Mixed Models for Longitudinal Ordinal and Nominal Data

Hedeker, D. (2005). Generalized linear mixed models. In B. Everitt & D. Howell (Eds.), Encyclopedia of Statistics in Behavioral Science. Wiley.

Chapters 10 & 11 in Hedeker, D. & Gibbons, R.D. (2006). Longitudinal Data Analysis. Wiley.

Why analyze as ordinal?

- Efficiency: Armstrong & Sloan (1989, Amer Jrn of Epid) report efficiency losses between 89% to 99% comparing an ordinal to continuous outcome, depending on the number of categories and distribution within the ordinal categories.

- Bias: continuous model can yield correlated residuals and regressors when applied to ordinal outcomes, because the continuous model does not take into account the ceiling and floor effects of the ordinal outcome. This can result in biased estimates of regression coefficients and is most critical when the ordinal variables is highly skewed.

- Logic: continuous model can yield predicted values outside of the range of the ordinal variable.

Proportional Odds Model - McCullagh (1980)

$$\log\left[\frac{P(\mathsf{y} \leq c)}{1 - P(\mathsf{y} \leq c)}\right] = \gamma_c - \boldsymbol{x}'\boldsymbol{\beta}$$

$$\begin{aligned} c &= 1, \ldots, C-1 \text{ for the } C \text{ categories of the ordinal outcome} \\ \boldsymbol{x} &= \text{vector of explanatory variables (plus the intercept)} \\ \gamma_c &= \text{thresholds; reflect cumulative odds when } \boldsymbol{x} = 0 \text{ (for identification: } \gamma_1 = 0 \text{ or } \beta_0 = 0) \end{aligned}$$

- positive association between x and y is reflected by $\beta > 0$
- the effect of x is assumed to be the same for each cumulative odds ratio
- odds that the response is greater than or equal to c (for fixed c) is multiplied by e^{β} for every unit change in x:

$$\left[\frac{1 - P(\mathsf{y} \leq c)}{P(\mathsf{y} \leq c)}\right] = e^{-\gamma_c} \times (e^{\beta})^x$$

Ordinal Model for Dichotomous Response: same as it ever was!

$$\log\left[\frac{P(\mathsf{y}=0)}{1-P(\mathsf{y}=0)}\right] = 0 - \boldsymbol{x}'\boldsymbol{\beta}$$

$$\frac{P(\mathsf{y}=0)}{1-P(\mathsf{y}=0)} = \exp(0 - \boldsymbol{x}'\boldsymbol{\beta})$$

$$\frac{1-P(\mathsf{y}=0)}{P(\mathsf{y}=0)} = [\exp(0 - \boldsymbol{x}'\boldsymbol{\beta})]^{-1}$$

$$\frac{1-P(\mathsf{y}=0)}{P(\mathsf{y}=0)} = \exp(\boldsymbol{x}'\boldsymbol{\beta})$$

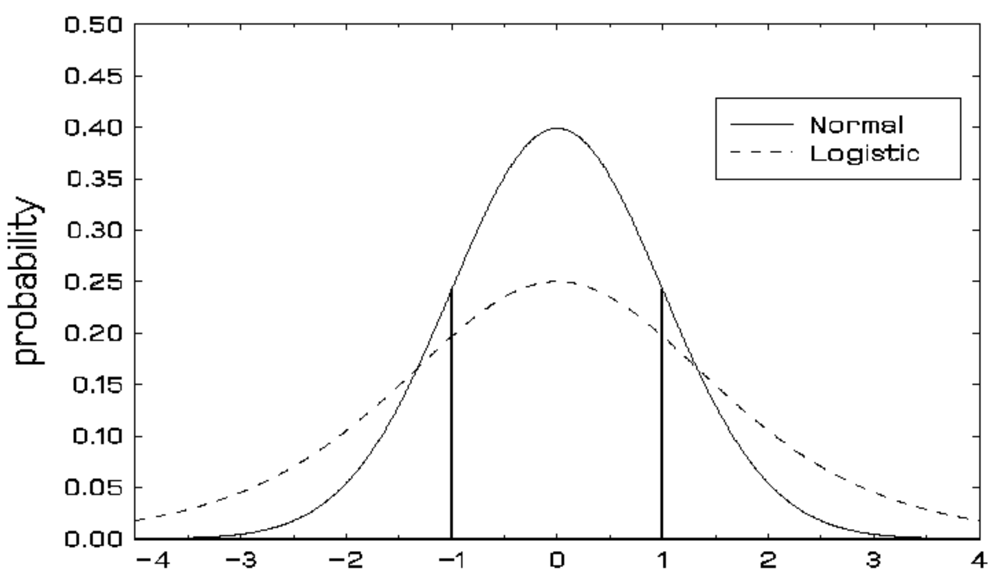
$$\log\left[\frac{P(\mathsf{y}=1)}{1-P(\mathsf{y}=1)}\right] = \boldsymbol{x}'\boldsymbol{\beta}$$

Ordinal Response and Threshold Concept

Continuous y_i - unobservable latent variable - related to ordinal response y_i via "threshold concept"

- series of threshold values $\gamma_1, \gamma_2, \ldots, \gamma_{C-1}$
- C = number of ordered categories
- $\gamma_0 = -\infty$ and $\gamma_C = \infty$

Response occurs in category c, $\quad \mathsf{y}_i = c \quad$ if $\quad \gamma_{c-1} < y_i < \gamma_c$



The Threshold Concept in Practice

"How was your day?"
(what is your level of satisfaction today?)

- Satisfaction may be continuous, but we sometimes emit an ordinal response:

Model for Latent Continuous Responses

Consider the model with p covariates for the latent response strength y_i ($i = 1, 2, \ldots, N$):

$$y_i = \boldsymbol{x}_i' \boldsymbol{\beta} + \varepsilon_i$$

- probit: $\varepsilon_i \sim$ standard normal (mean=0, variance=1)
- logistic: $\varepsilon_i \sim$ standard logistic (mean=0, variance=$\pi^2/3$)

$\Rightarrow \boldsymbol{\beta}$ estimates from logistic regression are larger (in abs. value) than from probit regression by approximately $\sqrt{\pi^2/3} = 1.8$

Underlying latent variable

- useful way of thinking of the problem
- not an essential assumption of the model

Mixed-effects ordinal logistic regression model

(Hedeker & Gibbons, 1994, 1996)

- $i = 1, \ldots N$ level-2 units (clusters or subjects)
- $j = 1, \ldots, n_i$ level-1 units (subjects or repeated observations)
- $c = 1, 2, \ldots, C$ response categories
- y_{ij} = ordinal response of level-2 unit i and level-1 unit j

How was your day? (asked repeatedly each day for a week)

Mixed-effects ordinal logistic regression model

$$\lambda_{ijc} = \log \left[\frac{P_{ijc}}{(1 - P_{ijc})} \right] = \gamma_c - (\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{v}_i)$$

- $P_{ijc} = \Pr\,(\mathsf{y}_{ij} \le c \mid \boldsymbol{v}\,;\, \gamma_c, \boldsymbol{\beta}, \boldsymbol{\Sigma}_v) \;=\; \frac{1}{1+\exp(-\lambda_{ijc})}$
- $p_{ijc} = \Pr\,(\mathsf{y}_{ij} = c \mid \boldsymbol{v}\,;\, \gamma_c, \boldsymbol{\beta}, \boldsymbol{\Sigma}_v) \;=\; P_{ijc} - P_{ijc-1}$

- $C - 1$ strictly increasing model thresholds γ_c
- $\boldsymbol{x}_{ij} = p \times 1$ covariate vector
- $\boldsymbol{z}_{ij} = r \times 1$ design vector for random effects
- $\boldsymbol{\beta} = p \times 1$ fixed regression parameters
- $\boldsymbol{v}_i = r \times 1$ random effects for level-2 unit $i \quad \sim N(\mathbf{0}, \boldsymbol{\Sigma}_v)$

Model for Latent Continuous Responses

Model with p covariates for the latent response strength y_{ij}:

$$y_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \upsilon_{0i} + \varepsilon_{ij}$$

where $\upsilon_{0i} \sim N(0, \sigma_\upsilon^2)$, and assuming

- $\varepsilon_{ij} \sim$ standard normal (mean 0 and $\sigma^2 = 1$) leads to mixed-effects ordinal probit regression
- $\varepsilon_{ij} \sim$ standard logistic (mean 0 and $\sigma^2 = \pi^2/3$) leads to mixed-effects ordinal logistic regression

Underlying latent variable

- not an essential assumption of the model
- useful for obtaining intra-class correlation (r)

$$r = \frac{\sigma_v^2}{\sigma_v^2 + \sigma^2}$$

and for design effect (d)

$$d = \frac{\sigma_v^2 + \sigma^2}{\sigma^2} = 1/(1 - r)$$

ratio of actual variance to the variance that would be obtained by simple random sampling (holding sample size constant)

Scaling of regression coefficients

Fixed-effects model

$\boldsymbol{\beta}$ estimates from logistic regression are larger (in abs. value) than from probit regression by approximately

$$\sqrt{\frac{\pi^2/3}{1}} = 1.8$$

because

- $V(y) = \sigma^2 = \pi^2/3$ for logistic
- $V(y) = \sigma^2 = 1$ for probit

Mixed-effects model

$\boldsymbol{\beta}$ estimates from mixed-effects (random intercepts) model are larger (in abs. value) than from fixed-effects model by approximately

$$\sqrt{d} = \sqrt{\frac{\sigma_v^2 + \sigma^2}{\sigma^2}}$$

because

- $V(y) = \sigma_v^2 + \sigma^2$ in mixed-effects (random intercepts) model
- $V(y) = \sigma^2$ in fixed-effects model

- difference depends on size of random-effects variance σ_v^2
- more complex for models with multiple random effects

Standardization of the random effects

With $\boldsymbol{\upsilon} = \boldsymbol{T}\boldsymbol{\theta}$ (where $\boldsymbol{T}\boldsymbol{T}' = \boldsymbol{\Sigma}_{\upsilon}$ is the Cholesky)

$$\lambda_{ijc} = \gamma_c - (\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{T}\boldsymbol{\theta}_i)$$

$\boldsymbol{\theta}_i$ are distributed multivariate standard normal

- $P_{ijc} = \Pr\,(\mathsf{y}_{ij} \leq c \mid \boldsymbol{\theta}\,;\, \gamma_c, \boldsymbol{\beta}, \boldsymbol{T}) \;=\; \frac{1}{1+\exp(-\lambda_{ijc})}$
- $p_{ijc} = \Pr\,(\mathsf{y}_{ij} = c \mid \boldsymbol{\theta}\,;\, \gamma_c, \boldsymbol{\beta}, \boldsymbol{T}) \;=\; P_{ijc} - P_{ijc-1}$

Conditional probability for response vector

$\mathsf{y}_i = n_i \times 1$ vector of responses from level-2 unit i

Conditional Probability

$$\ell(\mathsf{y}_i \mid \boldsymbol{\theta}; \gamma_c, \boldsymbol{\beta}, \boldsymbol{T}) = \ell_i = \prod_{j=1}^{n_i} \prod_{c=1}^{C} (p_{ijc})^{d_{ijc}}$$

where $d_{ijc} = 1$ if $\mathsf{y}_{ij} = c$ and 0 otherwise

Conditional Independence: responses from level-2 unit i are independent given $\boldsymbol{\theta}$

Maximum (Marginal) Likelihood Estimation

Marginal Probability

$$h(\mathsf{y}_i) = h_i = \int_{\boldsymbol{\theta}} \ell_i \, g(\boldsymbol{\theta}) \, d\boldsymbol{\theta}$$

$g(\boldsymbol{\theta})$ = multivariate standard normal density

Maximize the marginal log-likelihood from N level-2 units

$$\log L = \sum_i^N \log h_i \quad \text{with respect to} \quad \boldsymbol{\eta} = [\, \gamma_c \;\vdots\; \boldsymbol{\beta} \;\vdots\; \boldsymbol{T} \,]$$

Full-likelihood approach found in SAS PROC NLMIXED, STATA, MIXOR

Numerical Quadrature for integration over θ

- method to numerically perform an integration

$$\int_\theta f(\theta)g(\theta)d\theta \approx \sum_{q=1}^{Q} f(B_q)A(B_q)$$

where B_q $(q = 1, \ldots, Q)$ are the quadrature nodes or points
$A(B_q)$ $(q = 1, \ldots, Q)$ are the weights (sum $= 1$)

- the more points you use, the more accurate the approximation, but the more time it takes

