Lectures in Quantitative International Relations
Common Models in International Relations

Kevin A. Clarke
University of Rochester

Dublin City University, May 2007

Lectures

1. Introduction and Introduction to Maximum Likelihood Estimation
2. Some Common MLE Models Used in International Relations
3. Comparative Theory Testing
4. Choosing a Specfication

Overview of Lecture 2

1. The Probit Model
   - The Escalation of Great Power Militarized Disputes
   - Binary Dependent Variables
   - Numerical Maximization
   - Interpreting Probit Results

Overview of Lecture 2

1. The Probit Model
   - The Escalation of Great Power Militarized Disputes
   - Binary Dependent Variables
   - Numerical Maximization
   - Interpreting Probit Results
2. The Logit Model
   - Set-up
   - Interpretation

Overview of Lecture 2

1. The Probit Model
   - The Escalation of Great Power Militarized Disputes
   - Binary Dependent Variables
   - Numerical Maximization
   - Interpreting Probit Results
2. The Logit Model
   - Set-up
   - Interpretation
3. Ordered Probit and Logit
   - Set-up
   - Interpretation

Overview of Lecture 2

1. The Probit Model
   - The Escalation of Great Power Militarized Disputes
   - Binary Dependent Variables
   - Numerical Maximization
   - Interpreting Probit Results
2. The Logit Model
   - Set-up
   - Interpretation
3. Ordered Probit and Logit
   - Set-up
   - Interpretation
4. The Poisson Model
   - Set-up
   - Interpretation

Some International Relations (finally!)

Huth, Bennett, and Gelpi (1993) are interested in explaining the escalation of great-power militarized disputes.

They consider the population of great-power extended and direct immediate deterrence encounters from 1816-1984.

Some International Relations (finally!)

Huth, Bennett, and Gelpi (1993) are interested in explaining the escalation of great-power militarized disputes.

They consider the population of great-power extended and direct immediate deterrence encounters from 1816-1984.

A great-power deterrence encounter is defined by the explicit verbal threat of force or the movement and buildup of military forces in preparation for armed conflict by a challenging great power and a counterthreat by the defending great power.

Rational Deterrence Theory

Deterrence theory argues that the credibility of threats is the primary determinant of deterrence success or failure: the more credible the threat, the more likely deterrence will succeed.

Rational Deterrence Theory

Deterrence theory argues that the credibility of threats is the primary determinant of deterrence success or failure: the more credible the threat, the more likely deterrence will succeed.

Credibility is a function of two central variables: the balance of military capabilities between challenger and defender and the challenger and defender's level of resolve.

Rational Deterrence Theory

Deterrence theory argues that the credibility of threats is the primary determinant of deterrence success or failure: the more credible the threat, the more likely deterrence will succeed.

Credibility is a function of two central variables: the balance of military capabilities between challenger and defender and the challenger and defender's level of resolve.

Thus, escalation should be a function of military capabilities, the interests at stake, previous crisis behavior, and whether another dispute is ongoing.

Measurement of Variables

Dispute escalation: the failure of the deterrent policies of the great-power defender. Either the challenger may resort to the large-scale use of military force, or the defender may capitulate to the demands of the challenger regarding the central issues at stake in the dispute.

Measurement of Variables

Dispute escalation: the failure of the deterrent policies of the great-power defender. Either the challenger may resort to the large-scale use of military force, or the defender may capitulate to the demands of the challenger regarding the central issues at stake in the dispute.

Balance of conventional military capabilities: ratio comparing the capabilities of the challenger(s) to the total capabilities of the challenger(s) and defender(s).

Measurement of Variables

Defender possession of second strike nuclear capability: coded 1 if the defender possesses the capability to deliver nuclear weapons onto the population of the challenger following the absorption of a nuclear first strike, and 0 otherwise.

Measurement of Variables

Defender possession of second strike nuclear capability: coded 1 if the defender possesses the capability to deliver nuclear weapons onto the population of the challenger following the absorption of a nuclear first strike, and 0 otherwise.

Interests at stake: For the defender, coded 1 if the issues at stake in the dispute center on the control or acquisition of territory adjacent to, or part of, the homeland or colonial empire of the defender.

Same for challenger.

Measurement of Variables

Past behavior: coded 1 if the state, either challenger or defender, suffered a diplomatic put-down in a dispute with the same opponent within the past 10 years.

Measurement of Variables

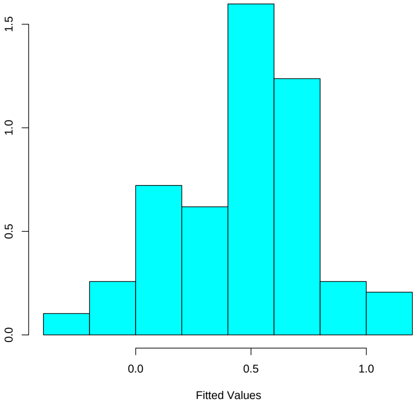
Past behavior: coded 1 if the state, either challenger or defender, suffered a diplomatic put-down in a dispute with the same opponent within the past 10 years.

Current dispute involvement: coded 1 if the state was either involved in at least one other militarized dispute within the past six months, or a war with a third party during the current or previous year.

