Political Methodology III: Model Based Inference

Justin Grimmer

Associate Professor
Department of Political Science
Stanford University

April 19th, 2017

$$
\begin{aligned}
\frac{\partial l(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial \theta_j} &= \frac{1}{f(\boldsymbol{Y}|\boldsymbol{\theta})}\frac{\partial f(\boldsymbol{Y}|\boldsymbol{\theta})}{\partial \theta_j} \\
E\left[\frac{\partial l(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial \theta_j}\right] &= \int \frac{1}{f(\boldsymbol{Y}|\boldsymbol{\theta})}\frac{\partial f(\boldsymbol{Y}|\boldsymbol{\theta})}{\partial \theta_j} f(\boldsymbol{Y}|\boldsymbol{\theta}) d\boldsymbol{Y} \\
&= \int \frac{\partial f(\boldsymbol{Y}|\boldsymbol{\theta})}{\partial \theta_j} d\boldsymbol{Y} \\
&= \frac{\partial}{\partial \theta_j} \int f(\boldsymbol{Y}|\boldsymbol{\theta}) d\boldsymbol{Y} = \frac{\partial}{\partial \theta_j} 1 = 0
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial l(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial \theta_j}\frac{\partial l(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial \theta_k} & = \frac{1}{f(\boldsymbol{Y}|\boldsymbol{\theta})}\frac{\partial f(\boldsymbol{Y}|\boldsymbol{\theta})}{\partial \theta_j}\frac{1}{f(\boldsymbol{Y}|\boldsymbol{\theta})}\frac{\partial f(\boldsymbol{Y}|\boldsymbol{\theta})}{\partial \theta_k} \\
E\left[\frac{\partial l(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial \theta_j}\frac{\partial l(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial \theta_k}\right] & = \int \frac{1}{(f(\boldsymbol{Y}|\boldsymbol{\theta}))^2}\frac{\partial f(\boldsymbol{Y}|\boldsymbol{\theta})}{\partial \theta_k}\frac{\partial f(\boldsymbol{Y}|\boldsymbol{\theta})}{\partial \theta_j} f(\boldsymbol{Y}|\boldsymbol{\theta}) d\boldsymbol{Y} \\
-E\left[\frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{Y})}{\partial \theta_j \partial \theta_k}\right] & = -\int \frac{\frac{\partial^2 f(\boldsymbol{Y}|\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} f(\boldsymbol{Y}|\boldsymbol{\theta}) - \frac{\partial f(\boldsymbol{Y}|\boldsymbol{\theta})}{\partial \theta_j}\frac{\partial f(\boldsymbol{Y}|\boldsymbol{\theta})}{\partial \theta_k}}{f(\boldsymbol{Y}|\boldsymbol{\theta})^2} f(\boldsymbol{Y}|\boldsymbol{\theta}) d\boldsymbol{Y}
\end{aligned}
$$

Model Based Inference

1) Likelihood inference
2) Logit/Probit
   a) DGP
   b) Optimization
   c) Quantities of Interest
   d) Perfect + Near Perfect Separation
3) Ordered Probit
   a) DGP
   b) Optimization
   c) Quantities of Interest

Civil Conflict and Political Instability (h/t: Yamamoto)

Fearon & Laitin (2003):

- Y_i: Civil conflict
- T_i: Political instability
- W_i: Geography (log % mountainous)

Estimated model:

$$\widehat{\Pr(Y_i = 1 \mid T_i, W_i)} = \text{logit}^{-1}\left(-2.84 + 0.91T_i + 0.35W_i\right)$$

Predicted probability:

$$\hat{\pi}(T_i = 1, W_i = 3.10) = 0.299$$

ATE:

$$\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n}\{\hat{\pi}(1, W_i) - \hat{\pi}(0, W_i)\} = 0.127$$

Civil Conflict and Political Instability (h/t: Yamamoto)

Fearon & Laitin (2003):

- Y_i: Civil conflict
- T_i: Political instability
- W_i: Geography (log % mountainous)

Estimated model:

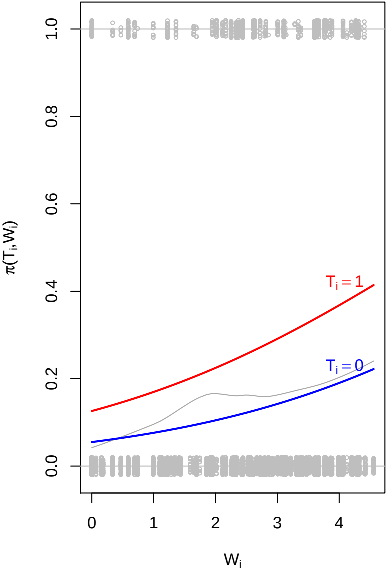
$$\widehat{\Pr(Y_i = 1 \mid T_i, W_i)} = \text{logit}^{-1}(-2.84 + 0.91T_i + 0.35W_i)$$

Predicted probability:

$$\hat{\pi}(T_i = 1, W_i = 3.10) = 0.299$$

ATE:

$$\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n}\{\hat{\pi}(1, W_i) - \hat{\pi}(0, W_i)\} = 0.127$$



Civil Conflict and Political Instability (h/t: Yamamoto)

Fearon & Laitin (2003):

- Y_i: Civil conflict
- T_i: Political instability
- W_i: Geography (log % mountainous)

Estimated model:

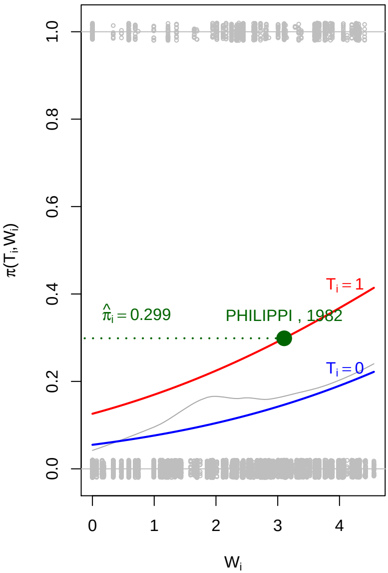
$$\widehat{\Pr(Y_i = 1 \mid T_i, W_i)}$$
$$= \text{logit}^{-1}\left(-2.84 + 0.91 T_i + 0.35 W_i\right)$$

Predicted probability:

$$\hat{\pi}(T_i = 1, W_i = 3.10) \ = \ 0.299$$

ATE:

$$\hat{\tau} \ = \ \frac{1}{n}\sum_{i=1}^{n}\{\hat{\pi}(1, W_i) - \hat{\pi}(0, W_i)\}$$
$$= \ 0.127$$



Civil Conflict and Political Instability (h/t: Yamamoto)

Fearon & Laitin (2003):

- Y_i: Civil conflict
- T_i: Political instability
- W_i: Geography (log % mountainous)

Estimated model:

