Econometrics I
TA Session 7

Jukina HATAKEYAMA *

May 28, 2024

Contents

*E-mail: u868710a@ecs.osaka-u.ac.jp

1 Review of F Statistic Test

1.1 Fisher Distribution under Normal Disturbance Term

The Fisher distribution is used for the F test. We can use this statistic to test linear restrictions of OLS models, structural change, seasonality and so on.

Definition 1.1. The Fisher distribution with q_1 and q_2 degrees of freedom, denoted as $F(q_1, q_2)$, is defined as the ratio between two χ^2 distributions normalized by their degrees of freedom:

$$Q_1 \sim \chi^2(q1), \quad Q_2 \sim \chi^2(q_2), \quad Q_1 \perp Q_2,$$
$$Z = \frac{Q_1/q_1}{Q_2/q_2} \sim F(q_1, q_2). \tag{1}$$

The density function $f(x)$ is given by

$$f(x) := \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1} \left(\frac{m}{n}x + 1\right)^{-\frac{m+n}{2}},$$

where $\Gamma(x)$ stands for a *Gamma function*, defined as follows:

$$\Gamma(x) = \int_0^\infty s^{x-1} e^{-s} ds.$$

Now we consider the classical regression model as follows:

$$y = Xb + u. \tag{2}$$

In this equation, we assume that $u|X \sim N_{\mathbb{R}^n}(0, \sigma^2 I_n)$. By using (1) , the normality of the disturbance terms, and Cochran's theorem, we can derive:

$$\frac{(\hat{b} - b)' X' X (\hat{b} - b)/K}{\hat{u}'\hat{u}/(n - K)} \sim F(K, n - K). \tag{3}$$

To prove this equation, keep in mind a following lemma.

Lemma 1.2. Suppose that a matrix $A \in \mathbb{R}^{k \times k}$ is idempotent. Then, we can establish following statements.

1. All eigen values of A are equal to 0 or 1.
2. $tr(A)$ is equal to numbers of the unit roots of A.
3. Assume the $k \times 1$ vector $X \sim N(0, I)$, then we have $X'AX \sim \chi^2(tr(A))$.

By applying the first and second statements, we can prove last one.[1].

[1]Please check Greene(2011):1040-1042, 1083-1084 if you need to review these two statements.

Proof. Suppose the case of the quadratic form like $q = X'AX$. By using the diagonal decomposition of A, we can rewrite q as follows:

$$q = X'C\Lambda C'X = Y'\Lambda Y = \sum_{i=1}^{n} \lambda_i y_i^2,$$

where $Y = X'C$ and this vector also follows the multivariate normal distribution because C is defined as $C'C = I$. Recall that a matrix A is idempotent, so its eigen values are equal to 1 or 0. Therefore, we can state $q = \sum_{i=1}^{J} y_i^2$, where J is the numbers of the unit roots, eigen values which are equal to 1. Thus, we can prove third statement by using the first and second ones. □

By using above lemma, we can show that (3) is surely established. At first, we calculate the numerator of (3).

$$\begin{aligned}(\hat{b} - b)'X'X(\hat{b} - b) &= [(X'X)^{-1}X'u]'X'X[(X'X)^{-1}X'u] \\ &= u'X(X'X)^{-1}X'u.\end{aligned}$$

Since $X(X'X)^{-1}X'$ is symmetric and idempotent and u follows the normal distribution, we have following relationship:

$$\frac{u'X(X'X)^{-1}X'u}{\sigma^2} \sim \chi^2(tr(X(X'X)^{-1}X')),$$

and the degrees of freedom is $tr(X(X'X)^{-1}X') = tr((X'X)^{-1}X'X)) = K$. Thererfore, we can conclude that:

$$\frac{u'X(X'X)^{-1}X'u}{\sigma^2} \sim \chi^2(K). \tag{4}$$

Next, we analyze the denominator of the (3).

$$\begin{aligned}\hat{u}'\hat{u} &= (M_x u)'(M_x u) \\ &= u'M_x u \\ &= u'(I_n - X(X'X)^{-1})X)u\end{aligned}$$

Therefore, we deduce the following relationship:

$$\frac{\hat{u}'}{\sigma}\frac{\hat{u}}{\sigma} = \frac{\hat{u}'\hat{u}}{\sigma^2} \sim \chi^2(tr(M_x)).$$

