Multiple Equation Models & Amelia

Outline

Outline

Multiple Equation Models

Multiple Equation Models

Occasionally, the system we are studying produces multiple outputs:

Multiple Equation Models

Occasionally, the system we are studying produces multiple outputs:

- The number of presidential vetoes and the number of Congressional overrides in any given year

Multiple Equation Models

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

Multiple Equation Models

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

Multiple Equation Models

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

Multiple Equation Models

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

Advantages:

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.

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.

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.

Multiple Equation Models

Multiple Equation Models

When is it better to use multiple equation models?

Multiple Equation Models

When is it better to use multiple equation models?

- When the possible system outcomes (the Y_i's) are either:

Multiple Equation Models

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}

Multiple Equation Models

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} When the number of hostile acts toward Israel is correlated with the number of hostile acts directed toward Palestinians

Multiple Equation Models

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} When the number of hostile acts toward Israel is correlated with the number of hostile acts directed toward Palestinians
  - (2) parametrically dependent: e.g., when $\theta_{j=1}$ and $\theta_{j=2}$ are deterministically related

Multiple Equation Models

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} When the number of hostile acts toward Israel is correlated with the number of hostile acts directed toward Palestinians
  - (2) parametrically dependent: e.g., when $\theta_{j=1}$ and $\theta_{j=2}$ are deterministically related When the mean of the number of hostile acts toward Israel is a function of the mean of hostile acts directed toward Palestinians

Multiple Equation Models

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} When the number of hostile acts toward Israel is correlated with the number of hostile acts directed toward Palestinians
  - (2) parametrically dependent: e.g., when $\theta_{j=1}$ and $\theta_{j=2}$ are deterministically related When the mean of the number of hostile acts toward Israel is a function of the mean of hostile acts directed toward Palestinians
- If not, you are probably ok using single equation models.

What is Stochastic Independence?

What is Stochastic Independence?

- With stochastic independence, we can factor the likelihood into two components

What is Stochastic Independence?

- With stochastic independence, we can factor the likelihood into two components
  - Full Likelihood

What is Stochastic Independence?

- With stochastic independence, we can factor the likelihood into two components
  - Full Likelihood

$$L(\theta, \alpha | y) = f(y | \theta, \alpha)$$

What is Stochastic Independence?

- With stochastic independence, we can factor the likelihood into two components
  - Full Likelihood

$$
\begin{aligned}
L(\theta, \alpha|y) &= f(y|\theta, \alpha) \\
&= \prod_{i=1}^{n} f(y_i|\theta_i, \alpha)
\end{aligned}
$$

What is Stochastic Independence?

- With stochastic independence, we can factor the likelihood into two components
  - Full Likelihood

$$
\begin{aligned}
L(\theta, \alpha | y) &= f(y|\theta, \alpha) \\
&= \prod_{i=1}^{n} f(y_i|\theta_i, \alpha)
\end{aligned}
$$

  - Factored Likelihood

What is Stochastic Independence?

- With stochastic independence, we can factor the likelihood into two components
  - Full Likelihood

$$\begin{aligned} L(\theta, \alpha | y) &= f(y|\theta, \alpha) \\ &= \prod_{i=1}^{n} f(y_i|\theta_i, \alpha) \end{aligned}$$

  - Factored Likelihood

$$L(\theta, \alpha | y) = \prod_{i=1}^{n} f(y_{1i}|\theta_{1i}, \alpha_1) f(y_{2i}|\theta_{2i}, \alpha_2)...$$

What is Stochastic Independence?

- With stochastic independence, we can factor the likelihood into two components
  - Full Likelihood

$$\begin{aligned} L(\theta, \alpha|y) &= f(y|\theta, \alpha) \\ &= \prod_{i=1}^{n} f(y_i|\theta_i, \alpha) \end{aligned}$$

  - Factored Likelihood

$$\begin{aligned} L(\theta, \alpha|y) &= \prod_{i=1}^{n} f(y_{1i}|\theta_{1i}, \alpha_1) f(y_{2i}|\theta_{2i}, \alpha_2)... \\ lnL(\theta, \alpha|y) &= \sum_{i=1}^{n} f(y_{1i}|\theta_{1i}, \alpha_1) + \sum_{i=1}^{n} f(y_{2i}|\theta_{2i}, \alpha_2)... \end{aligned}$$

What is Stochastic Independence?

What is Stochastic Independence?

- With stochastic dependence we can't do this.

What is Stochastic Independence?

- With stochastic dependence we can't do this.
  - Example:

What is Stochastic Independence?

- With stochastic dependence we can't do this.
  - Example:

$$L(\mu, \Sigma | y) \propto N(\vec{y_i} | \vec{\mu_i}, \Sigma)$$

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?

What is Parametric Independence?

- Parametric independence means we can treat one of the θ's as a constant.

What is Parametric Independence?

- Parametric independence means we can treat one of the θ's as a constant.
- So if we have

What is Parametric Independence?

