Econometrics Lecture 2: OLS Estimation With Matrix Algebra

R. G. Pierse

1 The Classical Linear Regression Model

The classical linear regression model can be written as

$$y_i = \sum_{j=1}^{k} \beta_j x_{ij} + u_i, \quad i = 1, \cdots, n$$

where y_i is the ith observation on the dependent variable y, β_j is the coefficient on the jth explanatory variable or *regressor* x_{ij}, and u_i is the ith observation on an unobserved error term u. There are n observations and k regressors. All n observations in the model can be written as the set of equations

$$\begin{aligned}
y_1 &= \beta_1 x_{11} + \beta_2 x_{12} + \beta_3 x_{13} + \cdots + \beta_k x_{1k} + u_1 \\
y_2 &= \beta_1 x_{21} + \beta_2 x_{22} + \beta_3 x_{23} + \cdots + \beta_k x_{2k} + u_2 \\
&\vdots \\
y_n &= \beta_1 x_{n1} + \beta_2 x_{n2} + \beta_3 x_{n3} + \cdots + \beta_k x_{nk} + u_n.
\end{aligned}$$

This system itself can be re-expressed succinctly using matrix notation as

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

or

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{1.1}$$

where $\mathbf{y}$ and $\mathbf{u}$ are $n \times 1$ column vectors, $\boldsymbol{\beta}$ is a $k \times 1$ column vector and $\mathbf{X}$ is an $n \times k$ matrix, each column corresponding to a different regressor. By convention, where the regression contains an intercept, this will be the first column of the matrix $\mathbf{X}$ will consist of a vector of ones, corresponding to an intercept in the model. β_1 will then represent the coefficient on this intercept.

The error term $\mathbf{u}$ is a vector of random variables. It has an associated mean vector given by

$$\mathrm{E}(\mathbf{u}) = \begin{bmatrix} \mathrm{E}(u_1) \\ \mathrm{E}(u_2) \\ \vdots \\ \mathrm{E}(u_n) \end{bmatrix}$$

where E represents the *expectations operator*. Note that, because E is a linear operator, it can be taken inside matrices.

The variances and covariances of the n elements of $\mathbf{u}$ can be arranged into the $n \times n$ symmetric matrix

$$\mathrm{var}(\mathbf{u}) = \mathbf{\Sigma} = \mathrm{E}(\mathbf{u} - \mathrm{E}(\mathbf{u}))(\mathbf{u} - \mathrm{E}(\mathbf{u}))'$$

$$= \begin{bmatrix} \mathrm{E}(u_1 - \mathrm{E}(u_1))^2 & \cdots & \mathrm{E}(u_1 - \mathrm{E}(u_1))(u_n - \mathrm{E}(u_n)) \\ \mathrm{E}(u_2 - \mathrm{E}(u_2))(u_1 - \mathrm{E}(u_1)) & \cdots & \mathrm{E}(u_2 - \mathrm{E}(u_2))(u_n - \mathrm{E}(u_n)) \\ \vdots & \ddots & \vdots \\ \mathrm{E}(u_n - \mathrm{E}(u_n))(u_1 - \mathrm{E}(u_1)) & \cdots & \mathrm{E}(u_n - \mathrm{E}(u_n))^2 \end{bmatrix}$$

$$= \begin{bmatrix} \mathrm{var}(u_1) & \mathrm{cov}(u_1 u_2) & \cdots & \mathrm{cov}(u_1 u_n) \\ \mathrm{cov}(u_2 u_1) & \mathrm{cov}(u_2) & \cdots & \mathrm{cov}(u_2 u_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(u_n u_1) & \mathrm{cov}(u_n u_2) & \cdots & \mathrm{var}(u_n) \end{bmatrix}.$$

This matrix is called the *variance-covariance matrix* of u. Note that the variances are the diagonal elements and the covariances the off-diagonal elements. The variance-covariance matrix is both *symmetric* and *positive-definite*.

