PS 2010: 11. Ordinary Least Squares(OLS)

Qing Chang

Department of Political Science
University of Pittsburgh

Fall 2023

- No homework but today's class is very important

Agenda

- OLS estimator
- Hypothesis Testing

OLS

- $y = X\hat{\beta} + \epsilon$
- Our goal is to minimize sum of distance
- One way is to minimize sum of squared distance
- That is $\epsilon^{'}\epsilon$
- We have $\epsilon = y - X\hat{\beta}$

$$\begin{bmatrix} \epsilon_1 & \epsilon_2 & \dots & \epsilon_n \end{bmatrix}_{1\times n} \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}_{n\times 1} = [\epsilon_1 \times \epsilon_1 + \epsilon_2 \times \epsilon_2 + \dots + \epsilon_n \times \epsilon_n]_{1\times 1}$$

OLS

$$\begin{aligned}
\epsilon'\epsilon &= (y - X\hat{\beta})'(y - X\hat{\beta}) \\
&= y'y - \hat{\beta}'X'y - y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta} \\
&= y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}
\end{aligned}$$

Where $\hat{\beta}'X'y$ and $y'X\hat{\beta}$ are scalars

OLS

$$\epsilon'\epsilon = y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}$$

So, the goal is to find $\hat{\beta}$ that minimize the above equation

$$\frac{\partial \epsilon'\epsilon}{\partial \hat{\beta}} = -2X'y + 2X'X\hat{\beta} = 0$$

Then

$$X'X\hat{\beta} = X'y$$

Finally

$$\hat{\beta} = (X'X)^{-1}X'y$$

Assumptions

- We make no assumptions to derive OLS estimator
- However, we need make some assumptions to guarantee OLS is BLUE
- And allow us to do inference

Gauss-Markov Assumptions

- **Assumption 1:** Linear relationship between dependent and independent variables

$$y = X\beta + e$$

- **Assumption 2:** No perfect multicollinearity in independent variables

$$rank(X) = k$$

- **Assumption 3:** Strict exogeneity or zero conditional mean assumption

$$E(e \mid X) = 0$$

- **Assumption 4:** Spherical error variance
  - **Homoskedasticity:** $E(e'e \mid X) = \sigma^2$
  - **No autocorrelation:** $E(e_i e_j \mid X) = 0$ for i, j = 1,2,3,...; $i \neq j$

Unbiaseness

- The Gauss-Markov Assumptions guarantee OLS estimator is BLUE
- Unbiasedness:

$$
\begin{aligned}
\hat{\beta} &= (X'X)^{-1}X'y \\
\hat{\beta} &= (X'X)^{-1}X'(X\beta + e) \\
\hat{\beta} &= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'e \\
\hat{\beta} &= \beta + (X'X)^{-1}X'e \\
E[\hat{\beta}] &= E[\beta + (X'X)^{-1}X'e] \\
&= \beta + (X'X)^{-1}X'E[e]
\end{aligned}
$$

- $E[e] = 0$ by assumption
- Therefore, we have $E[\hat{\beta}] = \beta$

Variance-covariance Structure

What is variance-covariance structure for $\hat{\beta}$ looks like, given Gauss-Markov Assumptions

$$\begin{aligned}
\hat{\beta} &= \beta + (X'X)^{-1}X'e \\
V[\hat{\beta}] &= V[(X'X)^{-1}X'e] \\
&= (X'X)^{-1}X'V[e]X(X'X)^{-1}
\end{aligned}$$

We have $V[e] = E(e'e) - [E(e)]^2 = E(e'e) = \sigma^2$

$$\begin{aligned}
V[\hat{\beta}] &= (X'X)^{-1}X'E(e'e)X(X'X)^{-1} \\
&= (X'X)^{-1}X'(\sigma^2 I)X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}
\end{aligned}$$

Variance-covariance

What does the variance-covariance matrix of the OLS estimator look like?

Remember that error variance looks like this

$$\sigma^2 I = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix}$$

Therefore, the variance-covariance matrix for the OLS estimator:

$$V[\hat{\beta}] = \begin{bmatrix} \text{var}\left(\hat{\beta}_1\right) & \text{cov}\left(\hat{\beta}_1, \hat{\beta}_2\right) & \dots & \text{cov}\left(\hat{\beta}_1, \hat{\beta}_k\right) \\ \text{cov}\left(\hat{\beta}_2, \hat{\beta}_1\right) & \text{var}\left(\hat{\beta}_2\right) & \dots & \text{cov}\left(\hat{\beta}_2, \hat{\beta}_k\right) \\ \vdots & \vdots & \vdots & \vdots \\ \text{cov}\left(\hat{\beta}_k, \hat{\beta}_1\right) & \text{cov}\left(\hat{\beta}_k, \hat{\beta}_2\right) & \dots & \text{var}\left(\hat{\beta}_k\right) \end{bmatrix}$$

Non-Spherical Errors (Not Important!)

