Multiple Regression Analysis

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$

1. Estimation

Ch.3 Multiple Regression: Estimation

1. Motivation for Multiple Regression
2. Mechanics & Interpretation of OLSE
3. The Expected value of the OLSE
4. The Variances of the OLSE
5. Efficiency of OLS: The Gauss-Markov Theorem

3.1 Motivation for Multiple Regression

The model with 2 regressors

- We are interested in the effect of *educ* on *wage*.

  $wage = \beta_0 + \beta_1\ educ + \beta_2\ exper + u$ (3.1)
- (3.1) takes *exper* out of the error term and put it explicitly in the equation.
- Thus, we don't have to assume *exper* is uncorrelated with *educ*. (Recall a zero conditional mean assumption, $E(u|\mathbf{x}) = 0$.)

The model with *k* Regressors

- The general ***multiple linear regression model*** in population is

  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u.$ (3.6)
- Still we need to make a zero conditional mean assumption, so now assume that

  $E(u|x_1,x_2, \ldots,x_k) = 0.$ (3.8)

3.2 Mechanics & Interpretation of OLS

Interpreting Multiple Regression

- The case with *k* independent variables

  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u,$

  so holding $x_2, \ldots, x_k$ fixed implies that

  $\frac{\partial y}{\partial x_1} = \beta_1$, that is each β_i has a *ceteris paribus* interpreta tion.

A "Partialling Out" Interpretation

Consider the case where $k = 2$, i.e.

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$, then

$\hat{\beta}_1 = \left(\sum \hat{r}_{i1} y_i\right) / \sum \hat{r}_{i1}^2$ (3.22)

where $\hat{r}_{i1}$ are the residuals from the estimated regression

$x_{i1} = \hat{\gamma}_0 + \hat{\gamma}_1 x_{i2} + \hat{r}_{i1}$

Cont. "Partialling Out"

- Previous equation implies that regressing y on x_1 and x_2 gives same effect of x_1 as regressing y on residuals from a regression of x_1 on x_2.
- This means only the part of x_{i1} that is uncorrelated with x_{i2} are being related to y_i so we're estimating the effect of x_1 on y after x_2 has been "***partialled out***".

Simple vs. Multiple Regression Estimates

Compare the simple regression $\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$

with the multiple regression $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$:

$$\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \tilde{\delta}_1 \quad (3.23)$$

where $\tilde{\delta}_1$ is the slope coefficient from the simple regression of x_{i2} on x_{i1}. Generally, $\tilde{\beta}_1 \neq \hat{\beta}_1$ unless:

$\hat{\beta}_2 = 0$ (i.e. no partial effect of x_2), or

$\tilde{\delta}_1 = 0$ (x_1 and x_2 are uncorrelated in the sample).

Cont. Simple vs. Multiple Reg. Estimate

$$\tilde{\beta}_1 = \frac{\sum (x_1 - \bar{x})y}{\sum (x_1 - \bar{x})^2} \text{ (simple)}$$

$$\hat{\beta}_1 = \frac{\sum (\hat{r}_1 - \bar{\hat{r}}_1)y}{\sum (\hat{r}_1 - \bar{\hat{r}}_1)^2} \text{ (multiple)}$$

where $\hat{r}_1 = x_1 - (\hat{\gamma}_0 + \hat{\gamma}_2 x_2)$, that is, the residual from a regression of x_1 on x_2, so $\bar{\hat{r}}_1 = 0$.

Goodness-of-Fit

- Just as in the simple regression, R^2 is defined to be (3.29).
- R^2 will usually increase with the number of independent variables. Therefore, it is not a good way to compare models.
- Adjusted R^2 is often used in multiple regression analysis (see the detail in ch.6).

$$\bar{R}^2 = 1 - (1 - R^2)\frac{(n-1)}{(n-k-1)}$$

3.3 Expected Values of OLS Estimators

Assumptions for Unbiasedness

1. Population model is linear in parameters:
   $y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u \quad (3.31)$
2. Random Sampling:
   We can use a random sample of n observations, $\{(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i)$: $i=1, 2, \ldots, n\}$, from the population model.
   Therefore, the sample model is
   $y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + u_i \quad (3.32)$

Cont. Assumptions for Unbiasedness

3. No Perfect Collinearity:
   None of the regressors is constant, and there are ***no exact linear*** relationships among them.
4. Zero Conditional Mean:
   All of the explanatory variables are ***exogenous***.
   $E(u|x_1, x_2, \ldots, x_k) = 0 \quad (3.36)$.