- For standard normal distribution, Gauss-Hermite quadrature

- does yield a likelihood value that can be used for LR tests

Quadrature: integration over $\boldsymbol{\theta}$

- Standard normal distribution - tabled values of Q nodes $\boldsymbol{B}_q$ and weights $A(\boldsymbol{B}_q)$

logit $$\lambda_{ijcq} = \gamma_c - (\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{T}\boldsymbol{B}_q)$$

probabilities $$p_{ijcq} = \frac{1}{1+\exp(-\lambda_{ijcq})} - \frac{1}{1+\exp(-\lambda_{ij,\,c-1,\,q})}$$

likelihoods $$\ell(\mathsf{y}_i \mid \boldsymbol{B}_q\,;\,\gamma_c,\boldsymbol{\beta},\boldsymbol{T}) \;=\; \ell_{iq} \;=\; \prod_{j=1}^{n_i}\prod_{c=1}^{C}(p_{ijcq})^{d_{ijc}}$$

$$h(\mathsf{y}_i) \approx \sum_{q=1}^{Q} \ell_{iq}\, A(\boldsymbol{B}_q)$$

Empirical Bayes estimates *(univariate case)*

$$\hat{\theta}_i = E(\theta_i \mid \mathsf{y}_i) \;=\; \frac{1}{h_i} \int_\theta \theta_i \, \ell_i \, g(\theta) \, d\theta \;\approx\; \frac{1}{h_i} \sum_{q=1}^{Q} B_q \, \ell_{iq} \, A(B_q)$$

The variance of this estimator is obtained as:

$$V(\hat{\theta}_i \mid \mathsf{y}_i) \;=\; \frac{1}{h_i} \int_\theta (\theta_i - \hat{\theta}_i)^2 \, \ell_i \, g(\theta) \, d\theta$$

$$\approx \frac{1}{h_i} \sum_{q=1}^{Q} (B_q - \hat{\theta}_i)^2 \, \ell_{iq} \, A(B_q)$$

At convergence, one more round of quadrature and the values of

- $h_i = h(\mathsf{y}_i)$ which vary by i units
- $\ell_{iq} = \ell(\mathsf{y}_i \mid B_q \,;\, \gamma_c, \boldsymbol{\beta}, \sigma_\upsilon)$ which vary by i units and quad pts

Subject-specific probabilities

estimated logits given $\boldsymbol{\theta}_i$

$$\hat{\lambda}_{ijc}(\boldsymbol{\theta}_i) = \hat{\gamma}_c - \left(\boldsymbol{x}'_{ij}\hat{\boldsymbol{\beta}} + \boldsymbol{z}'_{ij}\hat{\boldsymbol{T}}\,\boldsymbol{\theta}_i\right)$$

estimated probabilities

$$\hat{p}_{ijc}(\boldsymbol{\theta}_i) = \frac{1}{1+\exp(-\hat{\lambda}_{ijc}(\boldsymbol{\theta}_i))} - \frac{1}{1+\exp(-\hat{\lambda}_{ijc-1}(\boldsymbol{\theta}_i))}$$

Marginal probabilities

$$\hat{p}_{ijc} = \int_{\boldsymbol{\theta}} \hat{p}_{ijc}(\boldsymbol{\theta}_i)\; g(\boldsymbol{\theta})\; d\boldsymbol{\theta}$$

$$\approx \sum_{q=1}^{Q} \hat{p}_{ijc}(\boldsymbol{B}_q)\; A(\boldsymbol{B}_q)$$

Adaptive Quadrature

- adapt quadrature points and weights for each subject, and at each iteration, using EB estimates of their location $\hat{\theta}_i$ and uncertainty $s_i^2 = V(\hat{\theta}_i \mid \mathsf{y}_i)$
- requires fewer points to obtain accurate solution
- especially useful if subject random effects are very spread out (*i.e.*, ICC is high)

Adapted quadrature points and weights, from original points B_q and weights A_q, where $\phi(\cdot)$ = normal pdf

$$B_{iq} = \hat{\theta}_i + s_i B_q$$

$$A_{iq} \;=\; \sqrt{2\pi}\, s_i \exp(B_q^2/2)\, \phi(B_{iq})\, A_q$$

Multiple Random Effects

- quadrature solution must integrate over each random effect dimension (r = number of random effects)

$$\boldsymbol{B}_q = (B_{q1}, B_{q2}, \ldots, B_{qr}) = r - \text{dimension quad pt vector}$$

$$A(\boldsymbol{B}_q) = \prod_{h=1}^{r} A(B_{qh}) = \text{product of univariate weights}$$

- curse of dimensionality: Q^r total points, where Q is the number of points per dimension (*e.g.*, $Q = 10$ and $r = 3$ leads to evaluation at 1000 points)

- adaptive quadrature especially useful here, since Q can be lower

Other methods for integration of θ

Methods based on first- or second-order Taylor series expansions

- Marginal quasi-likelihood (MQL) involves expansion around the fixed part of the model
- Penalized or predictive quasi-likelihood (PQL) also includes the random part in its expansion
- Both are available in the SAS PROC GLIMMIX and MLwiN
- fast, but doesn't yield a likelihood for LR tests
- can yield downwardly biased estimates in certain situations (if N and/or n is small, or ICC is high), especially for MQL

Laplace approximation - Raudenbush et. al., (2000)

- a combination of a fully multivariate Taylor series expansion and Laplace approximation
- fast and computationally accurate
- yields a likelihood for LR tests
- available in HLM, though not for all models

Other methods

- Markov Chain Monte Carlo (MCMC) Bayesian approach (in BUGS)
- Maximum Simulated Likelihood (in some STATA programs) in econometric, transportation, political science literatures

Treatment-Related Change Across Time

Data from the NIMH Schizophrenia collaborative study on treatment related changes in overall severity. IMPS item 79, *Severity of Illness*, was scored as:

1 = normal or borderline mentally ill
2 = mildly or moderately ill
3 = markedly ill
4 = severely or among the most extremely ill

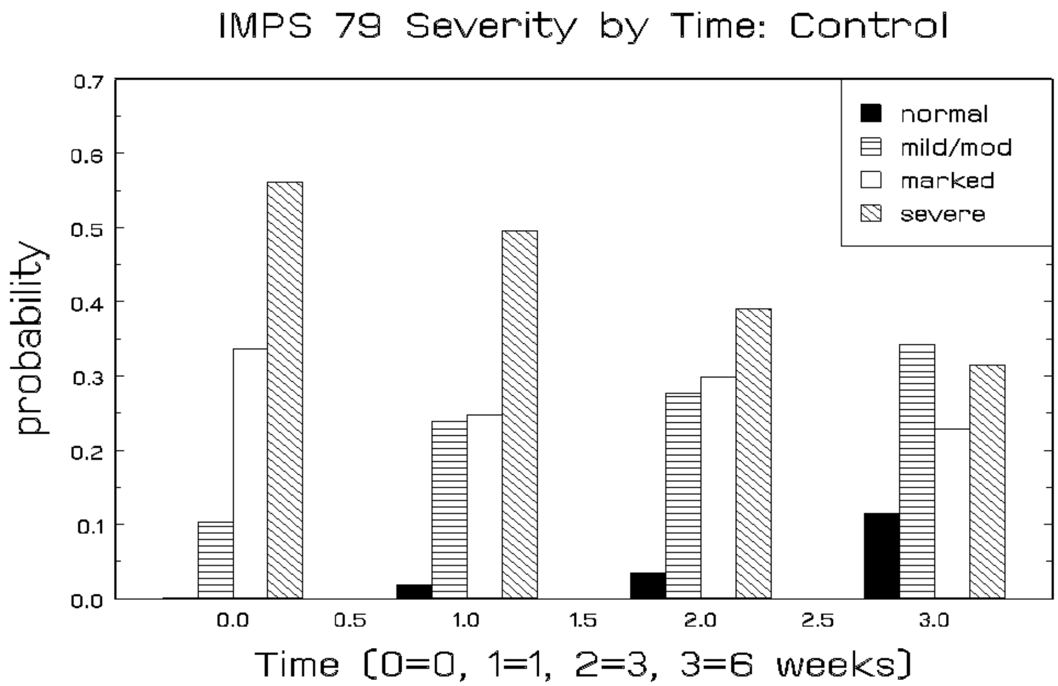
The experimental design and corresponding sample sizes:

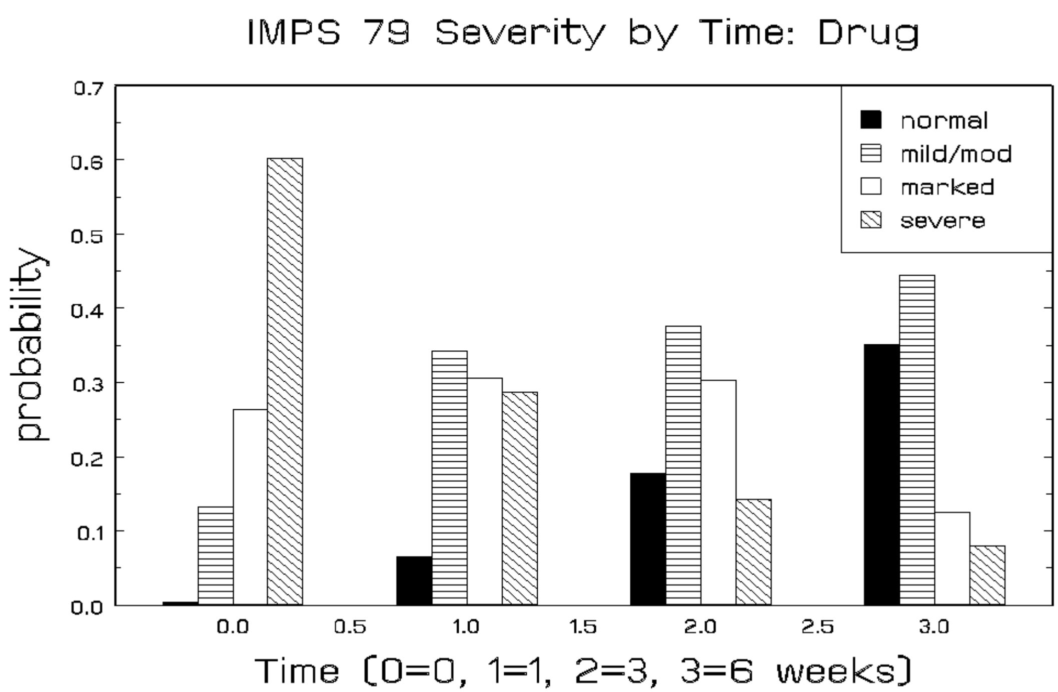
	Sample size at Week							
Group	0	1	2	3	4	5	6	*completers*
PLC (n=108)	107	105	5	87	2	2	70	65%
DRUG (n=329)	327	321	9	287	9	7	265	81%

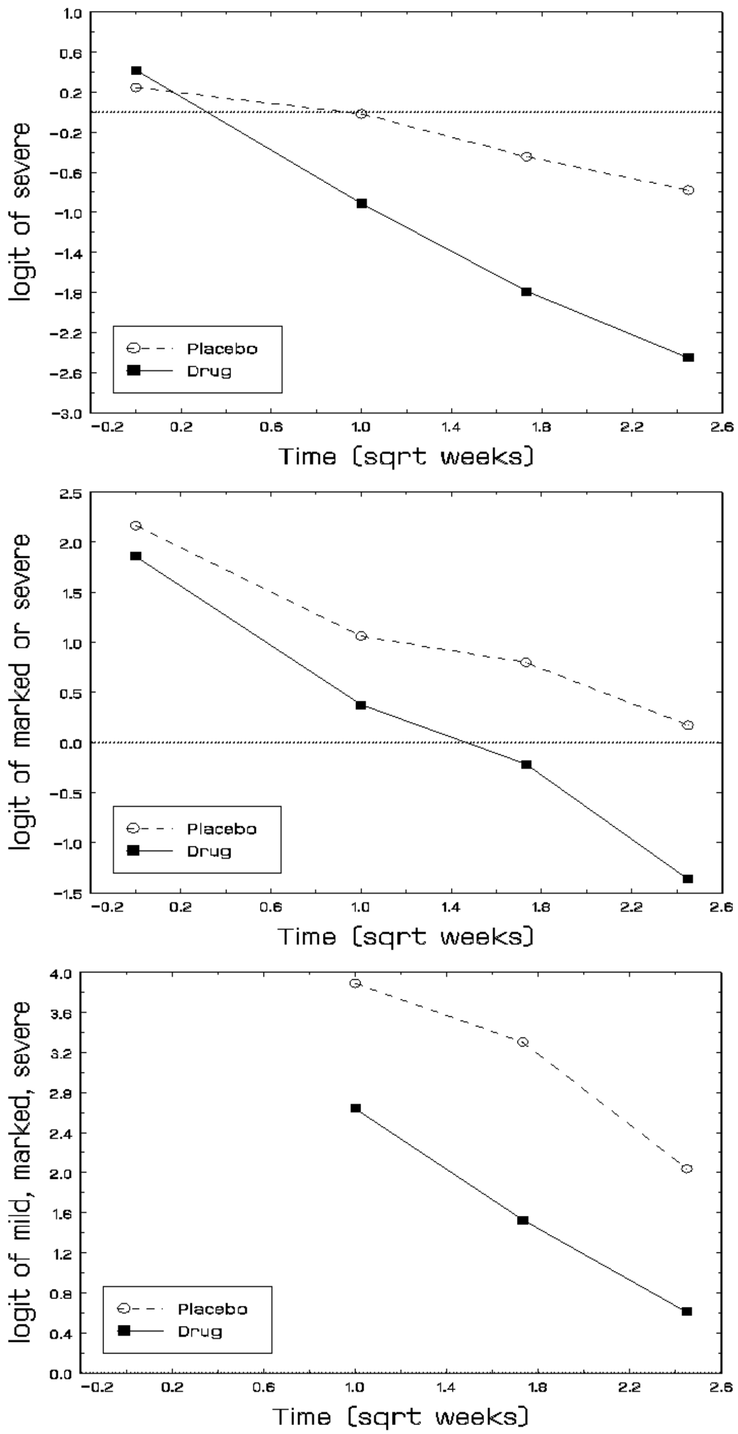
Drug = Chlorpromazine, Fluphenazine, or Thioridazine

Main question of interest:

- Was there differential improvement for the drug groups relative to the control group?