Regression Results, $n = 97$

Variable	Coefficient	S.E.	z	p-value
Balance of forces	0.354	0.238	1.49	0.14
Secure 2nd strike	-0.457	0.097	-4.70	0.00
Defender vital interests	-0.192	0.094	-2.03	0.05
Challenger vital interests	0.179	0.095	1.89	0.06
Defender backed down	0.232	0.108	2.15	0.03
Challenger backed down	-0.183	0.135	-1.35	0.18
Defender other dispute	0.171	0.095	1.81	0.07
Challenger other dispute	-0.005	0.094	-0.06	0.95
Constant	0.313	0.169	1.85	0.07
Log-Likelihood	-49.616			

Histogram of the Predicted Values



Binary Outcomes

When the outcome is binary, we can no longer use regular regression. There are two problems:

Binary Outcomes

When the outcome is binary, we can no longer use regular regression. There are two problems:

1. Regression does not constrain the predicted values to lie between 0 and 1.

Binary Outcomes

When the outcome is binary, we can no longer use regular regression. There are two problems:

1. Regression does not constrain the predicted values to lie between 0 and 1.
2. Regression is heteroscedastic (non-constant error variance) in this situation.

The Latent Variable

Although we only see if a dispute escalates or not, we can imagine an underlying variable, which is the propensity to escalate. We also assume that this latent variable is a function of some explanatory variables,

$$y_i^* = \mathbf{x}_i'\beta + \epsilon_i \quad \epsilon_i \sim N(\mathbf{0}, \sigma^2).$$

The Latent Variable

Although we only see if a dispute escalates or not, we can imagine an underlying variable, which is the propensity to escalate. We also assume that this latent variable is a function of some explanatory variables,

$$y_i^* = \mathbf{x}_i'\beta + \epsilon_i \quad \epsilon_i \sim N(\mathbf{0}, \sigma^2).$$

We only observe an escalation when the propensity crosses a threshold:

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

The Implication

$$\begin{aligned}\Pr(y_i = 1) &= \Pr(y_i^* > 0)\\ &= \Pr(\mathbf{x}_i'\beta + \epsilon_i > 0)\\ &= \Pr(\epsilon_i > -\mathbf{x}_i'\beta)\\ &= \Pr\left(\frac{\epsilon_i}{\sigma} > \mathbf{x}_i'\frac{\beta}{\sigma}\right)\end{aligned}$$

We write the probability in this way because now ϵ_i/σ is distributed standard normal (mean=0 and variance=1). That is, it has been normalized by subtracting the mean, 0, and dividing by the standard error, σ.

The Normal is Symmetric

Because the normal curve is symmetric,

$$\begin{aligned}\Pr(y_i = 1) &= \Pr\left(\frac{\epsilon_i}{\sigma} > -\mathbf{x}_i'\frac{\beta}{\sigma}\right)\\ &= \Pr\left(\frac{\epsilon_i}{\sigma} < \mathbf{x}_i'\frac{\beta}{\sigma}\right)\\ &= \Phi\left(\mathbf{x}_i'\frac{\beta}{\sigma}\right)\end{aligned}$$

where Φ indicates the standard normal distribution.

The Probability of Failure

If the probability of success is

$$\Pr(y_i = 1) = \Phi\left(\mathbf{x}_i'\beta\right)^{1},$$

[1] We set $\sigma = 1$ as a normalization.

The Probability of Failure

If the probability of success is

$$\Pr(y_i = 1) = \Phi\left(\mathbf{x}_i'\beta\right)^{1},$$

then the probability of failure must be

$$\Pr(y_i = 0) = 1 - \Pr(y_i = 1) = 1 - \Phi\left(\mathbf{x}_i'\beta\right).$$

[1] We set $\sigma = 1$ as a normalization.

The Likelihood

Remember that the likelihood is the product of the individual densities:

$$\begin{aligned} L(\mathbf{y};\beta) &= \Pr(y_1 = 0)\cdot\Pr(y_2 = 0)\cdots\Pr(y_m = 0)\cdot \\ &\quad \Pr(y_{m+1} = 1)\cdots\Pr(y_n = 1) \\ &= \prod_{i=1}^{m}\left[1-\Phi\left(\mathbf{x}_i'\beta\right)\right]\prod_{i=m+1}^{n}\left[\Phi\left(\mathbf{x}_i'\beta\right)\right] \\ &= \prod_{i=1}^{n}\Phi\left(\mathbf{x}_i'\beta\right)^{y_i}\left[1-\Phi\left(\mathbf{x}_i'\beta\right)\right]^{1-y_i} \end{aligned}$$

The Log-Likelihood

$$\ln L(\mathbf{y};\beta) = \sum_{i=1}^{n} \left\{ y_i \cdot \ln\left[\Phi\left(\mathbf{x}_i'\beta\right)\right] + (1-y_i)\cdot \ln\left[1-\Phi\left(\mathbf{x}_i'\beta\right)\right] \right\}$$

The Log-Likelihood

$$\ln L(\mathbf{y};\beta) = \sum_{i=1}^{n} \left\{ y_i \cdot \ln\left[\Phi\left(\mathbf{x}_i'\beta\right)\right] + (1-y_i)\cdot \ln\left[1-\Phi\left(\mathbf{x}_i'\beta\right)\right]\right\}$$

Just like our previous examples, we should take the derivative of the log-likelihood, set it equal to zero, and solve for $\hat{\beta}$.

