Lectures in Quantitative International Relations

Comparative Model Testing

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 Specification
Overview of Lecture 3

1 The Necessity of Being Comparative
Overview of Lecture 3

1. The Necessity of Being Comparative

2. Traditional Approaches
   - Nested Models
   - Nonnested Models
Overview of Lecture 3

1. The Necessity of Being Comparative
2. Traditional Approaches
   - Nested Models
   - Nonnested Models
3. Model Selection Criteria
   - The Kullback-Leibler Information Criteria
   - AIC and SIC
   - Example
Overview of Lecture 3

1. The Necessity of Being Comparative
2. Traditional Approaches
   - Nested Models
   - Nonnested Models
3. Model Selection Criteria
   - The Kullback-Leibler Information Criteria
   - AIC and SIC
   - Example
4. Model Selection Tests
   - The Vuong Test
   - The Distribution-Free Test
   - Example
Comparative model testing is necessary because political scientists (and other social scientists) have combined probabilistic falsificationism with a hypothetico-deductive approach to research.
Comparative model testing is necessary because political scientists (and other social scientists) have combined probabilistic falsificationism with a hypothetico-deductive approach to research.

Let’s spell out each in turn....
First some notation

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$</td>
<td>theory</td>
<td></td>
</tr>
<tr>
<td>$\land$</td>
<td>and (conjunction)</td>
<td></td>
</tr>
<tr>
<td>$K$</td>
<td>background conditions</td>
<td>$T \land K$</td>
</tr>
<tr>
<td>$H_0$</td>
<td>null hypothesis</td>
<td></td>
</tr>
<tr>
<td>$H_1$</td>
<td>alternative hypothesis</td>
<td></td>
</tr>
<tr>
<td>$\neg$</td>
<td>not</td>
<td>$\neg H_0$</td>
</tr>
<tr>
<td>$\rightarrow$</td>
<td>if...then</td>
<td>$A \rightarrow B$</td>
</tr>
<tr>
<td>$\leftrightarrow$</td>
<td>if and only if</td>
<td>$A \leftrightarrow B$</td>
</tr>
</tbody>
</table>
Classical or frequentist hypothesis testing is based on a probabilistic version of Popperian falsificationism.
Classical or frequentist hypothesis testing is based on a probabilistic version of Popperian falsificationism.

\[ H_0 \rightarrow \beta = 0 \]
\[ \beta \neq 0 \]

\[ \neg H_0 \]

where \( \beta \neq 0 \) means a p-value less than the significance level.
H-D simply means deriving a prediction from a theory and background conditions and then testing the prediction.
H-D simply means deriving a prediction from a theory and background conditions and then testing the prediction.

A qualitative example from Huth, Gelpi, and Bennett (1993):

Structural realism
Risk-acceptant leaders
Multipolarity

In a multipolar system, risk-acceptant leaders will be more likely to escalate
Now let’s combine statistical tests with H.D.

Step 1: Theory to hypothesis

\[ T \land K \rightarrow \neg H_0 \]

Step 2: Data to hypothesis

\[ H_0 \rightarrow \beta = 0 \]
\[ \beta \neq 0 \]

\[ \neg H_0 \]

Step 3: Hypothesis to theory

\[ T \land K \rightarrow \neg H_0 \]
\[ \neg H_0 \]
\[ \neg H_0 \]

\[ T \land K \]
Uh oh, a problem

The last step is false! Finding support for an alternative hypothesis implied by a theory does not mean that the theory is therefore true.
Uh oh, a problem

The last step is false! Finding support for an alternative hypothesis implied by a theory does not mean that the theory is therefore true.

Step 3: Hypothesis to theory

\[ T \land K \rightarrow \neg H_0 \]
\[ \neg H_0 \]
\[ \neg H_0 \rightarrow \neg H_0 \]

Why not? An infinite number of theories predict this alternative hypothesis.
Uh oh, a problem

The last step is false! Finding support for an alternative hypothesis implied by a theory does not mean that the theory is therefore true.

Step 3: Hypothesis to theory

\[ T \land K \rightarrow \neg H_0 \]

\[ \neg H_0 \]

\[ \neg H_0 \]

Why not? An infinite number of theories predict this alternative hypothesis.
Objection 1: we do something else in practice....

Step 1: Theory to hypothesis

\[ T \land K \rightarrow H_1 \]

Step 2: Data to hypothesis

\[ H_1 \rightarrow \text{Coefficient is correct} \]

Coefficient is correct

\[ H_1 \]

Step 3: Hypothesis to theory

\[ T \land K \rightarrow H_1 \]

\[ H_1 \]

\[ T \land K \]
Objection 2: being Bayesian solves the problem

Step 1: Theory to hypothesis

\[ T \land K \rightarrow H_1 \]

Step 2: Data to hypothesis

\[ H_1 \rightarrow P(H_1 | y) \text{ is high} \]
\[ P(H_1 | y) \text{ is high} \]

---

\( H_1 \)

Step 3: Hypothesis to theory

\[ T \land K \rightarrow H_1 \]
\[ H_1 \]

---

\( T \land K \)
Theory to data

\[ T \land K \rightarrow P(H_1|y) \text{ is high} \]

\[ P(H_1|y) \text{ is high} \]

\[ T \land K \]
So what is the problem?

When the hypothesis implies the data, \( H \rightarrow y \), the probability of the data given the hypothesis, \( P(y|H) \), is equal to one. Under these conditions, Bayes’ theorem simplifies to

\[
P(H|y) = \frac{P(H)P(y|H)}{P(y)} = \frac{P(H)}{P(y)}.
\]
So what is the problem?

When the hypothesis implies the data, $H \rightarrow y$, the probability of the data given the hypothesis, $P(y|H)$, is equal to one. Under these condition, Bayes’ theorem simplifies to

$$P(H|y) = \frac{P(H)P(y|H)}{P(y)} = \frac{P(H)}{P(y)}.$$

The inverse probability statement upon which we have pinned our hopes reduces to the ratio of the prior probability of the hypothesis to the probability of the data.
The biconditional to save the day!