- Under these 4 assumptions, the OLSE are unbiased estimators of the population parameters.
  $E(\hat{\beta}_j) = \beta_j, \; j = 0,1,\ldots,k \quad (3.37)$

Including Irrelevant Variables

- There is no effect on our parameter estimate, and OLSE remains unbiased.

$$Y_i = \beta_0 + \beta_1 X_1 + u \text{ (population)}$$
$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u \text{ (misspecification)}$$
$$\Rightarrow \mathrm{E}(\hat{\beta}_1) = \beta_1,\ \mathrm{E}(\hat{\beta}_2) = 0$$

- But including irrelevant variables can have undesirable effects on the *variances* of the OLSE.

Omitted Variables Bias

- OLSE will usually be biased with the problem of excluding a relevant variable.

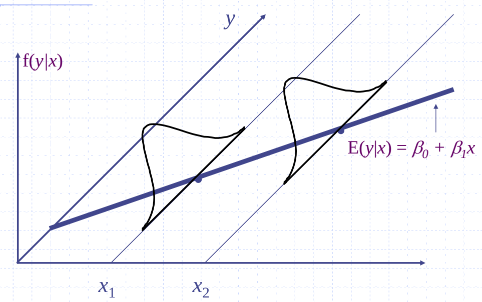
$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u \text{ (population)}$$
$$Y_i = \beta_0 + \beta_1 X_1 + u \text{ (misspecification)}$$
$$\Rightarrow \mathrm{E}(\tilde{\beta}_1) = \beta_1 + \beta_2 \tilde{\delta}_1 \neq \beta_1 \quad (3.45)$$

- Two cases where bias is equal to zero
  - $\beta_2 = 0$, that is x_2 doesn't really belong in model
  - x_1 and x_2 are uncorrelated in the sample
- Table 3.2 shows the possible sign of the bias.

3.4 Variance of the OLSE

- Now we know that the sampling distribution of our estimate is centered around the true parameter.
  - It's important to think about how spread out this distribution is.
  - It is much easier to think about this variance under an additional assumption, so
- Assume Var(u|x) = σ^2 (Homoskedasticity)

Homoskedasticity (simple regression)



Cont. Variance of OLSE

- Let $\boldsymbol{x}$ stand for (x_1, x_2,…x_k).
- Assuming that Var(u|$\boldsymbol{x}$) = σ^2 also implies that Var(y|$\boldsymbol{x}$) = σ^2.
- The 4 assumptions for unbiasedness, plus this homoskedasticity assumption are known as the ***Gauss-Markov assumptions***.

Cont. Variance of OLS

Given the Gauss-Markov Assumptions,

$$Var\left(\hat{\beta}_j\right) = \frac{\sigma^2}{SST_j\left(1 - R_j^2\right)}, \qquad (3.51)$$

where

$SST_j = \sum \left(x_{ij} - \bar{x}_j\right)^2$ and R_j^2 is the R^2 from regressing x_j on all other x's.

Components of OLS Variances

- The error variance: a larger σ^2 implies a larger variance for the OLS estimators
- The total sample variation: a larger SST_j implies a smaller variance for the estimators
- Linear relationships among the independent variables: a larger R_j^2 implies a larger variance for the estimators
  - **Multicollinearity**: High (but not perfect) correlation among regressors.

Variances in Misspecified Models

Consider again the misspecified model

$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$, so that

$$Var(\tilde{\beta}_1) = \sigma^2 / SST_1 \quad (3.55)$$

$$\Leftrightarrow Var(\hat{\beta}_1) = \sigma^2 / [SST_1 (1 - R_1^2)] \quad (3.54)$$

Thus, $Var(\tilde{\beta}_1) < Var(\hat{\beta}_1)$, unless x_1 and x_2 are uncorrelated, then they're the same.

Cont. Misspecified Models

- While the variance of the estimator is smaller for the misspecified model, unless $\beta_2 = 0$ the misspecified model is biased.
- As the sample size grows, the variance of each estimator shrinks to zero, making the variance difference less important.

Estimating the Error Variance

- Like simple regression model, we need to estimate σ^2 to obtain ***standard error*** (or the regression and estimators.)

$$\hat{\sigma}^2 = \left(\sum \hat{u}_i^2\right) / (n - k - 1) \quad (3.56)$$

$$se\left(\hat{\beta}_j\right) = \hat{\sigma} / \left[SST_j \left(1 - R_j^2\right)\right]^{1/2} \quad (3.58)$$

- The ***degrees of freedom*** (*df*) is:

$df = n - (k + 1)$ (3.57)

$= (\text{\# of obs.}) - (\text{\# of est. parameters})$

The Gauss-Markov Theorem

- Given our 5 Gauss-Markov Assumptions it can be shown that OLS is "BLUE."
  - **B**est (smallest variance)
  - **L**inear (see (3.59))
  - **U**nbiased
  - **E**stimator

  See the proof on appendix 3A.6.
- Thus, if the assumptions hold, use OLS.