
  - Yulan Ding, MS
  - Dawn Childs