These notes come from Chapters 6 and 7 in the textbook and cover the large-sample properties of least squares.

Linear Regression Notes 3: Asymptotic theory for least squares

Asymptotic Theory for Least Squares

The asymptotic theory for least squares applies both to linear projection model and to the linear CEF model. Therefore, in this section, we only use the weaker assumptions of the linear projection model. That is, we use the following assumptions throughout this section

Assumption 7.1

1. The variables $\{(Y_i, X_i)\}_{i=1}^n$ are iid
2. $\mathbb{E}[Y^2] < \infty$
3. $\mathbb{E}||X||^2 < \infty$
4. $\mathbb{E}[XX']$ is positive definite

Consistency of Least Squares Estimator

H: 7.2

Step 1: Weak Law of Large Numbers. Recall that

$$\hat{\beta} = \left(\frac{1}{n} \sum_{i=1}^n X_i X_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^n X_i Y_i \tag{1}$$

Next, notice that

$$\frac{1}{n} \sum_{i=1}^n X_i X_i' \xrightarrow{p} \mathbb{E}[XX']$$
$$\frac{1}{n} \sum_{i=1}^n X_i Y_i \xrightarrow{p} \mathbb{E}[XY]$$

which holds by the weak law of large numbers (which requires the iid assumption and that $\mathbb{E}[XX'] < \infty$ and $\mathbb{E}[XY] < \infty$, both of which hold by Assumption 7.1)

Step 2: Continuous Mapping Theorem. Next, notice that, we can write

$$\hat{\beta} = g(\hat{\mathbb{E}}[XX'], \hat{\mathbb{E}}[XY])$$

where $g(\mathbf{A}, b) = \mathbf{A}^{-1}b$. This is a continuous function of $\mathbf{A}$ and b at all values of the arguments such that $\mathbf{A}^{-1}$ exists. Assumption 7.1 includes that $\mathbb{E}[XX']$ is positive definite which implies that $\mathbb{E}[XX']^{-1}$ exists. Thus, $g(\mathbf{A}, b)$ is continuous at $\mathbf{A} = \mathbb{E}[XX']$ and we can apply the "convergence in probability" version of the CMT; that is,

$$\begin{aligned}\hat{\beta} &\xrightarrow{p} g(\mathbb{E}[XX'], \mathbb{E}XY) \\ &= \mathbb{E}[XX']^{-1}\mathbb{E}[XY] = \beta\end{aligned}$$

Asymptotic Normality

H: 7.3

For this section, we strengthen Assumption 7.1.

Assumption 7.2 In addition to Assumption 7.1

1. $\mathbb{E}[Y^4] < \infty$
2. $\mathbb{E}||X||^4 < \infty$

Next, we will establish the limiting distribution of $\hat{\beta}$. Plugging $Y_i = X_i'\beta + e_i$ into Equation 1 implies that

$$\begin{aligned}\hat{\beta} &= \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} \left(X_i(X_i'\beta + e_i)\right) \\ &= \beta + \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} X_i e_i\end{aligned}$$

Multiplying by $\sqrt{n}$ and re-arranging implies that

$$\sqrt{n}\left(\hat{\beta} - \beta\right) = \left(\frac{1}{n}\sum_{i=1}^{n} X_i X_i'\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i e_i \tag{2}$$

Step 1: Central Limit Theorem. First, notice that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i e_i \xrightarrow{d} N(0, \mathbf{\Omega})$$

where $\mathbf{\Omega} = \mathbb{E}[Xe(Xe)'] = \mathbb{E}[XX'e^2]$.

Let's explain carefully why the central limit theorem applies here. First, we have that (Y_i, X_i) are iid, which implies that any function of (Y_i, X_i) is also iid (and this includes $e_i = Y_i - X_i'\beta$ and

$X_i e_i$). Also, notice that $\mathbb{E}[Xe] = 0$ which implies that $\text{var}(Xe) = \Omega$. Finally, to invoke the central limit theorem, we need to show that our assumptions imply that all of the elements of Ω are finite (if you are interested in this, see the technical details section below).

Technical Details: As a step in this direction, let's first show that Assumption 7.2 implies that $\mathbb{E}[e^4] < \infty$.