Here, $tr(M_x)$ equals to $n - K$ and we have following result:

$$\frac{\hat{u}'\hat{u}}{\sigma^2} = \frac{(n-K)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-K), \tag{5}$$

where $\hat{\sigma}^2 = \hat{u}'\hat{u}/(n-K)$. Finally, we check whether $\hat{b} \perp \hat{u}$ is satisfied or not. Recall that we can say $X \perp Y$ is the same meaning as X and Y are independent (i.e. $f_{XY}(x, y) = f_X(x)f_Y(y)$) in the cases of $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$. In this class, since $u|X \sim N(0, \sigma^2 I_n)$ and $\hat{b}|X \sim N(b, \sigma^2(X'X)^{-1})$ are given, it is enough to prove $\mathrm{Cov}(\hat{u}, \hat{b}) = 0$.

$$\begin{aligned}\mathrm{Cov}(\hat{u}, \hat{b}) &= \mathbb{E}[\hat{u}(\hat{b} - b)'] \\ &= \mathbb{E}[M_x u((X'X)^{-1}X'u)'] \\ &= \mathbb{E}[M_x uu'X(X'X)^{-1}] \\ &= M_x\mathbb{E}[uu'|X]X(X'X)^{-1} \\ &= \sigma^2 M_x X(X'X)^{-1} = 0.\end{aligned}$$

We have used the law of iterated expectation and $M_x X = 0$ to derive this equation. Next, we apply the F distribution to test the estimators of the classical linear regression model. We usually use the F test to check whether the predictive variables can affect the dependent variable. In this test, a null hypothesis is given as $H_0 : b_2 = 0$ and we can check whether this hypothesis is rejected or not by (3). Here, we suppose that the model to estimate is the multiple regression model who has a constant term. Under H_0, we can rewrite (3) as follows:

$$\frac{\hat{b}'X'X\hat{b}/K}{\hat{\sigma^2}} \sim F(K, n-K).$$

1.2 Testing For Linear Restrictions

We also use the F statistic to test the linear restrictions. After explaining this test statistic, we explain the relationship between the F statistic and the R^2 coefficient.

Theorem 1.3. Under the assumptions of the classical linear regression model and the normality of the disturbance terms, we can perform a test of the null hypothesis $H_0 : Rb - r = 0$, $R \in \mathcal{M}_{q\times K}(\mathbb{R})$ with rank $q \leq K-1$ based on the Fisher statistic such as:

$$F = \frac{\hat{\Delta}'(R'(XX')^{-1}R)^{-1}\hat{\Delta}/q}{\hat{\sigma^2}} \sim F(q, n-K), \tag{6}$$

where $\hat{\sigma^2} = \frac{1}{n-K}\sum_{i=1}^{n} \hat{u_i}^2$ and $\hat{\Delta} = R\hat{b} - r$. In the above equation, q is the number of the restrictions.

Proof. We assume $u|X \sim N_{\mathbb{R}^n}(0, \sigma^2 I_n)$, which implies $\hat{b}|X \sim N_{\mathbb{R}^k}(b, \sigma^2(X'X)^{-1})$. Then, we can say $R\hat{b} - r|X \sim N_{\mathbb{R}^q}(0, \sigma^2 R(X'X)^{-1}R')$ and derive the Wald criterion such as:

$$W_0 = \frac{(R\hat{b}-r)'\{R(X'X)^{-1}R'\}^{-1}(R\hat{b}-r)}{\sigma^2} \sim \chi^2(q), \tag{7}$$

because of the following lemma.

Lemma 1.4. If a vector y follows the normal distribution, $y \sim N_{\mathbb{R}^n}(\mu, \Sigma)$, then we can derive $(y-\mu)'\Sigma^{-1}(y-\mu) \sim \chi^2(n)$.

The proof of the above lemma is omitted because we explained this theorem in the TA session #03. In addition, we know the following relationship by the previous subsection:

$$\frac{\hat{u}'\hat{u}}{\sigma^2} = \frac{(n-K)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-K).$$

Therefore, under the normality assumption on u and thereby $\hat{b} \perp \hat{u}$ (more accurately, by Cochran's theorem), we can derive a test statistic for linear restrictions as follows:

$$f = \frac{(R\hat{b}-r)'\{R(X'X)^{-1}R'\}^{-1}(R\hat{b}-r)/q}{\hat{u}'\hat{u}/(n-K)} \sim F(q, n-K). \tag{8}$$

□

You need not prove the independence of the numerator and denominator of the F-statistic if they take the forms $A'ZA$ and $B'ZB$, where A and B are vectors of certain values, and Z is a random variable.