- Parametric independence means we can treat one of the θ's as a constant.
- So if we have

$$lnL(\theta, \alpha|y) = \sum_{i=1}^{n} f(y_{1i}|\theta_{1i}, \alpha_1) + \sum_{i=1}^{n} f(y_{2i}|\theta_{2i}, \alpha_2)$$

What is Parametric Independence?

- Parametric independence means we can treat one of the θ's as a constant.
- So if we have

$$lnL(\theta, \alpha|y) = \sum_{i=1}^{n} f(y_{1i}|\theta_{1i}, \alpha_1) + \sum_{i=1}^{n} f(y_{2i}|\theta_{2i}, \alpha_2)$$

And the θ's are parametrically independent, when we maximize in terms of θ_{1i}, we can drop all terms that don't include θ_{1i}:

What is Parametric Independence?

- Parametric independence means we can treat one of the θ's as a constant.
- So if we have

$$lnL(\theta, \alpha|y) = \sum_{i=1}^{n} f(y_{1i}|\theta_{1i}, \alpha_1) + \sum_{i=1}^{n} f(y_{2i}|\theta_{2i}, \alpha_2)$$

And the θ's are parametrically independent, when we maximize in terms of θ_{1i}, we can drop all terms that don't include θ_{1i}:

$$lnL(\theta, \alpha|y) = \sum_{i=1}^{n} f(y_{1i}|\theta_{1i}, \alpha_1)$$

What is Parametric Independence?

What is Parametric Independence?

- With parametric dependence we can't do this.

What is Parametric Independence?

- With parametric dependence we can't do this.

$$lnL(\theta, \alpha|y) = \sum_{i=1}^{n} f(y_{1i}|\theta_{1i}, \alpha_1) + \sum_{i=1}^{n} f(y_{2i}|\theta_{2i}, \alpha_2)$$

What is Parametric Independence?

- With parametric dependence we can't do this.

$$lnL(\theta, \alpha|y) = \sum_{i=1}^{n} f(y_{1i}|\theta_{1i}, \alpha_1) + \sum_{i=1}^{n} f(y_{2i}|\theta_{2i}, \alpha_2)$$

- Say $\theta_{1i} = \theta_{2i}$, then we can't drop any terms because

What is Parametric Independence?

- With parametric dependence we can't do this.

$$lnL(\theta, \alpha|y) = \sum_{i=1}^{n} f(y_{1i}|\theta_{1i}, \alpha_1) + \sum_{i=1}^{n} f(y_{2i}|\theta_{2i}, \alpha_2)$$

- Say $\theta_{1i} = \theta_{2i}$, then we can't drop any terms because

$$lnL(\theta, \alpha|y) = \sum_{i=1}^{n} f(y_{1i}|\theta_{1i}, \alpha_1) + \sum_{i=1}^{n} f(y_{2i}|\theta_{1i}, \alpha_2)$$

What is Parametric Independence?

- With parametric dependence we can't do this.

$$lnL(\theta, \alpha|y) = \sum_{i=1}^{n} f(y_{1i}|\theta_{1i}, \alpha_1) + \sum_{i=1}^{n} f(y_{2i}|\theta_{2i}, \alpha_2)$$

- Say $\theta_{1i} = \theta_{2i}$, then we can't drop any terms because

$$lnL(\theta, \alpha|y) = \sum_{i=1}^{n} f(y_{1i}|\theta_{1i}, \alpha_1) + \sum_{i=1}^{n} f(y_{2i}|\theta_{1i}, \alpha_2)$$

Multiple Equation Models

Multiple Equation Models

- If you have stochastic and parametric independence, great. Just use a single equation model.

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:

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

A Closer Look: Parametric vs. Stochastic Dependence

The model:

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

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

A Closer Look: Parametric vs. Stochastic Dependence

The model:

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

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

A Closer Look: Parametric vs. Stochastic Dependence

The model:

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

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

Where,

A Closer Look: Parametric vs. Stochastic Dependence

The model:

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

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

Where,

x_1 demographics

A Closer Look: Parametric vs. Stochastic Dependence

The model:

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

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

Where,

x_1 demographics

x_2 candidate characteristics (affecting vote but not PID)

A Closer Look: Parametric vs. Stochastic Dependence

The model:

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

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

Where,

x_1 demographics

x_2 candidate characteristics (affecting vote but not PID)

x_3 parents PID (affecting PID but not vote)

A Closer Look: Parametric vs. Stochastic Dependence

The model:

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

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

Where,

x_1 demographics

x_2 candidate characteristics (affecting vote but not PID)

x_3 parents PID (affecting PID but not vote)

A Closer Look: Parametric vs. Stochastic Dependence

The model:

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

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

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:

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

$$\mu_{1i} = x_{1i}\beta_1 + x_{2i}\beta_2 \qquad \text{(vote)}$$
$$\mu_{2i} = x_{1i}\gamma_1 + x_{3i}\gamma_2 \qquad \text{(PID)}$$

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 regression?

