The Multilevel Logit Model for Binary Dependent Variables

Marco R. Steenbergen

Part I

The Single Level Logit Model: A Review

Motivating Example

- Imagine we are interested in voting for Labour in the 2001 British elections:

$$y_i = \begin{cases} 1 & \text{if voted for Labour} \\ 0 & \text{if voted for another party} \end{cases}$$

- We are interested in the effects of identification with the Labour party, ideological distance to Labour, and class.
- The quantity of interest is the probability of a Labour vote, π_i.

Formulation As a Generalized Linear Model

- Y_i is a binomial variable with mean $\mu_i = \pi_i$.
- Define the linear predictor as

$$\eta_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_P x_{Pi}$$

- We now need to link the mean to the linear predictor, which can be done as follows:

$$\begin{aligned} \eta_i &= \ln\left(\frac{\pi_i}{1-\pi_i}\right) = logit_i \\ \pi_i &= logit_i^{-1} = \frac{\exp(\eta_i)}{1+\exp(\eta_i)} \end{aligned}$$

where *logit* is known as the link function.

A Latent Variable Model

– We think of Y as a reflection of an underlying continuum Y^*, which remains unobserved.

– The latent variable is a linear function of the predictors:

$$\begin{aligned} y_i^* &= \beta_0 + \beta_1 x_{1i} + \cdots + \beta_P x_{Pi} + \epsilon_i \\ &= \eta_i + \epsilon_i \end{aligned}$$

with ϵ_i being a standard logistic variate: $\epsilon_i \sim \mathcal{L}(0, 1)$.

A Latent Variable Model Cont'd

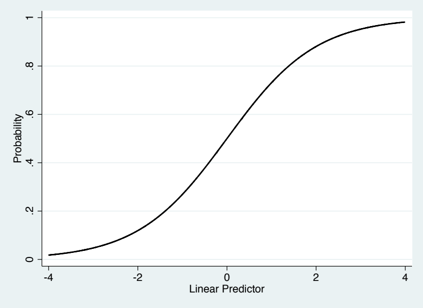
- The latent and observed dependent variables are linked as follows:

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

- Consequently,

$$\begin{aligned} \pi_i &= \Pr(y_i^* > 0) \\ &= \Pr(\eta_i + \epsilon_i > 0) \\ &= \Pr(\epsilon_i > -\eta_i) \\ &= \Pr(\epsilon \leq \eta_i) \\ &= \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \\ &= F(\eta_i) \end{aligned}$$

The Standard Logistic Distribution



Example: The Labour Vote in 2001

Parameter	Estimate	SE
Labour Identifier	4.27	0.17
Ideological Distance	-0.21	0.05
Middle Class	-0.45	0.22
Working Class	0.23	0.19
Constant	-1.89	0.15

Notes: $n = 1679$. Source: 2001 British Election Study. Estimated using `logit`.

Alternative Link Functions

- Probit:
  - ϵ_i follows the standard normal distribution
- Complementary log-log:
  - ϵ_i follows the Gumbel distribution

The Error Variance

- The error variance is fixed in the logit model to $Var(\epsilon) = \pi^2/3 \approx 3.29$.
- This is in order to fix the scale of Y^* and thereby of the βs.
- In multilevel extensions this means that no level-1 variance will be estimated.

Estimation

- Estimation proceeds via maximum likelihood.
- The observed dependent variable, Y, follows the binomial distribution (assuming a single trial):

$$f(y) = \pi^y (1-\pi)^{1-y}$$

- For n independent observations, the likelihood function is given by

$$L(y|\beta_0 \cdots \beta_P) = \prod_i f(y_i) = \prod_i \pi_i^{y_i} (1-\pi_i)^{1-y_i}$$

- This is optimized with respect to $\beta_0 \cdots \beta_P$.

Estimation Cont'd

- Optimization of the likelihood is done through a numeric optimizer.
- In particular, a hill-climbing algorithm such as Newton-Raphson is used.
- Here starting values are updated in successive steps, depending on the gradient.
- For the logit model, convergence is usually fast.

