AFRE 835: Introductory Econometrics

Chapter 3: Multiple Regression Analysis: Estimation

Spring 2017

Introduction

- The simple regression model illustrates many of the basic concepts in econometric modeling.
- However, we typically want to take into account how our dependent variable depends upon multiple regressors.
- For example, considering some earlier examples, we might expect
  - School performance (y) to depend, not only on class size, but also on teacher experience, innate ability, and the ability of classmates;
  - Wages earned (y) to depend, not only on years of schooling, but also on experience and innate ability;
  - Soybean yields (y) to depend, not only on fertilizer application rates, but also on rainfall, soil quality, and field slope;
  - Local pollution levels (y) to depend, not only on policies restricting vehicle usage, but also on mass transit options and local climate;
  - Housing prices (y) to depend, not only on local environmental amenities, but also on local shopping and schooling, as well as housing characteristics;
  - Health outcomes (y) to depend, not only on hospital visits, but also on a person's initial health conditions.

Introduction

- With additional regressors, we can more *explicitly* control for other factors that influence our dependent variable of interest.
- This, in turn, improves our ability to make claims about the *ceteris paribus* (all else equal) impact of one of our explanatory variables.

Outline

Models with Two Independent Variables

- Wooldridge (p. 69) provides two examples generalizing models introduced in chapter 2;
  - Modeling hourly wage as a function of years of education and years of experience.
$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + u \tag{1}$$
  - Modeling student test scores as a function of average per student expenditures and average family income
$$avgscore = \beta_0 + \beta_1 expend + \beta_2 avginc + u \tag{2}$$
  - In each case, we can now *explicitly* hold constant the effects of the second variable when trying to estimate the *ceteris paribus* impact of the first.
  - We still have to make assumptions of how u is related to each of the explanatory variables.

The General Regression Model with Two Independent Variables

- The general form of the model in this case becomes;
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \tag{3}$$
where β_0, β_1, and β_2 are all parameters to be estimated.
- The interpretations of the parameters are similar to what we found in the simple regression model.
  - β_0 is the intercept, measuring the value of y when $x_1 = 0$, $x_2 = 0$ and $u = 0$.
  - β_1 measures the change in y for each unit change in x_1 holding *all other factors fixed*; i.e.,
$$\beta_1 = \frac{\Delta y}{\Delta x_1} \text{ if } \Delta x_2 = 0 \text{ and } \Delta u = 0. \tag{4}$$
  - β_2 measures the change in y for each unit change in x_2 holding *all other factors fixed*; i.e.,
$$\beta_2 = \frac{\Delta y}{\Delta x_2} \text{ if } \Delta x_1 = 0 \text{ and } \Delta u = 0. \tag{5}$$

The General Regression Model with Two Independent Variables (cont'd)

- For each of the slope parameters, the interpretation requires holding all other factors fixed, including both the other observable factor (e.g., x_2 in the case of β_1) and the unobservable factors (i.e., u).
- Holding the observable factors fixed is straightforward, but holding the unobservable factors fixed is harder because they are unobservable.
- As in the simple regression model context, we instead have to rely on assumptions about the relationship between the observable and unobservable factors.
- The key assumption we will need is an extension of the zero conditional mean assumption; i.e.,

$$E(u|x_1, x_2) = 0. \tag{6}$$

- Given this assumption, $E(y|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$.

Nonlinear Models

- The multiple regression model can also be used to incorporate nonlinear effects of an independent variable.
- For example, we might allow the impact of education to be nonlinear by specifying:

$$wage = \beta_0 + \beta_1 educ + \beta_2 educ^2 + u \tag{7}$$

- In this case, β_1 no longer captures the change in *wage* due to a change in education, holding everything else fixed, since we cannot change *educ* without also changing $educ^2$.
- Instead,

$$\frac{\Delta wage}{\Delta educ} \approx \frac{\partial wage}{\partial educ} = \beta_1 + 2\beta_2 educ \tag{8}$$

- We might, for example, expect $\beta_1 > 0$, with $\beta_2 < 0$.

