Multicategory Choice Models

Brad Jones[1]

[1]Department of Political Science
University of California, Davis

May 12, 2009

Stereotype Model

- The model Anderson (1984) proposed has the apparently desirable feature of being a reduced rank version of the multinomial logit model.

Stereotype Model

- The model Anderson (1984) proposed has the apparently desirable feature of being a reduced rank version of the multinomial logit model.
- If a response variable has J categories, the full rank mnl gives $J-1$ nonredundant logits. So with K covariates, there are $K \times J-1$ regression parameters to interpret. The full rank mnl model is given by

$$\log\left[\frac{P(Y=y_j \mid \mathbf{x})}{P(Y=J \mid \mathbf{x})}\right]=\zeta_{ij}=\theta_j+\mathbf{x}'\beta_{\mathbf{j}}, \quad j=1,2,\ldots j-1, \quad (1)$$

Stereotype Model

- The model Anderson (1984) proposed has the apparently desirable feature of being a reduced rank version of the multinomial logit model.
- If a response variable has J categories, the full rank mnl gives $J-1$ nonredundant logits. So with K covariates, there are $K \times J - 1$ regression parameters to interpret. The full rank mnl model is given by

$$\log\left[\frac{P(Y = y_j \mid \mathbf{x})}{P(Y = J \mid \mathbf{x})}\right] = \zeta_{ij} = \theta_j + \mathbf{x}'\beta_{\mathbf{j}}, \quad j = 1, 2, \ldots j-1, \quad (1)$$

- Or equivalently, $J-1$ probabilities given by

$$P_{ij} = \frac{\exp(\zeta_{ij})}{1 + \sum_{j=1}^{J-1} \exp(\zeta_{ij})}. \quad (2)$$

Stereotype Model

- In contrast, the reduced rank model Anderson (1984) proposed is given by (in log-odds form)

$$\log\left[\frac{P(Y = y_j \mid \mathbf{x})}{P(Y = J \mid \mathbf{x})}\right] = \nu_{ij} = \theta_j + \phi_j(\mathbf{x}'\beta), \quad j = 1, 2, \ldots j-1, \tag{3}$$

Stereotype Model

- In contrast, the reduced rank model Anderson (1984) proposed is given by (in log-odds form)

$$\log\left[\frac{P(Y=y_j \mid \mathbf{x})}{P(Y=J \mid \mathbf{x})}\right] = \nu_{ij} = \theta_j + \phi_j(\mathbf{x}'\beta), \quad j = 1, 2, \ldots j-1, \tag{3}$$

- In probabilities as

$$P_{ij} = \frac{\exp(\nu_{ij})}{1 + \sum_{j=1}^{J-1} \exp(\nu_{ij})}. \tag{4}$$

Stereotype Model

- In contrast, the reduced rank model Anderson (1984) proposed is given by (in log-odds form)

$$\log\left[\frac{P(Y=y_j \mid \mathbf{x})}{P(Y=J \mid \mathbf{x})}\right] = \nu_{ij} = \theta_j + \phi_j(\mathbf{x}'\beta), \quad j = 1, 2, \ldots j-1, \tag{3}$$

- In probabilities as

$$P_{ij} = \frac{\exp(\nu_{ij})}{1 + \sum_{j=1}^{J-1}\exp(\nu_{ij})}. \tag{4}$$

- Equations (1) and (3) differ in one important respect. The ϕ parameters serve to rescale the regression parameters up or down in the rr-mnl model, while in the full rank model, each β can assume a unique value for each of the $J-1$ nonredundant logits.

Stereotype Model

- The interpretation of the regression parameters, β, are explicitly connected to the ϕ.

Stereotype Model

- The interpretation of the regression parameters, β, are explicitly connected to the ϕ.
- The log-odds ratios in equation (3) are given by $\phi_j \beta$.

Stereotype Model

- The interpretation of the regression parameters, β, are explicitly connected to the ϕ.
- The log-odds ratios in equation (3) are given by $\phi_j\beta$.
- Because of its connection to the mnl model, the log-odds ratios are referenced against an arbitrary baseline category.