Interpretation: General Comments

- The logit model is a nonlinear model, which means that the coefficients cannot be directly interpreted.
- We focus on two interpretation methods:
  1. Predicted probabilities
  2. Odds ratios

Predicted Probabilities

For given values of the predictors $X_1 \cdots X_P$, the predicted probability is

$$\hat{\pi}_i = \frac{\exp(\hat{\eta}_i)}{1 + \exp(\hat{\eta}_i)}$$

with

$$\hat{\eta}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_P x_{Pi}$$

Example: The Labour Vote in 2001

- Consider a working class voter who does not identify with Labour and who is at a median ideological distance from Labour (1 unit).
- Based on the earlier estimates, $\hat{\eta}_i = -1.87$ and $\hat{\pi}_i = 0.13$.

The Odds and Odds Ratio

- The odds are given by

$$\frac{\Pr(y_i = 1)}{\Pr(y_i = 0)} = \frac{\pi_i}{1 - \pi_i} = \exp(\eta_i)$$

- The odds ratio is the ratio of two odds, evaluated at different values of a predictor.
- Let X_p change by δ units, while holding all else constant. Then the odds ratio is given by

$$or = \exp(\beta_p \delta)$$

- When $\delta = 1$ this is referred to as the factor change in the odds.

Example: The Labour Vote in 2001

- The factor change in the odds due to an identification with Labour is $\exp(4.27) = 71.36$.
- This means that the odds of a Labour vote are 71 times higher for those who identify with Labour than those who do not.
- Note: This result does not depend on the values of the remaining covariates or the starting value of the covariate of interest.

Part II

Derivation of the Multilevel Logit Model

A Random Intercept Model

Level-1 Model

For unit i in context j,

$$logit_{ij} = \beta_{0j} + \beta_{1j} x_{ij}$$

Level-2 Model

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01} z_j + \delta_{0j} \\ \delta_{0j} &\sim \mathcal{N}(0, \tau_{00}) \\ \beta_{1j} &= \gamma_{10} \end{aligned}$$

A Random Intercept Model Cont'd

Mixed Model

$$logit_{ij} = \gamma_{00} + \gamma_{01} z_j + \gamma_{10} x_{ij} + \delta_{0j}$$

or

$$\pi_{ij} = \frac{\exp(\gamma_{00} + \gamma_{01} z_j + \gamma_{10} x_{ij} + \delta_{0j})}{1 + \exp(\gamma_{00} + \gamma_{01} z_j + \gamma_{10} x_{ij} + \delta_{0j})}$$

A Random Intercept Model Cont'd

Alternative Formulation

$$
\begin{aligned}
y_{ij}^* &= \beta_{0j} + \beta_{1j} x_{ij} + \epsilon_{ij} && \text{Level-1 Model} \\
\beta_{0j} &= \gamma_{00} + \gamma_{01} z_j + \delta_{0j} && \text{Level-2 Model} \\
\beta_{1j} &= \gamma_{10} && \\
y_{ij}^* &= \gamma_{00} + \gamma_{01} z_j + \gamma_{10} x_{ij} + \delta_{0j} + \epsilon_{ij} && \text{Mixed Model} \\
\epsilon_{ij} &\sim \mathcal{L}(0, 1) && \text{Error Distributions} \\
\delta_{0j} &\sim \mathcal{N}(0, \tau_{00}) && \\
\pi_{ij} &= F\left(\gamma_{00} + \gamma_{01} z_j + \gamma_{10} x_{ij} + \delta_{0j}\right) && \text{Choice Probability}
\end{aligned}
$$

The Intra-Class Correlation Revisited

– The usual formula for the intraclass correlation is

$$\rho = \frac{\tau_{00}}{\tau_{00} + \sigma^2}$$

where σ^2 is the level-1 error variance.

– In a multilevel logit model $\sigma^2 = \pi^2/3$ by assumption, so that the ICC is computed as

$$\rho = \frac{\tau_{00}}{\tau_{00} + \frac{\pi^2}{3}}$$

– This is the ICC for the latent response variable.

Example: The Labour Vote in 2001

– Consider again the 2001 BES data, which were previously analyzed as a single-level structure.

– In fact, the data can be broken down by district.