IMPS 79 Severity by Time: Control
probability
Time (0=0, 1=1, 2=3, 3=6 weeks)
normal
mild/mod
marked
severe
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.0
0.5
1.0
1.5
2.0
2.5
3.0


IMPS 79 Severity by Time: Drug
probability
Time (0=0, 1=1, 2=3, 3=6 weeks)
normal
mild/mod
marked
severe
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.0
0.5
1.0
1.5
2.0
2.5
3.0


logit of severe
Time (sqrt weeks)
Placebo
Drug
logit of marked or severe
Time (sqrt weeks)
Placebo
Drug
logit of mild, marked, severe
Time (sqrt weeks)
Placebo
Drug

Within-Subjects / Between-Subjects components

Within-subjects model - *level 1* $\quad (j = 1, \ldots, n_i \text{ obs})$

$$\lambda_{ijc} = \gamma_c - [b_{0i} \; + \; b_{1i}\sqrt{Week_j}]$$

Between-subjects model - *level 2* $\quad (i = 1, \ldots, N \text{ subjects})$

$$b_{0i} \; = \; \beta_0 \; + \; \beta_2 Grp_i \; + \; \upsilon_{0i}$$

$$b_{1i} \; = \; \beta_1 \; + \; \beta_3 Grp_i$$

$$\upsilon_{0i} \sim \mathcal{NID}(0, \sigma_\upsilon^2)$$

NIMH Schizophrenia Study: Severity of Illness (N = 437)
Ordinal Logistic ML Estimates (se) - *random intercept model*

	ML estimates	se	z	$p <$
intercept	5.844	0.331	17.68	.001
threshold_2	3.028	0.137	22.07	.001
threshold_3	5.142	0.182	28.28	.001
Drug (0 = plc; 1 = drug)	-0.055	0.313	-0.18	.86
Time (sqrt week)	-0.766	0.131	-5.86	.001
Drug by Time	-1.202	0.153	-7.88	.001
Intercept sd	1.928	0.118		

Intra-person correlation $= 1.928^2/(1.928^2 + \pi^2/3) = .53$

$-2 \log L = 3404.2$

Within-Subjects / Between-Subjects components

Within-subjects model - *level 1* $(j = 1, \ldots, n_i \text{ obs})$

$$logit_{cij} = \gamma_c - [b_{0i} \ + \ b_{1i}\sqrt{Week}_j]$$

Between-subjects model - *level 2* $(i = 1, \ldots, N \text{ subjects})$

$$b_{0i} \ = \ \beta_0 \ + \ \beta_2 Grp_i \ + \ \upsilon_{0i}$$

$$b_{1i} \ = \ \beta_1 \ + \ \beta_3 Grp_i \ + \ \upsilon_{1i}$$

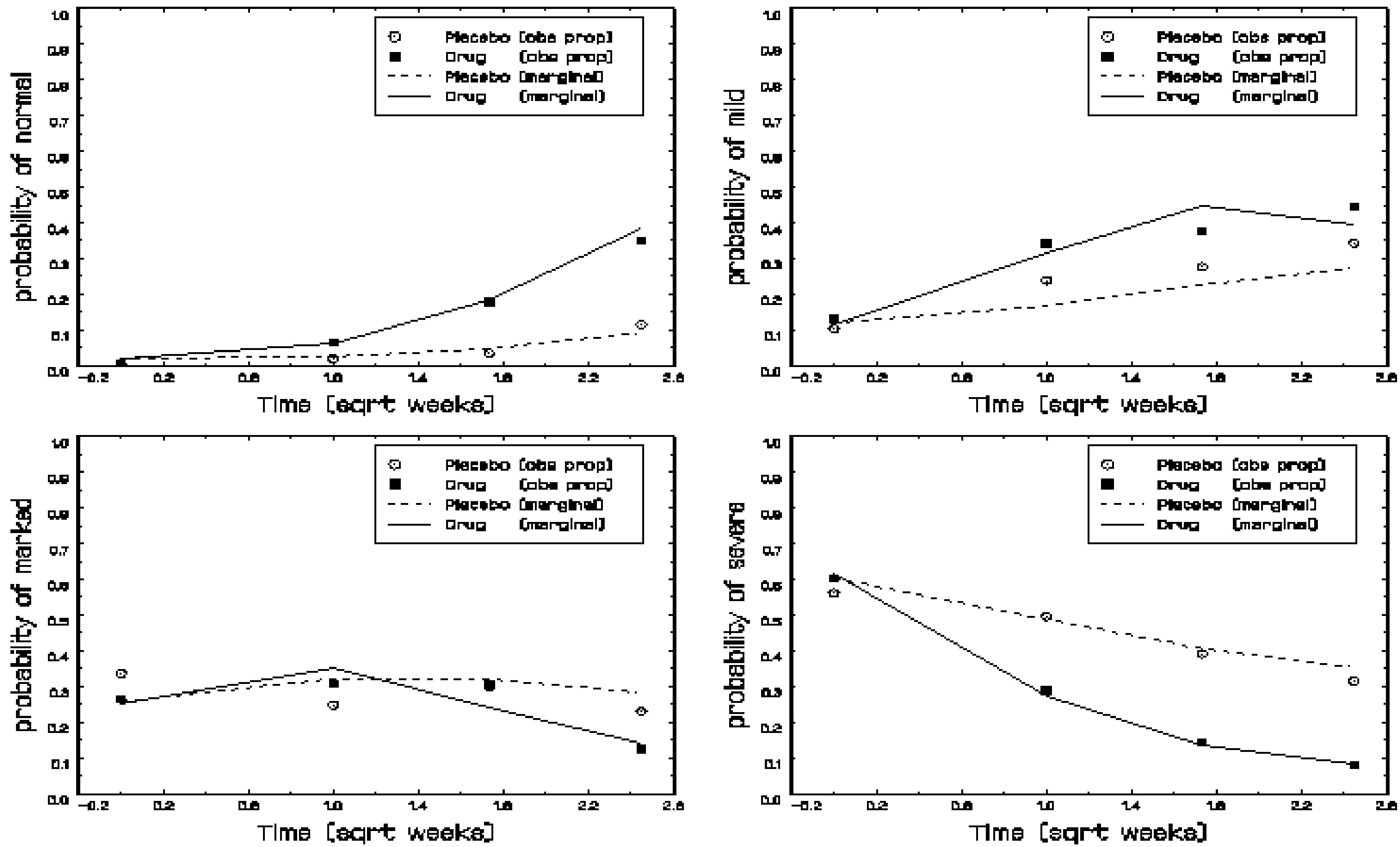
$$\boldsymbol{\upsilon}_i \sim \mathcal{NID}(\mathbf{0}, \boldsymbol{\Sigma})$$

Ordinal Logistic Random Intercept and Trend Model

	ML estimates	se	z	$p <$
intercept	7.283	0.467	15.59	.001
threshold_2	3.884	0.209	18.57	.001
threshold_3	6.478	0.290	22.36	.001
Drug (0=plc; 1 = drug)	0.056	0.388	0.15	.88
Time (sqrt week)	-0.879	0.216	-4.07	.001
Drug by Time	-1.684	0.250	-6.74	.001
Intercept var	6.847	1.282		
Int-Time covar	-1.447	0.515	$(r_{\upsilon_0 \upsilon_1} = -.40)$	
Time var	1.949	0.404		

$-2 \log L = 3326.5, \quad \chi^2_2 = 77.7, p < .001$

Model Fit of Observed Proportions



SAS NLMIXED code: random-intercepts ordinal logistic regression

```
DATA one; INFILE 'c:\schizx1.dat' ;
INPUT id imps79 imps79b imps79o int tx week sweek txswk ;

/* get rid of observations with missing values */
IF imps79 > -9;

PROC FORMAT;   VALUE tx 0 = 'placebo' 1 = 'drug';


PROC NLMIXED ;
PARMS b0=0 b1=0 b2=0 b3=0 sd=1 g2=1 g3=2;

z = b0 + b1*tx + b2*sweek + b3*txswk + u;

IF (imps79o=1) THEN p = 1 / (1 + EXP(-(0-z)));
ELSE IF (imps79o=2) THEN p = (1/(1 + EXP(-(g2-z)))) - (1/(1 + EXP(-(0-z))));
ELSE IF (imps79o=3) THEN p = (1/(1 + EXP(-(g3-z)))) - (1/(1 + EXP(-(g2-z))));
ELSE IF (imps79o=4) THEN p = 1 - (1 / (1 + EXP(-(g3-z))));

loglike = LOG(p);
MODEL imps79o ~ GENERAL(loglike);
RANDOM u ~ NORMAL(0,sd*sd) SUBJECT=id;
ESTIMATE 'icc' sd*sd/(3.289868134+sd*sd);
RUN;
```

SAS NLMIXED code: random-trend ordinal logistic regression
using data from the earlier NLMIXED example

```
PROC NLMIXED ;
PARMS b0=0 b1=0 b2=0 b3=0 v0=1 c01=0 v1=.5 g2=1 g3=2;

z = b0 + b1*tx + b2*sweek + b3*txswk + u0 + u1*sweek;

IF (imps79o=1) THEN p = 1 / (1 + EXP(-(0-z)));
ELSE IF (imps79o=2) THEN p = (1/(1 + EXP(-(g2-z)))) - (1/(1 + EXP(-(0-z))));
ELSE IF (imps79o=3) THEN p = (1/(1 + EXP(-(g3-z)))) - (1/(1 + EXP(-(g2-z))));
ELSE IF (imps79o=4) THEN p = 1 - (1 / (1 + EXP(-(g3-z))));

loglike = LOG(p);
MODEL imps79o ~ GENERAL(loglike);
RANDOM u0 u1 ~ NORMAL([0,0], [v0,c01,v1]) SUBJECT=id;

ESTIMATE 're corr' c01/SQRT(v0*v1);
RUN;
```

Model fit of observed marginal proportions

1. $\hat{\boldsymbol{y}}_i \ = \ \boldsymbol{X}_i \hat{\boldsymbol{\beta}}$

2. calculate "marginalization" vector

$$\hat{\boldsymbol{s}} = \frac{1}{\sigma} \left[\text{Diag}(\hat{V}(\boldsymbol{y}_i)) \right]^{1/2}$$

- $\hat{V}(\boldsymbol{y}_i) = \boldsymbol{Z}_i \hat{\boldsymbol{\Sigma}} \boldsymbol{Z}_i' + \sigma^2 \boldsymbol{I}_i$
- $\sigma = 1$ for probit and $\sigma = \pi/\sqrt{3}$ for logistic
- $\boldsymbol{Z}_i$ = design matrix for random effects
- for random-intercepts model $\boldsymbol{Z}_i = \boldsymbol{1}_i$, and so, $\hat{s} = \sqrt{\hat{\sigma}_{\upsilon}^2/\sigma^2 + 1}$

3. perform element-wise division

$$\hat{\mathbf{z}}_i = \hat{\boldsymbol{y}}_i \;/.\; \hat{\boldsymbol{s}}$$

4. $\hat{\boldsymbol{p}}_i = \Phi(\hat{\mathbf{z}}_i)$ for probit and $\hat{\boldsymbol{p}}_i = \Psi(\hat{\mathbf{z}}_i)$ for logistic

5. In practice, for logistic, $(15\pi)/(16\sqrt{3})$ works better than $\pi/\sqrt{3}$ as σ (Zeger *et al.*, 1988, Biometrics)