Models with k Independent Variables

- Extending the multiple regression model to include k independent variables proceeds in the obvious way, with

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u \tag{9}$$

- The key assumption we now need for the relationship between u and the x's becomes

$$E(u|x_1, x_2, \ldots, x_k) = 0. \tag{10}$$

- Given this, our *population regression function (PRF)* becomes

$$E(y|x_1, x_2, \ldots, x_k) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \tag{11}$$

- Departures from the PRF (the error term u) are essentially noise with an average value of zero in the population *regardless of the value of the observable regressors (the x's)*.

OLS Estimator

- Suppose that we have n observations selected at random from the population; i.e., $\{(y_i, x_{i1}, \ldots, x_{ik}) : i = 1, \ldots, n\}$, where now x_{ik} denotes the value of the kth variable for the i^{th} observation.
- The OLS estimator of our parameters minimizes the sum of squared residuals; i.e.,

$$\min_{\hat\beta_0, \hat\beta_1, \ldots, \hat\beta_k} \sum_{i=1}^{n} \hat u_i^2 = \min_{\hat\beta_0, \hat\beta_1, \ldots, \hat\beta_k} \sum_{i=1}^{n} \left(y_i - \hat\beta_0 - \hat\beta_1 x_{i1} - \cdots - \hat\beta_k x_{ik} \right)^2$$

where

$$\begin{aligned} \hat u_i &= y_i - \hat y_i \\ &= y_i - \hat\beta_0 - \hat\beta_1 x_{i1} - \cdots - \hat\beta_k x_{ik} \end{aligned} \tag{12}$$

denotes our fitted residual.

OLS Estimator (cont'd)

- The first order conditions for this minimization problem can be written as:

$$0 = \sum_{i=1}^{n} \left[y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \beta_k x_{ik} \right]$$

$$0 = \sum_{i=1}^{n} x_{i1} \left[y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \beta_k x_{ik} \right]$$

$$\vdots$$

$$0 = \sum_{i=1}^{n} x_{ik} \left[y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \beta_k x_{ik} \right]$$

- Just as was the case of the simple regression model, these can be written as the sample counterparts to the moment conditions $E(u) = 0$ and $Cov(x_j u) = 0, \quad j = 1, \ldots, k$.

An Explicit Formula for the OLS Estimator

- Writing out an explicit formula for the OLS estimator is easy to do using matrix algebra.
- Let

  $\boldsymbol{y} =$ $(y_1, \ldots, y_n)'$ be an $n \times 1$ column vector containing the dependent variables;

  $\boldsymbol{x}_i =$ $(1, x_{i1}, \ldots, x_{ik})$ be a $1 \times (k+1)$ row vector of regressors for individual i (including a constant) ;

  $\boldsymbol{x} =$ $[x_1' \; x_2' \; \cdots \; x_n']'$ be an $n \times (k+1)$ matrix, stacking the n row vectors of regressors x_1 through x_n.

  $\boldsymbol{u} =$ $(u_1, \ldots, u_n)'$ be an $n \times 1$ column vector containing the error terms; and

  $\boldsymbol{\beta} =$ $(\beta_0, \beta_1, \ldots, \beta_k)'$ be a $(k+1) \times 1$ column vector containing the error terms

An Explicit Formula for the OLS Estimator (cont'd)

- Our linear model then becomes

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots & \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_n \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \tag{13}$$

...or more simply

$$\boldsymbol{y} = \boldsymbol{x}\boldsymbol{\beta} + \boldsymbol{u} \tag{14}$$

An Explicit Formula for the OLS Estimator (cont'd)

- The OLS Estimator solves

$$\min_{\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k} \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik} \right)^2$$

$$= \min_{\hat{\boldsymbol{\beta}}} \left(\boldsymbol{y} - \boldsymbol{x}\hat{\boldsymbol{\beta}} \right)' \left(\boldsymbol{y} - \boldsymbol{x}\hat{\boldsymbol{\beta}} \right)$$

$$= \min_{\hat{\boldsymbol{\beta}}} \left[\boldsymbol{y}'\boldsymbol{y} - 2\hat{\boldsymbol{\beta}}'\boldsymbol{x}'\boldsymbol{y} + \hat{\boldsymbol{\beta}}'\boldsymbol{x}'\boldsymbol{x}\hat{\boldsymbol{\beta}} \right] \tag{15}$$