– For now, we estimate a model without level-2 covariates, namely:

$$logit_{ij} = \gamma_{00} + \gamma_{10} labid_{ij} + \gamma_{20} dist_{ij} + \gamma_{30} middle_{ij} + \gamma_{40} working_{ij} + \delta_{0j}$$

– We want to know to what extent the Labour vote varied across districts.

Example: The Labour Vote in 2001

Parameter	Estimate	SE
Labour Identifier	4.49	0.20
Ideological Distance	-0.23	0.05
Middle Class	-0.43	0.23
Working Class	0.21	0.20
Constant	-1.96	0.18
τ_{00}	0.45	0.19

Notes: $n = 1679$, $J = 127$. $\hat{\rho} = .12$. Source: 2001 British Election Study. Estimated using `gllamm`.

A Random Slope and Intercept Model

Level-1 Model

For unit i in context j,

$$logit_{ij} = \beta_{0j} + \beta_{1j} x_{ij}$$

A Random Slope and Intercept Model Cont'd

Level-2 Model

$$
\begin{aligned}
\beta_{0j} &= \gamma_{00} + \gamma_{01} z_j + \delta_{0j} \\
\beta_{1j} &= \gamma_{10} + \gamma_{11} z_j + \delta_{1j} \\
\begin{pmatrix} \delta_{0j} \\ \delta_{1j} \end{pmatrix} &\sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \\ \tau_{01} & \tau_{11} \end{bmatrix} \right) \\
&\sim \mathcal{N}(\mathbf{0}, \boldsymbol{T})
\end{aligned}
$$

A Random Intercept and Slope Model Cont'd

Mixed Model

$$logit_{ij} = \gamma_{00} + \gamma_{01} z_j + \gamma_{10} x_{ij} + \gamma_{11} z_j x_{ij} + \delta_{0j} + \delta_{1j} x_{ij}$$

or

$$\pi_{ij} = \frac{\exp(\gamma_{00} + \gamma_{01} z_j + \gamma_{10} x_{ij} + \gamma_{11} z_j x_{ij} + \delta_{0j} + \delta_{1j} x_{ij})}{1 + \exp(\gamma_{00} + \gamma_{01} z_j + \gamma_{10} x_{ij} + \gamma_{11} z_j x_{ij} + \delta_{0j} + \delta_{1j} x_{ij})}$$

A Random Intercept and Sope Model Cont'd

Alternative Formulation

$$
\begin{aligned}
y_{ij}^{*} &= \beta_{0j} + \beta_{1j}x_{ij} + \epsilon_{ij} && \text{Level-1 Model} \\
\beta_{0j} &= \gamma_{00} + \gamma_{01}z_j + \delta_{0j} && \text{Level-2 Model} \\
\beta_{1j} &= \gamma_{10} + \gamma_{11}z_j + \delta_{1j} && \\
y_{ij}^{*} &= \gamma_{00} + \gamma_{01}z_j + \gamma_{10}x_{ij} + \gamma_{11}z_j x_{ij} + && \text{Mixed Model} \\
&\quad \delta_{0j} + \delta_{1j}x_{ij} + \epsilon_{ij} && \\
\epsilon_{ij} &\sim \mathcal{L}(0,1) && \text{Error Distributions} \\
\delta_{0j}, \delta_{1j} &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{T}) && \\
\pi_{ij} &= F\left(\gamma_{00} + \gamma_{01}z_j + \gamma_{10}x_{ij} + \gamma_{11}z_j x_{ij} + \right. && \text{Choice Probability} \\
&\quad \left. \delta_{0j} + \delta_{1j}x_{ij}\right) &&
\end{aligned}
$$

Heterogeneity

- In the single-level logit model, $V[y_i^*] = \pi^2/3$, which is constant.
- Due to "causal" heterogeneity, the variance in the random intercept and slope model is not fixed. Rather,

$$V[y_{ij}^*] = \tau_{00} + 2\tau_{01}x_{ij} + \tau_{11}x_{ij}^2 + \frac{\pi^2}{3}$$

- This is important because the fixed variance assumption is used to fix the scale of the estimators.
- If that assumption is false, the scale of the estimators may be fixed incorrectly.