- Heteroskedasticity but no correlation between observations

$$\sigma^2 I = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix}$$

- Heteroskedasticity and correlation between observations

$$\Omega = \begin{bmatrix} \Omega_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \Omega_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \Omega_G \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdot & \cdot & 0 & 0 & 0 \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} & \cdot & \cdot & 0 & 0 & 0 \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 & \cdot & \cdot & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ 0 & 0 & 0 & \cdot & \cdot & \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ 0 & 0 & 0 & \cdot & \cdot & \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix}$$

Hypothesis Testing

- We now have our estimates and their corresponding variance
- we can conduct single or joint hypothesis testing
- To calculate statistic, we need one more assumption for error term

$$e \sim N\left[0, \sigma^2 I\right]$$

- For single parameter testing

$$H_0 : \beta_k = c$$
$$H_1 : \beta_k \neq c$$

$$T = \frac{\hat{\beta}_k - c}{\sqrt{v(\hat{\beta}_{kk})}}$$

- Usually, we set $c = 0$

Hypothesis Testing

- If sample is small then it is t, if large, then z
- The important point is: we need error variance σ^2 to calculate variance matrix for $\hat{\beta}$. However, we do not know it

$$V(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

- So, we use estimated error variance to replace

$$\hat{\sigma}^2 = \frac{(Y - \mathbf{X}\hat{\beta})'(Y - \mathbf{X}\hat{\beta})}{n-k} = \frac{\epsilon'\epsilon}{n-k}$$

- We also need calculate $t_{\alpha/2,n-k}$ for critical value
- Notice, we have k equal to the total parameters we estimate: independent variable + 1

Confidence Interval

- Because we have $SE(\beta_k)$, we can also construct confidence interval

$$\mathrm{CI}_{1-\alpha}(\hat{\beta}_k) = [\hat{\beta}_k - z_\alpha \, \mathrm{SE}(\hat{\beta}_k), \hat{\beta}_k + z_\alpha \, \mathrm{SE}(\hat{\beta}_k)]$$

- Change z to t in small sample

R-square

- How do you know if your model is a good model?
- How many proportion of variance in dependent variable is explained by your model?
- Ok, this means we need to know two things:
  - Total variance in dependent variable
  - Variance explained by our model
- Total variance $= \sum_i^n (y_i - \bar{y})^2$
- % of Explained variance = 1 - % of unexplained variance
- We know unexplained variance$= \sum_i^n \epsilon_i^2$

$$R^2 = 1 - \frac{\sum_i^n \epsilon_i^2}{\sum_i^n (y_i - \bar{y})^2}; R^2_{Adjust} = 1 - (1 - R^2) * \frac{n-1}{n-k}$$

Joint Hypotheses Testing

$$H_0 : \beta_1 = \beta_2 = ...\beta_k = 0$$
$$H_1 : \textit{One of them or All of them are non zero}$$

$$F = \frac{\text{Explained variance}/(k-1)}{\text{Unexplained variance}/(n-k)}$$

The critical value for significant level α is F With k-1 and n-k degree of freedom

Empirical Application

- RQ: Effects of economic openness and democracy on national income inequality
- Dependent variable: gini inequality
- Independent variables:
  - Democracy level: Freedom house
  - Openness 1: trade, (import+export)/GDP
  - Openness 2: FDI, total inflows of FDI/GDP
- Data: 1996, 183 observations

Model

$$Inequality_i = \beta_0 + \beta_1 Democracy + \beta_2 Trade + \beta_3 FDI + \epsilon_i$$

Where i represent each countries

OLS

- OLS estimator:
$$\hat{\beta} = (X'X)^{-1}X'y$$
- Variance-covariance matrix under homoskedasticity
$$\sigma^2(X'X)^{-1}$$
- Use sample variance to replace σ^2
$$\hat{\sigma}^2 = \frac{(Y - \mathbf{X}\hat{\beta})'(Y - \mathbf{X}\hat{\beta})}{n-k} = \frac{\epsilon'\epsilon}{n-k}$$

OLS

- Confidence Interval

$$\mathrm{CI}_{1-\alpha}(\hat{\beta}_k) = [\hat{\beta}_k - t_{\alpha,n-k}\,\mathrm{SE}(\hat{\beta}_k), \hat{\beta}_k + t_{\alpha,n-k}\,\mathrm{SE}(\hat{\beta}_k)]$$

- R square and adjust R square

$$R^2 = 1 - \frac{\sum_i^n \epsilon_i^2}{\sum_i^n (y_i - \bar{y})^2}; R^2_{Adjust} = 1 - (1 - R^2) * \frac{n-1}{n-k}$$

- F test

$$F = \frac{\text{Explained variance}/(k-1)}{\text{Unexplained variance}/(n-k)}$$

Practice Questions

Suppose we want to know how education quality is influenced by a country's democracy level and income level. After you collect data (let's say 100 observations), you want to run OLS to see if democracy and income have effects on education quality.

- Define your dependent and independent variables
- Then, let's define your dependent variable matrix as y
- Your independent variable matrix as X
- Let's define your model as $y = X\hat{\beta} + \epsilon$
- What are the dimensions of your X and y matrix?

Practice Questions

- Derive OLS estimator
- what assumptions do we need to ensure OLS is unbias?
- If we assume error term is homoskedasticity and equals to σ^2, what is the variance-covariance matrix for OLS estimator
- Does violation of homoskedasticity assumption effect the unbiaseness of OLS estimator?
- Vice versa?

Practice Questions

- Imagine that after you calculate variance covariance matrix for $\hat{\beta}$, you have a matrix as the following

$$\begin{bmatrix} 0.25 & 0.56 & 0.18 \\ 0.56 & 0.36 & 0 \\ 0.18 & 0 & 0.64 \end{bmatrix}$$

- What are standard errors for each of your $\hat{\beta}$?
- Let's imagine from your OLS estimation, you get coefficient for democracy is 3 and income level is 5
- Can you conduct a hypothesis testing for democracy? Calculate the t value using the coefficient and variance covariance matrix in the above?

Practice Questions

- If pre-select significant level as 5% and given ($t_{(0.975,97)} = 1.66$), you reject or fail to reject your null?
- Again, using the above information, can you construct a 95% confidence interval for the coefficient of income level?
- Can you explain what are R square and F test? Why do we want to use them?

Extra Material

- Intro to Causal Inference
- Empirical Methods