Stereotype Model

- The interpretation of the regression parameters, β, are explicitly connected to the ϕ.
- The log-odds ratios in equation (3) are given by $\phi_j\beta$.
- Because of its connection to the mnl model, the log-odds ratios are referenced against an arbitrary baseline category.
- The log-odds ratio for $Y = j$ are referenced to the fixed baseline category $Y = J$. In general, the odds between any response categories can be obtained by (shown in [**?**]).

$$\log\left(\frac{P_{ij}}{P_{ik}}\right) = \theta_j - \theta_k + (\phi_j - \phi_k)x_i'\beta.$$

Stereotype Logit

- Since the ϕ are almost always unknown, they must be estimated.

Stereotype Logit

- Since the ϕ are almost always unknown, they must be estimated.
- A corner constraint is commonly imposed on the ϕ for identification. For a J-category scale, then, $\phi_1 = 1$ and $\phi_J = 0$; the intermediary ϕ are estimated from the data.

Stereotype Logit

- Since the ϕ are almost always unknown, they must be estimated.
- A corner constraint is commonly imposed on the ϕ for identification. For a J-category scale, then, $\phi_1 = 1$ and $\phi_J = 0$; the intermediary ϕ are estimated from the data.
- Most statistical software impose this identification constraint. For a three-point scale, there is one free ϕ to estimate; for a seven-point scale, there are five free ϕ.

Stereotype Logit

- Since the ϕ are almost always unknown, they must be estimated.
- A corner constraint is commonly imposed on the ϕ for identification. For a J-category scale, then, $\phi_1 = 1$ and $\phi_J = 0$; the intermediary ϕ are estimated from the data.
- Most statistical software impose this identification constraint. For a three-point scale, there is one free ϕ to estimate; for a seven-point scale, there are five free ϕ.
- Yee and Hastie (2003) note that the vector of regression parameters can be expressed as

$$\mathbf{B} = [\phi_1(\beta), \phi_2(\beta), \ldots, \phi_J(\beta)] = \beta \times (1, \phi_2, \ldots, 0). \quad (5)$$

- Anderson (1984) argued that the ordinality condition could be evaluated subject to the condition that $\phi_1 = 1 > \phi_2 > \ldots \phi_J = 0$, noting again that the bottom and top scale scores have imposed ϕ of 1 and 0 respectively.

- Anderson (1984) argued that the ordinality condition could be evaluated subject to the condition that $\phi_1 = 1 > \phi_2 > \ldots \phi_J = 0$, noting again that the bottom and top scale scores have imposed ϕ of 1 and 0 respectively.
- This gives the interpretation that the probability of the baseline category is

$$P_{iJ} = \frac{1}{1 + \sum_{j=1}^{J-1} \exp(\nu_{ij})},$$

and the remaining probabilities are given by equation (4).

- Anderson (1984) argued that the ordinality condition could be evaluated subject to the condition that $\phi_1 = 1 > \phi_2 > \ldots \phi_J = 0$, noting again that the bottom and top scale scores have imposed ϕ of 1 and 0 respectively.
- This gives the interpretation that the probability of the baseline category is

$$P_{iJ} = \frac{1}{1 + \sum_{j=1}^{J-1} \exp(\nu_{ij})},$$

  and the remaining probabilities are given by equation (4).
- To evaluate distinguishability, standard errors for the ϕ can be inspected. Overlapping confidence intervals around the ϕ may give evidence that some scale categories are not distinguishable, given the model that is posited.

Stereotype Model

- The stereotype model can be built up in rank. Subject to the number of regressors, the dimensionality of the model can span the range $J - 1$ (full rank mnl) to $J - (J - 1)$ (rr-mnl given by equation [3]).

Stereotype Model

- The stereotype model can be built up in rank. Subject to the number of regressors, the dimensionality of the model can span the range $J-1$ (full rank mnl) to $J-(J-1)$ (rr-mnl given by equation [3]).
- It is easy to see how this is done by looking at the rank-1 mnl. As the dimensionality increases, the number of parameter vectors also increase (and the ϕ become more constrained as fewer can be uniquely identified).