Minkowski's Inequality: $(\mathbb{E}||X+Y||^p)^{1/p} \leq \left(\mathbb{E}||X||^p\right)^{1/p} + \left(\mathbb{E}||Y||^p\right)^{1/p}$

Schwarz Inequality: $|a'b| \leq ||a||\ ||b||$

Therefore,

$$\begin{aligned}
\mathbb{E}[e^4]^{1/4} &= \mathbb{E}[(Y - X'\beta)^4]^{1/4} \\
&\leq \mathbb{E}[Y^4]^{1/4} + \mathbb{E}[(X'\beta)^4]^{1/4} \\
&\leq \mathbb{E}[Y^4]^{1/4} + (\mathbb{E}||X||^4)^{1/4}||\beta|| \\
&< \infty
\end{aligned}$$

where the second line uses Minkowski's inequality, the third inequality holds by the Schwarz inequality; to be clear on this part, notice that $\mathbb{E}[(X'\beta)^4]^{1/4} = \mathbb{E}[|X'\beta|^4]^{1/4} \leq \mathbb{E}[(||X||\ ||\beta||)^4]^{1/4} = \mathbb{E}[||X||^4||\beta||^4]^{1/4} = \mathbb{E}[||X||^4]^{1/4}||\beta|| < \infty$. That $\mathbb{E}[e^4]^{1/4} < \infty$ implies that $\mathbb{E}[e^4] < \infty$.

Expectation Inequality: For a random vector $Y \in \mathbb{R}^m$ with $\mathbb{E}||Y|| < \infty$, $||\mathbb{E}[Y]|| \leq \mathbb{E}||Y||$.

Cauchy Schwarz Inequality: $\mathbb{E}||X'Y|| \leq \left(\mathbb{E}||X||^2\right)^{1/2}\left(\mathbb{E}||Y||^2\right)^{1/2}$

Next, the (j,l) element of $\boldsymbol{\Omega}$ is given by $\mathbb{E}[X_j X_l e^2]$ (we want to show that this is finite). Therefore, consider

$$\begin{aligned}
|\mathbb{E}[X_j X_l e^2]| &\leq \mathbb{E}|X_j X_l e^2| \\
&= \mathbb{E}[|X_j|\ |X_l| e^2] \\
&\leq \mathbb{E}[X_j^2 X_l^2]^{1/2}\mathbb{E}[e^4]^{1/2} \\
&\leq \left(\mathbb{E}[X_j^4]^{1/2}\mathbb{E}[X_l^4]^{1/2}\right)^{1/2}\mathbb{E}[e^4]^{1/2} \\
&= \mathbb{E}[X_j^4]^{1/4}\mathbb{E}[X_l^4]^{1/4}\mathbb{E}[e^4]^{1/2} \\
&< \infty
\end{aligned}$$

where the first equality holds by the expectation inequality, the second equality holds because of the absolute value, the third equality holds by the Cauchy-Schwarz inequality, the fourth equality holds by applying the Cauchy-Schwarz inequality again, the fifth equality holds immediately, and the last equality holds by Assumption 7.2 and because $\mathbb{E}[e^4] < \infty$ (which we showed right before).

Combining this with Equation 2, we have that

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathbb{E}[XX']^{-1}\, N(0, \mathbf{\Omega}) = N(0, \mathbf{V}_\beta)$$

where $\mathbf{V}_\beta = \mathbb{E}[XX']^{-1}\mathbf{\Omega}\mathbb{E}[XX']^{-1}$ and which holds by the continuous mapping theorem.

$\mathbf{V}_\beta$ is called the **asymptotic variance matrix** of $\hat{\beta}$. $\mathbb{E}[XX']^{-1}\mathbf{\Omega}\mathbb{E}[XX']^{-1}$ is called a "sandwich form". It is called this because $\mathbf{\Omega}$ is sandwiched by $\mathbb{E}[XX']^{-1}$ (sometimes $\mathbf{\Omega}$ is called the "meat" and $\mathbb{E}[XX']^{-1}$ is called the "bread"). Many asymptotic variance matrices have a similar form.

The previous result is the basis for hypothesis testing/inference, constructing confidence intervals, etc. To operationalize it, though, we need to construct an estimator of $\mathbf{V}_\beta$. Before doing that, let's introduce one relatively common simplification.

Homoskedasticity Assumption: $\mathbb{E}[e^2|X] = \sigma^2$.

Homoskedasticity says that the second moment of the error term does not vary across different values of X. This is often constrasted with **heteroskedasticity** which amounts to just not making the homoskedasticity assumption. Most applications in economics do not invoke the homoskedasticity assumption mainly because, often, we do not "need" it. That said, as we will see below, it is useful for simplifying some expressions and serves as a useful benchmark in many cases.