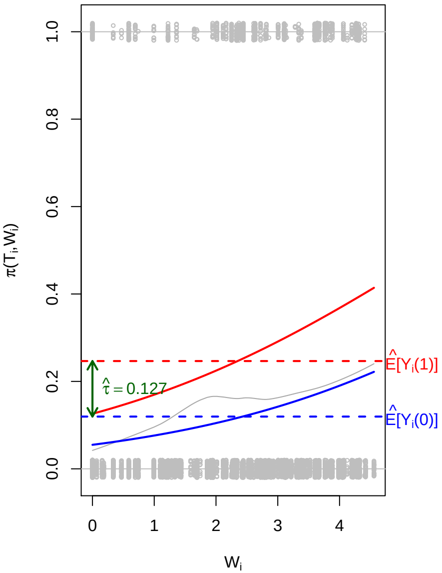
$$\begin{aligned}
&\widehat{\Pr(Y_i = 1 \mid T_i, W_i)} \\
&= \text{logit}^{-1}\left(-2.84 + 0.91 T_i + 0.35 W_i\right)
\end{aligned}$$

Predicted probability:

$$\hat{\pi}(T_i = 1, W_i = 3.10) \ = \ 0.299$$

ATE:

$$\begin{aligned}
\hat{\tau} &= \frac{1}{n}\sum_{i=1}^{n}\{\hat{\pi}(1, W_i) - \hat{\pi}(0, W_i)\} \\
&= 0.127
\end{aligned}$$



Civil Conflict and Political Instability (h/t: Yamamoto)

Fearon & Laitin (2003):

- Y_i: Civil conflict
- T_i: Political instability
- W_i: Geography (log % mountainous)

Estimated model:

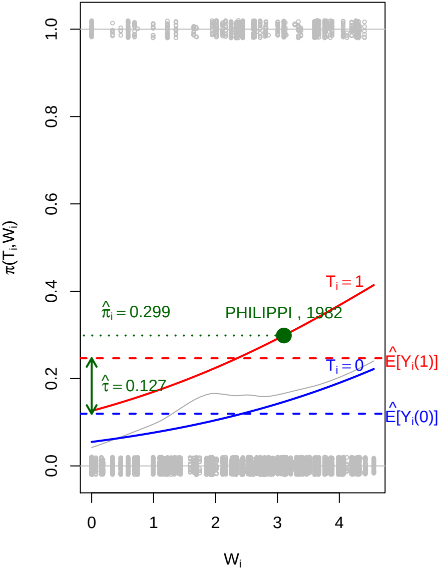
$$\widehat{\Pr(Y_i = 1 \mid T_i, W_i)} = \text{logit}^{-1}(-2.84 + 0.91 T_i + 0.35 W_i)$$

Predicted probability:

$$\hat{\pi}(T_i = 1, W_i = 3.10) = 0.299$$

ATE:

$$\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n}\{\hat{\pi}(1, W_i) - \hat{\pi}(0, W_i)\} = 0.127$$



Final Passage Vote for the Affordable Care Act

$Y_i =$ Vote on ACA

$X_{i1} =$ Democrat

$X_{i2} =$ DW-Nominate Score

Final Passage Vote for the Affordable Care Act

$Y_i =$ Vote on ACA

$X_{i1} =$ Democrat

$X_{i2} =$ DW-Nominate Score

	Nay	Yea
Republican	178	0
Democrat	34	219

Final Passage Vote for the Affordable Care Act

Y_i = Vote on ACA

X_{i1} = Democrat

X_{i2} = DW-Nominate Score

	Nay	Yea
Republican	178	0
Democrat	34	219

$$\Pr(Y_i = 1 | X_{i1} = 0) \quad = \quad 0$$

Final Passage Vote for the Affordable Care Act

```
> test_model<- glm(vote~dem + ideo,
data = as.data.frame(final_votes), family = binomial('logit'))
Warning message:
glm.fit: fitted probabilities numerically 0 or 1 occurred
```

Final Passage Vote for the Affordable Care Act

```
> test_model<- glm(vote~dem + ideo,
data = as.data.frame(final_votes), family = binomial('logit'))
Warning message:
glm.fit: fitted probabilities numerically 0 or 1 occurred
```

	ACA Vote
Intercept	-14.00
	(1670.439)
Democrat	11.67
	(1670.439)
Ideology	-16.86
	(2.71)

Final Passage Vote for the Affordable Care Act

```
> test_model<- glm(vote~dem + ideo,
data = as.data.frame(final_votes), family = binomial('logit'))
Warning message:
glm.fit: fitted probabilities numerically 0 or 1 occurred
```

	ACA Vote
Intercept	-14.00
	(1670.439)
Democrat	11.67
	(1670.439)
Ideology	-16.86
	(2.71)

$$E[Y_i|X_i \widehat{= (1, 0}, 0.7), \boldsymbol{\beta}^*] \quad = \quad 0$$

Final Passage Vote for the Affordable Care Act

```
> test_model<- glm(vote~dem + ideo,
data = as.data.frame(final_votes), family = binomial('logit'))
Warning message:
glm.fit: fitted probabilities numerically 0 or 1 occurred
```

	ACA Vote
Intercept	-14.00
	(1670.439)
Democrat	11.67
	(1670.439)
Ideology	-16.86
	(2.71)

$$E[Y_i|\widehat{X_i = (1,0},0.7),\beta^*] \quad = \quad 0$$

$$CI_{95} \quad = \quad [0,1] \text{ Coefficient Simulation}$$

Final Passage Vote for the Affordable Care Act

```
> test_model<- glm(vote~dem + ideo,
data = as.data.frame(final_votes), family = binomial('logit'))
Warning message:
glm.fit: fitted probabilities numerically 0 or 1 occurred
```

	ACA Vote
Intercept	-14.00
	(1670.439)
Democrat	11.67
	(1670.439)
Ideology	-16.86
	(2.71)

$$
\begin{aligned}
E[Y_i|X_i = \widehat{(1, 0, 0.7), \beta^*}] &= 0 \\
CI_{95} &= [0, 1] \text{ Coefficient Simulation} \\
CI_{95} &= [1e - 13, 2.3e - 10] \text{ Bootstrap}
\end{aligned}
$$

Final Passage Vote for the Affordable Care Act

```
> test_model<- glm(vote~dem + ideo,
data = as.data.frame(final_votes), family = binomial('logit'))
Warning message:
glm.fit: fitted probabilities numerically 0 or 1 occurred
```

	ACA Vote
Intercept	-14.00
	(1670.439)
Democrat	11.67
	(1670.439)
Ideology	-16.86
	(2.71)

$$\begin{aligned} E[Y_i \widehat{|X_i = (1,0}, 0.7), \boldsymbol{\beta}^*] &= 0 \\ CI_{95} &= [0,1] \text{ Coefficient Simulation} \\ CI_{95} &= [1e-13, 2.3e-10] \text{ Bootstrap} \end{aligned}$$

We have problems!

Is There Anything the Tea Party Doesn't Screw Up?