The Log-Likelihood

$$\ln L(\mathbf{y};\beta) = \sum_{i=1}^{n} \left\{ y_i \cdot \ln\left[\Phi\left(\mathbf{x}_i'\beta\right)\right] + (1-y_i)\cdot \ln\left[1-\Phi\left(\mathbf{x}_i'\beta\right)\right] \right\}$$

Just like our previous examples, we should take the derivative of the log-likelihood, set it equal to zero, and solve for $\hat{\beta}$.

Except we can't.

The Log-Likelihood

$$\ln L(\mathbf{y};\beta) = \sum_{i=1}^{n} \left\{ y_i \cdot \ln\left[\Phi\left(\mathbf{x}_i'\beta\right)\right] + (1 - y_i) \cdot \ln\left[1 - \Phi\left(\mathbf{x}_i'\beta\right)\right] \right\}$$

Just like our previous examples, we should take the derivative of the log-likelihood, set it equal to zero, and solve for $\hat{\beta}$.

Except we can't.
There is no analytical solution to this problem. We need to find another way.

Numerical Maximization

Most of the time, we can't solve $\partial \ln L / \partial \boldsymbol{\theta} = \mathbf{0}$.

When we can't find analytical solutions, we need to use numerical methods to find one.

Numerical Maximization

Most of the time, we can't solve $\partial \ln L / \partial \boldsymbol{\theta} = \mathbf{0}$.

When we can't find analytical solutions, we need to use numerical methods to find one.

Letting

$$g(\theta) = \frac{d \ln L(\theta; \mathbf{x})}{d\theta},$$

the problem is to solve for

$$g(\theta) = 0.$$

This is really a matter of finding the roots (zeros) of $g()$.

Finding roots

We can find the root of a function by turning the problem into an easier one.

Finding roots

We can find the root of a function by turning the problem into an easier one.

1. Choose a starting value, θ_0 (i.e., a guess).

Finding roots

We can find the root of a function by turning the problem into an easier one.

1. Choose a starting value, θ_0 (i.e., a guess).
2. Replace the function at that point with its tangent, $g(\theta)$.

Finding roots

We can find the root of a function by turning the problem into an easier one.

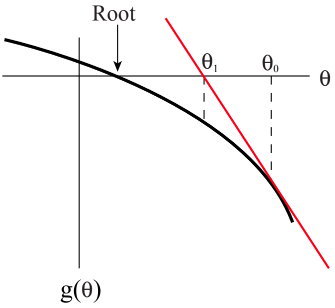
1. Choose a starting value, θ_0 (i.e., a guess).
2. Replace the function at that point with its tangent, $g(\theta)$.
3. Compute the root (zero), θ_1, of the tangent, which is easier than finding the root of the function, itself.

Finding roots

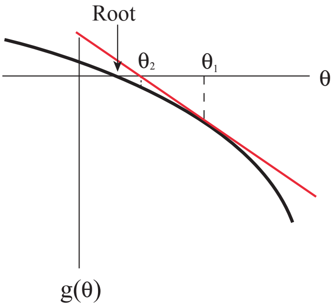
We can find the root of a function by turning the problem into an easier one.

1. Choose a starting value, θ_0 (i.e., a guess).
2. Replace the function at that point with its tangent, $g(\theta)$.
3. Compute the root (zero), θ_1, of the tangent, which is easier than finding the root of the function, itself.
4. Iterate the process using the new value, θ_1 as your new guess.

Newton's Method



Newton's Method



Demonstrations of Newton's Method

Demonstrations

The Mathematics

What's happening mathmetically:

$$g'(\theta_0) = \frac{\text{rise}}{\text{run}} = \frac{g(\theta_0) - 0}{\theta_0 - \theta_1} = \frac{0 - g(\theta_0)}{(\theta_1 - \theta_0)}$$

which equals

$$\theta_1 = \theta_0 - \frac{g(\theta_0)}{g'(\theta_0)}$$

The Mathematics

What's happening mathmetically:

$$g'(\theta_0) = \frac{\text{rise}}{\text{run}} = \frac{g(\theta_0) - 0}{\theta_0 - \theta_1} = \frac{0 - g(\theta_0)}{(\theta_1 - \theta_0)}$$

which equals

$$\theta_1 = \theta_0 - \frac{g(\theta_0)}{g'(\theta_0)}$$

The next guess is then

$$\theta_2 = \theta_1 - \frac{g(\theta_1)}{g'(\theta_1)}$$

From Root Finding to Maximization

To go from a root finding algorithm to a maximization algorithm, substitute $f'()$ for $g()$ and $f''()$ for $g'()$.

From Root Finding to Maximization

To go from a root finding algorithm to a maximization algorithm, substitute $f'()$ for $g()$ and $f''()$ for $g'()$.

1. Start with a guess, θ_0.

From Root Finding to Maximization

To go from a root finding algorithm to a maximization algorithm, substitute $f'()$ for $g()$ and $f''()$ for $g'()$.

1. Start with a guess, θ_0.
2. Calculate new guess, $\theta_1 = \theta_0 - f'(\theta)/f''(\theta)$.

From Root Finding to Maximization

To go from a root finding algorithm to a maximization algorithm, substitute $f'()$ for $g()$ and $f''()$ for $g'()$.