- Several matrix differentiation rules help in solving this problem:
  $\frac{\partial \boldsymbol{a}'\boldsymbol{b}}{\partial \boldsymbol{b}} = \frac{\partial \boldsymbol{b}'\boldsymbol{a}}{\partial \boldsymbol{b}} = \boldsymbol{a}$ where $\boldsymbol{a}$ and $\boldsymbol{b}$ are $n \times 1$ vectors.
  $\frac{\partial \boldsymbol{b}'\boldsymbol{A}\boldsymbol{b}}{\partial \boldsymbol{b}} = 2\boldsymbol{A}\boldsymbol{b}$ for a symmetric matrix $\boldsymbol{A}$.

An Explicit Formula for the OLS Estimator (cont'd)

- The resulting first order conditions for this maximization are:

$$\mathbf{0}_{(k+1)\times 1} = -2\boldsymbol{x}'\boldsymbol{y} + 2(\boldsymbol{x}'\boldsymbol{x})\hat{\boldsymbol{\beta}}$$
$$\Rightarrow$$
$$\hat{\boldsymbol{\beta}} = (\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{y} \tag{16}$$

where $\mathbf{0}_{(k+1)\times 1}$ is a $(k+1)\times 1$ column vector.

Interpreting OLS

- Recall that the population regression function (PRF), given the zero conditional mean assumption, is given by

$$E(y|x_1, x_2, \ldots, x_k) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \tag{17}$$

- Note that $\beta_0 = E(y|x_1 = 0, x_2 = 0, \ldots, x_k = 0)$.
- With OLS, and a random sample, we can construct an estimate of the PRF, the sample regression function (SRF) given by:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k \tag{18}$$

- The SRF can, in turn, be used for prediction and counterfactual analysis.
- For example, using the SRF, we can predict changes in y given changes in any or all of the regressors, with

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1 + \hat{\beta}_2 \Delta x_2 + \cdots + \hat{\beta}_k \Delta x_k \tag{19}$$

Interpreting OLS (cont'd)

- If we set all but $\Delta x_1 = 0$, we get

$$\Delta \hat{y} = \hat{\beta}_1 \Delta x_1 \text{ given } \Delta x_2 = \Delta x_3 = \cdots = \Delta x_k = 0 \tag{20}$$

  or equivalently

$$\frac{\Delta \hat{y}}{\Delta x_1} = \hat{\beta}_1 \text{ given } \Delta x_2 = \Delta x_3 = \cdots = \Delta x_k = 0 \tag{21}$$

- Thus, $\hat{\beta}_1$ provides an estimate of the *partial effect* of a change in x_1 on y, holding everything else constant.
- This is also known as the *ceteris paribus* effect of x_1.
- The other slope terms have a similar interpretations.

The "Partialling Out" Interpretation of $\hat{\beta}_k$

- Consider again the two regressor case, with

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \tag{22}$$

- The OLS estimator for β_1 can be constructed using 2-step process:
  1. Step 1: Estimate the following equation using OLS

$$x_1 = \delta_0 + \delta_1 x_2 + r_1 \tag{23}$$

     and form the residuals $\hat{r}_1 = x_1 - \hat{x}_1 = x_1 - \left(\hat{\delta}_0 + \hat{\delta}_1 x_2\right)$.
  2. Step 2: Estimate the following equation using OLS

$$y = \gamma_0 + \gamma_1 \hat{r}_1 + v \tag{24}$$

- One can show that $\hat{\beta}_1 = \hat{\gamma}_1$. This is numerically true, not just approximately true.