Stereotype Model

- The stereotype model can be built up in rank. Subject to the number of regressors, the dimensionality of the model can span the range $J-1$ (full rank mnl) to $J-(J-1)$ (rr-mnl given by equation [3]).
- It is easy to see how this is done by looking at the rank-1 mnl. As the dimensionality increases, the number of parameter vectors also increase (and the ϕ become more constrained as fewer can be uniquely identified).
- At full rank $(J-1)$, each β can assume a unique value in each of the $J-1$ logits. In this setting, no ϕ parameters can be estimated. It is also useful to point out that the rr-mnl with a single regressor is simply a reparameterized full-rank mnl. (Why?)

Stereotype Model

- An intermediate model can be obtained as

$$P_{ij} = \frac{\exp(\theta_j + \phi_{j1} \sum_{i=1}^{p} x_i \beta_{i1} + \phi_{j2} \sum_{i=1}^{p} x_i \beta_{i2})}{1 + \sum_{j=1}^{J-1} \exp(\theta_j + \phi_{j1} \sum_{i=1}^{p} x_i \beta_{i1} + \phi_{j2} \sum_{i=1}^{p} x_i \beta_{i2})}. \quad (6)$$

Stereotype Model

- An intermediate model can be obtained as

$$P_{ij} = \frac{\exp(\theta_j + \phi_{j1}\sum_{i=1}^{p} x_i\beta_{i1} + \phi_{j2}\sum_{i=1}^{p} x_i\beta_{i2})}{1 + \sum_{j=1}^{J-1}\exp(\theta_j + \phi_{j1}\sum_{i=1}^{p} x_i\beta_{i1} + \phi_{j2}\sum_{i=1}^{p} x_i\beta_{i2})}. \quad (6)$$

- Note what is going on here.

Stereotype Model

- An intermediate model can be obtained as

$$P_{ij} = \frac{\exp(\theta_j + \phi_{j1} \sum_{i=1}^{p} x_i \beta_{i1} + \phi_{j2} \sum_{i=1}^{p} x_i \beta_{i2})}{1 + \sum_{j=1}^{J-1} \exp(\theta_j + \phi_{j1} \sum_{i=1}^{p} x_i \beta_{i1} + \phi_{j2} \sum_{i=1}^{p} x_i \beta_{i2})}. \quad (6)$$

- Note what is going on here.
- Under this model, we have a two-dimensional stereotype model having the property that some covariates only exert an effect over one range of the scale while other covariates exert an effect over a different range of the scale.

Stereotype Model

- An intermediate model can be obtained as

$$P_{ij} = \frac{\exp(\theta_j + \phi_{j1}\sum_{i=1}^{p} x_i\beta_{i1} + \phi_{j2}\sum_{i=1}^{p} x_i\beta_{i2})}{1 + \sum_{j=1}^{J-1}\exp(\theta_j + \phi_{j1}\sum_{i=1}^{p} x_i\beta_{i1} + \phi_{j2}\sum_{i=1}^{p} x_i\beta_{i2})}. \quad (6)$$

- Note what is going on here.
- Under this model, we have a two-dimensional stereotype model having the property that some covariates only exert an effect over one range of the scale while other covariates exert an effect over a different range of the scale.
- Contrast this model with the full-rank MNL model. Under this model, each covariate exerts a unique effect on the scale score (relative to an arbitrary baseline).

Stereotype Model

- An intermediate model can be obtained as

$$P_{ij} = \frac{\exp(\theta_j + \phi_{j1} \sum_{i=1}^{p} x_i \beta_{i1} + \phi_{j2} \sum_{i=1}^{p} x_i \beta_{i2})}{1 + \sum_{j=1}^{J-1} \exp(\theta_j + \phi_{j1} \sum_{i=1}^{p} x_i \beta_{i1} + \phi_{j2} \sum_{i=1}^{p} x_i \beta_{i2})}. \tag{6}$$

- Note what is going on here.
- Under this model, we have a two-dimensional stereotype model having the property that some covariates only exert an effect over one range of the scale while other covariates exert an effect over a different range of the scale.
- Contrast this model with the full-rank MNL model. Under this model, each covariate exerts a unique effect on the scale score (relative to an arbitrary baseline).
- I can think of no political science analyses using this model.

