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, \( \pi_i \).
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:

$$\eta_i = \ln \left( \frac{\pi_i}{1 - \pi_i} \right) = \text{logit}_i$$

$$\pi_i = \text{logit}_i^{-1} = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}$$

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:

$$y_i^* = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_P x_{Pi} + \epsilon_i$$

$$= \eta_i + \epsilon_i$$

with $\epsilon_i$ 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,

\[ \pi_i = \Pr(y_i^* > 0) \]
\[ = \Pr(\eta_i + \epsilon_i > 0) \]
\[ = \Pr(\epsilon_i > -\eta_i) \]
\[ = \Pr(\epsilon \leq \eta_i) \]
\[ = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \]
\[ = F(\eta_i) \]
The Standard Logistic Distribution
Example: The Labour Vote in 2001

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labour Identifier</td>
<td>4.27</td>
<td>0.17</td>
</tr>
<tr>
<td>Ideological Distance</td>
<td>-0.21</td>
<td>0.05</td>
</tr>
<tr>
<td>Middle Class</td>
<td>-0.45</td>
<td>0.22</td>
</tr>
<tr>
<td>Working Class</td>
<td>0.23</td>
<td>0.19</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.89</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Alternative Link Functions

- Probit:
  - $\epsilon_i$ follows the standard normal distribution
- Complementary log-log:
  - $\epsilon_i$ follows the Gumbel distribution
The Error Variance

- The error variance is fixed in the logit model to $\text{Var}(\epsilon) = \pi^2/3 \approx 3.29$.
- This is in order to fix the scale of $Y^*$ and thereby of the $\beta$s.
- 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
  \[
  L(y|\beta_0 \cdots \beta_P) = \prod_i f(y_i) = \prod_i \pi_{yi}^y (1 - \pi_i)^{1-y_i}
  \]
- 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 $\delta$ 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 $j$,

$$\text{logit}_{ij} = \beta_0 + \beta_1 x_{ij}$$

Level-2 Model

$$\beta_0 = \gamma_0 + \gamma_1 z_j + \delta_0$$

$$\delta_0 \sim \mathcal{N}(0, \tau_0)$$

$$\beta_1 = \gamma_1$$
A Random Intercept Model Cont’d

Mixed Model

\[ \text{logit}_{ij} = \gamma_{00} + \gamma_{01}z_{j} + \gamma_{10}x_{ij} + \delta_{0j} \]

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

$$y^*_ij = \beta_{0j} + \beta_{1j} x_{ij} + \epsilon_{ij}$$  
**Level-1 Model**

$$\beta_{0j} = \gamma_{00} + \gamma_{01} z_j + \delta_{0j}$$  
**Level-2 Model**

$$\beta_{1j} = \gamma_{10}$$

$$y^*_ij = \gamma_{00} + \gamma_{01} z_j + \gamma_{10} x_{ij} + \delta_{0j} + \epsilon_{ij}$$  
**Mixed Model**

$$\epsilon_{ij} \sim \mathcal{N}(0, 1)$$  
**Error Distributions**

$$\delta_{0j} \sim \mathcal{N}(0, \tau_{00})$$

$$\pi_{ij} = F(\gamma_{00} + \gamma_{01} z_j + \gamma_{10} x_{ij} + \delta_{0j})$$  
**Choice Probability**
The Intra-Class Correlation Revisited

– The usual formula for the intraclass correlation is

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

where \( \sigma^2 \) 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:

\[
\text{logit}_{ij} = \gamma_{00} + \gamma_{10} \text{labid}_{ij} + \gamma_{20} \text{dist}_{ij} + \gamma_{30} \text{middle}_{ij} + \gamma_{40} \text{working}_{ij} + \delta_{0j}
\]

- We want to know to what extent the Labour vote varied across districts.
Example: The Labour Vote in 2001

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labour Identifier</td>
<td>4.49</td>
<td>0.20</td>
</tr>
<tr>
<td>Ideological Distance</td>
<td>-0.23</td>
<td>0.05</td>
</tr>
<tr>
<td>Middle Class</td>
<td>-0.43</td>
<td>0.23</td>
</tr>
<tr>
<td>Working Class</td>
<td>0.21</td>
<td>0.20</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.96</td>
<td>0.18</td>
</tr>
<tr>
<td>$\tau_{00}$</td>
<td>0.45</td>
<td>0.19</td>
</tr>
</tbody>
</table>

A Random Slope and Intercept Model

Level-1 Model
For unit $i$ in context $j$,

$$\text{logit}_{ij} = \beta_0 + \beta_1 x_{ij}$$
A Random Slope and Intercept Model Cont’d

