Advanced Quantitative Research Methodology, Lecture Notes: Multiple Equation Models[1]

Gary King
http://GKing.Harvard.Edu

April 20, 2013

Identification

Reading: Gary King. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. Ann Arbor: University of Michigan Press, 1998: Chapter 8.

Identification

Reading: Gary King. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. Ann Arbor: University of Michigan Press, 1998: Chapter 8.

Models that often don't make sense, even though it is hard to tell.

Example 1: Flat Likelihoods

Example 1: Flat Likelihoods

A (dumb) model:

Example 1: Flat Likelihoods

A (dumb) model:

$$Y_i \sim f_p(y_i|\lambda_i)$$

Example 1: Flat Likelihoods

A (dumb) model:

$$Y_i \sim f_p(y_i|\lambda_i)$$

$$\lambda_i = 1 + 0\beta$$

Example 1: Flat Likelihoods

A (dumb) model:

$$Y_i \sim f_p(y_i|\lambda_i)$$

$$\lambda_i = 1 + 0\beta$$

What do we know about β?

Example 1: Flat Likelihoods

A (dumb) model:

$$Y_i \sim f_p(y_i|\lambda_i)$$

$$\lambda_i = 1 + 0\beta$$

What do we know about β?

$$L(\lambda|y) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!}$$

Example 1: Flat Likelihoods

A (dumb) model:

$$Y_i \sim f_p(y_i|\lambda_i)$$

$$\lambda_i = 1 + 0\beta$$

What do we know about β?

$$L(\lambda|y) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!}$$

and the log-likelihood, with $(1 + 0\beta)$ substituted for λ_i:

Example 1: Flat Likelihoods

A (dumb) model:

$$Y_i \sim f_p(y_i|\lambda_i)$$

$$\lambda_i = 1 + 0\beta$$

What do we know about β?

$$L(\lambda|y) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!}$$

and the log-likelihood, with $(1 + 0\beta)$ substituted for λ_i:

$$\ln L(\beta|y) = \sum_{i=1}^{n} \{-(0\beta + 1) - y_i \ln(0\beta + 1)\}$$

Example 1: Flat Likelihoods

A (dumb) model:

$$Y_i \sim f_p(y_i|\lambda_i)$$

$$\lambda_i = 1 + 0\beta$$

What do we know about β?

$$L(\lambda|y) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!}$$

and the log-likelihood, with $(1 + 0\beta)$ substituted for λ_i:

$$\begin{aligned}
\ln L(\beta|y) &= \sum_{i=1}^{n} \{-(0\beta + 1) - y_i \ln(0\beta + 1)\} \\
&= \sum_{i=1}^{n} -1
\end{aligned}$$

Example 1: Flat Likelihoods

A (dumb) model:

$$Y_i \sim f_p(y_i|\lambda_i)$$

$$\lambda_i = 1 + 0\beta$$

What do we know about β?

$$L(\lambda|y) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!}$$

and the log-likelihood, with $(1 + 0\beta)$ substituted for λ_i:

$$\begin{aligned}\ln L(\beta|y) &= \sum_{i=1}^{n} \{-(0\beta + 1) - y_i \ln(0\beta + 1)\} \\ &= \sum_{i=1}^{n} -1 \\ &= -n\end{aligned}$$

Example 1: Flat Likelihoods

Example 1: Flat Likelihoods

1. An identified likelihood has a unique maximum.

Example 1: Flat Likelihoods

1. An identified likelihood has a unique maximum.
2. A likelihood function with a flat region or plateau at the maximum is not identified.

Example 1: Flat Likelihoods

1. An identified likelihood has a unique maximum.
2. A likelihood function with a flat region or plateau at the maximum is not identified.
3. A likelihood with a plateau can be informative, but a unique MLE doesn't exist

Example 2: Non-unique Reparameterization

Example 2: Non-unique Reparameterization

A model

Example 2: Non-unique Reparameterization

A model

$$Y_i \sim f_N(y_i|\mu_i, \sigma^2)$$

Example 2: Non-unique Reparameterization

A model

$$
\begin{aligned}
Y_i &\sim f_N(y_i|\mu_i, \sigma^2) \\
\mu_i &= x_{1i}\beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3,
\end{aligned}
$$

Example 2: Non-unique Reparameterization

A model

$$
\begin{aligned}
&Y_i \sim f_N(y_i|\mu_i, \sigma^2)\\
&\mu_i = x_{1i}\beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3, \qquad \text{where } x_{2i} = x_{3i}
\end{aligned}
$$

Example 2: Non-unique Reparameterization

A model

$$
\begin{aligned}
Y_i &\sim f_N(y_i|\mu_i, \sigma^2) \\
\mu_i &= x_{1i}\beta_1 + {\color{red}x_{2i}}\beta_2 + {\color{red}x_{3i}}\beta_3, \qquad \text{where } {\color{red}x_{2i}} = {\color{red}x_{3i}} \\
&= x_{1i}\beta_1 + {\color{red}x_{2i}}(\beta_2 + \beta_3)
\end{aligned}
$$

Example 2: Non-unique Reparameterization

A model

$$
\begin{aligned}
Y_i &\sim f_N(y_i|\mu_i, \sigma^2) \\
\mu_i &= x_{1i}\beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3, \qquad \text{where } x_{2i} = x_{3i} \\
&= x_{1i}\beta_1 + x_{2i}(\beta_2 + \beta_3)
\end{aligned}
$$

What is the (unique) MLE of β_2 and β_3? Different parameter values lead to the same values of μ and thus the same likelihood values:

Example 2: Non-unique Reparameterization

A model

$$
\begin{aligned}
Y_i &\sim f_N(y_i|\mu_i, \sigma^2) \\
\mu_i &= x_{1i}\beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3, \qquad \text{where } x_{2i} = x_{3i} \\
&= x_{1i}\beta_1 + x_{2i}(\beta_2 + \beta_3)
\end{aligned}
$$

What is the (unique) MLE of β_2 and β_3? Different parameter values lead to the same values of μ and thus the same likelihood values:

$$\mu_i = x_{1i}\beta_1 + x_{2i}(5 + 3)$$

Example 2: Non-unique Reparameterization

A model

$$\begin{aligned} Y_i &\sim f_N(y_i|\mu_i, \sigma^2) \\ \mu_i &= x_{1i}\beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3, \qquad \text{where } x_{2i} = x_{3i} \\ &= x_{1i}\beta_1 + x_{2i}(\beta_2 + \beta_3) \end{aligned}$$

What is the (unique) MLE of β_2 and β_3? Different parameter values lead to the same values of μ and thus the same likelihood values:

$$\mu_i = x_{1i}\beta_1 + x_{2i}(5 + 3)$$
$$\mu_i = x_{1i}\beta_1 + x_{2i}(3 + 5)$$

Example 2: Non-unique Reparameterization

A model

$$
\begin{aligned}
Y_i &\sim f_N(y_i|\mu_i, \sigma^2) \\
\mu_i &= x_{1i}\beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3, \qquad \text{where } x_{2i} = x_{3i} \\
&= x_{1i}\beta_1 + x_{2i}(\beta_2 + \beta_3)
\end{aligned}
$$

What is the (unique) MLE of β_2 and β_3? Different parameter values lead to the same values of μ and thus the same likelihood values:

$$
\begin{aligned}
\mu_i &= x_{1i}\beta_1 + x_{2i}(5+3) \\
\mu_i &= x_{1i}\beta_1 + x_{2i}(3+5) \\
\mu_i &= x_{1i}\beta_1 + x_{2i}(7+1)
\end{aligned}
$$

Example 2: Non-unique Reparameterization

A model

$$
\begin{aligned}
Y_i &\sim f_N(y_i|\mu_i, \sigma^2) \\
\mu_i &= x_{1i}\beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3, \qquad \text{where } x_{2i} = x_{3i} \\
&= x_{1i}\beta_1 + x_{2i}(\beta_2 + \beta_3)
\end{aligned}
$$

What is the (unique) MLE of β_2 and β_3? Different parameter values lead to the same values of μ and thus the same likelihood values:

$$
\begin{aligned}
\mu_i &= x_{1i}\beta_1 + x_{2i}(5 + 3) \\
\mu_i &= x_{1i}\beta_1 + x_{2i}(3 + 5) \\
\mu_i &= x_{1i}\beta_1 + x_{2i}(7 + 1)
\end{aligned}
$$

So $\{\beta_2 = 2, \beta_3 = 5\}$ gives the same likelihood as $\{\beta_2 = 5, \beta_3 = 2\}$.

Introduction to Multiple Equation Models

Introduction to Multiple Equation Models

1. Let Y_i be an $N \times 1$ vector for $i = 1, \ldots, n$

Introduction to Multiple Equation Models

1. Let Y_i be an $N \times 1$ vector for $i = 1, \ldots, n$
2. Y_i is jointly distributed

Introduction to Multiple Equation Models

1. Let Y_i be an $N \times 1$ vector for $i = 1, \ldots, n$
2. Y_i is jointly distributed

$$Y_i \sim f(y_i|\theta_i, \alpha)$$

Introduction to Multiple Equation Models

1. Let Y_i be an $N \times 1$ vector for $i = 1, \ldots, n$
2. Y_i is jointly distributed

$$Y_i \sim f(y_i|\theta_i, \alpha)$$

3. θ_i is $N \times 1$, α is usually $N \times N$

Introduction to Multiple Equation Models

1. Let Y_i be an $N \times 1$ vector for $i = 1, \ldots, n$
2. Y_i is jointly distributed
$$Y_i \sim f(y_i|\theta_i, \alpha)$$
3. θ_i is $N \times 1$, α is usually $N \times N$
4. Systematic components:

1. Let Y_i be an $N \times 1$ vector for $i = 1, \ldots, n$
2. Y_i is jointly distributed

$$Y_i \sim f(y_i|\theta_i, \alpha)$$

3. θ_i is $N \times 1$, α is usually $N \times N$
4. Systematic components:

$$\theta_{1i} = g_1(x_{1i}, \beta_1)$$

Introduction to Multiple Equation Models

1. Let Y_i be an $N \times 1$ vector for $i = 1, \ldots, n$
2. Y_i is jointly distributed
$$Y_i \sim f(y_i|\theta_i, \alpha)$$
3. θ_i is $N \times 1$, α is usually $N \times N$
4. Systematic components:
$$\theta_{1i} = g_1(x_{1i}, \beta_1)$$
$$\theta_{2i} = g_2(x_{2i}, \beta_2)$$

Introduction to Multiple Equation Models

1. Let Y_i be an $N \times 1$ vector for $i = 1, \ldots, n$
2. Y_i is jointly distributed

$$Y_i \sim f(y_i|\theta_i, \alpha)$$

3. θ_i is $N \times 1$, α is usually $N \times N$
4. Systematic components:

$$\begin{aligned}
\theta_{1i} &= g_1(x_{1i}, \beta_1) \\
\theta_{2i} &= g_2(x_{2i}, \beta_2) \\
&\vdots \\
\theta_{Ni} &= g_N(x_{Ni}, \beta_N)
\end{aligned}$$

When are Multiple Equation Models different from N separate equation-by-equation models?