Is There Anything the Tea Party Doesn't Screw Up?

$$L(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{Y}) = f(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\beta})$$

Is There Anything the Tea Party Doesn't Screw Up?

$$
\begin{aligned}
L(\boldsymbol{\beta}|\boldsymbol{X},\boldsymbol{Y}) &= f(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\beta}) \\
&= \sum_{i=1}^{N} Y_i \log\left(\frac{1}{1+\exp(-\boldsymbol{X}'\boldsymbol{\beta})}\right) + (1-Y_i)\log\left(1-\frac{1}{1+\exp(-\boldsymbol{X}'\boldsymbol{\beta})}\right)
\end{aligned}
$$

Is There Anything the Tea Party Doesn't Screw Up?

$$
\begin{aligned}
L(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{Y}) &= f(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\beta}) \\
&= \sum_{i=1}^{N} Y_i \log\left(\frac{1}{1+\exp(-\boldsymbol{X}'\boldsymbol{\beta})}\right) + (1 - Y_i)\log\left(1 - \frac{1}{1+\exp(-\boldsymbol{X}'\boldsymbol{\beta})}\right)
\end{aligned}
$$

Remember:

Is There Anything the Tea Party Doesn't Screw Up?

$$
\begin{aligned}
L(\boldsymbol{\beta}|\boldsymbol{X},\boldsymbol{Y}) &= f(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\beta}) \\
&= \sum_{i=1}^{N} Y_i \log\left(\frac{1}{1+\exp(-\boldsymbol{X}'\boldsymbol{\beta})}\right) + (1-Y_i)\log\left(1 - \frac{1}{1+\exp(-\boldsymbol{X}'\boldsymbol{\beta})}\right)
\end{aligned}
$$

Remember:

$$
\Pr(Y_i = 1|X_{i1} = 0) = 0
$$

Is There Anything the Tea Party Doesn't Screw Up?

$$\begin{aligned} L(\boldsymbol{\beta}|\boldsymbol{X},\boldsymbol{Y}) &= f(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{\beta}) \\ &= \sum_{i=1}^{N} Y_i \log\left(\frac{1}{1+\exp(-\boldsymbol{X}'\boldsymbol{\beta})}\right) + (1-Y_i)\log\left(1-\frac{1}{1+\exp(-\boldsymbol{X}'\boldsymbol{\beta})}\right) \end{aligned}$$

Remember:

$$\Pr(Y_i = 1|X_{i1} = 0) = 0$$

To fit data: set $\beta_0 \rightarrow -\infty$.

Perfect and Near-Perfect Separation

Perfect and Near-Perfect Separation

Perfect separation: one covariate perfectly separates 0's and 1's

Perfect and Near-Perfect Separation

Perfect separation: one covariate perfectly separates 0's and 1's
Near perfect separation: one covariate perfectly separates 0's or 1's

Perfect and Near-Perfect Separation

Perfect separation: one covariate perfectly separates 0's and 1's
Near perfect separation: one covariate perfectly separates 0's or 1's
Solution?:

Perfect and Near-Perfect Separation

Perfect separation: one covariate perfectly separates 0's and 1's
Near perfect separation: one covariate perfectly separates 0's or 1's
Solution?:
You need to make more assumptions

Add a Few Observations...

	Nay	Yea
Republican	178	0
Democrat	34	219

Add a Few Observations...

	Nay	Yea
Republican	178.5	0.5
Democrat	34.5	219.5

Penalized (Prior)-Logistic Regression

- Separation: causes coefficients to diverge
- Penalty (prior): force coefficients towards zero

Step 1: Standardize inputs (Gelman et al)

- Binary variables: mean 0, differ by 1.
  - Democrats: (30%). (0.3, -0.7)
- Other variables: mean 0, sd 0.5.

Penalized (Prior)-Logistic Regression

Step 2: Penalize Likelihood

Penalized (Prior)-Logistic Regression

Step 2: Penalize Likelihood

1) Firth's Penalty (Zorn 2005):

Penalized (Prior)-Logistic Regression

Step 2: Penalize Likelihood

1) Firth's Penalty (Zorn 2005):

$$L(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{Y}) = \prod_{i=1}^{N} \pi_i^{Y_i}(1-\pi_i)^{1-Y_i}|I(\boldsymbol{\beta})|^{1/2}$$

Penalized (Prior)-Logistic Regression

Step 2: Penalize Likelihood

1) Firth's Penalty (Zorn 2005):

$$L(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{Y}) = \prod_{i=1}^{N} \pi_i^{Y_i}(1-\pi_i)^{1-Y_i}|I(\boldsymbol{\beta})|^{1/2}$$

where:

Penalized (Prior)-Logistic Regression

Step 2: Penalize Likelihood

1) Firth's Penalty (Zorn 2005):

$$L(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{Y}) = \prod_{i=1}^{N} \pi_i^{Y_i}(1-\pi_i)^{1-Y_i} |I(\boldsymbol{\beta})|^{1/2}$$

where:

$$\pi_i = \frac{1}{1+\exp(-\boldsymbol{X}_i^{'}\boldsymbol{\beta})}$$

$$|I(\boldsymbol{\beta})| = \text{Determinant of Fisher's information at } \boldsymbol{\beta}$$

$$I(\boldsymbol{\beta}) = \boldsymbol{X}^{'}\boldsymbol{W}\boldsymbol{X}$$

$$\boldsymbol{W} = \begin{pmatrix} \pi_1(1-\pi_1) & 0 & \ldots & 0 \\ 0 & \pi_2(1-\pi_2) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \pi_N(1-\pi_N) \end{pmatrix}$$

Penalized (Prior)-Logistic Regression

$$
\begin{aligned}
L(\boldsymbol{\beta}|\boldsymbol{X},\boldsymbol{Y}) &= \prod_{i=1}^{N} \pi_i^{Y_i}(1-\pi_i)^{1-Y_i}|I(\boldsymbol{\beta})|^{1/2} \\
l(\boldsymbol{\beta}|\boldsymbol{X},\boldsymbol{Y}) &= \sum_{i=1}^{N} Y_i \log \pi_i + (1-Y_i)\log(1-\pi_i) + \frac{1}{2}\log(|I(\boldsymbol{\beta})|)
\end{aligned}
$$

Penalized (Prior)-Logistic Regression

```
jef_pri<- function(params, X, Y){
    beta<- params
    y.tilde<- X%*%beta
    y.prob<- plogis(y.tilde)
    temp<- matrix(0, nrow = length(Y), ncol=length(Y))
    part1<- Y%*%log(y.prob) + (1-Y)%*%log(1- y.prob)
    diag(temp)<- y.prob*(1-y.prob)
    part2<- 0.5*log(det(t(X)%*%temp%*%X))
    out<- part1 + part2
}
firth<- optim(rnorm(3), jef_pri, method = 'BFGS',
control=list(fnscale=-1), hessian=T,
X = cbind(1, dem, ideo), Y =clean[,3] )
```

Comparison

	ACA Vote (GLM)	Firth
Intercept	-14.00	-5.70
	(1670.439)	(24.68)
Democrat	11.67	3.30
	(1670.439)	(42.10)
Ideology	-16.86	-17.72
	(2.71)	(2.85)

Penalized (Prior) Logistic Regression

Penalized (Prior) Logistic Regression

Alternative Prior (Gelman et al 2005):

Penalized (Prior) Logistic Regression

Alternative Prior (Gelman et al 2005):

- Add 1/2 to success and failure: t distribution, 7 degrees of freedom and scale $= 2.5$

Penalized (Prior) Logistic Regression

Alternative Prior (Gelman et al 2005):

- Add 1/2 to success and failure: t distribution, 7 degrees of freedom and scale $= 2.5$
- Suggestion Cauchy (DOF $= 1$) with scale 2.5.

