Multiple Equation Models

Gov 2001 Section

Spring 2012
Outline

1. Review of Multiple Equation Models
2. Example: SURM
3. Multinomial Choice Models
4. Multiple Equation Models and Missing Data: A Preview
Outline

1. Review of Multiple Equation Models
2. Example: SURM
3. Multinomial Choice Models
4. Multiple Equation Models and Missing Data: A Preview
Occasionally, the system we are studying produces multiple outputs:

- The number of presidential vetoes and the number of Congressional overrides in any given year
- The number of hostile acts directed toward Israel and the number of hostile acts directed toward Palestinians
- The monthly unemployment rate in the United States and the monthly inflation rate.
Multiple Equation Models

You could build a separate model for each individual variable. Or, you could model them using multiple equation modeling.

Advantages:

- Allows more sophisticated, realistic modeling.
- Less restrictive in the assumptions it requires.
- May result in better estimates.
When is it better to use multiple equation models?

- When the possible system outcomes (the $Y_i$’s) are either:
  - (1) stochastically dependent: e.g., dependence in $Y_{1i}$ and $Y_{2i}$
  - (2) parametrically dependent: e.g., when $\theta_{j=1}$ and $\theta_{j=2}$ are deterministically related

- If not, you are probably ok using single equation models.
What is Stochastic Independence?

- With stochastic independence, we can factor the likelihood into two components
  - Full Likelihood
    \[
    L(\theta, \alpha|y) = f(y|\theta, \alpha)
    = \prod_{i=1}^{n} f(y_i|\theta_i, \alpha)
    \]
  - Factored Likelihood
    \[
    L(\theta, \alpha|y) = \prod_{i=1}^{n} f(y_{1i}|\theta_{1i}, \alpha_1)f(y_{2i}|\theta_{2i}, \alpha_2)...
    \]
    \[
    \ln L(\theta, \alpha|y) = \sum_{i=1}^{n} f(y_{1i}|\theta_{1i}, \alpha_1) + \sum_{i=1}^{n} f(y_{2i}|\theta_{2i}, \alpha_2)...
    \]
What is Stochastic Independence?

- With stochastic dependence we can’t do this.
  - Example:
    \[
    L(\mu, \Sigma|y) \propto N(\vec{y}_i|\vec{\mu}_i, \Sigma)
    \]
- The vector of \( y \)’s are jointly determined by the probability density function.
What is Parametric Independence?

- Parametric independence means we can treat one of the \( \theta \)'s as a constant.
- So if we have

\[
\ln L(\theta, \alpha | y) = \sum_{i=1}^{n} f(y_{1i}|\theta_{1i}, \alpha_{1}) + \sum_{i=1}^{n} f(y_{2i}|\theta_{2i}, \alpha_{2})
\]

And the \( \theta \)'s are parametrically independent, when we maximize in terms of \( \theta_{1i} \), we can drop all terms that don’t include \( \theta_{1i} \):

\[
\ln L(\theta, \alpha | y) = \sum_{i=1}^{n} f(y_{1i}|\theta_{1i}, \alpha_{1})
\]
What is Parametric Independence?

- With parametric dependence we can’t do this.

\[
\ln L(\theta, \alpha | y) = \sum_{i=1}^{n} f(y_{1i} | \theta_{1i}, \alpha_1) + \sum_{i=1}^{n} f(y_{2i} | \theta_{2i}, \alpha_2)
\]

- Say \( \theta_{1i} = \theta_{2i} \), then we can’t drop any terms because

\[
\ln L(\theta, \alpha | y) = \sum_{i=1}^{n} f(y_{1i} | \theta_{1i}, \alpha_1) + \sum_{i=1}^{n} f(y_{2i} | \theta_{1i}, \alpha_2)
\]
Multiple Equation Models

- If you have stochastic and parametric independence, great. Just use a single equation model.
- But if you don’t, then think about using a multiple equation model.
A Closer Look: Parametric vs. Stochastic Dependence