The "Partialling Out" Interpretation of $\hat{\beta}_k$ (cont'd)

- Mathematically,
$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \hat{r}_{i1} y_i}{\sum_{i=1}^{n} \hat{r}_{i1}^2} \tag{25}$$
- The residuals $\hat{r}_1$ measure the portion of x_1 that is not related to x_2; i.e., we have "partialled out" (or "controlled for") that part of x_1 that is related to x_2.
- $\hat{\beta}_1$ then captures that part of x_1 *unrelated to* x_2 that explains the variation in y.
- A similar result can be obtained for $\hat{\beta}_2$
- One can extend this result to the case with k regressors.
- The residuals $\hat{r}_1$ are then obtained by regressing x_1 on all the other regressors.

Mathematical Properties of the OLS Estimator

- The mathematical properties of the OLS estimator carry over from the simple regression model to the multiple regression model context in obvious ways.
- In particular, letting $\hat{u}_i \equiv y_i - \hat{y}_i$ denote residual for individual i,
  - The sample average of the residuals is zero and $\bar{y} = \bar{\hat{y}}$.
  - The sample covariance between each independent variable and $\hat{u}_i$ is zero and the sample covariance between $\hat{y}$ and $\hat{u}_i$ is zero.
  - The point $(\bar{x}_1, \ldots, \bar{x}_k, \bar{y})$ is on the fitted regression line.

Goodness of Fit

- We saw in the previous chapter that R^2 provides one measure of how well a given model "fits" the data, where

$$R^2 = \frac{SSE}{SST} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \in [0,1] \tag{26}$$

- The same metric can be used the multiple regression context.
- The R^2 has all of the same advantages and disadvantages we saw in the simple regression model.
- Another representation of the R^2 statistic is

$$R^2 = \frac{\left[\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})\right]}{\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right]\left[\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2\right]} \tag{27}$$

which is just the sample correlation between y_i and $\hat{y}_i$.

The Expected Value of the OLS Estimators

- Under conditions similar to those discussed in chapter 2, the OLS estimator will be unbiased.
- It is important to keep in mind the distinction between the estimator and the estimates.
  - An *estimator* is a rule for combining data to produce a numerical value for a population parameter; the form of the rule does not depend upon the particular sample obtained.
    ...In the case of OLS, the estimator is $\hat{\beta} = (\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{y}$.
  - An *estimate* is the numerical value taken on by an estimator for a particular sample of data.
  - The estimator is a random variable, because it is a function of random variables, whereas an estimate is not.

First Four Assumptions and the Unbiasedness of OLS

MLR.1 (Linear in Parameters): In the population, the dependent variable y is related to the independent variables $x_1, \ldots, x_k$ and the error u as

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u. \tag{28}$$

MLR.2 (Random Sampling): We have a random sample size of n, $\{(x_{i1}, \ldots, x_{ik}, y_i) : i = 1, \ldots, n\}$, following the population model (28).

MLR.3 (No Perfect Collinearity): None of the independent variables are constant and there are no exact *linear* relationships among the independent variables.

MLR.4 (Zero Conditional Mean): The error u has an expected value of zero given any value of the explanatory variable; i.e.,

$$E(u|x_1, \ldots, x_k) = 0 \tag{29}$$

- *Theorem 3.1*: Under assumptions MLR.1 through MLR.4, the OLS estimator is unbiased; i.e., $E(\hat{\beta}_j) = \beta_j$ for $j = 0, \ldots, k$.

Specification Errors

- There are two obvious specification errors that we might worry about:

1. Including an irrelevant variable.
   - Suppose the true model is

   $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \tag{30}$$

   and we estimate

   $$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u \tag{31}$$

   - OLS is still an unbiased estimator, since *unbiasedness* holds regardless of the true value of the parameters, even $\beta_j = 0$.
   - Including an irrelevant variable will, however, impact the variance of the OLS estimator.
2. Omitting a Relevant Variable
   - For example, if (30) is the true model and we estimate.

   $$y = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \tilde{u} \tag{32}$$

Omitted Variables Bias

- It turns out that there is a simple relationship between between the OLS estimator for $\tilde{\beta}_1$ and the OLS estimator for β_1.
- Specifically,

$$\hat{\hat{\beta}}_1 = \hat{\beta}_1 + \hat{\beta}_2\hat{\delta}_1 \tag{33}$$

where $\hat{\delta}_1$ is the OLS estimator for δ_1 in a simple regression of x_2 on x_1.
- One can, in turn, show that

$$E(\hat{\hat{\beta}}_1) = \beta_1 + \beta_2\delta_1 \tag{34}$$

- The implications:
  - OLS will be unbiased if either $\beta_2 = 0$ or $\delta_1 = 0$.
  - The size and direction of the bias will depend on the sign and size of $\beta_2\delta_1$
- With multiple regressors, the sign and size of the bias is less clear.