When the elements of Y_i are (conditional on X),

When are Multiple Equation Models different from N separate equation-by-equation models?

When the elements of Y_i are (conditional on X),

1. Stochastically dependent, or

When are Multiple Equation Models different from N separate equation-by-equation models?

When the elements of Y_i are (conditional on X),

1. Stochastically dependent, or
2. Parametrically dependent (shared parameters)

When are Multiple Equation Models different from N separate equation-by-equation models?

When the elements of Y_i are (conditional on X),

1. Stochastically dependent, or
2. Parametrically dependent (shared parameters)

Example and proof:

When are Multiple Equation Models different from N separate equation-by-equation models?

When the elements of Y_i are (conditional on X),

1. Stochastically dependent, or
2. Parametrically dependent (shared parameters)

Example and proof:

Suppose no ancillary parameters, and $N = 2$. The joint density:

When are Multiple Equation Models different from N separate equation-by-equation models?

When the elements of Y_i are (conditional on X),

1. Stochastically dependent, or
2. Parametrically dependent (shared parameters)

Example and proof:

Suppose no ancillary parameters, and $N = 2$. The joint density:

$$f(y|\theta) = \prod_{i=1}^{n} f(y_{1i}, y_{2i}|\theta_{1i}, \theta_{2i})$$

When are Multiple Equation Models different from N separate equation-by-equation models?

When the elements of Y_i are (conditional on X),

1. Stochastically dependent, or
2. Parametrically dependent (shared parameters)

Example and proof:

Suppose no ancillary parameters, and $N = 2$. The joint density:

$$f(y|\theta) = \prod_{i=1}^{n} f(y_{1i}, y_{2i}|\theta_{1i}, \theta_{2i})$$

(BTW, you now know how to form the likelihood for multiple equation models!)

Assuming stochastic independence lets us factor f:

Assuming stochastic independence lets us factor f:

$$P(y|\theta) = \prod_{i=1}^{n} f(y_{1i}, y_{2i}|\theta_{1i}, \theta_{2i})$$

Assuming stochastic independence lets us factor f:

$$\begin{aligned} P(y|\theta) &= \prod_{i=1}^{n} f(y_{1i}, y_{2i}|\theta_{1i}, \theta_{2i}) \\ &= \prod_{i=1}^{n} f(y_{1i}|\theta_{1i}) f(y_{2i}|\theta_{2i}) \end{aligned}$$

Assuming stochastic independence lets us factor f:

$$\begin{aligned} P(y|\theta) &= \prod_{i=1}^{n} f(y_{1i}, y_{2i}|\theta_{1i}, \theta_{2i}) \\ &= \prod_{i=1}^{n} f(y_{1i}|\theta_{1i}) f(y_{2i}|\theta_{2i}) \end{aligned}$$

with log-likelihood

Assuming stochastic independence lets us factor f:

$$
\begin{aligned}
P(y|\theta) &= \prod_{i=1}^{n} f(y_{1i}, y_{2i}|\theta_{1i}, \theta_{2i}) \\
&= \prod_{i=1}^{n} f(y_{1i}|\theta_{1i}) f(y_{2i}|\theta_{2i})
\end{aligned}
$$

with log-likelihood

$$
\ln L(\theta_1, \theta_2|y) = \sum_{i-1}^{n} \ln f(y_{1i}|\theta_{1i}) + \sum_{i=1}^{n} \ln f(y_{2i}|\theta_{2i})
$$

Assuming stochastic independence lets us factor f:

$$
\begin{aligned}
P(y|\theta) &= \prod_{i=1}^{n} f(y_{1i}, y_{2i}|\theta_{1i}, \theta_{2i}) \\
&= \prod_{i=1}^{n} f(y_{1i}|\theta_{1i}) f(y_{2i}|\theta_{2i})
\end{aligned}
$$

with log-likelihood

$$
\ln L(\theta_1, \theta_2|y) = \sum_{i-1}^{n} \ln f(y_{1i}|\theta_{1i}) + \sum_{i=1}^{n} \ln f(y_{2i}|\theta_{2i})
$$

Also assume parametric independence, and you can estimate the equations separately.

Seemingly Unrelated Regression Models (SURM)

Seemingly Unrelated Regression Models (SURM)

The model:

Seemingly Unrelated Regression Models (SURM)

The model:

1. Y_i, μ_i are $N \times 1$;

Seemingly Unrelated Regression Models (SURM)

The model:

1. Y_i, μ_i are $N \times 1$; Σ is $N \times N$

Seemingly Unrelated Regression Models (SURM)

The model:

1. Y_i, μ_i are $N \times 1$; Σ is $N \times N$; X_{ij} is $k_j \times 1$

Seemingly Unrelated Regression Models (SURM)

The model:

1. Y_i, μ_i are $N \times 1$; Σ is $N \times N$; X_{ij} is $k_j \times 1$; β_j is $k_j \times 1$

Seemingly Unrelated Regression Models (SURM)

The model:

1. Y_i, μ_i are $N \times 1$; Σ is $N \times N$; X_{ij} is $k_j \times 1$; β_j is $k_j \times 1$
2. $Y_i \sim N(y_i|\mu_i, \Sigma)$

Seemingly Unrelated Regression Models (SURM)

The model:

1. Y_i, μ_i are $N \times 1$; Σ is $N \times N$; X_{ij} is $k_j \times 1$; β_j is $k_j \times 1$
2. $Y_i \sim N(y_i|\mu_i, \Sigma)$
3. $\mu_{ij} = X_{ij}\beta_j$ for equation j, $j = 1, \ldots, N$

Seemingly Unrelated Regression Models (SURM)

The model:

1. Y_i, μ_i are $N \times 1$; Σ is $N \times N$; X_{ij} is $k_j \times 1$; β_j is $k_j \times 1$
2. $Y_i \sim N(y_i | \mu_i, \Sigma)$
3. $\mu_{ij} = X_{ij}\beta_j$ for equation j, $j = 1, \ldots, N$

Likelihood:

Seemingly Unrelated Regression Models (SURM)

The model:

1. Y_i, μ_i are $N \times 1$; Σ is $N \times N$; X_{ij} is $k_j \times 1$; β_j is $k_j \times 1$
2. $Y_i \sim N(y_i|\mu_i, \Sigma)$
3. $\mu_{ij} = X_{ij}\beta_j$ for equation j, $j = 1, \ldots, N$

Likelihood:

$$L(\beta, \Sigma) = \prod_{i=1}^{n} N(y_i|\mu_i, \Sigma)$$

Seemingly Unrelated Regression Models (SURM)

The model:

1. Y_i, μ_i are $N \times 1$; Σ is $N \times N$; X_{ij} is $k_j \times 1$; β_j is $k_j \times 1$
2. $Y_i \sim N(y_i|\mu_i, \Sigma)$
3. $\mu_{ij} = X_{ij}\beta_j$ for equation j, $j = 1, \ldots, N$

Likelihood:

$$L(\beta, \Sigma) = \prod_{i=1}^{n} N(y_i|\mu_i, \Sigma)$$

Seemingly Unrelated Regression Models (SURM)

The model:

1. Y_i, μ_i are $N \times 1$; Σ is $N \times N$; X_{ij} is $k_j \times 1$; β_j is $k_j \times 1$
2. $Y_i \sim N(y_i|\mu_i, \Sigma)$
3. $\mu_{ij} = X_{ij}\beta_j$ for equation j, $j = 1, \ldots, N$

Likelihood:

$$\begin{aligned} L(\beta, \Sigma) &= \prod_{i=1}^{n} N(y_i|\mu_i, \Sigma) \\ &= \prod_{i=1}^{n} (2\pi)^{-1} |\Sigma|^{-1/2} \exp\left[-\frac{1}{2}(y_i - \mu_i)'\Sigma^{-1}(y_i - \mu_i)\right] \end{aligned}$$

Notes:

Notes:

1. Programming in R is more complicated since `Y` is now a $n \times N$ matrix instead of an $n \times 1$ vector

Notes:

1. Programming in R is more complicated since `Y` is now a $n \times N$ matrix instead of an $n \times 1$ vector
2. Some computational tricks exist to make estimation a lot faster

Notes:

1. Programming in R is more complicated since `Y` is now a $n \times N$ matrix instead of an $n \times 1$ vector
2. Some computational tricks exist to make estimation a lot faster
3. If conditional on X, the Y's are stochastically independent of each other, and the β's are parametrically independent of each other, then SURM = equation-by-equation LS.

Notes:

1. Programming in R is more complicated since `Y` is now a $n \times N$ matrix instead of an $n \times 1$ vector
2. Some computational tricks exist to make estimation a lot faster
3. If conditional on X, the Y's are stochastically independent of each other, and the β's are parametrically independent of each other, then SURM = equation-by-equation LS.
4. Normally:

Notes:

1. Programming in R is more complicated since `Y` is now a $n \times N$ matrix instead of an $n \times 1$ vector
2. Some computational tricks exist to make estimation a lot faster
3. If conditional on X, the Y's are stochastically independent of each other, and the β's are parametrically independent of each other, then SURM = equation-by-equation LS.
4. Normally:
   (a) stochastic independence $[P(a,b) = P(a)P(b)] \implies$

Notes:

1. Programming in R is more complicated since `Y` is now a $n \times N$ matrix instead of an $n \times 1$ vector
2. Some computational tricks exist to make estimation a lot faster
3. If conditional on X, the Y's are stochastically independent of each other, and the β's are parametrically independent of each other, then SURM = equation-by-equation LS.
4. Normally:
   (a) stochastic independence $[P(a,b) = P(a)P(b)] \implies$
   (b) mean independence $[E(ab) = E(a)E(b)] \implies$

Notes:

1. Programming in R is more complicated since `Y` is now a $n \times N$ matrix instead of an $n \times 1$ vector
2. Some computational tricks exist to make estimation a lot faster
3. If conditional on X, the Y's are stochastically independent of each other, and the β's are parametrically independent of each other, then SURM = equation-by-equation LS.
4. Normally:
   (a) stochastic independence $[P(a, b) = P(a)P(b)] \implies$
   (b) mean independence $[E(ab) = E(a)E(b)] \implies$
   (c) uncorrelatedness (no linear relationship) $[\text{Corr}(a, b) = 0]$.