Level-2 Model

\[
\begin{align*}
\beta_{0j} &= \gamma_{00} + \gamma_{01} z_j + \delta_{0j} \\
\beta_{1j} &= \gamma_{10} + \gamma_{11} z_j + \delta_{1j} \\
(\delta_{0j} \quad \delta_{1j}) &\sim \mathcal{N}(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{bmatrix}) \\
&\sim \mathcal{N}(0, T)
\end{align*}
\]
A Random Intercept and Slope Model Cont’d

Mixed Model

\[ \logit_{ij} = \gamma_0 + \gamma_{01}Z_j + \gamma_{10}x_{ij} + \gamma_{11}Z_jx_{ij} + \delta_0 + \delta_1x_{ij} \]

or

\[ \pi_{ij} = \frac{\exp(\gamma_0 + \gamma_{01}Z_j + \gamma_{10}x_{ij} + \gamma_{11}Z_jx_{ij} + \delta_0 + \delta_1x_{ij})}{1 + \exp(\gamma_0 + \gamma_{01}Z_j + \gamma_{10}x_{ij} + \gamma_{11}Z_jx_{ij} + \delta_0 + \delta_1x_{ij})} \]
A Random Intercept and Sope Model Cont’d

Alternative Formulation

\[ y_{ij}^* = \beta_0 + \beta_1 x_{ij} + \epsilon_{ij} \]
\[ \beta_0 = \gamma_{00} + \gamma_{01} z_j + \delta_0 \]
\[ \beta_1 = \gamma_{10} + \gamma_{11} z_j + \delta_1 \]
\[ y_{ij}^* = \gamma_{00} + \gamma_{01} z_j + \gamma_{10} x_{ij} + \gamma_{11} z_j x_{ij} + \delta_0 + \delta_1 x_{ij} + \epsilon_{ij} \]
\[ \epsilon_{ij} \sim \mathcal{L}(0, 1) \]
\[ \delta_{0j}, \delta_{1j} \sim \mathcal{N}(0, \mathbf{T}) \]
\[ \pi_{ij} = F (\gamma_{00} + \gamma_{01} z_j + \gamma_{10} x_{ij} + \gamma_{11} z_j x_{ij} + \delta_0 + \delta_1 x_{ij}) \]
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 $x_{ij}$. Then, the level-1 model may be written as

$$\text{logit}_{ij} = x_{ij}^T \beta_j$$

where $\beta_j$ is a $(P + 1) \times 1$ column vector of random coefficients.

Level-2 Model

$$\beta_j = Z_j \gamma + \delta_j$$
$$\delta_j \sim \mathcal{N}(0, T)$$
A General Model Cont’d

Mixed Model in Logit Form

\[ \text{logit}_{ij} = x_{ij}^T Z_j \gamma + x_{ij}^T \delta_j \]

Mixed Model in Probability Form

\[ \pi_{ij} = \frac{\exp(x_{ij}^T Z_j \gamma + x_{ij}^T \delta_j)}{1 + \exp(x_{ij}^T Z_j \gamma + x_{ij}^T \delta_j)} \]
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 $\delta_j$ and that these are the only source of dependencies in the data. Then,

$$L(y, \gamma | \delta) = \prod_{i} \prod_{j} \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{1-y_{ij}}$$
The Likelihood Function Cont’d

The Unconditional Log-Likelihood

Of course, the elements of $\delta_j$ are unknown, which means they have to be integrated out of the likelihood function:

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

where $\Phi_{P+1}$ is the $P + 1$-variate normal distribution over the level-2 errors with covariance matrix $T$ 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

Adaptive Quadrature

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,
  \[ gllamm < xtmelogit < xtlogit \]