1.3 Coefficient of Determination and F Distribution

1.3.1 Coefficient of Determination

The coefficient of determination measures the portion of variance explained by the model and is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^n \hat{u_i}^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \in [0, 1]. \tag{9}$$

Lemma 1.5. In the matrix form, we we can rewrite R^2 as follows:

$$R^2 = 1 - \frac{\hat{u}'\hat{u}}{y'(I_n - \frac{1}{n}ii')y}, \tag{10}$$

where $i \in \mathbb{R}^n$ is a vector of 1s.

Here, an operator $M_i = (I_n - \frac{1}{n}ii')$ is a symmetric and idempotent matrix and this is an operator to measure the deviation of a vector $x \in \mathbb{R}^{n\times 1}$ from its mean:

$$\begin{aligned} M_i x &= (I_n - \frac{1}{n}ii')x \\ &= x - \frac{1}{n}ii'x \\ &= \begin{pmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{pmatrix}, \end{aligned} \tag{11}$$

because ii' is equal to a $n \times n$ matrix whose all elements are 1. The proof of this Lemma is given as follows.

Proof. In this lemma, we have to show that $\sum_{i=1}^n (y_i - \bar{y})^2 = y'(I_n - \frac{1}{n}ii')y$.

$$\begin{aligned} y'M_i y &= (y_1, \cdots, y_n) \begin{pmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix} \\ &= \sum_{i=1}^n y_i(y_i - \bar{y}) = \sum_{i=1}^n (y_i^2 - y_i\bar{y}) \\ &= \sum_{i=1}^n y_i^2 - \bar{y}\sum_{i=1}^n y_i = \sum_{i=1}^n y_i^2 - n\bar{y}^2 = \sum_{i=1}^n (y_i - \bar{y})^2 \end{aligned} \tag{12}$$

□

1.3.2 Relationship between R^2 and F Distribution

We now show that the R^2 can be used to test whether the predictive variables can explain the dependent variable. Consider the regression model as follows:

$$y = Xb + u,$$

where $X = (i \mid \tilde{x}) \in \mathbb{R}^{n\times K}$, $\tilde{x} = (X_2, \ldots, X_K) \in \mathbb{R}^{n\times K-1}$, $b = (b_1, b_2')' \in \mathbb{R}^K$. Then, we would like to test $H_0 : b_2 = 0$, where $b_2 \in \mathbb{R}^{K-1}$. Using the test for linear constraints, we take $R = (0 \mid I_{K-1}) \in \mathbb{R}^{K-1\times K}$, $r = 0 \in \mathbb{R}^{K-1}$ and the test statistic is given as follows:

$$f_0 = \frac{(R\hat{b} - r)'\{R(X'X)^{-1}R'\}^{-1}(R\hat{b} - r)/K - 1}{\hat{u}'\hat{u}/(n - K)} \sim F(K - 1, n - K). \tag{13}$$

At first, we are going to show $(R\hat{b} - r)'\{R(X'X)^{-1}R'\}^{-1}(R\hat{b} - r) = \hat{b}_2'X_2'M_iX_2\hat{b}_2$, where $M_i = I_n - \frac{1}{n}ii'$. $R(X'X)^{-1}R'$ is represented as follows:

$$R(X'X)^{-1}R' = (0\ I_{K-1}) \begin{pmatrix} i'i & i'\tilde{X} \\ \tilde{X}'i & \tilde{X}'\tilde{X} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ I_{K-1} \end{pmatrix} \tag{14}$$

In the above equation, an inverse of the partition matrix is rewritten as follows:[2]

$$\begin{pmatrix} i'i & i'\tilde{X} \\ \tilde{X}'i & \tilde{X}'\tilde{X} \end{pmatrix}^{-1} = \begin{pmatrix} \cdot & \cdots \\ \vdots & (\tilde{X}'M_i\tilde{X})^{-1} \end{pmatrix}.$$

Hence, we can derive $R(X'X)^{-1}R' = (\tilde{X}'M_i\tilde{X})^{-1}$. As a consequence, under $H_0 : b_2 = 0$, we have the following relationship:

$$\begin{aligned} f_0 &= \frac{(R\hat{b} - r)'\{R(X'X)^{-1}R'\}^{-1}(R\hat{b} - r)/K - 1}{\hat{u}'\hat{u}/(n - K)} \\ &= \frac{\hat{b}_2'\tilde{X}'M_i\tilde{X}\hat{b}_2/K - 1}{\hat{u}'\hat{u}/(n - K)} \sim F(K - 1, n - K). \end{aligned} \tag{15}$$