The model:

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

Where,

\[x_1\text{ demographics} \]
\[x_2\text{ candidate characteristics (affecting vote but not PID)} \]
\[x_3\text{ parents PID (affecting PID but not vote)} \]

Are \(Y_{1i}\) and \(Y_{2i}\) either parametrically or stochastically dependent, or both?
A Closer Look: Parametric vs. Stochastic Dependence

The model:

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

\[\mu_{1i} = x_{1i} \beta_1 + x_{2i} \beta_2 + \mu_{2i} \beta_3 \quad \text{(vote)}\]

Are \(Y_{1i}\) and \(Y_{2i}\) either parametrically or stochastically dependent, or both?
A Closer Look: Parametric vs. Stochastic Dependence

The model:

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

\[
\begin{align*}
\mu_{1i} &= x_{1i} \beta_1 + x_{2i} \beta_2 + \mu_{2i} \beta_3 \quad \text{(vote)} \\
\mu_{2i} &= x_{1i} \gamma_1 + x_{3i} \gamma_2 + \mu_{1i} \gamma_3 \quad \text{(PID)}
\end{align*}
\]

Are \(Y_{1i}\) and \(Y_{2i}\) either parametrically or stochastically dependent, or both?
A Closer Look: Parametric vs. Stochastic Dependence

The model:

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

\[
\begin{align*}
\mu_{1i} &= x_{1i} \beta_1 + x_{2i} \beta_2 + \mu_{2i} \beta_3 & \text{(vote)} \\
\mu_{2i} &= x_{1i} \gamma_1 + x_{3i} \gamma_2 + \mu_{1i} \gamma_3 & \text{(PID)}
\end{align*}
\]

Where,

Are \(Y_{1i}\) and \(Y_{2i}\) either parametrically or stochastically dependent, or both?
A Closer Look: Parametric vs. Stochastic Dependence

The model:

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

\[
\mu_{1i} = x_{1i}\beta_1 + x_{2i}\beta_2 + \mu_{2i}\beta_3 \quad \text{(vote)}
\]

\[
\mu_{2i} = x_{1i}\gamma_1 + x_{3i}\gamma_2 + \mu_{1i}\gamma_3 \quad \text{(PID)}
\]

Where,

\[x_1 \text{ demographics}\]

Are \(Y_{1i}\) and \(Y_{2i}\) either parametrically or stochastically dependent, or both?
A Closer Look: Parametric vs. Stochastic Dependence

The model:

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

$$\mu_{1i} = x_{1i} \beta_1 + x_{2i} \beta_2 + \mu_{2i} \beta_3 \quad \text{(vote)}$$

$$\mu_{2i} = x_{1i} \gamma_1 + x_{3i} \gamma_2 + \mu_{1i} \gamma_3 \quad \text{(PID)}$$

Where,

$x_1$ demographics
$x_2$ candidate characteristics (affecting vote but not PID)

Are $Y_{1i}$ and $Y_{2i}$ either parametrically or stochastically dependent, or both?
A Closer Look: Parametric vs. Stochastic Dependence

The model:

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

\[\mu_{1i} = x_{1i} \beta_1 + x_{2i} \beta_2 + \mu_{2i} \beta_3\] (vote)
\[\mu_{2i} = x_{1i} \gamma_1 + x_{3i} \gamma_2 + \mu_{1i} \gamma_3\] (PID)

Where,
- \(x_1\) demographics
- \(x_2\) candidate characteristics (affecting vote but not PID)
- \(x_3\) parents PID (affecting PID but not vote)

Are \(Y_{1i}\) and \(Y_{2i}\) either parametrically or stochastically dependent, or both?
A Closer Look: Parametric vs. Stochastic Dependence

The model:

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

Where,

- \(x_{1i}\) demographics
- \(x_{2i}\) candidate characteristics (affecting vote but not PID)
- \(x_{3i}\) parents PID (affecting PID but not vote)

Are \(Y_{1i}\) and \(Y_{2i}\) either parametrically or stochastically dependent, or both?
A Closer Look: Parametric vs. Stochastic Dependence

The model:

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

\[\mu_{1i} = x_{1i} \beta_1 + x_{2i} \beta_2\] (vote)

Are \(Y_{1i}\) and \(Y_{2i}\) either parametrically or stochastically dependent, or both?
A Closer Look: Parametric vs. Stochastic Dependence

The model:

\[(Y_{1i}, Y_{2i}) \sim N(\mu_{1i}, \mu_{2i}, \sigma_1, \sigma_2, \sigma_{12})\]

\[
\mu_{1i} = x_{1i}\beta_1 + x_{2i}\beta_2 \quad \text{(vote)}
\]

\[
\mu_{2i} = x_{1i}\gamma_1 + x_{3i}\gamma_2 \quad \text{(PID)}
\]

Are \(Y_{1i}\) and \(Y_{2i}\) either parametrically or stochastically dependent, or both?
A Closer Look: Parametric vs. Stochastic Dependence

The model:

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

\[
\begin{align*}
\mu_{1i} &= x_{1i} \beta_1 + x_{2i} \beta_2 \\
\mu_{2i} &= x_{1i} \gamma_1 + x_{3i} \gamma_2
\end{align*}
\]

(vote) \hspace{2cm} (PID)

Where,

Are \(Y_{1i}\) and \(Y_{2i}\) either parametrically or stochastically dependent, or both?
A Closer Look: Parametric vs. Stochastic Dependence

The model:

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

\[\mu_{1i} = x_{1i} \beta_1 + x_{2i} \beta_2\] (vote)

\[\mu_{2i} = x_{1i} \gamma_1 + x_{3i} \gamma_2\] (PID)

Where,

\[x_1\] demographics

Are \(Y_{1i}\) and \(Y_{2i}\) either parametrically or stochastically dependent, or both?
A Closer Look: Parametric vs. Stochastic Dependence

The model:

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

\[
\begin{align*}
\mu_{1i} &= x_{1i} \beta_1 + x_{2i} \beta_2 \\
\mu_{2i} &= x_{1i} \gamma_1 + x_{3i} \gamma_2
\end{align*}
\]

Where,

- $x_1$ demographics
- $x_2$ candidate characteristics (affecting vote but not PID)
- $x_3$ parent's PID (affecting PID but not vote)

Are $Y_{1i}$ and $Y_{2i}$ either parametrically or stochastically dependent, or both?
A Closer Look: Parametric vs. Stochastic Dependence

The model:

\[(Y_{1i}, Y_{2i}) \sim N(y_{1i}, y_{2i} | \mu_{1i}, \mu_{2i}, \sigma_{1}, \sigma_{2}, \sigma_{12})\]

\[\mu_{1i} = x_{1i}\beta_{1} + x_{2i}\beta_{2} \quad \text{(vote)}\]
\[\mu_{2i} = x_{1i}\gamma_{1} + x_{3i}\gamma_{2} \quad \text{(PID)}\]

Where,

\(x_{1}\) demographics

\(x_{2}\) candidate characteristics (affecting vote but not PID)

\(x_{3}\) parents PID (affecting PID but not vote)

Are \(Y_{1i}\) and \(Y_{2i}\) either parametrically or stochastically dependent, or both?
A Closer Look: Parametric vs. Stochastic Dependence

The model:

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

\[
\mu_{1i} = x_{1i} \beta_1 + x_{2i} \beta_2 \quad \text{(vote)}
\]

\[
\mu_{2i} = x_{1i} \gamma_1 + x_{3i} \gamma_2 \quad \text{(PID)}
\]

Where,

\(x_1\) demographics
\(x_2\) candidate characteristics (affecting vote but not PID)
\(x_3\) parents PID (affecting PID but not vote)