If the alternative hypothesis is true if and only if the theory is true, and the hypothesis is true if and only if the data comes out right, then then the theory is true if and only if the data come out right.
The biconditional to save the day!

If the alternative hypothesis is true if and only if the theory is true, and the hypothesis is true if and only if the data comes out right, then then the theory is true if and only if the data come out right.

\[ T \land K \leftrightarrow H_1 \]

\[ H_1 \leftrightarrow D \]

\[ T \land K \leftrightarrow D \]
The biconditional to save the day!

If the alternative hypothesis is true if and only if the theory is true, and the hypothesis is true if and only if the data comes out right, then then the theory is true if and only if the data come out right.

\[
T \land K \iff H_1
\]

\[
H_1 \iff D
\]

\[
T \land K \iff D
\]

How does this help us?
The biconditional in action

Step 1: Theory to hypothesis

\[(T_1 \land K) \equiv (T_2 \land K) \leftrightarrow H_0\]

Step 2: Data to hypothesis

\[H_0 \leftrightarrow \tau \approx 0\]
\[\tau > 0\]
\[\neg H_0\]

Step 3: Hypothesis to theory

\[(T_1 \land K) \equiv (T_2 \land K) \leftrightarrow H_0\]
\[\neg H_0\]
\[\neg \neg H_0\]
\[\neg H_0\]
\[(T_1 \land K) > (T_2 \land K).\]
Why comparative model testing?

Being able to make “if and only if” statements is a necessary condition for learning. As our theories provide little in the way of such statements, we need to be comparative.
Why comparative model testing?

Being able to make “if and only if” statements is a necessary condition for learning. As our theories provide little in the way of such statements, we need to be comparative.

Now let’s look at how traditional approaches handle comparative model testing. There are two cases: nested and nonnested.
Definition: Nested models

Two models are *nested* if one model can be reduced to the other model by imposing a set of linear restrictions on the parameter vector.
The Necessity of Being Comparative
Traditional Approaches
Model Selection Criteria
Model Selection Tests

Definition: Nested models

Two models are *nested* if one model can be reduced to the other model by imposing a set of linear restrictions on the parameter vector.

Consider, for example, these two models,

Model 1: \[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon_1 \]
Model 2: \[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon_2 \]
Definition: Nested models

Two models are *nested* if one model can be reduced to the other model by imposing a set of linear restrictions on the parameter vector.

Consider, for example, these two models,

Model 1: \[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon_1 \]
Model 2: \[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon_2 \]

By setting \( \beta_3 = \beta_4 = 0 \) in model 2, we get model 1.
Models may also be nested in terms of their functional forms. Consider two common duration models:

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Hazard Function, $\lambda(t)$</th>
<th>Survival Function, $S(t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>$\lambda$</td>
<td>$S(t) = e^{-\lambda t}$</td>
</tr>
<tr>
<td>Weibull</td>
<td>$\lambda p(\lambda t)^{p-1}$</td>
<td>$S(t) = e^{-(\lambda t)^p}$</td>
</tr>
</tbody>
</table>
Nested models and functional form

Models may also be nested in terms of their functional forms. Consider two common duration models:

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Hazard Function, $\lambda(t)$</th>
<th>Survival Function, $S(t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>$\lambda$</td>
<td>$S(t) = e^{-\lambda t}$</td>
</tr>
<tr>
<td>Weibull</td>
<td>$\lambda p(\lambda t)^{p-1}$</td>
<td>$S(t) = e^{-(\lambda t)^p}$</td>
</tr>
</tbody>
</table>

The Weibull is the Exponential when $p = 1$. 
Tests for discriminating between nested models

Discriminating between nested models is easily accomplished using a variety of standard statistical techniques, such as,
Tests for discriminating between nested models

Discriminating between nested models is easily accomplished using a variety of standard statistical techniques, such as,

- $Z$-tests (for a single restriction in OLS or GLM),
Tests for discriminating between nested models

Discriminating between nested models is easily accomplished using a variety of standard statistical techniques, such as,

- Z-tests (for a single restriction in OLS or GLM),
- Likelihood ratio tests (for a single or multiple restrictions in GLM),
Tests for discriminating between nested models

Discriminating between nested models is easily accomplished using a variety of standard statistical techniques, such as,

- *Z*-tests (for a single restriction in OLS or GLM),
- Likelihood ratio tests (for a single or multiple restrictions in GLM),
- *F*-tests (for a single or multiple restrictions in OLS).
Example: Do nukes make a difference?

Dep. variable: escalation or not (Huth, Gelpi, and Bennett (1993))

|                              | Estimate | Std. Error | z value | Pr(>|z|) |
|------------------------------|----------|------------|---------|----------|
| (Intercept)                  | -1.3739  | 0.5334     | -2.58   | 0.0100   |
| Balance of forces            | 1.4628   | 0.7691     | 1.90    | 0.0572   |
| Def. vital interests         | -0.3653  | 0.2995     | -1.22   | 0.2225   |
| Chall. vital interests       | 0.6136   | 0.3125     | 1.96    | 0.0496   |
| Def. backed down             | 0.8359   | 0.3571     | 2.34    | 0.0192   |
| Chall. backed down           | -0.9565  | 0.4504     | -2.12   | 0.0337   |
| Def. other dispute           | 0.7511   | 0.3063     | 2.45    | 0.0142   |
| Chall. other dispute         | -0.1457  | 0.3029     | -0.48   | 0.6304   |

The log-likelihood is -56.99703 (df=8).
### Example: Do nukes make a difference?

