(Optional) OLS Properties

Name: Jikhan Jeong

March 3, 2022

1. Positive Definite, Trace, Idempotent, Inverse, Matrix Derivative

1.1 Positive Definite

- X'X = Square Matrix = Positive Definite = Convex = Can be minimized
- In OLS assumptions, we requires $(X'X)^{-1} \rightarrow 0$ to derive consistency of Cov(β)

1.2 Trace

- (Example) Trace $\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} = \sum_{i=1}^{3} x_{ii} = x_{11} + x_{22} + x_{33}$
- Trace(scalar ×X) = scalar × Trace(X)
- Trace(X+Z) = Trace(X) + Trace(Z)
- Trace(X') = Trace(X)
- Trace(AB) = Trace(BA)
- We need it when Cov(Y) = $\sigma^2 I$ = Cov(ε) = E($\varepsilon\varepsilon'$) , where E[Trace(e'Me)] = Trace[M]E(e'e)]= (n-k)E(e'e), where n = sample number, k = variable numbers = column dim(X)

1.3 Idempotent Matrix

- We need it when Cov(Y) = $\sigma^2 I$ = Cov(ε) = E($\varepsilon\varepsilon'$) , where E[Trace(e'M'Me)] = E[Trace(e'Me)]
- P $\subset$Idempotent Matrix, then P'P = P
- P $\subset$Idempotent Matrix, then I – P = M, M $\subset$Idempotent Matrix

1.4 Inverse Matrix

- A = $(A^{-1})^{-1}$
- $(A')^{-1} = (A^{-1})'$
- $(AB)^{-1} = B^{-1}A^{-1}$
- $|A^{-1}| = \frac{1}{|A|}$

1.5 Matrix Derivative

- From Prof. Ron's Book p.440, Useful for OLS Loss minimization
- **z** : (k x 1) vector, **A** : (k x j) matrix, **w** : (j x 1) vector
- $\frac{\vartheta z' Aw}{\vartheta z} = Aw$
- $\frac{\vartheta z' Aw}{\vartheta w} = A'w$
- $\frac{\vartheta z' Az}{\vartheta z} = 2Az$, where k=j and A is a symmetric (A'=A)
- $\frac{\vartheta z' w}{\vartheta z} = w$, (k=j)
- $\frac{\vartheta z' w}{\vartheta w} = z$, (k=j)

2. Summarize the classical assumptions of GLM

- E(Y) = xβ and E(ε)=0
- Cov(Y) = $\sigma^2 I$ = Cov(ε) = E($\varepsilon\varepsilon'$)
- x is a fixed matrix of values with rank(x) = k, where k = unknown parameters

3. OLS objective function

- S = e'e, OLS minimize S
- S = e'e = (y–xβ)'(y–xβ)
- FOC β: –2X'y + 2X'X$\hat{\beta}$=0 $\mapsto$ X'y = X'X$\hat{\beta}$ $\mapsto$ (X'X)$^{-1}$X'y =$\hat{\beta}$
- SOC β: 2X'X is a positive definite

4. Gauss-Markov

- $\hat{\beta}$ in OLS is unbiased and efficient within the class of linear, unbiased estimator (BLUE)
- $\hat{\beta}$ = (X'X)$^{-1}$X'y =(X'X)$^{-1}$X'(Xβ+e) = (X'X)$^{-1}$X'Xβ+(X'X)$^{-1}$X'e
  = β+(X'X)$^{-1}$X'e
- E($\hat{\beta}$) = E(β+(X'X)$^{-1}$X'e) = β + 0 = β , unbias
- Cov($\hat{\beta}$) = E[($\hat{\beta}$–E($\hat{\beta}$))($\hat{\beta}$–E($\hat{\beta}$))'] = E[($\hat{\beta}$–β)($\hat{\beta}$–β)']= E[(X'X)$^{-1}$X'ee'X(X'X)$^{-1}$]
  = (X'X)$^{-1}$X'E(ee')X(X'X)$^{-1}$, where E(ee') = $\sigma^2 I$
  = σ^2(X'X)$^{-1}$X'X(X'X)$^{-1}$= σ^2(X'X)$^{-1}$
- Cov(Y) = E($\hat{e}$'$\hat{e}$) = E[(Y- $\hat{\beta}$X)(Y- $\hat{\beta}$X)']= E[(Y- $\hat{\beta}$X)(Y- $\hat{\beta}$X)'] = E[e'M'Me] = E[e'Me]
- = E[Tr(e'Me)]= E[Tr(M)e'e]= Tr(M)E(e'e) = Tr(I–P)σ^2 = (n–k)σ^2
- $\frac{\hat{e}'\hat{e}}{n-k} = s^2$, sample variance to make it unbiased estimator of $\sigma^2 \mapsto E(\frac{\hat{e}'\hat{e}}{n-k})$ = E(s^2) = σ^2

