1 September 2004

Statistical Inference in the Classical Linear Regression Model

A. Introduction

In this section, we will summarize the properties of estimators in the classical linear regression model previously developed, make additional distributional assumptions, and develop further properties associated with the added assumptions. Before presenting the results, it will be useful to summarize the structure of the model, and some of the algebraic and statistical results presented elsewhere.

B. Statement of the classical linear regression model

The classical linear regression model can be written in a variety of forms. Using summation notation we write it as

$y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \ldots + \varepsilon_t \ \forall\, t$	(linear model)	(1)
$E(\varepsilon_t \mid x_{t1}, x_{t2}, \ldots x_{tk}) = 0 \ \forall\, t$	(zero mean)	(2)
$\mathrm{Var}(\varepsilon_t \mid x_{t1},\ldots,x_{tk}) = \sigma^2 \ \forall t$	(homoskedasticity)	(3)
$E(\varepsilon_t\, \varepsilon_s) = 0 \ \ t \neq s$	(no autocorrelation)	(4)
x_{ti} is a known constant	(x's nonstochastic)	(5a)
No x_i is a linear combination of the other x's		(5b)
$\varepsilon_t \sim N(0,\sigma^2)$	(normality)	(6)

We can also write it in matrix notation as follows

$$
\begin{aligned}
&I \quad y = X\beta + \varepsilon \\
&II \quad E(\varepsilon \mid X) = 0 \\
&III \quad E(\varepsilon\varepsilon' \mid X) = \sigma^2 I \\
&IV \quad X \textit{ is a nonstochastic matrix of rank } k \\
&V \quad \varepsilon \sim N(0; \Sigma = \sigma^2 I)
\end{aligned} \tag{1}
$$

The ordinary least squares estimator of β in the model is given by

$$
\begin{aligned}
\hat{\beta} &= (X'X)^{-1}X'y \\
&= (X'X)^{-1}X'(X\beta + \epsilon) \\
&= \beta + (X'X)^{-1}X'\varepsilon
\end{aligned} \tag{2}
$$

The fitted value of y $(\hat{y})$ and the estimated vectors of residuals (e) in the model are defined by

$$
\begin{aligned}
\hat{y} &= X\hat{\beta} \\
e &= y - X\hat{\beta} \\
y &= X\hat{\beta} + e
\end{aligned} \tag{3}
$$

The variance of ϵ (σ^2) is usually estimated using the estimated residuals as

$$s^2 = \frac{\sum_{t=1}^{n} e_t^2}{n-k} = \frac{e'e}{n-k} \tag{4}$$

C. The fundamental matrices of linear regression

1. M - the residual creation matrix

The residuals from the least squares regression can be expressed as

$$\begin{aligned} e &= y - X\hat{\beta} \\ &= y - X(X'X)^{-1}X'y \\ &= (I - X(X'X)^{-1}X')y \\ &= M_X y \quad \textit{where} \\ M_X &= (I - X(X'X)^{-1}X') \end{aligned} \tag{5}$$

a. The matrix M_X is symmetric and idempotent.
b. $M_X X = 0$.
c. $e = M_X \epsilon$
d. $e'e = y'M_X y$.
e. $e'e = \epsilon' M_X \epsilon$.

2. P - the projection matrix

Consider a representation of the predicted value of y

$$\begin{aligned} \hat{y} &= y - e \\ &= y - My \\ &= (I - M)y \\ &= P_X y \quad \textit{where} \quad P_X = (I - M_X) \ \textit{is the projection matrix} \end{aligned} \tag{6}$$

a. P is symmetric and idempotent.
b. $P_X X = X$
c. $P_X M_X = 0$

3. A_n – The deviation transformation matrix

Consider the matrix A_n below which transforms a vector or matrix to deviations from the mean.

$$
\begin{aligned}
A_n &= I - \frac{1}{n}jj' \\
&= \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} - \frac{1}{n}\begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 - \frac{1}{n} & \frac{-1}{n} & \dots & \frac{-1}{n} \\ \frac{-1}{n} & 1 - \frac{1}{n} & \dots & \frac{-1}{n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{-1}{n} & \frac{-1}{n} & \dots & 1 - \frac{1}{n} \end{bmatrix}
\end{aligned} \tag{7}
$$

a. A_n is symmetric and idempotent.

b. $\sum_{t=1}^{n} (y_t - \bar{y})^2 = y' A_n y$

c. $A_n M_X = M_X$ (first column of X is a column of ones)

Proof:

First write A_n in a different fashion noting that the vector of ones we called j, is the same as the first column of the X matrix in a regression with a constant term.

$$
\begin{aligned}
A &= \begin{bmatrix} 1-\frac{1}{n} & \frac{-1}{n} & \cdots & \frac{-1}{n} \\ \frac{-1}{n} & 1-\frac{1}{n} & \cdots & \frac{-1}{n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{-1}{n} & \frac{-1}{n} & \cdots & 1-\frac{1}{n} \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} - \frac{1}{n} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} - \frac{1}{n} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \left(\begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix} \\
&= I - x_1 (x_1' x_1)^{-1} x_1'
\end{aligned}
\tag{8}
$$

Now consider the product of A_n and M_X

$$
\begin{aligned}
A_n M_X &= (I - x_1 (x_1' x_1)^{-1} x_1') M_X \\
&= M_X - x_1 (x_1' x_1)^{-1} x_1' M_X
\end{aligned}
\tag{9}
$$

From previous results, $M_X X = 0_{n \times k}$, which implies that $X` M_X = 0_{k \times n}$. This then implies that $x_1' M = 0_n'$. Given that this product is a row of zeroes, it is clear that the entire second term vanishes. This then implies

$$
A_n M_X = M_X \tag{10}
$$

D. Some results on traces of matrices

The trace of a square matrix is the sum of the diagonal elements and is denoted tr A or tr (A). We will state without proof some properties of the trace operator.

a. trace (I_n) = n
b. tr(kA) = k tr(A)
c. trace (A + B) = trace (A) + trace (B)

d. $tr(AB) = tr(BA)$ if both AB and BA are defined
e. trace (ABC) = trace (CAB) = trace (BCA)

The results in part e hold as along as the matrices involved are conformable, though the products may be different dimensions.

We will also use Theorem 17 from the lecture on characteristic roots and vectors. A proof of this theorem is given there.

Theorem 17: Let A be a square symmetric idempotent matrix of order n and rank r. Then the trace of A is equal to the rank of A, i.e., $tr(A) = r(A)$.

E. Some theorems on quadratic forms and normal variables (stated without proof)

1. Quadratic Form Theorem 1:

   If $y \sim N(\mu_y, \Sigma_y)$, then $z = Cy \sim N(\mu_z = C\mu_y; \Sigma_z = C\, \Sigma_y\, C')$ where C is a matrix of constants.

2. Quadratic Form Theorem 2

   Let the nx1 vector $y \sim N(0, I)$, then $y'y \sim \chi^2(n)$.

3. Quadratic Form Theorem 3: If $y \sim N(0,\sigma^2 I)$ and M is a symmetric idempotent matrix of rank m then

$$\frac{y'My}{\sigma^2} \sim \chi^2(trM) \tag{11}$$

   Corollary: If the nx1 vector $y \sim N(0,I)$ and the nxn matrix A is idempotent and of rank m, then $y'Ay \sim \chi^2(m)$.

4. Quadratic Form Theorem 4: If $y \sim N(0,\sigma^2 I)$, M is a symmetric idempotent matrix of order n, and L is a kxn matrix, then Ly and $y'My$ are independently distributed if $LM = 0$.

5. Quadratic Form Theorem 5: Let the nx1 vector $y \sim N(0,I)$, let A be an nxn idempotent matrix of rank m, let B be an nxn idempotent matrix of rank s, and suppose $BA = 0$. Then $y'Ay$ and $y'By$ are independently distributed χ^2 variables.