Dep. variable: escalation or not (Huth, Gelpi, and Bennett (1993))

|                          | Estimate | Std. Error | z value | Pr(>|z|) |
|--------------------------|----------|------------|---------|----------|
| (Intercept)              | −0.8857  | 0.5913     | −1.50   | 0.1341   |
| Balance of forces        | 1.4650   | 0.8158     | 1.80    | 0.0725   |
| Secure 2nd strike        | −1.7609  | 0.4132     | −4.26   | 0.0000   |
| Def. vital interests     | −0.8871  | 0.3702     | −2.40   | 0.0166   |
| Chall. vital interests   | 0.8293   | 0.3651     | 2.27    | 0.0231   |
| Def. backed down         | 1.0274   | 0.4153     | 2.47    | 0.0134   |
| Chall. backed down       | −0.7090  | 0.4911     | −1.44   | 0.1488   |
| Def. other dispute       | 0.7257   | 0.3496     | 2.08    | 0.0379   |
| Chall. other dispute     | −0.0173  | 0.3430     | −0.05   | 0.9598   |

The log-likelihood is -45.55808 (df=9).
Example: R commands

```r
huth1 <- glm(outcome~dispbof+defint+chint+
+riwhimp+chwhimp+riothdis+chothdis,
family=binomial(link=probit))

huth2 <- glm(outcome~dispbof+rinukes+defint+
+chint+riwhimp+chwhimp+riothdis+
+chothdis,family=binomial
(link=probit))
```
Example: R commands

```r
huth1 <- glm(outcome~dispbof+defint+chint
+riwhimp+chwhimp+riothdis+chothdis,
family=binomial(link=probit))

huth2 <- glm(outcome~dispbof+rinukes+defint
+chint+riwhimp+chwhimp+riothdis
+chothdis,family=binomial
(link=probit))

addterm(huth1,huth2,test="Chisq")
```
Example: R commands

```r
huth1 <- glm(outcome~dispbof+defint+chint
  +riwhimp+chwhimp+riothdis+chothdis,
  family=binomial(link=probit))

huth2 <- glm(outcome~dispbof+rinukes+defint
  +chint+riwhimp+chwhimp+riothdis
  +chothdis, family=binomial
  (link=probit))

addterm(huth1,huth2,test="Chisq")

<table>
<thead>
<tr>
<th>Df</th>
<th>LRT</th>
<th>Pr(Chi)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rinukes</td>
<td>1</td>
<td>22.88</td>
</tr>
</tbody>
</table>
```

\[ LRT = -2 \times [ -45.55808 - (-56.99703)] = -22.8779 \]
Definition: Nonnested models

Two models are *nonnested* if one model cannot be reduced to the other model by imposing a set of linear restrictions on the parameter vector.
Definition: Nonnested models

Two models are *nonnested* if one model cannot be reduced to the other model by imposing a set of linear restrictions on the parameter vector.

Model 1: \[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon_1 \]
Model 2: \[ Y = \beta_3 + \beta_4 X_4 + \beta_5 X_5 + \epsilon_2 \]
Definition: Nonnested models

Two models are *nonnested* if one model cannot be reduced to the other model by imposing a set of linear restrictions on the parameter vector.

Model 1: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon_1$

Model 2: $Y = \beta_3 + \beta_4 X_4 + \beta_5 X_5 + \epsilon_2$

There is no set of linear restrictions that we can impose on either model 1 or model 2 that will give us the other model.
Nonnested models and functional form

Two models may also be nonnested in terms of their functional forms.
Nonnested models and functional form

Two models may also be nonnested in terms of their functional forms.

\[
\begin{align*}
\Pr(Y = 1) &= \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2) = \int_{-\infty}^{\beta_0+\beta_1 X_1+\beta_2 X_2} \phi(t) \, dt \\
\Pr(Y = 1) &= \Lambda(\beta_0 + \beta_1 X_1 + \beta_2 X_2) = \frac{e^{\beta_0+\beta_1 X_1+\beta_2 X_2}}{1 + e^{\beta_0+\beta_1 X_1+\beta_2 X_2}}
\end{align*}
\]
Two models may also be nonnested in terms of their functional forms.

\[
\Pr(Y = 1) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2) = \int_{-\infty}^{\beta_0+\beta_1 X_1 + \beta_2 X_2} \phi(t) \, dt
\]

\[
\Pr(Y = 1) = \Lambda(\beta_0 + \beta_1 X_1 + \beta_2 X_2) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2}}
\]

Even when they have the same covariates, probits and logits are nonnested.
The traditional approach: “Super” models

- Model 1 includes $X_1$, $X_2$, and $X_3$.
- Model 2 includes $X_3$, $X_4$, and $X_5$. 

Will any of the traditional approaches taken to nested models work here?
The traditional approach: “Super” models

- Model 1 includes $X_1$, $X_2$, and $X_3$.
- Model 2 includes $X_3$, $X_4$, and $X_5$.

The combined “super” model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \epsilon$$
The traditional approach: “Super” models

- Model 1 includes $X_1$, $X_2$, and $X_3$.
- Model 2 includes $X_3$, $X_4$, and $X_5$.

The combined “super” model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \epsilon$$

Will any of the traditional approaches taken to nested models work here?
The traditional approaches and “super” models

\[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \epsilon \]
The traditional approaches and “super” models

\[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \epsilon \]

- Z-tests?
The traditional approaches and “super” models

\[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \epsilon \]

- Z-tests?
- Likelihood ratio tests?
The traditional approaches and “super” models

\[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \epsilon \]

- Z-tests?
- Likelihood ratio tests?
- F-tests?
More about the f-test and nonnested models

Model 1: \[ Y = X\beta + \epsilon_1 \]
Model 2: \[ Y = Z\gamma + \epsilon_1 \]
More about the f-test and nonnested models

Model 1: \[ Y = X\beta + \epsilon_1 \]
Model 2: \[ Y = Z\gamma + \epsilon_1 \]

Let \( \tilde{X} \) be the variables in \( X \), but not in \( Z \).
More about the f-test and nonnested models

Model 1: \[ Y = X\beta + \epsilon_1 \]
Model 2: \[ Y = Z\gamma + \epsilon_1 \]

Let \( \tilde{X} \) be the variables in \( X \), but not in \( Z \).
Let \( \tilde{Z} \) be the variables in \( Z \), but not in \( X \).
More about the f-test and nonnested models

Model 1: \[ Y = X\beta + \epsilon_1 \]
Model 2: \[ Y = Z\gamma + \epsilon_1 \]

Let \( \hat{X} \) be the variables in \( X \), but not in \( Z \).
Let \( \hat{Z} \) be the variables in \( Z \), but not in \( X \).
Let \( W \) be the variables the two models have in common.
More about the f-test and nonnested models

Model 1: \( Y = X\beta + \epsilon_1 \)
Model 2: \( Y = Z\gamma + \epsilon_1 \)

Let \( \tilde{X} \) be the variables in \( X \), but not in \( Z \).
Let \( \tilde{Z} \) be the variables in \( Z \), but not in \( X \).
Let \( W \) be the variables the two models have in common.