Notes:

1. Programming in R is more complicated since `Y` is now a $n \times N$ matrix instead of an $n \times 1$ vector
2. Some computational tricks exist to make estimation a lot faster
3. If conditional on X, the Y's are stochastically independent of each other, and the β's are parametrically independent of each other, then SURM = equation-by-equation LS.
4. Normally:
   (a) stochastic independence $[P(a, b) = P(a)P(b)] \implies$
   (b) mean independence $[E(ab) = E(a)E(b)] \implies$
   (c) uncorrelatedness (no linear relationship) $[\text{Corr}(a, b) = 0]$.
5. For the Normal, uncorrelatedness $\implies$ stochastic independence

Notes:

1. Programming in R is more complicated since `Y` is now a $n \times N$ matrix instead of an $n \times 1$ vector
2. Some computational tricks exist to make estimation a lot faster
3. If conditional on X, the Y's are stochastically independent of each other, and the β's are parametrically independent of each other, then SURM = equation-by-equation LS.
4. Normally:
   (a) stochastic independence $[P(a,b) = P(a)P(b)] \implies$
   (b) mean independence $[E(ab) = E(a)E(b)] \implies$
   (c) uncorrelatedness (no linear relationship) $[\text{Corr}(a,b) = 0]$.
5. For the Normal, uncorrelatedness $\implies$ stochastic independence
6. In the special case of the normal, identical explanatory variables also mean SURM = equation-by-equation LS.

Notes:

1. Programming in R is more complicated since `Y` is now a $n \times N$ matrix instead of an $n \times 1$ vector
2. Some computational tricks exist to make estimation a lot faster
3. If conditional on X, the Y's are stochastically independent of each other, and the β's are parametrically independent of each other, then SURM = equation-by-equation LS.
4. Normally:
   (a) stochastic independence $[P(a,b) = P(a)P(b)] \implies$
   (b) mean independence $[E(ab) = E(a)E(b)] \implies$
   (c) uncorrelatedness (no linear relationship) $[\text{Corr}(a,b) = 0]$.
5. For the Normal, uncorrelatedness $\implies$ stochastic independence
6. In the special case of the normal, identical explanatory variables also mean SURM = equation-by-equation LS.
7. $\implies$ identification of extra parameters in multiple equation models with identical X's comes solely from model assumptions

Reciprocal Causation

Reciprocal Causation

Stochastic component:

Reciprocal Causation

Stochastic component:

$$(Y_{1i}, Y_{2i}) \sim N(y_{1i}, y_{2i} | \mu_{1i}, \mu_{2i}, \sigma_1, \sigma_2, \sigma_{12})$$

Reciprocal Causation

Stochastic component:

$$(Y_{1i}, Y_{2i}) \sim N(y_{1i}, y_{2i} | \mu_{1i}, \mu_{2i}, \sigma_1, \sigma_2, \sigma_{12})$$

Systematic component:

Reciprocal Causation

Stochastic component:

$$(Y_{1i}, Y_{2i}) \sim N(y_{1i}, y_{2i}|\mu_{1i}, \mu_{2i}, \sigma_1, \sigma_2, \sigma_{12})$$

Systematic component:

$$\text{Vote:} \quad \mu_{1i} = {\color{red}x_{1i}}\beta_1 + {\color{blue}x_{2i}}\beta_2 + \mu_{2i}\beta_3$$

Reciprocal Causation

Stochastic component:

$$(Y_{1i}, Y_{2i}) \sim N(y_{1i}, y_{2i} | \mu_{1i}, \mu_{2i}, \sigma_1, \sigma_2, \sigma_{12})$$

Systematic component:

$$\begin{aligned} \text{Vote:} \quad & \mu_{1i} = x_{1i}\beta_1 + x_{2i}\beta_2 + \mu_{2i}\beta_3 \\ \text{PID:} \quad & \mu_{2i} = x_{1i}\gamma_1 + x_{3i}\gamma_2 + \mu_{1i}\gamma_3 \end{aligned}$$

Reciprocal Causation

Stochastic component:

$$(Y_{1i}, Y_{2i}) \sim N(y_{1i}, y_{2i} | \mu_{1i}, \mu_{2i}, \sigma_1, \sigma_2, \sigma_{12})$$

Systematic component:

$$\begin{aligned} \text{Vote:} \quad & \mu_{1i} = x_{1i}\beta_1 + x_{2i}\beta_2 + \mu_{2i}\beta_3 \\ \text{PID:} \quad & \mu_{2i} = x_{1i}\gamma_1 + x_{3i}\gamma_2 + \mu_{1i}\gamma_3 \end{aligned}$$

Where,

Reciprocal Causation

Stochastic component:

$$(Y_{1i}, Y_{2i}) \sim N(y_{1i}, y_{2i} | \mu_{1i}, \mu_{2i}, \sigma_1, \sigma_2, \sigma_{12})$$

Systematic component:

$$\begin{aligned} \text{Vote:} \quad & \mu_{1i} = x_{1i}\beta_1 + x_{2i}\beta_2 + \mu_{2i}\beta_3 \\ \text{PID:} \quad & \mu_{2i} = x_{1i}\gamma_1 + x_{3i}\gamma_2 + \mu_{1i}\gamma_3 \end{aligned}$$

Where,
x_1 demographics

Reciprocal Causation

Stochastic component:

$$(Y_{1i}, Y_{2i}) \sim N(y_{1i}, y_{2i} | \mu_{1i}, \mu_{2i}, \sigma_1, \sigma_2, \sigma_{12})$$

Systematic component:

$$\begin{aligned} \text{Vote:} \quad & \mu_{1i} = x_{1i}\beta_1 + x_{2i}\beta_2 + \mu_{2i}\beta_3 \\ \text{PID:} \quad & \mu_{2i} = x_{1i}\gamma_1 + x_{3i}\gamma_2 + \mu_{1i}\gamma_3 \end{aligned}$$

Where,

x_1 demographics

x_2 candidate characteristics (affecting vote but not PID)

Reciprocal Causation

Stochastic component:

$$(Y_{1i}, Y_{2i}) \sim N(y_{1i}, y_{2i} | \mu_{1i}, \mu_{2i}, \sigma_1, \sigma_2, \sigma_{12})$$

Systematic component:

$$\begin{aligned} \text{Vote:} \quad & \mu_{1i} = x_{1i}\beta_1 + x_{2i}\beta_2 + \mu_{2i}\beta_3 \\ \text{PID:} \quad & \mu_{2i} = x_{1i}\gamma_1 + x_{3i}\gamma_2 + \mu_{1i}\gamma_3 \end{aligned}$$

Where,

x_1 demographics

x_2 candidate characteristics (affecting vote but not PID)

x_3 parents PID (affecting PID but not vote)

Reciprocal Causation

Likelihood:

Reciprocal Causation

Likelihood:

$$f(y|\mu,\Sigma) = \prod_{i=1}^{n} N(y_i|\mu_i,\Sigma)$$

Reciprocal Causation

Likelihood:

$$
\begin{aligned}
f(y|\mu,\Sigma) &= \prod_{i=1}^{n} N(y_i|\mu_i,\Sigma) \\
&= \prod_{i-1}^{n} (2\pi)^{N/2}|\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(y_i-\mu_i)'\Sigma^{-1}(y_i-\mu_i)\right\}
\end{aligned}
$$

Reciprocal Causation

Likelihood:

$$f(y|\mu,\Sigma) = \prod_{i=1}^{n} N(y_i|\mu_i,\Sigma)$$

$$= \prod_{i-1}^{n} (2\pi)^{N/2}|\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(y_i-\mu_i)'\Sigma^{-1}(y_i-\mu_i)\right\}$$

Where,

Reciprocal Causation

Likelihood:

$$f(y|\mu,\Sigma) = \prod_{i=1}^{n} N(y_i|\mu_i,\Sigma)$$
$$= \prod_{i-1}^{n} (2\pi)^{N/2} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(y_i-\mu_i)'\Sigma^{-1}(y_i-\mu_i)\right\}$$

Where,

$$y_i = (y_{1i}, y_{2i})',$$

Reciprocal Causation

Likelihood:

$$f(y|\mu,\Sigma) = \prod_{i=1}^{n} N(y_i|\mu_i,\Sigma)$$
$$= \prod_{i-1}^{n} (2\pi)^{N/2}|\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(y_i-\mu_i)'\Sigma^{-1}(y_i-\mu_i)\right\}$$

Where,

$$y_i = (y_{1i}, y_{2i})',$$
$$\mu_i = (\mu_{1i}, \mu_{2i})'$$

Reciprocal Causation

Likelihood:

$$
\begin{aligned}
f(y|\mu,\Sigma) &= \prod_{i=1}^{n} N(y_i|\mu_i,\Sigma) \\
&= \prod_{i-1}^{n} (2\pi)^{N/2}|\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(y_i-\mu_i)'\Sigma^{-1}(y_i-\mu_i)\right\}
\end{aligned}
$$

Where,

$$
\begin{aligned}
y_i &= (y_{1i}, y_{2i})', \\
\mu_i &= (\mu_{1i}, \mu_{2i})' \\
\Sigma &= \begin{pmatrix} \sigma_1 & \sigma_{12} \\ \sigma_{12} & \sigma_2 \end{pmatrix}
\end{aligned}
$$

How to substitute in the systematic component?

How to substitute in the systematic component?

1. Standard models are reparameterized

How to substitute in the systematic component?

1. Standard models are reparameterized
2. Time series models are recursively reparameterized

How to substitute in the systematic component?