4.1 (Proof E[e'Me] = σ^2Tr(M), Econometrics(Hayashi, p31)

- e'Me = $\sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}\varepsilon_i\varepsilon_j$, derivation of quadratic form
- E(e'Me|X) =$\sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}E(\varepsilon_i\varepsilon_j|X)$, where $E(m\varepsilon_i\varepsilon_j|X)$=$m_{ij}E(\varepsilon_i\varepsilon_j|X)$ as m_{ij} is function of X
  =$\sigma^2\sum_{i=1}^{n} m_{ii}$, as the classical assumptions of OLS = $\sigma^2 I$ = Cov(ε) = E($\varepsilon\varepsilon'$)
  =$\sigma^2 \cdot Trace(M)$

5. Consistent estimator of θ

- If $\hat{\theta}$ converges in probability to θ, then we say it is a consistent estimator
- plim $\hat{\beta} \to \beta$, consistent estimator as follows:
  plim$\hat{\beta}$ = β+ plim(X'X)$^{-1}$X'e = β+ plim(X'X$\frac{n}{n}$)$^{-1}$X'e =β+ plim(X'X$\frac{1}{n}$)$^{-1}\frac{X'e}{n}$
  =β + plim(X'X$\frac{1}{n}$)$^{-1}\cdot$0 =β
- plim $s^2 = \frac{\hat{e}'\hat{e}}{n-k} = \to \sigma^2$, consistent estimator as follows:
  $s^2 = \frac{\hat{e}'\hat{e}}{n-k} = \frac{e'Me}{n-k} = \frac{e'(I-P)e}{n-k} = \frac{e'e-e'X(X'X)^{-1}X'e}{n-k} = \frac{e'e}{n-k} - \frac{e'X(X'X)^{-1}X'e}{n-k}$
  plim $s^2 \to$ plim($\frac{e'e}{n-k}$) – plim ($\frac{e'X(X'X)^{-1}X'e}{n-k}$) $\to \sigma^2$– plim ($\frac{e'X(X'X)^{-1}X'e}{n-k}$)
  $\to \sigma^2$, where plim ($\frac{e'X(X'X)^{-1}X'e}{n-k}$) = 0 as $\frac{(X'X)^{-1}}{n} \to 0$

6. R^2

- Simple regression case, 1 exogeneous variable“
- R^2 measures the precentage of variations in $\hat{Y}$ can be explained by the regression model
- SST = total sum of squares = (Y–$\bar{Y}$)(Y–$\bar{Y}$)'
- SSE = error sum of squares = (Y–$\hat{Y}$)(Y–$\hat{Y}$)'
- SSR = sum of square residuals = SST – SSE = ($\hat{Y}$–$\bar{Y}$)($\hat{Y}$–$\bar{Y}$)'
- R^2= $\frac{SSR}{SST}$= $\frac{SST-SSE}{SST}$ = 1 – $\frac{SSE}{SST}$

Reference

- Mittelhammer, R.C. and Mittelhammer, R.C., 1996. Mathematical statistics for economics and business (Vol. 78). New York: Springer.
- Fomby, T.B., Hill, R.C. and Johnson, S.R., 2012. Advanced econometric methods. Springer Science & Business Media.
- Hayashi, F., 2000. Econometrics. 2000. Princeton University Press.