Combined: \( Y = \tilde{X}\beta + \tilde{Z}\gamma + W\sigma + \epsilon \)
Combined: $Y = \tilde{X}\beta + \tilde{Z}\gamma + W\sigma + \epsilon$
More about the f-test and nonnested models

Combined: \[ Y = \tilde{X}\beta + \tilde{Z}\gamma + W\sigma + \epsilon \]

- Testing \( \beta \) or \( \gamma \) does not test the full models.
More about the f-test and nonnested models

Combined: $Y = \hat{X}\beta + \hat{Z}\gamma + W\sigma + \epsilon$

- Testing $\beta$ or $\gamma$ does not test the full models.
- Testing $\beta$ and $\sigma$ or $\gamma$ and $\sigma$ does not test model 1 against model 2.
Combined: $Y = \tilde{X}\beta + \tilde{Z}\gamma + W\sigma + \epsilon$

- Testing $\beta$ or $\gamma$ does not test the full models.
- Testing $\beta$ and $\sigma$ or $\gamma$ and $\sigma$ does not test model 1 against model 2.
- The f-test discriminates between either model 1 or model 2 and a hybrid model that is neither model 1 or model 2.
More about the f-test and nonnested models

Combined: \( Y = \tilde{X}\beta + \tilde{Z}\gamma + W\sigma + \epsilon \)

- Testing \( \beta \) or \( \gamma \) does not test the full models.
- Testing \( \beta \) and \( \sigma \) or \( \gamma \) and \( \sigma \) does not test model 1 against model 2.
- The f-test discriminates between either model 1 or model 2 and a hybrid model that is neither model 1 or model 2.
- The model, itself, is atheoretic.
Traditional approaches

So traditional approaches to model discrimination do not work for the nonnested case.
So traditional approaches to model discrimination do not work for the nonnested case.

While there are numerous techniques we might use, let's consider model selection criteria and model selection tests.
So traditional approaches to model discrimination do not work for the nonnested case.

While there are numerous techniques we might use, let’s consider model selection criteria and model selection tests.

Model selection criteria
Traditional approaches

So traditional approaches to model discrimination do not work for the nonnested case.

While there are numerous techniques we might use, let’s consider model selection criteria and model selection tests.

Model selection criteria

AIC (Akaike’s information criteria)
So traditional approaches to model discrimination do not work for the nonnested case.

While there are numerous techniques we might use, let’s consider model selection criteria and model selection tests.

**Model selection criteria**
- **AIC** (Akaike’s information criteria)
- **SIC** (Schwarz’s Bayesian information criteria)
Traditional approaches

So traditional approaches to model discrimination do not work for the nonnested case.

While there are numerous techniques we might use, let’s consider model selection criteria and model selection tests.

Model selection criteria
- AIC (Akaike’s information criteria)
- SIC (Schwarz’s Bayesian information criteria)

Model selection tests
Traditional approaches

So traditional approaches to model discrimination do not work for the nonnested case.

While there are numerous techniques we might use, let’s consider model selection criteria and model selection tests.

Model selection criteria
- AIC (Akaike’s information criteria)
- SIC (Schwarz’s Bayesian information criteria)

Model selection tests
- Vuong test and my distribution-free test.
Kullback-Leibler discrepancy

The criteria and tests we will look at are all based upon this quantity, which is similar to a measure of “distance.”
The criteria and tests we will look at are all based upon this quantity, which is similar to a measure of “distance.”

\[ I(g : f) = E_Y \ln \left( \frac{g(Y)}{f(Y)} \right) = \int_{-\infty}^{\infty} \ln \left( \frac{g(Y)}{f(Y)} \right) g(Y) dY \]
Kullback-Leibler discrepancy

The criteria and tests we will look at are all based upon this quantity, which is similar to a measure of “distance.”

\[
I(g : f) = E_Y \ln \left\{ \frac{g(Y)}{f(Y)} \right\} = \int_{-\infty}^{\infty} \ln \left\{ \frac{g(Y)}{f(Y)} \right\} g(Y) dY
\]

\[
I(g : f) > 0 \quad \text{for} \quad f \neq g
\]
\[
I(g : f) = 0 \quad \text{for} \quad f = g
\]
More on the Kullback-Leibler discrepancy

\[ I(g : f) = E_Y \ln \left\{ \frac{g(Y)}{f(Y|\theta)} \right\} \]
\[ = E_Y \ln g(Y) - E_Y \ln f(Y|\theta) \]

If \( g(Y) \) is the true sampling distribution of \( Y \) and \( f(Y|\theta) \) is a particular model, we want an \( f(Y|\theta) \) that is as large as possible to minimize the KL discrepancy. That is, we want a model that is as close to the true model as possible.
More on the Kullback-Leibler discrepancy

\[ l(g : f) = E_Y \ln \left\{ \frac{g(Y)}{f(Y|\theta)} \right\} \]

\[ = E_Y \ln g(Y) - E_Y \ln f(Y|\theta) \]

If \( g(Y) \) is the true sampling distribution of \( Y \) and \( f(Y|\theta) \) is a particular model, we want an \( f(Y|\theta) \) that is as large as possible to minimize the KL discrepancy. That is, we want a model that is as close to the true model as possible.
The Kullback-Leibler Information Criteria

Model Selection Criteria

Model selection criteria estimate $l(g : f) = E_Y \ln g(Y) - E_Y \ln f(Y)$. The model for which this estimate is minimized is the "best" model.

AIC: $-2 \ln (\text{maximum likelihood}) + 2 \times (\text{number of parameters})$

SIC: $-2 \ln (\text{maximum likelihood}) + \ln T \times (\text{number of parameters})$
Model selection criteria

\[ l(g : f) = E_Y \ln g(Y) - E_Y \ln f(Y) \]

Model selection criteria estimate \(-E_Y \ln f(Y)\). The model for which this estimate is minimized is the “best” model.
Model selection criteria

\[ I(g : f) = E_Y \ln g(Y) - E_Y \ln f(Y) \]

Model selection criteria estimate \(-E_Y \ln f(Y)\). The model for which this estimate is minimized is the “best” model.