6. Quadratic Form Theorem 6 (Craig's Theorem) If $y \sim N(\mu, \Omega)$ where Ω is positive definite, then $q_1 = y'Ay$ and $q_2 = y'By$ are independently distributed iff $A\Omega B = 0$.

7. Quadratic Form Theorem 7 If y is a nx1 random variable and $y \sim N(\mu,\Sigma)$ then $(y - \mu)'\Sigma^{-1}(y - \mu) \sim \chi^2(n)$.

8. Quadratic Form Theorem 8: Let $y \sim N(0, I)$. Let M be a nonrandom idempotent matrix of dimension nxn ($rank(M)=r \leq n$). Let A be a nonrandom matrix such that $AM = 0$. Let $t_1 = My$ and let $t_2 = Ay$. Then t_1 and t_2 are independent random vectors.

F. Some finite sample properties of the ordinary least squares estimator in the classical linear regression model can be derived without specific assumptions about the exact distribution of the error term

1. Unbiasedness of $\hat{\beta}$

   Given the properties of the model, we can show that $\hat{\beta}$ is unbiased as follows if X is a nonstochastic matrix of full rank

$$
\begin{aligned}
\hat{\beta} &= (X'X)^{-1}X'y \\
&= (X'X)^{-1}X'(X\beta + \epsilon) \\
&= \beta + (X'X)^{-1}X'\varepsilon \\
E[\hat{\beta}] &= \beta + E[(X'X)^{-1}X'\varepsilon] \\
&= \beta + (X'X)^{-1}X' E(\varepsilon) \\
&= \beta
\end{aligned}
\tag{12}
$$

2. Variance of y

We know that y_t depends on the constants x_t and β, and on the stochastic error, ϵ_t. We write this as

$$y_t = x_t'\beta + \varepsilon_t \quad (t = 1,\ldots,n). \tag{13}$$

This implies that

$$Var(y_t) = Var(\varepsilon_t) = \sigma^2, \; (t = 1,\ldots,n). \tag{14}$$

Furthermore with $E(\varepsilon_t\ \varepsilon_s) = 0 \;\; t \neq s$, i.e., the covariance between y_t and y_{t+s} is zero, implying that

$$Var(y) = Var(\varepsilon) = \sigma^2 I \tag{15}$$

3. Variance of $\hat{\beta}$

We can determine the variance of $\hat{\beta}$ by writing it out and then using the information we have on the variance of y and the formula for the variance of any quadratic form.

$$
\begin{aligned}
\hat{\beta} &= (X'X)^{-1}X'y = Cy \\
Var(\hat{\beta}) &= (X'X)^{-1}X'\ Var(y)\, X(X'X)^{-1} \\
&= (X'X)^{-1}X'\ \sigma^2 I\, X(X'X)^{-1} \\
&= \sigma^2\, (X'X)^{-1}X'X(X'X)^{-1} \\
&= \sigma^2\, (X'X)^{-1} \\
&= \sigma^2\, CC'
\end{aligned}
\tag{16}
$$

4. $\hat{\beta}$ is the best linear unbiased estimator of β

We can show that $\hat{\beta}$ is the best linear unbiased estimator of β by showing that any other linear unbiased estimator has a variance which is larger that the variance of $\hat{\beta}$ by a positive definite matrix. The least squares estimator is given by

$$\hat{\beta} = (X'X)^{-1}X'y = Cy \tag{17}$$

Consider another linear unbiased estimator given by $\beta^* = Gy$. Linearity is imposed by the linear form of β^*. We can determine the restrictions on G for β^* to be unbiased by writing it out as follows.

$$\begin{aligned} y &= X\beta + \varepsilon \\ \beta^* &= Gy \\ E(\beta^*) &= E(Gy) = E(GX\beta + G\varepsilon) \\ &= E(GX\beta) + GE(\varepsilon) \\ &= E(GX\beta) \\ &\Rightarrow GX = I \quad \textit{if } \beta^* \textit{ is unbiased} \end{aligned} \tag{18}$$

The variance of β^* is similar to the variance of $\hat{\beta}$

$$Var(\beta^*) = \sigma^2 G G' \tag{19}$$

Now let D = G - C = G - $(X'X)^{-1}X'$, so that G = D +C. Now rewrite the variance of β^* as

$$\begin{aligned} Var(\beta^*) &= \sigma^2 GG' \\ &= \sigma^2[D+C]\,[D+C]' \\ &= \sigma^2[DD' + CD' + DC' + CC'] \\ &= \sigma^2[DD' + (X'X)^{-1}X'D' + DX(X'X)^{-1} + (X'X)^{-1}X'X(X'X)^{-1}] \\ &= \sigma^2[DD' + (X'X)^{-1}X'D' + DX(X'X)^{-1} + (X'X)^{-1}] \end{aligned} \tag{20}$$

Now substitute in equation 20 for $D = G - (X'X)^{-1}X'$ and $D' = G' - X(X'X)^{-1}$ noting that $GX = I_k$ and $X'G' = I_k$.

$$\begin{aligned} Var(\beta^*) &= \sigma^2[DD' + (X'X)^{-1}X'D' + DX(X'X)^{-1} + (X'X)^{-1}] \\ &= \sigma^2 DD' \\ &+ \sigma^2 (X'X)^{-1}X'\left[G' - X(X'X)^{-1}\right] \\ &+ \sigma^2\left[G - (X'X)^{-1}X'\right]X(X'X)^{-1} \\ &+ \sigma^2\,(X'X)^{-1} \\ &= \sigma^2 DD' + \sigma^2(X'X)^{-1} - \sigma^2(X'X)^{-1} + \sigma^2(X'X)^{-1} - \sigma^2(X'X)^{-1} + \sigma^2(X'X)^{-1} \\ &= \sigma^2 DD' + \sigma^2(X'X)^{-1} \\ &= \sigma^2 DD' + Var(\hat{\beta}) \end{aligned} \tag{21}$$

The variance of β^* is thus the variance of $\hat{\beta}$ plus a matrix that can be shown to be positive definite.

5. Unbiasedness of s^2

Given the properties of the model, we can show that s^2 is an unbiased estimator of σ^2. First write $e'e$ as a function of ϵ.

$$
\begin{aligned}
e &= M_X y \\
&= M_X(X\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\
&= M_X\boldsymbol{\varepsilon} \quad \textit{since} \quad MX = 0 \\
\Rightarrow e'e &= \boldsymbol{\varepsilon}' M_X' M_X \boldsymbol{\varepsilon} \\
&= \boldsymbol{\varepsilon}' M_X \boldsymbol{\varepsilon}
\end{aligned}
\tag{22}
$$

Now take the expected value of e′e, use the property of the trace operator that tr (ABC) = tr (BCA), and then simplify

$$
\begin{aligned}
e'e &= \boldsymbol{\varepsilon}' M \boldsymbol{\varepsilon} \\
E\left(\frac{e'e}{n-k}\right) &= E\left(\frac{\boldsymbol{\varepsilon}' M \boldsymbol{\varepsilon}}{n-k}\right) \\
&= E\left[\frac{tr(\boldsymbol{\varepsilon}' M \boldsymbol{\varepsilon})}{n-k}\right] \\
&= E\left[\frac{tr(M \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')}{n-k}\right] \\
&= \frac{tr(ME[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'])}{n-k} \\
&= \frac{tr(M\sigma^2 I)}{n-k} \\
&= \frac{\sigma^2 tr(M)}{n-k} \\
&= \frac{\sigma^2(n-k)}{n-k} = \sigma^2
\end{aligned}
\tag{23}
$$

We find the trace of M using the properties on sums, products, and identity matrices.

$$
\begin{aligned}
tr(M_X) &= tr(I - X(X'X)^{-1}X') \\
&= tr(I_n) - tr(X(X'X)^{-1}X') \\
&= tr(I_n) - tr(X'X(X'X)^{-1}) \\
&= tr(I_n) - tr(I_k) \\
&= n-k
\end{aligned}
\tag{24}
$$