6. Logistic is approximate; relies on cumulative Gaussian approximation to the logistic function

SAS IML code: computing marginal probabilities - ordinal models

```
TITLE1 'NIMH Schizophrenia Data - Estimated Marginal Probabilities';
PROC IML;
/* Results from NLMIXED analysis:  random intercept model */;

 x0  = { 1 0 0.00000 0,
         1 0 1.00000 0,
         1 0 1.73205 0,
         1 0 2.44949 0};
 x1  = { 1 1 0.00000 0.00000,
         1 1 1.00000 1.00000,
         1 1 1.73205 1.73205,
         1 1 2.44949 2.44949};

 sdu    = {1.928};
 beta   = {5.844, -.055, -.766, -1.202};
 thresh = {3.028, 5.142};
```

```
/* Approximate Marginalization Method */;

 pi   = 3.141592654;
 nt   = 4;
 ivec = J(nt,1,1);
 zvec = J(nt,1,1);
 evec = (15/16)**2 * (pi**2)/3 * ivec;

/* nt by nt matrix with evec on the diagonal and zeros elsewhere */;
emat = diag(evec);

/* variance-covariance matrix of underlying latent variable */;
vary = zvec * sdu*sdu * T(zvec) + emat;

sdy = sqrt(vecdiag(vary) / vecdiag(emat));
```

```
za0 = (0 - x0*beta) / sdy ;
zb0 = (thresh[1] - x0*beta) / sdy;
zc0 = (thresh[2] - x0*beta) / sdy;
za1 = (0 - x1*beta) / sdy;
zb1 = (thresh[1] - x1*beta) / sdy;
zc1 = (thresh[2] - x1*beta) / sdy;

grp0a = 1 / ( 1 + EXP(0 - za0));
grp0b = 1 / ( 1 + EXP(0 - zb0));
grp0c = 1 / ( 1 + EXP(0 - zc0));
grp1a = 1 / ( 1 + EXP(0 - za1));
grp1b = 1 / ( 1 + EXP(0 - zb1));
grp1c = 1 / ( 1 + EXP(0 - zc1));

print 'Random intercept model';
print 'Approximate Marginalization Method';
print 'marginal prob for group 0 - catg 1' grp0a [FORMAT=8.4];
print 'marginal prob for group 0 - catg 2' (grp0b-grp0a) [FORMAT=8.4];
print 'marginal prob for group 0 - catg 3' (grp0c-grp0b) [FORMAT=8.4];
print 'marginal prob for group 0 - catg 4' (1-grp0c) [FORMAT=8.4];
print 'marginal prob for group 1 - catg 1' grp1a [FORMAT=8.4];
print 'marginal prob for group 1 - catg 2' (grp1b-grp1a) [FORMAT=8.4];
print 'marginal prob for group 1 - catg 3' (grp1c-grp1b) [FORMAT=8.4];
print 'marginal prob for group 1 - catg 4' (1-grp1c) [FORMAT=8.4];
```

```
/* Random Intercept and Trend Model */;

 varu   = {6.847 -1.447,
            -1.447 1.949};
 beta   = { 7.283, .056, -.879, -1.684};
 thresh = {3.884, 6.478};

/* Approximate Marginalization Method */;

 pi   = 3.141592654;
 nt   = 4;
 ivec = J(nt,1,1);
 zmat = {1 0.00000,
         1 1.00000,
         1 1.73205,
         1 2.44949};
 evec = (15/16)**2 * (pi**2)/3 * ivec;

/* nt by nt matrix with evec on the diagonal and zeros elsewhere */;
emat = diag(evec);

/* variance-covariance matrix of underlying latent variable */;
vary = zmat * varu * T(zmat) + emat;
sdy = sqrt(vecdiag(vary) / vecdiag(emat));
```

```
za0 = (0 - x0*beta) / sdy ;
zb0 = (thresh[1] - x0*beta) / sdy;
zc0 = (thresh[2] - x0*beta) / sdy;
za1 = (0 - x1*beta) / sdy;
zb1 = (thresh[1] - x1*beta) / sdy;
zc1 = (thresh[2] - x1*beta) / sdy;

grp0a = 1 / ( 1 + EXP(0 - za0));
grp0b = 1 / ( 1 + EXP(0 - zb0));
grp0c = 1 / ( 1 + EXP(0 - zc0));
grp1a = 1 / ( 1 + EXP(0 - za1));
grp1b = 1 / ( 1 + EXP(0 - zb1));
grp1c = 1 / ( 1 + EXP(0 - zc1));

print 'Random intercept and trend model';
print 'Approximate Marginalization Method';
print 'marginal prob for group 0 - catg 1' grp0a [FORMAT=8.4];
print 'marginal prob for group 0 - catg 2' (grp0b-grp0a) [FORMAT=8.4];
print 'marginal prob for group 0 - catg 3' (grp0c-grp0b) [FORMAT=8.4];
print 'marginal prob for group 0 - catg 4' (1-grp0c) [FORMAT=8.4];
print 'marginal prob for group 1 - catg 1' grp1a [FORMAT=8.4];
print 'marginal prob for group 1 - catg 2' (grp1b-grp1a) [FORMAT=8.4];
print 'marginal prob for group 1 - catg 3' (grp1c-grp1b) [FORMAT=8.4];
print 'marginal prob for group 1 - catg 4' (1-grp1c) [FORMAT=8.4];
```

Proportional and Non-proportional Odds

Proportional Odds model

$$\log\left[\frac{P(\mathsf{y}_{ij} \le c)}{1 - P(\mathsf{y}_{ij} \le c)}\right] = \gamma_c - \left[\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{\upsilon}_i\right]$$

$$\text{with } \boldsymbol{\upsilon}_i \sim N(\mathbf{0}, \boldsymbol{T}\boldsymbol{T}' = \boldsymbol{\Sigma})$$

$$= \gamma_c - \left[\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{T}\boldsymbol{\theta}_i\right] \quad \text{with } \boldsymbol{\theta}_i \sim N(\mathbf{0}, \boldsymbol{I})$$

- relationship between the explanatory variables and the cumulative logits does not depend on c
- effects of $\boldsymbol{x}$ variables DO NOT vary across the $C - 1$ cumulative logits

Hedeker & Mermelstein (1998, *Mult Behav Res*) extension:

$$\log\left[\frac{P(\mathsf{y}_{ij} \leq c)}{1 - P(\mathsf{y}_{ij} \leq c)}\right] = \gamma_{c(0)} - \left[\boldsymbol{u}'_{ij}\boldsymbol{\gamma}_c + \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{T}\boldsymbol{\theta}_i\right]$$

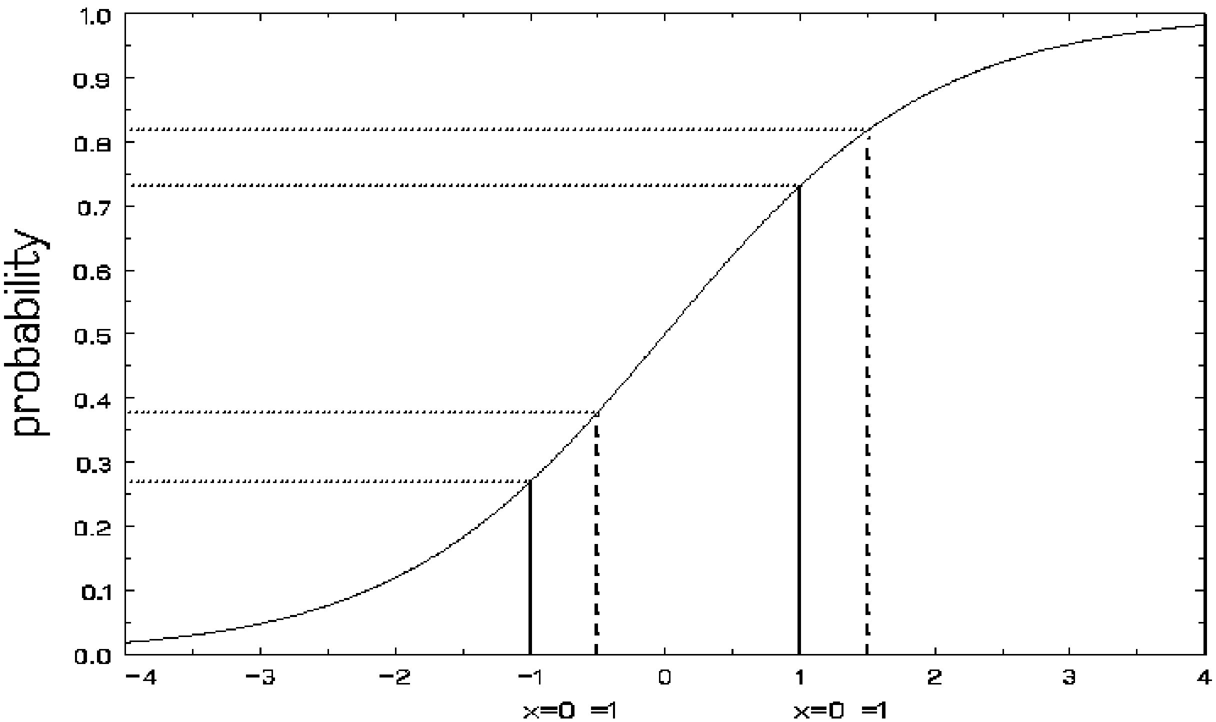
$\boldsymbol{u}_{ij} = h \times 1$ vector for the set of h covariates for which proportional odds is not assumed

- effects of $\boldsymbol{u}$ variables DO vary across the $C - 1$ cumulative logits
- more flexible model for ordinal response relations

Proportional Odds Assumption: covariate effects are the same across all cumulative logits

	Response			
group	Absent	Mild	Severe	total
control	27	46	27	100
cumulative odds	$\frac{27}{73} = .37$	$\frac{73}{27} = 2.7$		
logit	*-1*	*1*		
treatment	38	44	18	100
cumulative odds	$\frac{38}{62} = .61$	$\frac{82}{18} = 4.6$		
logit	*-.5*	*1.5*		

$\Rightarrow$ *group difference = .5 for both cumulative logits*



$$\log\left[\frac{P(\mathsf{y}_{ij}=1)}{P(\mathsf{y}_{ij}=2 \text{ or } 3)}\right] = 0 - [\beta_0 + x\,\beta_1]$$

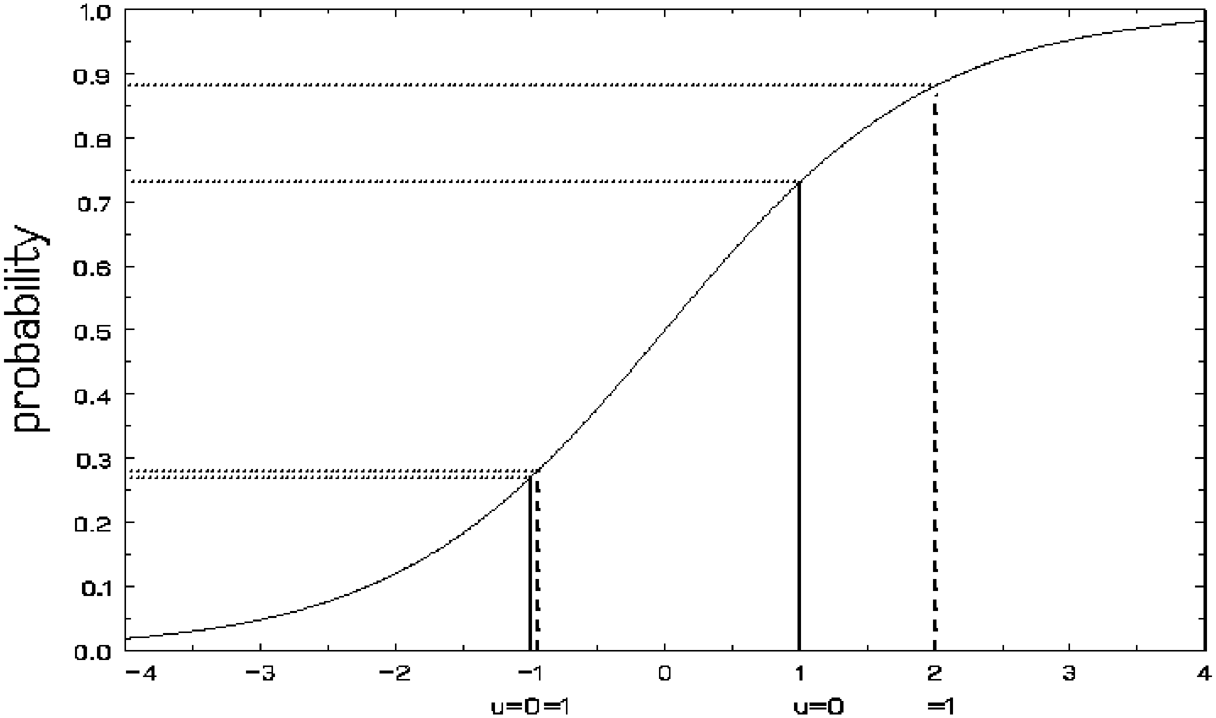
$$\log\left[\frac{P(\mathsf{y}_{ij}=1 \text{ or } 2)}{P(\mathsf{y}_{ij}=3)}\right] = \gamma_2 - [\beta_0 + x\,\beta_1]$$

$\beta_0 = 1\,, \quad \gamma_2 = 2\,, \quad \beta_1 = -.5$
(covariate effect is same for both cumulative logits)

Non-Proportional Odds: covariate effects vary across the cumulative logits

group	*Response* Absent	Mild	Severe	total
control	27	46	27	100
cumulative odds	$\frac{27}{73} = .37$	$\frac{73}{27} = 2.7$		
logit	*-1*	*1*		
treatment	28	60	12	100
cumulative odds	$\frac{28}{72} = .39$	$\frac{88}{12} = 7.3$		
logit	*-.95*	*2*		

$\Rightarrow$ *UNEQUAL group difference across cumulative logits*



$$\log\left[\frac{P(\mathsf{y}_{ij}=1)}{P(\mathsf{y}_{ij}=2 \text{ or } 3)}\right] \;=\; 0-[\beta_0+u\,\gamma_1]$$

$$\log\left[\frac{P(\mathsf{y}_{ij}=1 \text{ or } 2)}{P(\mathsf{y}_{ij}=3)}\right] \;=\; \gamma_{2\,(0)}-[\beta_0+u\,\gamma_2]$$