AIC: \(-2 \ln(\text{maximum likelihood}) + 2(\text{number of parameters})\)
Model selection criteria

\[ l(g : f) = E_Y \ln g(Y) - E_Y \ln f(Y) \]

Model selection criteria estimate \(-E_Y \ln f(Y)\). The model for which this estimate is minimized is the “best” model.

**AIC:** \(-2 \ln(\text{maximum likelihood}) + 2(\text{number of parameters})\)

**SIC:** \(-2 \ln(\text{maximum likelihood}) + \ln T(\text{number of parameters})\)
Example: Realism v. rational deterrence

Huth, Gelpi, and Bennett (1993) test the relative explanatory power of structural realism and rational deterrence theory on the escalation of militarized disputes among great powers from 1816 to 1985.
Example: Realism v. rational deterrence

Huth, Gelpi, and Bennett (1993) test the relative explanatory power of structural realism and rational deterrence theory on the escalation of militarized disputes among great powers from 1816 to 1985.

**Structural realism**
Focuses on the attributes of the international system.
Example: Realism v. rational deterrence

Huth, Gelpi, and Bennett (1993) test the relative explanatory power of structural realism and rational deterrence theory on the escalation of militarized disputes among great powers from 1816 to 1985.

**Structural realism**
Focuses on the attributes of the international system.

**Rational deterrence**
Focuses on the resolve and relative military capabilities of adversarial states.
Example: Results on structural realism

Dependent variable: escalation or not

|                          | Estimate | Std. Error | z value | Pr(>|z|) |
|--------------------------|----------|------------|---------|----------|
| (Intercept)              | -1.5668  | 0.7777     | -2.01   | 0.0439   |
| System uncertainty 1     | 0.7772   | 0.1866     | 4.16    | 0.0000   |
| System size*risk         | -0.6900  | 0.2766     | -2.49   | 0.0126   |
| System uncertainty 2     | -0.0244  | 0.1760     | -0.14   | 0.8898   |
| System diffusion*risk    | 0.1750   | 0.2589     | 0.68    | 0.4990   |
| Risk-acceptant           | 0.8184   | 1.2580     | 0.65    | 0.5153   |
Example: Results on rational deterrence

Dependent variable: escalation or not

|                                | Estimate | Std. Error | z value | Pr(>|z|) |
|--------------------------------|----------|------------|---------|----------|
| (Intercept)                    | −0.8857  | 0.5913     | −1.50   | 0.1341   |
| Balance of forces              | 1.4650   | 0.8158     | 1.80    | 0.0725   |
| Secure 2nd strike              | −1.7609  | 0.4132     | −4.26   | 0.0000   |
| Def. vital interests           | −0.8871  | 0.3702     | −2.40   | 0.0166   |
| Chall. vital interests         | 0.8293   | 0.3651     | 2.27    | 0.0231   |
| Def. backed down               | 1.0274   | 0.4153     | 2.47    | 0.0134   |
| Chall. backed down             | −0.7090  | 0.4911     | −1.44   | 0.1488   |
| Def. other dispute             | 0.7257   | 0.3496     | 2.08    | 0.0379   |
| Chall. other dispute           | −0.0173  | 0.3430     | −0.05   | 0.9598   |
Example: Model comparison

```r
huth.test <- mod.sel(huth1, huth2)
```
Example: Model comparison

```r
huth.test <- mod.sel(huth1,huth2)
```

Selection Criteria

<table>
<thead>
<tr>
<th></th>
<th>Model One</th>
<th>Model Two</th>
</tr>
</thead>
<tbody>
<tr>
<td>LogLik</td>
<td>-55.933</td>
<td>-45.558</td>
</tr>
<tr>
<td>AIC</td>
<td>123.865</td>
<td>109.116</td>
</tr>
<tr>
<td>BIC</td>
<td>139.313</td>
<td>132.289</td>
</tr>
</tbody>
</table>
Problems with model selection criteria
Problems with model selection criteria

- Sampling properties are unknown.
Problems with model selection criteria

- Sampling properties are unknown.
- No statement of uncertainty.
Problems with model selection criteria

- Sampling properties are unknown.
- No statement of uncertainty.
- No ability to choose neither model.
Model selection tests

Model selection tests are also based on the Kullback-Leibler discrepancy. Written a little differently to emphasize the covariates:

$$KLIC \equiv E_0[\ln h_0(Y_i|X_i)] - E_0[\ln f(Y_i|X_i; \beta^*)]$$
Model selection tests are also based on the Kullback-Leibler discrepancy. Written a little differently to emphasize the covariates:

$$\text{KLIC} \equiv E_0[\ln h_0( Y_i | X_i)] - E_0[\ln f( Y_i | X_i; \beta_*)]$$

We want to minimize this “distance,” and we do that by choosing the model that maximizes $E_0[\ln f( Y_i | X_i; \beta_*)]$. 
Model selection tests are also based on the Kullback-Leibler discrepancy. Written a little differently to emphasize the covariates:

$$KLIC \equiv E_0[\ln h_0(Y_i|X_i)] - E_0[\ln f(Y_i|X_i; \beta^*)]$$

We want to minimize this “distance,” and we do that by choosing the model that maximizes $E_0[\ln f(Y_i|X_i; \beta^*)]$.

To compare two models, then, it is natural to look at the ratio of their likelihoods.
Log-likelihoods

Before going further, let’s consider the log-likelihood that is reported every time you run a generalized linear model. That log-likelihood is the sum of the log-likelihoods for each individual observation.
Before going further, let’s consider the log-likelihood that is reported every time you run a generalized linear model. That log-likelihood is the sum of the log-likelihoods for each individual observation.