Notice that, under homoskedasticity, we can simplify the expression for $\mathbf{\Omega}$ (I use the notation $\mathbf{\Omega}_0$ to indicate that this is the expression for $\mathbf{\Omega}$ under homoskedasticity):

$$\mathbf{\Omega} = \mathbb{E}\Big[XX' \underbrace{\mathbb{E}[e^2|X]}_{\sigma^2}\Big] = \sigma^2\mathbb{E}[XX']$$

where the first equality holds by the law of iterated expectations, and the second equality holds by homoskedasticity. Plugging this back in to the expression for $\mathbf{V}_\beta$, it will also simplify (again, I switch the notation to indicate the asymptotic variance of $\sqrt{n}(\hat{\beta} - \beta)$ under homoskedasticity):

$$\mathbf{V}_0 = \sigma^2\mathbb{E}[XX']^{-1}$$

which holds by plugging in $\mathbf{\Omega}_0$ into the expression for $\mathbf{V}_\beta$ and cancelling.

Consistency of Error Variance Estimators

H: 7.5

Next, we consider estimating $\sigma^2 = \mathbb{E}[e^2]$. Using the analogy principle would suggest estimating σ^2 by

$$\frac{1}{n}\sum_{i=1}^{n} e_i^2$$

but this estimator is infeasible since we do not observe e_i. Instead, let's consider the estimator

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{e}_i^2$$

where $\hat{e}_i$ is the **residual** that is defined as

$$\hat{e}_i = Y_i - X_i'\hat{\beta}$$

which is the difference between the actual outcome and $X_i\hat{\beta}$ (the fitted value from the regression). Notice that, $\hat{e}_i$ is something that we can actually recover because it depends on the estimated $\hat{\beta}$ rather than, say, the population parameter β. Notice that, by plugging in $Y_i = X_i'\beta$ into the expression for $\hat{e}_i$, we have that

$$\begin{aligned}\hat{e}_i &= Y_i - X_i'\hat{\beta} \\ &= X_i'\beta + e_i - X_i'\hat{\beta} \\ &= e_i - X_i'(\hat{\beta} - \beta)\end{aligned}$$

which implies that

$$\hat{e}_i^2 = e_i^2 - 2e_iX_i'(\hat{\beta} - \beta) + (\hat{\beta} - \beta)'X_iX_i'(\hat{\beta} - \beta)$$

so that

$$\frac{1}{n}\sum_{i=1}^{n} \hat{e}_i^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2 - 2\left(\frac{1}{n}\sum_{i=1}^{n} e_iX_i'\right)(\hat{\beta} - \beta) + (\hat{\beta} - \beta)\left(\frac{1}{n}\sum_{i=1}^{n} X_iX_i'\right)(\hat{\beta} - \beta)$$

Then, since,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} e_i^2 &\xrightarrow{p} \mathbb{E}[e^2] = \sigma^2 \\ \hat{\beta} - \beta &\xrightarrow{p} 0 \\ \frac{1}{n}\sum_{i=1}^{n} X_iX_i' &\xrightarrow{p} \mathbb{E}[XX']\end{aligned}$$

it follows by the continuous mapping theorem that

$$\hat{\sigma}^2 \xrightarrow{p} \sigma^2$$

Heteroskedastic Covariance Matrix Estimation

H: 7.7

Next, we consider estimating $\mathbf{V}_\beta$. The natural estimator is

$$\hat{\mathbf{V}}_\beta = \left(\frac{1}{n}\sum_{i=1}^n X_i X_i'\right)^{-1} \hat{\mathbf{\Omega}} \left(\frac{1}{n}\sum_{i=1}^n X_i X_i'\right)^{-1}$$

where $\hat{\mathbf{\Omega}}$ is an estimate of $\mathbf{\Omega}$ given by

$$\hat{\mathbf{\Omega}} = \frac{1}{n}\sum_{i=1}^n X_i X_i' \hat{e}_i^2$$

We aim to show that $\hat{\mathbf{\Omega}}$ is consistent for $\mathbf{\Omega}$. To this end, notice that

$$\hat{\mathbf{\Omega}} = \frac{1}{n}\sum_{i=1}^n X_i X_i' e_i^2 + \frac{1}{n}\sum_{i=1}^n X_i X_i'(\hat{e}_i^2 - e_i^2)$$

which holds by adding and subtracting terms. Then, notice that

$$\frac{1}{n}\sum_{i=1}^n X_i X_i' e_i^2 \xrightarrow{p} \mathbb{E}[XX'e^2] = \mathbf{\Omega}$$

It remains to show be shown that

$$\frac{1}{n}\sum_{i=1}^n X_i X_i'(\hat{e}_i^2 - e_i^2) \xrightarrow{p} 0$$

Given our earlier result on $\hat{\sigma}^2$ being consistent for σ^2, it is perhaps not surprising that this term converges to 0 though the arguments are more challenging (if you are interested, see the technical details below).

Technical Details: To start with, let me briefly introduce some useful concepts related to matrix norms and useful inequalities for matrix norms. Below, $\mathbf{A}$ and $\mathbf{B}$ are notation for matrices.