A Closer Look: Parametric vs. Stochastic Dependence

The model:

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

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

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

Multiple Equation Example

Multiple Equation Example

Multiple Equation Example

Let's look at the `Grunfeld` data in `Zelig`.

Multiple Equation Example


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

Multiple Equation Example

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
  - Ige and Iw = Gross investment for GE and W, respectively;

Multiple Equation Example


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
  - Ige and Iw = Gross investment for GE and W, respectively;
  - Fge and Fw=Market value of Firm at beginning of the year;

Multiple Equation Example


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
  - Ige and Iw = Gross investment for GE and W, respectively;
  - Fge and Fw=Market value of Firm at beginning of the year;
  - Cge and Cw= Capital stock measure at beginning of the year.

Multiple Equation Example

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
  - Ige and Iw = Gross investment for GE and W, respectively;
  - Fge and Fw=Market value of Firm at beginning of the year;
  - Cge and Cw= Capital stock measure at beginning of the year.
- We are interested in modeling investment as a function of market value and capital stock.

Multiple Equation Example

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
  - Ige and Iw = Gross investment for GE and W, respectively;
  - Fge and Fw=Market value of Firm at beginning of the year;
  - Cge and Cw= 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

Multiple Equation Example

Let's look at some notation first.

Multiple Equation Example

Let's look at some notation first.

- Y: $J \times n$ matrix with $j = 1, ..., J$ and $i = 1, ..., n$

Multiple Equation Example

Let's look at some notation first.

- Y: $J \times n$ matrix with $j = 1, ..., J$ and $i = 1, ..., n$
  - j indexes dependent variables (2)

Multiple Equation Example

Let's look at some notation first.

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

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:

Multiple Equation Example

Let's look at some notation first.

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

For the `Grunfeld` data:

```
 Ige     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
```

Multiple Equation Example

Let's look at some notation first.

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

For the `Grunfeld` data:

```
 Ige    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
```

$Y_{1j} =$ (`33.1, 12.93`)

Multiple Equation Example

Let's look at some notation first.

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

For the `Grunfeld` data:

```
 Ige    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
```

$Y_{1j} = ($ 33.1, 12.93 $)$
$Y_{i1} = ($ 33.1, 45.0, 77.2, 44.6, 48.1, 74.4... $)^T$

Multiple Equation Example

Additional notation

Multiple Equation Example

Additional notation

- X: We're going to separate X into two parts:

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.

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.

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:

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:

X_{i1}

Fge	Cge
1170.6	97.8
2015.8	104.4
2803.3	118.0
2039.7	156.2
2256.2	172.6
2132.2	186.6

X_{i2}

Fw	Cw
191.5	1.8
516.0	0.8
729.0	7.4
560.4	18.1
519.9	23.5
628.5	26.5

Multiple Equation Example

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

Multiple Equation Example

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

- Stochastic Component

Multiple Equation Example

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

- Stochastic Component

$$\vec{Y}_i \sim N(\vec{y}_i | \vec{\mu}_i, \Sigma)$$

Multiple Equation Example

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

- Stochastic Component

$$\vec{Y}_i \sim N(\vec{y}_i | \vec{\mu}_i, \Sigma)$$

  where

Multiple Equation Example

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

- Stochastic Component

$$\vec{Y}_i \sim N(\vec{y}_i | \vec{\mu}_i, \Sigma)$$

  where
  - $\vec{Y}_i$ and $\vec{\mu}_i$ are $J \times 1$

Multiple Equation Example

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

- Stochastic Component

$$\vec{Y}_i \sim N(\vec{y}_i|\vec{\mu}_i, \Sigma)$$

  where
  - $\vec{Y}_i$ and $\vec{\mu}_i$ are $J \times 1$
  - Σ is $J \times J$

Multiple Equation Example

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

- Stochastic Component

$$\vec{Y}_i \sim N(\vec{y}_i | \vec{\mu}_i, \Sigma)$$

  where
  - $\vec{Y}_i$ and $\vec{\mu}_i$ are $J \times 1$
  - Σ is $J \times J$
  - Σ is symmetric and full; Why? Stochastic Dependence

Multiple Equation Example

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

- Stochastic Component

$$\vec{Y}_i \sim N(\vec{y}_i | \vec{\mu}_i, \Sigma)$$

  where
  - $\vec{Y}_i$ and $\vec{\mu}_i$ are $J \times 1$
  - Σ is $J \times J$
  - Σ is symmetric and full; Why? Stochastic Dependence
- Systematic Component

