Ch. 6 Linear Regression Model

(July 4, 2017)

1 Introduction

The (multiple) linear model is used to study the relationship between a dependent variable (Y) and several independent variables ($X_1, X_2, ..., X_k$). That is[1]

$$\begin{aligned} Y &= f(X_1, X_2, ..., X_k) + \varepsilon \quad assume\ linear\ function \\ &= \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + \varepsilon \\ &= \mathbf{x}'\boldsymbol{\beta} + \varepsilon \end{aligned}$$

where Y is the dependent or explained variable, $\mathbf{x} = [X_1\ X_2.....; X_k]'$ are the independent or the explanatory variables and $\boldsymbol{\beta} = [\beta_1\ \beta_2..... \beta_k]'$ are unknown coefficients that we are interested in learning about, either through estimation or through hypothesis testing. The term ε is an unobservable random disturbance.

Suppose we have a sample of size T observations[2] on the scalar dependent variable Y_t and the vector of explanatory variables $\mathbf{x}_t = (X_{t1}, X_{t2}, ..., X_{tk})'$, i.e.

$$Y_t = \mathbf{x}_t'\boldsymbol{\beta} + \varepsilon_t, \quad t = 1, 2, ..., T,$$

then in matrix form, this relationship can be written as

$$\begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_T \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & . & . & . & X_{1k} \\ X_{21} & X_{22} & . & . & . & X_{2k} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ X_{T1} & X_{T2} & . & . & . & X_{Tk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_T \end{bmatrix}, \ or$$

[1] When it is common to include that an intercept in the regression, then we have $X_1 = 1$ and then $Y = \beta_1 + \beta_2 X_2 + ... + \beta_k X_k + \varepsilon$.

[2] Recall from Chapter 2 that we cannot postulate the probability model Φ if the sample is non-random. The probability model must be defined in terms of their sample joint distribution.

$$\begin{aligned}\mathbf{y} &= \begin{bmatrix}\mathbf{x}_1' \\ \mathbf{x}_2' \\ . \\ . \\ . \\ \mathbf{x}_T'\end{bmatrix} \cdot \boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},\end{aligned}$$

where $\mathbf{y}$ is $T \times 1$ vector, $\mathbf{X}$ is an $T \times k$ matrix with rows $\mathbf{x}_t'$,[3] $\boldsymbol{\beta}$ is $k \times 1$ and $\boldsymbol{\varepsilon}$ is an $T \times 1$ vector with element ε_t.

Notation.

A linear regression model with k explanatory variables is

$$Y_{(1\times 1)} = \mathbf{x}'_{(1\times k)}\boldsymbol{\beta}_{(k\times 1)} + \varepsilon_{(1\times 1)}.$$

A sample of size T of the above linear regression model is

$$\mathbf{y}_{(T\times 1)} = \mathbf{X}_{(T\times k)}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_{(T\times 1)}. \qquad \blacksquare$$

Our goal is to regard last equation as a parametric probability and sampling model, and try to inference the unknown β_i's and the parameters contained in $\boldsymbol{\varepsilon}$.

[3] When $X_1 = 1$ for all t, then $\mathbf{X} = \begin{bmatrix} 1 & X_{12} & . & . & . & X_{1k} \\ 1 & X_{22} & . & . & . & X_{2k} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & X_{T2} & . & . & . & X_{Tk} \end{bmatrix}$.

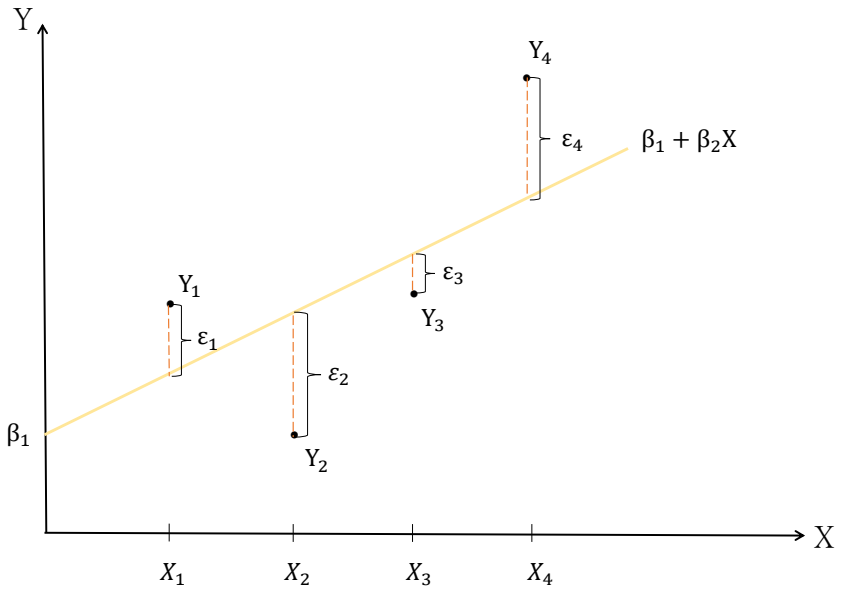

Figure (6-1). Population Regression Function.

1.1 The Probability Model: Gaussian Linear Model

Assume that $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Omega})$, if $\boldsymbol{X}$ are not stochastic, then by results from "functions of random variables" (n $\Rightarrow$ n transformation) we have $\mathbf{y} \sim N(\boldsymbol{X\beta},\ \boldsymbol{\Omega})$. That is, we have specified a probability and sampling model for $\mathbf{y}$ to be

(Probability and Sampling Model)

$$\begin{aligned}\mathbf{y} &\sim N\left(\begin{bmatrix} X_{11} & X_{12} & . & . & . & X_{1k} \\ X_{21} & X_{22} & . & . & . & X_{2k} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ X_{T1} & X_{T2} & . & . & . & X_{Tk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_k \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} & . & . & . & \sigma_{1T} \\ \sigma_{21} & \sigma_2^2 & . & . & . & \sigma_{2T} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \sigma_{T1} & \sigma_{T2} & . & . & . & \sigma_T^2 \end{bmatrix}\right) \\ &\equiv N(\mathbf{X\beta}, \boldsymbol{\Omega}).\end{aligned}$$

That is the sample joint density function is

$$f(\mathbf{y};\boldsymbol{\theta}) = (2\pi)^{-T/2}|\boldsymbol{\Omega}|^{-1/2}\exp(-1/2)(\mathbf{y}-\mathbf{X\beta})'\boldsymbol{\Omega}^{-1}(\mathbf{y}-\mathbf{X\beta}),$$

where $\boldsymbol{\theta} = (\beta_1, \beta_2, ..., \beta_k, \sigma_1^2, \sigma_{12}, ..., \sigma_T^2)'$. It is easy to see that the number of parameters in $\boldsymbol{\theta}$ is large than the sample size, T. Therefore, some restrictions must be imposed in the probability and sampling model for the purpose of estimation as we shall see in the subsequence.

One kind of restriction on $\boldsymbol{\theta}$ is that $\boldsymbol{\Omega}$ is a scalar matrix, then maximize the likelihood of the sample model $f(\boldsymbol{\theta}; \mathbf{x})$ (w.r.t. $\boldsymbol{\beta}$) is equivalent to minimize the equation $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ $(= \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = \sum_{t=1}^{T} \varepsilon_t^2$, a sums of squared residuals), this constitutes the foundation of ordinary least square estimation.

1.2 Assumptions of the Classical Model

The classical linear regression model consists of a set of assumptions about how a data set will be produced by an underlying "data-generating process". The theory will usually specify a precise, deterministic relationship between the dependent variable and the independent variables.

Assumption 1. (Classical Ideal Conditions)

(a). The model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ is correct; (no problem of model misspecification)

(b). $\text{Rank}(\mathbf{X}) = k$; (for model identification)

(c). $\mathbf{X}$ is nonstochastic[4] and $\lim_{T\to\infty}(\mathbf{X}'\mathbf{X}/T) = \underline{\mathbf{Q}}$, where $\underline{\mathbf{Q}}$ is a finite and nonsingular matrix.

(d). $E(\boldsymbol{\varepsilon}) = \mathbf{0}$. (This condition can easily be satisfied by adding a constant in the regression.)

(e). $Var(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2 \cdot \mathbf{I}$. (Disturbance have same variance and are not autocorrelated)

(f). $\boldsymbol{\varepsilon}$ is normal distributed. ■

The above six assumptions are usually called the classical ordinary least squares assumption or *the ideal conditions*.

[4]Therefore, regression comes first from experimental science.

2 Estimation: OLS Estimator

2.1 Population and Sample Regression Function

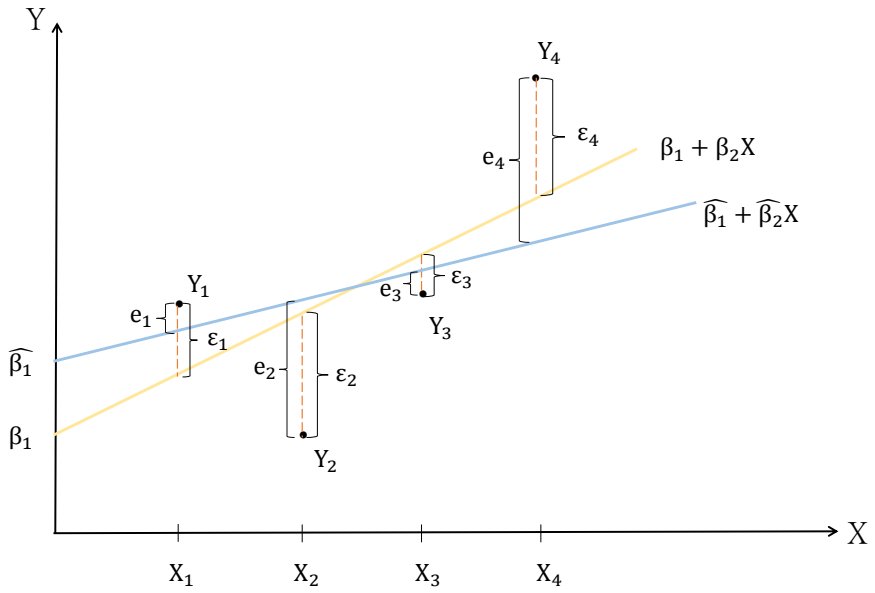

Figure (6-2). Population and Sample Regression Function.

2.2 Estimation of $\boldsymbol{\beta}$

Let us first consider the *ordinary least square estimator* (OLS) which is the value for $\boldsymbol{\beta}$ that minimizes the *sum of squared errors* (or residuals)[5] denoted as SSE which is defined as

$$\begin{aligned} SSE(\boldsymbol{\beta}) &= \sum_{t=1}^{T}(y_t - \mathbf{x}_t'\boldsymbol{\beta})^2 \\ &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}. \end{aligned}$$

[5]Remember the principal of estimation at Ch. 3.