Frobenius/Matrix Norm $||\mathbf{A}|| = ||\text{vec}(\mathbf{A})||$
Schwarz Inequality $||\mathbf{AB}|| \leq ||\mathbf{A}|| \, ||\mathbf{B}||$
Triangle Inequality: $||\mathbf{A}+\mathbf{B}|| \leq ||\mathbf{A}|| + ||\mathbf{B}||$
Holder's Inequality: For any $p > 1$ and $q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$, $\mathbb{E}||X'Y|| \leq \left(\mathbb{E}||X||^p\right)^{1/p}\left(\mathbb{E}||Y||^q\right)^{1/q}$

The Frobenius norm is a matrix norm (there are others) that "converts" the matrix into a vector and then applies the Euclidean norm to that vector. The next two inequalities say that versions of the Schwarz and triangle inequalities apply to matrices. Next, notice that

$$\begin{aligned}\left|\left|\frac{1}{n}\sum_{i=1}^{n} X_i X_i'(\hat{e}_i^2 - e_i^2)\right|\right| &\leq \frac{1}{n}\sum_{i=1}^{n} ||X_i X_i'(\hat{e}_i^2 - e_i^2)|| \\ &\leq \frac{1}{n}\sum_{i=1}^{n} ||X_i||^2 |\hat{e}_i^2 - e_i^2| \end{aligned} \tag{3}$$

where the first inequality holds by the triangle inequality and the second inequality holds by applying the Schwarz inequality twice. Now consider

$$\begin{aligned}|\hat{e}_i^2 - e_i^2| &= |-2e_i X_i'(\hat{\beta} - \beta) + (\hat{\beta} - \beta)' X_i X_i'(\hat{\beta} - \beta)| \\ &\leq 2|e_i X_i'(\hat{\beta} - \beta)| + (\hat{\beta} - \beta)' X_i X_i'(\hat{\beta} - \beta) \\ &= 2|e_i| \; |X_i'(\hat{\beta} - \beta)| + |(\hat{\beta} - \beta)' X_i|^2 \\ &\leq 2|e_i| \; ||X_i|| \; ||\hat{\beta} - \beta|| + ||X_i||^2 \; ||\hat{\beta} - \beta||^2\end{aligned}$$

where the first equality holds by plugging in from above the difference between $\hat{e}_i^2$ and e_i^2, the second inequality holds by the triangle inequality (the second term is positive because it is quadratic), the third equality holds by properties of absolute value, the fourth inequality holds by the Schwarz inequality. Using this expression back in Equation 3 implies that

$$\left|\left|\frac{1}{n}\sum_{i=1}^{n} X_i X_i'(\hat{e}_i^2 - e_i^2)\right|\right| \leq 2\left(\frac{1}{n}\sum_{i=1}^{n} ||X_i||^3 \; |e_i|\right) \; ||\hat{\beta} - \beta|| + \frac{1}{n}\sum_{i=1}^{n} ||X_i||^4 \; ||\hat{\beta} - \beta||^2$$

The second term converges to 0 because $n^{-1}\sum_{i=1}^{n} ||X_i||^4 \xrightarrow{p} \mathbb{E}[X^4]$ and because $||\hat{\beta} - \beta|| \xrightarrow{p} 0$. For the first term $||\hat{\beta} - \beta|| \xrightarrow{p} 0$, and then consider

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} ||X_i||^3 \; |e_i| \xrightarrow{p} \mathbb{E}[||X||^3 \; |e|] & \\ &\leq \mathbb{E}\left[(||X||^3)^{4/3}\right]^{3/4} \mathbb{E}[e^4]^{1/4} \\ &= \mathbb{E}[||X||^4]^{3/4}\mathbb{E}[e^4]^{1/4} \\ &< \infty\end{aligned}$$

where the first equality holds by the weak law of large numbers, the second equality holds using Holder's inequality (using $||X||^3$ and $|e|$ and setting $p = 4/3$ and $q = 4$), the third equality by canceling the inside exponents, and the last inequality by Assumption 7.2 and that we showed that $\mathbb{E}[e^4] < \infty$.

Thus, we have shown that $\hat{\mathbf{\Omega}} \xrightarrow{p} \mathbf{\Omega}$. It immediately follows from the weak law of large numbers and the continuous mapping theorem that $\hat{\mathbf{V}}_\beta \xrightarrow{p} \mathbb{E}[XX']^{-1}\mathbf{\Omega}\mathbb{E}[XX']^{-1} = \mathbf{V}_\beta$. This implies that $\hat{\mathbf{V}}_\beta$ is consistent for $\mathbf{V}_\beta$.