Multiple Equation Example

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

- Stochastic Component

$$\vec{Y}_i \sim N(\vec{y_i} | \vec{\mu}_i, \Sigma)$$

  where
  - $\vec{Y}_i$ and $\vec{\mu}_i$ are $J \times 1$
  - Σ is $J \times J$
  - Σ is symmetric and full; Why? Stochastic Dependence
- Systematic Component

$$\mu_{i1} = X_{i1} \vec{\beta}$$

Multiple Equation Example

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

- Stochastic Component

$$\vec{Y}_i \sim N(\vec{y}_i | \vec{\mu}_i, \Sigma)$$

  where
  - $\vec{Y}_i$ and $\vec{\mu}_i$ are $J \times 1$
  - Σ is $J \times J$
  - Σ is symmetric and full; Why? Stochastic Dependence
- Systematic Component

$$\mu_{i1} = X_{i1}\vec{\beta}$$
$$\mu_{i2} = X_{i2}\vec{\gamma}$$

Multiple Equation Example

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

- Stochastic Component

$$\vec{Y}_i \sim N(\vec{y}_i | \vec{\mu}_i, \Sigma)$$

  where
  - $\vec{Y}_i$ and $\vec{\mu}_i$ are $J \times 1$
  - Σ is $J \times J$
  - Σ is symmetric and full; Why? Stochastic Dependence
- Systematic Component

$$\mu_{i1} = X_{i1}\vec{\beta}$$
$$\mu_{i2} = X_{i2}\vec{\gamma}$$

  where

Multiple Equation Example

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

- Stochastic Component

$$\vec{Y_i} \sim N(\vec{y_i}|\vec{\mu_i}, \Sigma)$$

  where
  - $\vec{Y_i}$ and $\vec{\mu_i}$ are $J \times 1$
  - Σ is $J \times J$
  - Σ is symmetric and full; Why? Stochastic Dependence
- Systematic Component

$$\mu_{i1} = X_{i1}\vec{\beta}$$
$$\mu_{i2} = X_{i2}\vec{\gamma}$$

  where
  - β: Estimates that predict GE mean

Multiple Equation Example

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

- Stochastic Component

$$\vec{Y}_i \sim N(\vec{y}_i | \vec{\mu}_i, \Sigma)$$

  where
  - $\vec{Y}_i$ and $\vec{\mu}_i$ are $J \times 1$
  - Σ is $J \times J$
  - Σ is symmetric and full; Why? Stochastic Dependence
- Systematic Component

$$\mu_{i1} = X_{i1}\vec{\beta}$$
$$\mu_{i2} = X_{i2}\vec{\gamma}$$

  where
  - β: Estimates that predict GE mean
  - γ: Estimates that predict Westinghouse mean

Multiple Equation Example

- The Likelihood

Multiple Equation Example

- The Likelihood

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

Multiple Equation Example

- The Likelihood

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

Multiple Equation Example

- The Likelihood

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

- What if Σ had off-diagonal elements that were zero? Stochastic Independence (so using independent models ok)

Multiple Equation Example

We can operationalize this model in `Zelig`

Multiple Equation Example

We can operationalize this model in Zelig

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

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`

Multiple Equation Example

We can operationalize this model in `Zelig`

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

Multiple Equation Example

We can operationalize this model in `Zelig`

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

z.out

systemfit results
method: SUR

Coefficients:
mu1_(Intercept)         mu1_Fge         mu1_Cge mu2_(Intercept)         mu
    -27.7193171       0.0383102       0.1390363      -1.2519882       0.057
```

Multiple Equation Example

We can operationalize this model in `Zelig`

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

z.out

systemfit results
method: SUR

Coefficients:
mu1_(Intercept)         mu1_Fge         mu1_Cge mu2_(Intercept)         mu
    -27.7193171       0.0383102       0.1390363      -1.2519882       0.057
```

and then we can use the usual tricks to interpret this

Multiple Equation Example

Are the μ's correlated?

Multiple Equation Example

Are the μ's correlated?

```
summary(z.out)
```

⋮

```
The correlations of the residuals
         mu1      mu2
mu1 1.000000 0.765043
mu2 0.765043 1.000000
```

⋮

Multiple Equation Example

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

Multiple Equation Example

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

s.out <- sim(z.out, x=x.out)
```

Multiple Equation Example

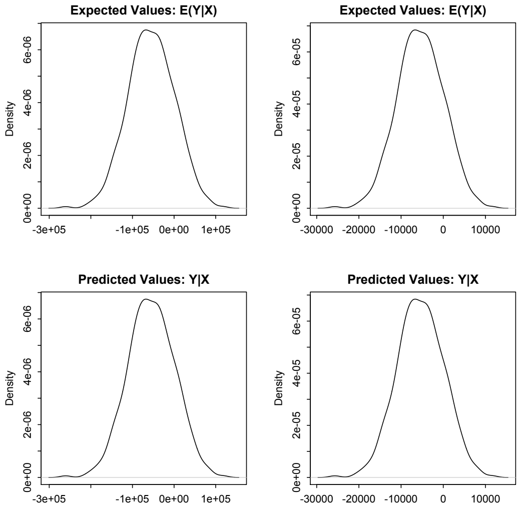
```
x.out <- setx(z.out, x=list(Fge=0, Fw = 800))