In the (10), we have $M_i\hat{u} = M_iy - M_iX\hat{b}$. Because of $i'\hat{u} = 0$[3] and $M_iX = (0 \mid M_i\tilde{X})$, we can calculate M_iy as follows:

$$\begin{aligned} M_iy &= M_iX\hat{b} + M_i\hat{u} \\ &= M_i\tilde{X}\hat{b}_2 + \hat{u}. \end{aligned}$$

Therefore, we can show that $y'M_iy = \hat{b}_2'\tilde{X}'M_i\tilde{X}\hat{b}_2 + \hat{u}'\hat{u}$. Hence, the (15) is represented as follows:

$$\begin{aligned} f_0 &= \frac{\hat{b}_2'\tilde{X}'M_i\tilde{X}\hat{b}_2/(K - 1)}{\hat{u}'\hat{u}/(n - K)} \\ &= \frac{R^2y'M_iy/(K - 1)}{(1 - R^2)y'M_iy/(n - K)} \\ &= \frac{R^2/(K - 1)}{(1 - R^2)/(n - K)} \sim F(K - 1, n - K). \end{aligned} \tag{16}$$

As a consequence, using the R^2 coefficient, the null hypothesis $H_0 : b_2 = 0$ can be easily tested.

[2] More discussion for the inverse matrix of the partition matrix is explained in the Appendix.

[3] A sum of residuals are equal to zero by definition.

2 Constrained OLS

2.1 Deriviation of COLS estimator

Suppose that we estimate an econometric model which has a restriction on its parameters. If we have a priori knowledge that there are some restrictions in the model, we should estimate constrained OLS (:COLS) estimator.

Theorem 2.1. The constrained least squares estimator satisfies following optimization problem such that:

$$\tilde{b} = \arg\min_{b} \frac{1}{2}||y - Xb||_2^2 \quad s.t. \quad Rb = r,$$

where $b \in \mathbb{R}^K$, $R \in \mathbb{R}^{q \times K}$ and $r \in \mathbb{R}^q$. The constrained least squares estimator is given as follows:

$$\tilde{b} = \hat{b} + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\hat{b}), \tag{17}$$

where $\hat{b}$ is called as the unconstrained least squares estimator.

The criterion that we minimize is given as follows:

$$(y - Xb)'(y - Xb) \quad s.t. \quad Rb = r.$$

We can solve the above minimization problem through first order conditions of the Lagrangian function as follows:

$$\mathcal{L}(b, \lambda) = (y - Xb)'(y - Xb) - 2\lambda'(Rb - r). \tag{18}$$

The first order conditions of (18) is derived as follows:

$$\nabla_b \mathcal{L}(\cdot, \cdot) = -2X'(y - X\tilde{b}) - 2R'\lambda^* = 0,$$
$$\nabla_\lambda \mathcal{L}(\cdot, \cdot) = -2(R\tilde{b} - r) = 0.$$

By using these first order conditions, we can proof the above theorem. More details of the proof is explained in the Appendix.

2.2 Properties of COLS estimator

Next, we introduce the expectation and variance of COLSE. Since $Rb = r$, we can easily show that COLSE is a unbiased estimator.

$$\begin{aligned}\mathbb{E}[\tilde{b}|X] &= \mathbb{E}[\hat{b}|X] + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\hat{b}) \\ &= b + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - Rb) = b\end{aligned}$$

For our convenience, we denote $\tilde{b} - b$ as follows:

$$\begin{aligned}\tilde{b} - b &= (\hat{b} - b) + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(Rb - R\hat{b}) \\ &= [I_K - (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R](\hat{b} - b) := W(\hat{b} - b).\end{aligned}$$

Thus, the variance of $\tilde{b}$ is written as:

$$\begin{aligned}
\mathrm{Var}(\tilde{b}|X) &= \mathbb{E}[(\tilde{b}-b)(\tilde{b}-b)'|X] \\
&= \mathbb{E}[W(\hat{b}-b)(\hat{b}-b)'W'|X] \\
&= W\mathbb{E}[(\hat{b}-b)(\hat{b}-b)'|X]W' \\
&= W\mathrm{Var}(\hat{b}|X)W' \\
&= \sigma^2 W(X'X)^{-1}W'. \qquad (19)
\end{aligned}$$

Expanding with respect to W, the variance of COLSE is smaller than the variance of the any other (unrestricted) OLSE:

$$\begin{aligned}
\mathrm{Var}(\tilde{b}|X) &= \sigma^2(X'X)^{-1} - \sigma^2(X'X)^{-1}R'(R(X'X)^{-1}R')R(X'X)^{-1} \\
&= \mathrm{Var}(\hat{b}|X) - \sigma^2(X'X)^{-1}R'(R(X'X)^{-1}R')R(X'X)^{-1}. \qquad (20)
\end{aligned}$$

Remember that the above properties are satisfied if the restriction(s) are certainly correct.