Penalized (Prior) Logistic Regression

Alternative Prior (Gelman et al 2005):

- Add 1/2 to success and failure: t distribution, 7 degrees of freedom and scale $= 2.5$
- Suggestion Cauchy (DOF $= 1$) with scale 2.5.
  - Prioritize $\beta < 0.5$

Penalized (Prior) Logistic Regression

Alternative Prior (Gelman et al 2005):

- Add 1/2 to success and failure: t distribution, 7 degrees of freedom and scale = 2.5
- Suggestion Cauchy (DOF = 1) with scale 2.5.
  - Prioritize $\beta < 0.5$
  - Occasionally allows very large values (Cauchy)

```
lst<- function(x, nu, mu, sigma2){
    part1<- lgamma( (nu + 1)/2)
    part2<- lgamma(nu/2)
    part3<- sqrt(pi *nu*sqrt(sigma2))
    part4<- 1 + (1/nu)*(( (x- mu)^2)/sigma2)
    part4<- ( - (nu + 1)/2)*log(part4)
    out<- part4
    return(out)
}
```

Penalized (Prior) Logistic Regression

Alternative Prior (Gelman et al 2005):

- Add 1/2 to success and failure: t distribution, 7 degrees of freedom and scale = 2.5
- Suggestion Cauchy (DOF = 1) with scale 2.5.
  - Prioritize $\beta < 0.5$
  - Occasionally allows very large values (Cauchy)

```
lst<- function(x, nu, mu, sigma2){
     part1<- lgamma( (nu + 1)/2)
     part2<- lgamma(nu/2)
     part3<- sqrt(pi *nu*sqrt(sigma2))
     part4<- 1 + (1/nu)*(( (x- mu)^2)/sigma2)
     part4<- ( - (nu + 1)/2)*log(part4)
     out<- part4
     return(out)
}
```

```
log_t<- function(params, X, Y, nu, mu, sigma2){
beta<- params
prior<- 0
for(k in 2:ncol(X)){
prior<- prior + lst(beta[k], nu, mu, sigma2)
}
prior<- prior + lst(beta[1], 1, 0, 10)
y.tilde<- X%*%beta
y.prob<- plogis(y.tilde)
out<- Y%*%log(y.prob) + (1- Y)%*%log(1- y.prob)
out<- out + prior
return(out)
}
```

```
log_t<- function(params, X, Y, nu, mu, sigma2){
beta<- params
prior<- 0
for(k in 2:ncol(X)){
prior<- prior + lst(beta[k], nu, mu, sigma2)
}
prior<- prior + lst(beta[1], 1, 0, 10)
y.tilde<- X%*%beta
y.prob<- plogis(y.tilde)
out<- Y%*%log(y.prob) + (1- Y)%*%log(1- y.prob)
out<- out + prior
return(out)
}
cauch<- optim(rnorm(3), log_t, method='BFGS',
control=list(fnscale=-1), hessian=T,
X= cbind(1, dem, ideo), Y = clean[,3], nu = 1,
mu = 0, sigma2=2.5)
```

```
log_t<- function(params, X, Y, nu, mu, sigma2){
beta<- params
prior<- 0
for(k in 2:ncol(X)){
prior<- prior + lst(beta[k], nu, mu, sigma2)
}
prior<- prior + lst(beta[1], 1, 0, 10)
y.tilde<- X%*%beta
y.prob<- plogis(y.tilde)
out<- Y%*%log(y.prob) + (1- Y)%*%log(1- y.prob)
out<- out + prior
return(out)
}
cauch<- optim(rnorm(3), log_t, method='BFGS',
control=list(fnscale=-1), hessian=T,
X= cbind(1, dem, ideo), Y = clean[,3], nu = 1,
mu = 0, sigma2=2.5)
```

In practice: `bayesglm` in `library(arm)` is awesome!

Comparison

	ACA Vote (GLM)	Firth	Cauchy
Intercept	-14.00	-5.70	-3.23
	(1670.439)	(24.68)	(0.99)
Democrat	11.67	3.30	-0.19
	(1670.439)	(42.10)	(1.12)
Ideology	-16.86	-17.72	-16.25
	(2.71)	(2.85)	(2.65)

Ordered Outcome Data

Modeling Ordered Outcomes (ht: Yamamoto)

- Suppose that the J choices are ordered in a substantively meaningful way
- Examples:
  - "Likert scale" in survey questions ("strongly agree", "agree", etc.)
  - Party positions (extreme left, center left, center, right, extreme right)
  - Levels of democracy (autocracy, anocracy, democracy)
  - Health status (healthy, sick, dying, dead)
- Why not use continuous outcome models?
  $\longrightarrow$ Don't want to assume equal distances between levels
- Why not use categorical outcome models? (on Monday)
  $\longrightarrow$ Don't want to waste information about ordering

Modeling Ordered Outcomes (ht: Yamamoto)

- Suppose that the J choices are ordered in a substantively meaningful way
- Examples:
  - "Likert scale" in survey questions ("strongly agree", "agree", etc.)
  - Party positions (extreme left, center left, center, right, extreme right)
  - Levels of democracy (autocracy, anocracy, democracy)
  - Health status (healthy, sick, dying, dead)
- Why not use continuous outcome models?
  $\longrightarrow$ Don't want to assume equal distances between levels
- Why not use categorical outcome models? (on Monday)
  $\longrightarrow$ Don't want to waste information about ordering

Modeling Ordered Outcomes (ht: Yamamoto)

- Suppose that the J choices are ordered in a substantively meaningful way
- Examples:
  - "Likert scale" in survey questions ("strongly agree", "agree", etc.)
  - Party positions (extreme left, center left, center, right, extreme right)
  - Levels of democracy (autocracy, anocracy, democracy)
  - Health status (healthy, sick, dying, dead)
- Why not use continuous outcome models?
  $\longrightarrow$ Don't want to assume equal distances between levels
- Why not use categorical outcome models? (on Monday)
  $\longrightarrow$ Don't want to waste information about ordering

Modeling Ordered Outcomes (ht: Yamamoto)

- Suppose that the J choices are ordered in a substantively meaningful way
- Examples:
  - "Likert scale" in survey questions ("strongly agree", "agree", etc.)
  - Party positions (extreme left, center left, center, right, extreme right)
  - Levels of democracy (autocracy, anocracy, democracy)
  - Health status (healthy, sick, dying, dead)
- Why not use continuous outcome models?
  $\longrightarrow$ Don't want to assume equal distances between levels
- Why not use categorical outcome models? (on Monday)
  $\longrightarrow$ Don't want to waste information about ordering

Modeling Ordered Outcomes (ht: Yamamoto)

- Suppose that the J choices are ordered in a substantively meaningful way
- Examples:
  - "Likert scale" in survey questions ("strongly agree", "agree", etc.)
  - Party positions (extreme left, center left, center, right, extreme right)
  - Levels of democracy (autocracy, anocracy, democracy)
  - Health status (healthy, sick, dying, dead)
- Why not use continuous outcome models?
  $\longrightarrow$ Don't want to assume equal distances between levels
- Why not use categorical outcome models? (on Monday)
  $\longrightarrow$ Don't want to waste information about ordering

Modeling Ordered Outcomes (ht: Yamamoto)