The first order conditions for a minimum are

$$\frac{\partial SSE(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = 0.$$

If $\mathbf{X}'\mathbf{X}$ is nonsingular (in fact it is positive definite which is satisfied by the Assumption (1b) of ideal condition and p.41 of Ch.1 result (c)), the system of k equations in k unknown can be uniquely solved for the ordinary least squares (OLS) estimator

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \left[\sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t'\right]^{-1} \sum_{t=1}^{T} \mathbf{x}_t'\mathbf{y}_t. \tag{6-1}$$

To ensure that $\hat{\boldsymbol{\beta}}$ is indeed a solution of minimization, we require that

$$\frac{\partial^2 SSE(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} = 2\mathbf{X}'\mathbf{X}$$

must be a positive definite matrix. This condition is satisfied by p.41 of Ch.1 result (c).

Definition. (Residuals)
The OLS residuals are defined by the $T \times 1$ vector $\mathbf{e}$,

$$\mathbf{e} \;=\; \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}. \qquad \blacksquare$$

It is obvious that

$$\mathbf{X}'\mathbf{e} = \mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{0}, \tag{6-2}$$

i.e., the regressors are orthogonal to the OLS residual. Therefore, if one of the regressors is a constant term, the sum of the residuals is zero since the first element of $\mathbf{X}'\mathbf{e}$ would be

$$\begin{bmatrix} 1 & 1 & . & . & . & 1 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ . \\ e_T \end{bmatrix} = \sum_{t=1}^{T} e_t = 0. \; (a\ scalar)$$

Exercise 1.[6]
Reproduce the results on p.25 from the data in Table 3.1 (p.23) of Greene 6th edition.■

2.3 Estimation of σ^2

At this moment, we arrive at the following notation:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ &= \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{e}. \end{aligned}$$

Therefore, to estimate the variance of ε, i.e., σ^2, a simple and intuitive idea is that to use information obtained from sample such as $\mathbf{e}$.[7] To serve as a proxy for $\boldsymbol{\varepsilon}$, we must establish the relationship between $\mathbf{e}$ and $\boldsymbol{\varepsilon}$.

Definition. (Residuals Maker Matrix)
The matrix $\mathbf{M_X} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is symmetric and idempotent. $\mathbf{M_X}$ produces the vector of least square residuals in the regression of $\mathbf{y}$ on $\mathbf{X}$. Furthermore, $\mathbf{M_X X} = \mathbf{0}$ and $\mathbf{M_X e} = \mathbf{e}$. ■

Theorem.
$\mathbf{e} = \mathbf{M_X y} = \mathbf{M_X}\boldsymbol{\varepsilon}$.

Proof.

[6] All the data can be downed from: http://pages.stern.nyu.edu/∼wgreene/Text/econometricanalysis.htm
[7] To estimate the variance of ε, we would use the estimator $\sum \varepsilon_t^2/T$. However, ε_t is not observed directly, hence we use the information from e_t.

By definition,

$$\begin{aligned}
\mathbf{e} &= \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \\
&= \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
&= (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y} \\
&= \mathbf{M}_{\mathbf{X}}\mathbf{y} \\
&= \mathbf{M}_{\mathbf{X}}\mathbf{X}\boldsymbol{\beta} + \mathbf{M}_{\mathbf{X}}\boldsymbol{\varepsilon} \\
&= \mathbf{M}_{\mathbf{X}}\boldsymbol{\varepsilon}. \qquad \blacksquare
\end{aligned}$$

Using the fact that $\mathbf{M}_{\mathbf{X}}$ is symmetric and idempotent we have the following relation between $\mathbf{e}$ and $\boldsymbol{\varepsilon}$.

Lemma.
$\mathbf{e}'\mathbf{e} = \boldsymbol{\varepsilon}'\mathbf{M}_{\mathbf{X}}'\mathbf{M}_{\mathbf{X}}\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}'\mathbf{M}_{\mathbf{X}}\boldsymbol{\varepsilon}$. ■

Theorem. (The Expectation of the Sums of squared Residuals)
$E(\mathbf{e}'\mathbf{e}) = \sigma^2(T-k)$.

Proof.

$$\begin{aligned}
E(\mathbf{e}'\mathbf{e}) &= E(\boldsymbol{\varepsilon}'\mathbf{M}_{\mathbf{X}}\boldsymbol{\varepsilon}) \\
&= E[trace\ (\boldsymbol{\varepsilon}'\mathbf{M}_{\mathbf{X}}\boldsymbol{\varepsilon})] \quad (since\ \boldsymbol{\varepsilon}'\mathbf{M}_{\mathbf{X}}\boldsymbol{\varepsilon},\ is\ a\ scalar,\ equals\ its\ trace) \\
&= E[trace\ (\mathbf{M}_{\mathbf{X}}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')] \\
&= trace\ E(\mathbf{M}_{\mathbf{X}}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')] \quad (Why\ ?) \\
&= trace\ (\mathbf{M}_{\mathbf{X}}\sigma^2\mathbf{I}_T) \\
&= \sigma^2\ trace\ (\mathbf{M}_{\mathbf{X}}),
\end{aligned}$$

but

$$\begin{aligned} trace(\ \mathbf{M_X}) &= trace\ (\mathbf{I}_T) - trace\ (\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}') \\ &= trace\ (\mathbf{I}_T) - trace\ ((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}) \\ &= trace\ (\mathbf{I}_T) - trace\ (\mathbf{I}_k) \\ &= T - k. \end{aligned}$$

Hence $E(\mathbf{e}'\mathbf{e}) = \sigma^2(T-k)$. ■

Result. (An Unbiased Estimator of σ^2 from $\mathbf{e}$)
An unbiased estimator of σ^2 is therefore suggested as

$$s^2 = \frac{\mathbf{e}'\mathbf{e}}{T-k}.$$ ■

The squared root of s^2, i.e. $s = \sqrt{s^2}$ is sometimes called the "*standard error of the regression.*"

Definition. (Projection Matrix)
The matrix $\mathbf{P_X} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is symmetric and idempotent. $\mathbf{P_X}$ produces the fitted values in least square residuals in the regression of $\mathbf{y}$ on $\mathbf{X}$.[8] Furthermore, $\mathbf{P_X}\mathbf{X} = \mathbf{X}$ and $\mathbf{P_X}\mathbf{e} = \mathbf{0}$. ■

The vector $\mathbf{X}\boldsymbol{\beta}$ is always in the column space of $\mathbf{X}$, and $\mathbf{y}$ is unlikely to be in the column space. So, we project $\mathbf{y}$ onto a vector $\mathbf{p}(= \mathbf{P_X}\mathbf{y})$ in the column space of $\mathbf{X}$ and solve $\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{p}$.

Theorem.
$\mathbf{P_X}\mathbf{M_X} = \mathbf{0}$ and $\mathbf{P_X} + \mathbf{M_X} = \mathbf{I}$.

Proof.
By definition

$$\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}\right] = \left[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\mathbf{y} = \mathbf{P_X}\mathbf{y}.$$

[8] It creates the projection of the vector $\mathbf{y}$ into the column space of $\mathbf{X}$.

Hence

$$\begin{aligned}\mathbf{P_X M_X} &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right] \\ &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= \mathbf{0}.\end{aligned}$$

■

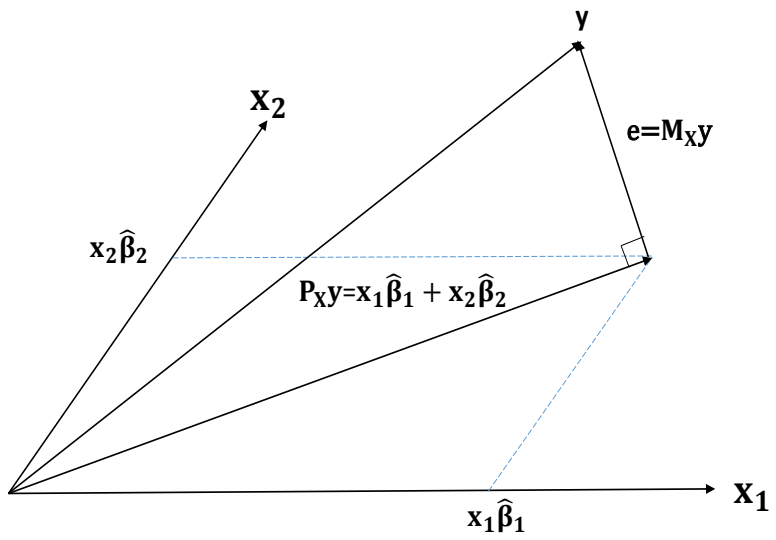

Figure (6-3). The orthogonal projection of $\mathbf{y}$ onto span$(\mathbf{x}_1, \mathbf{x}_2)$.

2.4 Partitioned Regression Estimation

It is common to specify a multiple regression model, when in fact, interest centers on only one of a subset of the full set of variables.[9] Let $k_1 + k_2 = k$ we can express the OLS result in isolation as

$$\begin{aligned}\mathbf{y} &= \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{e} \\ &= \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{bmatrix} + \mathbf{e} \\ &= \mathbf{X}_1\hat{\boldsymbol{\beta}}_1 + \mathbf{X}_2\hat{\boldsymbol{\beta}}_2 + \mathbf{e},\end{aligned}$$

where $\mathbf{X}_1$ and $\mathbf{X}_2$ are $T \times k_1$ and $T \times k_2$, respectively; $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ are $k_1 \times 1$ and $k_2 \times 1$, respectively.

What is the algebraic solution for $\hat{\boldsymbol{\beta}}_2$? Denote $\mathbf{M}_1 = \mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'$, then

$$\begin{aligned}\mathbf{M}_1\mathbf{y} &= \mathbf{M}_1\mathbf{X}_1\hat{\boldsymbol{\beta}}_1 + \mathbf{M}_1\mathbf{X}_2\hat{\boldsymbol{\beta}}_2 + \mathbf{M}_1\mathbf{e} \\ &= \mathbf{M}_1\mathbf{X}_2\hat{\boldsymbol{\beta}}_2 + \mathbf{e}, \qquad (6\text{-}3)\end{aligned}$$

where we have used the fact that $\mathbf{M_1}\mathbf{X}_1 = 0$ and $\mathbf{M_1}\mathbf{e} = \mathbf{e}$.

Multiplying $\mathbf{X}'_2$ on the above equation (6-3) and using the fact that

$$\mathbf{X}'\mathbf{e} = \begin{bmatrix} \mathbf{X}_1' \\ \mathbf{X}_2' \end{bmatrix} \mathbf{e} = \begin{bmatrix} \mathbf{X}_1'\mathbf{e} \\ \mathbf{X}_2'\mathbf{e} \end{bmatrix} = \mathbf{0},$$

we have

$$\mathbf{X}_2'\mathbf{M}_1\mathbf{y} = \mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2\hat{\boldsymbol{\beta}}_2 + \mathbf{X}_2'\mathbf{e} = \mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2\hat{\boldsymbol{\beta}}_2. \qquad (6\text{-}4)$$

Therefore $\hat{\boldsymbol{\beta}}_2$ can be expressed in isolation as

$$\begin{aligned}\hat{\boldsymbol{\beta}}_2 &= (\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{M}_1\mathbf{y} \\ &= (\mathbf{X}_2^{*'}\mathbf{X}_2^*)^{-1}\mathbf{X}_2^{*'}\mathbf{y}^*,\end{aligned}$$

where $\mathbf{X}_2^* = \mathbf{M}_1\mathbf{X}_2$ and $\mathbf{y}^* = \mathbf{M}_1\mathbf{y}$, are vectors of residual from the regression of $\mathbf{X}_2$ and $\mathbf{y}$ on $\mathbf{X}_1$, respectively.