What would we have to restrict for these to be estimated as two separate regressions?
A Closer Look: Parametric vs. Stochastic Dependence

The model:

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

\[
\mu_{1i} = x_{1i} \beta_1 + x_{2i} \beta_2 \quad \text{(vote)}
\]

\[
\mu_{2i} = x_{1i} \gamma_1 + x_{3i} \gamma_2 \quad \text{(PID)}
\]

Where,

- \(x_1\) demographics
- \(x_2\) candidate characteristics (affecting vote but not PID)
- \(x_3\) parents PID (affecting PID but not vote)

This is a “Seemingly Unrelated Regression Model”
Outline

1. Review of Multiple Equation Models
2. Example: SURM
3. Multinomial Choice Models
4. Multiple Equation Models and Missing Data: A Preview
Let’s look at the Grunfeld data in Zelig.

- Observations from 1935 to 1954 of 7 variables for two firms: General Electric and Westinghouse. The variables are
  - $I_{ge}$ and $I_{w} =$ Gross investment for GE and W, respectively;
  - $F_{ge}$ and $F_{w} =$ Market value of Firm at beginning of the year;
  - $C_{ge}$ and $C_{w} =$ Capital stock measure at beginning of the year.

- We are interested in modeling investment as a function of market value and capital stock.

- Why use a multiple equations model?
Multiple Equation Example

Let’s look at some notation first.

- \( Y \): \( J \times n \) matrix with \( j = 1, \ldots, J \) and \( i = 1, \ldots, n \)
  - \( j \) indexes dependent variables (2)
  - \( i \) indexes observations (20 years)

For the Grunfeld data:

\[
\begin{array}{cc}
\text{Ige} & \text{Iw} \\
33.1 & 12.93 \\
45.0 & 25.90 \\
77.2 & 35.05 \\
44.6 & 22.89 \\
48.1 & 18.84 \\
74.4 & 28.57 \\
\end{array}
\]

\[
Y_{1j} = (33.1, 12.93) \\
Y_{i1} = (33.1, 45.0, 77.2, 44.6, 48.1, 74.4 \ldots)^T
\]
Multiple Equation Example

Additional notation

- **X**: We’re going to separate X into two parts:
  - $X_{i1}$: the covariates specific to the mean for dependent variable 1, in this case GE.
  - $X_{i2}$: the covariates specific to the mean for dependent variable 2, in this case Westinghouse.

Again, for the Grunfeld data:

<table>
<thead>
<tr>
<th>$X_{i1}$</th>
<th>$X_{i2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fge</td>
<td>Cge</td>
</tr>
<tr>
<td>1170.6</td>
<td>97.8</td>
</tr>
<tr>
<td>2015.8</td>
<td>104.4</td>
</tr>
<tr>
<td>2803.3</td>
<td>118.0</td>
</tr>
<tr>
<td>2039.7</td>
<td>156.2</td>
</tr>
<tr>
<td>2256.2</td>
<td>172.6</td>
</tr>
<tr>
<td>2132.2</td>
<td>186.6</td>
</tr>
<tr>
<td></td>
<td>Fw</td>
</tr>
<tr>
<td>191.5</td>
<td>1.8</td>
</tr>
<tr>
<td>516.0</td>
<td>0.8</td>
</tr>
<tr>
<td>729.0</td>
<td>7.4</td>
</tr>
<tr>
<td>560.4</td>
<td>18.1</td>
</tr>
<tr>
<td>519.9</td>
<td>23.5</td>
</tr>
<tr>
<td>628.5</td>
<td>26.5</td>
</tr>
</tbody>
</table>
Multiple Equation Example

Let’s analyze the data using the Seemingly Unrelated Regression Model (SURM). Here, our data are multivariate normal:

- **Stochastic Component**

  \[
  \tilde{Y}_i \sim N(\tilde{y}_i|\tilde{\mu}_i, \Sigma)
  \]

  where
  - \(\tilde{Y}_i\) and \(\tilde{\mu}_i\) are \(J \times 1\)
  - \(\Sigma\) is \(J \times J\)
  - \(\Sigma\) is symmetric and full; Why? Stochastic Dependence

- **Systematic Component**

  \[
  \mu_{i1} = X_{i1}\beta \\
  \mu_{i2} = X_{i2}\gamma
  \]

  where
  - \(\beta\): Estimates that predict GE mean
  - \(\gamma\): Estimates that predict Westinghouse mean
Multiple Equation Example

- The Likelihood

\[
L(\mu, \Sigma) = \prod_{i=1}^{n} N(\vec{y}_i | \vec{\mu}_i, \Sigma)
\]

\[
= \prod_{i=1}^{n} (2\pi)^{-\frac{j}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} (\vec{y}_i - \vec{\mu}_i)' \Sigma^{-1} (\vec{y}_i - \vec{\mu}_i) \right]
\]

- What if \( \Sigma \) had off-diagonal elements that were zero? Stochastic Independence (so using independent models ok)
Multiple Equation Example

We can operationalize this model in Zelig

data(grunfeld)
formula <- list(mu1 = Ige ~ Fge + Cge,
               mu2 = Iw ~ Fw + Cw)

z.out <- zelig(formula = formula,
               model = "sur", data = grunfeld)
Multiple Equation Example

We can operationalize this model in Zelig

```r
z.out <- zelig(formula = formula,
                model = "sur", data = grunfeld)

z.out
```

systemfit results
method: SUR

Coefficients:

<table>
<thead>
<tr>
<th></th>
<th>mu1_(Intercept)</th>
<th>mu1_Fge</th>
<th>mu1_Cge</th>
<th>mu2_(Intercept)</th>
<th>mu2_Fw</th>
<th>mu2_Cw</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-27.7193171</td>
<td>0.0383102</td>
<td>0.1390363</td>
<td>-1.2519882</td>
<td>0.057</td>
<td></td>
</tr>
</tbody>
</table>

and then we can use the usual tricks to interpret this
Multiple Equation Example

Are the $\mu$'s correlated?

summary(z.out)

The correlations of the residuals

\begin{array}{c c}
\mu_1 & \mu_2 \\
\mu_1 & 1.000000 & 0.765043 \\
\mu_2 & 0.765043 & 1.000000 \\
\end{array}
Multiple Equation Example

```r
x.out <- setx(z.out, x=list(Fge=0, Fw = 800))
s.out <- sim(z.out, x=x.out)
summary(s.out)
```

Model: sur
Number of simulations: 1000

Values of X
(Intercept)  Fge  Cge  Fw  Cw
1 1 1941.325 400.16 191.5 1.8

Expected Values: E(Y|X)

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>sd</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-54611.349</td>
<td>55970.197</td>
<td>-158049.19</td>
<td>50144.240</td>
</tr>
<tr>
<td>2</td>
<td>-5388.513</td>
<td>5521.624</td>
<td>-15592.78</td>
<td>4945.616</td>
</tr>
</tbody>
</table>

Predicted Values: Y|X

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>sd</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-54611.349</td>
<td>55970.197</td>
<td>-158049.19</td>
<td>50144.240</td>
</tr>
<tr>
<td>2</td>
<td>-5388.513</td>
<td>5521.624</td>
<td>-15592.78</td>
<td>4945.616</td>
</tr>
</tbody>
</table>
Multiple Equation Example

plot(s.out)
Outline

1. Review of Multiple Equation Models
2. Example: SURM
3. Multinomial Choice Models
4. Multiple Equation Models and Missing Data: A Preview
Multinomial models

Suppose you have a choice among different, non-ordered things:

- Democrat, Republican, Independent
- Republican candidates for president
- at a restaurant between beef, chicken, or vegetarian
- in what to wear – a blue, yellow, or green sweater

The choice sets are not ordered. But we can use generalizations of ordered logit and ordered probit to analyze.
Multinomial models

We’ll use a running example of vote choice between three candidates.