1. Start with a guess, θ_0.
2. Calculate new guess, $\theta_1 = \theta_0 - f'(\theta)/f''(\theta)$.
3. Repeat.

Probit Results, $n = 97$

Variable	Coefficient	S.E.	z	p-value
Balance of forces	1.465	0.816	1.80	0.07
Secure 2nd strike	-1.761	0.413	-4.26	0.00
Defender vital interests	-0.887	0.370	-2.40	0.02
Challenger vital interests	0.829	0.365	2.27	0.02
Defender backed down	1.027	0.415	2.47	0.01
Challenger backed down	-0.709	0.491	-1.44	0.15
Defender other dispute	0.726	0.350	2.08	0.04
Challenger other dispute	-0.017	0.343	-0.05	0.99
Constant	-0.886	0.591	-1.50	0.13
Log-Likelihood	-45.558			

Interpreting Results

Interpreting linear regression results is simple:

A 1-unit change in X_k produces, on average, a $\hat{\beta}_k$ change in y.

Interpreting Results

Interpreting linear regression results is simple:

A 1-unit change in X_k produces, on average, a $\hat{\beta}_k$ change in y.

The same interpretation holds for probit...but only for the latent variable, y^*:

A 1-unit change in X_k produces, on average, a $\hat{\beta}_k$ change in y^*.

Interpreting Results

Interpreting linear regression results is simple:

A 1-unit change in X_k produces, on average, a $\hat{\beta}_k$ change in y.

The same interpretation holds for probit...but only for the latent variable, y^*:

A 1-unit change in X_k produces, on average, a $\hat{\beta}_k$ change in y^*.

Of course, we want to know the effect on y, itself.

Another Look at the Difference

In the standard linear regression model,

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i,$$

the effect of x_1 is given by the partial derivative,

$$\frac{\partial y_i}{\partial x_{i1}} = \beta_1 .$$

Another Look at the Difference

In the standard linear regression model,

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i,$$

the effect of x_1 is given by the partial derivative,

$$\frac{\partial y_i}{\partial x_{i1}} = \beta_1 .$$

In the probit model, the effect of x_1 is also the partial derivative,

$$\frac{\partial \Pr(y_i = 1)}{\partial x_{i1}} = \beta_1 \cdot \phi\left(\beta_0 + \sum_{i=1}^{k} \beta_i x_{ij}\right) .$$

The Upshot

The effect of a variable in a probit model...

The Upshot

The effect of a variable in a probit model...

depends.

The Upshot

The effect of a variable in a probit model...

depends.

It depends on the level of the other variables. Thus to interpret probit coefficients, we have to think about the other variables.

First Differences

The easiest way to interpret probit results is to use what are called first differences.

First Differences

The easiest way to interpret probit results is to use what are called first differences.

The idea is to change a single variable and see what effect it has while holding constant all the other variables at some meaningful value.

First Differences

The easiest way to interpret probit results is to use what are called first differences.

The idea is to change a single variable and see what effect it has while holding constant all the other variables at some meaningful value.

For binary variables, this is usually zero. The mean is often used for continuous variables.

The Effect of a Secure Second-Strike

"Secure second-strike" can take on two values: 0 and 1. What is the probability of escalation when it is 0?

The Effect of a Secure Second-Strike

"Secure second-strike" can take on two values: 0 and 1. What is the probability of escalation when it is 0?

We'll hold "balances of forces" at its mean, 0.479, and the binary variables at 0.

$$
\begin{aligned}
y^* &= -0.886 + 1.465 \times 0.479 - 1.761 \times 0 \\
&= -0.184
\end{aligned}
$$

The Effect of a Secure Second-Strike

"Secure second-strike" can take on two values: 0 and 1. What is the probability of escalation when it is 0?

We'll hold "balances of forces" at its mean, 0.479, and the binary variables at 0.

$$\begin{aligned} y^* &= -0.886 + 1.465 \times 0.479 - 1.761 \times 0 \\ &= -0.184 \end{aligned}$$

But note that this is not a probability; it is a z-score. Remember that in a probit, $Pr(y_i = 1) = \Phi(\mathbf{x}_i'\beta)$. What we have here is just $\mathbf{x}_i'\hat{\beta}$.

The Effect of a Secure Second-Strike

To get the probability, we have to feed the linear prediction, -0.184, into the cumulative standard normal distribution,

$$\Pr(y = 1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\mathbf{x}'\boldsymbol{\beta}} \exp\left(-\frac{z^2}{2}\right) dz.$$

When we do this, the probability of escalation when the defender does not have nuclear weapons is 0.427.

The Effect of a Secure Second-Strike

To get the probability, we have to feed the linear prediction, -0.184, into the cumulative standard normal distribution,

$$\Pr(y = 1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\mathbf{x}'\boldsymbol{\beta}} \exp\left(-\frac{z^2}{2}\right) dz.$$

When we do this, the probability of escalation when the defender does not have nuclear weapons is 0.427.

How about when the defender does have nuclear weapons?

The Effect of a Secure Second-Strike

$$
\begin{aligned}
y^* &= -0.886 + 1.465 \times 0.479 - 1.761 \times 0 \\
&= -0.184 \\
\Pr(y = 1) &= \Phi(-0.184) \\
&= 0.427
\end{aligned}
$$

The Effect of a Secure Second-Strike

$$\begin{aligned} y^* &= -0.886 + 1.465 \times 0.479 - 1.761 \times 0 \\ &= -0.184 \\ \Pr(y = 1) &= \Phi(-0.184) \\ &= 0.427 \end{aligned}$$

$$\begin{aligned} y^* &= -0.886 + 1.465 \times 0.479 - 1.761 \times 1 \\ &= -1.945 \\ \Pr(y = 1) &= \Phi(-1.945) \\ &= 0.026 \end{aligned}$$

The Effect of a Secure Second-Strike

So the possession by the defender of a secure second-strike decreases the probability of escalation by 40% (0.427-0.026=0.401).