- Suppose that the J choices are ordered in a substantively meaningful way
- Examples:
  - "Likert scale" in survey questions ("strongly agree", "agree", etc.)
  - Party positions (extreme left, center left, center, right, extreme right)
  - Levels of democracy (autocracy, anocracy, democracy)
  - Health status (healthy, sick, dying, dead)
- Why not use continuous outcome models?
  $\longrightarrow$ Don't want to assume equal distances between levels
- Why not use categorical outcome models? (on Monday)
  $\longrightarrow$ Don't want to waste information about ordering

Ordered Logit and Probit Models (ht: Yamamoto)

- Again, the latent variable representation: $Y_i^* = X_i'\beta + \epsilon_i$
- Assume that Y_i^* gives rise to Y_i based on the following scheme:

$$Y_i = \begin{cases} 1 & \text{if} \quad -\infty(=\psi_0) < Y_i^* \leq \psi_1, \\ 2 & \text{if} \quad \psi_1 < Y_i^* \leq \psi_2, \\ \vdots & \qquad \vdots \\ J & \text{if} \quad \psi_{J-1} < Y_i^* \leq \infty(=\psi_J) \end{cases}$$

  where $\psi_1, ..., \psi_{J-1}$ are the threshold parameters to be estimated
- If X_i contains an intercept, one of the ψ's must be fixed for identifiability (typically $\psi_1 = 0$)
- $\epsilon_j \sim_{\text{iid}}$ logistic $\Rightarrow$ the ordered logit model:

$$\Pr(Y_i \leq j \mid X_i) = \frac{\exp(\psi_j - X_i'\beta)}{1 + \exp(\psi_j - X_i^\top \beta)}$$

- $\epsilon_j \sim_{\text{iid}}$ Normal$(0, 1) \Rightarrow$ the ordered probit model:

$$\Pr(Y_i \leq j \mid X_i) = \Phi\left(\psi_j - X_i'\beta\right)$$

Ordered Logit and Probit Models (ht: Yamamoto)

- Again, the latent variable representation: $Y_i^* = X_i'\beta + \epsilon_i$
- Assume that Y_i^* gives rise to Y_i based on the following scheme:

$$Y_i = \begin{cases} 1 & \text{if} \quad -\infty(=\psi_0) < Y_i^* \leq \psi_1, \\ 2 & \text{if} \quad \psi_1 < Y_i^* \leq \psi_2, \\ \vdots & \qquad\qquad \vdots \\ J & \text{if} \quad \psi_{J-1} < Y_i^* \leq \infty(=\psi_J) \end{cases}$$

where $\psi_1, ..., \psi_{J-1}$ are the threshold parameters to be estimated

- If X_i contains an intercept, one of the ψ's must be fixed for identifiability (typically $\psi_1 = 0$)
- $\epsilon_j \sim_{\text{iid}}$ logistic $\Rightarrow$ the ordered logit model:

$$\Pr(Y_i \leq j \mid X_i) = \frac{\exp(\psi_j - X_i'\beta)}{1 + \exp(\psi_j - X_i^\top\beta)}$$

- $\epsilon_j \sim_{\text{iid}}$ Normal$(0, 1) \Rightarrow$ the ordered probit model:

$$\Pr(Y_i \leq j \mid X_i) = \Phi\left(\psi_j - X_i'\beta\right)$$

Ordered Logit and Probit Models (ht: Yamamoto)

- Again, the latent variable representation: $Y_i^* = X_i'\beta + \epsilon_i$
- Assume that Y_i^* gives rise to Y_i based on the following scheme:

$$Y_i = \begin{cases} 1 & \text{if} \quad -\infty(=\psi_0) < Y_i^* \leq \psi_1, \\ 2 & \text{if} \quad \psi_1 < Y_i^* \leq \psi_2, \\ \vdots & \qquad \vdots \\ J & \text{if} \quad \psi_{J-1} < Y_i^* \leq \infty(=\psi_J) \end{cases}$$

where $\psi_1, ..., \psi_{J-1}$ are the threshold parameters to be estimated

- If X_i contains an intercept, one of the ψ's must be fixed for identifiability (typically $\psi_1 = 0$)
- $\epsilon_j \sim_{\text{iid}}$ logistic $\Rightarrow$ the ordered logit model:

$$\Pr(Y_i \leq j \mid X_i) = \frac{\exp(\psi_j - X_i'\beta)}{1 + \exp(\psi_j - X_i^\top\beta)}$$

- $\epsilon_j \sim_{\text{iid}}$ Normal$(0, 1)$ $\Rightarrow$ the ordered probit model:

$$\Pr(Y_i \leq j \mid X_i) = \Phi\left(\psi_j - X_i'\beta\right)$$

Ordered Logit and Probit Models (ht: Yamamoto)

- Again, the latent variable representation: $Y_i^* = X_i'\beta + \epsilon_i$
- Assume that Y_i^* gives rise to Y_i based on the following scheme:

$$Y_i = \begin{cases} 1 & \text{if} \quad -\infty(=\psi_0) < Y_i^* \leq \psi_1, \\ 2 & \text{if} \quad \psi_1 < Y_i^* \leq \psi_2, \\ \vdots & \qquad\qquad \vdots \\ J & \text{if} \quad \psi_{J-1} < Y_i^* \leq \infty(=\psi_J) \end{cases}$$

where $\psi_1, ..., \psi_{J-1}$ are the threshold parameters to be estimated

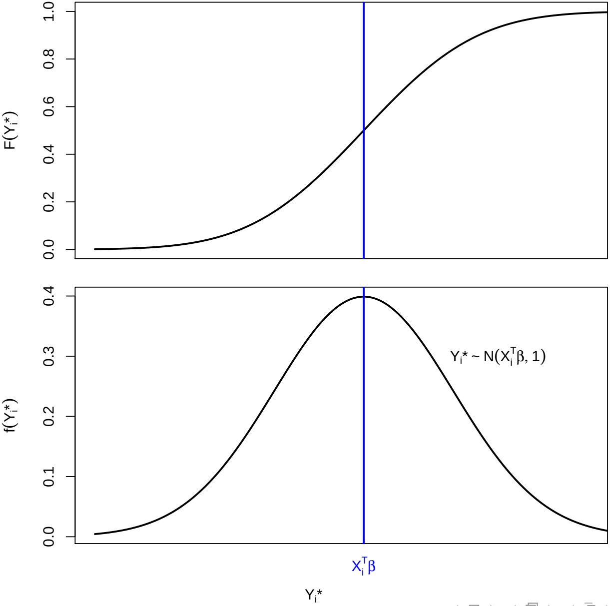
- If X_i contains an intercept, one of the ψ's must be fixed for identifiability (typically $\psi_1 = 0$)
- $\epsilon_j \sim_{\text{iid}}$ logistic $\Rightarrow$ the ordered logit model:

$$\Pr(Y_i \leq j \mid X_i) = \frac{\exp(\psi_j - X_i'\beta)}{1 + \exp(\psi_j - X_i^\top\beta)}$$