6. Covariance of $\hat{\beta}$ and e

Given the properties of the model, we can show that the covariance of $\hat{\beta}$ and e is zero. First write both $\hat{\beta}$ and e as functions of ϵ from equations 2 and 5.

$$\begin{aligned} y &= X\beta + \varepsilon \\ \hat{\beta} &= \beta + (X'X)^{-1}X'\varepsilon \\ e &= M_X y \end{aligned} \tag{25}$$

Remember that $\hat{\beta}$ has an expected valued of β because it is unbiased. We can show that e has an expected valued of zero as follows

$$\begin{aligned} y &= X\beta + \varepsilon \\ M_X &= (I - X(X'X)^{-1}X') \\ e &= M_X y \\ &= M_X(X\beta + \varepsilon) \\ &= 0 + M_X\varepsilon, \quad M_X X = 0 \\ &= M_X\varepsilon \\ E(e) &= E(M_X\varepsilon) \\ &= M_X E(\epsilon) \\ &= 0 \end{aligned} \tag{26}$$

We then have

$$\begin{aligned} \hat{\beta} - E(\hat{\beta}) &= \hat{\beta} - \beta = (X'X)^{-1}X'\varepsilon \\ e - E(e) &= e \end{aligned} \tag{27}$$

Now compute the covariance directly

$$\begin{aligned} Cov(\beta hat,\ e) &= E[(\hat{\beta} - \beta)(e - 0)'] \\ &= E[(X'X)^{-1}X'\varepsilon\varepsilon'M'] \\ &= (X'X)^{-1}X'E(\varepsilon\varepsilon')M, \quad M \textit{ is symmetric} \\ &= (X'X)^{-1}X'\sigma^2 IM \\ &= \sigma^2(X'X)^{-1}X'M \\ &= \sigma^2(X'X)^{-1}X'(I - X(X'X)^{-1}X') \\ &= \sigma^2[(X'X)^{-1}X' - (X'X)^{-1}X'X(X'X)^{-1}X'] \\ &= \sigma^2[(X'X)^{-1}X' - (X'X)^{-1}X'] \\ &= \sigma^2[0] = 0 \end{aligned} \tag{28}$$

G. Distribution of $\hat{\beta}$ given normally distributed errors

1. introduction

Now make the assumption that $\varepsilon_t \sim N(0,\sigma^2)$ or $\boldsymbol{\varepsilon} \sim N(0;\Sigma = \sigma^2 I)$. Given that

$$\begin{aligned} & y = X\beta + \varepsilon \\ & X \text{ is a nonstochastic matrix of rank } k \\ & \varepsilon \sim N(0;\Sigma = \sigma^2 I) \end{aligned} \tag{29}$$

then y is also distributed normally because we are simply adding a constant vector to the random vector ϵ. The error vector ϵ is not transformed in forming y. Given $E(\epsilon) = 0$, $E(y) = X\beta$, and $Var(y) = \sigma^2 I$, we then have

$$y \sim N(X\beta;\Sigma = \sigma^2 I) \tag{30}$$

2. exact distribution of $\hat{\beta}$

We can write $\hat{\beta}$ as a linear function of the normal random variable y from equation 2 as follows

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'y \\ &= Cy, \quad C = (X'X)^{-1}X' \end{aligned} \tag{31}$$

We can find its distribution by applying Quadratic Form Theorem 1. From this theorem $E(\hat{\beta}) = CE(y)$ and $Var(\hat{\beta}) = C\,Var(y)\,C'$. Substituting we obtain

$$\begin{aligned} E(\hat{\beta}) &= (X'X)^{-1}X'E(y) \\ &= (X'X)^{-1}X'X\beta \\ &= \beta \\ Var(\hat{\beta}) &= (X'X)^{-1}X'\,Var(y)\,X(X'X)^{-1} \\ &= (X'X)^{-1}X'\,\sigma^2 I X(X'X)^{-1} \\ &= \sigma^2 (X'X)^{-1} \end{aligned} \tag{32}$$

Therefore we have

$$\hat{\beta} \sim N\left(\beta, \sigma^2 (X'X)^{-1}\right) \tag{33}$$

We can also show this by viewing $\hat{\beta}$ directly as a function of ϵ and then applying the theorem.

$$\begin{aligned} \hat{\beta} &= \beta + (X'X)^{-1}X'\varepsilon \\ &= \beta + C\varepsilon \\ \varepsilon &\sim N(0, \sigma^2 I) \\ \Rightarrow \quad \hat{\beta} &\sim N(\beta, C\sigma^2 I C') \\ \Rightarrow \quad \hat{\beta} &\sim N(\beta, \sigma^2 (X'X)^{-1}) \end{aligned} \tag{34}$$

H. Distribution of s^2

Consider the quantity

$$\frac{(n-k)s^2}{\sigma^2} \tag{35}$$

This can be written

$$\begin{aligned} \frac{(n-k)s^2}{\sigma^2} &= \frac{(n-k)\frac{e'e}{(n-k)}}{\sigma^2} \\ &= \frac{e'e}{\sigma^2} \\ &= \frac{\varepsilon' M_X \varepsilon}{\sigma^2} \\ &= \left[\frac{\varepsilon}{\sigma}\right]' M_X \left[\frac{\varepsilon}{\sigma}\right] \end{aligned} \tag{36}$$

The random variable (ε/σ) is a standard normal variable with mean zero and variance I. The matrix M_X is symmetric and idempotent. By Theorem 3 on quadratic forms, this ratio is distributed as a χ^2 variable with (n-k) degrees of freedom, that is

$$\frac{(n-k)s^2}{\sigma^2} \sim \chi^2(n-k) \tag{37}$$

where we found the trace of M_X in equation 24.

Given that $\frac{(n-k)s^2}{\sigma^2} \sim \chi^2(n-k)$, we can use information on the properties of chi-squared random variables to find the variance of s^2. First remember that the variance of a χ^2 variable is equal to twice its degrees of freedom, i.e., Var $(\chi^2(\nu)) = 2\nu$. Now rearrange equation 37 as as follows

$$
\begin{aligned}
Var\left(\frac{(n-k)s^2}{\sigma^2}\right) &= 2(n-k)\\
\Rightarrow \frac{(n-k)^2}{(\sigma^2)^2}Var(s^2) &= 2(n-k)\\
\Rightarrow Var(s^2) &= \frac{2\sigma^4}{n-k}
\end{aligned} \tag{38}
$$

I. sampling distribution of $\hat{\beta} - \beta$

1. sample variance of $\hat{\beta}$

We showed in equation 34 that

$$
\hat{\beta} \sim N\left(\beta; \sigma^2 (X'X)^{-1}\right) \tag{39}
$$

We can write the variance of $\hat{\beta}$ as

$$
\sigma^2(X'X)^{-1} = \begin{pmatrix} \sigma^2_{\hat{\beta}_1} & \sigma_{\hat{\beta}_1\hat{\beta}_2} & \cdots & \sigma_{\hat{\beta}_1\hat{\beta}_k} \\ \sigma_{\hat{\beta}_2\hat{\beta}_1} & \sigma^2_{\hat{\beta}_2} & \cdots & \sigma_{\hat{\beta}_2\hat{\beta}_k} \\ \vdots & & & \\ \sigma_{\hat{\beta}_k\hat{\beta}_1} & \sigma_{\hat{\beta}_k\hat{\beta}_2} & & \sigma^2_{\hat{\beta}_k} \end{pmatrix} \tag{40}
$$

We can estimate this using s^2 as an estimate of σ^2

$$
s^2(X'X)^{-1} = \begin{pmatrix} s^2_{\hat{\beta}_1} & s_{\hat{\beta}_1\hat{\beta}_2} & \cdots & s_{\hat{\beta}_1\hat{\beta}_k} \\ s_{\hat{\beta}_2\hat{\beta}_1} & s^2_{\hat{\beta}_2} & \cdots & s_{\hat{\beta}_2\hat{\beta}_k} \\ \vdots & & & \\ s_{\hat{\beta}_k\hat{\beta}_1} & s_{\hat{\beta}_k\hat{\beta}_2} & & s^2_{\hat{\beta}_k} \end{pmatrix} \tag{41}
$$

Note that the individual variances of the coefficients are equal to an element of $(X'X)^{-1}$, say s_{ii} times s^2. Using s_{ij} for the ijth element of $(X'X)^{-1}$ is a sometimes confusing notation, but seems to be standard.