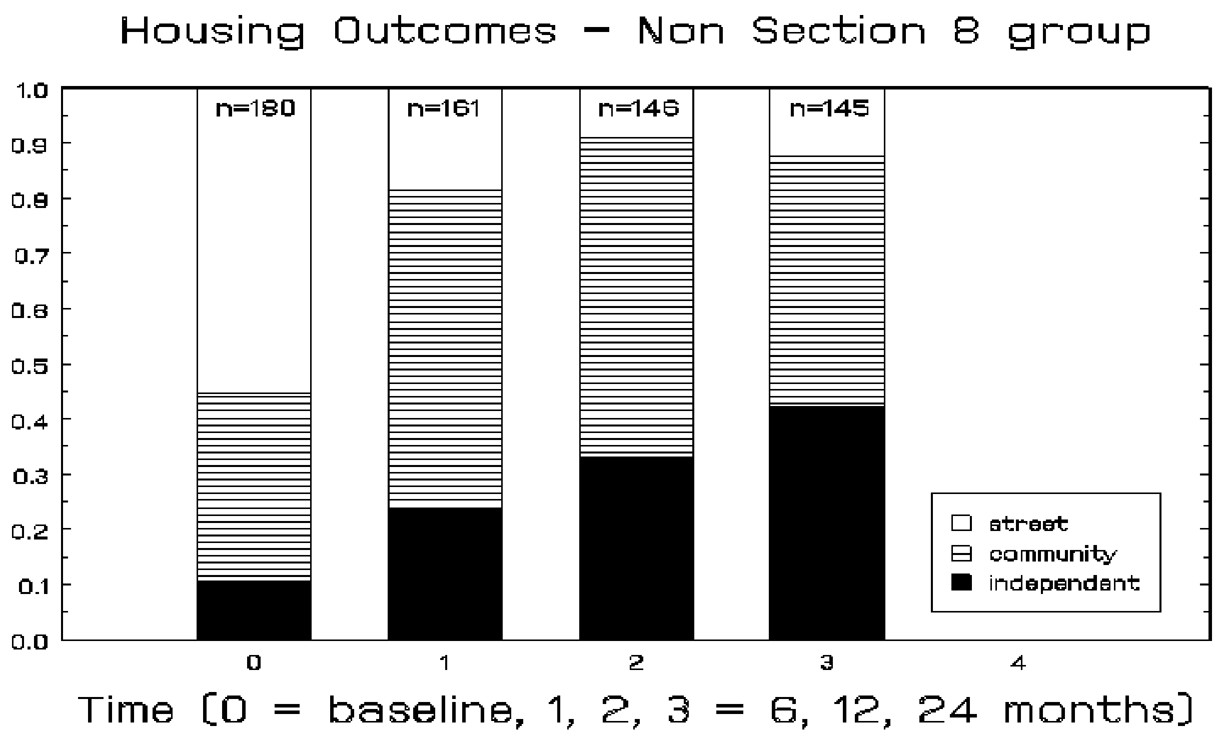
$\beta_0 = 1\,, \quad \gamma_{2(0)} = 2\,, \quad \gamma_1 = -.05\,, \quad \gamma_2 = -.1$

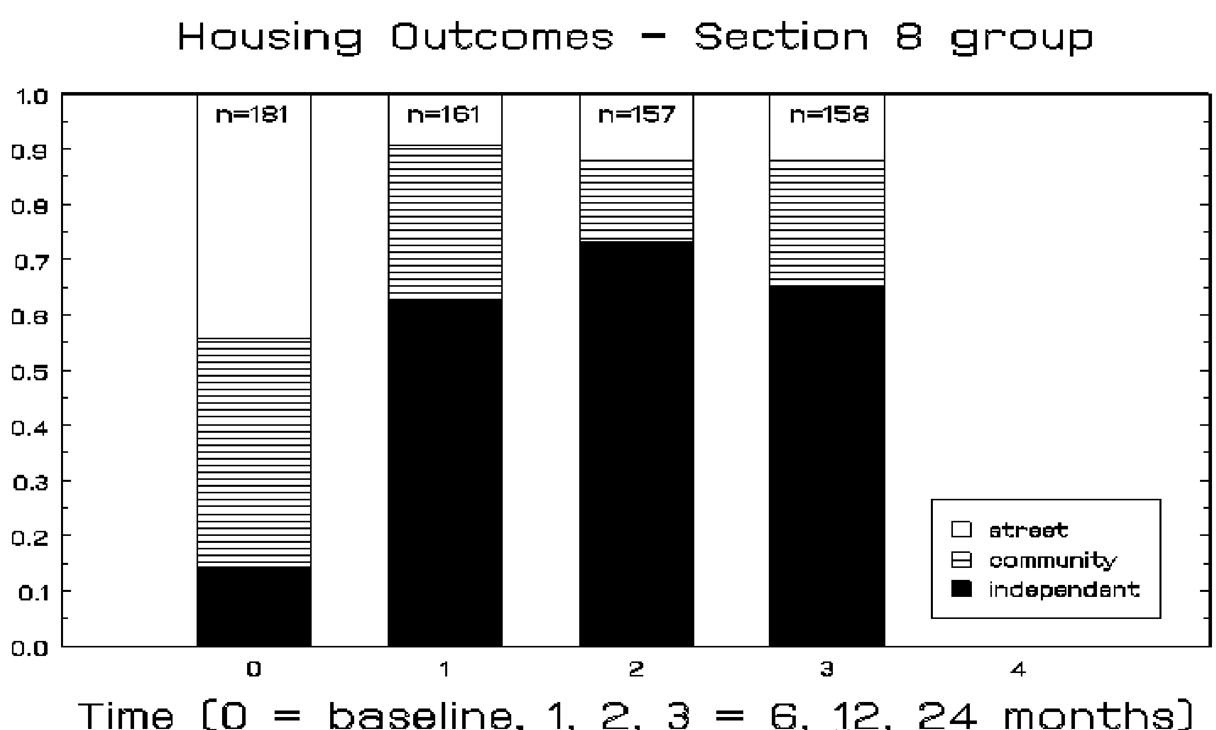
(covariate effect varies across the cumulative logits)

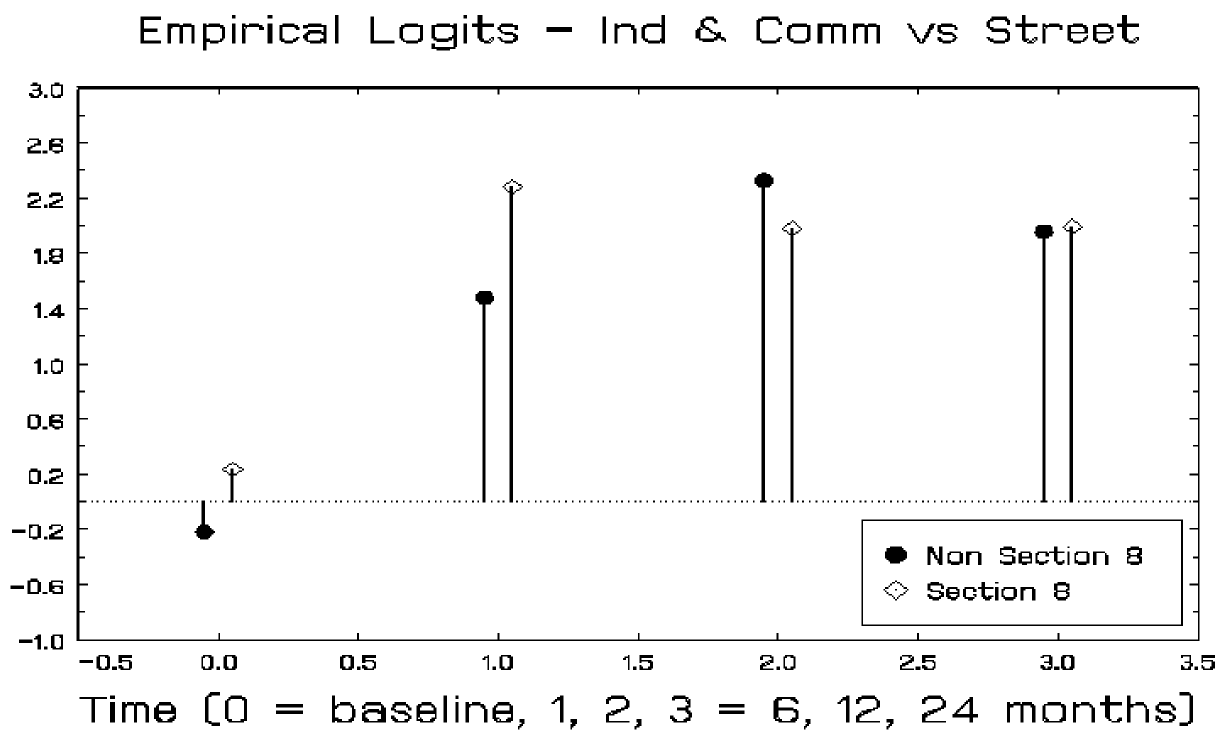
San Diego Homeless Research Project (Hough)

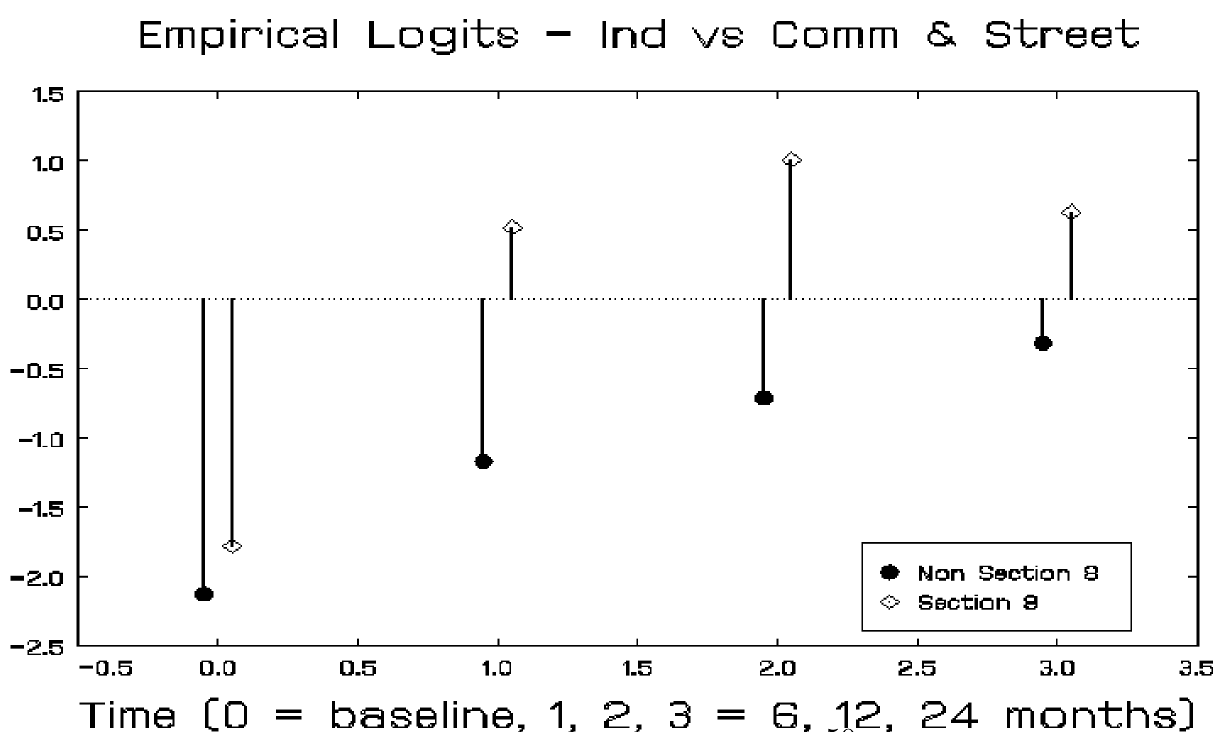
- 361 mentally ill subjects who were
  - homeless or
  - at very high risk of becoming homeless
- 2 conditions: HUD Section 8 rental certificates (yes/no)
- baseline and 6, 12, and 24 month follow-ups
- Categorical outcome: housing status
  - streets / shelters ($\mathsf{y} = 1$)
  - community / institutions ($\mathsf{y} = 2$)
  - independent ($\mathsf{y} = 3$)

Question: Do Section 8 certificates influence housing status across time?