### Log-likelihoods

<table>
<thead>
<tr>
<th>Observations</th>
<th>Log-likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs. 1</td>
<td>-0.54498</td>
</tr>
<tr>
<td>Obs. 2</td>
<td>-0.55547</td>
</tr>
<tr>
<td>Obs. 3</td>
<td>-0.42883</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Obs. 97</td>
<td>-0.18510</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>-55.93255</strong></td>
</tr>
</tbody>
</table>
Log-likelihoods

How do we get these individual log-likelihoods?
Log-likelihoods

How do we get these individual log-likelihoods?

Probit:

$$\log\text{-Lik}_i = y_i \times \log(\hat{p}) + (1 - y_i) \times \log(\hat{p})$$
Log-likelihoods

How do we get these individual log-likelihoods?

Probit:

\[
\text{log-Lik}_i = y_i \ast \log(\hat{\rho}) + (1 - y_i) \ast \log(\hat{\rho})
\]

Poisson:

\[
\text{log-Lik}_i = \beta'x_i \ast y_i - \exp(\beta'x_i) - \log(\Gamma(y_i + 1))
\]
Log-likelihood ratios

From the individual log-likelihoods, we can get the individual log-likelihood ratios by simply differencing.

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs. 1</td>
<td>-0.54498</td>
<td>-0.50802</td>
<td>-0.03696</td>
</tr>
<tr>
<td>Obs. 2</td>
<td>-0.55547</td>
<td>-0.37072</td>
<td>-0.18476</td>
</tr>
<tr>
<td>Obs. 3</td>
<td>-0.42883</td>
<td>-0.38157</td>
<td>-0.04726</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Obs. 97</td>
<td>-0.18510</td>
<td>-0.14135</td>
<td>-0.04375</td>
</tr>
<tr>
<td>Total</td>
<td>-55.93255</td>
<td>-45.55808</td>
<td>-10.37447</td>
</tr>
</tbody>
</table>
Log-likelihood ratios

If the two models are equivalent, the difference between their log-likelihoods should be zero. If one model is closer to the true specification, then the difference between their log-likelihoods will be significantly different from zero.
Log-likelihood ratios

If the two models are equivalent, the difference between their log-likelihoods should be zero. If one model is closer to the true specification, then the difference between their log-likelihoods will be significantly different from zero.

The difference between the two tests we’re going to talk about lies in whether we consider the average or the median.
Log-likelihood ratios

If the two models are equivalent, the difference between their log-likelihoods should be zero. If one model is closer to the true specification, then the difference between their log-likelihoods will be significantly different from zero.

The difference between the two tests we’re going to talk about lies in whether we consider the average or the median.

- **Vuong**: is the average individual log-likelihood ratio different from zero?
Log-likelihood ratios

If the two models are equivalent, the difference between their log-likelihoods should be zero. If one model is closer to the true specification, then the difference between their log-likelihoods will be significantly different from zero.

The difference between the two tests we’re going to talk about lies in whether we consider the average or the median.

- **Vuong**: is the average individual log-likelihood ratio different from zero?
- **Distribution-free**: is the median individual log-likelihood ratio different from zero?
Vuong test

Vuong proves under general conditions that the expected value given in the null hypothesis can be consistently estimated by \((1/n)\) times the likelihood ratio statistic,

\[
\frac{1}{n} LR_n(\hat{\beta}_n, \hat{\gamma}_n) \xrightarrow{a.s.} E_0 \left[ \ln \frac{f(Y_i|X_i; \beta_\star)}{g(Y_i|Z_i; \gamma_\star)} \right].
\]
Vuong proves under general conditions that the expected value given in the null hypothesis can be consistently estimated by \((1/n)\) times the likelihood ratio statistic,

\[
\frac{1}{n} LR_n(\hat{\beta}_n, \hat{\gamma}_n) \xrightarrow{a.s.} E_0 \left[ \ln \frac{f(Y_i|X_i; \beta_*)}{g(Y_i|Z_i; \gamma_*)} \right].
\]

After standardization, the likelihood ratio statistic is asymptotically normally distributed. So the actual test is,

under \(H_0\) : \[
\frac{LR_n(\hat{\beta}_n, \hat{\gamma}_n)}{(\sqrt{n})\hat{\omega}_n} \xrightarrow{D} N(0, 1)
\]
The variance is calculated in the normal way — the sum of the squares minus the square of the sum.
Vuong test

The variance is calculated in the normal way — the sum of the squares minus the square of the sum.

\[ \hat{\omega}_n^2 \equiv \frac{1}{n} \sum_{i=1}^{n} \left[ \ln \frac{f(Y_i|X_i; \hat{\beta}_n)}{g(Y_i|Z_i; \hat{\gamma}_n)} \right]^2 - \left[ \frac{1}{n} \sum_{i=1}^{n} \ln \frac{f(Y_i|X_i; \hat{\beta}_n)}{g(Y_i|Z_i; \hat{\gamma}_n)} \right]^2 \]
Vuong test

The variance is calculated in the normal way — the sum of the squares minus the square of the sum.

\[ \hat{\omega}_n^2 \equiv \frac{1}{n} \sum_{i=1}^{n} \left[ \ln \frac{f(Y_i|X_i; \hat{\beta}_n)}{g(Y_i|Z_i; \hat{\gamma}_n)} \right]^2 - \left[ \frac{1}{n} \sum_{i=1}^{n} \ln \frac{f(Y_i|X_i; \hat{\beta}_n)}{g(Y_i|Z_i; \hat{\gamma}_n)} \right]^2 \]

So what does the Vuong test amount to?
The variance is calculated in the normal way — the sum of the squares minus the square of the sum.

\[ \hat{\omega}_n^2 \equiv \frac{1}{n} \sum_{i=1}^{n} \left[ \ln \frac{f(Y_i | X_i; \hat{\beta}_n)}{g(Y_i | Z_i; \hat{\gamma}_n)} \right]^2 - \left[ \frac{1}{n} \sum_{i=1}^{n} \ln \frac{f(Y_i | X_i; \hat{\beta}_n)}{g(Y_i | Z_i; \hat{\gamma}_n)} \right]^2 \]

So what does the Vuong test amount to?
A paired z-test on the individual log-likelihoods!
### The Vuong Test

#### Example

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs. 1</td>
<td>-0.54498</td>
<td>-0.50802</td>
<td>-0.03696</td>
</tr>
<tr>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
<tr>
<td>Obs. 97</td>
<td>-0.18510</td>
<td>-0.14135</td>
<td>-0.04375</td>
</tr>
</tbody>
</table>

| Total            | -55.93255| -45.55808| -10.37447  |

**Z** = \( \frac{\text{Observed difference} - 0}{\text{S.D.diff}/\sqrt{n}} \)

\( Z \sim N(0,1) \)

**Clarke Lectures in Quantitative International Relations**
### Vuong test

The Vuong test is a distribution-free test for model selection. It is used to compare two non-nested models.