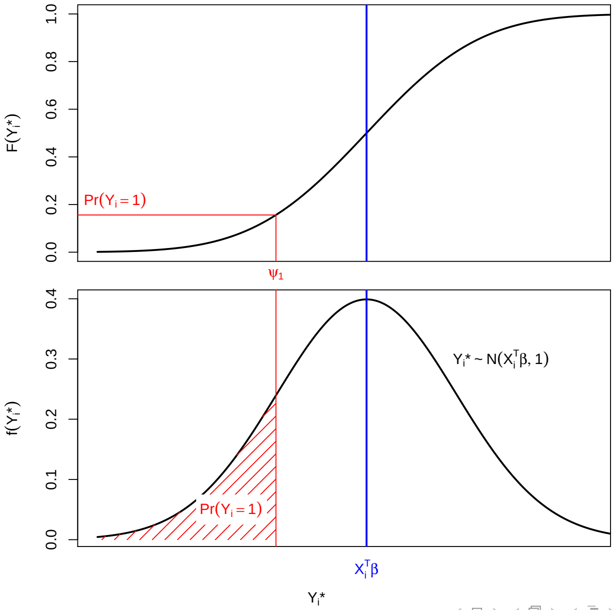
- $\epsilon_j \sim_{\text{iid}}$ Normal$(0,1) \Rightarrow$ the ordered probit model:

$$\Pr(Y_i \leq j \mid X_i) = \Phi\left(\psi_j - X_i'\beta\right)$$

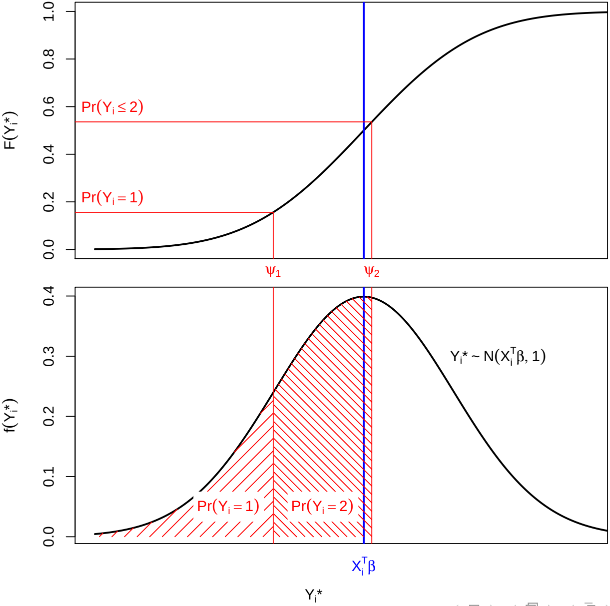
Ordered Logit and Probit Models (ht: Yamamoto)



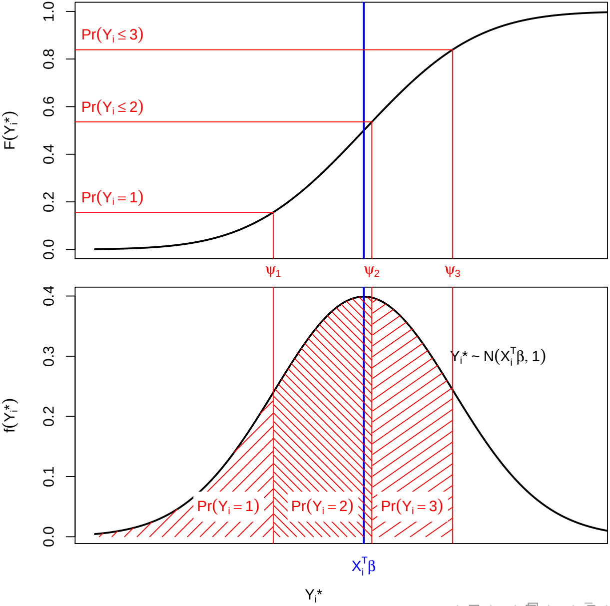
Ordered Logit and Probit Models (ht: Yamamoto)



Ordered Logit and Probit Models (ht: Yamamoto)



Ordered Logit and Probit Models (ht: Yamamoto)



Ordered Logit and Probit

Ordered Logit and Probit

$$
\begin{aligned}
P(Y_i = J) &= \int_{\psi_{j-1}}^{\psi_j} \phi(\tilde{y} | \boldsymbol{X}_i^{'} \boldsymbol{\beta}) d\tilde{y} \\
&= \Phi(\psi_j | \boldsymbol{X}_i^{'} \boldsymbol{\beta}) - \Phi(\psi_{j-1} | \boldsymbol{X}_i^{'} \boldsymbol{\beta})
\end{aligned}
$$

Ordered Logit and Probit

$$
\begin{aligned}
P(Y_i = J) &= \int_{\psi_{j-1}}^{\psi_j} \phi(\tilde{y}|\boldsymbol{X}_i^{'}\boldsymbol{\beta}) d\tilde{y} \\
&= \Phi(\psi_j|\boldsymbol{X}_i^{'}\boldsymbol{\beta}) - \Phi(\psi_{j-1}|\boldsymbol{X}_i^{'}\boldsymbol{\beta})
\end{aligned}
$$

Implies a likelihood of:

Ordered Logit and Probit

$$
\begin{aligned}
P(Y_i = J) &= \int_{\psi_{j-1}}^{\psi_j} \phi(\tilde{y}|\boldsymbol{X}_i^{'}\boldsymbol{\beta}) d\tilde{y} \\
&= \Phi(\psi_j|\boldsymbol{X}_i^{'}\boldsymbol{\beta}) - \Phi(\psi_{j-1}|\boldsymbol{X}_i^{'}\boldsymbol{\beta})
\end{aligned}
$$

Implies a likelihood of:

$$
L(\boldsymbol{\beta}, \boldsymbol{\Psi}|\boldsymbol{X}, \boldsymbol{Y}) = \prod_{i=1}^{N} \left[\prod_{j=1}^{J} [\Phi(\psi_j|\boldsymbol{X}_i^{'}\boldsymbol{\beta}) - \Phi(\psi_{j-1}|\boldsymbol{X}_i^{'}\boldsymbol{\beta})]^{I(Y_i=j)} \right]
$$

Ordered Logit and Probit

$$
\begin{aligned}
P(Y_i = J) &= \int_{\psi_{j-1}}^{\psi_j} \phi(\tilde{y}|\boldsymbol{X}_i^{'}\boldsymbol{\beta}) d\tilde{y} \\
&= \Phi(\psi_j|\boldsymbol{X}_i^{'}\boldsymbol{\beta}) - \Phi(\psi_{j-1}|\boldsymbol{X}_i^{'}\boldsymbol{\beta})
\end{aligned}
$$

Implies a likelihood of:

$$
\begin{aligned}
L(\boldsymbol{\beta}, \boldsymbol{\Psi}|\boldsymbol{X}, \boldsymbol{Y}) &= \prod_{i=1}^{N} \left[\prod_{j=1}^{J} [\Phi(\psi_j|\boldsymbol{X}_i^{'}\boldsymbol{\beta}) - \Phi(\psi_{j-1}|\boldsymbol{X}_i^{'}\boldsymbol{\beta})]^{I(Y_i=j)} \right] \\
l(\boldsymbol{\beta}, \boldsymbol{\Psi}|\boldsymbol{X}, \boldsymbol{Y}) &= \sum_{i=1}^{N} \left(\sum_{j=1}^{J} I(Y_i = j) \log \left[\Phi(\psi_j|\boldsymbol{X}_i^{'}\boldsymbol{\beta}) - \Phi(\psi_{j-1}|\boldsymbol{X}_i^{'}\boldsymbol{\beta}) \right] \right)
\end{aligned}
$$