1. Standard models are reparameterized
2. Time series models are recursively reparameterized
3. This model requires *multiple equation reparameterization*.

$$\mu_{1i} = x_{1i}\beta_1 + x_{2i}\beta_2 + \mu_{2i}\beta_3$$

$$
\begin{aligned}
\mu_{1i} &= x_{1i}\beta_1 + x_{2i}\beta_2 + \mu_{2i}\beta_3 \\
&= x_{1i}\beta_1 + x_{2i}\beta_2 + (x_{1i}\gamma_1 + x_{3i}\gamma_2 + \mu_{1i}\gamma_3)\beta_3
\end{aligned}
$$

$$
\begin{aligned}
\mu_{1i} &= x_{1i}\beta_1 + x_{2i}\beta_2 + \mu_{2i}\beta_3 \\
&= x_{1i}\beta_1 + x_{2i}\beta_2 + (x_{1i}\gamma_1 + x_{3i}\gamma_2 + \mu_{1i}\gamma_3)\beta_3 \\
&= x_{1i}\beta_1 + x_{2i}\beta_2 + x_{1i}\gamma_1\beta_3 + x_{3i}\gamma_2\beta_3 + \mu_{1i}\gamma_3\beta_3
\end{aligned}
$$

$$
\begin{aligned}
\mu_{1i} &= x_{1i}\beta_1 + x_{2i}\beta_2 + \mu_{2i}\beta_3 \\
&= x_{1i}\beta_1 + x_{2i}\beta_2 + (x_{1i}\gamma_1 + x_{3i}\gamma_2 + \mu_{1i}\gamma_3)\beta_3 \\
&= x_{1i}\beta_1 + x_{2i}\beta_2 + x_{1i}\gamma_1\beta_3 + x_{3i}\gamma_2\beta_3 + \mu_{1i}\gamma_3\beta_3 \\
&= \left(\frac{1}{1-\gamma_3\beta_3}\right)[x_{1i}\beta_1 + (x_{1i}\gamma_1 + x_{3i}\gamma_2)\beta_3 + x_{2i}\beta_2]
\end{aligned}
$$

$$
\begin{aligned}
\mu_{1i} &= x_{1i}\beta_1 + x_{2i}\beta_2 + \mu_{2i}\beta_3 \\
&= x_{1i}\beta_1 + x_{2i}\beta_2 + (x_{1i}\gamma_1 + x_{3i}\gamma_2 + \mu_{1i}\gamma_3)\beta_3 \\
&= x_{1i}\beta_1 + x_{2i}\beta_2 + x_{1i}\gamma_1\beta_3 + x_{3i}\gamma_2\beta_3 + \mu_{1i}\gamma_3\beta_3 \\
&= \left(\frac{1}{1-\gamma_3\beta_3}\right)[x_{1i}\beta_1 + (x_{1i}\gamma_1 + x_{3i}\gamma_2)\beta_3 + x_{2i}\beta_2]
\end{aligned}
$$

$$
\mu_{2i} = x_{1i}\gamma_1 + x_{3i}\gamma_2 + \mu_{1i}\gamma_3
$$

$$
\begin{aligned}
\mu_{1i} &= x_{1i}\beta_1 + x_{2i}\beta_2 + \mu_{2i}\beta_3 \\
&= x_{1i}\beta_1 + x_{2i}\beta_2 + (x_{1i}\gamma_1 + x_{3i}\gamma_2 + \mu_{1i}\gamma_3)\beta_3 \\
&= x_{1i}\beta_1 + x_{2i}\beta_2 + x_{1i}\gamma_1\beta_3 + x_{3i}\gamma_2\beta_3 + \mu_{1i}\gamma_3\beta_3 \\
&= \left(\frac{1}{1-\gamma_3\beta_3}\right)[x_{1i}\beta_1 + (x_{1i}\gamma_1 + x_{3i}\gamma_2)\beta_3 + x_{2i}\beta_2]
\end{aligned}
$$

$$
\begin{aligned}
\mu_{2i} &= x_{1i}\gamma_1 + x_{3i}\gamma_2 + \mu_{1i}\gamma_3 \\
&= x_{1i}\gamma_1 + x_{3i}\gamma_2 + (x_{1i}\beta_1 + x_{2i}\beta_2 + \mu_{2i}\beta_3)\gamma_3
\end{aligned}
$$

$$\begin{aligned}
\mu_{1i} &= x_{1i}\beta_1 + x_{2i}\beta_2 + \mu_{2i}\beta_3 \\
&= x_{1i}\beta_1 + x_{2i}\beta_2 + (x_{1i}\gamma_1 + x_{3i}\gamma_2 + \mu_{1i}\gamma_3)\beta_3 \\
&= x_{1i}\beta_1 + x_{2i}\beta_2 + x_{1i}\gamma_1\beta_3 + x_{3i}\gamma_2\beta_3 + \mu_{1i}\gamma_3\beta_3 \\
&= \left(\frac{1}{1-\gamma_3\beta_3}\right)[x_{1i}\beta_1 + (x_{1i}\gamma_1 + x_{3i}\gamma_2)\beta_3 + x_{2i}\beta_2]
\end{aligned}$$

$$\begin{aligned}
\mu_{2i} &= x_{1i}\gamma_1 + x_{3i}\gamma_2 + \mu_{1i}\gamma_3 \\
&= x_{1i}\gamma_1 + x_{3i}\gamma_2 + (x_{1i}\beta_1 + x_{2i}\beta_2 + \mu_{2i}\beta_3)\gamma_3 \\
&= x_{1i}\gamma_1 + x_{3i}\gamma_2 + (x_{1i}\beta_1 + x_{2i}\beta_2)\gamma_3 + \mu_{2i}\beta_3\gamma_3
\end{aligned}$$

$$
\begin{aligned}
\mu_{1i} &= x_{1i}\beta_1 + x_{2i}\beta_2 + \mu_{2i}\beta_3 \\
&= x_{1i}\beta_1 + x_{2i}\beta_2 + (x_{1i}\gamma_1 + x_{3i}\gamma_2 + \mu_{1i}\gamma_3)\beta_3 \\
&= x_{1i}\beta_1 + x_{2i}\beta_2 + x_{1i}\gamma_1\beta_3 + x_{3i}\gamma_2\beta_3 + \mu_{1i}\gamma_3\beta_3 \\
&= \left(\frac{1}{1-\gamma_3\beta_3}\right)[x_{1i}\beta_1 + (x_{1i}\gamma_1 + x_{3i}\gamma_2)\beta_3 + x_{2i}\beta_2]
\end{aligned}
$$

$$
\begin{aligned}
\mu_{2i} &= x_{1i}\gamma_1 + x_{3i}\gamma_2 + \mu_{1i}\gamma_3 \\
&= x_{1i}\gamma_1 + x_{3i}\gamma_2 + (x_{1i}\beta_1 + x_{2i}\beta_2 + \mu_{2i}\beta_3)\gamma_3 \\
&= x_{1i}\gamma_1 + x_{3i}\gamma_2 + (x_{1i}\beta_1 + x_{2i}\beta_2)\gamma_3 + \mu_{2i}\beta_3\gamma_3 \\
&= \left(\frac{1}{1-\beta_3\gamma_3}\right)[x_{1i}\gamma_1 + x_{3i}\gamma_2 + (x_{1i}\beta_1 + x_{2i}\beta_2)\gamma_3]
\end{aligned}
$$

Suppose we drop X_2 and X_3

Suppose we drop X_2 and X_3

$$\mu_{1i} = \left(\frac{1}{1-\gamma_3\beta_3}\right)[x_{1i} + x_{1i}\gamma_1\beta_3]$$

Suppose we drop X_2 and X_3

$$\mu_{1i} = \left(\frac{1}{1-\gamma_3\beta_3}\right)[x_{1i} + x_{1i}\gamma_1\beta_3]$$
$$= x_{1i}\left(\frac{\beta_1+\gamma_1\beta_3}{1-\gamma_3\beta_3}\right)$$

Suppose we drop X_2 and X_3

$$
\begin{aligned}
\mu_{1i} &= \left(\frac{1}{1-\gamma_3\beta_3}\right)[x_{1i} + x_{1i}\gamma_1\beta_3] \\
&= x_{1i}\left(\frac{\beta_1 + \gamma_1\beta_3}{1-\gamma_3\beta_3}\right)
\end{aligned}
$$

Result:

Suppose we drop X_2 and X_3

$$\begin{aligned}\mu_{1i} &= \left(\frac{1}{1-\gamma_3\beta_3}\right)[x_{1i} + x_{1i}\gamma_1\beta_3] \\ &= x_{1i}\left(\frac{\beta_1 + \gamma_1\beta_3}{1-\gamma_3\beta_3}\right)\end{aligned}$$

Result:

- Not identified

Suppose we drop X_2 and X_3

$$
\begin{aligned}
\mu_{1i} &= \left(\frac{1}{1-\gamma_3\beta_3}\right)[x_{1i} + x_{1i}\gamma_1\beta_3] \\
&= x_{1i}\left(\frac{\beta_1+\gamma_1\beta_3}{1-\gamma_3\beta_3}\right)
\end{aligned}
$$

Result:

- Not identified
- $\{\beta_1 = 1,\ \gamma_1 = 30,\ \beta_3 = 555,\ \gamma_3 = -30\} \implies \mu_{1i} = -x_i.$

Suppose we drop X_2 and X_3

$$
\begin{aligned}
\mu_{1i} &= \left(\frac{1}{1-\gamma_3\beta_3}\right)[x_{1i} + x_{1i}\gamma_1\beta_3] \\
&= x_{1i}\left(\frac{\beta_1+\gamma_1\beta_3}{1-\gamma_3\beta_3}\right)
\end{aligned}
$$

Result:

- Not identified
- $\{\beta_1 = 1,\ \gamma_1 = 30,\ \beta_3 = 555,\ \gamma_3 = -30\} \implies \mu_{1i} = -x_i.$
- $\{\beta_1 = 1,\ \gamma_1 = 1,\ \beta_3 = -999,\ \gamma_3 = -1\} \implies \mu_{1i} = -x_i.$

Suppose we drop X_2 and X_3

$$\mu_{1i} = \left(\frac{1}{1-\gamma_3\beta_3}\right)[x_{1i} + x_{1i}\gamma_1\beta_3]$$
$$= x_{1i}\left(\frac{\beta_1 + \gamma_1\beta_3}{1-\gamma_3\beta_3}\right)$$

Result:

- Not identified
- $\{\beta_1 = 1, \gamma_1 = 30, \beta_3 = 555, \gamma_3 = -30\} \implies \mu_{1i} = -x_i$.
- $\{\beta_1 = 1, \gamma_1 = 1, \beta_3 = -999, \gamma_3 = -1\} \implies \mu_{1i} = -x_i$.
- As long as $\beta_1 = 1$ and $\gamma_1 = -\gamma_3$, for any value of β_3, $\mu_{1i} = -x_i$.