2. distribution of $\dfrac{\hat{\beta}_i - \beta_i}{\sigma}$

First consider the moments of $\hat{\beta} - \beta$. From equation 2 write $\hat{\beta}$ as a function of ϵ

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'y \\ &= (X'X)^{-1}X'(X\beta + \epsilon) \\ &= \beta + (X'X)^{-1}X'\varepsilon\end{aligned} \tag{42}$$

As usual define $C = (X'X)^{-1}X'$ and write (42) as

$$\begin{aligned}\hat{\beta} &= \beta + C\varepsilon \\ \Rightarrow \hat{\beta} - \beta &= C\varepsilon\end{aligned} \tag{43}$$

Now compute the mean and variance of $\hat{\beta} - \beta$

$$\begin{aligned}E(C\varepsilon) &= CE(\varepsilon) = 0 \\ Var(C\varepsilon) &= CVar(\varepsilon)C' \\ &= (X'X)^{-1}X'\,\sigma^2 I X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}\end{aligned} \tag{44}$$

We noted previously that

$$\left(\frac{\varepsilon}{\sigma}\right) \sim N\left(0, \left(\frac{1}{\sigma^2}\right)\sigma^2 I\right) \sim N(0, I) \tag{45}$$

Now consider the moments of $\dfrac{\hat{\beta} - \beta}{\sigma}$. This can be written in a manner similar to (44) using the matrix C as follows

$$\begin{aligned}\frac{\hat{\beta} - \beta}{\sigma} &= C\left(\frac{\varepsilon}{\sigma}\right) \\ E\left(C\left(\frac{\varepsilon}{\sigma}\right)\right) &= C\,\frac{E(\varepsilon)}{\sigma} = 0 \\ Var\left(C\left(\frac{\varepsilon}{\sigma}\right)\right) &= C\,Var\left(\frac{\varepsilon}{\sigma}\right)C' \\ &= (X'X)^{-1}X'\left(\frac{1}{\sigma^2}\right)\sigma^2 I\,X(X'X)^{-1} \\ &= (X'X)^{-1}\end{aligned} \tag{46}$$

Given that $\dfrac{\epsilon}{\sigma}$ is distributed normally, this implies that

$$\frac{\hat{\beta} - \beta}{\sigma} = C\left(\frac{\varepsilon}{\sigma}\right) \sim N(0, (X'X)^{-1}) \tag{47}$$

Now consider a single element of $\hat{\beta}$ say $\hat{\beta}_i$

$$\frac{\hat{\beta} - \beta}{\sigma} \sim N(0, (X'X)^{-1})$$
$$\Rightarrow \frac{\hat{\beta}_i - \beta_i}{\sigma} \sim N(0, s_{ii}) \tag{48}$$

To create a N(0, 1) variable, we divide the left hand side by $\sqrt{s_{ii}}$, the appropriate element on the diagonal of $(X'X)^{-1}$. Doing so we obtain

$$\frac{\hat{\beta}_i - \beta_i}{\sigma} \sim N(0, s_{ii})$$
$$\Rightarrow \frac{\hat{\beta}_i - \beta_i}{\sigma \sqrt{s_{ii}}} = \frac{\hat{\beta}_i - \beta_i}{\sigma_{\beta_i}} \sim N(0, 1) \tag{49}$$

3. distribution of $\dfrac{(n-k)\, s^2_{\hat{\beta}_i}}{\sigma^2_{\hat{\beta}_i}}$

We start by recalling the discussion of the distribution of s^2 from equation 37

$$\frac{(n - k)s^2}{\sigma^2} \sim \chi^2(n - k) \tag{50}$$

Now multiply the numerator and denominator of this expression by s_{ii} as follows

$$\frac{(n - k)s^2}{\sigma^2} \sim \chi^2(n - k)$$
$$\Rightarrow \frac{(n - k)s^2 s_{ii}}{\sigma^2 s_{ii}} \sim \chi^2(n - k) \tag{51}$$
$$\Rightarrow \frac{(n - k)s^2_{\hat{\beta}_i}}{\sigma^2_{\hat{\beta}_i}} \sim \chi^2(n - k)$$

Given that the numerator and denominator are multiplied by the same thing, the distribution does not change.

4. distribution of $\dfrac{\hat{\beta}_i - \beta_i}{s_{\hat{\beta}_i}}$

We start by dividing the expression in equation 51 by (n - k) and then taking its square root as follows

$$\chi d = \sqrt{\left(\frac{(n - k) s^2_{\hat{\beta}_i}}{(n-k)\, \sigma^2_{\hat{\beta}_i}} \right)}$$

$$= \left(\frac{s^2_{\hat{\beta}_i}}{\sigma^2_{\hat{\beta}_i}} \right)^{\frac{1}{2}} \tag{52}$$

$$= \left(\frac{s_{\hat{\beta}_i}}{\sigma_{\hat{\beta}_i}} \right)$$

Now form the ratio of equation 49 and equation 52 which we denote t

$$t = \frac{\left(\dfrac{\hat{\beta}_i - \beta_i}{\sigma_{\hat{\beta}_i}} \right)}{\left(\dfrac{s_{\hat{\beta}_i}}{\sigma_{\hat{\beta}_i}} \right)} \tag{53}$$

$$= \frac{\hat{\beta}_i - \beta_i}{s_{\hat{\beta}_i}}$$

Equation 53 is the ratio of a N(0, 1) variable from equation 49 and the square root of a chi-squared random variable divided by its degrees of freedom from equation 52. If we can show that there two variables are independent, then the expression in equation 53 is distributed as a **t** random variable with n - k degrees of freedom. Given that multiplying the numerator and denominator of (50) by the constant s_{ii} to obtain (51), and the denominator of (48) by $\sqrt{s_{ii}}$ will not affect independence, we will show independence of the terms in (53) by showing independence of (48) and (50).

These two equations are both functions of the same standard normal variable. We can show that they are independent as follows. First write $\frac{\hat{\beta} - \beta}{\sigma}$ as a function of $\frac{\epsilon}{\sigma}$ as in equation 47

$$\frac{\hat{\beta} - \beta}{\sigma} = C \left(\frac{\varepsilon}{\sigma} \right) \tag{54}$$

Then write $\frac{(n - k)s^2}{\sigma^2}$ as a function of $\frac{\epsilon}{\sigma}$ as in equation 36

$$\frac{(n-k)\, s^2}{\sigma^2} = \left[\frac{\varepsilon}{\sigma} \right]' M_X \left[\frac{\varepsilon}{\sigma} \right] \tag{55}$$

We showed that $\frac{\epsilon}{\sigma}$ has a mean of zero and a variance of 1 in equation 45. Now consider Quadratic Form Theorem 4, which we repeat here for convenience.

> Quadratic Form Theorem 4: If $y \sim N(0,\sigma^2 I)$, M is a symmetric idempotent matrix of order n, and L is a kxn matrix, then Ly and $y'My$ are independently distributed if LM = 0.