Ordered Logit and Probit

$$
\begin{aligned}
P(Y_i = J) &= \int_{\psi_{j-1}}^{\psi_j} \phi(\tilde{y}|\boldsymbol{X}_i^{'}\boldsymbol{\beta}) d\tilde{y} \\
&= \Phi(\psi_j|\boldsymbol{X}_i^{'}\boldsymbol{\beta}) - \Phi(\psi_{j-1}|\boldsymbol{X}_i^{'}\boldsymbol{\beta})
\end{aligned}
$$

Implies a likelihood of:

$$
\begin{aligned}
L(\boldsymbol{\beta}, \boldsymbol{\Psi}|\boldsymbol{X}, \boldsymbol{Y}) &= \prod_{i=1}^{N} \left[\prod_{j=1}^{J} [\Phi(\psi_j|\boldsymbol{X}_i^{'}\boldsymbol{\beta}) - \Phi(\psi_{j-1}|\boldsymbol{X}_i^{'}\boldsymbol{\beta})]^{I(Y_i=j)} \right] \\
l(\boldsymbol{\beta}, \boldsymbol{\Psi}|\boldsymbol{X}, \boldsymbol{Y}) &= \sum_{i=1}^{N} \left(\sum_{j=1}^{J} I(Y_i = j) \log \left[\Phi(\psi_j|\boldsymbol{X}_i^{'}\boldsymbol{\beta}) - \Phi(\psi_{j-1}|\boldsymbol{X}_i^{'}\boldsymbol{\beta}) \right] \right)
\end{aligned}
$$

fit with `polr` package

Calculating Quantities of Interest (ht: Yamamoto)

- Predicted probability:

$$
\begin{aligned}
\pi_{ij}(X_i) &\equiv \Pr(Y_i = j \mid X_i) = \Pr(Y_i \leq j \mid X_i) - \Pr(Y_i \leq j-1 \mid X_i) \\
&= \begin{cases} \frac{\exp(\psi_j - X_i^{'}\beta)}{1+\exp(\psi_j - X_i^{'}\beta)} - \frac{\exp(\psi_{j-1} - X_i^{'}\beta)}{1+\exp(\psi_{j-1} - X_i^{'}\beta)} & \text{for logit} \\ \Phi\left(\psi_j - X_i^{'}\beta\right) - \Phi\left(\psi_{j-1} - X_i^{'}\beta\right) & \text{for probit} \end{cases}
\end{aligned}
$$

- ATE (APE): $\tau_j = \mathsf{E}\left[\pi_j(T_i = 1, W_i) - \pi_j(T_i = 0, W_i)\right]$

- Estimate β and ψ via MLE, plug the estimates in, replace E with $\frac{1}{n}\sum$, and compute CI by delta or MC or bootstrap

- Note that $X_i^{'}\beta$ appears both before and after the minus sign in π_{ij}
  $\longrightarrow$ Direction of effect of X_i on Y_{ij} is ambiguous (except top and bottom)
  $\longrightarrow$ Again, calculate quantities of interest, not just coefficients

Calculating Quantities of Interest (ht: Yamamoto)

- Predicted probability:

$$\begin{aligned}
\pi_{ij}(X_i) &\equiv \Pr(Y_i = j \mid X_i) = \Pr(Y_i \leq j \mid X_i) - \Pr(Y_i \leq j-1 \mid X_i) \\
&= \begin{cases} \frac{\exp(\psi_j - X_i^{'}\beta)}{1+\exp(\psi_j - X_i^{'}\beta)} - \frac{\exp(\psi_{j-1} - X_i^{'}\beta)}{1+\exp(\psi_{j-1} - X_i^{'}\beta)} & \text{for logit} \\ \Phi\left(\psi_j - X_i^{'}\beta\right) - \Phi\left(\psi_{j-1} - X_i^{'}\beta\right) & \text{for probit} \end{cases}
\end{aligned}$$

- ATE (APE): $\tau_j = \mathsf{E}\left[\pi_j(T_i = 1, W_i) - \pi_j(T_i = 0, W_i)\right]$

- Estimate β and ψ via MLE, plug the estimates in, replace E with $\frac{1}{n}\sum$, and compute CI by delta or MC or bootstrap

- Note that $X_i^{'}\beta$ appears both before and after the minus sign in π_{ij}
  $\longrightarrow$ Direction of effect of X_i on Y_{ij} is ambiguous (except top and bottom)
  $\longrightarrow$ Again, calculate quantities of interest, not just coefficients

Calculating Quantities of Interest (ht: Yamamoto)

- Predicted probability:

$$\begin{aligned}
\pi_{ij}(X_i) &\equiv \Pr(Y_i = j \mid X_i) = \Pr(Y_i \leq j \mid X_i) - \Pr(Y_i \leq j-1 \mid X_i) \\
&= \begin{cases} \frac{\exp(\psi_j - X_i^{'}\beta)}{1+\exp(\psi_j - X_i^{'}\beta)} - \frac{\exp(\psi_{j-1} - X_i^{'}\beta)}{1+\exp(\psi_{j-1} - X_i^{'}\beta)} & \text{for logit} \\ \Phi\left(\psi_j - X_i^{'}\beta\right) - \Phi\left(\psi_{j-1} - X_i^{'}\beta\right) & \text{for probit} \end{cases}
\end{aligned}$$

- ATE (APE): $\tau_j = \mathsf{E}\left[\pi_j(T_i = 1, W_i) - \pi_j(T_i = 0, W_i)\right]$

- Estimate β and ψ via MLE, plug the estimates in, replace E with $\frac{1}{n}\sum$, and compute CI by delta or MC or bootstrap

- Note that $X_i^{'}\beta$ appears both before and after the minus sign in π_{ij}
  $\longrightarrow$ Direction of effect of X_i on Y_{ij} is ambiguous (except top and bottom)
  $\longrightarrow$ Again, calculate quantities of interest, not just coefficients

Calculating Quantities of Interest (ht: Yamamoto)

- Predicted probability:

$$\pi_{ij}(X_i) \equiv \Pr(Y_i = j \mid X_i) = \Pr(Y_i \le j \mid X_i) - \Pr(Y_i \le j-1 \mid X_i)$$
$$= \begin{cases} \frac{\exp(\psi_j - X_i^{'}\beta)}{1+\exp(\psi_j - X_i^{'}\beta)} - \frac{\exp(\psi_{j-1} - X_i^{'}\beta)}{1+\exp(\psi_{j-1} - X_i^{'}\beta)} & \text{for logit} \\ \Phi\left(\psi_j - X_i^{'}\beta\right) - \Phi\left(\psi_{j-1} - X_i^{'}\beta\right) & \text{for probit} \end{cases}$$

- ATE (APE): $\tau_j = \mathsf{E}\left[\pi_j(T_i = 1, W_i) - \pi_j(T_i = 0, W_i)\right]$

- Estimate β and ψ via MLE, plug the estimates in, replace E with $\frac{1}{n}\sum$, and compute CI by delta or MC or bootstrap