- Suppose for each individual $i$, we observe her vote with

$$V_{ij} = (V_{i1}, V_{i2}, V_{i3})$$

where $V_{ij} = 1$ if the person votes for candidate $j$ and is zero otherwise.

- We could use the ordered logit/probit from the beginning of the semester. But this assumes candidates are ordered along a single dimension.

- Multinomial choice models relax this assumption; we assume there is no order to the choice under consideration, just categories.

Why is this a Multiple Equation Model?
First, let’s use a multinomial logit model to model this choice.

This model has stochastic component

\[ \tilde{V}_i \sim Multinomial(\tilde{v}_i|\tilde{\pi}_i) \]

where

\[ Pr(V_{ij} = 1|X_i, \beta_j) = \pi_{ij} \]

In our example, \( \tilde{\pi}_i = (\pi_{i1}, \pi_{i2}, \pi_{i3}) \). We can write the systematic component as

\[ \pi_{ij} = \frac{\exp(X_i\beta_j)}{\sum_{k=1}^{3} \exp(X_i\beta_k)} \]
Suppose we have $n$ voters and 3 candidates.
The likelihood model for the data, having observed an $n \times K$ matrix of covariates $X$ and an $n \times 3$ matrix of votes $V$ is:

$$L(\beta | X, V) \propto \prod_{i=1}^{n} \prod_{j=1}^{3} \pi_{ij}^{V_{ij}}$$

We can operationalize this in R using Zelig or the mlogit package.
Mutlinomial Logit

Let’s look at the 1988 presidential vote in Mexico.

Our hypothesis is that vote for the ruling PRI party varies with age.

```r
z.out <- zelig(as.factor(vote88) ~ age + female,
               model = "mlogit", data = mexico)
```

```r
z.out
Coefficients:
(Intercept):1 (Intercept):2 age:1 age:2 female:1
0.104832885 -0.500933247 0.016842559 0.008101289 0.265989231 0.289899347
```

As usual, this is a bit difficult to interpret...
Multinomial Logit

...so we can look to first differences

```r
x.young <- setx(z.out, age = min(mexico$age))
x.old <- setx(z.out, age = max(mexico$age))
s.out <- sim(z.out, x1 = x.old, x = x.young)
summary(s.out)
```

First Differences: $\Pr(Y=k|X1) - \Pr(Y=k|X)$

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>sd</th>
<th>2.5%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Pr(Y=1)$</td>
<td>0.1831760</td>
<td>0.06264429</td>
<td>0.0618293</td>
<td>0.30277899</td>
</tr>
<tr>
<td>$\Pr(Y=2)$</td>
<td>-0.0406195</td>
<td>0.05047512</td>
<td>-0.1350517</td>
<td>0.06689080</td>
</tr>
<tr>
<td>$\Pr(Y=3)$</td>
<td>-0.1425565</td>
<td>0.04747605</td>
<td>-0.2322232</td>
<td>-0.04139632</td>
</tr>
</tbody>
</table>
Multinomial Logit

plot(s.out)

Predicted Values: $Y=k|X$

$Y=3$

$Y=2$

$Y=1$

Predicted Probabilities: $Pr(Y=k|X)$

First Differences: $Pr(Y=k|X_1) - Pr(Y=k|X)$
The multinomial logit assumes independence of irrelevant alternatives (IIA).

My choice between Donald Trump and Sarah Palin does not depend on the presence of candidate Mitt Romney.

My choice of transportation between a red bus and a taxi does not depend on the presence of a blue bus.

