Multiple Regression Analysis: Estimation. Wooldridge (2013), Chapter 3

- Variance of OLS Estimators
- The Gauss-Markov Theorem
- Variance of the OLS Estimators - Misspecified Models
- Estimating the Error Variance
- Incorporating Non-linearities

Multiple Regression Analysis: Inference. Wooldridge (2013), Chapter 4 and Chapter 6 (section 6.4)

- Introduction
- Assumptions of the Classical Linear Model
- The t Test
- Hypothesis testing - one-sided alternatives
- Hypothesis testing - two-sided alternatives
- Confidence Intervals

Multiple Regression Analysis: Estimation. Wooldridge (2013), Chapter 3

- Variance of OLS Estimators
- The Gauss-Markov Theorem
- Variance of the OLS Estimators - Misspecified Models
- Estimating the Error Variance
- Incorporating Non-linearities

The Variances of the OLS Estimator conditional on the sample values $\{x_i : i = 1, 2, \ldots, n\}$ are given by

$$\begin{aligned} Var\left(\hat{\beta}_0\right) &= \frac{\sigma^2}{(n-1)\, n\, S_x^2} \sum_{i=1}^{n} x_i^2, \\ Var\left(\hat{\beta}_1\right) &= \frac{\sigma^2}{(n-1)\, S_x^2}, \end{aligned}$$

where $S_x^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})^2$.

Proof that $Var\left(\hat{\beta}_1|\tilde{x}\right) = \frac{\sigma^2}{(n-1)S_x^2}$, where $\tilde{x} = (x_1, x_2, ..., x_n)$

$$\begin{aligned}
Var\left(\hat{\beta}_1|\tilde{x}\right) &= Var\left(\frac{\sum_{i=1}^n (x_i-\bar{x})\, y_i}{\sum_{i=1}^n (x_i-\bar{x})^2}|\tilde{x}\right) \\
&= \frac{1}{\left[\sum_{i=1}^n (x_i-\bar{x})^2\right]^2} Var\left(\sum_{i=1}^n (x_i-\bar{x})\, y_i|\tilde{x}\right).
\end{aligned}$$

Now notice that for $i \neq j$

$$\begin{aligned}
cov\left((x_i-\bar{x})\, y_i, (x_j-\bar{x})\, y_j|\tilde{x}\right) &= \\
(x_i-\bar{x})\left(x_j-\bar{x}\right) cov\left(y_i, y_j|\tilde{x}\right) &= 0
\end{aligned}$$

because y_i and y_j are independent.

Therefore

$$\begin{aligned} Var\left(\hat{\beta}_1|\tilde{x}\right) &= \frac{1}{\left[\sum_{i=1}^n (x_i-\bar{x})^2\right]^2} Var\left(\sum_{i=1}^n (x_i-\bar{x})\, y_i|\tilde{x}\right) \\ &= \frac{1}{\left[\sum_{i=1}^n (x_i-\bar{x})^2\right]^2} \sum_{i=1}^n Var\left((x_i-\bar{x})\, y_i|\tilde{x}\right) \\ &= \frac{1}{\left[\sum_{i=1}^n (x_i-\bar{x})^2\right]^2} \sum_{i=1}^n (x_i-\bar{x})^2\, Var\left(y_i|\tilde{x}\right). \end{aligned}$$

Now

$$\begin{aligned} Var\left(y_i|\tilde{x}\right) &= Var\left(y_i|x_1, x_2, ..., x_i, ..., x_n\right) \\ &= Var\left(y_i|x_i\right) \end{aligned}$$

because y_i is independent from x_j for $j \neq i$ because we assumed that we use a random sample $\{(x_i, y_i)\}_{i=1}^n$.

Now

$$\begin{aligned} Var\left(y_i|x_i\right) &= Var\left(\beta_0+\beta_1 x_i+u_i|x_i\right) \\ &= Var\left(u_i|x_i\right) \\ &= \sigma^2 \end{aligned}$$