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)
       mean        sd       2.5%     97.5%
1 -54611.349 55970.197 -158049.19 50144.240
2  -5388.513  5521.624  -15592.78  4945.616

Predicted Values: Y|X
       mean        sd       2.5%     97.5%
1 -54611.349 55970.197 -158049.19 50144.240
2  -5388.513  5521.624  -15592.78  4945.616
```

Multiple Equation Example

```
plot(s.out)
```

Multiple Equation Example

```
plot(s.out)
```



Outline

Multinomial models

Multinomial models

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

Multinomial models

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


Multinomial models

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

- Democrat, Republican, Independent

Multinomial models

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

- Democrat, Republican, Independent
- Republican candidates for president

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

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

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.

Multinomial models

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

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

Multinomial models

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

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

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

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.

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 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.

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?

Multinomial Logit

Multinomial Logit

- First, let's use a multinomial logit model to model this choice.

Multinomial Logit

- First, let's use a multinomial logit model to model this choice.
- This model has stochastic component

Multinomial Logit

- First, let's use a multinomial logit model to model this choice.
- This model has stochastic component

$$\vec{V}_i \sim Multinomial(\vec{v}_i|\vec{\pi}_i)$$

Multinomial Logit

- First, let's use a multinomial logit model to model this choice.
- This model has stochastic component

$$\vec{V}_i \sim Multinomial(\vec{v}_i|\vec{\pi}_i)$$

where

Multinomial Logit

- First, let's use a multinomial logit model to model this choice.
- This model has stochastic component

$$\vec{V}_i \sim Multinomial(\vec{v}_i|\vec{\pi}_i)$$

where

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

Multinomial Logit

- First, let's use a multinomial logit model to model this choice.
- This model has stochastic component

$$\vec{V}_i \sim Multinomial(\vec{v}_i|\vec{\pi}_i)$$

where

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

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

Multinomial Logit

- First, let's use a multinomial logit model to model this choice.
- This model has stochastic component

$$\vec{V}_i \sim Multinomial(\vec{v}_i|\vec{\pi}_i)$$

where

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

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

$$\pi_{ij} = \frac{exp(X_i\beta_j)}{\sum_{k=1}^{3} exp(X_i\beta_k)}$$

Multinomial Logit

- Suppose we have n voters and 3 candidates.

Multinomial Logit

- 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:

Multinomial Logit

- 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:

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

Multinomial Logit

- 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.

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.

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.

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

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.

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

```
z.out
Coefficients:
(Intercept):1 (Intercept):2         age:1         age:2      female:1
  0.104832885  -0.500933247   0.016842559   0.008101289   0.265989231    0.2
```

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.

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

```
z.out
Coefficients:
(Intercept):1 (Intercept):2         age:1         age:2      female:1
  0.104832885  -0.500933247   0.016842559   0.008101289   0.265989231    0.2
```

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

Multinomial Logit

...so we can look to first differences

```
x.young <- setx(z.out, age = min(mexico$age))
```

Multinomial Logit

...so we can look to first differences

```
x.young <- setx(z.out, age = min(mexico$age))

x.old<- setx(z.out, age = max(mexico$age))
```

Multinomial Logit

...so we can look to first differences

```
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)
```

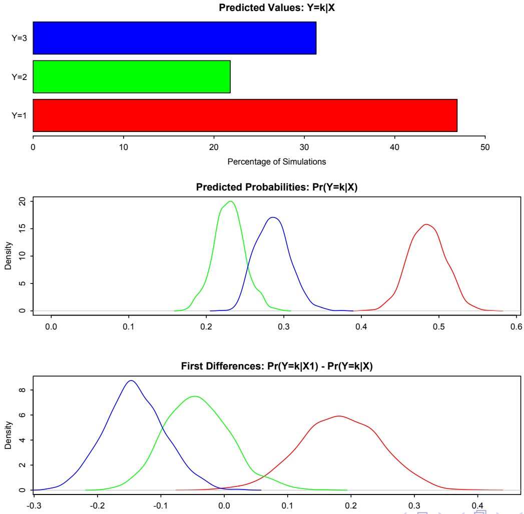
Multinomial Logit

...so we can look to first differences

```
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)
              mean         sd        2.5%       97.5%
Pr(Y=1)  0.1831760 0.06264429  0.0618293  0.30277899
Pr(Y=2) -0.0406195 0.05047512 -0.1350517  0.06689080
Pr(Y=3) -0.1425565 0.04747605 -0.2322232 -0.04139632
```

Multinomial Logit