A General Model

Level-1 Model

Collect information about all of the covariates (and the constant) in the $(P+1) \times 1$ column vector $\boldsymbol{x}_{ij}$. Then, the level-1 model may be written as

$$logit_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta}_j$$

where $\boldsymbol{\beta}_j$ is a $(P+1) \times 1$ column vector of random coefficients.

Level-2 Model

$$\begin{aligned} \boldsymbol{\beta}_j &= \boldsymbol{Z}_j \gamma + \boldsymbol{\delta}_j \\ \boldsymbol{\delta}_j &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{T}) \end{aligned}$$

A General Model Cont'd

Mixed Model in Logit Form

$$logit_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{Z}_j \gamma + \boldsymbol{x}_{ij}^T \boldsymbol{\delta}_j$$

Mixed Model in Probability Form

$$\pi_{ij} = \frac{\exp(\boldsymbol{x}_{ij}^T \boldsymbol{Z}_j \gamma + \boldsymbol{x}_{ij}^T \boldsymbol{\delta}_j)}{1 + \exp(\boldsymbol{x}_{ij}^T \boldsymbol{Z}_j \gamma + \boldsymbol{x}_{ij}^T \boldsymbol{\delta}_j)}$$

Part III

Estimation

III.A

Estimation Theory

General Comment

- Relative to the HLM, estimation of the multilevel logit model is much more complicated.
- It requires special tools, either in the form of numerical integration or simulation; we focus on the former.

The Likelihood Function

The Conditional Likelihood

Imagine we know the elements of the vector $\boldsymbol{\delta}_j$ and that these are the only source of dependencies in the data. Then,

$$L(\boldsymbol{y}, \gamma | \boldsymbol{\delta}_j) = \prod_i \prod_j \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{1 - y_{ij}}$$

The Likelihood Function Cont'd

The Unconditional Log-Likelihood

Of course, the elements of $\boldsymbol{\delta}_j$ are unknown, which means they have to be integrated out of the likelihood function:

$$L(\boldsymbol{y}|\gamma, \boldsymbol{\delta}_j) = \int_{\boldsymbol{\delta}_j} \Phi_{P+1} L(\boldsymbol{y}, \gamma|\boldsymbol{\delta}_j) d\boldsymbol{\delta}_j$$

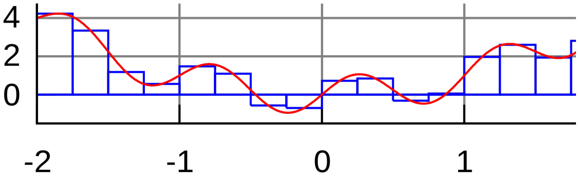
where Φ_{P+1} is the $P+1$-variate normal distribution over the level-2 errors with covariance matrix $\boldsymbol{T}$ and the integral is of dimension $P+1$.

The Difficulty

- With the exception of the random intercept model, the likelihood function is difficult to evaluate.
- One solution is to first use numerical integration of the likelihood and then optimize it; this is what `gllamm` does.
- Note that numerical integration is computer intensive and can be very slow.

Numerical Integration: Basic Ideas

Rectangular Integration



Source: wikipedia.org

Numerical Integration: Basic Ideas Cont'd

- With the rectangular approximation, the integration points are equally spaced.
- The problem here is that, unless the number of integration points is large, the approximation can be very crude.
- Gaussian quadrature is a major improvement over rectangular approximation.
- This can be further improved through adaptive quadrature.

Numerical Integration: Gaussian Quadrature

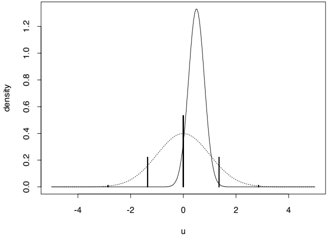
- With Gaussian quadrature, a smaller number of well-chosen points are used to improve the performance of the integration.
- These points have fixed locations and weights and are combined using

$$\int f(x)dx \approx \sum_q w_q f(x_q)$$

  where q denotes an integration point and w is the weight.
- In general, if the integrand is a polynomial of order $2k-1$, then k integration points suffice for exact integration.