2.3 F Distribution for OLS and COLS

In the case of the restricted OLS, the test statistic f_0 in (13) is rewritten by the simple representation which we use the residual of COLS. By using (17), the numerator of (13) is equal to $(\hat{b}-\tilde{b})'X'X(\hat{b}-\tilde{b})$. Moreover, the numerator is represented as another equation such that:

$$\begin{aligned}(y-X\tilde{b})'(y-X\tilde{b}) &= (y-X\hat{b}-X(\tilde{b}-\hat{b}))'(y-X\hat{b}-X(\tilde{b}-\hat{b}) \\ &= (y-X\hat{b})'(y-X\hat{b})+(\tilde{b}-\hat{b})'X'X(\tilde{b}-\hat{b}) \\ &- (y-X\hat{b})'X(\tilde{b}-\hat{b})-(\tilde{b}-\hat{b})'X'(y-X\hat{b}) \\ &= (y-X\hat{b})'(y-X\hat{b})+(\tilde{b}-\hat{b})'X'X(\tilde{b}-\hat{b}),\end{aligned}$$

where $X'(y-X\hat{b})=X'\hat{u}=0$ is used. In short, the part of numerator in (13) is rewritten as follows:

$$\begin{aligned}(R\hat{b}-r)'\{R(X'X)^{-1}R'\}(R\hat{b}-r) &= (\hat{b}-\tilde{b})'X'X(\hat{b}-\tilde{b}) \\ &= (y-X\tilde{b})'(y-X\tilde{b})-(y-X\hat{b})'(y-X\hat{b}) \\ &= \tilde{u}'\tilde{u}-\hat{u}'\hat{u}.\end{aligned}$$

As a consequence, we can obtain f_0 by the simple calculation such that:

$$\begin{aligned}f_0 &= \frac{(R\hat{b}-r)'\{R(X'X)^{-1}R'\}^{-1}(R\hat{b}-r)/K-1}{\hat{u}'\hat{u}/(n-K)} \\ &= \frac{\tilde{u}'\tilde{u}-\hat{u}'\hat{u}/q}{\hat{u}'\hat{u}/(n-K)}. \qquad (21)\end{aligned}$$

3 R Exercise

In this subsection, we will explain how to use R. Today, we use daily data of the Turkish stock market from 2009/1/5 to 2009/6/26. Consider the following regression model:

$$\begin{aligned}\Delta log(Turkish\ Stock\ Market)_t = a + b_{21}\Delta log(US\ Stock\ Market)_{t-1} \\ + b_{22}\Delta log(EU\ Stock\ Market)_{t-1} + u_t.\end{aligned}$$

where Turkish Stock Market implies Istanbul stock exchange national 100 index, US Stock Market implies Standard & Poor's 500 return index, and EU Stock Market implies MSCI European index[4]. In this class, we test a null hypothesis $H_0 : b_2 = (b_{21}, b_{22})' = 0$ by the F statistic. Although some packages to test OLS models with restrictions may exist, we calculate the F statistic directly.

[4]The dataset for this model is uploaded in the UCI Machine Learning Repository, `https://archive.ics.uci.edu/ml/datasets/ISTANBUL+STOCK+EXCHANGE`. This data is used in Akbilgic, O., Bozdogan, H., Balaban, M.E., (2013) "A novel Hybrid RBF Neural Networks model as a forecaster", Statistics and Computing. We reduce the estimated period for our convinience.

```
library(stargazer)

dataset<-read.csv("data_akbilgic.csv",header=TRUE,stringsAsFactors=FALSE)
turkish<-dataset[,2]
US<-dataset[,4]
EUlag<-dataset[,9]
lnturkish<-diff(turkish)
lnUSlag<-diff(US)
lnEUlag<-diff(EUlag)
#Be careful that this model is the predict model: time series analysis.
#Because these index may not have the weak stationarity,
#we take the log. diff. .

lnturkish<-lnturkish[2:124]
lnUSlag<-lnUSlag[1:123]
lnEUlag<-lnEUlag[1:123]

variableset<-data.frame(lnturkish,lnUSlag,lnEUlag)
stock.lm<-lm(lnturkish~lnUSlag+lnEUlag,data=variableset)
summary(stock.lm)

#Firstly, derive the estimators of coefficients.
hat_b<-coef(stock.lm)

#Suppose that we want to check H0:b_2=0, where b2 is
#the coef. vec. of log.diff.(US)&log.diff.(EU).
#Convert the data frame into a matrix.
#Please make a matrix R and vector r.

x<-as.matrix(variableset)
R<-matrix(c(0,1,0,0,0,1),2,3,byrow=TRUE)
r<-c(0,0)

#numerator
A<-(R%*%hat_b)-r
B<-R%*%solve(t(x)%*%x)%*%t(R)
nume<-(t(A)%*%solve(B)%*%A)/2

#denominator
hat_u<-residuals(stock.lm)
denom<-(t(hat_u)%*%hat_u)/120

#F statistic: 2 restrictions/ n =123, K=3
F=nume/denom
F
#F=10.46914/ H_0 is rejected in 5% conf. interval.

stargazer(stock.lm,title="Turkish Stock Market",style="all",type="latex")
```