#### Model Selection Tests

- **Model Selection Criteria**
- **Model Selection Tests**
  - The Vuong Test
  - The Distribution-Free Test

### Example

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs. 1</td>
<td>-0.54498</td>
<td>-0.50802</td>
<td>-0.03696</td>
</tr>
<tr>
<td>Obs. 97</td>
<td>-0.18510</td>
<td>-0.14135</td>
<td>-0.04375</td>
</tr>
<tr>
<td>Total</td>
<td>-55.93255</td>
<td>-45.55808</td>
<td>-10.37447</td>
</tr>
</tbody>
</table>

The Vuong test statistic is given by:

$$ Z = \frac{\text{Observed difference} - 0}{\text{S.D.}_{\text{diff}}/\sqrt{n}} = \frac{LR_n(\hat{\beta}_n, \hat{\gamma}_n) - 0}{(\sqrt{n})\hat{\omega}_n} \sim N(0, 1) $$
We can always make the log-likelihood larger by adding additional variables to our model. Just like with AIC and BIC, we need a penalty adjustment for a model with too many variables.
The Necessity of Being Comparative
Traditional Approaches
Model Selection Criteria
Model Selection Tests

Vuong test: Adjustment

We can always make the log-likelihood larger by adding additional variables to our model. Just like with AIC and BIC, we need a penalty adjustment for a model with too many variables.

For the Vuong test, this adjustment is:

\[ L\tilde{R}_n(\hat{\beta}_n, \hat{\gamma}_n) \equiv LR_n(\hat{\beta}_n, \hat{\gamma}_n) - \left[ \left( \frac{p}{2} \right) \ln n - \left( \frac{q}{2} \right) \ln n \right], \]

where \( p \) and \( q \) are the number of parameters estimated in the two models. For the Huth et al. case, \( p = 6 \) and \( q = 9 \).
Vuong test: Problem

It can be shown that the Vuong test is very powerful if the underlying distribution of the individual log-likelihood ratios is normally distributed.
Vuong test: Problem

It can be shown that the Vuong test is very powerful if the underlying distribution of the individual log-likelihood ratios is normally distributed.

If the underlying distribution is not normal, then we can do better. If we measure the kurtosis of the distribution...

<table>
<thead>
<tr>
<th>Distance</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50</td>
</tr>
<tr>
<td>0.3</td>
<td>5.19</td>
</tr>
<tr>
<td>0.4</td>
<td>5.26</td>
</tr>
<tr>
<td>0.5</td>
<td>5.24</td>
</tr>
<tr>
<td>0.6</td>
<td>5.32</td>
</tr>
<tr>
<td>0.7</td>
<td>5.40</td>
</tr>
<tr>
<td>0.8</td>
<td>5.47</td>
</tr>
<tr>
<td>0.9</td>
<td>5.58</td>
</tr>
</tbody>
</table>
Vuong test: Problem

...it looks much more like a double exponential or Laplace distribution.
Distribution-free test

If faced with data from a double-exponential or Laplace distribution, you would replace the normal $z$ or $t$ test with a distribution-free test such as the sign test, the paired sign test, or a signed-rank test.
Distribution-free test

If faced with data from a double-exponential or Laplace distribution, you would replace the normal $z$ or $t$ test with a distribution-free test such as the sign test, the paired sign test, or a signed-rank test.

The paired sign test is used to test the null hypothesis that the probability of a random variable from the population of paired differences being greater than zero is equal to the probability of the random variable being less than zero.
Distribution-free test

If faced with data from a double-exponential or Laplace distribution, you would replace the normal $z$ or $t$ test with a distribution-free test such as the sign test, the paired sign test, or a signed-rank test.

The paired sign test is used to test the null hypothesis that the probability of a random variable from the population of paired differences being greater than zero is equal to the probability of the random variable being less than zero.

In our case,

$$H_0 : \Pr_0 \left[ \ln \frac{f(Y_i|X_i; \beta^*)}{g(Y_i|Z_i; \gamma^*)} > 0 \right] = 0.5.$$
Distribution-free test

Letting $d_i = \ln f(Y_i|X_i; \hat{\beta}_n) - \ln g(Y_i|Z_i; \hat{\gamma}_n)$, the test statistic is

$$B = \sum_{i=1}^{n} I(0, +\infty)(d_i),$$

which is simply the number of positive differences, and it is distributed Binomial with parameters $n$ and $\theta = 0.5$. 
Distribution-free test

Letting \( d_i = \ln f(Y_i|X_i; \hat{\beta}_n) - \ln g(Y_i|Z_i; \hat{\gamma}_n) \), the test statistic is

\[
B = \sum_{i=1}^{n} I_{(0, +\infty)}(d_i),
\]

which is simply the number of positive differences, and it is distributed Binomial with parameters \( n \) and \( \theta = 0.5 \).

If model \( f \) is “better” than model \( g \), \( B \) will be significantly larger than its expected value under the null hypothesis \( (n/2) \).
Distribution-free test: How to....

1. Run model $f$, saving the individual log-likelihoods.
2. Run model $g$, saving the individual log-likelihoods.
3. Compute the differences and count the number of positive and negative values.
4. The number of positive differences is distributed binomial ($n$, $p = \ldots$).
Distribution-free test: How to....

1. Run model $f$, saving the individual log-likelihoods.
Distribution-free test: How to....

1. Run model $f$, saving the individual log-likelihoods.
2. Run model $g$, saving the individual log-likelihoods.
Distribution-free test: How to....