Stereotype Model: Estimation

- Estimation of model (3) entails estimating the β, the ϕ and the θ parameters.

Stereotype Model: Estimation

- Estimation of model (3) entails estimating the β, the ϕ and the θ parameters.
- Since the β are a multiplicative function of the ϕ, estimation of the stereotype model is nontrivial.

Stereotype Model: Estimation

- Estimation of model (3) entails estimating the β, the ϕ and the θ parameters.
- Since the β are a multiplicative function of the ϕ, estimation of the stereotype model is nontrivial.
- Anderson (1984) proposed to directly estimate the likelihood function, given by

$$L(x_k, \theta_j, \phi_j, \beta) = \prod_{i=1}^{N} \prod_{j=1}^{J} P_j(x_i)^{I(Y=j)}$$

where I is an index function.

Stereotype Model: Estimation

- Kuss (2006) proposed the use of GLS estimation for the stereotype model. This approach proceeds by treating Y as a J-dimensional response vector,

$$\begin{aligned} Y &= 1 \iff Y^* = (1, 0, \ldots, 0)' \\ Y &= 2 \iff Y^* = (0, 1, \ldots, 0)' \\ &\vdots \\ Y &= J \iff Y^* = (0, 0, \ldots, 1)'. \end{aligned}$$

Stereotype Model: Estimation

- Kuss (2006) proposed the use of GLS estimation for the stereotype model. This approach proceeds by treating Y as a J-dimensional response vector,

$$\begin{aligned} Y &= 1 \iff Y^* = (1, 0, \ldots, 0)' \\ Y &= 2 \iff Y^* = (0, 1, \ldots, 0)' \\ &\vdots \\ Y &= J \iff Y^* = (0, 0, \ldots, 1)'. \end{aligned}$$

- The stereotype model is written as a nonlinear model

$$y_i^* = f(x_i, \theta_j, \phi_j, \beta) + \epsilon_i,$$

where $E(\epsilon_i) = 0$ and $E(\epsilon\epsilon') = \Sigma$.

Stereotype Model: Estimation

- Kuss (2006) proposed the use of GLS estimation for the stereotype model. This approach proceeds by treating Y as a J-dimensional response vector,

$$\begin{aligned} Y &= 1 \iff Y^* = (1, 0, \ldots, 0)' \\ Y &= 2 \iff Y^* = (0, 1, \ldots, 0)' \\ &\vdots \\ Y &= J \iff Y^* = (0, 0, \ldots, 1)'. \end{aligned}$$

- The stereotype model is written as a nonlinear model

$$y_i^* = f(x_i, \theta_j, \phi_j, \beta) + \epsilon_i,$$

where $E(\epsilon_i) = 0$ and $E(\epsilon\epsilon') = \Sigma$.

- Corner constraints are imposed, $\phi_1 = 0$ and $\phi_J = 1$, and $\theta_j = 0$ (thus assigning the baseline category to J).

Stereotype Model: Estimation

- Let f have logistic dependency and then GLS can then be applied to minimize the sum of the squared error, given by

$$SSE(\theta_j, \phi_j, \beta, \Sigma) = \sum_{i=1}^{N} [y_i^* - f(x_i, \theta_J, \phi_j, \beta)]' \Sigma^{-1} [y_i^* - f(x_i, \theta_j, \phi_j, \beta)].$$

Stereotype Model: Estimation

- Let f have logistic dependency and then GLS can then be applied to minimize the sum of the squared error, given by

$$SSE(\theta_j, \phi_j, \beta, \Sigma) = \sum_{i=1}^{N} [y_i^* - f(x_i, \theta_J, \phi_j, \beta)]' \Sigma^{-1} [y_i^* - f(x_i, \theta_j, \phi_j, \beta)].$$

- An alternative approach to estimating the models was originally proposed by Greenland (1994).

Stereotype Model: Estimation

- Let f have logistic dependency and then GLS can then be applied to minimize the sum of the squared error, given by

$$SSE(\theta_j, \phi_j, \beta, \Sigma) = \sum_{i=1}^{N} [y_i^* - f(x_i, \theta_J, \phi_j, \beta)]' \Sigma^{-1} [y_i^* - f(x_i, \theta_j, \phi_j, \beta)].$$

- An alternative approach to estimating the models was originally proposed by Greenland (1994).
- Greenland (1994) proposed the use of an alternating algorithm for estimation of β and ϕ.