Suppose we drop X_2 and X_3

$$\begin{aligned}\mu_{1i} &= \left(\frac{1}{1-\gamma_3\beta_3}\right)[x_{1i} + x_{1i}\gamma_1\beta_3] \\ &= x_{1i}\left(\frac{\beta_1+\gamma_1\beta_3}{1-\gamma_3\beta_3}\right)\end{aligned}$$

Result:

- Not identified
- $\{\beta_1 = 1,\ \gamma_1 = 30,\ \beta_3 = 555,\ \gamma_3 = -30\} \implies \mu_{1i} = -x_i$.
- $\{\beta_1 = 1,\ \gamma_1 = 1,\ \beta_3 = -999,\ \gamma_3 = -1\} \implies \mu_{1i} = -x_i$.
- As long as $\beta_1 = 1$ and $\gamma_1 = -\gamma_3$, for any value of β_3, $\mu_{1i} = -x_i$.
- All results are highly sensitive to X_2 and X_3

Multinomial Choice Models

Multinomial Choice Models

1. $k \geq 2$ nominal choices, from which one is chosen.

Multinomial Choice Models

1. $k \geq 2$ nominal choices, from which one is chosen.
2. Example: choice among candidates by voters; dishes at a restaurant by customers; war/peace/limited war by countries, etc

Multinomial Choice Models

1. $k \geq 2$ nominal choices, from which one is chosen.
2. Example: choice among candidates by voters; dishes at a restaurant by customers; war/peace/limited war by countries, etc
3. Generalizations of (binary) logit and probit models

Multinomial Choice Models

1. $k \geq 2$ nominal choices, from which one is chosen.
2. Example: choice among candidates by voters; dishes at a restaurant by customers; war/peace/limited war by countries, etc
3. Generalizations of (binary) logit and probit models

Multinomial Probit

Multinomial Choice Models

1. $k \geq 2$ nominal choices, from which one is chosen.
2. Example: choice among candidates by voters; dishes at a restaurant by customers; war/peace/limited war by countries, etc
3. Generalizations of (binary) logit and probit models

Multinomial Probit

$$U_i^* \sim N(u_i^* | \mu_i, \Sigma)$$

Multinomial Choice Models

1. $k \geq 2$ nominal choices, from which one is chosen.
2. Example: choice among candidates by voters; dishes at a restaurant by customers; war/peace/limited war by countries, etc
3. Generalizations of (binary) logit and probit models

Multinomial Probit

$$U_i^* \sim N(u_i^*|\mu_i, \Sigma)$$
$$\mu_{ij} = x_{ij}\beta_j$$

Multinomial Choice Models

1. $k \geq 2$ nominal choices, from which one is chosen.
2. Example: choice among candidates by voters; dishes at a restaurant by customers; war/peace/limited war by countries, etc
3. Generalizations of (binary) logit and probit models

Multinomial Probit

$$U_i^* \sim N(u_i^*|\mu_i, \Sigma)$$
$$\mu_{ij} = x_{ij}\beta_j$$

with observation mechanism:

Multinomial Choice Models

1. $k \geq 2$ nominal choices, from which one is chosen.
2. Example: choice among candidates by voters; dishes at a restaurant by customers; war/peace/limited war by countries, etc
3. Generalizations of (binary) logit and probit models

Multinomial Probit

$$U_i^* \sim N(u_i^* | \mu_i, \Sigma)$$
$$\mu_{ij} = x_{ij}\beta_j$$

with observation mechanism:

$$Y_{ij} = \begin{cases} 1 & \text{if } U_{ij}^* > U_{ij'}^* \; \forall \; j \neq j' \\ 0 & \text{otherwise} \end{cases}$$

The stochastic component:

The stochastic component:

$$\Pr(Y_{ij} = 1) = \pi_{ij}, \qquad \text{s.t.} \qquad \sum_{j=1}^{J} \pi_{ij} = 1$$

The stochastic component:

$$\Pr(Y_{ij} = 1) = \pi_{ij}, \qquad \text{s.t.} \qquad \sum_{j=1}^{J} \pi_{ij} = 1$$

Systematic component. Let $Y_{ij}^* = U_{ij}^* - U_{ij'}^*$, so the observation mechanism is

The stochastic component:

$$\Pr(Y_{ij} = 1) = \pi_{ij}, \qquad \text{s.t.} \qquad \sum_{j=1}^{J} \pi_{ij} = 1$$

Systematic component. Let $Y_{ij}^* = U_{ij}^* - U_{ij'}^*$, so the observation mechanism is

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

The stochastic component:

$$\Pr(Y_{ij} = 1) = \pi_{ij}, \qquad \text{s.t.} \qquad \sum_{j=1}^{J} \pi_{ij} = 1$$

Systematic component. Let $Y^*_{ij} = U^*_{ij} - U^*_{ij'}$, so the observation mechanism is

$$Y_{ij} = \begin{cases} 1 & \text{if } Y^*_{ij} > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\pi_{ij} = \Pr(y_{ij} = 1)$$

The stochastic component:

$$\Pr(Y_{ij} = 1) = \pi_{ij}, \qquad \text{s.t.} \qquad \sum_{j=1}^{J} \pi_{ij} = 1$$

Systematic component. Let $Y^*_{ij} = U^*_{ij} - U^*_{ij'}$, so the observation mechanism is

$$Y_{ij} = \begin{cases} 1 & \text{if } Y^*_{ij} > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} \pi_{ij} &= \Pr(y_{ij} = 1) \\ &= \Pr(Y^*_{i1} \leq 0, \ldots, {\color{red} Y^*_{ij} > 0}, \ldots, Y^*_{iJ} \leq 0) \end{aligned}$$

The stochastic component:

$$\Pr(Y_{ij} = 1) = \pi_{ij}, \qquad \text{s.t.} \qquad \sum_{j=1}^{J} \pi_{ij} = 1$$

Systematic component. Let $Y_{ij}^* = U_{ij}^* - U_{ij'}^*$, so the observation mechanism is

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

$$\begin{aligned} \pi_{ij} &= \Pr(y_{ij} = 1) \\ &= \Pr(Y_{i1}^* \leq 0, \ldots, Y_{ij}^* > 0, \ldots, Y_{iJ}^* \leq 0) \\ &= \int_{-\infty}^{0} \cdots \int_{0}^{\infty} \cdots \int_{-\infty}^{0} N(y|\mu_i, \Sigma) dy_{i1} \cdots dy_{ij} \cdots dy_{iJ} \end{aligned}$$

Computational and Estimation issues

Computational and Estimation issues

1. No analytical solution is known to the integral

Computational and Estimation issues

1. No analytical solution is known to the integral
2. Doing it numerically with > 4 or 5 choices would take forever

Computational and Estimation issues

1. No analytical solution is known to the integral
2. Doing it numerically with > 4 or 5 choices would take forever
3. The problem has been solved to at least 9-12 choices by simulation (simple version: draw from normal and count fraction in regions)

Computational and Estimation issues

1. No analytical solution is known to the integral
2. Doing it numerically with > 4 or 5 choices would take forever
3. The problem has been solved to at least 9-12 choices by simulation (simple version: draw from normal and count fraction in regions)
4. All elements of Σ are not identified.

Computational and Estimation issues

1. No analytical solution is known to the integral
2. Doing it numerically with > 4 or 5 choices would take forever
3. The problem has been solved to at least 9-12 choices by simulation (simple version: draw from normal and count fraction in regions)
4. All elements of Σ are not identified.
5. The entire model is only weakly identified

Computational and Estimation issues

1. No analytical solution is known to the integral
2. Doing it numerically with > 4 or 5 choices would take forever
3. The problem has been solved to at least 9-12 choices by simulation (simple version: draw from normal and count fraction in regions)
4. All elements of Σ are not identified.
5. The entire model is only weakly identified
6. Model is a straightforward generalization of SURM, or of univariate probit

From Binary Logit to Multinomial Logit

From Binary Logit to Multinomial Logit

1. Suppose the data are Y_{ij} for group i ($i = 1, \ldots, n$) and person within group j ($j = 1, 2$ for simplicity, but it generalizes).

From Binary Logit to Multinomial Logit

1. Suppose the data are Y_{ij} for group i ($i = 1, \ldots, n$) and person within group j ($j = 1, 2$ for simplicity, but it generalizes).
2. Suppose also that there's a single 1 (e.g., among heterosexual couples, does the husband or wife balance the checkbook, assuming one of the two does?):

From Binary Logit to Multinomial Logit

1. Suppose the data are Y_{ij} for group i ($i = 1, \dots, n$) and person within group j ($j = 1, 2$ for simplicity, but it generalizes).
2. Suppose also that there's a single 1 (e.g., among heterosexual couples, does the husband or wife balance the checkbook, assuming one of the two does?):

$$\sum_{j=1}^{2} Y_{ij} = 1 \qquad \text{for all } j$$

From Binary Logit to Multinomial Logit

1. Suppose the data are Y_{ij} for group i ($i = 1, \ldots, n$) and person within group j ($j = 1, 2$ for simplicity, but it generalizes).
2. Suppose also that there's a single 1 (e.g., among heterosexual couples, does the husband or wife balance the checkbook, assuming one of the two does?):

$$\sum_{j=1}^{2} Y_{ij} = 1 \qquad \text{for all } j$$

3. The usual binary logit gives the *unconditional* probability:

From Binary Logit to Multinomial Logit

1. Suppose the data are Y_{ij} for group i ($i = 1, \ldots, n$) and person within group j ($j = 1, 2$ for simplicity, but it generalizes).
2. Suppose also that there's a single 1 (e.g., among heterosexual couples, does the husband or wife balance the checkbook, assuming one of the two does?):

$$\sum_{j=1}^{2} Y_{ij} = 1 \qquad \text{for all } j$$

3. The usual binary logit gives the *unconditional* probability:

$$\pi_{ij} = \frac{1}{1 + e^{-X_{ij}\beta}}$$

From Binary Logit to Multinomial Logit

1. Suppose the data are Y_{ij} for group i ($i = 1, \ldots, n$) and person within group j ($j = 1, 2$ for simplicity, but it generalizes).
2. Suppose also that there's a single 1 (e.g., among heterosexual couples, does the husband or wife balance the checkbook, assuming one of the two does?):

$$\sum_{j=1}^{2} Y_{ij} = 1 \qquad \text{for all } j$$

3. The usual binary logit gives the *unconditional* probability:

$$\pi_{ij} = \frac{1}{1 + e^{-X_{ij}\beta}} = \frac{e^{X_{ij}\beta}}{1 + e^{X_{ij}\beta}}$$

From Binary Logit to Multinomial Logit

1. Suppose the data are Y_{ij} for group i ($i = 1, \dots, n$) and person within group j ($j = 1, 2$ for simplicity, but it generalizes).
2. Suppose also that there's a single 1 (e.g., among heterosexual couples, does the husband or wife balance the checkbook, assuming one of the two does?):

$$\sum_{j=1}^{2} Y_{ij} = 1 \qquad \text{for all } j$$

3. The usual binary logit gives the *unconditional* probability:

$$\pi_{ij} = \frac{1}{1 + e^{-X_{ij}\beta}} = \frac{e^{X_{ij}\beta}}{1 + e^{X_{ij}\beta}}$$

where $X_{ij}\beta = \alpha + \beta x_{ij}$.