The detail of the result is written in the Appendix.

4 Appendix

4.1 Review of a Partition Matrix

Partitioning the matrix to make some groups of elements is useful. For instance, we can make blocks of the elements of matrix X as follows.

$$X = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{21} \end{pmatrix}$$

In this lecture note, we calculate an inverse of the partition matrix.

Theorem 4.1. Suppose that X_{11} and X_{22} is the square and positive definite matrix in the above example. Then, we can derive the inverse matrix of X as follows:

$$X^{-1} = \begin{pmatrix} X_{11} + X_{11}^{-1}X_{12}Y_{22}X_{21}X_{11}^{-1} & -X_{11}^{-1}X_{12}Y_{22} \\ -Y_{22}X_{21}X_{11}^{-1} & Y_{22} \end{pmatrix}, \tag{A1}$$

with $Y_{22} = (X_{22} - X_{21}X_{11}^{-1}X_{12})^{-1}$. Or alternatively, we can represent X^{-1} as following matrix such as:

$$X^{-1} = \begin{pmatrix} Y_{11} & -Y_{11}X_{12}X_{22}^{-1} \\ -X_{22}^{-1}X_{21}Y_{11} & X_{22}^{-1} + X_{22}^{-1}X_{21}Y_{11}X_{12}X_{22}^{-1} \end{pmatrix}, \tag{A2}$$

with $Y_{11} = (X_{11} - X_{21}X_{11}^{-1}X_{12})^{-1}$.

Now we explain how to proof (A1).

Proof. Firstly, we can decompose X as a product of the lower and upper triangular matrices such as:

$$X = \begin{pmatrix} I & 0 \\ S & I \end{pmatrix} \begin{pmatrix} P & Q \\ 0 & R \end{pmatrix} = \begin{pmatrix} P & Q \\ SP & SQ + R \end{pmatrix}.$$

In the above equation, $P = X_{11},\ Q = X_{12},\ S = X_{21}X_{11}^{-1}$, and $R = X_22 - X_{21}X_{11}^{-1}X_{12}$.

In the same manner, we can decompose the second matrix in (RHS) of the above equation as follows:

$$\begin{pmatrix} X_{11} & X_{12} \\ 0 & X_{22} - X_{21}X_{11}^{-1}X_{12} \end{pmatrix} = \begin{pmatrix} X_{11} & 0 \\ 0 & X_{22} - X_{21}X_{11}^{-1}X_{12} \end{pmatrix} \begin{pmatrix} I & X_{11}^{-1}X_{12} \\ 0 & I \end{pmatrix}.$$

Therefore, we can rewrite X as a product of a lower triangular matrix, a diagonal matrix and a upper triangular matrix. This is a kind of the LU decomposition.

$$X = \begin{pmatrix} I & 0 \\ X_{21}X_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} X_{11} & 0 \\ 0 & X_{22} - X_{21}X_{11}^{-1}X_{12} \end{pmatrix} \begin{pmatrix} I & X_{11}^{-1}X_{12} \\ 0 & I \end{pmatrix} \tag{A3}$$

In the above equation, we can apply useful formulas as follows:

$$\begin{pmatrix} I & 0 \\ X_{21}X_{11}^{-1} & I \end{pmatrix}^{-1} = \begin{pmatrix} I & 0 \\ -X_{21}X_{11}^{-1} & I \end{pmatrix},$$

$$\begin{pmatrix} X_{11} & 0 \\ 0 & X_{22} - X_{21}X_{11}^{-1}X_{12} \end{pmatrix}^{-1} = \begin{pmatrix} X_{11}^{-1} & 0 \\ 0 & (X_{22} - X_{21}X_{11}^{-1}X_{12})^{-1} \end{pmatrix}$$

and

$$\begin{pmatrix} I & X_{11}^{-1}X_{12} \\ 0 & I \end{pmatrix}^{-1} = \begin{pmatrix} I & -X_{11}^{-1}X_{12} \\ 0 & I \end{pmatrix}.$$