Apply the theorem with $\frac{\epsilon}{\sigma}$ in the place of y, M_X is the place of M and C in the place of L. If CM_X= 0, then the numerator and denominator of equation 53 are independent. We can show this as follows

$$\begin{aligned} CM_X &= (X'X)^{-1}X'(I - X(X'X)^{-1}X') \\ &= [(X'X)^{-1}X' - (X'X)^{-1}X'X(X'X)^{-1}X'] \\ &= [(X'X)^{-1}X' - (X'X)^{-1}X'] = 0 \end{aligned} \tag{56}$$

What we have then shown is that

$$\begin{aligned} \frac{\hat{\beta}_i - \beta_i}{s_{\hat{\beta}_i}} &= \frac{\hat{\beta}_i - \beta_i}{[s^2 (X'X)^{-1}_{ii}]^{\frac{1}{2}}} \\ &= \frac{\hat{\beta}_i - \beta_i}{[s^2 s_{ii}]^{\frac{1}{2}}} \\ &= \frac{N(0,1)}{\sqrt{\frac{\chi^2(v)}{v}}} \sim t(v) \end{aligned} \tag{57}$$

Hypotheses of the form H_0: $\beta_i = \beta^0$ can be tested using the result

$$\frac{\hat{\beta}_i - \beta_i^0}{s_{\hat{\beta}_i}} \sim t(n-k) \tag{58}$$

J. independence of $\hat{\beta}$ and e under normality

First use equation 43 and equation 26 to write $\hat{\beta}$ and e as functions of ϵ as follows

$$\begin{aligned} \hat{\beta} &= \beta + (X'X)^{-1}X'\varepsilon \\ &= \beta + C\varepsilon \\ &\Rightarrow \hat{\beta} - \beta = C\varepsilon \\ e &= M_X y \\ &= M_X(X\beta + \varepsilon) = M_X\varepsilon \end{aligned} \tag{59}$$

Now consider application of Theorem 8 on quadratic forms. Given possible confusion with the variable y in the theorem as stated earlier and the y in our model, we restate the theorem with u replacing y as follows

> Quadratic Form Theorem 8: Let u~ N(0, I). Let M be a nonrandom idempotent matrix of dimension nxn (rank(M)=r ≤ n). Let A be a nonrandom matrix such that AM = 0. Let t_1 = Mu and let t_2 = Au. Then t_1 and t_2 are independent random vectors.

We will let M_X replace M and C replace A when we apply the theorem. Now let u = (1/σ)ε or ε = σu. Clearly u ~ N(0, I). Now rewrite the expressions in equation 59 replacing ϵ with σu as follows

$$\begin{aligned} \hat{\beta} - \beta &= C\sigma u \\ e &= M_X \sigma u \end{aligned} \tag{60}$$

Now define the new variables z_1 and z_2 as

$$\begin{aligned} z_1 &= \frac{1}{\sigma} e = M_X u \\ z_2 &= \frac{1}{\sigma}(\hat{\beta} - \beta) = Cu \end{aligned} \tag{61}$$

The theorem states that if $CM_X = 0$, then z_1 and z_2 are independent and so are $e = \sigma z_1$ and $\hat{\beta} = \beta + \sigma z_2$. We have shown previously that $CM_X = 0$ as follows

$$\begin{aligned} C &= (X'X)^{-1}X' \\ M_X &= (I - X(X'X)^{-1}X') \\ CM_X &= (X'X)^{-1}X'(I - X(X'X)^{-1}X') \\ &= [(X'X)^{-1}X' - (X'X)^{-1}X'X(X'X)^{-1}X'] \\ &= [(X'X)^{-1}X' - (X'X)^{-1}X'] = 0 \end{aligned} \tag{62}$$

So the estimate of β, $\hat{\beta}$, is independent of the error term in the regression equation.

K. distribution of certain quadratic forms representing sums of squares

1. We will consider the statistical distribution of the following quadratic forms.

$$\begin{aligned} SST &= \sum_{t=1}^{n} (y_t - \bar{y})^2 = (y - \bar{y})'(y - \bar{y}) \\ SSE &= \sum_{t=1}^{n} (y_t - \hat{y}_t)^2 = e'e \\ SSR &= \sum_{t=1}^{n} (\hat{y}_t - \bar{y})^2 = (\hat{y} - \bar{y})'(\hat{y} - \bar{y}) \end{aligned} \tag{63}$$

We will be able to show that they are chi-squared variables and thus useful in performing statistical tests. It will be useful to write SST in terms of the deviation matrix A_n. When the matrix n×n matrix A_n premultiplies any n-vector y, the resulting n-vector is each element of y minus the mean of the y's. Specifically,

$$A_n = I - \frac{1}{n}jj'$$

$$= \begin{bmatrix} 1-\frac{1}{n} & \frac{-1}{n} & \cdots & \frac{-1}{n} \\ \frac{-1}{n} & 1-\frac{1}{n} & \cdots & \frac{-1}{n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{-1}{n} & \frac{-1}{n} & \cdots & 1-\frac{1}{n} \end{bmatrix} \tag{64}$$

$$\Rightarrow A_n y = \begin{bmatrix} 1-\frac{1}{n} & \frac{-1}{n} & \cdots & \frac{-1}{n} \\ \frac{-1}{n} & 1-\frac{1}{n} & \cdots & \frac{-1}{n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{-1}{n} & \frac{-1}{n} & \cdots & 1-\frac{1}{n} \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix}$$

Clearly then,

$$y'A_n'A_n y = \begin{pmatrix} y_1 - \bar{y} & y_2 - \bar{y} & \cdots & y_n - \bar{y} \end{pmatrix} = \begin{pmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix} \tag{65}$$

$$= \sum_{t=1}^{n} (y_t - \bar{y})^2$$

Given that A_n is symmetric and idempotent, we can also write

$$y'A_n y = \sum_{t=1}^{n} (y_t - \bar{y})^2 \tag{66}$$

2. distribution of SSE = $(n - k)s^2$

Given that s^2 = SSE/(n - k), we already know that it is a chi-squared variable. The demonstration is obvious given that $\frac{\epsilon}{\sigma}$ is a N(0, 1) variable. First write SSE as $e'e$ and remember that we can write $e'e$ as a function of ϵ from equation 22

$$\begin{aligned} e &= M_X y \\ &= M_X (X\beta + \varepsilon) \\ &= M_X \varepsilon \quad since \quad MX = 0 \\ \Rightarrow SSE = e'e &= \varepsilon' M_X' M_X \varepsilon \\ &= \varepsilon' M_X \varepsilon \end{aligned} \tag{67}$$

Now rewrite SSE using the residual matrix M_X. Consider now the following expression and its distribution

$$\frac{SSE}{\sigma^2} = \frac{e'e}{\sigma^2} = \frac{\varepsilon' M_X \varepsilon}{\sigma^2} \tag{68}$$

By appropriately rearranging equation 68, we can invoke Quadratic Form Theorem 3 as before. Divide each element of ε in (68) by σ to obtain a standard normal variable and then rewrite as follows

$$\begin{aligned} &\frac{\varepsilon}{\sigma} \sim N(0, 1) \\ &\frac{SSE}{\sigma^2} = \left[\frac{\varepsilon}{\sigma}\right]' M_X \left[\frac{\varepsilon}{\sigma}\right] \sim \chi^2(tr\ M_X) \sim \chi^2(n-k) \end{aligned} \tag{69}$$

3. distribution of SSR

Write SSR as the difference between SST and SSE

$$\begin{aligned} SST &= \sum_{t=1}^{n} (y_t - \bar{y})^2 = (y - \bar{y})'(y - \bar{y}) = y'Ay \\ SSE &= \sum_{t=1}^{n} (y_t - \hat{y}_t)^2 = e'e = y'My \\ SSR &= \sum_{t=1}^{n} (\hat{y}_t - \bar{y})^2 = (\hat{y} - \bar{y})'(\hat{y} - \bar{y}) = \hat{\beta}'X'AX\hat{\beta} \\ &= SST - SSE = y'Ay - y'My \end{aligned} \tag{70}$$

Because SSR measures the sum of squares due to the inclusion of the slope coefficients (β_2, β_3, ... , β_k) we need to consider the model with this fact explicitly represented

$$\begin{aligned} y &= X\beta + \varepsilon \\ &= x_1\beta_1 + X_2\beta_2 + \varepsilon \\ &= \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \beta_1 + \begin{pmatrix} x_{12} & x_{13} & \cdots & x_{1k} \\ x_{22} & x_{23} & \cdots & x_{2k} \\ \vdots & & \vdots & \\ x_{n2} & x_{n3} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \end{aligned} \quad \textbf{(71)}$$