Numerical Integration: Gaussian Quadrature Cont'd

Gaussian Quadrature

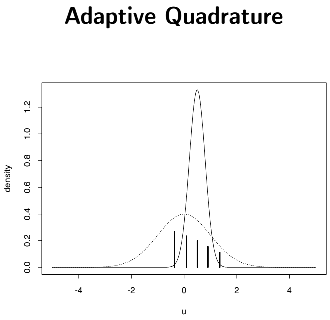


Source: Rabe-Hesketh et al. (2002)

Numerical Integration: Adaptive Quadrature

- As the illustration shows, when the distribution is peaked, the selection of quadrature points may not be ideal when the function is strongly peaked.
- In multilevel analysis, such peakedness can arise when cluster sizes are large.
- In this case, adaptive quadrature may perform better.
- Here the integration points are chosen under the peak.
- The result can be a better approximation with fewer quadrature points.

Numerical Integration: Adaptive Quadrature Cont'd



Source: Rabe-Hesketh et al. (2002)

Numerical Integration: Practical Considerations

- With several random effects, it is possible to use spherical quadrature to further improve things.
- There is no consensus on how many integration points are sufficient.
- There is consensus that more points is better.
- In case of doubt, try the same analysis specifying different numbers of integration points.

EB Estimates

- The level-2 error terms can be estimated using empirical Bayes estimation.
- Specifically, they are equal to the mean of the posterior with the ML estimates plugged in.

III.B

Using Stata

Estimation Routines

- Stata offers three different estimation routines:
  1. `xtlogit` (or `xtprobit`) for random intercept models
  2. `xtmelogit` for random coefficient models
  3. `gllamm` for random coefficient models
- In terms of speed,

$$\text{gllamm} < \text{xtmelogit} < \text{xtlogit}$$

- In terms of versatility,

$$\text{xtlogit} < \text{xtmelogit} < \text{gllamm}$$

gllamm Syntax for a Random Intercept Model

The basic syntax is:

gllamm y [x], i(name) link(logit) family(binom) nip(#) adapt

Here:

- y is the name of the binary outcome variable
- x is a list of the level-1 and level-2 covariates, as well as any cross-level interactions
- i(name) specifies the name of the level-2 units
- link(logit) specifies the logit link function (for probit, this would be link(probit))
- family(binom) specifies that Y follows the binomial distribution
- nip(#) specifies the number of integration points
- adapt asks for adaptive quadrature

gllamm Syntax for a Random Slope and Intercept Model

For a model with a single predictor x, the syntax is:

```
gen cons=1
eq inter:  cons
eq slope:  x
gllamm y x,(name) link(logit) family(binom) nip(#)
adapt eqs(inter slope) nrf(2) [nocorrel] [ip(m)]
```

Here

- nrf(2) specifies the number of random effects
- nocorrel constrains all covariance components to 0
- ip(m) calls for spherical quadrature

`gllamm` Syntax for Empirical Bayes Residuals

The EB residuals and their standard errors are obtained via

gllapred eb, u

This stores the means and standard deviations in variables with the stubs `ebm` and `ebs`, respectively.

Example: The Labour Vote in 2001

Part IV

Interpretation

General Comments

- The most prominent ways of interpreting hierarchical logit models is via (1) odds ratios or (2) predicted probabilities.
- Most frequently, scholars focus on the fixed effects.
- Odds ratio interpretations are based on the logit.
- Since the logit is a linear function, this form of interpretation is relatively simple.
- With predicted probabilities, things become more complex, since they are nonlinear functions.

IV.A

Odds Ratios

Models Without Interactions

– Consider the model $logit_{ij} = \gamma_{00} + \gamma_{01} z_j + \gamma_{10} x_{ij} + \delta_{0j} + \delta_{1j} x_{ij}$.

– A fixed effects interpretation ignores the last two terms, based on the assumption that they average to 0 and therefore do not contribute to the odds ratio.