```
plot(s.out)
```



Multinomial Logit

- The multinomial logit assumes independence of irrelevant alternatives (IIA).

Multinomial Logit

- 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.

Multinomial Logit

- 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.

Multinomial Logit

- 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:

Multinomial Logit

- 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:

$$\frac{\pi_i(1)}{\pi_i(2)} =$$

Multinomial Logit

- 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:

$$\frac{\pi_i(1)}{\pi_i(2)} = \frac{\frac{exp(X_i\beta_1)}{\sum exp(X_i\beta_j)}}{\frac{exp(X_i\beta_2)}{\sum exp(X_i\beta_j)}}$$

Multinomial Logit

- 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:

$$\frac{\pi_i(1)}{\pi_i(2)} = \frac{\frac{exp(X_i\beta_1)}{\sum exp(X_i\beta_j)}}{\frac{exp(X_i\beta_2)}{\sum exp(X_i\beta_j)}} = \frac{exp(X_i\beta_1)}{exp(X_i\beta_2)}$$

Multinomial Probit

Multinomial Probit

- If you think you may be violating IIA, then use a multinomial probit, which relaxes the IIA assumption.

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:

Multinomial Probit

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

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

Multinomial Probit

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

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

where Σ is the 3×3 var-cov matrix.

Multinomial Probit

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

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

where Σ is the 3×3 var-cov matrix.

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

Multinomial Probit

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

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

where Σ is the 3×3 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

Multinomial Probit

And now we can use the same stochastic component

$$V_i \sim Multinomial(v_i|\pi_i)$$

Multinomial Probit

And now we can use the same stochastic component

$$V_i \sim Multinomial(v_i|\pi_i)$$

and the systematic component

Multinomial Probit

And now we can use the same stochastic component

$$V_i \sim Multinomial(v_i|\pi_i)$$

and the systematic component

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

Multinomial Probit

And now we can use the same stochastic component

$$V_i \sim Multinomial(v_i|\pi_i)$$

and the systematic component

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

Multinomial Probit

And now we can use the same stochastic component

$$V_i \sim Multinomial(v_i|\pi_i)$$

and the systematic component

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

such that

Multinomial Probit

And now we can use the same stochastic component

$$V_i \sim Multinomial(v_i|\pi_i)$$

and the systematic component

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

such that

$$\sum_{j=1}^{3} \pi_{ik} = 1$$

Multinomial Probit

To compute π_{ij}, we need to calculate the following:

Multinomial Probit

To compute π_{ij}, we need to calculate the following:

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

Multinomial Probit

To compute π_{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 $j = 3$:

Multinomial Probit

To compute π_{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 π_{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!

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

Multinomial Probit

To compute π_{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!

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

Multinomial Probit

To compute π_{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 π_{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 π_{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.

Multinomial Probit

To compute π_{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.