Housing Outcomes - Non Section 8 group
n=180
n=161
n=146
n=145
street
community
independent
Time (0 = baseline, 1, 2, 3 = 6, 12, 24 months)


Housing Outcomes - Section 8 group
n=181
n=161
n=157
n=158
street
community
independent
Time (0 = baseline, 1, 2, 3 = 6, 12, 24 months)


Empirical Logits - Ind & Comm vs Street
3.0
2.6
2.2
1.8
1.4
1.0
0.6
0.2
-0.2
-0.6
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Non Section 8
Section 8
Time (0 = baseline, 1, 2, 3 = 6, 12, 24 months)


Empirical Logits - Ind vs Comm & Street
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
-2.5
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Non Section 8
Section 8
Time (0 = baseline, 1, 2, 3 = 6, 12, 24 months)

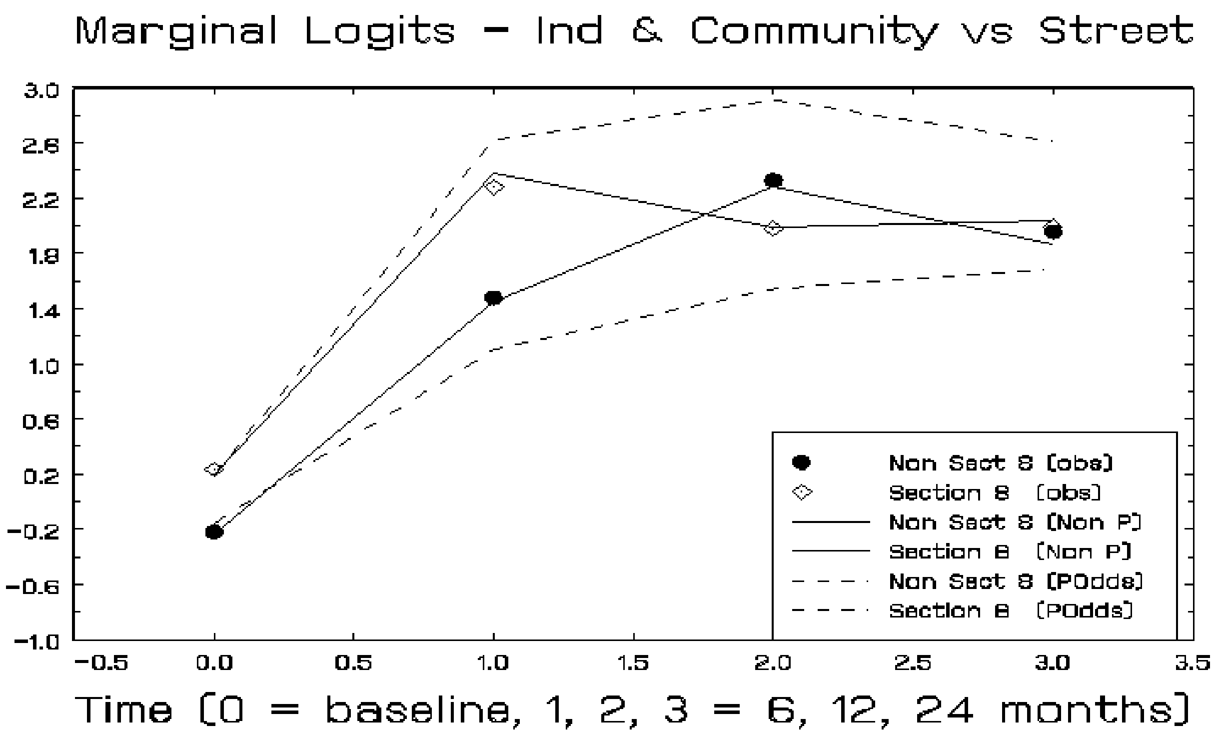
Housing status across time: 1289 observations within 361 subjects
Ordinal Mixed Regression Model estimates and standard errors (se)

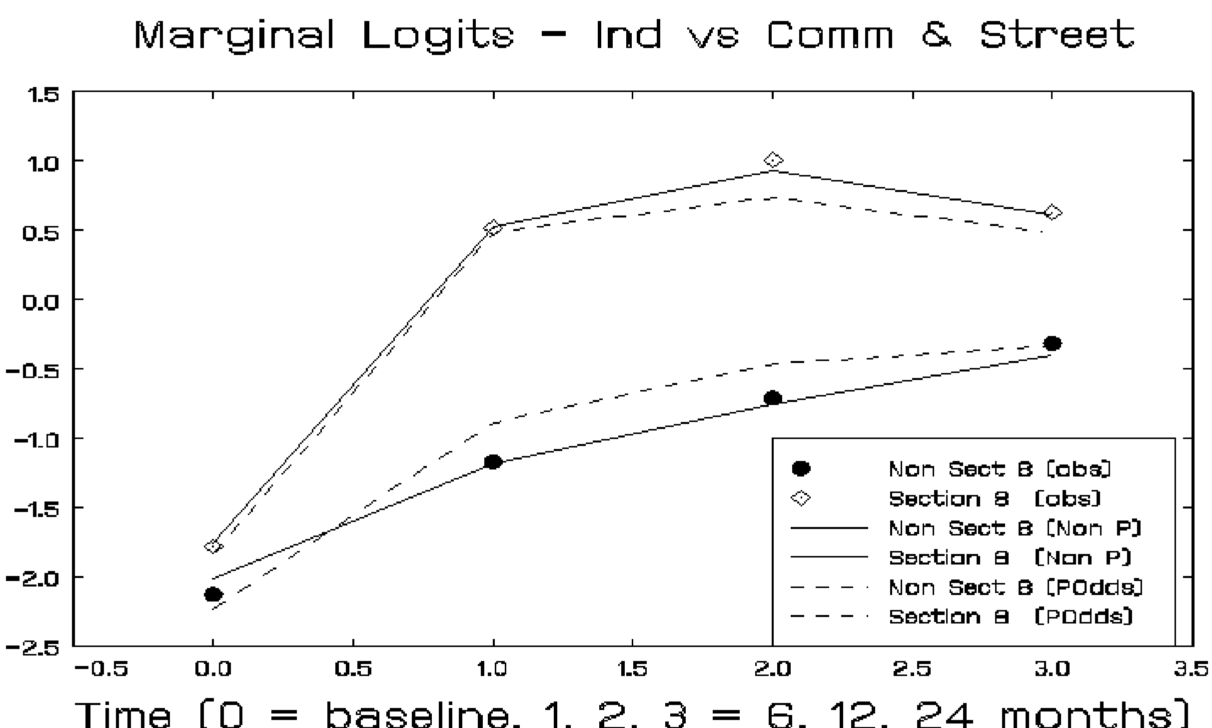
	Proportional Odds Model		Non-Proportional Odds			
			Non-street[1]		Independent[2]	
term	estimate	se	estimate	se	estimate	se
intercept	- .219	.198	- .318	.207		
threshold	**2.738**	.130			**2.375**	.280
t1 (6 month)	**1.731**	.234	**2.288**	.302	**1.077**	.343
t2 (12 month)	**2.310**	.246	**3.333**	.385	**1.643**	.339
t3 (24 month)	**2.494**	.253	**2.814**	.347	**2.143**	.337
section 8 (y=1)	*.494*	.275	**.584**	.293	.323	.393
section 8 by t1	**1.403**	.341	.569	.466	**2.014**	.470
section 8 by t2	**1.170**	.353	-.948	.505	**2.009**	.475
section 8 by t3	*.634*	.348	-.365	.479	**1.064**	.463
subject sd	**1.446**	.120	**1.445**	.120	($ICC \approx .4$)	
$-2 \log L$	2275.2		2223.1		($\chi^2_7 = 52.1$)	

bold indicates $p < .05$ *italic* indicates $.05 < p < .10$

[1] = independent + community vs street

[2] = independent vs community + street


Marginal Logits - Ind & Community vs Street
3.0
2.6
2.2
1.8
1.4
1.0
0.6
0.2
-0.2
-0.6
-1.0
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Non Sect 8 (obs)
Section 8 (obs)
Non Sect 8 (Non P)
Section 8 (Non P)
Non Sect 8 (POdds)
Section 8 (POdds)
Time (0 = baseline, 1, 2, 3 = 6, 12, 24 months)


Marginal Logits - Ind vs Comm & Street
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
-2.5
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Non Sect 8 (obs)
Section 8 (obs)
Non Sect 8 (Non P)
Section 8 (Non P)
Non Sect 8 (POdds)
Section 8 (POdds)
Time (0 = baseline, 1, 2, 3 = 6, 12, 24 months)

SAS NLMIXED code: random-intercepts proportional odds model

```
DATA one; INFILE 'c:\sdhouse.dat' ;
INPUT id housing int section8 time1 time2 time3 sect8t1 sect8t2 sect8t3;

/* recode missing values */
IF housing = 999 THEN housing = .;

PROC NLMIXED ;
PARMS b0=0 b1=0 b2=0 b3=0 b4=0 b5=0 b6=0 b7=0 sd=1 g2=1;

z = b0 + b1*time1 + b2*time2 + b3*time3 + b4*section8
       + b5*sect8t1 + b6*sect8t2 + b7*sect8t3 + u;

IF (housing=0) THEN p = 1 / (1 + EXP(-(0-z)));
ELSE IF (housing=1) THEN p = (1/(1 + EXP(-(g2-z)))) - (1/(1 + EXP(-(0-z))));
ELSE IF (housing=2) THEN p = 1 - (1 / (1 + EXP(-(g2-z))));

loglike = LOG(p);
MODEL housing ~ GENERAL(loglike);
RANDOM u ~ NORMAL(0,sd*sd) SUBJECT=id;

ESTIMATE 'icc' sd*sd/(3.289868134+sd*sd);
RUN;
```

SAS NLMIXED code: random-intercepts non-proportional odds model
using data from the earlier NLMIXED example

```
PROC NLMIXED ;
PARMS b0=0 ga1=0 ga2=0 ga3=0 ga4=0 ga5=0 ga6=0 ga7=0
           gb1=0 gb2=0 gb3=0 gb4=0 gb5=0 gb6=0 gb7=0 sd=1 g2=1;

za = b0 + ga1*time1 + ga2*time2 + ga3*time3 + ga4*section8
        + ga5*sect8t1 + ga6*sect8t2 + ga7*sect8t3 + u;
zb = b0 + gb1*time1 + gb2*time2 + gb3*time3 + gb4*section8
        + gb5*sect8t1 + gb6*sect8t2 + gb7*sect8t3 + u;

IF (housing=0) THEN p = 1 / (1 + EXP(-(0-za)));
ELSE IF (housing=1) THEN p = (1/(1 + EXP(-(g2-zb)))) - (1/(1 + EXP(-(0-za))));
ELSE IF (housing=2) THEN p = 1 - (1 / (1 + EXP(-(g2-zb))));

loglike = LOG(p);
MODEL housing ~ GENERAL(loglike);
RANDOM u ~ NORMAL(0,sd*sd) SUBJECT=id;

ESTIMATE 'icc' sd*sd/(3.289868134+sd*sd);
RUN;
```

SAS IML code: computing marginal probabilities - ordinal models

```
TITLE1 'San Diego Homeless Data - Estimated Marginal Probabilities';
PROC IML;
/* Results from NLMIXED analysis:  proportional odds model */;

 x0 = { 0 0 0 0 0 0 0,
        1 0 0 0 0 0 0,
        0 1 0 0 0 0 0,
        0 0 1 0 0 0 0};
 x1 = { 0 0 0 1 0 0 0,
        1 0 0 1 1 0 0,
        0 1 0 1 0 1 0,
        0 0 1 1 0 0 1};

 int    = {-.219};
 sd     = {1.446};
 beta   = {1.731, 2.310, 2.494, 0.494, 1.403, 1.170, .634};
 thresh = {2.738};
```

```
/* number of quadrature points, quadrature nodes & weights */
nq = 10;
bq = { -4.85946282833231, -3.58182348355193, -2.48432584163895,
       -1.46598909439116, -0.48493570751550,  0.48493570751550,
        1.46598909439116,  2.48432584163895,  3.58182348355193,
        4.85946282833231};
wq = { 0.00000431065265, 0.00075807095698, 0.01911158107317,
       0.13548370704150, 0.34464234526294, 0.34464234526294,
       0.13548370704150, 0.01911158107317, 0.00075807095698,
       0.00000431065265};

/* initialize to zero */
 grp0a  =  J(4,1,0);
 grp0b  =  J(4,1,0);
 grp1a  =  J(4,1,0);
 grp1b  =  J(4,1,0);
```

```
DO q = 1 to nq;

    za0 = 0 - (int + x0*beta + sd*bq[q]);
    zb0 = thresh - (int + x0*beta + sd*bq[q]);
    za1 = 0 - (int + x1*beta + sd*bq[q]);
    zb1 = thresh - (int + x1*beta + sd*bq[q]);

    grp0a = grp0a + ( 1 / ( 1 + EXP(0 - za0)))*wq[q];
    grp0b = grp0b + ( 1 / ( 1 + EXP(0 - zb0)))*wq[q];

    grp1a = grp1a + ( 1 / ( 1 + EXP(0 - za1)))*wq[q];
    grp1b = grp1b + ( 1 / ( 1 + EXP(0 - zb1)))*wq[q];

END;

print 'Proportional odds model';
print 'Quadrature method - 10 points';
print 'marginal prob for group 0 - catg 1' grp0a [FORMAT=8.4];
print 'marginal prob for group 0 - catg 2' (grp0b-grp0a) [FORMAT=8.4];
print 'marginal prob for group 0 - catg 3' (1-grp0b) [FORMAT=8.4];
print 'marginal prob for group 1 - catg 1' grp1a [FORMAT=8.4];
print 'marginal prob for group 1 - catg 2' (grp1b-grp1a) [FORMAT=8.4];
print 'marginal prob for group 1 - catg 3' (1-grp1b) [FORMAT=8.4];
```

```
/* Non-Proportional Odds Model */;

 int    = {-.318};
 sd     = {1.445};
 gam1   = {2.288, 3.333, 2.814, .584, .569, -.948, -.365};
 gam2   = {1.077, 1.643, 2.143, .323, 2.014, 2.009, 1.064};
 thresh = {2.375};

/* initialize to zero */
 grp0a = J(4,1,0);
 grp0b = J(4,1,0);
 grp1a = J(4,1,0);
 grp1b = J(4,1,0);
```

```
DO q = 1 to nq;

    za0 = 0 - (int + x0*gam1 + sd*bq[q]);
    zb0 = thresh - (int + x0*gam2 + sd*bq[q]);
    za1 = 0 - (int + x1*gam1 + sd*bq[q]);
    zb1 = thresh - (int + x1*gam2 + sd*bq[q]);

    grp0a = grp0a + ( 1 / ( 1 + EXP(0 - za0)))*wq[q];
    grp0b = grp0b + ( 1 / ( 1 + EXP(0 - zb0)))*wq[q];

    grp1a = grp1a + ( 1 / ( 1 + EXP(0 - za1)))*wq[q];
    grp1b = grp1b + ( 1 / ( 1 + EXP(0 - zb1)))*wq[q];

END;

print 'Non-proportional odds model';
print 'Quadrature method - 10 points';
print 'marginal prob for group 0 - catg 1' grp0a [FORMAT=8.4];
print 'marginal prob for group 0 - catg 2' (grp0b-grp0a) [FORMAT=8.4];
print 'marginal prob for group 0 - catg 3' (1-grp0b) [FORMAT=8.4];
print 'marginal prob for group 1 - catg 1' grp1a [FORMAT=8.4];
print 'marginal prob for group 1 - catg 2' (grp1b-grp1a) [FORMAT=8.4];
print 'marginal prob for group 1 - catg 3' (1-grp1b) [FORMAT=8.4];
```