By taking the inverse of both sides in (A3), we can prove (A1). □

Finally, we introduce the special case of the partition matrix such as:

$$X^* = \begin{pmatrix} A & B \\ B' & D \end{pmatrix}.$$

In this case, we can calculate the inverse matrix of X^* as follows:

$$\begin{pmatrix} A & B \\ B' & D \end{pmatrix}^{-1} = \begin{pmatrix} E & F \\ F' & G \end{pmatrix}. \tag{A4}$$

with the following notations:

$$\begin{aligned} E &= A^{-1} + A^{-1}B(D - B'A^{-1}B)^{-1}B'A, \\ F &= A^{-1}B(D - B'A^{-1}B)^{-1}, \\ G &= D^{-1} + D^{-1}B'(A - BD^{-1}B')^{-1}BD^{-1}. \end{aligned}$$

4.2 Deriviation of COLSE

The first order conditions of (18) is derived as follows:

$$\frac{\partial \mathcal{L}(\cdot,\cdot)}{\partial b} = -2X'(y - Xb) - 2R'\lambda = 0, \tag{A5}$$

$$\frac{\partial \mathcal{L}(\cdot,\cdot)}{\partial \lambda} = -2(Rb - r) = 0. \tag{A6}$$

By premultiplying $(X'X)^{-1}$ into the (A5), we have $\tilde{b}$ as follows:

$$\tilde{b} = (X'X)^{-1}X'y + (X'X)^{-1}R'\tilde{\lambda} = \hat{b} + (X'X)^{-1}R'\tilde{\lambda} \tag{A7}$$

Multiplying by R from the left side of the above equation, $R\tilde{b} = R\hat{b} + R(X'X)^{-1}R'\tilde{\lambda}$. Since $R\tilde{b} = r$ must be satisfied, then we also calculate r as $r = R\hat{b} + R(X'X)^{-1}R'\tilde{\lambda}$. Hence, solving the above equation w.r.t. $\tilde{\lambda}$, we obtain $\tilde{\lambda} = [R(X'X)^{-1}R']^{-1}(r - R\hat{b})$. Plugging $\tilde{\lambda}$ into (A7), the COLSE becomes as follows:

$$\tilde{b} = \hat{b} + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r - R\hat{b}). \tag{A8}$$

The orthogonality condition of the OLSE, $X'\hat{u} = 0$, establishes the above equation such that:

$$(X'X)\tilde{b} - R'\tilde{\lambda} = X'y, \quad R\tilde{b} = r,$$

because we have following equation by multiplying $(X'X)$ to the both sides of (A8):

$$\begin{aligned}(X'X)\tilde{b} - R'\tilde{\lambda} &= X'X\hat{b} \\ &= X'\underline{\hat{y}} \\ &= X'(y - \hat{u}).\end{aligned}$$

Equivalently, we can write the orthogonality condition as a matrix form such as:

$$\begin{pmatrix} X'X & R' \\ R & 0 \end{pmatrix} \begin{pmatrix} \tilde{b} \\ -\tilde{\lambda} \end{pmatrix} = \begin{pmatrix} X'y \\ r \end{pmatrix}.$$

The solution of $\tilde{b}$ and $\tilde{\lambda}$ are given as following matrix representation such as:

$$\begin{pmatrix} \tilde{b} \\ -\tilde{\lambda} \end{pmatrix} = \begin{pmatrix} X'X & R' \\ R & 0 \end{pmatrix}^{-1} \begin{pmatrix} X'y \\ r \end{pmatrix}. \tag{A9}$$

Therefore, we can derive $\tilde{b}$ as (17) because of the representation of (A4).

4.3 Result of R Exercise

Table 1: Turkish Stock Market

	Dependent variable:
	lnturkish
lnUSlag	0.215^{**}
	(0.091)
	t = 2.367
	p = 0.020
lnEUlag	-0.600^{***}
	(0.120)
	t = −4.999
	p = 0.00001
Constant	−0.0002
	(0.002)
	t = −0.101
	p = 0.920
Observations	123
R^2	0.174
Adjusted R^2	0.161
Residual Std. Error	0.027 (df = 120)
F Statistic	12.680^{***} (df = 2; 120) (p = 0.00002)
Note:	*p<0.1; **p<0.05; ***p<0.01

4.4 Answer for Questions

Consider the case of a parameter θ_0 to be estimated and sequence of its estimator $\hat{\theta_n}$. If $\hat{\theta_n}$ has an asymptotic normality, $\sqrt{n}(\hat{\theta_n} - \theta_0) \xrightarrow[n\to\infty]{d} N(0, \Sigma)$, we can derive the following theorem.