From Binary Logit to Multinomial Logit

1. Suppose the data are Y_{ij} for group i ($i = 1, \ldots, n$) and person within group j ($j = 1, 2$ for simplicity, but it generalizes).
2. Suppose also that there's a single 1 (e.g., among heterosexual couples, does the husband or wife balance the checkbook, assuming one of the two does?):

$$\sum_{j=1}^{2} Y_{ij} = 1 \qquad \text{for all } j$$

3. The usual binary logit gives the *unconditional* probability:

$$\pi_{ij} = \frac{1}{1 + e^{-X_{ij}\beta}} = \frac{e^{X_{ij}\beta}}{1 + e^{X_{ij}\beta}}$$

   where $X_{ij}\beta = \alpha + \beta x_{ij}$.
4. Applying binary logit: telling the computer you have $2n$ observations (husbands and wives) when you really only have n (families) — known as the double barreled data extender!

5. We instead want the *conditional* probability (i.e., the prob it is the husband given that it is either the husband or wife) and so use the usual rule: $\Pr(A|B) = \Pr(AB)/\Pr(B)$. Thus,

5. We instead want the *conditional* probability (i.e., the prob it is the husband given that it is either the husband or wife) and so use the usual rule: $\Pr(A|B) = \Pr(AB)/\Pr(B)$. Thus,

$$\Pr(Y_{i1} = 1|Y_{i1} + Y_{i2} = 1) = \frac{\Pr(Y_{i1} = 1 \text{ and } Y_{i1} + Y_{i2} = 1)}{\Pr(Y_{i1} + Y_{i2} = 1)}$$

5. We instead want the *conditional* probability (i.e., the prob it is the husband given that it is either the husband or wife) and so use the usual rule: $\Pr(A|B) = \Pr(AB)/\Pr(B)$. Thus,

$$\begin{aligned}\Pr(Y_{i1} = 1|Y_{i1} + Y_{i2} = 1) &= \frac{\Pr(Y_{i1} = 1 \text{ and } Y_{i1} + Y_{i2} = 1)}{\Pr(Y_{i1} + Y_{i2} = 1)} \\ &= \frac{\Pr(Y_{i1} = 1)\Pr(Y_{i2} = 0)}{\Pr(Y_{i1} = 1)\Pr(Y_{i2} = 0) + \Pr(Y_{i1} = 0)\Pr(Y_{i2} = 1)}\end{aligned}$$

5. We instead want the *conditional* probability (i.e., the prob it is the husband given that it is either the husband or wife) and so use the usual rule: $\Pr(A|B) = \Pr(AB)/\Pr(B)$. Thus,

$$
\begin{aligned}
\Pr(Y_{i1} = 1 | Y_{i1} + Y_{i2} = 1) &= \frac{\Pr(Y_{i1} = 1 \text{ and } Y_{i1} + Y_{i2} = 1)}{\Pr(Y_{i1} + Y_{i2} = 1)} \\
&= \frac{\Pr(Y_{i1} = 1)\Pr(Y_{i2} = 0)}{\Pr(Y_{i1} = 1)\Pr(Y_{i2} = 0) + \Pr(Y_{i1} = 0)\Pr(Y_{i2} = 1)} \\
&= \frac{\left(\frac{e^{X_{i1}\beta}}{1 + e^{X_{i1}\beta}}\right)\left(\frac{1}{1 + e^{X_{i2}\beta}}\right)}{\left(\frac{e^{X_{i1}\beta}}{1 + e^{X_{i1}\beta}}\right)\left(\frac{1}{1 + e^{X_{i2}\beta}}\right) + \left(\frac{1}{1 + e^{X_{i1}\beta}}\right)\left(\frac{e^{X_{i2}\beta}}{1 + e^{X_{i2}\beta}}\right)}
\end{aligned}
$$

And so all the denominators drop out, leaving

And so all the denominators drop out, leaving

$$= \frac{e^{X_{i1}\beta}}{e^{X_i 1\beta} + e^{X_{i2}\beta}}$$

And so all the denominators drop out, leaving

$$= \frac{e^{X_{i1}\beta}}{e^{X_i 1\beta} + e^{X_{i2}\beta}}$$

Letting $X_{ij}\beta = \alpha + x_{ij}\beta$,
(What's α if every family has exactly one checkbook balancer?)

And so all the denominators drop out, leaving

$$= \frac{e^{X_{i1}\beta}}{e^{X_i 1\beta} + e^{X_{i2}\beta}}$$

Letting $X_{ij}\beta = \alpha + x_{ij}\beta$,
(What's α if every family has exactly one checkbook balancer?)

$$= \frac{e^{\alpha + x_{i1}\beta}}{e^{\alpha + x_i 1\beta} + e^{\alpha + x_{i2}\beta}}$$

And so all the denominators drop out, leaving

$$= \frac{e^{X_{i1}\beta}}{e^{X_i 1\beta} + e^{X_{i2}\beta}}$$

Letting $X_{ij}\beta = \alpha + x_{ij}\beta$,
(What's α if every family has exactly one checkbook balancer?)

$$= \frac{e^{\alpha + x_{i1}\beta}}{e^{\alpha + x_i 1\beta} + e^{\alpha + x_{i2}\beta}}$$
$$= \frac{e^{\alpha} e^{x_{i1}\beta}}{e^{\alpha} e^{x_{i1}\beta} + e^{\alpha} e^{x_{i2}\beta}}$$

And so all the denominators drop out, leaving

$$= \frac{e^{X_{i1}\beta}}{e^{X_i 1\beta} + e^{X_{i2}\beta}}$$

Letting $X_{ij}\beta = \alpha + x_{ij}\beta$,
(What's α if every family has exactly one checkbook balancer?)

$$= \frac{e^{\alpha + x_{i1}\beta}}{e^{\alpha + x_i 1\beta} + e^{\alpha + x_{i2}\beta}}$$
$$= \frac{e^{\alpha} e^{x_{i1}\beta}}{e^{\alpha} e^{x_{i1}\beta} + e^{\alpha} e^{x_{i2}\beta}}$$

and so α has no effect and can be dropped, leaving:

And so all the denominators drop out, leaving

$$= \frac{e^{X_{i1}\beta}}{e^{X_i 1\beta} + e^{X_{i2}\beta}}$$

Letting $X_{ij}\beta = \alpha + x_{ij}\beta$,
(What's α if every family has exactly one checkbook balancer?)

$$= \frac{e^{\alpha + x_{i1}\beta}}{e^{\alpha + x_i 1\beta} + e^{\alpha + x_{i2}\beta}}$$
$$= \frac{e^{\alpha} e^{x_{i1}\beta}}{e^{\alpha} e^{x_{i1}\beta} + e^{\alpha} e^{x_{i2}\beta}}$$

and so α has no effect and can be dropped, leaving:

$$= \frac{e^{x_{i1}\beta}}{e^{x_{i1}\beta} + e^{x_{i2}\beta}}$$

What about a family-level variable X that doesn't distinguish husbands and wives? Let $x_{ij} = (A_{ij}, B_{ij})$ but $B_{i1} = B_{i2}$. Then:

What about a family-level variable X that doesn't distinguish husbands and wives? Let $x_{ij} = (A_{ij}, B_{ij})$ but $B_{i1} = B_{i2}$. Then:

$$= \frac{e^{A_{i1}\beta_A + B_{i1}\beta_B}}{e^{A_{i1}\beta_A + B_{i1}\beta_B} + e^{A_{i2}\beta_A + B_{i2}\beta_B}}$$

What about a family-level variable X that doesn't distinguish husbands and wives? Let $x_{ij} = (A_{ij}, B_{ij})$ but $B_{i1} = B_{i2}$. Then:

$$= \frac{e^{A_{i1}\beta_A + B_{i1}\beta_B}}{e^{A_{i1}\beta_A + B_{i1}\beta_B} + e^{A_{i2}\beta_A + B_{i2}\beta_B}}$$

$$= \frac{e^{A_{i1}\beta_A} e^{B_{i1}\beta_B}}{e^{A_{i1}\beta_A} e^{B_{i1}\beta_B} + e^{A_{i2}\beta_A} e^{B_{i2}\beta_B}}$$

What about a family-level variable X that doesn't distinguish husbands and wives? Let $x_{ij} = (A_{ij}, B_{ij})$ but $B_{i1} = B_{i2}$. Then:

$$
\begin{aligned}
&= \frac{e^{A_{i1}\beta_A + B_{i1}\beta_B}}{e^{A_{i1}\beta_A + B_{i1}\beta_B} + e^{A_{i2}\beta_A + B_{i2}\beta_B}} \\
&= \frac{e^{A_{i1}\beta_A} e^{B_{i1}\beta_B}}{e^{A_{i1}\beta_A} e^{B_{i1}\beta_B} + e^{A_{i2}\beta_A} e^{B_{i2}\beta_B}}
\end{aligned}
$$

and so B drops out, leaving

What about a family-level variable X that doesn't distinguish husbands and wives? Let $x_{ij} = (A_{ij}, B_{ij})$ but $B_{i1} = B_{i2}$. Then:

$$= \frac{e^{A_{i1}\beta_A + B_{i1}\beta_B}}{e^{A_{i1}\beta_A + B_{i1}\beta_B} + e^{A_{i2}\beta_A + B_{i2}\beta_B}}$$

$$= \frac{e^{A_{i1}\beta_A} e^{B_{i1}\beta_B}}{e^{A_{i1}\beta_A} e^{B_{i1}\beta_B} + e^{A_{i2}\beta_A} e^{B_{i2}\beta_B}}$$

and so B drops out, leaving

$$= \frac{e^{A_{i1}\beta_A}}{e^{A_{i1}\beta_A} + e^{A_{i2}\beta_A}}$$

The moral of this mathematical story is:

The moral of this mathematical story is:

1. The constant term and any slope on a variable that doesn't vary within groups is not identified.

The moral of this mathematical story is:

1. The constant term and any slope on a variable that doesn't vary within groups is not identified.
2. These parameters cannot be estimated under the conditional logit model (no amount of data will help).