Mixed-effects Multinomial Logistic Regression Model for Nominal Responses (Hedeker, 2003)

y_{ij} = nominal response of level-2 unit i and level-1 unit j

$$\log \frac{p_{ijc}}{p_{ij1}} = \boldsymbol{u}'_{ij}\boldsymbol{\gamma}_c + \boldsymbol{z}'_{ij}\boldsymbol{T}_c\boldsymbol{\theta}_i \qquad c = 2, 3, \ldots C$$

- $C - 1$ contrasts to reference cell ($c = 1$)
- regression effects $\boldsymbol{\gamma}_c$ vary across contrasts
- random-effects variance terms $\boldsymbol{T}_c$ vary across contrasts

For example, with C = 3

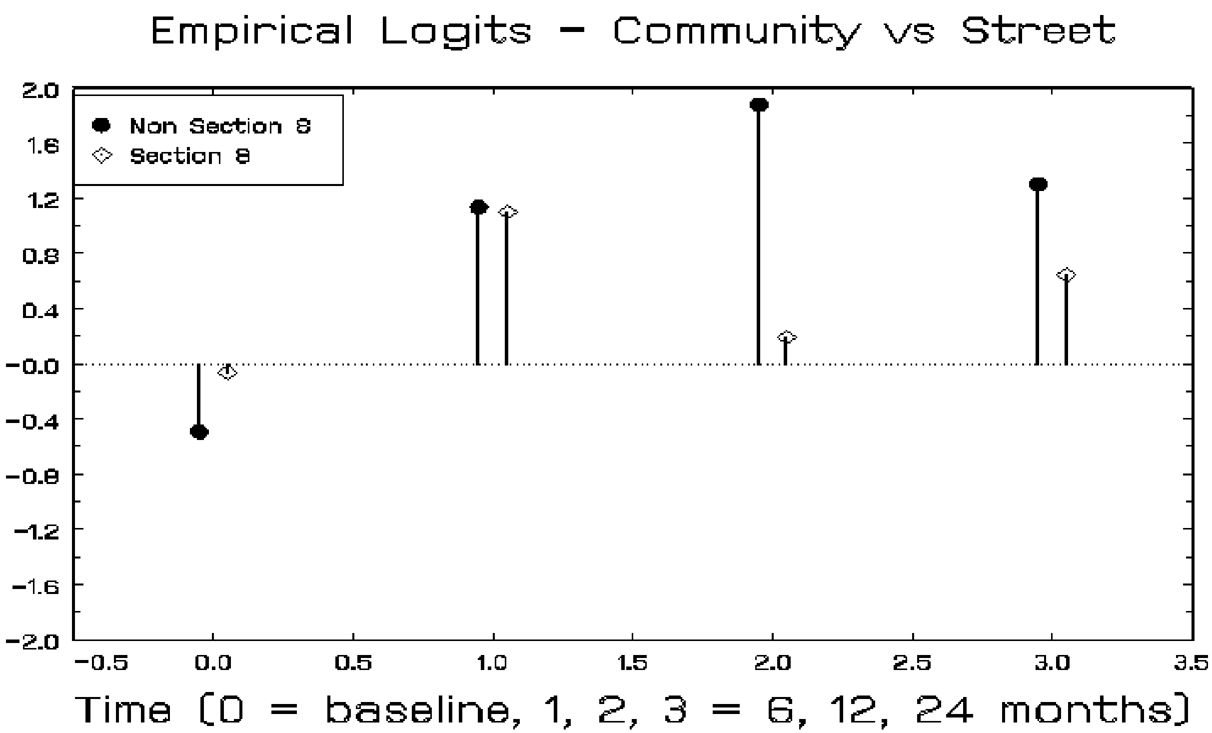
contrast	ordinal	nominal
c1	2 & 3 vs 1	2 vs 1
c2	3 vs 1 & 2	3 vs 1

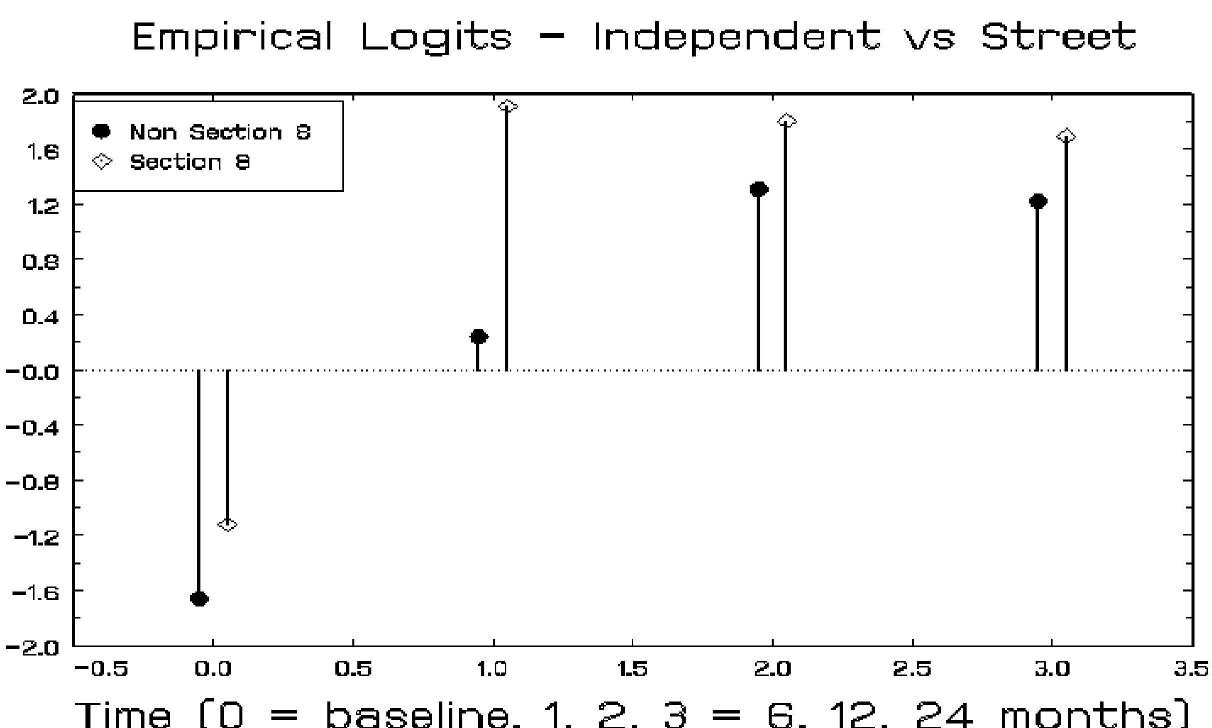
Model in terms of the category probabilities

$$p_{ijc} = \Pr(\mathsf{y}_{ij} = c \mid \boldsymbol{\theta}) = \frac{\exp(\mathrm{z}_{ijc})}{1 + \Sigma_{h=2}^{C} \exp(\mathrm{z}_{ijh})} \quad \text{for } c = 2, 3, \ldots, C$$

$$p_{ij1} = \Pr(\mathsf{y}_{ij} = 1 \mid \boldsymbol{\theta}) = \frac{1}{1 + \Sigma_{h=2}^{C} \exp(\mathrm{z}_{ijh})}$$

where the multinomial logit $\mathrm{z}_{ijc} = \boldsymbol{u}'_{ij}\boldsymbol{\gamma}_c + \boldsymbol{z}'_{ij}\boldsymbol{T}_c\,\boldsymbol{\theta}_i$

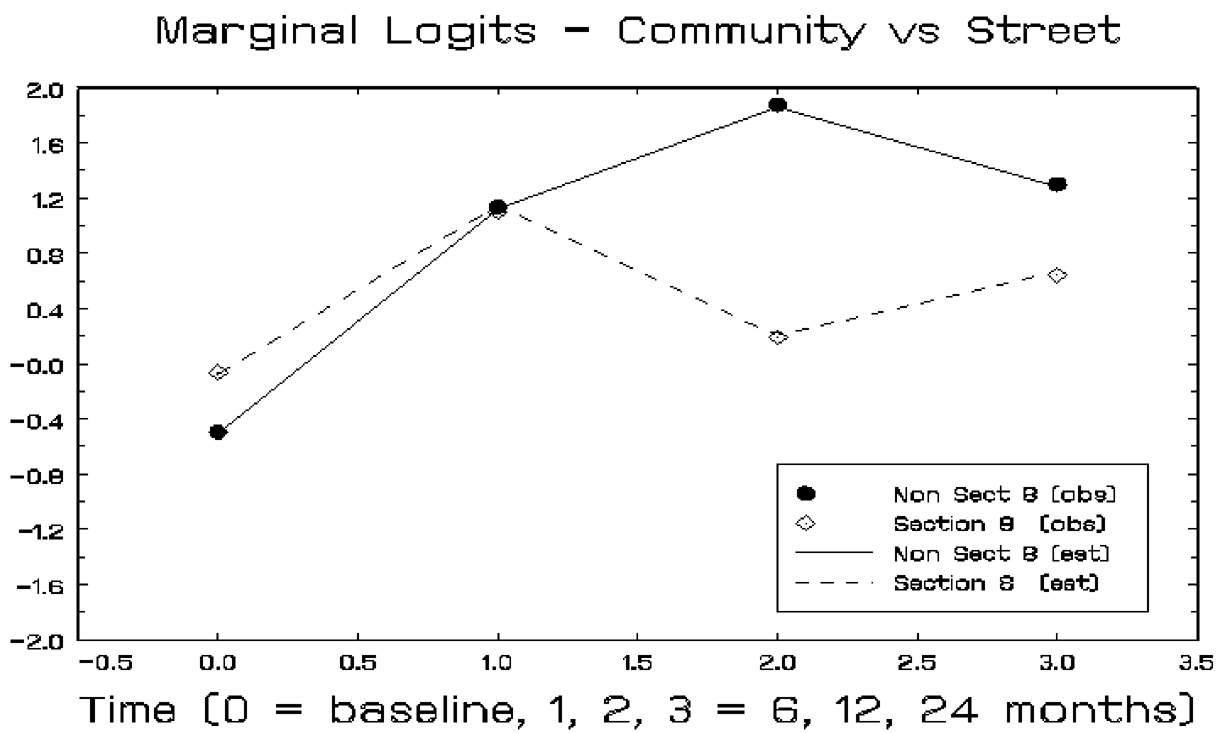

Empirical Logits - Community vs Street
Non Section 8
Section 8
2.0
1.6
1.2
0.8
0.4
-0.0
-0.4
-0.8
-1.2
-1.6
-2.0
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Time (0 = baseline, 1, 2, 3 = 6, 12, 24 months)

