ES465 – Econometrics

Dr. Todd Richard Yarbrough

Lecture Note #4

IV. Multivariate Linear Regression

1. The Model

The multiple linear regression model is:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u \tag{1}$$

there are k independent variables

The key assumption for the general multiple regression model is easy to state in terms of a condition expectation:

$$E(u|x_1, x_2, ..., x_k) \tag{2}$$

This means, at minimum, that all factors in the unobserved error term be uncorrelated with the explanatory variables. If equation (2) holds, then we say that the OLS estimates are "unbiased".

So, the OLS estimator will run the following regression:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + ... + \hat{\beta}_k x_k \tag{3}$$

and try to minimize:

$$\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_2 - ... - \hat{\beta}_k x_k)^2 \tag{4}$$

That is to say, that the OLS estimates of $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k$ are chosen such that they minimize the sum of squared residuals. Recall,

$$\hat{y} - y_i = u_i$$

$$SSR = \sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_2 - ... - \hat{\beta}_k x_k)^2$$

There are always $k + 1$ OLS estimates, including the intercept $\hat{\beta}_0$.

2. Interpretation

Recall from above:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + ... + \hat{\beta}_k x_k \tag{5}$$

Rewritten in terms of changes:

$$\Delta\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \Delta x_1 + \hat{\beta}_2 \Delta x_2 + ... + \hat{\beta}_k \Delta x_k \tag{6}$$

So,

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1 \tag{7}$$

if we hold $x_2, x_3, ..., x_k$ constant. This is another example of the "all else equal" or Ceteris Paribus assumption.

We can also think about this concept with respect to "partialling out" estimates from our OLS regression. Suppose that we have a model with two independent variables:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \tag{8}$$

Then one can express $\hat{\beta}_1$ like so:

$$\hat{\beta}_1 = (\sum_{i=1}^{n} \hat{r}_{i1}) / (\sum_{i=1}^{1} \hat{r}_{i1}^2) \tag{9}$$

where $\hat{r}_{i1}$ are the OLS residuals from a simple regression of x_1 on x_2 from our given sample.

3. The Gauss-Markov Assumptions

Just like with the single regression OLS, we have properties and assumptions that hold when performing OLS with mulitple variables. Let's first start out with our 3 OLS properties from Lecture 3:

OLS.1: The sample average of the residuals is zero. That is, $y_i - \hat{y}_i$

OLS.2: The sample covariance between each x variable and the OLS residuals is zero.

OLS.3: The point $(\bar{x}_1, \bar{x}_2, ..., \bar{x}_k, y)$ is always on the OLS regression line: $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1\bar{x}_1 + \hat{\beta}_2\bar{x}_2 + \hat{\beta}_k\bar{x}_k$

Now, let's turn to the Multiple Linear Regression Assumptions:

MLR.1: Our model is linear in its parameters and can be written as: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u$

MLR.2: Our model is estimated with randomly sampled with n observations, $\{(x_{i1}, x_{i2}, ..., x_{ik}) : i = 1, 2, ..., n\}$

MLR.3: No Perfect Collinearity – In the sample, none of the independent variables is constant, and there are no exact linear relationship among the independent variables.

MLR.4: Zero Conditional Mean:

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

The Theory of Unbiased Estimators in OLS

Under ***MLR.1-MLR.4*** we have:

$$E(\hat{\beta}_j) = \beta_j, j = 0, 1, ..., k \tag{11}$$

This just says that our OLS estimates are unbiased and are therefore our population parameters as well. We will of course have a hard time meeting all of these assumptions, and will in the future run tests to make sure that they hold. Remember, we're trying to generalize our OLS estimates to the population at large, and this can only occur if we first assume MLR.1-MLR.4 hold.

Be sure to read up on *Omitted Variable Bias* in Chpt. 3 of the Wooldridge text (pg. 88-92).

MLR.5: Homoskedasticity – The error u has the same variance given any values of the explanatory variables. In other words, $Var(u|x_1, ..., x_k) = \sigma^2$

We are also interested in the sample variance of our OLS slope estimates:

$$Var(\hat{\beta}_i) = \frac{\sigma^2}{SST_j(1 - R_j^2)} \tag{12}$$

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

One particularly important issue is "**multicollinearity**". This refers to high, but not perfect, correlation between two or more independent variables.

Under MLR.1 - MLR.5 (The Gauss-Markov Assumptions) we can say that

$$E(\hat{\sigma}^2) = \sigma^2$$

where $\hat{\sigma}^2$ is the estimated variance and $\hat{\sigma}$ is the estimated standard deviation, and is also called the **Standard Error of the Regression** (SER).

And, we also have the standard error of our coefficient estimates:

$$se(\hat{\beta}_j) = \sigma/[SST_j(1 - R^2)]^{\frac{1}{2}} \tag{13}$$

When the Gauss-Markov Assumptions hold, we say that our estimates are the Best Linear Unbiased Estimators (BLUE).