1.1 Assumptions of the Classical Model

The standard assumptions on $\mathbf{u}$ in the classical linear model are given in the familiar form by:

$$\mathrm{E}(u_i) = 0, \quad i = 1, \cdots, n$$

$$\mathrm{E}(u_i^2) = \sigma^2, \quad i = 1, \cdots, n$$

$$\mathrm{E}(u_i u_j) = 0, \quad i, j = 1, \cdots, n \quad j \neq i.$$

The first of these assumptions can be re-expressed in matrix terms as

$$\mathrm{E}(\mathbf{u}) = \mathbf{0} \tag{A1}$$

where the right-hand side is a zero vector is of order $n \times 1$. The second and third assumptions can be re-expressed in matrix terms in the single assumption

$$\text{var}(\mathbf{u}) = \text{E}(\mathbf{u}\mathbf{u}') = \sigma^2 \mathbf{I}_n. \tag{A2}$$

The diagonality of this matrix corresponds to the assumption of zero autocovariance while the constancy of the diagonal elements corresponds to the assumption of constant variance or *homoscedasticity*.

In addition we make the assumptions on the regressors that

$$\text{The } n \times k \text{ matrix } \mathbf{X} \text{ has rank } k \tag{A3}$$

and that

$$\text{The matrix } \mathbf{X} \text{ is } \textit{fixed in repeated sampling}. \tag{A4}$$

The first of these assumptions is that no single regressor can be expressed as an exact linear function of the other regressors. This is the assumption of *no perfect collinearity* in the regressors. The second assumption allows $\mathbf{X}$ to be treated as non-random so that it can be taken outside the expectations operator.

2 The Ordinary Least Squares Estimator

Let $\mathbf{b}$ be an estimator of the unknown parameter vector $\boldsymbol{\beta}$. Then

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \tag{2.1}$$

where $\mathbf{e}$ is an $n \times 1$ vector of residuals that are not explained by the regression.

The *OLS* estimator $\widehat{\boldsymbol{\beta}}$ is the estimator $\mathbf{b}$ that minimises the sum of squared residuals $\mathbf{s} = \mathbf{e}'\mathbf{e} = \sum_{i=1}^{n} e_i^2$.

$$\min_{\mathbf{b}} \ \mathbf{s} = \mathbf{e}'\mathbf{e} = (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b})$$

or, expanding the last expression,

$$\mathbf{s} = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{b} - \mathbf{b}'\mathbf{X}'\mathbf{y} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}.$$

Differentiating $\mathbf{s}$ with respect to $\mathbf{b}$ gives the vector of first order conditions:

$$\frac{\partial \mathbf{s}}{\partial \mathbf{b}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{0} \tag{2.2}$$

or, rearranging, the vector of normal equations

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}. \tag{2.3}$$

Note that this can also be written as

$$\mathbf{X}'(\mathbf{y}-\mathbf{X}\mathbf{b}) = \mathbf{X}'\mathbf{e} = \mathbf{0}.$$

On the assumption that the matrix $\mathbf{X}$ is of rank k, the $k \times k$ symmetric matrix $\mathbf{X}'\mathbf{X}$ will be of full rank and its inverse $(\mathbf{X}'\mathbf{X})^{-1}$ will exist. Premultiplying (2.3) by this inverse gives the expression for the *OLS* estimator $\widehat{\boldsymbol{\beta}}$:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \tag{2.4}$$

3 *OLS* Predictor and Residuals

The regression equation

$$\mathbf{y} = \mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{e}$$

separates the dependent variable into two components: the predicted part $\mathbf{X}\widehat{\boldsymbol{\beta}}$ and the residuals $\mathbf{e}$. Firstly, consider the *OLS* residuals