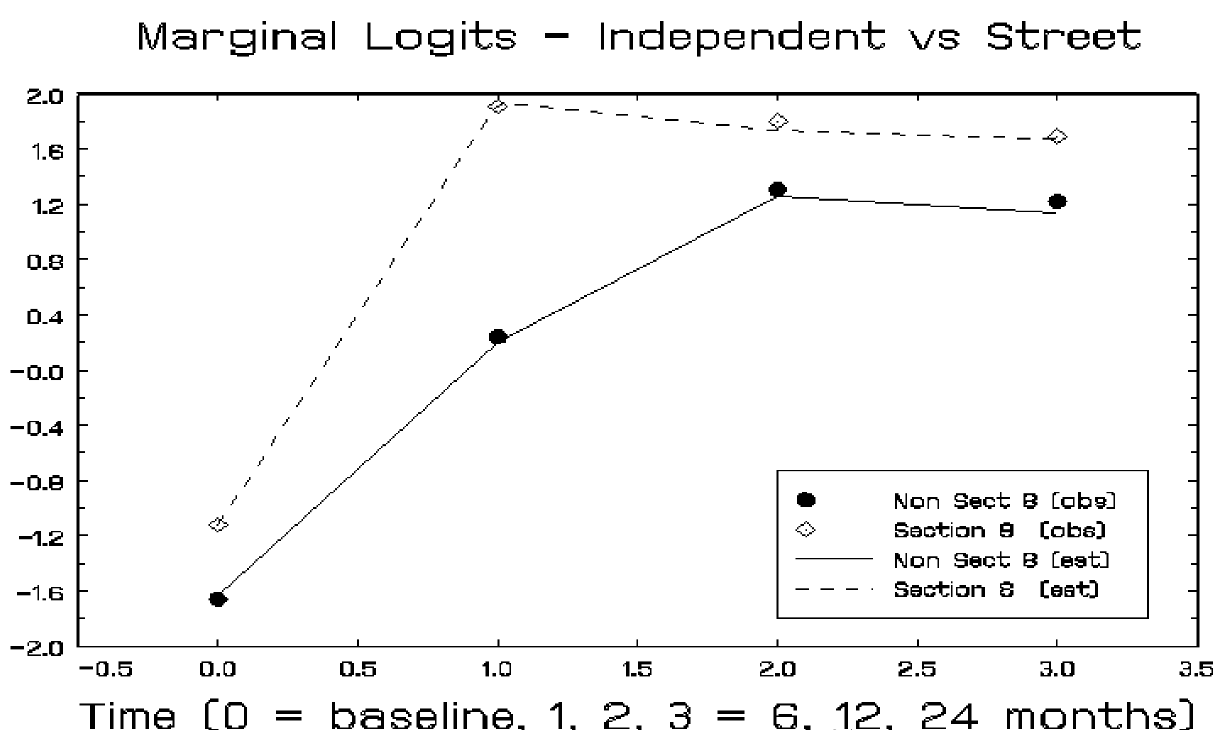

Empirical Logits - Independent vs Street
Non Section 8
Section 8
2.0
1.6
1.2
0.8
0.4
-0.0
-0.4
-0.8
-1.2
-1.6
-2.0
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Time (0 = baseline, 1, 2, 3 = 6, 12, 24 months)

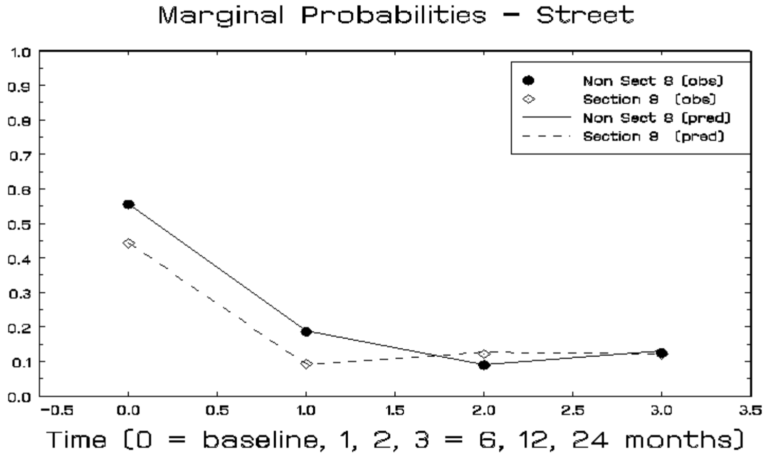
Housing status across time: 1289 observations within 361 subjects
Nominal Mixed Regression Model estimates & standard errors (se)

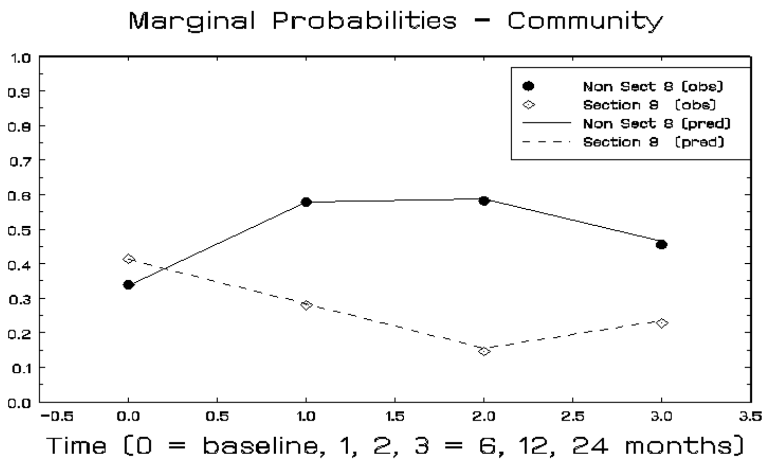
	Community vs Street		Independent vs Street	
term	estimate	se	estimate	se
intercept	- .452	.185	**-2.683**	.378
t1 (6 month)	**1.943**	.300	**2.683**	.436
t2 (12 month)	**2.822**	.384	**4.089**	.497
t3 (24 month)	**2.261**	.355	**4.099**	.470
section 8 (y=1)	**.521**	.262	.782	.481
section 8 by t1	-.134	.462	**2.000**	.619
section 8 by t2	**-1.916**	.517	.546	.650
section 8 by t3	**-.951**	.481	.303	.619
subject sd	**.874**	.187	**2.332**	.218
$-2 \log L$	2218.8			

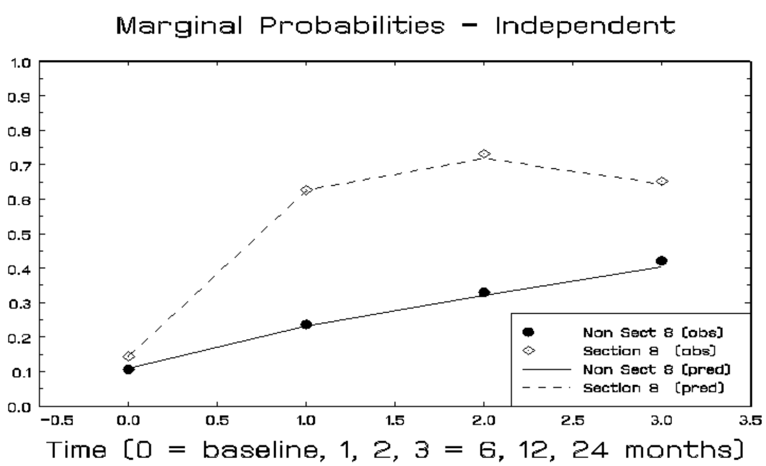
bold indicates $p < .05$ *italic* indicates $.05 < p < .10$


Marginal Logits - Community vs Street
2.0
1.6
1.2
0.8
0.4
-0.0
-0.4
-0.8
-1.2
-1.6
-2.0
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Non Sect 8 (obs)
Section 8 (obs)
Non Sect 8 (est)
Section 8 (est)
Time (0 = baseline, 1, 2, 3 = 6, 12, 24 months)


Marginal Logits - Independent vs Street
2.0
1.6
1.2
0.8
0.4
-0.0
-0.4
-0.8
-1.2
-1.6
-2.0
-0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
Non Sect 8 (obs)
Section 8 (obs)
Non Sect 8 (est)
Section 8 (est)
Time (0 = baseline, 1, 2, 3 = 6, 12, 24 months)


Marginal Probabilities – Street
Non Sect 8 (obs)
Section 8 (obs)
Non Sect 8 (pred)
Section 8 (pred)
Time (0 = baseline, 1, 2, 3 = 6, 12, 24 months)


Marginal Probabilities – Community
Non Sect 8 (obs)
Section 8 (obs)
Non Sect 8 (pred)
Section 8 (pred)
Time (0 = baseline, 1, 2, 3 = 6, 12, 24 months)


Marginal Probabilities – Independent
Non Sect 8 (obs)
Section 8 (obs)
Non Sect 8 (pred)
Section 8 (pred)
Time (0 = baseline, 1, 2, 3 = 6, 12, 24 months)

SAS NLMIXED code: random-intercepts nominal model with reference cell contrasts
using data from the earlier NLMIXED example

```
PROC NLMIXED ;
PARMS ga0=0 ga1=0 ga2=0 ga3=0 ga4=0 ga5=0 ga6=0 ga7=0 sda=1
      gb0=0 gb1=0 gb2=0 gb3=0 gb4=0 gb5=0 gb6=0 gb7=0 sdb=1;

za = ga0 + ga1*time1 + ga2*time2 + ga3*time3 + ga4*section8
         + ga5*sect8t1 + ga6*sect8t2 + ga7*sect8t3 + sda*u;
zb = gb0 + gb1*time1 + gb2*time2 + gb3*time3 + gb4*section8
         + gb5*sect8t1 + gb6*sect8t2 + gb7*sect8t3 + sdb*u;

IF (housing=0) THEN p = 1 / (1 + EXP(za) + EXP(zb));
ELSE IF (housing=1) THEN p = EXP(za) / (1 + EXP(za) + EXP(zb));
ELSE IF (housing=2) THEN p = EXP(zb) / (1 + EXP(za) + EXP(zb));

loglike = LOG(p);
MODEL housing ~ GENERAL(loglike);
RANDOM u ~ NORMAL(0,1) SUBJECT=id;

ESTIMATE 'icca' sda*sda/(3.289868134+sda*sda);
ESTIMATE 'iccb' sdb*sdb/(3.289868134+sdb*sdb);
RUN;
```

SAS IML code: computing marginal probabilities - nominal model

```
TITLE1 'San Diego Homeless Data - Estimated Marginal Probabilities';
PROC IML;
/* Results from MIXNO analysis */;

 u0  =  { 0 0 0 0 0 0 0,
          1 0 0 0 0 0 0,
          0 1 0 0 0 0 0,
          0 0 1 0 0 0 0};
 u1  =  { 0 0 0 1 0 0 0,
          1 0 0 1 1 0 0,
          0 1 0 1 0 1 0,
          0 0 1 1 0 0 1};

 inta  =  {-.452};
 sda   =  {.874};
 gama  =  {1.943, 2.822, 2.261, .521, -.134, -1.916, -.951};

 intb  =  {-2.673};
 sdb   =  { 2.332};
 gamb  =  { 2.683, 4.089, 4.099, .782, 2.000, .546, .303};
```

```
/* number of quadrature points, quadrature nodes & weights */
nq = 10;
bq = { -4.85946282833231, -3.58182348355193, -2.48432584163895,
       -1.46598909439116, -0.48493570751550,  0.48493570751550,
        1.46598909439116,  2.48432584163895,  3.58182348355193,
        4.85946282833231};
wq = { 0.00000431065265, 0.00075807095698, 0.01911158107317,
       0.13548370704150, 0.34464234526294, 0.34464234526294,
       0.13548370704150, 0.01911158107317, 0.00075807095698,
       0.00000431065265};

/* initialize to zero */
 grp0a  =  J(4,1,0);    grp0b = J(4,1,0);    grp0c = J(4,1,0);
 grp1a  =  J(4,1,0);    grp1b = J(4,1,0);    grp1c = J(4,1,0);
```

```
DO q = 1 to nq;
    za0 = u0*gama + inta + sda*bq[q];
    zb0 = u0*gamb + intb + sdb*bq[q];
    za1 = u1*gama + inta + sda*bq[q];
    zb1 = u1*gamb + intb + sdb*bq[q];

    grp0a = grp0a + 1 / (1+(EXP(za0)+EXP(zb0)))*wq[q];
    grp0b = grp0b + EXP(za0) / (1+(EXP(za0)+EXP(zb0)))*wq[q];
    grp0c = grp0c + EXP(zb0) / (1+(EXP(za0)+EXP(zb0)))*wq[q];

    grp1a = grp1a + 1 / (1+(EXP(za1)+EXP(zb1)))*wq[q];
    grp1b = grp1b + EXP(za1) / (1+(EXP(za1)+EXP(zb1)))*wq[q];
    grp1c = grp1c + EXP(zb1) / (1+(EXP(za1)+EXP(zb1)))*wq[q];

END;

print 'Quadrature method - 10 points';
print 'marginal prob for group 0 - category 1' grp0a [FORMAT=8.4];
print 'marginal prob for group 0 - category 2' grp0b [FORMAT=8.4];
print 'marginal prob for group 0 - category 3' grp0c [FORMAT=8.4];
print 'marginal prob for group 1 - category 1' grp1a [FORMAT=8.4];
print 'marginal prob for group 1 - category 2' grp1b [FORMAT=8.4];
print 'marginal prob for group 1 - category 3' grp1c [FORMAT=8.4];
```

Summary

Models for longitudinal categorical data as developed as models for continuous data

- Proportional odds models
- Non and partial proportional odds models
- Nominal models (with Reference-cell or Helmert contrasts)
- Scaling models (Hedeker, Berbaum, & Mermelstein, 2006)

Available software includes SAS PROC NLMIXED, STATA, SuperMix, ...

SuperMix

- Free student and 15-day trial editions
  http://www.ssicentral.com/supermix/downloads.html

- Datasets and examples
  http://www.ssicentral.com/supermix/examples.html

- Manual and documentation in PDF form
  http://www.ssicentral.com/supermix/resources.html