[9] See for example, Pesaran (2007).

Theorem. (Frisch-Waugh)
In a linear least squares regression of vector $\mathbf{y}$ on two sets of variables, $\mathbf{X}_1$ and $\mathbf{X}_2$, the subvector $\hat{\boldsymbol{\beta}}_2$ is the set of coefficients obtained when the residuals from a regression of $\mathbf{y}$ on $\mathbf{X}_1$ alone are regressed on the set of residuals obtained when each column of $\mathbf{X}_2$ is regressed on $\mathbf{X}_1$. ■

This process is common called *partialing out* or *netting out* the effects of $\mathbf{X}_1$. For this reason, the coefficients in a multiple regression are often called the *partial regression coefficients*.

Example.
Consider a simple regression with a constant,

$$\mathbf{y} = \mathbf{i}\hat{\beta}_1 + \mathbf{X}_2\hat{\boldsymbol{\beta}}_2 + \mathbf{e},$$

where $\mathbf{i} = [1, ..., 1]'$, i.e. a column of one's, then the slope estimator $\hat{\boldsymbol{\beta}}_2$ can also be obtained from a data-demeaned regression without constant, i.e.

$$\begin{aligned}\hat{\boldsymbol{\beta}}_2 &= (\mathbf{X}_2'\mathbf{M}_i\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{M}_i\mathbf{y} \\ &= (\mathbf{X}_2^{*'}\mathbf{X}_2^{*})^{-1}\mathbf{X}_2^{*'}\mathbf{y}^{*},\end{aligned}$$

where $\mathbf{M}_i = \mathbf{I} - \mathbf{i}(\mathbf{i}'\mathbf{i})^{-1}\mathbf{i} = \mathbf{M}_0$ (the demeaned matrix earlier) and $\mathbf{X}_2^{*} = \mathbf{M}_i\mathbf{X}_2$ and $\mathbf{y}^{*} = \mathbf{M}_i\mathbf{y}$, are vectors of demeaned data of $\mathbf{X}$ and $\mathbf{y}$, respectively. ■

Exercise 2.
Reproduce the results $\hat{\beta}_2, ..., \hat{\beta}_5$ on p.25 from data in Table 3.1 (p.23) of Greene 6th edition by the demeaned data. ■

2.4.1 Partial Regression and Partial Correlation Coefficients

Consider the OLS estimation of the regression

$$y_t = \hat{\beta}_1 + \hat{\beta}_1 X_t + \hat{\beta}_2 Z_t + e_t, \quad t = 1, 2, ..., T$$

or

$$\mathbf{y} = \hat{\beta}_0\mathbf{i} + \hat{\beta}_1\mathbf{x} + \hat{\beta}_2\mathbf{z} + \mathbf{e}.$$

It is the characteristic of the regression that is implied by the term partial regression coefficients. The way we obtain $\hat{\beta}_2$, we have seen, is first to regression $\mathbf{y}$ and $\mathbf{z}$ on $\mathbf{i}$ and $\mathbf{x}$, and then to compute the residuals from this regression. By construction, $\mathbf{x}$ will not have any power in explaining variation in these residuals. Therefore, any correlation between $\mathbf{y}$ and $\mathbf{z}$ after this "purging" is independent of (or after removing the effect of) $\mathbf{x}$.

The same principle can be applied to the correlation between two variables. To continue our example, to what extent can we assert that this correlation reflects a direct relation rather than that both $\mathbf{z}$ and $\mathbf{y}$ tend, on average, to rise as $\mathbf{x}$ increases. To find out, we use the *partial correlation coefficient*, which is computed along the same line as the partial regression coefficient.

Definition. (Partial Correlation Coefficients)
The partial correlation coefficient between $\mathbf{y}$ and $\mathbf{z}$, controlling for the effect of $\mathbf{x}$, is obtained as follows:

(a). $\mathbf{y}^*$ =the residuals in a regression of $\mathbf{y}$ on a constant and $\mathbf{x}$, i.e., $\mathbf{y}^* = \mathbf{M}_{\tilde{\mathbf{x}}}\mathbf{y}$.

(b). $\mathbf{z}^*$ =the residuals in a regression of $\mathbf{z}$ on a constant and $\mathbf{x}$, i.e., $\mathbf{z}^* = \mathbf{M}_{\tilde{\mathbf{x}}}\mathbf{z}$, where $\tilde{\mathbf{X}} = (\mathbf{i}, \mathbf{X})$.

(c). The partial correlation r^*_{yz} is the simple correlation between $\mathbf{y}^*$ and $\mathbf{z}^*$, calculated as

$$r^*_{yz} = \frac{\mathbf{z}^{*'}\mathbf{y}^*}{\sqrt{\mathbf{z}^{*'}\mathbf{z}^*}\sqrt{\mathbf{y}^{*'}\mathbf{y}^*}} = \frac{\mathbf{z}'\mathbf{M}_{\tilde{\mathbf{x}}}\mathbf{y}}{\sqrt{\mathbf{z}'\mathbf{M}_{\tilde{\mathbf{x}}}\mathbf{z}}\sqrt{\mathbf{y}'\mathbf{M}_{\tilde{\mathbf{x}}}\mathbf{y}}}.$$ ■

It is noted here that in this linear regression $\mathbf{y} = \hat{\beta}_0\mathbf{i} + \hat{\beta}_1\mathbf{x} + \hat{\beta}_2\mathbf{z} + \mathbf{e}$,

$$\hat{\beta}_2 = \frac{\mathbf{z}^{*'}\mathbf{y}^*}{\sqrt{\mathbf{z}^{*'}\mathbf{z}^*}\sqrt{\mathbf{z}^{*'}\mathbf{z}^*}},$$

so r^*_{yz} and $\hat{\beta}_2$ will have the same signs. However, the simple correlation between $\mathbf{y}$ and $\mathbf{z}$ is obtained as

$$r_{yz} = \frac{\mathbf{z}'\mathbf{M_i}\mathbf{y}}{\sqrt{\mathbf{z}'\mathbf{M_i}\mathbf{z}}\sqrt{\mathbf{y}'\mathbf{M_i}\mathbf{y}}},$$

there is *no necessary relation* between the simple and partial correlation coefficients.

Exercise 3.

Reproduce the results in Table 3.2 (p.31) of Greene 6th edition. ■

2.5 The Linearly Restricted Least Squares Estimators

Suppose that we explicitly imposes the linear restrictions of the hypothesis in the regression (take the example of LM test). The restricted least square estimator is obtained as the solution to

$$Minimize_{\boldsymbol{\beta}}\ SSE(\boldsymbol{\beta}) \ = \ (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) \quad subject\ to\ \mathbf{R}\boldsymbol{\beta}=\mathbf{q},$$

where $\mathbf{R}$ is a known $J \times k$ matrix and $\mathbf{q}$ is known values of these linear restrictions.

A Lagrangian function for this problem can be written as

$$L^*(\boldsymbol{\beta},\lambda) = (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) + 2\boldsymbol{\lambda}'(\mathbf{R}\boldsymbol{\beta}-\mathbf{q}),\ where\ \ \boldsymbol{\lambda}\ is\ J\times 1.$$

The solutions $\hat{\boldsymbol{\beta}}^*$ and $\hat{\boldsymbol{\lambda}}$ will satisfy the necessary conditions

$$\begin{aligned}
\frac{\partial L^*}{\partial \hat{\boldsymbol{\beta}}^*} &= -2\mathbf{X}'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}^*) + \mathbf{2R}'\hat{\boldsymbol{\lambda}} = \mathbf{0}, \\
\frac{\partial L^*}{\partial \hat{\boldsymbol{\lambda}}} &= 2(\mathbf{R}\hat{\boldsymbol{\beta}}^* - \mathbf{q}) = \mathbf{0}. \quad (remember\ \ \frac{\partial \mathbf{a}'\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a})
\end{aligned}$$

Dividing through by 2 and expanding terms produces the partitioned matrix equation

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{R}' \\ \mathbf{R} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}^* \\ \hat{\boldsymbol{\lambda}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{q} \end{bmatrix},$$

or expressed by the simple notation that

$$\mathbf{W}\hat{\mathbf{d}}^* = \mathbf{v}.$$

Assuming that the partitioned matrix in brackets is nonsingular, then

$$\hat{\mathbf{d}}^* = \mathbf{W}^{-1}\mathbf{v}.$$

Using the partition inverse rule of

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}_{11}^{-1}(\mathbf{I} + \mathbf{A}_{12}\mathbf{F}_2\mathbf{A}_{21}\mathbf{A}_{11}^{-1}) & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{F}_2 \\ -\mathbf{F}_2\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{F}_2 \end{bmatrix},$$

where $\mathbf{F}_2 = (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}$, we have the restricted least squared estimator

$$\hat{\boldsymbol{\beta}}^* = \hat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q}), \tag{6-5}$$

and

$$\hat{\boldsymbol{\lambda}} = [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q}).$$

Exercise 4.

Show that $Var(\hat{\boldsymbol{\beta}}^*) - Var(\hat{\boldsymbol{\beta}})$ is a nonpositive definite matrix. ■

The above result of exercise holds whether or not the restriction are true. One way to interpret this reduction in variance is as the value of the information contained in the restriction.

Let

$$\mathbf{e}^* = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}^*, \tag{6-6}$$

i.e., the residuals vector from the restricted least square estimator, then using the familiar device,

$$\mathbf{e}^* = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}(\hat{\boldsymbol{\beta}}^* - \hat{\boldsymbol{\beta}}) = \mathbf{e} - \mathbf{X}(\hat{\boldsymbol{\beta}}^* - \hat{\boldsymbol{\beta}}),$$

the "restricted" sums of squared residuals is

$$\mathbf{e}'^*\mathbf{e}^* = \mathbf{e}'\mathbf{e} + (\hat{\boldsymbol{\beta}}^* - \hat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}}^* - \hat{\boldsymbol{\beta}}). \tag{6-7}$$

Because $\mathbf{X}'\mathbf{X}$ is a positive definite matrix,

$$\mathbf{e}'^*\mathbf{e}^* \geq \mathbf{e}'\mathbf{e}. \tag{6-8}$$

2.6 Measurement of Goodness of Fit

Denote the dependent variable's "fitted value" from independent variables and *OLS* estimator, $\hat{\mathbf{y}}$, to be $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$, that is

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{e}.$$

Writing $\mathbf{y}$ as its fitted values, plus its residual, provided another way to interpret an *OLS* regression.

Lemma.
$\mathbf{y}'\mathbf{y} = \hat{\mathbf{y}}'\hat{\mathbf{y}} + \mathbf{e}'\mathbf{e}$.