Theorem 4.2. Suppose any continuous, differentiable function $g : \mathbb{R}^d \to \mathbb{R}^s$. Let $X_1, X_2, \ldots, X_n$ a sequence of d-dimensional random variables. If $\hat{\theta_n}$ has an asymptotic normality, we can state that:

$$\sqrt{n}[g(\hat{\theta_n}) - g(\theta_0)] \xrightarrow[n\to\infty]{d} N(0, D_g(\theta_0)\Sigma D_g'(\theta_0)), \tag{A10}$$

where $D_g(\theta)$ is a Jacobian matrtix of θ.

Suppose that there is a parameter $\bar{\theta} \in (\theta_0, \hat{\theta_n})$. Then, we can use a first order Taylor expansion such as:

$$g(\hat{\theta_n}) = g(\theta_0) + D_g(\bar{\theta})(\hat{\theta_n} - \theta_0). \tag{A11}$$

Please recall that the (A11) is established because of the mean value theorem.

Theorem 4.3. Let $f(x)$ be continuous on the interval $[a, b]$ and it can be differentiated at the interval of (a, b). Then, there is at least one point $c \in [a, b]$ that satisfies following equation:

$$\frac{f(b) - f(a)}{b - a} = f'(c). \tag{A12}$$

By calculating the above equation, we can get $f(b) = f(a) + f'(c)(b - a)$. In addition, a general case of the Taylor expansion is given as follows:

Theorem 4.4. Suppose a function $f(x)$, which is continuous on an interval $[a, b]$, can be differentiated n times in an interval (a, b). Then, there is a point $c \in (a, b)$ which satisfies following equation:

$$\begin{aligned} f(b) = f(a) &+ \frac{f'(a)}{1!}(b - a) + \frac{f''(a)}{2!}(b - a)^2 + \cdots \\ &+ \frac{f^{(n-1)}(a)}{(n-1)!}(b - a)^{(n-1)} + \frac{f^{(n)}(c)}{n!}(b - a)^n. \end{aligned} \tag{A13}$$

This equation implies that the first order Taylor expansion is same as the mean value theorem. Now, we prove the (A10).

Proof. Because of the mean value expansion, we can state:

$$\begin{aligned} \sqrt{n}[g(\hat{\theta_n}) - g(\theta_0)] &= \sqrt{n} D_g(\bar{\theta})(\hat{\theta_n} - \theta_0) \\ &= \sqrt{n} D_g(\theta_0)(\hat{\theta_n} - \theta_0) + \sqrt{n}[D_g(\bar{\theta}) - D_g(\theta_0)](\hat{\theta_n} - \theta_0) \end{aligned} \tag{A14}$$

Here, if $\hat{\theta_n} \xrightarrow[n\to\infty]{p} \theta_0$ is given, $\bar{\theta} \xrightarrow[n\to\infty]{p} \theta_0$ is established. Therefore, the second term of (RHS) in the (A14) is calculated as:

$$[D_g(\bar{\theta}) - D_g(\theta_0)]\sqrt{n}(\hat{\theta_n} - \theta_0) = o_p(1)O_p(1) = o_p(1), \tag{A15}$$

because $\sqrt{n}(\hat{\theta_n} - \theta_0)$ converges in distribution.[5] By the (A14) and the (A15), we can derive:

$$\sqrt{n}[g(\hat{\theta_n}) - g(\theta_0)] = \sqrt{n} D_g(\theta_0)(\hat{\theta_n} - \theta_0) + o_p(1). \tag{A16}$$

Since we can apply the property of multivariate normal distribution in this equation, we can say $\sqrt{n} D_g(\theta_0)(\hat{\theta_n} - \theta_0) \xrightarrow[n\to\infty]{d} N(0, D_g(\theta_0)\Sigma D'_g(\theta_0))$. We can now apply the asymptotic equivalence lemma of the main text.

Lemma 4.5. Let $\{x_n\}$ and $\{y_n\}$ be sequences of $n \times 1$ random vectors. If $z_n \xrightarrow[n\to\infty]{d} z$ and $x_n - z_n \xrightarrow[n\to\infty]{p} 0$, then $x_n \xrightarrow[n\to\infty]{d} z$.

By using this lemma, we can derive (A10). □

[5] The lemma 3.5 in the main textbook implies that a $K \times 1$ vector x_n is $O_p(1)$ if x_n converges x in distribution.