

# Structural Equation Models<sup>1</sup>

STA431 Spring 2015

---

<sup>1</sup>See last slide for copyright information.

# Structural Equation Models

- An extension of multiple regression.
- Can incorporate measurement error.
- More than one regression-like equation.
- All the variables are random.
- An independent variable in one equation can be the dependent variable in another equation.

# Measurement Error

- What you see is not what you really want.
- *Latent variable*: A random variable whose values cannot be directly observed.
- Contrast with *Observable variable*.
- Usually, interest is in relationships between latent variables.
- But all you can see are the observable variables.

# Doubly Labeled Water

Participants drink water that is enriched with respect to two isotopes, and urine samples allow the measurement of energy expenditure (Graphics used without permission).

*Measurement Error in Nonlinear Models: Carroll et al., 2006, p. 8*

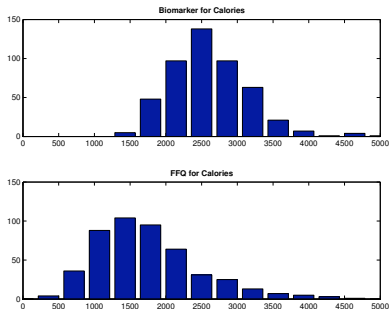
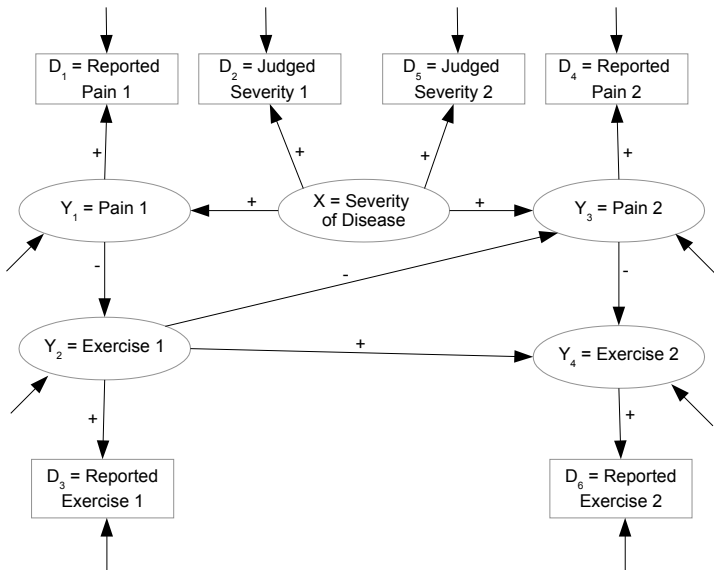


Figure 1.5 *OPEN Study data, histograms of energy (calories) using a biomarker (top panel) and a food frequency questionnaire (bottom panel). Note how individuals report far fewer calories than they actually consume.*

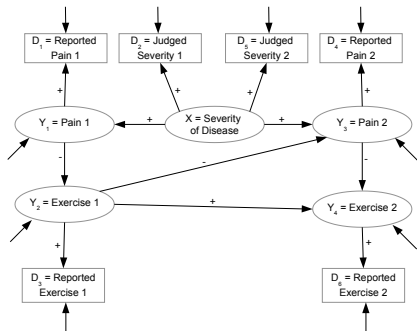
# Path diagrams



# Comments

- Latent variables are in ovals, observable variables are in boxes.
- Error terms seem to come from nowhere – often not shown.
- There is real modeling here. Lots of theoretical input is required.
- These are *causal* models: Models of influence.
- But the data are usually observational.
- Omitted variables can cause problems – more later.

# Path diagrams correspond to systems of equations



$$Y_{i,1} = \beta_{0,1} + \beta_1 X_i + \epsilon_{i,1}$$

$$Y_{i,2} = \beta_{0,2} + \beta_2 Y_{i,1} + \epsilon_{i,2}$$

$$Y_{i,3} = \beta_{0,3} + \beta_3 X_i + \beta_4 Y_{i,2} + \epsilon_{i,3}$$

$$Y_{i,4} = \beta_{0,4} + \beta_5 Y_{i,2} + \beta_6 Y_{i,3} + \epsilon_{i,4}$$

$$D_{i,1} = \lambda_{0,1} + \lambda_1 Y_{i,1} + e_{i,1}$$

$$D_{i,2} = \lambda_{0,2} + \lambda_2 X_i + e_{i,2}$$

$$D_{i,3} = \lambda_{0,3} + \lambda_3 Y_{i,2} + e_{i,3}$$

$$D_{i,4} = \lambda_{0,4} + \lambda_4 Y_{i,3} + e_{i,4}$$

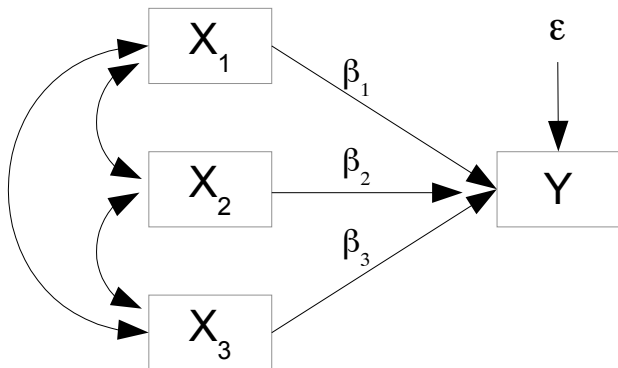
$$D_{i,5} = \lambda_{0,5} + \lambda_2 X_i + e_{i,5}$$

$$D_{i,6} = \lambda_{0,6} + \lambda_5 Y_{i,4} + e_{i,6}$$

Multivariate normal model is standard.

## Regression with observable variables

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \beta_3 X_{i,3} + \epsilon_i$$





# Tools

- Scalar variance-covariance calculations
- Matrices
- Random vectors
- Multivariate normal
- Maximum likelihood
- A little large-sample theory
- SAS

## Copyright Information

This slide show was prepared by **Jerry Brunner**, Department of Statistical Sciences, University of Toronto. Except for the picture taken from Carroll et al.'s *Measurement error in non-linear models*, it is licensed under a **Creative Commons Attribution - ShareAlike 3.0 Unported License**. Use any part of it as you like and share the result freely. The L<sup>A</sup>T<sub>E</sub>X source code is available from the course website:

<http://www.utstat.toronto.edu/~brunner/oldclass/431s15>