Stereotype Model: Estimation

- Let f have logistic dependency and then GLS can then be applied to minimize the sum of the squared error, given by

$$SSE(\theta_j, \phi_j, \beta, \Sigma) = \sum_{i=1}^{N} [y_i^* - f(x_i, \theta_J, \phi_j, \beta)]' \Sigma^{-1} [y_i^* - f(x_i, \theta_j, \phi_j, \beta)].$$

- An alternative approach to estimating the models was originally proposed by Greenland (1994).
- Greenland (1994) proposed the use of an alternating algorithm for estimation of β and ϕ.
- This algorithm the criss-cross or zigzag method and starts by setting the β to some start-value and estimating ϕ.

Stereotype Model: Estimation

- Let f have logistic dependency and then GLS can then be applied to minimize the sum of the squared error, given by

$$SSE(\theta_j, \phi_j, \beta, \Sigma) = \sum_{i=1}^{N} [y_i^* - f(x_i, \theta_J, \phi_j, \beta)]' \Sigma^{-1} [y_i^* - f(x_i, \theta_j, \phi_j, \beta)].$$

- An alternative approach to estimating the models was originally proposed by Greenland (1994).
- Greenland (1994) proposed the use of an alternating algorithm for estimation of β and ϕ.
- This algorithm the criss-cross or zigzag method and starts by setting the β to some start-value and estimating ϕ.
- The resultant $\hat{\phi}$ are then fixed and the $\hat{\beta}$ are reestimated, fixed, and the $\hat{\phi}$ are updated.

Stereotype Model: Estimation

- This alternating procedure (hence the name) may be carried out by a series of generalized linear model estimates (Lall et. al. 2002).

Stereotype Model: Estimation

- This alternating procedure (hence the name) may be carried out by a series of generalized linear model estimates (Lall et. al. 2002).
- Kuss (2005) provides some simulation results indicating that each of these methods, MLE, GLS, and the alternating method, performed reasonably well under a variety of criterion.

Stereotype Model: Estimation

- This alternating procedure (hence the name) may be carried out by a series of generalized linear model estimates (Lall et. al. 2002).
- Kuss (2005) provides some simulation results indicating that each of these methods, MLE, GLS, and the alternating method, performed reasonably well under a variety of criterion.
- Illustration: 2004 National Election Studies (NES) to examine white, non-Latino evaluations of the effect of immigration on job availability.

Stereotype Model: Estimation

- This alternating procedure (hence the name) may be carried out by a series of generalized linear model estimates (Lall et. al. 2002).
- Kuss (2005) provides some simulation results indicating that each of these methods, MLE, GLS, and the alternating method, performed reasonably well under a variety of criterion.
- Illustration: 2004 National Election Studies (NES) to examine white, non-Latino evaluations of the effect of immigration on job availability.
- The dependent variable is constructed from a question about whether it is extremely likely ($Y = 1$), very likely ($Y = 2$), somewhat likely ($Y = 3$), or not likely at all ($Y = 4$) that current immigration levels will take jobs away from people already in the United States.

Table: RR-MNL for Immigration Attitudes

Variable	Coefficient
Moral Traditionalism	3.946
	(0.902)
Group Difference	3.469
	(0.779)
Hispanic Traits	2.379
	(1.153)
Ideology	-0.452
	(0.349)
Economic Evaluation	0.967
	(0.513)
θ_1	-4.285
	(0.888)
θ_2	-2.785
	(0.804)
θ_3	-0.607
	(0.645)
ϕ_1	1
ϕ_2	0.749
	(0.109)
ϕ_3	0.392
	(0.097)
ϕ_4	0
n	774
log-likelihood	-969.5284

Stereotype Model: Illustration

- For purposes of identification, category $J = 4$ ("not very likely") is treated as the baseline category and so the corner constraints are $\phi_1 = 1$ and $\phi_4 = 0$.