Why do you need IIA? Because, under multinomial logit, your decision between options 1 and 2 never depends on option 3:

\[
\frac{\pi_i(1)}{\pi_i(2)} = \frac{\exp(X_i \beta_1)}{\exp(X_i \beta_2)}
\]
Multinomial Probit

- If you think you may be violating IIA, then use a multinomial probit, which relaxes the IIA assumption.
- Usually more difficult to compute, though, and no Zelig function :(
Multinomial Probit

Suppose that voter $i$ receives utility from voting for candidate $j$:

$$U_{ij}^* \sim N(u_{ij}^* | \mu_{ij}, \Sigma)$$

where $\Sigma$ is the $3 \times 3$ var-cov matrix.

The voter will vote for whoever gives her the highest utility, so we have the following observation mechanism

$$V_{ij} = \begin{cases} 
1 & \text{if } U_{ij}^* > U_{ij'}^* \forall j \neq j' \\
0 & \text{otherwise.}
\end{cases}$$
Multinomial Probit

And now we can use the same stochastic component

\[ V_i \sim \text{Multinomial}(v_i | \pi_i) \]

and the systematic component

\[ \Pr(V_{ij} = 1) = \pi_{ij} \]

such that

\[ \sum_{j=1}^{3} \pi_{ik} = 1 \]
Multinomial Probit

To compute $\pi_{ij}$, we need to calculate the following:

$$\pi_{ij} = \Pr(U_{ij}^* > U_{ij'}^*) \forall j' \neq j$$

which is a messy series of integrals. For example, for $j = 3$:

$$\Pr(U_{i3}^* > U_{ij}^* \forall j \neq 3) = \int_{-\infty}^{\infty} \int_{-\infty}^{U_{i3}} \int_{-\infty}^{U_{i3}} f(U_1, U_2, U_3) dU_1 dU_2 dU_3.$$
To compute $\pi_{ij}$, we need to calculate the following:

$$\pi_{ij} = Pr(U_{ij}^* > U_{ij'}^*) \forall j' \neq j$$

which is a messy series of integrals. For example, for $j = 3$:

But it’s less messy than it looks!

$$Pr(U_{i3}^* > U_{ij}^* \forall j \neq 3) = \int_{-\infty}^{\infty} \int_{-\infty}^{U_{i3}} \int_{-\infty}^{U_{i3}} f(U_1, U_2, U_3) dU_1 dU_2 dU_3.$$
To compute $\pi_{ij}$, we need to calculate the following:

$$
\pi_{ij} = Pr(U_{ij}^* > U_{ij'}^*) \forall j' \neq j
$$

which is a messy series of integrals. For example, for $j = 3$:

$$
Pr(U_{i3}^* > U_{ij}^* \forall j \neq 3) = \int_{-\infty}^{\infty} \int_{-\infty}^{U_{i3}} \int_{-\infty}^{U_{i3}} f(U_1, U_2, U_3) dU_1 dU_2 dU_3.
$$
To compute $\pi_{ij}$, we need to calculate the following:

$$\pi_{ij} = Pr(U_{ij}^* > U_{ij'}^*) \forall j' \neq j$$

which is a messy series of integrals. For example, for $j = 3$:

$$Pr(U_{i3}^* > U_{ij}^* \forall j \neq 3) = \int_{-\infty}^{\infty} \int_{-\infty}^{U_{i3}} \int_{-\infty}^{U_{i3}} f(U_1, U_2, U_3) dU_1 dU_2 dU_3.$$
To compute \( \pi_{ij} \), we need to calculate the following:

\[
\pi_{ij} = \Pr(U_{ij}^* > U_{ij'}^*) \forall j' \neq j
\]

which is a messy series of integrals. For example, for \( j = 3 \):

\[
\Pr(U_{i3}^* > U_{ij}^* \forall j \neq 3) = \int_{-\infty}^{\infty} \int_{-\infty}^{U_{i3}} \int_{-\infty}^{U_{i3}} f(U_1, U_2, U_3) \, dU_1 \, dU_2 \, dU_3.
\]
Multinomial Probit