The moral of this mathematical story is:

1. The constant term and any slope on a variable that doesn't vary within groups is not identified.
2. These parameters cannot be estimated under the conditional logit model (no amount of data will help).
3. Say again? Any explantory variable that is constant within groups must be dropped from the equation.

The moral of this mathematical story is:

1. The constant term and any slope on a variable that doesn't vary within groups is not identified.
2. These parameters cannot be estimated under the conditional logit model (no amount of data will help).
3. Say again? Any explantory variable that is constant within groups must be dropped from the equation.
4. This makes sense since the group-level intercepts (the dummy variables) are being conditioned out.

The moral of this mathematical story is:

1. The constant term and any slope on a variable that doesn't vary within groups is not identified.
2. These parameters cannot be estimated under the conditional logit model (no amount of data will help).
3. Say again? Any explantory variable that is constant within groups must be dropped from the equation.
4. This makes sense since the group-level intercepts (the dummy variables) are being conditioned out.
5. Say again? This makes sense since we're asking (e.g.,) whether the husband or wife does the checkbook in each marriage. We don't have any variation (given this question) whether the husband from marriage 1 has a higher probability than the husband from marriage 2.

Applications of logit and conditional logit:

Applications of logit and conditional logit:

1. Multinomial Choice:

Applications of logit and conditional logit:

1. Multinomial Choice:
   (a) $\sum_j Y_{ij} = 1$, and Y's are all 0s or 1s.

Applications of logit and conditional logit:

1. Multinomial Choice:
   (a) $\sum_j Y_{ij} = 1$, and Y's are all 0s or 1s.
   (b) Example: Vote for 1 of 5 parties

Applications of logit and conditional logit:

1. Multinomial Choice:
   (a) $\sum_j Y_{ij} = 1$, and Y's are all 0s or 1s.
   (b) Example: Vote for 1 of 5 parties
   (c) Example: consumer choice among several brands of orange juice

Applications of logit and conditional logit:

1. Multinomial Choice:
   (a) $\sum_j Y_{ij} = 1$, and Y's are all 0s or 1s.
   (b) Example: Vote for 1 of 5 parties
   (c) Example: consumer choice among several brands of orange juice
2. What to do with Time Series, Cross-Sectional data, Example: war/peace in each country-year

Applications of logit and conditional logit:

1. Multinomial Choice:
   (a) $\sum_j Y_{ij} = 1$, and Y's are all 0s or 1s.
   (b) Example: Vote for 1 of 5 parties
   (c) Example: consumer choice among several brands of orange juice
2. What to do with Time Series, Cross-Sectional data, Example: war/peace in each country-year
   (a) Reading: King, Gary. "Proper Nouns and Methodological Propriety: Pooling Dyads in International Relations Data," [concluding comment in a symposium on the analysis of dyadic international conflict data, with papers by Donald Green, Soo Yeon Kim, and David Yoon; John Oneal and Bruce Russett; and Nathaniel Beck and Jonathan Katz], *International Organization*, Vol. 55, No. 2 (Fall, 2001): Pp. 497–507.

(b) Could run binary logit, but need to control for hard-to-measure country-specific variables (e.g., war-proneness, historical animosity, etc.)

(b) Could run binary logit, but need to control for hard-to-measure country-specific variables (e.g., war-proneness, historical animosity, etc.)

(c) One possible solution: include dummy variables for countries to control for *all* country-specific variables

(b) Could run binary logit, but need to control for hard-to-measure country-specific variables (e.g., war-proneness, historical animosity, etc.)
(c) One possible solution: include dummy variables for countries to control for *all* country-specific variables
(d) But, for MLE's to be consistent, the number of parameters must stay fixed as n increases (or not increase faster than n).

(b) Could run binary logit, but need to control for hard-to-measure country-specific variables (e.g., war-proneness, historical animosity, etc.)
(c) One possible solution: include dummy variables for countries to control for *all* country-specific variables
(d) But, for MLE's to be consistent, the number of parameters must stay fixed as n increases (or not increase faster than n).
(e) This makes sense: if the estimates do not encompass more information as n increases, the sampling distribution will not collapse to a spike.

(b) Could run binary logit, but need to control for hard-to-measure country-specific variables (e.g., war-proneness, historical animosity, etc.)
(c) One possible solution: include dummy variables for countries to control for *all* country-specific variables
(d) But, for MLE's to be consistent, the number of parameters must stay fixed as n increases (or not increase faster than n).
(e) This makes sense: if the estimates do not encompass more information as n increases, the sampling distribution will not collapse to a spike.
(f) The inconsistency is serious: $\hat{\beta}$ can be off by a factor of 2.

(b) Could run binary logit, but need to control for hard-to-measure country-specific variables (e.g., war-proneness, historical animosity, etc.)
(c) One possible solution: include dummy variables for countries to control for *all* country-specific variables
(d) But, for MLE's to be consistent, the number of parameters must stay fixed as n increases (or not increase faster than n).
(e) This makes sense: if the estimates do not encompass more information as n increases, the sampling distribution will not collapse to a spike.
(f) The inconsistency is serious: $\hat{\beta}$ can be off by a factor of 2.
(g) The theory: As T increases, we're ok, but as N increases we have a problem.

(b) Could run binary logit, but need to control for hard-to-measure country-specific variables (e.g., war-proneness, historical animosity, etc.)
(c) One possible solution: include dummy variables for countries to control for *all* country-specific variables
(d) But, for MLE's to be consistent, the number of parameters must stay fixed as n increases (or not increase faster than n).
(e) This makes sense: if the estimates do not encompass more information as n increases, the sampling distribution will not collapse to a spike.
(f) The inconsistency is serious: $\hat{\beta}$ can be off by a factor of 2.
(g) The theory: As T increases, we're ok, but as N increases we have a problem.
(h) The practice: normally N is large but T is small, so many dummy variables must be included.

(b) Could run binary logit, but need to control for hard-to-measure country-specific variables (e.g., war-proneness, historical animosity, etc.)
(c) One possible solution: include dummy variables for countries to control for *all* country-specific variables
(d) But, for MLE's to be consistent, the number of parameters must stay fixed as n increases (or not increase faster than n).
(e) This makes sense: if the estimates do not encompass more information as n increases, the sampling distribution will not collapse to a spike.
(f) The inconsistency is serious: $\hat{\beta}$ can be off by a factor of 2.
(g) The theory: As T increases, we're ok, but as N increases we have a problem.
(h) The practice: normally N is large but T is small, so many dummy variables must be included.
(i) Ethan Katz (*Political Analysis*, 2000) showed that if $T \geq 16$, you're ok.

(b) Could run binary logit, but need to control for hard-to-measure country-specific variables (e.g., war-proneness, historical animosity, etc.)
(c) One possible solution: include dummy variables for countries to control for *all* country-specific variables
(d) But, for MLE's to be consistent, the number of parameters must stay fixed as n increases (or not increase faster than n).
(e) This makes sense: if the estimates do not encompass more information as n increases, the sampling distribution will not collapse to a spike.
(f) The inconsistency is serious: $\hat{\beta}$ can be off by a factor of 2.
(g) The theory: As T increases, we're ok, but as N increases we have a problem.
(h) The practice: normally N is large but T is small, so many dummy variables must be included.
(i) Ethan Katz (*Political Analysis*, 2000) showed that if $T \geq 16$, you're ok. [Ethan was an undergraduate in this class when he did this research.]

(j) The alternative:

(j) The alternative:

   i. Include dummy variables for countries to control for *all* country-specific variables

(j) The alternative:

   i. Include dummy variables for countries to control for *all* country-specific variables
   ii. Condition on the total for country j over time, $\sum_i Y_{ij}$

(j) The alternative:

    i. Include dummy variables for countries to control for *all* country-specific variables
    ii. Condition on the total for country j over time, $\sum_i Y_{ij}$
    iii. Use conditional logit (sometimes known as "clogit" or Chamberlain's logit) and don't estimate the country dummies

(j) The alternative:

   i. Include dummy variables for countries to control for *all* country-specific variables
   ii. Condition on the total for country j over time, $\sum_i Y_{ij}$
   iii. Use conditional logit (sometimes known as "clogit" or Chamberlain's logit) and don't estimate the country dummies
   iv. This is weird since it predicts the present, conditional on knowing the future

(j) The alternative:

  i. Include dummy variables for countries to control for *all* country-specific variables
  ii. Condition on the total for country j over time, $\sum_i Y_{ij}$
  iii. Use conditional logit (sometimes known as "clogit" or Chamberlain's logit) and don't estimate the country dummies
  iv. This is weird since it predicts the present, conditional on knowing the future
  v. Its ok anyway, since the procedure will produce estimates of the same slopes as in the binary logit

(j) The alternative:

   i. Include dummy variables for countries to control for *all* country-specific variables
   ii. Condition on the total for country j over time, $\sum_i Y_{ij}$
   iii. Use conditional logit (sometimes known as "clogit" or Chamberlain's logit) and don't estimate the country dummies
   iv. This is weird since it predicts the present, conditional on knowing the future
   v. Its ok anyway, since the procedure will produce estimates of the same slopes as in the binary logit
   vi. But the procedure won't estimate the coefficients on the country dummies or the constant term, and so we can't calculate probabilities, first differences, etc. — only logit slope coefficients.