Stereotype Model: Illustration

- For purposes of identification, category $J = 4$ ("not very likely") is treated as the baseline category and so the corner constraints are $\phi_1 = 1$ and $\phi_4 = 0$.
- The ϕ_j parameters correspond to the score parameters; the θ_j correspond to the intercepts for the $J - 1$ categories.

Stereotype Model: Illustration

- For purposes of identification, category $J = 4$ ("not very likely") is treated as the baseline category and so the corner constraints are $\phi_1 = 1$ and $\phi_4 = 0$.
- The ϕ_j parameters correspond to the score parameters; the θ_j correspond to the intercepts for the $J - 1$ categories.
- To interpret the model, first consider the ϕ_j. The estimates suggest the response variable is ordinal, satisfying the condition $\phi_1 = 1 > \phi_2 > \phi_3 > \phi_4 = 0$.

Stereotype Model: Illustration

- For purposes of identification, category $J = 4$ ("not very likely") is treated as the baseline category and so the corner constraints are $\phi_1 = 1$ and $\phi_4 = 0$.
- The ϕ_j parameters correspond to the score parameters; the θ_j correspond to the intercepts for the $J - 1$ categories.
- To interpret the model, first consider the ϕ_j. The estimates suggest the response variable is ordinal, satisfying the condition $\phi_1 = 1 > \phi_2 > \phi_3 > \phi_4 = 0$.
- The standard errors on the estimated ϕ are small.

Stereotype Model: Distinguishability

- The confidence interval for ϕ_2 does not span 1 giving some evidence of distinguishability between $Y = 1$ and $Y = 2$.

Stereotype Model: Distinguishability

- The confidence interval for ϕ_2 does not span 1 giving some evidence of distinguishability between $Y = 1$ and $Y = 2$.
- Some possible questions emerge about "strict" distinguishability between categories 2 and 3. The lower tail of the 95 percent confidence interval for ϕ_2 overlaps, slightly, with the lower tail for ϕ_3.

Stereotype Model: Distinguishability

- The confidence interval for ϕ_2 does not span 1 giving some evidence of distinguishability between $Y = 1$ and $Y = 2$.
- Some possible questions emerge about "strict" distinguishability between categories 2 and 3. The lower tail of the 95 percent confidence interval for ϕ_2 overlaps, slightly, with the lower tail for ϕ_3.
- The confidence interval for ϕ_3 does not contain 0, suggesting distinguishability between $Y = 3$ and $Y = 4$. The 90 percent confidence intervals suggest strict distinguishability holds.

Stereotype Model: Distinguishability

- A Wald χ^2 test of distinguishability between the ϕ suggest that each category is distinguishable. Under the null, the hypothesis is $\phi_j = \phi_k$. The p-values on these tests are:

$$\begin{aligned} \phi_1 &= \phi_2 : \quad p = .021 \\ \phi_2 &= \phi_3 : \quad p = .001 \\ \phi_3 &= \phi_4 : \quad p = .000 \end{aligned} \tag{7}$$

Stereotype Model: Distinguishability

- A Wald χ^2 test of distinguishability between the ϕ suggest that each category is distinguishable. Under the null, the hypothesis is $\phi_j = \phi_k$. The *p*-values on these tests are:

$$\begin{aligned} \phi_1 &= \phi_2 : \quad p = .021 \\ \phi_2 &= \phi_3 : \quad p = .001 \\ \phi_3 &= \phi_4 : \quad p = .000 \end{aligned} \tag{7}$$

- The ϕ_j give information about the "distance" between categories.

Stereotype Model: Distinguishability

- A Wald χ^2 test of distinguishability between the ϕ suggest that each category is distinguishable. Under the null, the hypothesis is $\phi_j = \phi_k$. The p-values on these tests are:

$$\begin{aligned} \phi_1 &= \phi_2 : \quad p = .021 \\ \phi_2 &= \phi_3 : \quad p = .001 \\ \phi_3 &= \phi_4 : \quad p = .000 \end{aligned} \tag{7}$$

- The ϕ_j give information about the "distance" between categories.
- The difference between the baseline category ($J = 4$) and category $J = 3$ ("somewhat likely") is largest ($\phi_4 = .39$).