To compute $\pi_{ij}$, we need to calculate the following:

$$\pi_{ij} = \Pr(U_{ij}^* > U_{ij'}^*) \forall j' \neq j$$

which is a messy series of integrals. For example, for $j = 3$:

$$\Pr(U_{i3}^* > U_{ij}^* \forall j \neq 3) = \int_{-\infty}^{\infty} \int_{-\infty}^{U_{i3}} \int_{-\infty}^{U_{i3}} f(U_1, U_2, U_3) \, dU_1 \, dU_2 \, dU_3.$$

- We no longer need to assume IIA because we have included other options directly in this integral.
- However, this is hard to identify because of potential correlations between unobserved latent dimensions.
- Note: Imai and Van Dyck’s MNP package will calculate multinomial probit using Bayesian methods.
Outline

1. Review of Multiple Equation Models
2. Example: SURM
3. Multinomial Choice Models
4. Multiple Equation Models and Missing Data: A Preview
This could be our data matrix:

<table>
<thead>
<tr>
<th>Observation</th>
<th>Age</th>
<th>Education</th>
<th>Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13</td>
<td>4</td>
<td>$0</td>
</tr>
<tr>
<td>2</td>
<td>56</td>
<td>32</td>
<td>$100,000</td>
</tr>
<tr>
<td>3</td>
<td>24</td>
<td>15</td>
<td>$30,000</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>12</td>
<td>$20,000</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Wait, it’s a multiple equation model!

<table>
<thead>
<tr>
<th>Observation</th>
<th>Age</th>
<th>Education</th>
<th>Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_{1j} =$</td>
<td>(13</td>
<td>4</td>
<td>$0)</td>
</tr>
<tr>
<td>$Y_{2j} =$</td>
<td>(56</td>
<td>32</td>
<td>$100,000)</td>
</tr>
<tr>
<td>$Y_{3j} =$</td>
<td>(24</td>
<td>15</td>
<td>$30,000)</td>
</tr>
<tr>
<td>$Y_{4j} =$</td>
<td>(100</td>
<td>12</td>
<td>$20,000)</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

We could model this as:

\[
\begin{align*}
\mu_i, \text{Age} & \quad \mu_i, \text{Education} & \quad \mu_i, \text{Income} \\
\Rightarrow \bar{\mu}_i, \Sigma
\end{align*}
\]
We can model this the same way we’ve modeled multiple equation models:

\[ L(\mu, \Sigma|D) \propto \prod_{i=1}^{n} N(D_i|\mu, \Sigma) \]

- As usual, we can make \( \mu \) a function of some covariates
Multiple Equation Models and Missing Data: A Preview

But aren’t some of the data missing?

<table>
<thead>
<tr>
<th>Observation</th>
<th>Age</th>
<th>Education</th>
<th>Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_{1j}$</td>
<td>13</td>
<td>4</td>
<td>?</td>
</tr>
<tr>
<td>$Y_{2j}$</td>
<td>56</td>
<td>32</td>
<td>$100,000$</td>
</tr>
<tr>
<td>$Y_{3j}$</td>
<td>24</td>
<td>?</td>
<td>$30,000$</td>
</tr>
<tr>
<td>$Y_{4j}$</td>
<td>?</td>
<td>12</td>
<td>$20,000$</td>
</tr>
</tbody>
</table>

...
We’ll integrate out the missing components:

\[
L(\mu, \Sigma|D_{obs}) \propto \prod_{i=1}^{n} \int N(D_i|\mu, \Sigma) dD_{mis}
\]

\[
= \prod_{i=1}^{n} N(D_{i,obs}|\mu_{obs}, \Sigma_{obs})
\]

- Then we can maximize the likelihood to find the parameters of interest.
- Given those parameters of interest, we estimate the missing data.
- We go back in forth until convergence.