- In terms of versatility,
  \[ xtlogit < xtmelogit < gllamm \]
**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:

```plaintext
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

[Diagram showing empirical Bayes residuals]
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 \( \text{logit}_{ij} = \gamma_{00} + \gamma_{01}z_j + \gamma_{10}x_{ij} + \delta_{0j} + \delta_{1j}x_{ij} \).

– 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,

\[
\exp(\gamma_{01}) \quad \text{factor change due to } Z \\
\exp(\gamma_{10}) \quad \text{factor change due to } X
\]
Example: The Labour Vote in 2001

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Estimate</th>
<th>Factor Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labour Identifier</td>
<td>4.49</td>
<td>89.17</td>
</tr>
<tr>
<td>Ideological Distance</td>
<td>-0.23</td>
<td>0.80</td>
</tr>
<tr>
<td>Middle Class</td>
<td>-0.43</td>
<td>0.65</td>
</tr>
<tr>
<td>Working Class</td>
<td>0.21</td>
<td>1.23</td>
</tr>
</tbody>
</table>

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

\[ \hat{\gamma}_{pq} \pm z_{.975} \hat{se}_{\hat{\gamma}_{pq}} \]

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

  Lower Bound \[ \exp \left( \hat{\gamma}_{pq} - z_{.975} \hat{se}_{\hat{\gamma}_{pq}} \right) \]

  Upper Bound \[ \exp \left( \hat{\gamma}_{pq} + z_{.975} \hat{se}_{\hat{\gamma}_{pq}} \right) \]
# Example: The Labour Vote in 2001

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Estimate</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labour Identifier</td>
<td>89.17</td>
<td>60.47</td>
</tr>
<tr>
<td>Ideological Distance</td>
<td>0.80</td>
<td>0.72</td>
</tr>
<tr>
<td>Middle Class</td>
<td>0.65</td>
<td>0.41</td>
</tr>
<tr>
<td>Working Class</td>
<td>1.23</td>
<td>0.83</td>
</tr>
</tbody>
</table>

**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:

\[
\text{logit}_{ij} = \gamma_{00} + \gamma_{01}z_j + \gamma_{10}x_{ij} + \gamma_{11}z_jx_{ij} + \delta_0 + \delta_1jx_{ij}
\]

- 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:

\[
\frac{\partial E[\text{logit}]}{\partial x} = \gamma_{10} + \gamma_{11}z
\]
\[
\frac{\partial E[\text{logit}]}{\partial z} = \gamma_{01} + \gamma_{11}x
\]
Example: The Labour Vote in 2001

- Consider the following model

\[
\text{logit}_{ij} = \beta_0 + \beta_1 \text{labid}_{ij} + \beta_2 \text{dist}_{ij} + \beta_3 \text{middle}_{ij} + \beta_4 \text{working}_{ij}
\]

\[
\beta_0 = \gamma_{00} + \gamma_{01} \text{lab}_{01} + \delta_0 \\
\beta_1 = \gamma_{10} + \gamma_{11} \text{lab}_{01} + \delta_1 \\
\beta_2 = \gamma_{20} \\
\beta_3 = \gamma_{30} \\
\beta_4 = \gamma_{40}
\]

- Or

\[
\text{logit}_{ij} = \gamma_{00} + \gamma_{01} \text{lab}_{01} + \gamma_{10} \text{labid}_{ij} + \gamma_{11} \text{lab}_{01} \text{labid}_{ij} + \gamma_{20} \text{dist}_{ij} + \gamma_{30} \text{middle}_{ij} + \gamma_{40} \text{working}_{ij} + \delta_0 + \delta_1 \text{labid}_{ij}
\]
## Example Cont’d

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Estimate</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labour Share in District</td>
<td>0.03</td>
<td>0.01</td>
</tr>
<tr>
<td>Labour Identifier</td>
<td>3.46</td>
<td>0.54</td>
</tr>
<tr>
<td>Identifier (\times) District Share</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>Ideological Distance</td>
<td>-0.23</td>
<td>0.05</td>
</tr>
<tr>
<td>Middle Class</td>
<td>-0.47</td>
<td>0.23</td>
</tr>
<tr>
<td>Working Class</td>
<td>0.01</td>
<td>0.20</td>
</tr>
<tr>
<td>Constant</td>
<td>-3.06</td>
<td>0.41</td>
</tr>
<tr>
<td>(\tau_{00})</td>
<td>0.07</td>
<td>0.21</td>
</tr>
<tr>
<td>(\tau_{11})</td>
<td>0.51</td>
<td>0.47</td>
</tr>
<tr>
<td>(\tau_{01})</td>
<td>-0.15</td>
<td>0.27</td>
</tr>
</tbody>
</table>

**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:
  
  \texttt{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 in 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

\[ \Pr(y_{ij} = 1 | x_{ij}, Z_j) = \int_{\delta_j} \Pr(y_{ij} = 1 | x_{ij}, Z_j, \delta_j) \phi(\delta_j | T) d\delta_j \]

Note: Although it is sometimes explained this way, it is not correct to set \( \delta_j = 0 \) and then compute the predicted probability. (The average of the inverse logit \( \neq \) inverse logit of the average.)
gllamm can compute the marginal predicted probabilities using

\texttt{gllapred name, mu marginal}
Example: The Labour Vote in 2001

- Consider the random intercept model

\[ \text{logit}_{ij} = \gamma_0 + \gamma_{10} \text{labid}_{ij} + \gamma_{20} \text{dist}_{ij} + \gamma_{30} \text{middle}_{ij} + \gamma_{40} \text{working}_{ij} + \delta_{0j} \]

- The marginal predicted probabilities are given by integrating over \( \delta_{0j} \).
- 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

\[ Pr(y_{ij} = 1|x_{ij}, Z_j, \delta_j) = \frac{\exp(x_{ij}^T Z_j \gamma + x_{ij}^T \delta_j)}{1 + \exp(x_{ij}^T Z_j \gamma + x_{ij}^T \delta_j)} \]

Stata

\textit{gllapred name, mu}
Example Cont’d

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