because $Var\left(u_i|x_i\right)=\sigma^2$, therefore

$$\begin{aligned} Var\left(\hat{\beta}_1|\tilde{x}\right) &= \frac{1}{\left[\sum_{i=1}^n\left(x_i-\bar{x}\right)^2\right]^2}\sum_{i=1}^n\left(x_i-\bar{x}\right)^2 Var\left(y_i|\tilde{x}\right) \\ &= \frac{1}{\left[\sum_{i=1}^n\left(x_i-\bar{x}\right)^2\right]^2}\sum_{i=1}^n\left(x_i-\bar{x}\right)^2\sigma^2 \\ &= \frac{\sigma^2}{\left[\sum_{i=1}^n\left(x_i-\bar{x}\right)^2\right]^2}\sum_{i=1}^n\left(x_i-\bar{x}\right)^2 \\ &= \frac{\sigma^2}{\sum_{i=1}^n\left(x_i-\bar{x}\right)^2} \end{aligned}$$

$$
\begin{aligned}
Var\left(\hat{\beta}_1|\tilde{x}\right) &= \frac{\sigma^2}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2} \\
&= \frac{\sigma^2}{\left(n-1\right)\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2} \\
&= \frac{\sigma^2}{\left(n-1\right)S_x^2}
\end{aligned}
$$

Given our 5 Gauss-Markov Assumptions it can be shown that OLS is "BLUE":

- Best (have minimum variance, such that $Var(\hat{\beta}_j) \leq Var(\hat{\beta}_j^*)$, $j = 1, ..., k$ where $\hat{\beta}_j^*$ is any alternative estimator.
- Linear - weighted sum of the dependent variable.
- Unbiased- $E(\hat{\beta}_j) = \beta_j$, $E(\hat{\beta}_j^*) = \beta_j$, $j = 1, ..., k$.
- Estimator.

Thus, if the assumptions hold, use OLS.

We prove here the The Gauss-Markov Theorem in the case of the simple linear regression model for the estimator of the slope parameter.

An estimator is said to be linear if it can be written as a simple weighted sum of the dependent variable, where the weights do not depend on this variable.

Consider the estimator for the slope coefficient

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \\ &= \frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\sum_{i=1}^{n}x_i(x_i-\bar{x})} \\ &= \sum_{i=1}^{n} w_i y_i, \end{aligned}$$

where

$$w_i = \frac{x_i-\bar{x}}{\sum_{i=1}^{n}x_i(x_i-\bar{x})}.$$

Outline of the proof:

1. Consider an alternative linear unbiased estimator.
2. Show that the new estimator can never have smaller variance than the OLS estimator.

Step 1: An alternative linear estimator for the slope coefficient will have the form

$$\bar{\beta}_1 = \sum_{i=1}^{n} k_i y_i.$$

where k_i is a function of the regressors. Unbiasedness means that $E\left(\bar{\beta}_1\right) = \beta_1$ and this requires that the weights should satisfy

$$\sum_{i=1}^{n} k_i = 0 \text{ and } \sum_{i=1}^{n} k_i x_i = 1.$$

Step 2: Notice that conditional on the sample values $\{x_i : i = 1, 2, \dots, n\}$ we have

$$Var\left(\bar{\beta}_1\right) = \sigma^2 \sum_{i=1}^{n} k_i^2, \; Var\left(\hat{\beta}_1\right) = \frac{\sigma^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}.$$

Hence

$$\begin{aligned} Var\left(\bar{\beta}_1\right) - Var\left(\hat{\beta}_1\right) &= \sigma^2 \sum_{i=1}^{n} k_i^2 - \frac{\sigma^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} \\ &= \sigma^2 \left(\sum_{i=1}^{n} k_i^2\right) [1 - \frac{1}{\sum_{i=1}^{n} k_i^2 \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}] \\ &= \sigma^2 \left(\sum_{i=1}^{n} k_i^2\right) [1 - \frac{\left(\sum_{i=1}^{n} k_i x_i\right)^2}{\sum_{i=1}^{n} k_i^2 \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}] \\ &= \sigma^2 \left(\sum_{i=1}^{n} k_i^2\right) [1 - correlation(x_i, k_i)^2] \geq 0 \end{aligned}$$

Suppose that we know that the model is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

where $E(u|x_1, x_2) = 0$.

- Consider again $\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$ so that

$$Var(\tilde{\beta}_1) = \frac{\sigma^2}{(n-1)\, S^2_{x_1}}.$$

  where $S^2_{x_1}$ is the sample variance of x_1.
- Recall that

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{(n-1)\, S^2_{x_1}\, (1 - R^2_1)}.$$

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

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

where the $SST_j = \sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2$ and R_j^2 is the R^2 from the regressing x_j on all other $x's$.

