ECON 5350 Class Notes
Least Squares

1 Introduction

We are interested in estimating the population parameters from the regression equation

$$Y = X\beta + \epsilon.$$

The population values are β, σ^2 and ϵ. Their sample counterparts are b, $\hat{\sigma}^2$ and e. The sample counterpart to the error term (ϵ) is called the residual (e). The two are related according to

$$Y = X\beta + \epsilon = Xb + e.$$

2 Least Squares

2.1 The Problem

We want to estimate the parameter β by choosing a fitting criterion that makes the sample regression line as close as possible to the data points. Our criterion is

$$\min\ e'e = (Y - Xb)'(Y - Xb) = Y'Y - b'X'Y - Y'Xb + b'X'Xb. \tag{1}$$

The criterion is minimized by choosing b. Taking the (vector) derivative with respect to b and setting equal to zero gives

$$\frac{\partial e'e}{\partial b} = -2X'Y + 2X'Xb = 0. \tag{2}$$

Provided $X'X$ is nonsingular (guaranteed by Classical assumption two), we solve to get

$$b = (X'X)^{-1}X'Y. \tag{3}$$

The second-order condition gives

$$\frac{\partial^2(e'e)}{\partial b \partial b'} = 2X'X$$

which satisfies the condition for a minimum since $X'X$ is a positive-definite matrix if X is of full rank (Greene A-114).

2.2 Example: UW Enrollment and Energy Prices

Consider the bivariate regression over the sample period 1957-2006 where the variables are

- Y = UW resident undergraduate enrollment &
- X = price of oil.

Assume the population regression equation is

$$y_t = \beta_1 + \beta_2 x_t + \epsilon_t.$$

The objective is to choose b_1 and b_2 to minimize

$$\sum_{t=1}^{T} e_t^2 = \sum_{t=1}^{T} (y_t - b_1 - b_2 x_t)^2$$

which gives the two first-order conditions

$$\frac{\partial(\sum_t e_t^2)}{\partial b_1} = -2\sum_t (y_t - b_1 - b_2 x_t) = 0 \quad (4)$$

$$\frac{\partial(\sum_t e_t^2)}{\partial b_2} = -2\sum_t (y_t - b_1 - b_2 x_t) x_t = 0. \quad (5)$$

Equations (4) and (5) can be arranged to produce the **normal equations**

$$\begin{aligned} \sum_t y_t &= T b_1 + b_2 \sum_t x_t \\ \sum_t y_t x_t &= b_1 \sum_t x_t + b_2 \sum_t x_t^2. \end{aligned}$$

Finally, solving for b_1 and b_2 gives

$$\begin{aligned} b_1 &= \bar{y} - b_2 \bar{x} \\ b_2 &= \frac{\sum_t (y_t - \bar{y})(x_t - \bar{x})}{\sum_t (x_t - \bar{x})^2}. \end{aligned}$$

This is the same answer you get via matrix algebra $b = (b_1, b_2)' = (X'X)^{-1}(X'Y)$ for appropriately defined X and Y. See MATLAB example 10 for more details.

2.3 Algebra of Least Squares

Consider the normal equations

$$X'(Y - Xb) = X'e = 0. \quad (6)$$

Three interesting results from equation 6 (assuming a constant term).

1. First column of X implies $\sum_i e_i = 0$. Positive and negative residuals exactly cancel out.

2. $\sum_i e_i = 0$ implies that $\bar{e} = \bar{Y} - \bar{X}b = 0$, which implies $\bar{Y} = \bar{X}b$. The regression hyperplane passes through the sample mean.

3. $\hat{Y}'e = (Xb)'e = b'X'e = 0$. The fitted values are orthogonal to the residuals.

2.4 Partitioned and Partial Regressions

Let a regression have two sets of explanatory variables, X_1 and X_2, such that

$$Y = X_1\beta_1 + X_2\beta_2 + \epsilon.$$

The normal equations can be written in partitioned form as

$$\begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} X_1'Y \\ X_2'Y \end{bmatrix}.$$

Solving for b_2 gives

$$\begin{aligned} b_2 &= [X_2'(I - X_1(X_1'X_1)^{-1}X_1')X_2]^{-1}[X_2'(I - X_1(X_1'X_1)^{-1}X_1')Y] \\ &= [X_2'M_1X_2]^{-1}[X_2'M_1Y], \end{aligned}$$

where $M_1 = I - X_1(X_1'X_1)^{-1}X_1'$ can be interpreted as a **residual-maker matrix**, (i.e., premultiplying any conformable matrix by M_1 will generate the residuals associated with a regression on X_1). Note the following:

- Define $e_{Y1} = M_1Y$.
- Define $e_{21} = M_1X_2$.
- M_1 is symmetric and idempotent (i.e., $M_1 = M_1'M_1 = M_1M_1$).

This implies that we can write

$$\begin{aligned} b_2 &= [X_2'M_1X_2]^{-1}[X_2'M_1Y] \\ &= [e_{21}'e_{21}]^{-1}[e_{21}'e_{Y1}]. \end{aligned}$$

This is the result that makes multiple regression analysis so powerful for applied economics. We can interpret b_2 as the impact of X_2 on Y while "partialing or netting out" the effect of X_1. The results for b_1 are analogous.

2.5 Goodness of Fit and Analysis of Variance

We will now assess how well the regression model fits the data. Begin by writing the sample regression equation $Y = Xb + e$ in deviation from its mean form using the following matrix

$$M^0 = (I_n - \frac{1}{n}ii') = \begin{bmatrix} 1-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & \vdots & & \vdots \\ -\frac{1}{n} & -\frac{1}{n} & \cdots & 1-\frac{1}{n} \end{bmatrix}$$

where i is the unit column vector. We can then write

$$Y - \bar{Y} = M^0 Y = M^0(Xb + e) = M^0 Xb + e. \tag{7}$$

Premultiplying (7) by itself transposed, and noting that M^0 is a symmetric and idempotent matrix, gives

$$(Y - \bar{Y})'(Y - \bar{Y}) = Y' M^0 Y = b' X' M^0 X b + e' e$$

or $SST = SSR + SSE$, where the three terms stand for total, regression and error sum of squares, respectively.

A natural measure of goodness of fit is

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}.$$

A few notes about R^2

- $0 \leq R^2 \leq 1$.
- By adding additional explanatory variables, you can never make R^2 smaller.
- An alternative measure is $\bar{R}^2 = 1 - \frac{SSE/(n-k)}{SST/(n-1)}$, the adjusted R^2. This measure adds a penalty for additional explanatory variables.
- Be cautious interpreting R^2 when no constant is included.
- Value of R^2 will depend on the type of data (e.g., cross-sectional data tends to produce low R^2s and time series data often produces high R^2s).
- Comparing R^2s requires comparable dependent variables.