- Note that $X_i^{'}\beta$ appears both before and after the minus sign in π_{ij}
  $\longrightarrow$ Direction of effect of X_i on Y_{ij} is ambiguous (except top and bottom)
  $\longrightarrow$ Again, calculate quantities of interest, not just coefficients

Calculating Quantities of Interest (ht: Yamamoto)

- Predicted probability:

$$
\begin{aligned}
\pi_{ij}(X_i) &\equiv \Pr(Y_i = j \mid X_i) = \Pr(Y_i \leq j \mid X_i) - \Pr(Y_i \leq j-1 \mid X_i) \\
&= \begin{cases} \frac{\exp(\psi_j - X_i'\beta)}{1+\exp(\psi_j - X_i'\beta)} - \frac{\exp(\psi_{j-1} - X_i'\beta)}{1+\exp(\psi_{j-1} - X_i'\beta)} & \text{for logit} \\ \Phi\left(\psi_j - X_i'\beta\right) - \Phi\left(\psi_{j-1} - X_i'\beta\right) & \text{for probit} \end{cases}
\end{aligned}
$$

- ATE (APE): $\tau_j = \mathsf{E}\left[\pi_j(T_i = 1, W_i) - \pi_j(T_i = 0, W_i)\right]$

- Estimate β and ψ via MLE, plug the estimates in, replace E with $\frac{1}{n}\sum$, and compute CI by delta or MC or bootstrap

- Note that $X_i'\beta$ appears both before and after the minus sign in π_{ij}
  $\longrightarrow$ Direction of effect of X_i on Y_{ij} is ambiguous (except top and bottom)
  $\longrightarrow$ Again, calculate quantities of interest, not just coefficients

Calculating Quantities of Interest (ht: Yamamoto)

- Predicted probability:

$$
\begin{aligned}
\pi_{ij}(X_i) &\equiv \Pr(Y_i = j \mid X_i) = \Pr(Y_i \leq j \mid X_i) - \Pr(Y_i \leq j-1 \mid X_i) \\
&= \begin{cases} \frac{\exp(\psi_j - X_i^{'}\beta)}{1+\exp(\psi_j - X_i^{'}\beta)} - \frac{\exp(\psi_{j-1} - X_i^{'}\beta)}{1+\exp(\psi_{j-1} - X_i^{'}\beta)} & \text{for logit} \\ \Phi\left(\psi_j - X_i^{'}\beta\right) - \Phi\left(\psi_{j-1} - X_i^{'}\beta\right) & \text{for probit} \end{cases}
\end{aligned}
$$

- ATE (APE): $\tau_j = \mathsf{E}\left[\pi_j(T_i = 1, W_i) - \pi_j(T_i = 0, W_i)\right]$

- Estimate β and ψ via MLE, plug the estimates in, replace E with $\frac{1}{n}\sum$, and compute CI by delta or MC or bootstrap

- Note that $X_i^{'}\beta$ appears both before and after the minus sign in π_{ij}
  $\longrightarrow$ Direction of effect of X_i on Y_{ij} is ambiguous (except top and bottom)
  $\longrightarrow$ Again, calculate quantities of interest, not just coefficients

Immigration and Media Priming(ht: Yamamoto)

Brader, Valentino and Suhay (2008):

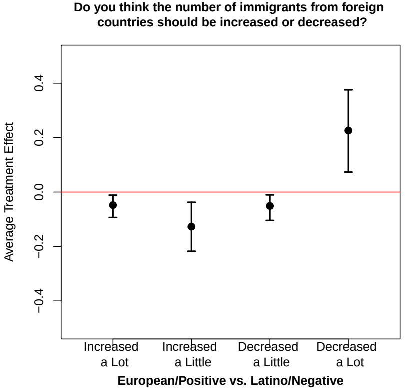
- Y_i: Ordinary response to question about increasing immigration
- T_{1i}, T_{2i}: Media cues (immigrant ethnicity $\times$ story tone)
- W_i: Respondent age and income

Estimated coefficients:

```
Coefficients:
          Value  s.e.    t
tone       0.27  0.32  0.85
eth       -0.33  0.32 -1.02
ppage      0.01  0.02  1.40
ppincimp   0.00  0.03  0.06
tone:eth   0.90  0.46  2.16

Intercepts:
    Value  s.e.    t
1|2 -1.93  0.58  -3.32
2|3 -0.12  0.55  -0.21
3|4  1.12  0.56   2.01
```

ATE:



Immigration and Media Priming(ht: Yamamoto)

Brader, Valentino and Suhay (2008):

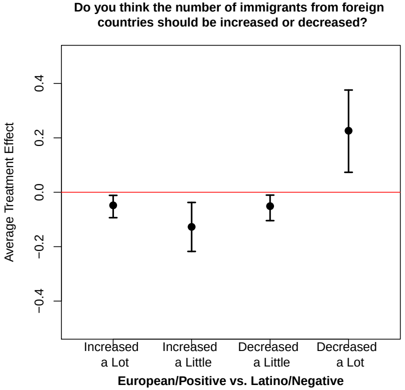
- Y_i: Ordinary response to question about increasing immigration
- T_{1i}, T_{2i}: Media cues (immigrant ethnicity $\times$ story tone)
- W_i: Respondent age and income

Estimated coefficients:

```
Coefficients:
          Value  s.e.   t
tone       0.27  0.32  0.85
eth       -0.33  0.32 -1.02
ppage      0.01  0.02  1.40
ppincimp   0.00  0.03  0.06
tone:eth   0.90  0.46  2.16

Intercepts:
    Value  s.e.    t
1|2 -1.93  0.58  -3.32
2|3 -0.12  0.55  -0.21
3|4  1.12  0.56   2.01
```

ATE:



Immigration and Media Priming(ht: Yamamoto)

Brader, Valentino and Suhay (2008):

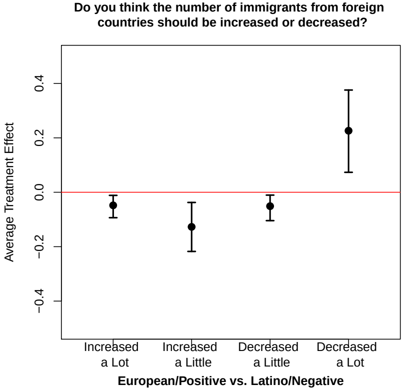
- Y_i: Ordinary response to question about increasing immigration
- T_{1i}, T_{2i}: Media cues (immigrant ethnicity $\times$ story tone)
- W_i: Respondent age and income

Estimated coefficients:

```
Coefficients:
          Value  s.e.   t
tone       0.27  0.32  0.85
eth       -0.33  0.32 -1.02
ppage      0.01  0.02  1.40
ppincimp  0.00  0.03  0.06
tone:eth  0.90  0.46  2.16

Intercepts:
    Value  s.e.    t
1|2 -1.93  0.58  -3.32
2|3 -0.12  0.55  -0.21
3|4  1.12  0.56   2.01
```

ATE:



Model Based Inference

1) Likelihood inference
2) Logit/Probit
   a) DGP
   b) Optimization
   c) Quantities of Interest
   d) Perfect + Near Perfect Separation
3) Ordered Probit
   a) DGP
   b) Optimization
   c) Quantities of Interest
4) Choice Models: Multinomial Logit/Probit