Proof.
Using the fact that $\mathbf{X}'\mathbf{y} = \mathbf{X}'(\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{e}) = \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}$, we have

$$\begin{aligned} \mathbf{e}'\mathbf{e} &= \mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \\ &= \mathbf{y}'\mathbf{y} - \hat{\mathbf{y}}'\hat{\mathbf{y}}. \end{aligned}$$

■

There are three measurements of variation which are defined as following:
(a). *SST* (Sums of Squared Total variation)= $\sum_{t=1}^{T}(Y_t - \bar{Y})^2 = \mathbf{y}'\mathbf{M_0}\mathbf{y}$,
(b). *SSR* (Sums of Squared Regression variation)=$\sum_{t=1}^{T}(\hat{Y}_t - \bar{\hat{Y}})^2 = \hat{\mathbf{y}}'\mathbf{M}_0\hat{\mathbf{y}}$,
(c). *SSE* (Sums of Squared Error variation)=$\sum_{t=1}^{T}(Y_t - \hat{Y}_t)^2 = \mathbf{e}'\mathbf{e}$,
where $\bar{Y} = \frac{1}{T}\sum_{t=1}^{T} Y_t$ and $\bar{\hat{Y}} = \frac{1}{T}\sum_{t=1}^{T} \hat{Y}_t$.

Lemma.
If one of the regressor is a constant, then $\bar{Y} = \bar{\hat{Y}}$.

Proof.
Writing $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{e}$, and using the fact that $\mathbf{i}'\mathbf{e} = 0$ we obtain the results, where $\mathbf{i}$ is a column of 1s. ■

Lemma.
If one of the regressor is a constant, then $SST = SSR + SSE$.

Proof.
Multiplying $\mathbf{M}_0$ (or call it as $\mathbf{M}_i$) on $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{e}$ we have

$$\mathbf{M}_0\mathbf{y} = \mathbf{M}_0\hat{\mathbf{y}} + \mathbf{M}_0\mathbf{e} = \mathbf{M}_0\hat{\mathbf{y}} + \mathbf{e}, \quad since \;\; \mathbf{M}_0\mathbf{e} = \mathbf{e} \; (why \; ?).$$

Therefore,

$$\begin{aligned} \mathbf{y}'\mathbf{M}_0\mathbf{y} &= \hat{\mathbf{y}}'\mathbf{M}_0'\hat{\mathbf{y}} + 2\hat{\mathbf{y}}'\mathbf{M}_0'\mathbf{e} + \mathbf{e}'\mathbf{e} \\ &= \hat{\mathbf{y}}'\mathbf{M}_0'\hat{\mathbf{y}} + \mathbf{e}'\mathbf{e} \\ &= SSR + SSE, \end{aligned}$$

using the fact that $\hat{\mathbf{y}}'\mathbf{M}_0'\mathbf{e} = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{e} = 0$. ■

Definition. (R^2)
If one of the regressor is a constant, the coefficient of determination is defined as

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}.$$ ■

One kind of restriction is of the "exclusion restrictions" form that

$$\mathbf{R}\boldsymbol{\beta} = 0,$$

where for example $\mathbf{R} = (0, 1, 0, ..., 0)$. It is to say that a particular explanatory variable has no partial effect on the dependent variable or we may think it as a model with fewer regressors (without X_2, but with the same dependent variable). Because $\mathbf{e}'^*\mathbf{e}^* \geq \mathbf{e}'\mathbf{e}$ from Eq. (6-8), it is apparent that the coefficient of determination from this restricted model, say R^{2*} is smaller. (Thus the R^2 in the longer regression cannot be smaller.) It is tempting to exploit this result by just adding variables to the model; R^2 will continue to rise to its limit. In view of this result, we sometimes report an adjusted R^2, which is computed as follow.

Definition. (Adjusted R^2)

$$\bar{R}^2 = 1 - \frac{\mathbf{e}'\mathbf{e}/(T-k)}{\mathbf{y}'\mathbf{M}_0\mathbf{y}/(T-1)}.$$

Exercise 5.

Reproduce the results of R^2 in Table 3.4 (p.35) of Greene 6th edition. ■

3 Statistical Properties of OLS

We now investigate the statistical properties of the estimator of parameters, $\hat{\boldsymbol{\beta}}$ and s^2 from OLS.

3.1 Finite Sample Properties

3.1.1 Unbiasedness

Based on the six classical assumptions, the expected value of $\hat{\boldsymbol{\beta}}$ and s^2 are

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= E[(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y})] = E[(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}))] = E[\boldsymbol{\beta}+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}] \\ &= \boldsymbol{\beta}+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{\varepsilon}) = \boldsymbol{\beta}, \qquad (\textit{using Assumption } (b) \textit{ and } (c).) \end{aligned}$$

and by construction

$$E(s^2) = \frac{E(\mathbf{e}'\mathbf{e})}{T-k} = \frac{(T-k)\sigma^2}{T-k} = \sigma^2.$$

Therefore both $\hat{\boldsymbol{\beta}}$ and s^2 are unbiased estimators.

3.1.2 Efficiency

To investigate the efficiency of these two estimators, we first show their variance-covariance. The variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ is

$$\begin{aligned} Var(\hat{\boldsymbol{\beta}}) &= E[(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'] \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}']\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\sigma^2\mathbf{I})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}. \qquad (\textit{using Assumption } (1b) \textit{ and } (1d).) \end{aligned}$$

With Assumption (1f) and using properties of idempotent quadratic form, we have

$$\frac{(T-k)s^2}{\sigma^2} = \frac{\mathbf{e}'\mathbf{e}}{\sigma^2} = \frac{\boldsymbol{\varepsilon}'\mathbf{M}_{\mathbf{X}}\boldsymbol{\varepsilon}}{\sigma^2} \sim \chi^2_{(T-k)}, \tag{6-9}$$

that is

$$Var\left(\frac{\mathbf{e}'\mathbf{e}}{\sigma^2}\right) = 2(T-k)$$

or $Var(\mathbf{e}'\mathbf{e}) = 2(T-k)\sigma^4$. The variance of s^2 $(=\frac{\mathbf{e}'\mathbf{e}}{T-k})$ is therefore

$$Var(s^2) = Var\left(\frac{\mathbf{e}'\mathbf{e}}{T-k}\right) = \frac{2\sigma^4}{T-k}.$$

It is time to show the efficiency of OLS estimators.

Theorem. (Gauss-Markov)
The OLS estimator $\hat{\boldsymbol{\beta}}$ is the best linear unbiased estimator (**BLUE**) of $\boldsymbol{\beta}$.

Proof.
Consider any estimator linear in $\mathbf{y}$, say $\tilde{\boldsymbol{\beta}} = \mathbf{C}\mathbf{y}$. Let $\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{D}$. Then

$$\begin{aligned} E(\tilde{\boldsymbol{\beta}}) &= E[((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{D})(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon})] \\ &= \boldsymbol{\beta} + \mathbf{D}\mathbf{X}\boldsymbol{\beta}, \end{aligned}$$

so that for $\tilde{\boldsymbol{\beta}}$ to be unbiased we require $\mathbf{D}\mathbf{X} = \mathbf{0}$. Then the covariance matrix of $\tilde{\boldsymbol{\beta}}$ is

$$\begin{aligned} E[(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'] &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{D}]\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} + \mathbf{D}'] \\ &= \sigma^2[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} + \mathbf{D}\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &\quad + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{I}\mathbf{D}' + \mathbf{D}\mathbf{I}\mathbf{D}'] \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} + \sigma^2\mathbf{D}\mathbf{D}', \quad \textit{since } \mathbf{D}\mathbf{X} = \mathbf{0}. \end{aligned}$$

Since $\mathbf{D}\mathbf{D}'$ is a positive semidefinite matrix (see Ch1, p.40), which shows that the covariance matrix of $\tilde{\boldsymbol{\beta}}$ equals the covariance matrix of $\hat{\boldsymbol{\beta}}$ plus a positive semidefinite matrix. Hence $\hat{\boldsymbol{\beta}}$ is efficient relative to any other linear unbiased estimator of $\boldsymbol{\beta}$. ■

In fact we can go a step further in the discussion of the efficiency of OLS estimators even it is compared with any other estimator both linear and nonlinear.

Theorem. (Cramér-Rao Bounds of OLS Estimators)
Let the linear regression $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ satisfy classical assumptions, then the Cramér-Rao lower bounds for the unbiased estimator of $\boldsymbol{\beta}$ and σ^2 are $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ and $2\sigma^4/T$,

respectively.

Proof.
The log-likelihood is

$$\ln L(\boldsymbol{\beta}, \sigma^2; \mathbf{y}) = -\frac{T}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}).$$

Therefore,

$$\begin{aligned}
\frac{\partial L}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma^2}(\mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\boldsymbol{\beta}) = \frac{1}{\sigma^2}\mathbf{X}'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}),\\
\frac{\partial L}{\partial \sigma^2} &= \frac{T}{2\sigma^2} + \frac{1}{2\sigma^4}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}),\\
\frac{\partial^2 L}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} &= -\frac{1}{\sigma^2}\mathbf{X}'\mathbf{X}; \quad -E\left[\frac{\partial^2 L}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}\right] = \frac{\mathbf{X}'\mathbf{X}}{\sigma^2};\\
\frac{\partial^2 L}{\partial (\sigma^2)^2} &= \frac{T}{2\sigma^4} - \frac{1}{\sigma^6}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}); \quad -E\left[\frac{\partial^2 L}{\partial (\sigma^2)^2}\right] = \frac{T}{2\sigma^4}\ (how\ ?);\\
\frac{\partial^2 L}{\partial \boldsymbol{\beta}\partial \sigma^2} &= -\frac{1}{\sigma^4}\mathbf{X}'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}); \quad -E\left[\frac{\partial^2 L}{\partial \boldsymbol{\beta}\partial \sigma^2}\right] = \mathbf{0}.
\end{aligned}$$

Hence, the information matrix is

$$\mathbf{I}_T(\boldsymbol{\beta}, \sigma^2) = \begin{bmatrix} \frac{\mathbf{X}'\mathbf{X}}{\sigma^2} & \mathbf{0} \\ \mathbf{0} & \frac{T}{2\sigma^4} \end{bmatrix},$$

in turn, the Cramér-Rao lower bounds for the unbiased estimator of $\boldsymbol{\beta}$ and σ^2 are $\sigma^2(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}$ and $2\sigma^4/T$, respectively. ■

From above theorem, the OLS $\hat{\boldsymbol{\beta}}$ is an absolutely efficient estimator while s^2 is not. However, it can be shown that s^2 is indeed minimum variance unbiased efficient through the alternative approach of complete, sufficient statistics. See for example, Schmidt (1976), p.14.

3.1.3 Distribution (Exact) of $\hat{\boldsymbol{\beta}}$ and s^2

We now investigate the finite sample distribution of the OLS estimators.

Theorem. (Finite Sample's Distribution of $\hat{\boldsymbol{\beta}}$)
$\hat{\boldsymbol{\beta}}$ has a multivariate normal distribution with mean $\boldsymbol{\beta}$ and covariance matrix $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.