Multinomial Probit

To compute π_{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

Multiple Equation Models and Missing Data: A Preview

Multiple Equation Models and Missing Data: A Preview

This could be our data matrix:

Multiple Equation Models and Missing Data: A Preview

This could be our data matrix:

Observation	*Age*	*Education*	*Income*
1	13	4	\$0
2	56	32	\$100,000
3	24	15	\$ 30,000
4	100	12	\$ 20,000
⋮	⋮	⋮	⋮

Multiple Equation Models and Missing Data: A Preview

Wait, it's a multiple equation model!

Multiple Equation Models and Missing Data: A Preview

Wait, it's a multiple equation model!

Observation	*Age*	*Education*	*Income*
$Y_{1j} =$	(13	4	\$0)
$Y_{2j} =$	(56	32	\$100,000)
$Y_{3j} =$	(24	15	\$ 30,000)
$Y_{4j} =$	(100	12	\$ 20,000)
$\vdots$	$\vdots$	$\vdots$	$\vdots$

Multiple Equation Models and Missing Data: A Preview

Wait, it's a multiple equation model!

Observation	*Age*	*Education*	*Income*
$Y_{1j} =$	(13	4	\$0)
$Y_{2j} =$	(56	32	\$100,000)
$Y_{3j} =$	(24	15	\$ 30,000)
$Y_{4j} =$	(100	12	\$ 20,000)
$\vdots$	$\vdots$	$\vdots$	$\vdots$

We could model this as:

$$\mu_{i,Age} \qquad \mu_{i,Education} \qquad \mu_{i,Income}$$

Multiple Equation Models and Missing Data: A Preview

Wait, it's a multiple equation model!

Observation	*Age*	*Education*	*Income*
$Y_{1j} =$	(13	4	\$0)
$Y_{2j} =$	(56	32	\$100,000)
$Y_{3j} =$	(24	15	\$ 30,000)
$Y_{4j} =$	(100	12	\$ 20,000)
$\vdots$	$\vdots$	$\vdots$	$\vdots$

We could model this as:

$$\mu_{i,Age} \qquad \mu_{i,Education} \qquad \mu_{i,Income}$$

$$\Rightarrow \vec{\mu_i},\ \Sigma$$

Multiple Equation Models and Missing Data: A Preview

We can model this the same way we've modeled multiple equation models:

Multiple Equation Models and Missing Data: A Preview

We can model this the same way we've modeled multiple equation models:

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

Multiple Equation Models and Missing Data: A Preview

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 μ a function of some covariates

Multiple Equation Models and Missing Data: A Preview

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 μ a function of some covariates (what predicts our data?).

Multiple Equation Models and Missing Data: A Preview

But aren't some of the data missing?

Multiple Equation Models and Missing Data: A Preview

But aren't some of the data missing?

Observation	*Age*	*Education*	*Income*
$Y_{1j} =$	(13	4	?)
$Y_{2j} =$	(56	32	\$100,000)
$Y_{3j} =$	(24	?	\$ 30,000)
$Y_{4j} =$	(?	12	\$ 20,000)
$\vdots$	$\vdots$	$\vdots$	$\vdots$

Multiple Equation Models and Missing Data: A Preview

We'll integrate out the missing components:

Multiple Equation Models and Missing Data: A Preview

We'll integrate out the missing components:

$$
\begin{aligned}
L(\mu, \Sigma | D_{obs}) &\propto \prod_{i=1}^{n} \int N(D_i | \mu, \Sigma) dD_{mis} \\
&= \prod_{i=1}^{n} N(D_{i,obs} | \mu_{obs}, \Sigma_{obs})
\end{aligned}
$$

Multiple Equation Models and Missing Data: A Preview

We'll integrate out the missing components:

$$\begin{aligned} L(\mu, \Sigma | D_{obs}) &\propto \prod_{i=1}^{n} \int N(D_i | \mu, \Sigma) dD_{mis} \\ &= \prod_{i=1}^{n} N(D_{i,obs} | \mu_{obs}, \Sigma_{obs}) \end{aligned}$$

- Then we can maximize the likelihood to find the parameters of interest.

Multiple Equation Models and Missing Data: A Preview

We'll integrate out the missing components:

$$\begin{aligned} L(\mu, \Sigma | D_{obs}) &\propto \prod_{i=1}^{n} \int N(D_i | \mu, \Sigma) dD_{mis} \\ &= \prod_{i=1}^{n} N(D_{i,obs} | \mu_{obs}, \Sigma_{obs}) \end{aligned}$$

- Then we can maximize the likelihood to find the parameters of interest.
- From there we'll simulate, just like we've done before to estimate the missing data.

Outline

Basic Syntax

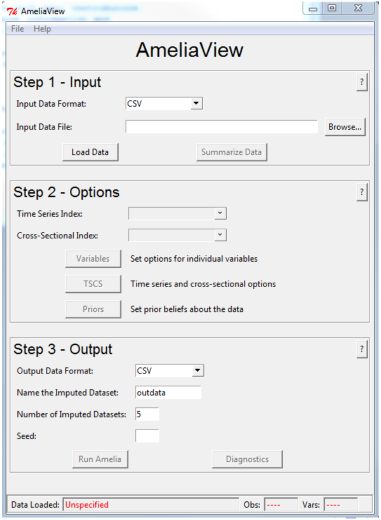
```
library(Amelia)
data(africa)