Now we need to consider the properties of A_n in relation to this rewritten model. Note that A_n multiplied by a column of constants will yield the zero vector because the mean of the column will equal each element of the column. This is specifically true for x_1 in equation 71.

$$A_n x_1 = \begin{bmatrix} 1-\frac{1}{n} & \frac{-1}{n} & \cdots & \frac{-1}{n} \\ \frac{-1}{n} & 1-\frac{1}{n} & \cdots & \frac{-1}{n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{-1}{n} & \frac{-1}{n} & \cdots & 1-\frac{1}{n} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \textbf{(72)}$$

$$\begin{aligned} \Rightarrow A_n y &= A_n x_1 \beta_1 + A_n X_2 \beta_2 + A_n \varepsilon \\ &= A_n X_2 \beta_2 + A_n \varepsilon \end{aligned}$$

This then implies that we can obtain β_2 by a regression of deviations of variables from their column means. It also means that we can write the vector of deviations of each element of y from is mean as

$$\begin{aligned} A_n y &= \begin{pmatrix} y_1 - \bar{y} \\ y_2 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix} = A_n (X_2\beta_2 + \varepsilon) \\ &= A_n X_2 \beta_2 + A_n \varepsilon \end{aligned} \quad \textbf{(73)}$$

Now construct SSR using this information

$$
\begin{aligned}
SSR &= y'A_n y - y'M_X y \\
&= y'(AX_2\beta_2 + A\varepsilon) - \varepsilon' M\varepsilon
\end{aligned}
\tag{74}
$$

Now substitute from equation 71 for y in equation 74

$$
\begin{aligned}
SSR &= y'(AX_2\beta_2 + A\varepsilon) - \varepsilon' M\varepsilon \\
&= [\beta_1' x_1' + \beta_2' X_2' + \varepsilon'][AX_2\beta_2 + A\varepsilon] - \varepsilon' M\varepsilon \\
&= (\beta_1' x_1' AX_2\beta_2 + \beta_1' x_1' A\varepsilon) \\
&\quad + (\beta_2' X_2' AX_2\beta_2 + \beta_2' X_2' A\varepsilon) \\
&\quad + (\varepsilon' AX_2\beta_2 + \varepsilon' A\varepsilon) - \varepsilon' M\varepsilon
\end{aligned}
\tag{75}
$$

The terms containing $x_1' A$ will be zero from equation 72. We can also reverse the order in terms as they are conformable given that we are computing a scalar, so we have

$$
\begin{aligned}
SSR &= (\beta_1' x_1' AX_2\beta_2 + \beta_1' x_1' A\varepsilon) \\
&\quad + (\beta_2' X_2' AX_2\beta_2 + \beta_2' X_2' A\varepsilon) \\
&\quad + (\varepsilon' AX_2\beta_2 + \varepsilon' A\varepsilon) - \varepsilon' M\varepsilon \\
&= \beta_2' X_2' AX_2\beta_2 + 2\beta_2' X_2' A\varepsilon + \varepsilon' A\varepsilon - \varepsilon' M\varepsilon \\
&= \varepsilon' A\varepsilon - \varepsilon' M\varepsilon \qquad \text{if} \;\; \beta_2 = 0 \\
&= \varepsilon'(A - M)\varepsilon
\end{aligned}
\tag{76}
$$

Now we want to find the distribution of the ratio

$$
\begin{aligned}
\frac{SSR}{\sigma^2} &= \frac{\varepsilon'(A_n - M_X)\varepsilon}{\sigma^2} \\
&= \left(\frac{\varepsilon}{\sigma}\right)' (A_n - M_X) \left(\frac{\varepsilon}{\sigma}\right)
\end{aligned}
\tag{77}
$$

We know from equation 45 that $\frac{\epsilon}{\sigma}$ is a N(0, 1) variable. If we apply Quadratic Form Theorem 3, we then obtain

$$
\begin{aligned}
\frac{SSR}{\sigma^2} &= \left(\frac{\varepsilon}{\sigma}\right)' (A_n - M_X) \left(\frac{\varepsilon}{\sigma}\right) \\
&\sim \chi^2(tr\,(A_n - M_X)), \qquad \textit{if } (A_n - M_X) \textit{ is symmetric and idempotent}
\end{aligned}
\tag{78}
$$

Clearly (A - M) is symmetric, given that A and M are both symmetric. To check if it is idempotent write it out as

$$\begin{aligned}(A-M)(A-M) &= AA - MA - AM + MM \\ &= A - 2AM + M \\ &= A - M \quad \mathit{if} \quad AM = M\end{aligned} \tag{79}$$

Then remember from equation 10 that

$$A_n M_X = M_X \tag{80}$$

So we have

$$\begin{aligned}(A - M)(A - M) &= A - 2AM + M \\ &= A - 2M + M \\ &= A - M\end{aligned} \tag{81}$$

We need only to determine the trace of (A - M). The trace of the sum of matrices is equal to the sum of the traces.

$$\begin{aligned} tr\ (A) &= \begin{bmatrix} 1-\frac{1}{n} & \frac{-1}{n} & \ldots & \frac{-1}{n} \\ \frac{-1}{n} & 1-\frac{1}{n} & \ldots & \frac{-1}{n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{-1}{n} & \frac{-1}{n} & \ldots & 1-\frac{1}{n} \end{bmatrix} \\ &= n - \frac{n}{n} = n - 1 \end{aligned} \tag{82}$$

Now find the trace of M

$$\begin{aligned} trM &= trI_n - tr[X(X'X)^{-1}X'] \\ &= trI_n - tr[X'X(X'X)^{-1}], \quad \mathit{rearrange\ order} \\ &= trI_n - trI_k \\ &= n - k \end{aligned} \tag{83}$$

Combining the information from equations 82 and 83 we obtain

$$\begin{aligned} tr\ (A-M) &= (n - 1) - (n - k) \\ &= k - 1 \end{aligned} \tag{84}$$

To summarize

$$\frac{SSR}{\sigma^2} = \left[\frac{\varepsilon}{\sigma}\right]'(A - M)\left[\frac{\varepsilon}{\sigma}\right] \sim \chi^2(tr\ (A - M)) \tag{85}$$

$$\textit{if all slope coefficients are zero}$$

4. distribution of SST

We showed in equation 66 that SST can be written as

$$SST = \sum_{t=1}^{n} (y_t - \bar{y})^2 = y'A_n y \tag{86}$$

As discussed earlier in the section on probability distributions, the sum of two independent χ^2 variables is also χ^2 with degrees of freedom equal to the sum of the degrees of freedom of the variables in the sum. If SSE and SSR are independent, then SST will be distributed as a χ^2 variable with [(n - k) + (k - 1) = (n - 1)] degrees of freedom. The question is if SSR and SSE are independent. To show independence we will use Quadratic Form Theorem 5 which we repeat here for convenience.

> Quadratic Form Theorem 5: Let the nx1 vector y~N(0,I), let A be an nxn idempotent matrix of rank m, let B be an nxn idempotent matrix of rank s, and suppose BA = 0. Then y′Ay and y′By are independently distributed χ^2 variables.