Proof.
By Assumption (1c),(1d) and (1f), we know that $\boldsymbol{\varepsilon} \sim N(0, \sigma^2\mathbf{I})$. Therefore by the results from linear function of a normal vector (Ch 2, p.62) we have[10]

$$\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} \sim N(\boldsymbol{\beta}, (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}),$$

or

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}). \qquad \blacksquare$$

Theorem. (Finite Sample's Distribution of s^2)
s^2 is distributed as a χ^2 distribution multiplied by a constant,

$$s^2 \sim \frac{\sigma^2}{(T-k)} \cdot \chi^2_{T-k}.$$

Proof.
As we have shown that $\frac{\mathbf{e}'\mathbf{e}}{\sigma^2} \sim \chi^2_{(T-k)}$, this result follows immediately. $\blacksquare$

Theorem. (Independence of $\hat{\boldsymbol{\beta}}$ and s^2)
$\hat{\boldsymbol{\beta}}$ and s^2 are independent.

Proof.
$s^2 = \boldsymbol{\varepsilon}'\mathbf{M}_\mathbf{X}\boldsymbol{\varepsilon}/(T-k)$ and $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$. Since $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}_\mathbf{X} = 0$, it implies that $\hat{\boldsymbol{\beta}}$ and s^2 are independent from the results on p.65 of Ch.2. $\blacksquare$

[10]Here, it should be noted that $(\mathbf{X}'\mathbf{X}) = \sum \mathbf{x}_t\mathbf{x}_t'$ is a infinite series, it is quite possible that $(\mathbf{X}'\mathbf{X}) \to \infty$, and so $(\mathbf{X}'\mathbf{X})^{-1} \to 0$ (indeed this is he requirement for the consistency of $\hat{\boldsymbol{\beta}}$.). So the distribution here is only true for a finite sample. For a infinity sample, the distribution of $\hat{\boldsymbol{\beta}}$ degenerate to a point.

3.2 Asymptotic Properties

We now investigate the properties of the OLS estimators when the sample goes to infinity $T \to \infty$.

3.2.1 Consistency

Theorem.
The OLS estimator $\hat{\boldsymbol{\beta}}$ is consistent.

Proof.
Denote $\lim_{T\to\infty}(\mathbf{X}'\mathbf{X}/T) = \lim_{T\to\infty}(\sum_{t=1}^{T}\mathbf{x}_t'\mathbf{x}_t)/T$ by $\underline{\mathbf{Q}}$ and assume that $\underline{\mathbf{Q}}$ is finite and nonsingular.[11] (What does it mean ?) Then $\lim_{T\to\infty}(\mathbf{X}'\mathbf{X}/T)^{-1}$ is also finite. Therefore for large sample, the variance of $\hat{\boldsymbol{\beta}}$

$$\begin{aligned}\lim_{T\to\infty}\sigma^2(\mathbf{X}'\mathbf{X})^{-1} &= \lim_{T\to\infty}\frac{\sigma^2}{T}\left(\frac{\mathbf{X}'\mathbf{X}}{T}\right)^{-1}\\ &= \lim_{T\to\infty}\frac{\sigma^2}{T}\underline{\mathbf{Q}}^{-1}\\ &= \mathbf{0}.\end{aligned}$$

Since $\hat{\boldsymbol{\beta}}$ is unbiased and its covariance matrix, $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, vanishes asymptotically, it converges in mean-squared error to $\boldsymbol{\beta}$, which implies $\hat{\boldsymbol{\beta}}$ converges in probability to $\boldsymbol{\beta}$. Therefore, $\hat{\boldsymbol{\beta}}$ is consistent.

Alternative Proof.
Note that

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} = \boldsymbol{\beta} + \left(\frac{\mathbf{X}'\mathbf{X}}{T}\right)^{-1}\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{T}.$$

Since $E(\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{T}) = \mathbf{0}$. Also $Var(\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{T}) = E(\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{T})(\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{T})' = \frac{\sigma^2}{T}(\frac{\mathbf{X}'\mathbf{X}}{T})$, so that

$$\lim_{T\to\infty} Var\left(\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{T}\right) = \lim_{T\to\infty}\frac{\sigma^2}{T}\underline{\mathbf{Q}} = \mathbf{0}.$$

But the fact that $E(\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{T}) = \mathbf{0}$ and $\lim_{T\to\infty} Var(\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{T}) = \mathbf{0}$ imply that $plim\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{T} = \mathbf{0}$. Therefore

$$plim\ \hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \underline{\mathbf{Q}}^{-1}plim\ \frac{\mathbf{X}'\boldsymbol{\varepsilon}}{T} = \boldsymbol{\beta}. \qquad \blacksquare$$

[11] That is, we say that $\mathbf{X}'\mathbf{X}$ is $O(T)$.

Theorem.
The OLS estimator s^2 is consistent.

Proof.
Since $E(s^2) = \sigma^2$ and $\lim_{T\to\infty} Var(s^2) = \lim_{T\to\infty} \frac{2\sigma^4}{T-k} = 0$, the result is trivial. ■

Example.
The following is the code to generate the linear regression

$$Y_t = 2 + 3X_t + \varepsilon_t,\ t = 1, 2, ..., T,$$

where $X_t \sim i.i.d.\ N(3, 2)$ and $\varepsilon_t \sim i.i.d. N(0, 1)$.
It can see that the OLS is unbiased and consistent from a repeat of 1000 trials.
(a). Plot $T = 100$ for unbiasedness.
(b). Plot, $T = 30, 50, 100, 200, 500$ for consistency. ■

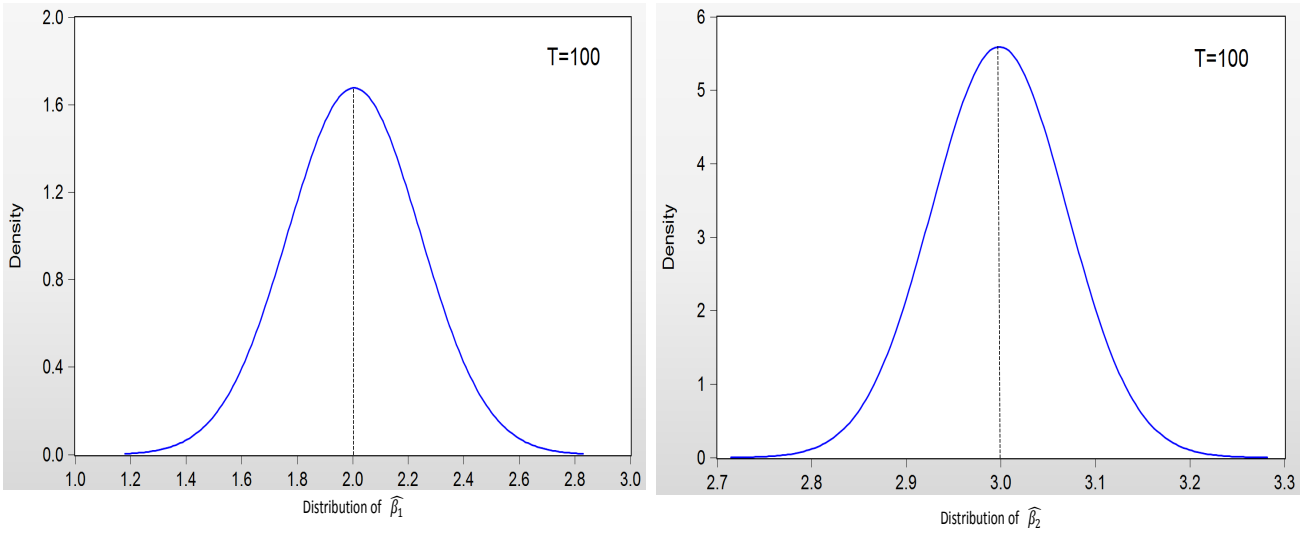

Figure (6-4a). Unbiasedness of $\hat{\beta}_1$ and $\hat{\beta}_2$ when regressors **x** are exogenous.

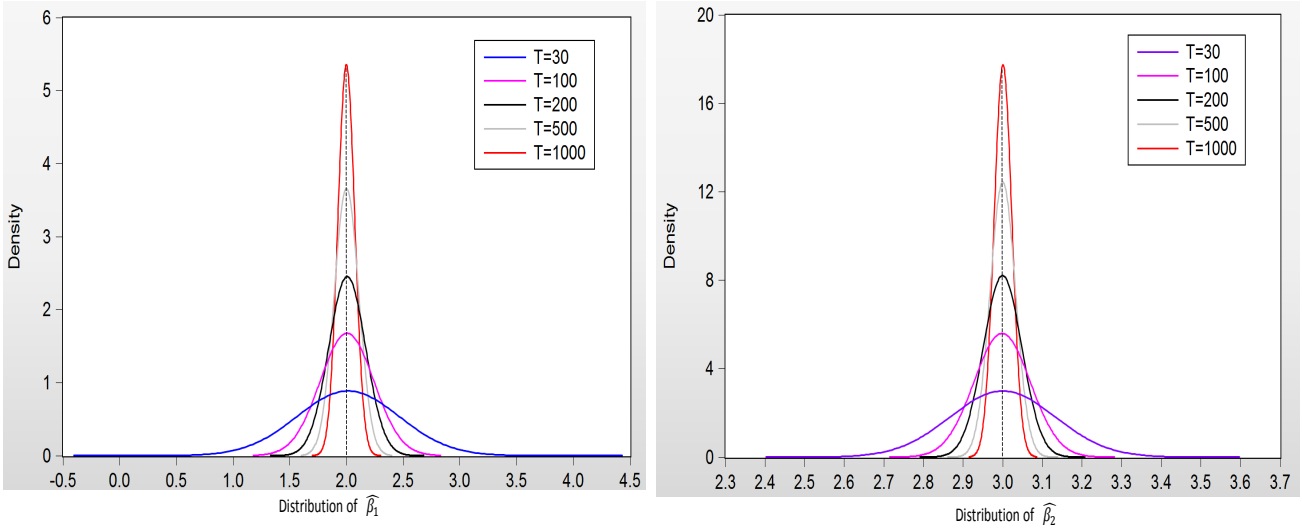

Figure (6-4b). Consistency of $\hat{\beta}_1$ and $\hat{\beta}_2$ when **x** are exogenous.