The Effect of a Secure Second-Strike

So the possession by the defender of a secure second-strike decreases the probability of escalation by 40% (0.427-0.026=0.401).

What if we left all the binary variables at 1 instead of 0?

The Effect of a Secure Second-Strike

$$
\begin{aligned}
y^* &= -0.886 + 1.465(0.479) - 1.761(0) - 0.887(1) \\
&\quad +0.829(1) + 1.027(1) - 0.709(1) + 0.726(1) \\
&\quad -0.017(1) \\
&= 0.785 \\
\Pr(y = 1) &= \Phi(0.785) = 0.784
\end{aligned}
$$

The Effect of a Secure Second-Strike

$$\begin{aligned} y^* &= -0.886 + 1.465(0.479) - 1.761(0) - 0.887(1) \\ &\quad +0.829(1) + 1.027(1) - 0.709(1) + 0.726(1) \\ &\quad -0.017(1) \\ &= 0.785 \\ \Pr(y=1) &= \Phi(0.785) = 0.784 \end{aligned}$$

$$\begin{aligned} y^* &= -0.886 + 1.465(0.479) - 1.761(1) - 0.887(1) \\ &\quad +0.829(1) + 1.027(1) - 0.709(1) + 0.726(1) \\ &\quad -0.017(1) \\ &= -0.976 \\ \Pr(y=1) &= \Phi(-1.945) = 0.165 \end{aligned}$$

The Effect of a Secure Second-Strike

When the other binary variables are set at 0, the defender possession of nuclear weapons decreases the probability of escalation by 40% (0.427-0.026=0.401).

The Effect of a Secure Second-Strike

When the other binary variables are set at 0, the defender possession of nuclear weapons decreases the probability of escalation by 40% (0.427-0.026=0.401).

When the other binary variables are set at 1, the defender possession of nuclear weapons decreases the probability of escalation by almost 62% (0.784-0.165=0.619).

The Effect of a Secure Second-Strike

When the other binary variables are set at 0, the defender possession of nuclear weapons decreases the probability of escalation by 40% (0.427-0.026=0.401).

When the other binary variables are set at 1, the defender possession of nuclear weapons decreases the probability of escalation by almost 62% (0.784-0.165=0.619).

This is what I mean when I say the effect of a variable depends on the level of the other variables. Interpretation is as much an art as a science.

The Logit Model

Another approach to estimating models with binary dependent variables is logit.

The Logit Model

Another approach to estimating models with binary dependent variables is logit.

The development of logit is identical to probit. Where in probit the error term, ϵ_i, follows a Normal distribution, for logit, the error term follows an extreme value distribution.

The Logit Model

Another approach to estimating models with binary dependent variables is logit.

The development of logit is identical to probit. Where in probit the error term, ϵ_i, follows a Normal distribution, for logit, the error term follows an extreme value distribution.

The probability of escalation is

$$\Pr(y_i = 1) = \Lambda(\mathbf{x}_i'\beta) = \frac{\exp(\mathbf{x}_i'\beta)}{1 + \exp(\mathbf{x}_i'\beta)}$$

Logit Results, $n = 97$

Variable	Coefficient	S.E.	z	p-value
Balance of forces	2.424	1.382	1.75	0.08
Secure 2nd strike	-2.927	0.746	-3.92	0.00
Defender vital interests	-1.463	0.643	-2.28	0.02
Challenger vital interests	1.363	0.628	2.17	0.03
Defender backed down	1.718	0.721	2.38	0.02
Challenger backed down	-1.172	0.833	-1.41	0.16
Defender other dispute	1.215	0.601	2.02	0.04
Challenger other dispute	-0.09	0.586	2.02	0.88
Constant	-1.437	1.001	-1.44	0.15
Log-Likelihood	-45.917			

Interpretation

Although the logit coefficients look different from the probit coefficients, they mean the exact same thing.

Consider again, the effect of secure second-strike.

$$\begin{aligned} y^* &= -1.437 + 2.424 \times 0.479 - 2.927 \times 0 \\ &= -0.276 \end{aligned}$$

$$\begin{aligned} y^* &= -1.437 + 2.424 \times 0.479 - 2.927 \times 1 \\ &= -3.206 \end{aligned}$$

Interpretation

Although the logit coefficients look different from the probit coefficients, they mean the exact same thing.

Consider again, the effect of secure second-strike.

$$\begin{aligned} y^* &= -1.437 + 2.424 \times 0.479 - 2.927 \times 0 \\ &= -0.276 \end{aligned}$$

$$\begin{aligned} y^* &= -1.437 + 2.424 \times 0.479 - 2.927 \times 1 \\ &= -3.206 \end{aligned}$$

Again, these are not probabilities.

Interpretation

For logit, instead of feeding the linear predictors into the cumulative normal, we feed them into the logit:

$$\Pr(y = 1) = \frac{\exp(\mathbf{x}_i'\hat{\beta})}{1 + \exp(\mathbf{x}_i'\hat{\beta})}.$$

Interpretation

For logit, instead of feeding the linear predictors into the cumulative normal, we feed them into the logit:

$$\Pr(y = 1) = \frac{\exp(\mathbf{x}_i'\hat{\beta})}{1 + \exp(\mathbf{x}_i'\hat{\beta})}.$$

For nukes=0, $\Pr(y = 1) = \Lambda(-0.276) = 0.432$.