1. Run model $f$, saving the individual log-likelihoods.
2. Run model $g$, saving the individual log-likelihoods.
3. Compute the differences and count the number of positive and negative values.
Distribution-free test: How to....

1. Run model $f$, saving the individual log-likelihoods.
2. Run model $g$, saving the individual log-likelihoods.
3. Compute the differences and count the number of positive and negative values.
4. The number of positive differences is distributed binomial($n, p = .5$).
As we are working with the individual log-likelihood ratios, we cannot apply the same correction to the “summed” log-likelihood ratio as Vuong did for his test.
Distribution-free test: Adjustment

As we are working with the individual log-likelihood ratios, we cannot apply the same correction to the “summed” log-likelihood ratio as Vuong did for his test.

We can, however, apply the *average* correction to the individual log-likelihood ratios. So we subtract the following factors from each individual log-likelihood:
As we are working with the individual log-likelihood ratios, we cannot apply the same correction to the “summed” log-likelihood ratio as Vuong did for his test.

We can, however, apply the average correction to the individual log-likelihood ratios. So we subtract the following factors from each individual log-likelihood:

- Model $f$: $[(p/2n) \ln n]$
- Model $g$: $[(q/2n) \ln n]$
Huth et al. (1993) data revisited

<table>
<thead>
<tr>
<th>Model One</th>
<th>Model Two</th>
</tr>
</thead>
<tbody>
<tr>
<td>LogLik</td>
<td>-55.933</td>
</tr>
<tr>
<td>AIC</td>
<td>123.865</td>
</tr>
<tr>
<td>BIC</td>
<td>139.313</td>
</tr>
</tbody>
</table>

Model Selection Tests

<table>
<thead>
<tr>
<th>Statistic</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vuong</td>
<td>0.435</td>
</tr>
<tr>
<td>Clarke</td>
<td>0.543</td>
</tr>
</tbody>
</table>

*The log-likelihoods for model two are subtracted from the log-likelihoods for model one.*
### Huth et al. (1993) data revisited

#### Selection criteria

<table>
<thead>
<tr>
<th></th>
<th>Model One</th>
<th>Model Two</th>
</tr>
</thead>
<tbody>
<tr>
<td>LogLik</td>
<td>-55.933</td>
<td>-45.558</td>
</tr>
<tr>
<td>AIC</td>
<td>123.865</td>
<td>109.116</td>
</tr>
<tr>
<td>BIC</td>
<td>139.313</td>
<td>132.289</td>
</tr>
</tbody>
</table>

The Vuong Test

<table>
<thead>
<tr>
<th>Statistic</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vuong</td>
<td>0.435</td>
</tr>
<tr>
<td>Clarke</td>
<td>0.543</td>
</tr>
</tbody>
</table>

*The log-likelihoods for model two are subtracted from the log-likelihoods for model one.*
Huth et al. (1993) data revisited

### Selection criteria

<table>
<thead>
<tr>
<th>Model</th>
<th>LogLik</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model One</td>
<td>-55.933</td>
<td>123.865</td>
<td>139.313</td>
</tr>
<tr>
<td>Model Two</td>
<td>-45.558</td>
<td>109.116</td>
<td>132.289</td>
</tr>
</tbody>
</table>

### Model Section Tests*

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Statistic</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vuong</td>
<td>-0.781</td>
<td>0.435</td>
</tr>
<tr>
<td>Clarke</td>
<td>45.000</td>
<td>0.543</td>
</tr>
</tbody>
</table>

*The log-likelihoods for model two are subtracted from the log-likelihoods for model one.
R commands

```r
huth1 <- glm(outcome~nuncp1+runcp13+nuncp2 +runcp23+risk23pm, family=binomial (link=probit))
```

Clarke
Lectures in Quantitative International Relations
The Necessity of Being Comparative

Traditional Approaches

Model Selection Criteria

Model Selection Tests

The Vuong Test

The Distribution-Free Test

Example

R commands

```r
huth1 <- glm(outcome~nuncp1+runcp13+nuncp2 +runcp23+risk23pm, family=binomial (link=probit))
huth2 <- glm(outcome~dispbof+rinukes+defint +chint+riwhimp+chwhimp+riothdis +chothdis, family=binomial (link=probit))
```
R commands

```r
huth1 <- glm(outcome~nuncp1+runcp13+nuncp2 + runcp23+risk23pm, family=binomial (link=probit))
huth2 <- glm(outcome~dispbof+rinukes+defint + chint+riwhimp+chwhimp+riothdis + chothdis, family=binomial (link=probit))

huth.test <- mod.sel(huth1,huth2)
```
R commands

huth1 <- glm(outcome~nuncp1+runcp13+nuncp2 +runcp23+risk23pm, family=binomial (link=probit))
huth2 <- glm(outcome~dispbof+rinukes+defint +chint+riwhimp+chwhimp+riothdis +chothdis, family=binomial (link=probit))

huth.test <- mod.sel(huth1,huth2)
summary(huth.test)
Test comparison: Getting it right

Difference in the probability of choosing the right model.
Test comparison: Getting it wrong

Difference in the probability of choosing the wrong model.

![Graph showing difference in power in the wrong direction versus distance of the alternative from the null for different sample sizes (N=50, N=100, N=200, N=500, N=1000).]
Test comparison: Balancing the good and bad

Linear combination of errors.

\[
\begin{align*}
0.5a(d) &= 0.5b(d) \\
0.33a(d) &= 0.67b(d) \\
0.17a(d) &= 0.83b(d) \\
0.09a(d) &= 0.91b(d) \\
0.06a(d) &= 0.94b(d)
\end{align*}
\]
The Necessity of Being Comparative

Traditional Approaches

Model Selection Criteria

Model Selection Tests

The Vuong Test

The Distribution-Free Test

Example

Take Home Message
Being comparative is a necessary condition for making reliable inferences.
Being comparative is a necessary condition for making reliable inferences.

Whether the rival models are nested or nonnested, there are techniques for discriminating between them.
Take Home Message

- Being comparative is a necessary condition for making reliable inferences.
- Whether the rival models are nested or nonnested, there are techniques for discriminating between them.
- The distribution-free test has greater power than the Vuong when the underlying distribution is not normal.