Section 2 Notes

Elizabeth Stone and Charles Wang

January 15, 2009

1 Ordinary Least Squares and Its Distribution

In econometrics, we care about some economic question (e.g. what are the determinants of income - age, education, race, gender?), and we collect some data in order to test our hypotheses.

We begin by making an assumption about how the data (which you have collected) is being generated. Typically, we assume that the data that you collect $\{y_i, \overrightarrow{x}_i\}$ satisfy the following linear equation:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi} + \varepsilon_i$$

or in matrix notation

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i$$

$$\text{where } \mathbf{x}_i^T = (1, x_{1i}, x_{2i}, ..., x_{pi})_{p\times 1} \text{ and } \boldsymbol{\beta} = \left(\beta_0, \beta_1, ..., \beta_p\right)^T$$

and we typically assume the error term satisfy the following assumption (called "mean independence"):

$$E\left(\varepsilon_i | \mathbf{x}_i\right) = 0$$

Note1: The mean independence assumption is KEY for econometrician. We can ALWAYS write any $\{y_i, \overrightarrow{x}_i\}$ to satisfy the equation $y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i$. (Just define ε_i as the $y_i - \mathbf{x}_i^T \boldsymbol{\beta}$). However, ε_i may not satisfy the special condition of mean independence.

Note2: Keep in mind that the mean independence condition implies that $Cov\left(\varepsilon_i, \mathbf{x}_i\right) = 0$ (i.e. the error is uncorrelated with each of the right hand side variables)

1.1 Deriving the OLS Formula

Again, we know that each observation $\left\{y_i, \mathbf{x}_i^T\right\}$ is generated from the following equation

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i$$

Goal: Come up with an expression for β
Steps:

1. Multiply both sides by $\mathbf{x}_i$

$$\mathbf{x}_i y_i = \mathbf{x}_i \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{x}_i \varepsilon_i$$

2. Take expectations on both sides of the equation

$$E\left(\mathbf{x}_i y_i\right) = E\left(\mathbf{x}_i \mathbf{x}_i^T\right) \boldsymbol{\beta} + E\left(\mathbf{x}_i \varepsilon_i\right)$$

3. Since $E\left(\varepsilon_i|\mathbf{x}_i\right)=0 \Rightarrow E\left(\varepsilon_i\mathbf{x}_i\right)=0$, we have

$$E\left(\mathbf{x}_i y_i\right) = E\left(\mathbf{x}_i\mathbf{x}_i^T\right)\boldsymbol{\beta}$$

4. Since $E\left(\mathbf{x}_i\mathbf{x}_i^T\right)$ is a kxk matrix that is invertible (we can assume this for now), we can multiply both sides by its inverse to get:

$$\boldsymbol{\beta} = E\left(\mathbf{x}_i\mathbf{x}_i^T\right)^{-1} E\left(\mathbf{x}_i y_i\right)$$

   This is an expression for the population coefficients $\boldsymbol{\beta}$.

When we collect data and run an ordinary least squares (OLS) regression, our estimate for $\boldsymbol{\beta}$ is very similar, except that we replace expectation by a sample average.

$$\hat{\boldsymbol{\beta}}_{OLS} = \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^T\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i y_i\right)$$

where i indexes observations of $\left\{y_i, \mathbf{x}_i^T\right\}$ and N is the total number of observations.

Since we know that $y_i = \mathbf{x}_i^T\boldsymbol{\beta}+\varepsilon_i$, we can rewrite this expression as:

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{OLS} &= \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^T\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\left[\underbrace{\mathbf{x}_i^T\boldsymbol{\beta}+\varepsilon_i}_{y_i}\right]\right) \\
&= \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^T\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^T\right)\boldsymbol{\beta} + \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^T\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\varepsilon_i\right) \\
&= \boldsymbol{\beta} + \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^T\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\varepsilon_i\right) \quad (***)
\end{aligned}$$

1. **THIS IS THE KEY FORMALA TO KNOW TO SHOW CONSISTENCY AND ASYMPTOTIC DISTRIBUTION OF OLS!**

1.2 Showing Consistency (i.e. $\hat{\beta}_{OLS} \rightarrow_P \beta$)

Consistency is a very basic property that we would like our regression coefficients to have. It says that if I have an infinite amount of data, then I should be able to compute the TRUE population coefficients ($\boldsymbol{\beta}$) exactly.

1. Start with our expression of OLS estimator: $\hat{\boldsymbol{\beta}}_{OLS} = \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^T\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i y_i\right)$

2. Plug in the "true" expression for y_i : $\hat{\boldsymbol{\beta}}_{OLS} = \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^T\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\left[\underbrace{\mathbf{x}_i^T\boldsymbol{\beta}+\varepsilon_i}_{y_i}\right]\right) = \boldsymbol{\beta} + \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^T\right)^{-1} \Big($

3. Use law of large numbers to figure out what happens when we have a very large number of data points (as N$\rightarrow\infty$)

$$\text{As N} \rightarrow \infty,\ \beta \text{ does not change}$$

$$\text{As N} \rightarrow \infty,\ \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^T \rightarrow_P E\left(\mathbf{x}_i\mathbf{x}_i^T\right) \quad \text{so} \quad \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^T\right)^{-1} \rightarrow_P E\left(\mathbf{x}_i\mathbf{x}_i^T\right)^{-1}$$

$$\text{As N} \rightarrow \infty,\ \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\varepsilon_i\right) \rightarrow_P E\left(\mathbf{x}_i\varepsilon_i\right) = 0$$

4. Putting everything together, as N$\rightarrow_P$ ∞, we have...

$$\hat{\beta} = \boldsymbol{\beta} + \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i \varepsilon_i \right) \rightarrow_P \beta$$

Note: Not all OLS coefficients are consistent for the true parameters. The above steps are typically the steps that we go through to show consistency.