– Interpretation is then analogous to the single-level logit model. Specifically,

$$\begin{array}{ll} \exp(\gamma_{01}) & \text{factor change due to } Z \\ \exp(\gamma_{10}) & \text{factor change due to } X \end{array}$$

Example: The Labour Vote in 2001

Predictor	Estimate	Factor Change
Labour Identifier	4.49	89.17
Ideological Distance	-0.23	0.80
Middle Class	-0.43	0.65
Working Class	0.21	1.23

Notes: Model estimates expressed as odds ratios using `gllamm, eform` after the estimation.

Confidence Intervals for the Factor Changes

- The 95% confidence interval for the effect of a covariate on the logit is given by

$$\hat{\gamma}_{pq} \pm z_{.975}\widehat{se}_{\hat{\gamma}_{pq}}$$

- We can use an endpoint transformation to obtain the 95% confidence interval for the factor change.

$$\begin{array}{ll} \text{Lower Bound} & \exp\left(\hat{\gamma}_{pq} - z_{.975}\widehat{se}_{\hat{\gamma}_{pq}}\right) \\ \text{Upper Bound} & \exp\left(\hat{\gamma}_{pq} + z_{.975}\widehat{se}_{\hat{\gamma}_{pq}}\right) \end{array}$$

Example: The Labour Vote in 2001

Predictor	Estimate	95% CI	
Labour Identifier	89.17	60.47	131.49
Ideological Distance	0.80	0.72	0.88
Middle Class	0.65	0.41	1.02
Working Class	1.23	0.83	1.83

Notes: Model estimates expressed as odds ratios using `gllamm, eform` after the estimation.

Models With Interactions

– Consider the following model with a cross-level interaction:

$$logit_{ij} = \gamma_{00} + \gamma_{01} z_j + \gamma_{10} x_{ij} + \gamma_{11} z_j x_{ij} + \delta_{0j} + \delta_{1j} x_{ij}$$

– Averaging over level-1 and level-2 units allows us to drop the error terms at the end.

– A better handle on the interaction is obtained via the simple slope on the logit:

$$\frac{\partial E[logit]}{\partial x} = \gamma_{10} + \gamma_{11} z$$

$$\frac{\partial E[logit]}{\partial z} = \gamma_{01} + \gamma_{11} x$$

Example: The Labour Vote in 2001

- Consider the following model

$$
\begin{aligned}
logit_{ij} &= \beta_{0j} + \beta_{1j} labid_{ij} + \beta_{2j} dist_{ij} + \beta_{3j} middle_{ij} + \beta_{4j} working_{ij} \\
\beta_{0j} &= \gamma_{00} + \gamma_{01} lab01_j + \delta_{0j} \\
\beta_{1j} &= \gamma_{10} + \gamma_{11} lab01_j + \delta_{1j} \\
\beta_{2j} &= \gamma_{20} \\
\beta_{3j} &= \gamma_{30} \\
\beta_{4j} &= \gamma_{40}
\end{aligned}
$$

- Or

$$
\begin{aligned}
logit_{ij} &= \gamma_{00} + \gamma_{01} lab01_j + \gamma_{10} labid_{ij} + \gamma_{11} lab01_j labid_{ij} + \\
&\quad \gamma_{20} dist_{ij} + \gamma_{30} middle_{ij} + \gamma_{40} working_{ij} + \delta_{0j} + \delta_{1j} labid_{ij}
\end{aligned}
$$

Example Cont'd

Predictor	Estimate	SE
Labour Share in District	0.03	0.01
Labour Identifier	3.46	0.54
Identifier $\times$ District Share	0.02	0.01
Ideological Distance	-0.23	0.05
Middle Class	-0.47	0.23
Working Class	0.01	0.20
Constant	-3.06	0.41
τ_{00}	0.07	0.21
τ_{11}	0.51	0.47
τ_{01}	-0.15	0.27

Notes: $J = 127$, $N = 1679$. Source: 2001 BES.

Testing Specific Values of the Simple Slope

- Imagine I want to know the effect of identification with Labour on the logit when the Labour vote share is 50%.
- In Stata, we issue:

```
lincom labid+inter*50
```

- The resulting simple slope estimate is 4.58 with a standard error of .25.
- This can be translated into a factor change through exponentiation.