To show independence, we we must show that product of the matrices in the two quadratic forms $\left[\frac{\varepsilon}{\sigma}\right]' M_X \left[\frac{\varepsilon}{\sigma}\right]$ and $\left[\frac{\varepsilon}{\sigma}\right]'(A_n - M_X)\left[\frac{\varepsilon}{\sigma}\right]$ is zero. Specifically we have,

$$\begin{aligned} \frac{SSE}{\sigma^2} &= \left[\frac{\varepsilon}{\sigma}\right]' M_X \left[\frac{\varepsilon}{\sigma}\right] \sim \chi^2(n - k) \\ \frac{SSR}{\sigma^2} &= \left[\frac{\varepsilon}{\sigma}\right]' (A_n - M_X) \left[\frac{\varepsilon}{\sigma}\right] \sim \chi^2(k - 1) \end{aligned} \tag{87}$$

So we must show that the product of the matrices M_X and (A_n - M_X) is zero,

$$\begin{aligned} (A_n - M_X)M_X &= A_n M_X - M_X M_X \\ &= A_n M_X - M_X, \quad M_X \textit{ is idempotent} \\ &= M_X - M_X, \quad \textit{from equation } 10 \\ &= 0 \end{aligned} \tag{88}$$

Therefore

$$\begin{aligned} &\frac{SST}{\sigma^2} = \frac{SSR}{\sigma^2} + \frac{SSE}{\sigma^2} \sim \chi^2 \textit{ and has degrees of freedom} \\ &(n-1) = (k-1) + (n-k) \end{aligned} \tag{89}$$

L. Tests for significance of the regression

Suppose we want to test the following hypothesis:

H_o: $\beta_2 = \beta_3 = \ldots = \beta_k = 0$

This hypothesis tests for the statistical significance of <u>overall</u> explanatory power, i.e., $y_t = \beta_1 + \varepsilon_t$ (all nonintercept coefficients = 0). The best way to test this is by using information on the sum of squares due to the regression, the error, and overall. Recall that the total sum of squares can be partitioned as:

$$\begin{aligned} SST &= \sum_{t=1}^{n} (y_t - \bar{y})^2 = \sum_{t=1}^{n} (y_t - \hat{y}_t)^2 + \sum_{t=1}^{n} (\hat{y}_t - \bar{y})^2 \\ &= SSE + SSR \end{aligned} \tag{90}$$

Dividing both sides of the equation by σ^2 yields quadratic forms which have chi-square distributions as above. From the section on probability distributions, we know that the ratio of two chi-square variables, each divided by its degrees of freedom, is a F random variable. This result provides the basis for using

$$F = \frac{\frac{SSR}{k-1}}{\frac{SSE}{n-k}} = \frac{\chi^2(k-1)\,(n-k)}{\chi^2(n-k)\,(k-1)} \sim F(k-1,\, n-k) \tag{91}$$

to test the hypothesis that $\beta_2 = \beta_3 = \ldots = \beta_k = 0$.

Also note that

$$\frac{SSR}{SSE} = \frac{SSR}{SST - SSR} = \frac{\frac{SSR}{SST}}{1 - \frac{SSR}{SST}} = \frac{R^2}{1 - R^2} \tag{92}$$

hence, the F statistic can be rewritten as

$$F = \frac{\frac{R^2}{k-1}}{\frac{1-R^2}{n-k}} = \left(\frac{n-k}{k-1}\right)\left(\frac{R^2}{1-R^2}\right) \sim F(k-1, n-k) \tag{93}$$

If the computed F statistic is larger than the tabled value, then we reject the hypothesis that β_2 (all slope coefficients) is zero.

M. Tests of a single linear restriction on β

1. idea

Sometimes we want to test a hypothesis regarding a linear combination of the β_i's in the classical linear regression model. Such a hypothesis can be written

$$\delta'\beta = \gamma \quad \textit{where} \quad \delta' \quad \textit{is} \quad 1 \times k \tag{94}$$

For example, to test that $\beta_2 = \beta_3$ in a model with 4 regressors (X is n × 4), we might define δ' and γ as follows

$$\begin{aligned} \delta' &= \begin{bmatrix} 0 & 1 & -1 & 0 \end{bmatrix}, \quad \gamma = 0 \\ \delta'\beta &= \gamma \\ \Rightarrow \begin{bmatrix} 0 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} &= 0 \end{aligned} \tag{95}$$

2. distribution of $\delta'\hat{\beta}$

We have previously (equation 40) shown that the variance of $\hat{\beta}$ is given by

$$\sigma^2(X'X)^{-1} = \begin{pmatrix} \sigma^2_{\hat{\beta}_1} & \sigma_{\hat{\beta}_1\hat{\beta}_2} & \cdots & \sigma_{\hat{\beta}_1\hat{\beta}_k} \\ \sigma_{\hat{\beta}_2\hat{\beta}_1} & \sigma^2_{\hat{\beta}_2} & \cdots & \sigma_{\hat{\beta}_2\hat{\beta}_k} \\ \vdots & & & \\ \sigma_{\hat{\beta}_k\hat{\beta}_1} & \sigma_{\hat{\beta}_k\hat{\beta}_2} & & \sigma^2_{\hat{\beta}_k} \end{pmatrix} \tag{96}$$

This then implies that the variance of $\delta'\hat{\beta}$ is given by

$$Var(\delta'\hat{\beta}) = \delta'\sigma^2(X'X)^{-1}\delta \tag{97}$$

So if the null hypothesis that $\delta'\hat{\beta} = \delta'\beta$ is true, then $\delta'\hat{\beta}$ is distributed normally with mean $\delta'\beta$ and variance $\sigma^2\delta'(X'X)^{-1}\delta$, i.e.,

$$\delta'\hat{\beta} \sim N\big(\delta'\beta;\, \sigma^2\delta'(X'X)^{-1}\delta\big) \tag{98}$$

This then implies that

$$\frac{\delta'\hat{\beta} - \delta'\beta}{\sqrt{\sigma^2 \delta'(X'X)^{-1}\delta}} \sim N(0;1)$$
$$\Rightarrow \frac{\delta'\hat{\beta} - \delta'\beta}{\sigma\sqrt{\delta'(X'X)^{-1}\delta}} \sim N(0;1) \tag{99}$$

3. estimating the variance of $\delta'\hat{\beta}$

The variance of $\delta'\hat{\beta}$ is $\sigma^2_{\delta'\hat{\beta}} = \sigma^2\delta'(X'X)^{-1}\delta$. We can estimate this variance as

$$s^2_{\delta'\hat{\beta}} = s^2\delta'(X'X)^{-1}\delta \tag{100}$$

4. distribution of $\delta'\hat{\beta}$

From equation 37 we know that

$$\frac{(n-k)s^2}{\sigma^2} \sim \chi^2(n-k) \tag{101}$$

If we multiply the numerator and denominator of equation 101 by $\delta'(X'X)^{-1}\delta$ we obtain

$$\frac{(n-k)s^2\delta'(X'X)^{-1}\delta}{\sigma^2\delta'(X'X)^{-1}\delta} \sim \chi^2(n-k)$$
$$\Rightarrow \frac{(n-k)s^2_{\delta'\hat{\beta}}}{\sigma^2_{\delta'\hat{\beta}}} \sim \chi^2(n-k) \tag{102}$$

Now consider the ratio of the statistic in equation 99 with the square root of the statistic in equation 102 divided by (n - k). Writing this out and simplifying we obtain

$$\frac{\dfrac{\delta'\hat{\beta} - \delta'\beta}{\sigma\sqrt{\delta'(X'X)^{-1}\delta}}}{\sqrt{\dfrac{(n-k)s^2\delta'(X'X)^{-1}\delta}{(n-k)\sigma^2\delta'(X'X)^{-1}\delta}}} = \frac{\delta'\hat{\beta} - \delta'\beta}{\sqrt{s^2\delta'(X'X)^{-1}\delta}}$$
$$= \frac{\delta'\hat{\beta} - \delta'\beta}{\sqrt{s^2_{\delta'\hat{\beta}}}}$$
$$= \frac{\delta'\hat{\beta} - \delta'\beta}{s_{\delta'\hat{\beta}}} \tag{103}$$