The Gauss Code for the generating the figures above is:

```
new; /*Open the operation of Gauss program*/
for T(100,500,100); /*setting number of periods T are 100, 200,300,400,500*/
a1 = ones(T,1); /*setting a1 is a matrix of T × 1 which all of elements
are equal to one */
a_hat1 = zeros(1000,1); /*setting a_hat1 is a zero matrix of 1000 × 1*/
b_hat11 = zeros(1000,1); /*setting b_hat11 is a zero matrix of 1000 × 1*/
yt1 = zeros(T,1);/*setting yt1 is a matrix of T × 1 which all of elements
are equal to zero*/
for j(1,1000,1); /*Loop 1000 times*/
xt1 = 3 + (2^0.5) * rndn(T,1); /* setting xt1 is a N(3,2) matrix of T × 1*/
e1 = rndn(T,1);
for k(1,T,1); /* Loop T times*/
yt1[k] = 2 + 3 * xt1[k] + e1[k]; /* setting model of yt1*/
endfor; /* end of loop*/
x1 = a1~xt1[1:T,1]; /* setting x1 is a matrix which a1 and xt1 array horizontally*/
b_hat1 = (inv(x1' * x1)) * x1' * yt1[1 : T,1]; /* setting b_hat1 is coefficient of
estimate*/
a_hat1[j] = b_hat1[1,1]; /* List all of alpha */
b_hat11[j] = b_hat1[2,1]; /* List all of beta */
endfor; /* end of loop*/
print "-----------------alpha---------------";
print a_hat1; /* calculate alpha */
print "-----------------beta----------------";
print b_hat11; /* calculate beta*/
endfor; /* end of loop */
```

■

3.2.2 Asymptotically Normality

Since by assumption, $\mathbf{X}'\mathbf{X}$ is $O(T)$, therefore $(\mathbf{X}'\mathbf{X})^{-\mathbf{1}} \to \mathbf{0}$. The exact distribution of $\hat{\boldsymbol{\beta}}$, i.e., $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-\mathbf{1}})$ will degenerate to a point in large sample. To express the limiting distribution of $\hat{\boldsymbol{\beta}}$, we need the following theorem.

Theorem. (Limiting Distribution of $\hat{\boldsymbol{\beta}}$)
The asymptotic distribution of $\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is $N(\mathbf{0}, \sigma^2\underline{\mathbf{Q}}^{-1})$, where $\underline{\mathbf{Q}} = \lim_{T\to\infty}(\frac{\mathbf{X}'\mathbf{X}}{T})$.

Proof.
For any sample size T, the distribution of $\sqrt{T}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is $N\left(\mathbf{0}, \sigma^2\left(\frac{\mathbf{X}'\mathbf{X}}{T}\right)^{-1}\right)$. The above limiting results is therefore trivial. ■

Theorem. (Limiting Distribution of s^2)
The asymptotic distribution of $\sqrt{T}(s^2 - \sigma^2)$ is $N(0, 2\sigma^4)$.

Proof.
Since the distribution of $\mathbf{e}'\mathbf{e}/\sigma^2$ is χ^2 with $(T-k)$ degree of freedom. Therefore

$$\frac{\mathbf{e}'\mathbf{e}}{\sigma^2} = \sum_{t=1}^{T-k} v_t^2,$$

where the v_t^2 are *i.i.d.* χ^2 with one degree of freedom. But this is a sum of *i.i.d.* with mean 1 and variance 2. According to the Lindberg-Levy central limit theorem it follows that

$$\frac{1}{\sqrt{T-k}}\sum_{t=1}^{T-k}\left(\frac{v_t^2 - 1}{\sqrt{2}}\right) \xrightarrow{L} N(0,1).$$

But this is equivalent to saying that

$$\frac{1}{\sqrt{T-k}}\left(\frac{\mathbf{e}'\mathbf{e}}{\sigma^2} - (T-k)\right) \xrightarrow{L} N(0,2),$$

or that

$$\sqrt{T-k}(s^2 - \sigma^2) \xrightarrow{L} N(0, 2\sigma^4),$$

or that

$$\sqrt{T}(s^2 - \sigma^2) \xrightarrow{L} N(0, 2\sigma^4).$$ ■

From above results, we find that although variance of s^2 does not attains Cramér-Rao lower bound in finite sample, however it does in large sample.

Theorem.
s^2 is asymptotically efficient.

Proof.
The asymptotic variance of s^2 is $2\sigma^4/T$, which equals the Cramér-Rao lower bound.■

4 Hypothesis Testing in Finite Sample

4.1 Tests of a Single Linear Combination on $\boldsymbol{\beta}$: Tests based on t Distribution

This section covers the very important topic of testing hypothesis about any single linear restriction in the population regression function.

Lemma.

Let R be a $1 \times k$ vector, and define s^* by

$$s^* = \sqrt{s^2 R(\mathbf{X}'\mathbf{X})^{-1}R'},$$

then $R(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})/s^*$ has a t distribution with $(T-k)$ degrees of freedom.

Proof.

As we have seen in an early result that in finite sample,

$$\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}).$$

Clearly $R(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})$ is a scalar random variable with zero mean and variance $\sigma^2 R(\mathbf{X}'\mathbf{X})^{-1}R'$; call this variance σ^{2*}. Then $\frac{R(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{\sigma^*} \sim N(0,1)$, but this test statistics is not a *pivot* sine it contains the unknown parameter σ. We need some transformation of this statistics to remove the parameter.

We know that in Eq. (6-9), $(T-k)s^2/\sigma^2 \sim \chi^2_{T-k}$, therefore,

$$\frac{s^{2*}}{\sigma^{2*}} = \frac{s^2}{\sigma^2} \sim \chi^2_{T-k}/(T-k).$$

Finally, then,

$$\frac{R(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{s^*} = \frac{R(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\sigma^*}{\sqrt{s^{2*}/\sigma^{2*}}} = \frac{N(0,1)}{\sqrt{\chi^2_{T-k}/(T-k)}} \sim t_{T-k}. \tag{6-10}$$

The above results is established upon the numerator and denominator being independent. This condition is shown to be true at section 3.1.3. of this Chapter. ■

Theorem. (Test of a Single Linear Combination on $\boldsymbol{\beta}$)
Let R be a known $1\times k$ vector, and r be a known scalar. Then under the null hypothesis that $H_0:\ R\boldsymbol{\beta}=r$, the test statistics

$$\frac{R\hat{\boldsymbol{\beta}}-r}{s^*}\sim t_{T-k}.$$

Proof.
Under the null hypothesis,

$$\frac{R\hat{\boldsymbol{\beta}}-r}{s^*}=\frac{R\hat{\boldsymbol{\beta}}-R\boldsymbol{\beta}}{s^*}\sim t_{T-k}.$$ ■

Corollary. (Test of Significance of β_i)
Let β_i be the i-th elements of $\boldsymbol{\beta}$, and denote

$$s_{\hat{\beta}_i}=\sqrt{s^2(\mathbf{X}'\mathbf{X})^{-1}_{ii}},$$

which is called the "*standard error of the coefficients estimated*", then under the null hypothesis that $H_0:\ \beta_i=0$, the test statistics

$$t\text{-ratio}=\frac{\hat{\beta}_i}{s_{\hat{\beta}_i}}\sim t_{T-k}.$$

Proof.
This is a special case of last Theorem, with $r=0$ and R being a vector of zeros except for a one in the i-th position. ■

Exercise 6.
Reproduce the results in Table 4.2 (p.54) of Greene 6th edition. ■

4.2 Tests of Several Linear Restrictions on $\boldsymbol{\beta}$: Tests based on F Distribution

Frequently, we wish to test multiple hypotheses about the underlying parameters $\boldsymbol{\beta}$. We now consider a set of J linear restrictions of the form

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{q},$$

against the alternative hypothesis,

$$H_1 : \mathbf{R}\boldsymbol{\beta} \neq \mathbf{q}.$$

Each row of $\mathbf{R}$ is the coefficients in a linear restriction on the coefficient vector. Hypothesis testing of the sort suggested in the preceding section can be approached from two viewpoints. First, having computed a set of parameter estimates, we can ask whether the estimates come reasonably close to satisfying the restrictions implied by the hypothesis. An alternative approach might proceed as follows. Suppose that we impose the restrictions implied by the theory. Since unrestricted least squares is, by definition, "least squares," this imposition must lead to a loss of fit. We can then ascertain whether this loss of fit results merely from sampling error or whether it is so large as to cast doubt on the validity of the restrictions. The two approaches are equivalent.

4.2.1 Test of Multiple Linear Combination on $\boldsymbol{\beta}$ from Wald Test

We first consider the tests constructed from the unrestricted OLS estimators.

Theorem. (Test of Multiple Linear Combination on $\boldsymbol{\beta}$, Wald Test)
Let $\mathbf{R}$ be a known matrix of dimension $m \times k$ and rank m, $\mathbf{q}$ a known $m \times 1$ vector. Then under the null hypothesis that $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{q}$, the statistics

$$\begin{aligned} F\text{-ratio} &= \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})/m}{\mathbf{e}'\mathbf{e}/(T-k)} \\ &= \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})'[s^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})}{m} \sim F_{m,T-k}. \end{aligned}$$

Proof.
From the liner function of a normal vector, we have

$$\mathbf{R}\hat{\boldsymbol{\beta}} \sim N(\mathbf{R}\boldsymbol{\beta}, \sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}').$$

Further by the quadratic form of normal vector (Sec. 6.2.2 of Ch. 2) we have

$$(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta})'[\sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta}) \sim \chi^2_m.$$

Then under the null hypothesis that $\mathbf{R}\boldsymbol{\beta} = q$, the test statistics

$$(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})'[\sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q}) \sim \chi^2_m. \tag{6-11}$$

However, this test statistics in (6-11) is not a *pivot* sine it contains the unknown parameter σ^2. We need some transformation of this statistics to remove the parameter as in the single test.

Recall that $(T-k)s^2/\sigma^2 = \mathbf{e}'\mathbf{e}/\sigma^2 \sim \chi^2_{T-k}$, therefore the numerator and the denominator of the statistics in the following are trying to remove out the unknown parameter σ^2 from (6-11) such that

$$\begin{aligned} &\frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})'[\sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})/m}{(T-k)s^2/\sigma^2(T-k)} \\ &= \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})'[s^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})}{m} \end{aligned} \tag{6-12}$$

or

$$\begin{aligned} &\frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})'[\sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})/m}{\mathbf{e}'\mathbf{e}/\sigma^2(T-k)} \\ &= \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})/m}{\mathbf{e}'\mathbf{e}/(T-k)}, \end{aligned} \tag{6-13}$$

the numerator and denominator of (6-12) and (6-13) are distributed as χ^2_m/m and $\chi^2_{T-k}/(T-k)$, respectively. This statistics in (6-12) and (6-13) are distributed as a $F_{m,T-k}$ if this two χ^2 are independent. Indeed, it is the case as can be proven in the same line as the single test. ■

Clearly the appropriate rejection region is the upper tail of the F distribution. That is, rejection should be based on large deviation of $\mathbf{R}\hat{\boldsymbol{\beta}}$ from $\mathbf{q}$, and hence on large values of the test statistics.

Exercise 7.

Show the two χ^2 are independent in the last Theorem. ■

4.2.2 Test of Multiple Linear Combination on $\boldsymbol{\beta}$ from Loss of Fit

A different type of hypothesis test statistics focused on the fit of the regression. Recalling that $\hat{\boldsymbol{\beta}}^* = \hat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})$ and $\mathbf{e}'^*\mathbf{e}^* = \mathbf{e}'\mathbf{e} + (\hat{\boldsymbol{\beta}}^* - \hat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}}^* - \hat{\boldsymbol{\beta}})$, where $\hat{\boldsymbol{\beta}}^*$ and $\mathbf{e}^*$ are estimators and residuals from the restricted least squares errors. We find that

$$\begin{aligned}
&\mathbf{e}'^*\mathbf{e}^* - \mathbf{e}'\mathbf{e} = (\hat{\boldsymbol{\beta}}^* - \hat{\boldsymbol{\beta}})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}}^* - \hat{\boldsymbol{\beta}}) \\
&= (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q}) \\
&= (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q}).
\end{aligned}$$

Result.