> africa[1,]
  year      country gdp_pc  infl trade civlib population
1 1972 Burkina Faso    377 -2.92 29.69    0.5    5848380

a.out <- amelia(x = africa, cs = ‘‘country’’, ts = ‘‘year’’,
  logs = ‘‘gdp_pc’’, m = 5)

write.amelia(obj=a.out, file.stem = ‘‘a.out’’)
```

`a.out` is a list of 5 imputed datasets, each of which can be accessed using `a.out$imputations[[i]]`.

GUI

You can use the GUI by typing:

```
AmeliaView()
```



Basic Syntax

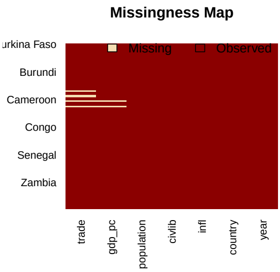
```
z.out.imp <- zelig(trade ~ log(population) + log(gdp_pc) + infl
  + civlib, data = a.out$imputations, model = "ls")
summary(z.out.imp)

Coefficients:
                   Value Std. Error t-stat   p-value
(Intercept)     112.4097   45.47779  2.472 1.345e-02
log(population) -17.8513    2.36646 -7.543 4.595e-14
log(gdp_pc)      31.3875    2.31108 13.581 1.834e-41
infl              0.2605    0.05836  4.463 8.075e-06
civlib           27.2104    6.67210  4.078 4.59$e-05
```

`Zelig` will automatically combine the results of the different models, but if a model you are using isn't programmed in `Zelig`, it isn't hard to combine your estimates using the equations I showed you before.

Diagnostics

The missingness map gives an overall sense of the shape and extent of the missingness.

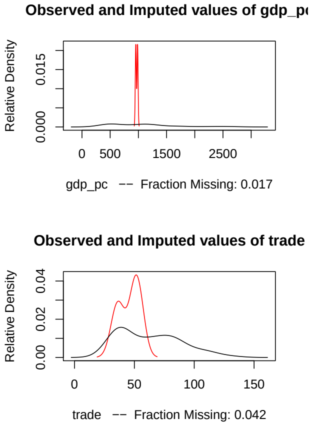
```
missmap(a.out)
```



Diagnostics

Plotting the Amelia object contrasts empirical and imputed densities.

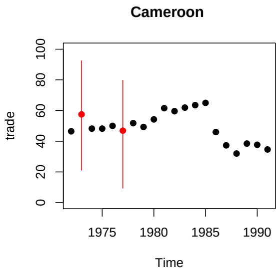
```
plot(a.out)
```



Diagnostics

Plotting Time-series cross-sectional plots

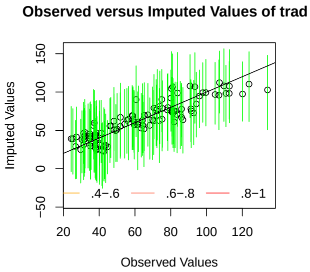
```
tscsPlot(a.out, var = "trade", cs = "Burundi")
```



Diagnostics

Overimputation for a specific variable tests the imputation model by imagining that each observation is missing and generating some imputations to check performance.

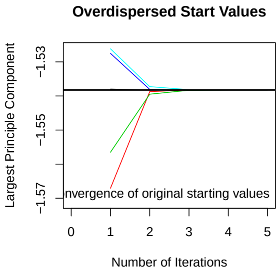
```
overimpute(a.out, var = ``trade'')
```



Diagnostics

The disperse function starts the algorithm at some unlikely values to check that AMELIA II hasn't found a local rather a global maximum for the likelihood of the complete data.

```
disperse(a.out, dims = 1, m = 5)
```



Considerations: Transformations

Recall that our model assumes multivariate normal data. This suggests some issues:

Considerations: Transformations

Recall that our model assumes multivariate normal data. This suggests some issues:

1. Ordinal variables: imputation is faster and more informative if ordinal variables are permitted to be continuous, but if undesirable use `ords` argument to constrain.

Considerations: Transformations

Recall that our model assumes multivariate normal data. This suggests some issues:

1. Ordinal variables: imputation is faster and more informative if ordinal variables are permitted to be continuous, but if undesirable use `ords` argument to constrain.
2. Nominal (unordered) variables: AMELIA II automatically converts into factors and imputes accordingly if a variable is passed to the `noms` argument.

Considerations: Transformations

Recall that our model assumes multivariate normal data. This suggests some issues:

1. Ordinal variables: imputation is faster and more informative if ordinal variables are permitted to be continuous, but if undesirable use `ords` argument to constrain.
2. Nominal (unordered) variables: AMELIA II automatically converts into factors and imputes accordingly if a variable is passed to the `noms` argument.
3. Various transformations (logarithmic, root) are pre-programmed to make skewed distributions more normal.

Considerations: Time Series/Panel Data

1. Use the `ts` and `cs` arguments to tell AMELIA II the panel structure. It uses this information for constructing time serieses correctly.

Considerations: Time Series/Panel Data

1. Use the `ts` and `cs` arguments to tell AMELIA II the panel structure. It uses this information for constructing time serieses correctly.
2. `polytime` can be used to add in polynomial time trends (up to a cubic) and `intercs = TRUE` creates a separate trend for each panel.

Considerations: Time Series/Panel Data

1. Use the `ts` and `cs` arguments to tell AMELIA II the panel structure. It uses this information for constructing time serieses correctly.
2. `polytime` can be used to add in polynomial time trends (up to a cubic) and `intercs = TRUE` creates a separate trend for each panel.
3. Add in lags and leads if you think it will help predict the missing data. Use `lags` and `leads` with some variable.

Things to remember

1. Set the seed!

Things to remember

1. Set the seed!
2. Include any variable in the analysis model in your imputation model.

Things to remember

1. Set the seed!
2. Include any variable in the analysis model in your imputation model.
3. Don't impute things that don't make sense.

Things to remember

1. Set the seed!
2. Include any variable in the analysis model in your imputation model.
3. Don't impute things that don't make sense.
4. Check diagnostics (and think carefully about applicability).

Things to remember

1. Set the seed!
2. Include any variable in the analysis model in your imputation model.
3. Don't impute things that don't make sense.
4. Check diagnostics (and think carefully about applicability).
5. Remember transformations, polynomials and data structure.

Gary's Papers

King, Gary; James Honaker; Ann Joseph; Kenneth Scheve. 2001. "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation," *APSR* 95, 1 (March, 2001): 49-69.

James Honaker and Gary King. "What to do about Missing Values in Time Series Cross-Section Data," *AJPS* 54, 2 (April, 2010): 561-581

Amelia II: A Program for Missing Data