- We don't know what the error variance, σ^2, is, because we don't observe the errors, u_i.
- What we observe are the residuals, $\hat{u}_i$.
- We can use the residuals to form an estimate of the error variance:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\hat{u}_i^2}{n-k-1},$$

thus

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

- $se\left(\hat{\beta}_j\right)$ is called the standard error of $\hat{\beta}_j$.
- The square root of $\hat{\sigma}^2$ is called the regression standard error, or standard error of the regression
- $df = n - (k+1)$, or $df = n - k - 1$. df (i.e. degrees of freedom) is the (number of observations) – (number of estimated parameters). Therefore $\hat{\sigma}^2 = SSR/df$.

So far we have focussed on linear relationships between the dependent and independent variable, however in applied work in Economics we often encounter regression equations where the dependent variable appears in logarithmic form.
Why is this done?
Recall the Wages-Education example:

$$Wages = \beta_0 + \beta_1 Education + u,$$

$E\left[u|Education\right] = 0$. The sample regression function obtained was

$$\widehat{Wages} = -1.60468 + 0.81395 \times Education.$$

Notice that this implies that:

- For a person with 6 years of Education, an additional year will increase the hourly wages by \$0.81395.
- For a person with 15 years of Education, an additional year will increase the hourly wages by \$0.81395.

Conclusion: This may not be reasonable.

- In empirical research it has been found that a better characterization of how the wages change is to assume that each year of education increases wages by a constant percentage.
- A model that gives (approximately) a constant percentage effect is:

$$\log(Wages) = \beta_0 + \beta_1 Education + u,$$

Why?
- The key reason lies in the following fact: If $\Delta y / y$ is close to zero:

$$\log(y + \Delta y) - \log(y) \cong \frac{\Delta y}{y}$$

that is the difference between the natural logarithm of $y + \Delta y$ and the natural logarithm of y is the percentage change divided by 100.
- Consider the linear regression model that where the dependent variable is in the logarithm form (known as *log-linear model*)

$$\log(y) = \beta_0 + \beta_1 x + u.$$

- Let us drop the error term u for simplicity

$$\log(y) = \beta_0 + \beta_1 x$$

  and denote Δy be the change in y when x changes by Δx.

- One can show that in this model

$$\frac{\Delta y}{y} \cong \beta_1 \Delta x$$

  In words: a unit change in x is associated with a $100 \times \beta_1\%$ expected change in y.

Running the regression of log(*Wages*) on *Education* we obtain:

Dependent variable: log(*Wages*)

Estimation Method: Ordinary Least Squares

Regressors	Estimates
Intercept	0.98237
Education	0.08262

- Hence, an additional year of education is expected to increase the hourly wages by 8.262%.
- For a person with 6 years of Education, an additional year is expected to increase the hourly wages by $0.08262 \times wages_6$ dollars, where $wages_6$ are the wages of that person.
- For a person with 15 years of Education, an additional year is expected to increase the hourly wages by $0.08262 \times wages_{15}$ dollars, where $wages_{15}$ are the wages of that person.

Other cases: *Linear-Log model*

- x is in logarithms and y is not, that is

$$y = \beta_0 + \beta_1 \log(x) + u.$$

  Denote Δy be the change in y when x changes by Δx.

- *Ignoring the error term* one can show that

$$\Delta y \cong \beta_1 \Delta x / x.$$

In words: a 1% change in x is associated with a $0.01 \times \beta_1$ expected change in y.

Other cases: *Log-Log model* or *constant elasticity model*

- Both x and y are in logarithms, that is

$$\log(y) = \beta_0 + \beta_1 \log(x) + u.$$

  Denote Δy be the change in y when x changes by Δx.

- *Ignoring the error term* one can show that

$$\Delta y / y \cong \beta_1 \Delta x / x.$$

In words: a 1% change in x is associated with a β_1% expected change in y (β_1 is the elasticity of y with respect to x).

Example: Economists often fit models that take logs of variables such as:

$$\log(Output) = \beta_0 + \beta_1 \log(Labour) + \beta_2 \log(Capital) + u,$$

Ignoring the error term u, this model corresponds to the *Cobb-Douglas production function*. That is, it is equivalent to

$$\begin{aligned} Output &= A \times Labour^{\beta_1} \times Capital^{\beta_2}, \\ A &= \exp(\beta_0). \end{aligned}$$

Thus:

- - β_1 is the elasticity of *Output* with respect to *Labour*.
  - β_2 is the elasticity of *Output* with respect to *Capital*.