Under the null hypothesis that $H_0: \mathbf{R}\boldsymbol{\beta} = q$ we will have the third F-ratio statistics from (6-13) that would also distributed as $F_{m,T-k}$, that is

$$\begin{aligned}
&\frac{(\mathbf{e}'^*\mathbf{e}^* - \mathbf{e}'\mathbf{e})/m}{\mathbf{e}'\mathbf{e}/(T-k)} \\
&= \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})/m}{\mathbf{e}'\mathbf{e}/(T-k)} \sim F_{m,T-k}.
\end{aligned} \tag{6-14}$$

Finally, by dividing the numerator and denominator of (6-14) by $\sum_t (Y_i - \bar{y})^2$, we obtain the fourth F-ratio statistics from (6-14)

$$\frac{(R^2 - R^{2*})/m}{(1-R^2)/(T-k)} = \frac{(\frac{\mathbf{e}'^*\mathbf{e}^*}{\mathbf{y}'\mathbf{M_0}\mathbf{y}} - \frac{\mathbf{e}'\mathbf{e}}{\mathbf{y}'\mathbf{M_0}\mathbf{y}})/m}{(\frac{\mathbf{e}'\mathbf{e}}{\mathbf{y}'\mathbf{M_0}\mathbf{y}})/(T-k)} \sim F_{m,T-k}, \tag{6-15}$$

where R^{2*} is the R-square under the restriction estimation. ■

A special set of exclusion restrictions is routinely tested by most regression packages. In a model with constant (i.e. $X_1 = 1$) and $k-1$ independent variables, we can write the null hypothesis as

$H_0: X_2, ..., X_k$ do not help to explain Y.

This null hypothesis is, in a way, very pessimistic. It states that *none* of the explanatory variables has an effect on Y. Stated in terms of the parameters, the null is that all slope parameters are zero:

$$H_0: \ \beta_2 = \beta_3 = \cdots = \beta_k = 0,$$

and the alternative is that at least one of the β_i is different from zero. This is a special case of Eq.(6-15) with $m = k - 1$) and $R^{2*} = 0$.

Corollary. (Test of the Significance of a Regression)
If all the slope's coefficients (except for constant term) are zero, then $\mathbf{R}$ is $(k-1) \times k$ ($m = k - 1$), $\mathbf{q} = \mathbf{0}$. Under this circumstance, $R^{2*} = 0$. The test statistics to test the significance of the regression that $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{0}$ is therefore from (6-15) that under the null hypothesis

$$\frac{R^2/(k-1)}{(1-R^2)/(T-k)} \sim F_{k-1,T-k}. \qquad \blacksquare$$

Exercise 8.

(a). Reproduce the results in Table 5.2 (p.91) of Greene 6th edition.

(b). Using the above four F-ratio statistics to compute the test statistics results $F_{3,21} = 1.768$ at the same page. $\blacksquare$

5 Prediction

In the context of a regression model, a *prediction* is an estimate of a future value of the dependent variable made *conditional* on the corresponding future values of the independent variables.

Let us consider a set of T_0 observations *not* included in the original sample of T observations. Specifically, Let $\mathbf{X}_0$ denote these T_0 observations on the regressors, and $\mathbf{y}_0$ these observations on $\mathbf{y}$. If the model obeys the classical assumptions, the *OLS* estimator $\hat{\boldsymbol{\beta}}_T$ is BLUE for $\boldsymbol{\beta}$. An obvious predictor for $\mathbf{y}_0$ is therefore

$$\hat{\mathbf{y}}_0 = \mathbf{X}_0\hat{\boldsymbol{\beta}}_T,$$

where $\hat{\boldsymbol{\beta}}_T = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is the OLS estimator based on the original T observations.

Definition. (Prediction Error).
The prediction error of $\mathbf{y}_0$ is defined by

$$\begin{aligned}\mathbf{v}_0 &= \mathbf{y}_0 - \mathbf{X}_0\hat{\boldsymbol{\beta}}_T \\ &= \mathbf{X}_0(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_T) + \boldsymbol{\varepsilon}_0.\end{aligned}$$ ■

Theorem. (Mean and Variance of Prediction Error)
$E(\mathbf{v}_0) = 0$ and $Var(\mathbf{v}_0) = \sigma^2(\mathbf{I} + \mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0')$.

Proof.

$$E(\mathbf{v}_0) = E(\mathbf{y}_0 - \mathbf{X}_0\hat{\boldsymbol{\beta}}) = E[\mathbf{X}_0(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_T) + \boldsymbol{\varepsilon}_0] = \mathbf{0},$$

and

$$\begin{aligned}Var(\mathbf{v}_0) &= E(\mathbf{v}_0\mathbf{v}_0') \\ &= E\{[\mathbf{X}_0(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \boldsymbol{\varepsilon}_0][\mathbf{X}_0(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \boldsymbol{\varepsilon}_0]'\} \\ &= E\{[\mathbf{X}_0((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}) + \boldsymbol{\varepsilon}_0][((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon})'\mathbf{X}_0' + \boldsymbol{\varepsilon}_0']\} \\ &= \sigma^2(\mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0') + \sigma^2\mathbf{I}_{T_0} \quad since\ E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}_0' = 0), \\ &= \sigma^2(\mathbf{I}_{T_0} + \mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0').\end{aligned}$$ ■

Corollary.
When $T_0 = 1$,

$$Var(v_0) = \sigma^2(1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0),$$

where $\mathbf{x}_0'$ is $1 \times k$.[12] ■

The prediction error variance can be estimated by using s^2 in place of σ^2.

Theorem.
Suppose that we wish to predict a single value of Y_0 ($T_0 = 1$) associated with a regressor $\mathbf{X}_{0_{(1\times k)}} = \mathbf{x}_0'$, then

$$\frac{v_0}{s^2(1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0)} \sim t_{T-k},$$

where $v_0 = Y_0 - \mathbf{x}_0'\hat{\boldsymbol{\beta}}$.

Proof.
Because $\mathbf{v}_0$ is a linear function of normal vector, $\mathbf{v}_0$ is also normally distributed,

$$\mathbf{v}_0 \sim N(\mathbf{0}, \sigma^2(\mathbf{I} + \mathbf{X}_0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_0'),$$

and

$$v_0 \sim N(0, \sigma^2(1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0).$$

Then as the proof in the t-ratio statistics we have

$$\frac{v_0/[\sigma^2(1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0)]}{(T-k)s^2/\sigma^2(T-k)} = \frac{v_0}{s^2(1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0)} \sim t_{T-k}.$$ ■

[12] Since by assumption that $\mathbf{X}'\mathbf{X}$ is $O(T)$, then $(\mathbf{X}'\mathbf{X})^{-1} \to \mathbf{0}$. The forecast error variance become progressive smaller as we accumulate more data.

Corollary.
The forecast interval for Y_0 would be formed using

$$forecast\ interval = \hat{Y}_0 \pm t_{\alpha/2} \cdot s^2(1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0). \quad \blacksquare$$

Example.
See Example 5.4 (p.99) of Greene 6th edition. ■

5.1 Measuring the Accuracy of Forecasts

Various measures have been proposed for assessing the predictive accuracy of forecast models. Two that are based on the residuals from the forecasts are the root mean squared error

$$RMSE = \sqrt{\frac{1}{T_0}\sum_i (Y_i - \hat{Y}_i)^2}$$

and the mean absolute error

$$MAE = \frac{1}{T_0}\sum_i |Y_i - \hat{Y}_i|,$$

where T_0 is the number of periods being forecasted.

It is needed to keep in mind that however the $RMSE$ and MAE are also random variables. To compare predictive accuracy we need a test statistics to test "*equality of forecast accuracy*". See for example Diebold and Mariano (1995, JBES, p. 253).

6 Tests of Structural Change

6.1 Chow Test

One of the more common applications of the F-ratio tests is in tests of structural change. In specifying a regression model, we assume that its assumptions apply to all the observations in our sample. It is straightforward, however, to test the hypothesis that some or all of the regression coefficients are different in different subsets of the data.

Theorem. (Different Parameter Vectors; Chow Test)
Suppose that one has T_1 observations on a regression equation

$$\mathbf{y}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1, \tag{6-16}$$

and T_2 observations on another regression equation

$$\mathbf{y}_2 = \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2. \tag{6-17}$$

Suppose that $\mathbf{X}_1$ and $\mathbf{X}_2$ are made up of k regressors. Let SSE_1 denotes the sum of squared errors in the regression of $\mathbf{y}_1$ on $\mathbf{X}_1$ and SSE_2 denotes the sum of squared errors in the regression of $\mathbf{y}_2$ on $\mathbf{X}_2$. Finally let the "joint regression" equation be

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix} \tag{6-18}$$

and SSE be the sum of squared errors in the joint regression.[13] Then under the null hypothesis that $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ and assume that $\boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix}$ is distributed as $N(0, \sigma^2 I_T)$, the statistics

$$\frac{(SSE - SSE_1 - SSE_2)/k}{(SSE_1 + SSE_2)/(T - 2k)}$$

is distributed as $F_{k,T-2k}$.

Proof.
The "separated regression" model in (6-16) and (6-17) can be written together as

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix} = \mathbf{X}\boldsymbol{\beta}^* + \boldsymbol{\varepsilon}. \tag{6-19}$$

[13] We think the model to be $Y_t = \mathbf{x}_t'\boldsymbol{\beta} + \varepsilon_t,\ t = 1, 2, ..., T_1, \underbrace{T_1 + 1, ..., T}_{T_2}$.

Formally, the joint regression model (6-18) can be regarded as a restriction $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ is $\mathbf{R}\boldsymbol{\beta}^* = \mathbf{q}$, where $\mathbf{R} = [\mathbf{I}_k \mid -\mathbf{I}_k]$ and $\mathbf{q} = \mathbf{0}$ on the "separated " model in (6-19). The general result given earlier can be applied directly.

The OLS of the separated model is therefore

$$\begin{bmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{bmatrix} = \hat{\boldsymbol{\beta}}^* = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{bmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2'\mathbf{X}_2' \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}_1' & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2' \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}$$
$$= \begin{bmatrix} (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}_1 \\ (\mathbf{X}_2'\mathbf{X}_2')^{-1}\mathbf{X}_2'\mathbf{y}_2 \end{bmatrix},$$

and the residual is

$$\mathbf{e} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} - \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1 - \mathbf{X}_1\hat{\boldsymbol{\beta}}_1 \\ \mathbf{y}_2 - \mathbf{X}_2\hat{\boldsymbol{\beta}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix}.$$

The sum of squared residual of the separate regression is $\mathbf{e}'\mathbf{e} = \mathbf{e}_1'\mathbf{e}_1 + \mathbf{e}_2'\mathbf{e}_2 = SSE_1 + SSE_2$, which is the sums of squared residuals from the addition of the "separated regression" and can be regarded as "errors from unrestricted model" relative to the joint regression (6-18). The results is apparent from (6-14). ■

Example.