Interpretation

For logit, instead of feeding the linear predictors into the cumulative normal, we feed them into the logit:

$$\Pr(y=1)=\frac{\exp(\mathbf{x}_i'\hat{\beta})}{1+\exp(\mathbf{x}_i'\hat{\beta})}.$$

For nukes=0, $\Pr(y=1)=\Lambda(-0.276)=0.432$.
For nukes=1, $\Pr(y=1)=\Lambda(-3.206)=0.039$.

Interpretation

For logit, instead of feeding the linear predictors into the cumulative normal, we feed them into the logit:

$$\Pr(y = 1) = \frac{\exp(\mathbf{x}_i'\hat{\beta})}{1 + \exp(\mathbf{x}_i'\hat{\beta})}.$$

For nukes=0, $\Pr(y = 1) = \Lambda(-0.276) = 0.432$.
For nukes=1, $\Pr(y = 1) = \Lambda(-3.206) = 0.039$.

The difference is 39.3%, which indicates that defender possession of nuclear weapons decreases the probability of escalation by about 40%.

Interpretation

For logit, instead of feeding the linear predictors into the cumulative normal, we feed them into the logit:

$$\Pr(y=1) = \frac{\exp(\mathbf{x}_i'\hat{\beta})}{1+\exp(\mathbf{x}_i'\hat{\beta})}.$$

For nukes=0, $\Pr(y=1) = \Lambda(-0.276) = 0.432$.
For nukes=1, $\Pr(y=1) = \Lambda(-3.206) = 0.039$.

The difference is 39.3%, which indicates that defender possession of nuclear weapons decreases the probability of escalation by about 40%.

This is precisely what probit told us; which you use makes no difference.

Discrimination in International Relations

Why do some ethnic groups in conflict receive more external support than others do?

Saideman (2002) argues that groups with ethnic ties to actors in positions of power elsewhere are more likely to receive external assistance.

The data come from the Minorities at Risk Dataset (Phase III), where minorities "at risk" are defined as those ethnic groups that as groups gain from or are hurt by systematic discriminatory treatment compared to other groups in society.

The Dependent Variable

The dependent variable measures the highest level of support given to an ethnic group—the intensity.

Value	Label	Minorities at Risk labels
0	None	No support received
1	Low	Ideological encouragement
2	Moderate	Non-military financial support
3	Strong	Funds for military supplies
4	Intense	Blockades, combat units, etc.

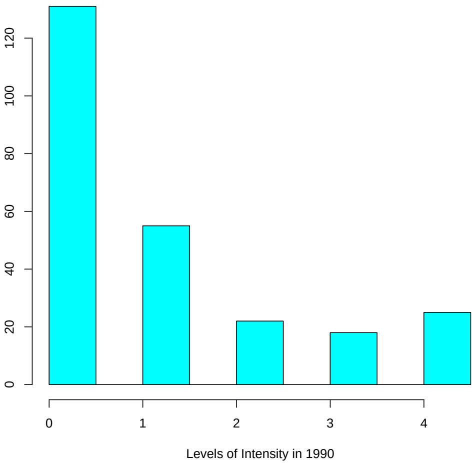
The Dependent Variable

The dependent variable measures the highest level of support given to an ethnic group—the intensity.

Value	Label	Minorities at Risk labels
0	None	No support received
1	Low	Ideological encouragement
2	Moderate	Non-military financial support
3	Strong	Funds for military supplies
4	Intense	Blockades, combat units, etc.

Once again, this dependent variable is discrete and not continuous.

Intensity of Support



The Latent Variable

We'll start again with a latent variable:

$$y_i^* = \mathbf{x}_i'\beta + \epsilon_i,$$

where **x** doesn't include an intercept (this is the identifying assumption).

The Latent Variable

We'll start again with a latent variable:

$$y_i^* = \mathbf{x}_i' \boldsymbol{\beta} + \epsilon_i,$$

where **x** doesn't include an intercept (this is the identifying assumption).

We get different outcomes as y^* crosses unknown thresholds called "cut-points" or α_j. For very low y^* intensity is none, for $y^* > \alpha_1$ intensity is low, for $y^* > \alpha_2$ intensity is moderate, and so on.

Probabilities

Let $y_i = j$ if $\alpha_{j-1} < y_i^* \leq \alpha_j$.

The probability of being category j is

$$\begin{aligned} Pr[y_i = j] &= \Pr[\alpha_{j-1} < y_i^* \leq \alpha_j] \\ &= \Pr[\alpha_{j-1} < \mathbf{x}_i' \boldsymbol{\beta} + \epsilon_i \leq \alpha_j] \\ &= \Pr[\alpha_{j-1} - \mathbf{x}_i' \beta < \epsilon_i \leq \alpha_j - \mathbf{x}_i' \beta] \\ &= F(\alpha_j - \mathbf{x}_i' \beta) - F(\alpha_{j-1} - \mathbf{x}_i' \beta) \end{aligned}$$

where F is the cdf of ϵ_i.

The Log-likelihood Function

We get estimates by maximizing the following log-likelihood function,

$$\ln L(\mathbf{y}; \alpha, \beta) = \sum_{i=1}^{N} \sum_{j=1}^{m} y_{ij} \ln p_{ij},$$

where p_{ij} was defined in the previous slide.