The Variance of the OLS Estimators

- Assumption MLR.5 (Homoskedasticity): The error u has the same variance given any values of the explanatory variables; i.e., $Var(u|x_1, \ldots, x_k) = \sigma^2$.
- Theorem 3.2: Under Assumptions MLR.1 through MLR.5

$$Var(\hat{\beta}_j|\boldsymbol{x}) = \frac{\sigma^2}{SST_j(1-R_j^2)} \tag{35}$$

where

$$SST_j = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2 \tag{36}$$

and R_j^2 is the R^2 from a regression of x_j on all the other covariates (including a constant).
- Equation (35) makes clear what factors influence $Var(\hat{\beta}_j|\boldsymbol{x})$.

The Variance of the OLS Estimators (cont'd)

- In matrix notation, using only assumptions MLR.1 through MLR.4, we have

$$\begin{aligned} Cov(\hat{\boldsymbol{\beta}}|\boldsymbol{x}) &= E\{[\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}][\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}]'|\boldsymbol{x}\} \\ &= E\{[(\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{y} - \boldsymbol{\beta}][(\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{y} - \boldsymbol{\beta}]'|\boldsymbol{x}\} \\ &= E\{[\boldsymbol{\beta} + (\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{u} - \boldsymbol{\beta}][\boldsymbol{\beta} + (\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{u} - \boldsymbol{\beta}]'|\boldsymbol{x}\} \\ &= E\{[(\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{u}][(\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{u}]'|\boldsymbol{x}\} \\ &= (\boldsymbol{x}'\boldsymbol{x})^{-1}E\{\boldsymbol{x}'\boldsymbol{u}\boldsymbol{u}'\boldsymbol{x}|\boldsymbol{x}\}(\boldsymbol{x}'\boldsymbol{x})^{-1} \end{aligned} \tag{37}$$

- These are the so-called *robust* variances.
- Adding assumption MLR.5, equation (37) reduces to

$$Cov(\hat{\boldsymbol{\beta}}|\boldsymbol{x}) = \sigma^2(\boldsymbol{x}'\boldsymbol{x})^{-1} \tag{38}$$

Estimating σ^2

- In practice, one will almost never know σ^2.
- Assumptions MLR.1 through MLR.5 are jointly known as *The Gauss-Markov assumptions*.
- Theorem 3.3: Under the Gauss-Markov assumptions, $\hat{\sigma}^2$ is an unbiased estimator of σ^2; i.e.,

$$E\left(\hat{\sigma}^2\right) = \sigma^2. \tag{39}$$

where

$$\hat{\sigma}^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}\hat{u}^2 = \frac{SSR}{n-k-1} \tag{40}$$

- The *standard error of* $\hat{\beta}_j$ becomes

$$se(\hat{\beta}_j) = \frac{\hat{\sigma}}{\sqrt{SST_j(1-R_j^2)}} \tag{41}$$

The Gauss-Markov Theorem

- Theorem 3.4: Under Assumptions MLR.1 through MLR.5, the OLS estimator $(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)$ are the best linear unbiased estimators (BLUE's) of $\beta_0, \beta_1, \ldots, \beta_k$, respectively, where
  - *best* means lowest variance;
  - *linear* refers to the fact that the estimator, say $\check{\beta}$, is a linear function of the data on the dependent variable; e.g.,

$$\check{\beta}_j = \sum_{i=1}^{n} w_{ij} y_i \tag{42}$$

- So, among all possible linear unbiased estimators, OLS provides the lowest variance estimator given assumptions MLR.1 through MLR.5.