Stereotype Model: Distinguishability

- A Wald χ^2 test of distinguishability between the ϕ suggest that each category is distinguishable. Under the null, the hypothesis is $\phi_j = \phi_k$. The p-values on these tests are:

$$\begin{aligned} \phi_1 &= \phi_2: \quad p = .021 \\ \phi_2 &= \phi_3: \quad p = .001 \\ \phi_3 &= \phi_4: \quad p = .000 \end{aligned} \tag{7}$$

- The ϕ_j give information about the "distance" between categories.
- The difference between the baseline category ($J = 4$) and category $J = 3$ ("somewhat likely") is largest ($\phi_4 = .39$).
- The distance between scale score $J = 1$ and $J = 2$ is the smallest ($\phi_2 = .25$).

Stereotype Model: Distinguishability

- Given the construction of the scale, this result might be expected: category four is distinct from the remaining three categories as it is the only response option allowing respondents to claim immigrants will *not* likely take away jobs.

Stereotype Model: Distinguishability

- Given the construction of the scale, this result might be expected: category four is distinct from the remaining three categories as it is the only response option allowing respondents to claim immigrants will *not* likely take away jobs.
- Since the other categories record negative views toward immigrants and the chances they will take away jobs, the difference between the ϕ_j are smallest for them.

Stereotype Model: Distinguishability

- Given the construction of the scale, this result might be expected: category four is distinct from the remaining three categories as it is the only response option allowing respondents to claim immigrants will *not* likely take away jobs.
- Since the other categories record negative views toward immigrants and the chances they will take away jobs, the difference between the ϕ_j are smallest for them.
- That is, the scale is more compressed around $Y = 1, 2, 3$ and more spread above $Y = 3$. Suppose instead of this being a survey item on immigration attitudes, it measured party placement.

Stereotype Model: Distinguishability

- Given the construction of the scale, this result might be expected: category four is distinct from the remaining three categories as it is the only response option allowing respondents to claim immigrants will *not* likely take away jobs.
- Since the other categories record negative views toward immigrants and the chances they will take away jobs, the difference between the ϕ_j are smallest for them.
- That is, the scale is more compressed around $Y = 1, 2, 3$ and more spread above $Y = 3$. Suppose instead of this being a survey item on immigration attitudes, it measured party placement.
- The ϕ may be useful in making claims about the relative distance among parties in a multiparty system.

Stereotype Model: Distinguishability

- With respect to the regression parameters, this model is rank-1 so interpretation of them follows the result shown in (5). Specifically, the probabilities are obtained by equation (4),

$$
\begin{aligned}
P_{i1} &= \frac{\exp(\nu_1)}{1+\exp(\nu_1)+\exp(\nu_2)+\exp(\nu_3)} \\
P_{i2} &= \frac{\exp(\nu_2)}{1+\exp(\nu_1)+\exp(\nu_2)+\exp(\nu_3)} \\
P_{i3} &= \frac{\exp(\nu_3)}{1+\exp(\nu_1)+\exp(\nu_2)+\exp(\nu_3)} \\
P_{i4} &= \frac{1}{1+\exp(\nu_1)+\exp(\nu_2)+\exp(\nu_3)}.
\end{aligned}
$$

Stereotype Model: Distinguishability

- With respect to the regression parameters, this model is rank-1 so interpretation of them follows the result shown in (5). Specifically, the probabilities are obtained by equation (4),

$$
\begin{aligned}
P_{i1} &= \frac{\exp(\nu_1)}{1+\exp(\nu_1)+\exp(\nu_2)+\exp(\nu_3)} \\
P_{i2} &= \frac{\exp(\nu_2)}{1+\exp(\nu_1)+\exp(\nu_2)+\exp(\nu_3)} \\
P_{i3} &= \frac{\exp(\nu_3)}{1+\exp(\nu_1)+\exp(\nu_2)+\exp(\nu_3)} \\
P_{i4} &= \frac{1}{1+\exp(\nu_1)+\exp(\nu_2)+\exp(\nu_3)}.
\end{aligned}
$$

- Note the import of the ϕ (which is contained in ν). The ϕ give the separation for the β over the scale points.