Incorporating Non-linearities - Why use log models?

- Log models are invariant to the scale of the variables.
- They give a direct estimate of elasticity (if both the dependent variable and regressors are in logarithms).
- For models with $y > 0$, the conditional distribution is often heteroskedastic or asymmetric, while $\log(y)$ is much less so.
- The distribution of $\log(y)$ is more narrow, limiting the effect of outliers.

- What types of variables are often used in log form?
  - Values measured in a currency that must be positive.
  - Very large variables, such as population.
- What types of variables are often used in level form?
  - Variables measured in years. **Example:** Education, Experience, tenure, age, etc.
  - Variables that are a proportion or percent.

Multiple Regression Analysis: Inference. Wooldridge (2013), Chapter 4 and Chapter 6 (section 6.4)

- Introduction
- Assumptions of the Classical Linear Model
- The t Test
- Hypothesis testing - one-sided alternatives
- Hypothesis testing - two-sided alternatives
- Confidence Intervals

Examples of test of hypothesis

Consider the model:

$$bwgth = \beta_0 + \beta_1 cigs + \beta_2 educ + \beta_3 npvis + \beta_4 age + u,$$

where

bwgth	-birth weight,
cigs	-cigarettes smoked per day while pregnant,
educ	-years of schooling for the mother,
npvis	-total number of prenatal visits,
age	-Age of the mother.

$$bwgth = \beta_0 + \beta_1 cigs + \beta_2 educ + \beta_3 npvis + \beta_4 age + u,$$

- Is the partial effect of *age* relevant after controlling for *cigs*, *education* and *npvis*?

$$H_0 : \beta_4 = 0 \; vs \; H_1 : \beta_4 \neq 0,$$

[Individual statistical significance]

- Is the effect of smoking 10 cigarettes canceled by the effect of one more prenatal visit?

$$H_0 : 10\beta_1 + \beta_3 = 0 \; vs \; H_1 : 10\beta_1 + \beta_3 \neq 0,$$

[single linear combination of parameters]

$$bwgth = \beta_0 + \beta_1 cigs + \beta_2 educ + \beta_3 npvis + \beta_4 age + u,$$

- Are the partial effects of *age*, *education* and *npvis* jointly irrelevant after controlling for the number of cigarettes smoked?

$$H_0 : \beta_2 = \beta_3 = \beta_4 = 0$$

vs

$$H_1 : \beta_2 \neq 0 \text{ and/or } \beta_3 \neq 0 \text{ and/or } \beta_4 \neq 0,$$

[jointly statistical significance; Exclusion restrictions]

- Is there any variable in the equation relevant to explain the birth weight?

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$

vs

$$H_1 : \beta_1 \neq 0 \text{ and/or } \beta_2 \neq 0 \text{ and/or } \beta_3 \neq 0 \text{ and/or } \beta_4 \neq 0,$$

[Overall significance of the regression]

- So far, we know that given the Gauss-Markov assumptions, OLS is BLUE,
- In order to do classical hypothesis testing, we need to add another assumption (beyond the Gauss-Markov assumptions),
- Assume that u is independent of $x_1, x_2, \ldots, x_k$ and u is normally distributed with zero mean and variance σ^2: $u \sim N\left(0, \sigma^2\right)$.

- Under CLM, OLS is not only BLUE, but is the minimum variance unbiased estimator:
  - BLUE means that the OLS estimator is the most efficient among the class of linear unbiased estimators.
  - Under CLM the OLS estimator is the most efficient among all unbiased estimators.
- We can summarize the population assumptions of CLM as follows

$$y|x \sim N(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k, \sigma^2).$$

- While for now we just assume normality, clear that sometimes not the case.
- Large samples will let us drop normality.

Under the CLM assumptions, conditional on the sample values of the independent variables for $j = 0, ..., k$

$$\hat{\beta}_j \sim N(\beta_j, Var(\hat{\beta}_j)),$$

so that

$$\frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)} \sim N(0, 1),$$

where $sd(\hat{\beta}_j) = \sqrt{Var(\hat{\beta}_j)}$.

$\hat{\beta}_j$ is distributed normally because it is a linear combination of independent errors that have the normal distribution.