(j) The alternative:

   i. Include dummy variables for countries to control for *all* country-specific variables
   ii. Condition on the total for country j over time, $\sum_i Y_{ij}$
   iii. Use conditional logit (sometimes known as "clogit" or Chamberlain's logit) and don't estimate the country dummies
   iv. This is weird since it predicts the present, conditional on knowing the future
   v. Its ok anyway, since the procedure will produce estimates of the same slopes as in the binary logit
   vi. But the procedure won't estimate the coefficients on the country dummies or the constant term, and so we can't calculate probabilities, first differences, etc. — only logit slope coefficients.

(k) The alternative is used, but its not a very happy solution!

Multinomial Logit

Multinomial Logit

The stochastic component:

The stochastic component:

$$\Pr(Y_{ij} = 1) = \pi_{ij}, \qquad \text{s.t.} \qquad \sum_{j=1}^{J} \pi_{ij} = 1$$

Multinomial Logit

The stochastic component:

$$\Pr(Y_{ij} = 1) = \pi_{ij}, \qquad \text{s.t.} \qquad \sum_{j=1}^{J} \pi_{ij} = 1$$

$$\pi_{ij} = \frac{e^{x_{ij}\beta_j}}{\sum_{k=1}^{J} e^{x_{ik}\beta_k}}$$

Multinomial Logit

The stochastic component:

$$\Pr(Y_{ij} = 1) = \pi_{ij}, \qquad \text{s.t.} \qquad \sum_{j=1}^{J} \pi_{ij} = 1$$

$$\pi_{ij} = \frac{e^{x_{ij}\beta_j}}{\sum_{k=1}^{J} e^{x_{ik}\beta_k}}$$

Likelihood:

Multinomial Logit

The stochastic component:

$$\Pr(Y_{ij} = 1) = \pi_{ij}, \qquad \text{s.t.} \qquad \sum_{j=1}^{J} \pi_{ij} = 1$$

$$\pi_{ij} = \frac{e^{x_{ij}\beta_j}}{\sum_{k=1}^{J} e^{x_{ik}\beta_k}}$$

Likelihood:

$$L(\beta|y) = \prod_{i=1}^{n} \left[\prod_{j=1}^{J} \pi_{ij}^{y_{ij}} \right]$$

Independence of Irrelevant Alternatives (IIA)

Independence of Irrelevant Alternatives (IIA)

1. Under MNL:

1. Under MNL:

$$\frac{\pi_{i1}}{\pi_{i2}} = \frac{\frac{e^{x_{i1}\beta_1}}{\sum_{k=1}^{J} e^{x_{ik}\beta_k}}}{\frac{e^{x_{i2}\beta_2}}{\sum_{k=1}^{J} e^{x_{ik}\beta_k}}} = \frac{e^{x_{i1}\beta_1}}{e^{x_{i2}\beta_2}}$$

which is not a function of choices 3,4,5, etc.

Independence of Irrelevant Alternatives (IIA)

1. Under MNL:

$$\frac{\pi_{i1}}{\pi_{i2}} = \frac{\frac{e^{x_{i1}\beta_1}}{\sum_{k=1}^{J} e^{x_{ik}\beta_k}}}{\frac{e^{x_{i2}\beta_2}}{\sum_{k=1}^{J} e^{x_{ik}\beta_k}}} = \frac{e^{x_{i1}\beta_1}}{e^{x_{i2}\beta_2}}$$

   which is not a function of choices 3,4,5, etc.

2. Under MNP, probability ratios are always a function of all alternatives

Independence of Irrelevant Alternatives (IIA)

1. Under MNL:

$$\frac{\pi_{i1}}{\pi_{i2}} = \frac{\frac{e^{x_{i1}\beta_1}}{\sum_{k=1}^{J} e^{x_{ik}\beta_k}}}{\frac{e^{x_{i2}\beta_2}}{\sum_{k=1}^{J} e^{x_{ik}\beta_k}}} = \frac{e^{x_{i1}\beta_1}}{e^{x_{i2}\beta_2}}$$

   which is not a function of choices 3,4,5, etc.
2. Under MNP, probability ratios are always a function of all alternatives
3. The red-bus-blue-bus problem: buses and candidates.

Independence of Irrelevant Alternatives (IIA)

1. Under MNL:

$$\frac{\pi_{i1}}{\pi_{i2}} = \frac{\frac{e^{x_{i1}\beta_1}}{\sum_{k=1}^{J} e^{x_{ik}\beta_k}}}{\frac{e^{x_{i2}\beta_2}}{\sum_{k=1}^{J} e^{x_{ik}\beta_k}}} = \frac{e^{x_{i1}\beta_1}}{e^{x_{i2}\beta_2}}$$

   which is not a function of choices 3,4,5, etc.

2. Under MNP, probability ratios are always a function of all alternatives
3. The red-bus-blue-bus problem: buses and candidates.
4. Does it matter empirically?

Independence of Irrelevant Alternatives (IIA)

1. Under MNL:

$$\frac{\pi_{i1}}{\pi_{i2}} = \frac{\frac{e^{x_{i1}\beta_1}}{\sum_{k=1}^{J} e^{x_{ik}\beta_k}}}{\frac{e^{x_{i2}\beta_2}}{\sum_{k=1}^{J} e^{x_{ik}\beta_k}}} = \frac{e^{x_{i1}\beta_1}}{e^{x_{i2}\beta_2}}$$

   which is not a function of choices 3,4,5, etc.

2. Under MNP, probability ratios are always a function of all alternatives
3. The red-bus-blue-bus problem: buses and candidates.
4. Does it matter empirically? It can, but less often given estimation uncertainty

A Ternary Diagram: Replicating Domínguez and McCann (1996) from King, Tomz, and Wittenberg (2000)

A Ternary Diagram: Replicating Domínguez and McCann (1996) from King, Tomz, and Wittenberg (2000)

1. A replication of Jorge Domínguez and James McCann (1996)

A Ternary Diagram: Replicating Domínguez and McCann (1996) from King, Tomz, and Wittenberg (2000)

1. A replication of Jorge Domínguez and James McCann (1996)
2. Salinas (of the ruling PRI) won the 1988 Mexican presidential election. Domínguez and McCann focus on what drives individual voting behavior.

A Ternary Diagram: Replicating Domínguez and McCann (1996) from King, Tomz, and Wittenberg (2000)

1. A replication of Jorge Domínguez and James McCann (1996)
2. Salinas (of the ruling PRI) won the 1988 Mexican presidential election. Domínguez and McCann focus on what drives individual voting behavior.
3. We focus on the question that motivated them in the first place: if voters had thought the PRI was weakening, who would have won?

A Ternary Diagram: Replicating Domínguez and McCann (1996) from King, Tomz, and Wittenberg (2000)

1. A replication of Jorge Domínguez and James McCann (1996)
2. Salinas (of the ruling PRI) won the 1988 Mexican presidential election. Domínguez and McCann focus on what drives individual voting behavior.
3. We focus on the question that motivated them in the first place: if voters had thought the PRI was weakening, who would have won?
4. Their model was MNL with 3 choices (Salinas from the PRI, Clouthier (from the PAN, a right-wing party), and Cuauhtémoc Cárdenas (head of a leftist coalition) and 31 explanatory variables:

A Ternary Diagram: Replicating Domínguez and McCann (1996) from King, Tomz, and Wittenberg (2000)

1. A replication of Jorge Domínguez and James McCann (1996)
2. Salinas (of the ruling PRI) won the 1988 Mexican presidential election. Domínguez and McCann focus on what drives individual voting behavior.
3. We focus on the question that motivated them in the first place: if voters had thought the PRI was weakening, who would have won?
4. Their model was MNL with 3 choices (Salinas from the PRI, Clouthier (from the PAN, a right-wing party), and Cuauhtémoc Cárdenas (head of a leftist coalition) and 31 explanatory variables:

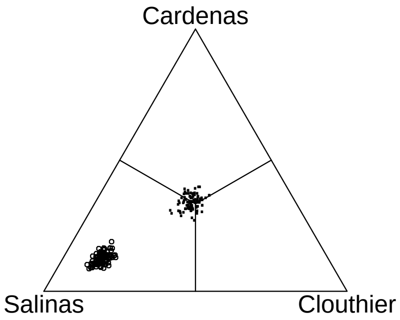
$$Y_i \sim \text{Multinomial}(\pi_i)$$

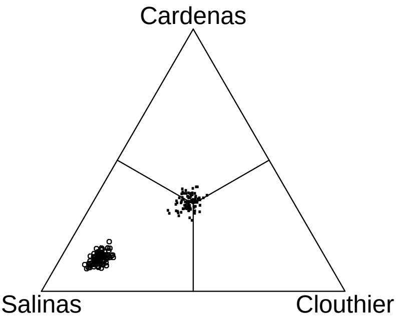
A Ternary Diagram: Replicating Domínguez and McCann (1996) from King, Tomz, and Wittenberg (2000)

1. A replication of Jorge Domínguez and James McCann (1996)
2. Salinas (of the ruling PRI) won the 1988 Mexican presidential election. Domínguez and McCann focus on what drives individual voting behavior.
3. We focus on the question that motivated them in the first place: if voters had thought the PRI was weakening, who would have won?
4. Their model was MNL with 3 choices (Salinas from the PRI, Clouthier (from the PAN, a right-wing party), and Cuauhtémoc Cárdenas (head of a leftist coalition) and 31 explanatory variables:

$$Y_i \sim \text{Multinomial}(\pi_i)$$

$$\pi_i = \frac{e^{X_i\beta_j}}{\sum_{k=1}^{3} e^{X_i\beta_k}} \qquad \text{where } j = 1, 2, 3 \text{ candidates.}$$


Cardenas
Salinas
Clouthier



5. Each point in the figure is an election outcome drawn randomly from a world in which all voters believe Salinas' PRI party is strengthing (for the "o"'s in the bottom left) or weakening (for the "."'s in the middle), with other variables held constant at their means. (100 of each).

6. When voters thought the PRI was strengthing, they win easily.

6. When voters thought the PRI was strengthing, they win easily.
7. When voters believe the PRI is weakening, the election is a tossup. The figure also supports the argument that, despite much voter fraud, Salinas probably did defeat a divided opposition in 1988.

6. When voters thought the PRI was strengthing, they win easily.
7. When voters believe the PRI is weakening, the election is a tossup. The figure also supports the argument that, despite much voter fraud, Salinas probably did defeat a divided opposition in 1988.
8. The PRI in fact lost the next election (finally, after 72 years in power)