The Log-likelihood Function

We get estimates by maximizing the following log-likelihood function,

$$\ln L(\mathbf{y}; \alpha, \beta) = \sum_{i=1}^{N} \sum_{j=1}^{m} y_{ij} \ln p_{ij},$$

where p_{ij} was defined in the previous slide.

We get ordered probit or logit by assuming that ϵ_i is distributed either logistic or normal.

Independent Variables

Ethnic politics

- racial differences
- linguistic differences
- religious differences
- does ethnic kin dominate adjoining state?

Independent Variables

Ethnic politics

- racial differences
- linguistic differences
- religious differences
- does ethnic kin dominate adjoining state?

Vulnerability

- African states
- is group separatist?
- other separatist groups in host state
- separatists in nearby state

Independent Variables

Independent Variables

Realism

- relative power of host state

Independent Variables

Realism

- relative power of host state

Control variables

- regime type
- rebellion

Ordered Logit Results, $n = 186$

Variable	Coefficient	S.E.	z	p-value
Racial differences	0.123	0.157	0.79	0.43
Linguistic differences	-0.247	0.432	-0.57	0.57
Religious differences	-0.085	0.126	-0.67	0.50
Ethnic kin dominate adjoining	0.841	0.376	2.23	0.03
African states	-0.110	0.461	-0.24	0.81
Is group separatist	0.385	0.390	0.99	0.32
Other separatist groups	-0.135	0.103	-1.31	0.19
Separatists in nearby state	0.123	0.040	3.06	0.00
Relative power of host	-0.006	0.052	-0.11	0.91
Regime type	-0.049	0.026	-1.90	0.06
Rebellion	0.287	0.087	3.30	0.00
Cut 1	0.923	0.460		
Cut 2	2.111	0.485		
Cut 3	2.795	0.504		
Cut 4	3.378	0.528		
Log-Likelihood	-211.160			

Interpretation

We could use first differences here again, but let's try something different. Let's plot the effect that a variable, say *Does ethnic kin dominate adjoining state?*, has on the probability of different levels of the dependent variable.

$$\Pr(y = j|\mathbf{x}) = F(\hat{\alpha}_j - \mathbf{x}_i'\hat{\beta}) - F(\hat{\alpha}_{j-1} - \mathbf{x}_i'\hat{\beta})$$

Interpretation

We could use first differences here again, but let's try something different. Let's plot the effect that a variable, say *Does ethnic kin dominate adjoining state?*, has on the probability of different levels of the dependent variable.

$$\Pr(y = j|\mathbf{x}) = F(\hat{\alpha}_j - \mathbf{x}_i'\hat{\beta}) - F(\hat{\alpha}_{j-1} - \mathbf{x}_i'\hat{\beta})$$

Thus, the probability of being in the "no support" category is

$$\Pr(y = 1|\mathbf{x}) = F(\hat{\alpha}_1 - \mathbf{x}_i'\hat{\beta}),$$

where $F()$ is the cdf of the logit distribution.

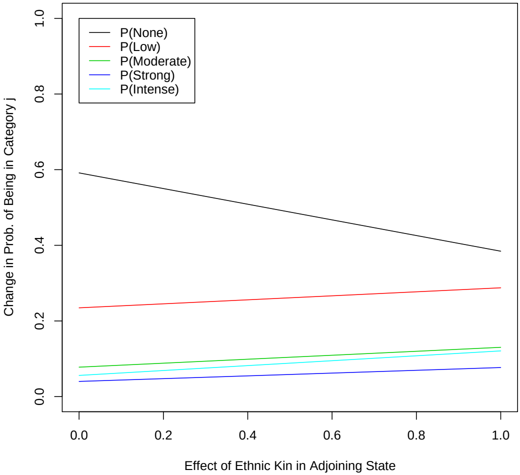
Interpretation

The probabilities of being in the other categories are:

$$
\begin{aligned}
\Pr(y=1|\mathbf{x}) &= F(\hat{\alpha}_1 - \bar{\mathbf{x}}_i'\hat{\beta}) \\
\Pr(y=2|\mathbf{x}) &= F(\hat{\alpha}_2 - \bar{\mathbf{x}}_i'\hat{\beta}) - F(\hat{\alpha}_1 - \bar{\mathbf{x}}_i'\hat{\beta}) \\
\Pr(y=3|\mathbf{x}) &= F(\hat{\alpha}_3 - \bar{\mathbf{x}}_i'\hat{\beta}) - F(\hat{\alpha}_2 - \bar{\mathbf{x}}_i'\hat{\beta}) \\
\Pr(y=4|\mathbf{x}) &= F(\hat{\alpha}_4 - \bar{\mathbf{x}}_i'\hat{\beta}) - F(\hat{\alpha}_3 - \bar{\mathbf{x}}_i'\hat{\beta}) \\
\Pr(y=5|\mathbf{x}) &= 1 - F(\hat{\alpha}_4 - \bar{\mathbf{x}}_i'\hat{\beta})
\end{aligned}
$$

We'll set the other variables at their mean and let *ethnic kin* change from 0 to 1.

Graph of Effects



Count data

Sometimes the dependent variable of interest is a non-negative integer or count that we wish to explain or analyze in terms of a set of regressors.