See Example 7.6 (p.136) of Greene 5th edition. ■

6.2 Alternative Tests of Model Stability

The Chow test described in last section assumes that the process underlying the data is stable up to a known transition point, at which it makes a discrete change to a new, but rather thereafter stable, structure. However, the change to a new regime might be more gradual and less obvious. Brown, Durbin and Evans (1975) proposed a test for model stability based on recursive residuals.

Suppose that the sample contains a total of T observations. The tth recursive residual is the ex post prediction error for Y_t when the regression is estimated using only the first $t-1$ observation:

$$\mathfrak{e}_t = Y_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}}_{t-1},$$

where $\mathbf{x}_t$ is vector of regressors associated with observation Y_t and $\hat{\boldsymbol{\beta}}_{t-1}$ is the OLS computed using the first $t-1$ observation. The forecast error variance of this residual is

$$\sigma_{ft}^2 = \sigma^2(1 + \mathbf{x}_t'(\mathbf{X}_{t-1}'\mathbf{X}_{t-1})^{-1}\mathbf{x}_t),$$

Let the rth scaled residual be

$$w_r = \frac{\mathfrak{e}_r}{\sqrt{1 + \mathbf{x}_r'(\mathbf{X}_{r-1}'\mathbf{X}_{r-1})^{-1}\mathbf{x}_r}}.$$

Under the hypothesis that the coefficient ($\boldsymbol{\beta}$) remain constant during the full sample period, $w_r \sim N(0, \sigma^2)$ and is independent of w_s for $s \neq r$.[14] Evidence that the distribution of w_r is changing over time against the hypothesis of model stability. Brown et al. (1975) suggest two test based on w_r.

Theorem. (CUSUM Test)
The CUSUM test is based on the cumulated sums of the recursive residuals:

$$W_t = \sum_{r=k+1}^{r=t} \frac{w_r}{\hat{\sigma}},$$

where[15]

$$\hat{\sigma}^2 = \frac{\sum_{r=k+1}^{T}(w_r - \bar{w})^2}{T-k+1}, \quad \text{and} \quad \bar{w} = \frac{\sum_{r=k+1}^{T} w_r}{T-k}.$$

Under the null hypothesis, W_t has a mean zero and a variance of approximately the number of residuals being summed (because each term has variance 1 and they are independent. The test is performed by plotting W_t against t. ■

Theorem. (CUSUMSQ Test)
An alternative similar test is based on the squares of the recursive residuals. The CUSUM of squares (CUSUMSQ) test used

$$S_t = \frac{\sum_{r=k+1}^{t} w_r^2}{\sum_{r=k+1}^{T} w_r^2}.$$

[14] For detailed proof, please refers to p. 54 of Harvey's book (1990).

[15] It is because the variance of w is σ^2. Hence to estimate the variance of w, we use its sample moments.

Under the null hypothesis S_t has a beta distribution with a mean $(t-k)/(T-k)$. ■

Example.

See Figures 7.6 (p.138) of Greene 5th edition. ■

7 Mixed-Frequency Regression Model

Macroeconomic data is typically sampled at monthly or quarterly frequencies while financial time series are sampled at almost arbitrarily higher frequencies. Despite the fact that most economic time series are not sampled at the same frequency the typical practice of estimating econometric models involves aggregating all variables to the same (low) frequency using an equal weighting scheme. However, there is no a priori reason why one should (i) ignore the fact that the variables involved in empirical models are in fact generated from processes of different/mixed frequencies and (ii) estimate econometric models based on an aggregation scheme of equal weights. In fact one would expect that for most time series declining weights would be a more appropriate aggregation scheme and that an equal weighting scheme may lead to information loss and thereby to inefficient and/or biased estimates.

7.1 Unrestricted Mixed-Frequency Model

A mixed-frequency regression model is described by

$$\begin{aligned} Y_t &= \beta\left(\sum_{k=0}^{m-1}\pi_{k+1}X_{t-(k/m)}^{(m)}\right)+\varepsilon_t, \\ &= \beta\left(\pi_1 X_{t-(0/m)}^{(m)}+\pi_2 X_{t-(1/m)}^{(m)}+...+\pi_m X_{t-((m-1)/m)}^{(m)}\right)+\varepsilon_t;\quad t=1,2,...,T, \end{aligned} \tag{6-20}$$

where β is a scalar. The index t represent the low frequency and runs from 1 to T. The superscript (m) denotes that the series are observed at the higher frequency. For example, if the regressor are observed $m = 12$ months per year, then k/m lags k months from December of each year. Hence $Y_t = Y_{t-(0/m)}^{(m)}$ is the annul data observed at December.

The weights parameters π's assigned weight to each high-frequency regressors within the low-frequency period. For example, an annual average of each month (*simple average or flat sampling*) has weights of $1/m$ for each of these high-frequency regressors. End-of-period sampling (a special case of selective of skip sampling), provides another example, in which the first weight π_1 is a unit, while the remaining weights are zeros.

7.2 MIDAS and CO-MIDAS Model

As the sampling rate m, increase, equation (6-20) leads to parameter proliferation. For example, for data sampled at a daily frequency (working day) for use in a monthly model, then $m = 20$. To solve the problem of parameter proliferation while preserving the time information from the high-frequency data, Ghysels, Santa-Clare, and Valkanov (2004) propose using a parsimonious nonlinear specification MIDAS (**MI**xed **DA**ta **S**ampling) model:

$$\begin{aligned} Y_t &= \beta\left(\sum_{k=0}^{m-1}\pi_{k+1}(\boldsymbol{\gamma})X_{t-(k/m)}^{(m)}\right)+\eta_t, \\ &= \beta\left(\pi_1(\boldsymbol{\gamma})X_{t-(0/m)}^{(m)}+\pi_2(\boldsymbol{\gamma})X_{t-(1/m)}^{(m)}+...+\pi_m(\boldsymbol{\gamma})X_{t-((m-1)/m)}^{(m)}\right)+\eta_t, \end{aligned} \tag{6-21}$$

where the function $\pi_k(\boldsymbol{\gamma})$ is a polynomial that determines the weights for temporal aggregation. Ghysels et al. (2005) suggest an exponential Almon specification:

$$\pi_s(\boldsymbol{\gamma}) = \pi_s(\gamma_1,\gamma_2) = \frac{\exp(\gamma_1 s+\gamma_2 s^2)}{\sum_{j=1}^{m}\exp(\gamma_1 j+\gamma_2 j^2)}. \tag{6-22}$$

In this case, simple average is obtained when $\gamma_1 = \gamma_2 = 0$. Figure 7 show s various parameterizations of exponential Almon polynomial weighting function .

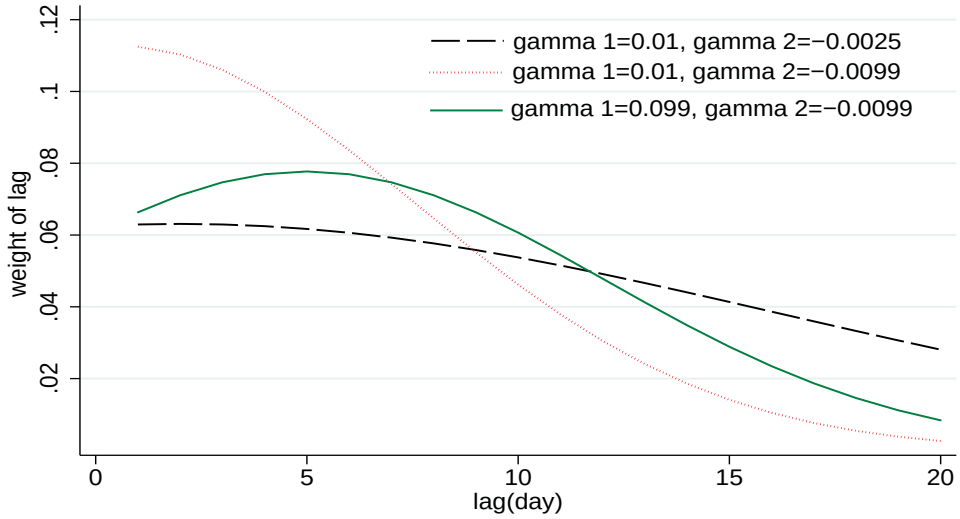

Figure 7. Exponential Almon Polynomial Weighting Function

If $Y_t \sim I(0)$ and $X^{(m)}_{t-(k/m)} \sim I(0), \forall k = 0, 1, ..., m-1$, and $\eta_t \sim I(0)$, it is the MIDAS model of Ghysels et al. (2004, 2006). While $Y_t \sim I(1)$ and $X^{(m)}_{t-(k/m)} \sim I(1), \forall k = 0, 1, ..., m-1$, and $\eta_t \sim I(0)$, Miller (2014) call (6-21) a cointegration MIDAS (CoMIDAS). That is $(Y_t, \boldsymbol{x}_t)$ are mixed frequency cointegrated with equilibrium error η_t.

7.2.1 Estimation of Parameters

In order to analyze the statistical properties of a nonlinear least square (NLS) estimator from the (Co)MIDAS regression in (6-21), the linear model in (6-20) may be written as very simply as

$$Y_t = \boldsymbol{\alpha}' \boldsymbol{x}_t + \varepsilon_t, t = 1, 2, ..., T, \tag{6-23}$$

where $\boldsymbol{\alpha} = [\beta\pi_1, \beta\pi_2, ..., \beta\pi_m]'$ and $\boldsymbol{x}_t = [X^{(m)}_{t-(0/m)}, X^{(m)}_{t-(1/m)}, ..., X^{(m)}_{t-((m-1)/m)}]'$. The (Co)MIDAS model therefore can also be rewritten as

$$Y_t = \boldsymbol{g}'(\boldsymbol{\theta}) \boldsymbol{x}_t + \eta_t, \tag{6-24}$$

where $\boldsymbol{g}(\boldsymbol{\theta}) = [\beta\pi_1(\boldsymbol{\gamma}), \beta\pi_2(\boldsymbol{\gamma}), ..., \beta\pi_m(\boldsymbol{\gamma})]'$ and $\boldsymbol{\theta} = [\beta, \gamma_1, \gamma_2]'$.

To estimate the parameters of the CoMIDAS model in (6-24), Miller (2014) consider the NLS estimator. The NLS objective function may be written as

$$Q_T(\boldsymbol{\theta}) = \frac{1}{2} \sum_t \left(\varepsilon_t - (\boldsymbol{g}(\boldsymbol{\theta}) - \boldsymbol{\alpha})' \boldsymbol{x}_t \right)^2,$$

and the NLS estimator is defined to be $\hat{\boldsymbol{\theta}}_{NLS} = \text{argmin}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} Q_T(\boldsymbol{\theta})$. Of course, numerical optimization is used to find $\hat{\boldsymbol{\theta}}_{NLS}$ in practice.

NSYSU from a Hill Overlooking.

End of this Chapter