Do We Still Need Random Effects?

- Inspection shows that the variance and covariance components are all around their standard error sin size, or smaller.
- This suggests they are not statistically significant.
- To test this accurately, we can use a LR test.

Do We Still Need Random Effects? Cont'd

– Syntax:

```
eq inter:  cons
eq slope:  labid
gllamm labvote lab01 labid lr_dist middle working
inter, i(polldist) link(logit) family(binom)
eqs(inter slope) nip(15) ip(m) adapt
est store full
logit labvote lab01 labid lr_dist middle working
inter
lrtest full, force
```

– This yields $\chi^2_3 = 2.12$ with a (conservative) p-value of .55.

– This is evidence that there is no residual heterogeneity.

IV.B

Predicted Probabilities

Two Types of Predicted Probabilities

1. Marginal predicted probabilities:
   - Predicted probabilities with the random effects integrated out
   - Allows one to focus on the fixed effects only
2. Conditional predicted probabilities:
   - Predicted probabilities conditional on a particular level-2 unit's error components
   - Allows one to focus on the random effects

Marginal Predicted Probabilities

Formula

$$\Pr(y_{ij} = 1 | \boldsymbol{x}_{ij}, \boldsymbol{Z}_j) \quad = \quad \int_{\boldsymbol{\delta_j}} \Pr(y_{ij} = 1 | \boldsymbol{x}_{ij}, \boldsymbol{Z}_j, \boldsymbol{\delta}_j) \phi(\boldsymbol{\delta}_j | \boldsymbol{T}) d\boldsymbol{\delta}_j$$

Note: Although it is sometimes explained this way, it is not correct to set $\boldsymbol{\delta}_j = \mathbf{0}$ and then compute the predicted probability. (The average of the inverse logit $\neq$ inverse logit of the average.)

Obtaining Marginal Predicted Probabilities in Stata

`gllamm` can compute the marginal predicted probabilities using

gllapred name, mu marginal

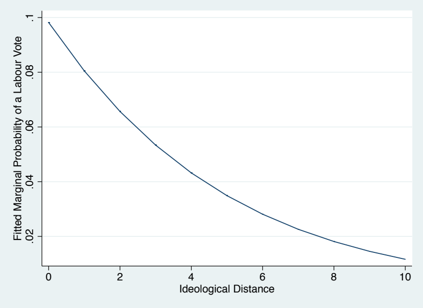
Example: The Labour Vote in 2001

- Consider the random intercept model

$$logit_{ij} = \gamma_{00} + \gamma_{10} labid_{ij} + \gamma_{20} dist_{ij} + \gamma_{30} middle_{ij} + \gamma_{40} working_{ij} + \delta_{0j}$$

- The marginal predicted probabilities are given by integrating over δ_{0j}.
- We want to plot them by ideological distance.

Example Cont'd



Note: Plot is for middle class voters who do not identify with the Labour party.

Conditional Predicted Probabilities

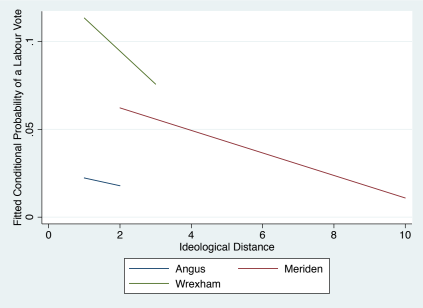
Formula

$$\Pr(y_{ij} = 1 | \boldsymbol{x}_{ij}, \boldsymbol{Z}_j, \boldsymbol{\delta}_j) = \frac{\exp(\boldsymbol{x}_{ij}^T \boldsymbol{Z}_j \gamma + \boldsymbol{x}_{ij}^T \boldsymbol{\delta}_j)}{1 + \exp(\boldsymbol{x}_{ij}^T \boldsymbol{Z}_j \gamma + \boldsymbol{x}_{ij}^T \boldsymbol{\delta}_j)}$$

Stata

```
gllapred name, mu
```

Example Cont'd



Note: Plot is for middle class voters who do not identify with the Labour party.