IV. Multiple Regression.

A. Introduction
B. CRM
C. Estimation
D. Interpretation of Parameter Estimates
E. Properties of Estimators
F. Estimator for σ^2 and Variances for $\hat{\beta}s$
G. Inference in Multiple Regression
H. Goodness of Fit

C. Estimation of population parameters.

1. How? What method??
2. Process: Describe in words.

D. Interpretation of Parameter Estimates.

1. Partial Effects – we now have several variables affecting Y. Each X has a ***partial effect on Y***.

2. Experimental vs Social Sciences

- ***Experimental Sciences*** – in physical and biological sciences, factors are physically held constant.

What do we do in econometrics?

- Gather data from individuals (markets, firms, etc.) **after choices have been made**.
- ***Holding other factors constant*** – we have to include them in the model.

3. Specification Bias

- This is what happens when ***we don't hold other factors constant.***
- Y is affected by X_1 and X_2.
- Suppose we don't include X_2 in the model – we violated CRM assumption #1.
- **The estimate of β_1 will be biased**.
- Correlation between X_1 and X_2 provides a path for X_2 to follow and ***bias*** the estimate of β_1

Specification Bias – what happens when we omit an important variable from our model.

Truth: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$

You specify: $Y_i = \alpha_0 + \alpha_1 X_{1i} + u_i$

Estimator: $\hat{\alpha}_1 = \frac{\sum x_{1i} Y_i}{\sum x_{1i}^2}$

Evaluate using the true model to see the effects of the ***specification mistake***:

$$E[\hat{\alpha}_1] = \beta_1 + \beta_2 \frac{\sum x_{1i} X_{2i}}{\sum x_{1i}^2}$$

Regression Analysis: sales versus prose, pcarn, dinc

The regression equation is
sales = 13355 – 3628 prose + 2634 pcarn - 19.3 dinc

Predictor	Coef	SE Coef	T	P
Constant	13355	6485	2.06	0.062
prose	-3628.2	635.6	-5.71	0.000
pcarn	2634	1013	2.60	0.023
dinc	-19.25	30.69	-0.63	0.542

S = 1076.29 R-Sq = 77.8% R-Sq(adj) = 72.2%

What if we don't hold other factors constant?

Regression Analysis: sales versus prose

The regression equation is
sales = 16899 – 2979 prose

Predictor	Coef	SE Coef	T	P
Constant	16899	1985	8.52	0.000
prose	–2978.5	630.0	–4.73	0.000

S = 1312.19   R–Sq = 61.5%   R–Sq(adj) = 58.7%

E. Properties of OLS Estimators

1. ***Linear:*** The OLS estimators are still linear estimators.
2. ***Unbiased:*** If CRMAs # 1 – 3 are correct, the OLS estimators are unbiased estimators.
3. ***Minimum Variance:*** If CRMAs # 1 – 5 are correct, the OLS estimators are the ***Best Linear Unbiased Estimators.***

(Gauss-Markov Theorem)

Draw a graph illustrating properties 2 & 3.

F. Estimator for σ^2 and Variances of $\hat{\beta}s$

1. Our ***simple regression*** estimator for σ^2:

2. What changes in ***multiple regression***?

3. Standard errors for $\hat{\beta}s$
   - Formula – i.e.,

$$s_{\hat{\beta}_1} = \sqrt{\frac{\hat{\sigma}^2 \sum x_{2i}^2}{\left(\sum x_{1i}^2\right)\left(\sum x_{2i}^2\right) - \left(\sum x_{1i} x_{1i}\right)^2}}$$

$$s_{\hat{\beta}_2} = \sqrt{\frac{\hat{\sigma}^2 \sum x_{1i}^2}{\left(\sum x_{1i}^2\right)\left(\sum x_{2i}^2\right) - \left(\sum x_{1i} x_{1i}\right)^2}}$$

Yuck! But the software takes care of these.

G. Inference in Multiple Regression.

1. Confidence Intervals – same deal:

2. Hypothesis Tests – Standardized Tests.

H. Goodness of Fit – R^2

1. Multiple Coefficient of Determination – R^2
2. Calculation – exactly the same.

   R^2 = ESS / TSS = 1 – (RSS / TSS)
3. Adjusted R^2:

$$\bar{R}^2 = 1 - \frac{RSS}{TSS}\left(\frac{n-1}{n-K-1}\right)$$

R^2 – used for assessing how well ***a model*** fits.

Adj. R^2 – used to ***compare two models***.

Same sample – same Y values.