Equation 103 is ratio of a N(0, 1) variable from equation 99 and the square root of a chi-squared random variable divided by its degrees of freedom from equation 102. If the numerator and denominator in equation 103 are independent then the statistic is distributed as a t random variable with (n - k) degrees of freedom. But we showed in part J (equations 60-62) that under normality $\hat{\beta}$ and **e** are independent. Given that s^2 is only a function of e, the numerator and denominator must be independent. We can also show this in a manner similar to that used to show that $\frac{\hat{\beta}_i - \beta_i}{s_{\hat{\beta}_i}}$ is distributed as a t random variable in equations 57 and 58. First write $\delta'\beta$ as a function of ϵ as follows

$$\begin{aligned}
\delta'\hat{\beta} &= \delta'(X'X)^{-1}X'y \\
&= \delta'(X'X)^{-1}X'(X\beta + \epsilon) \\
&= \delta'\beta + \delta'(X'X)^{-1}X'\epsilon \\
\Rightarrow \delta'\hat{\beta} - \delta'\beta &= \delta'(X'X)^{-1}X'\epsilon \\
&= \delta'C\epsilon \\
\Rightarrow \frac{\delta'\hat{\beta} - \delta'\beta}{\sigma} &= \delta'C\left(\frac{\epsilon}{\sigma}\right)
\end{aligned} \tag{104}$$

We can then apply Quadratic Form Theorem 8 as previously given that the denominator in equation 103 is just $\frac{(n-k)s^2}{\sigma^2}$ as previously. Therefore

$$\frac{\delta'\hat{\beta} - \delta'\beta}{\sqrt{\delta' s^2 (X'X)^{-1}\delta}} = \frac{\delta'\hat{\beta} - \gamma}{\sqrt{s^2_{\delta'\hat{\beta}}}} \sim t(n-k). \tag{105}$$

where $\gamma = \delta'\beta$. Such a test involves running a regression and constructing the estimated values of $\delta'\hat{\beta}$ and the variance of $\delta'\hat{\beta}$ from $\hat{\beta}$ and $s^2(X'X)^{-1}$.

N. Tests of several linear restrictions on β

1. idea

Consider a set of m linear constraints on the coefficients denoted by

$$R\beta = r \quad R \textit{ is } mxk, \; r \textit{ is } mx1 \tag{106}$$

To test this we need to discover how far $R\hat{\beta}$ is from r. To understand the intuition define a new variable d as $r - R\hat{\beta}$. This variable should be close to zero if the hypothesis is true. Note that d is normally distributed since it is a linear function of the normal variable $\hat{\beta}$. Its mean and variance are given as

$$\begin{aligned}
E(d) &= E(r - R\hat{\beta}) = 0 \quad \textit{if the hypothesis is true} \\
Var(d) &= Var(r - R\hat{\beta}) = R\,Var(\hat{\beta})R' = \sigma^2 R(X'X)^{-1}R'
\end{aligned} \tag{107}$$

2. A possible test statistic for testing H_0: d = 0

Consider Quadratic Form Theorem 7 which is as follows.

Quadratic Form Theorem 7 If y is a nx1 random variable and $y \sim N(\mu,\Sigma)$ then $(y - \mu)'\Sigma^{-1}(y - \mu) \sim \chi^2(n)$.

A possible test statistic is

$$\begin{aligned}(d-0)'[Var(d)]^{-1}(d-0) &= d'[\sigma^2 R(X'X)^{-1}R']^{-1}d \\ &= \frac{d'[R(X'X)^{-1}R']^{-1}d}{\sigma^2} \sim \chi^2(m)\end{aligned} \quad \textbf{(108)}$$

The problem with this is that σ^2 is not known.

3. A more useful test statistic

Consider the following test statistic

$$\frac{\frac{(r-R\hat{\beta})'[R(X'X)^{-1}R']^{-1}(r-R\hat{\beta})}{m}}{\frac{SSE}{n-k}} = \frac{\frac{(r-R\hat{\beta})'[R(X'X)^{-1}R']^{-1}(r-R\hat{\beta})}{m}}{s^2} \quad \textbf{(109)}$$

We can show that it is distributed as an F by showing that the numerator and denominator are independent chi-square variables.

First consider the numerator. We will show that we can write it as $\epsilon'Q\epsilon$ where Q is symmetric and idempotent. First write $r - R\hat{\beta}$ in the following useful manner.

$$\textit{If } r = R\beta \textit{ then } (r - R\hat{\beta}) = R\beta - R\hat{\beta} = R(\beta - \hat{\beta}) \quad \textbf{(110)}$$

Now write $R(\beta - \hat{\beta})$ as a function of ϵ

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + \varepsilon) \\ &= \beta + (X'X)^{-1}X'\varepsilon \\ \Rightarrow \hat{\beta} - \beta &= (X'X)^{-1}X'\varepsilon \\ \Rightarrow \beta - \hat{\beta} &= -(X'X)^{-1}X'\varepsilon \\ \Rightarrow R(\beta - \hat{\beta}) &= -R(X'X)^{-1}X'\varepsilon\end{aligned} \quad \textbf{(111)}$$

Then write out the numerator of equation 109 as follows

$$\begin{aligned}(r-R\hat{\beta})'[R(X'X)^{-1}R']^{-1}(r-R\hat{\beta}) &= \varepsilon'X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X'\varepsilon \\ &= \varepsilon'Q\varepsilon, \\ Q &= X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X'\end{aligned} \quad \textbf{(112)}$$

Notice by inspection that the matrix Q is symmetric. It is idempotent, as can be seen by writing it out as follows

$$\begin{aligned} Q &= X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X' \\ QQ &= X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X' \; X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X' \\ &= X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X' \\ &= X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X' \end{aligned} \tag{113}$$

Now find the trace of Q

$$\begin{aligned} trace\, Q &= tr\,[X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X'] \\ &= tr\,[[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X'X(X'X)^{-1}R'] \\ &= tr\,[[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}R'] \\ &= tr\, I_m = m \end{aligned} \tag{114}$$

Now remember from equation 45 that

$$\left(\frac{\varepsilon}{\sigma}\right) \sim N\left(0,\; \left(\frac{1}{\sigma^2}\right)\sigma^2 I\right) \sim N(0, I) \tag{115}$$

We then have

$$\begin{aligned} \left[\frac{\varepsilon}{\sigma}\right]' Q \left[\frac{\varepsilon}{\sigma}\right] &= \frac{\varepsilon' Q \varepsilon}{\sigma^2} \\ &\sim \chi^2(m) \end{aligned} \tag{116}$$

from Quadratic Form Theorem 3. Now consider the denominator. We can show that it distributed as χ^2(n-k) using Quadratic Form Theorem 3 as follows

$$\begin{aligned}
\frac{SSE}{n-k} &= \frac{e'e}{n-k} \\
e'e &= \varepsilon' M_X \varepsilon \\
\Rightarrow SSE &= \varepsilon' M_X \varepsilon \\
&= \left[\frac{\varepsilon}{\sigma}\right]' M_X \left[\frac{\varepsilon}{\sigma}\right] \\
&= \frac{\varepsilon' M_X \varepsilon}{\sigma^2} \sim \chi^2(n-k)
\end{aligned} \tag{117}$$

The last step follows because M_X is symmetric and idempotent.. Independence follows from Quadratic Form Theorem 5 because $M_X Q = 0$.

$$\begin{aligned}
M_X &= (I - X(X'X)^{-1}X') \\
Q &= X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X' \\
M_X Q &= (I - X(X'X)^{-1}X')\,X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X' \\
&= X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X' \\
&- X(X'X)^{-1}X'X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X' \\
&= X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X' \\
&- X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}X' \\
&= 0
\end{aligned} \tag{118}$$

Or we can simply remember that $M_X X = 0$ from previously, and note that the leading term in Q is X.

The test statistic is then distributed as an F with m and n-k degrees of freedom. We reject the null hypothesis that the set of linear restrictions holds if the computed value of the statistic is larger than the tabled value.

A random variable distributed as an F(1, n-k) is the square of a random variable distributed as a t(n-k), so when there is a single linear linear restriction on $\hat{\beta}$ (m = 1), a t-test based on 105 and an F test based on equation 109 give the same result.

$\hat{\beta}$, $\tilde{\beta}$, and β^{ℓ} are all

a. unbiased
b. consistent
c. minimum variance of all unbiased estimators
d. normally distributed

1. $\sigma^2(X'X)^{-1}$ can be shown to be the Cramer-Rao matrix that is minimum variance.