Count data

Sometimes the dependent variable of interest is a non-negative integer or count that we wish to explain or analyze in terms of a set of regressors.

One example is the number of interstate disputes that occur in a given year. This count is bounded by zero and rarely rises beyond a few disputes.

Count data

Sometimes the dependent variable of interest is a non-negative integer or count that we wish to explain or analyze in terms of a set of regressors.

One example is the number of interstate disputes that occur in a given year. This count is bounded by zero and rarely rises beyond a few disputes.

The data this time come from Pollins (*APSR*, March 1996).

Armed Conflict and the Economic Long Wave

Many systemic explanations of militarized disputes have concentrated on the association of long-term repeated cycles in the international system and the coincidence of particular subperiods in these cycles with the onset of militarized disputes.

Armed Conflict and the Economic Long Wave

Many systemic explanations of militarized disputes have concentrated on the association of long-term repeated cycles in the international system and the coincidence of particular subperiods in these cycles with the onset of militarized disputes.

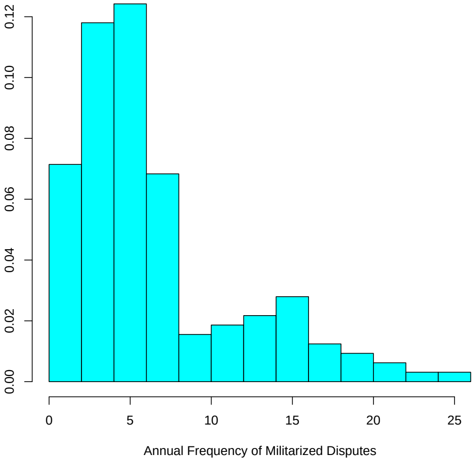
Goldstein (1988), for instance, argues that repeated patterns of rise and decline in global economic activity affect the number of interstate militarized disputes.

Goldstein's Phases

The level of conflict, according to Goldstein, coincides with upswing phases in the global economy, while downswing phases are relatively peaceful.

Stagnation	Rebirth	Expansion	War
1816	1831	1849	1861
1873	1885	1894	1911
1921	1934	1941	1969

The Dependent Variable



The Poisson Distribution

The Poisson distribution for the number of occurrences of a event has density

$$\Pr(Y = y) = \frac{\exp(-\mu)\mu^y}{y!}, \quad y = 0, 1, 2, \ldots,$$

where μ is the intensity or rate parameter.

The Poisson Distribution

The Poisson distribution for the number of occurrences of a event has density

$$\Pr(Y = y) = \frac{\exp(-\mu)\mu^y}{y!}, \quad y = 0, 1, 2, \ldots,$$

where μ is the intensity or rate parameter.

The Poisson distribution has an interesting property known as equidispersion—the mean and the variance are both μ.

A Poisson Regression

To turn the Poisson distribution into a regression model, we make the intensity parameter a function of covariates,

$$\mu_i = \exp(\mathbf{x}_i'\beta), \quad i = 1, \ldots, N.$$

A Poisson Regression

To turn the Poisson distribution into a regression model, we make the intensity parameter a function of covariates,

$$\mu_i = \exp(\mathbf{x}_i'\beta), \quad i = 1, \ldots, N.$$

The log-likelihood is therefore

$$\ln L[\beta; \mathbf{y}, \mathbf{X}] = \sum_{i=1}^{N} \left\{ y_i \mathbf{x}_i'\beta - \exp(\mathbf{x}_i'\beta) - \ln y_i! \right\}.$$

Poisson Results, $n = 161$

Variable	Coefficient	S.E.	z	p-value
Stagnation	0.645	0.100	6.24	0.00
Rebirth	0.857	0.096	8.91	0.00
Expansion	1.068	0.085	12.52	0.00
War	1.100	0.105	10.49	0.00
D_{t-1}	0.061	0.009	6.57	0.00
$Members_t$	0.006	0.001	4.21	0.00
Log-Likelihood	-368.569			

Interpretation

We can interpret poisson coefficients using first differences just as we did with probit and logit coefficients. Hold continuous variables at their mean, binary variables at 0, and move the variable of interest by a substantively meaningful unit.

Interpretation

We can interpret poisson coefficients using first differences just as we did with probit and logit coefficients. Hold continuous variables at their mean, binary variables at 0, and move the variable of interest by a substantively meaningful unit.

Changing "stagnation" from 0 to 1, for instance, produces a

$$
\begin{aligned}
\frac{\Delta \Pr(Y = y)}{\Delta x_k} &= \exp(0.645(1) + 0.061(-0.385) + 0.006(53.752)) \\
&- \exp(0.645(0) + 0.061(-0.385) + 0.006(53.752)) \\
&= 1.212
\end{aligned}
$$

change in the dependent variable.

Interpretation

So, when going from a period of no stagnation to stagnation, the expected number of militarized disputes in a year increases by 1.212, holding all other variables constant.

Other Models Often Seen in IR

Other Models Often Seen in IR

- multinomial logit and probit

Other Models Often Seen in IR

- multinomial logit and probit
- negative binomial regression

Other Models Often Seen in IR

- multinomial logit and probit
- negative binomial regression
- tobit

Other Models Often Seen in IR

- multinomial logit and probit
- negative binomial regression
- tobit
- Heckman selection models

Other Models Often Seen in IR

- multinomial logit and probit
- negative binomial regression
- tobit
- Heckman selection models
- strategic models

Time for...

Coffee!