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}.$$

From the definition of the OLS estimator (2.4)

$$\begin{aligned}\mathbf{e} &= \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= (\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y} \\ &= \mathbf{M}\mathbf{y}\end{aligned}$$

where $\mathbf{M} = (\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$ is a *symmetric idempotent* matrix satisfying

$$\mathbf{M}\mathbf{M} = \mathbf{M} \quad \text{and} \quad \mathbf{M}'\mathbf{M} = \mathbf{M}.$$

$\mathbf{M}$ has the property that $\mathbf{M}\mathbf{X} = \mathbf{0}$ since

$$\begin{aligned}\mathbf{M}\mathbf{X} &= (\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{X} \\ &= \mathbf{X} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} \\ &= \mathbf{X} - \mathbf{X} = \mathbf{0}.\end{aligned}$$

Therefore, substituting from (1.1)

$$\begin{aligned}\mathbf{e} &= \mathbf{M}\mathbf{y} \\ &= \mathbf{M}(\boldsymbol{\beta} + \mathbf{u}) \\ &= \mathbf{M}\mathbf{u}.\end{aligned} \tag{3.1}$$

Finally, it is useful to determine the rank of the matrix $\mathbf{M}$. Since $\mathbf{M}$ is idempotent, it follows that $\text{rank}(\mathbf{M}) = \text{tr}(\mathbf{M})$ and by the rules of the algebra of traces

$$\begin{aligned}
\text{tr}(\mathbf{M}) &= \text{tr}(\mathbf{I}_n - \mathbf{X}\,(\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}') \\
&= \text{tr}(\mathbf{I}_n) - \text{tr}(\mathbf{X}\,(\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}') = \text{tr}(\mathbf{I}_n) - \text{tr}(\mathbf{X}'\mathbf{X}\,(\mathbf{X}'\mathbf{X})^{-1}) \\
&= \text{tr}(\mathbf{I}_n) - \text{tr}(\mathbf{I}_k) = n - k
\end{aligned}$$

so that

$$\text{rank}(\mathbf{M}) = n - k.$$

Consider now the OLS predictor

$$\mathbf{X}\widehat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{I} - \mathbf{M})\mathbf{y}.$$

Note that predictor and residuals are orthogonal components since

$$\mathbf{M}(\mathbf{I} - \mathbf{M}) = \mathbf{0}.$$

4 Properties of the OLS Estimator

Substituting the OLS expression (2.4) into the model (1.1) gives

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) \\
&= (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{u} \\
&= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{u} \qquad (4.1)
\end{aligned}$$

4.1 The OLS Estimator $\widehat{\boldsymbol{\beta}}$ is Unbiased

The property that the OLS estimator is unbiased or that

$$\text{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$

will now be proved.

Proposition 4.1. $\text{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$

Proof. Taking expectations of (4.1),

$$\begin{aligned}
\text{E}(\widehat{\boldsymbol{\beta}}) &= \boldsymbol{\beta} + \text{E}((\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\mathbf{u}) \\
&= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\,\mathbf{X}'\,\text{E}(\mathbf{u})
\end{aligned}$$

where, in the second line, $\mathbf{X}$ is taken out of the expectation because of assumption (A4), but, from assumption (A1), the second term is zero so that

$$\text{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta} \qquad (4.2)$$

and the estimator $\widehat{\boldsymbol{\beta}}$ is *unbiased.* □

4.2 The Variance of the OLS Estimator

From (4.1) and (4.2)

$$\widehat{\boldsymbol{\beta}} - \mathrm{E}(\widehat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$$

and it follows that

$$\begin{aligned} \mathrm{var}(\widehat{\boldsymbol{\beta}}) &= \mathrm{E}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\mathrm{E}(\mathbf{u}\mathbf{u}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

from assumption (A4). Then, using assumption (A2)

$$\begin{aligned} \mathrm{var}(\widehat{\boldsymbol{\beta}}) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\sigma^2\mathbf{I}_n)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

and, cancelling $\mathbf{X}'\mathbf{X}$ and its inverse $(\mathbf{X}'\mathbf{X})^{-1}$

$$\mathrm{var}(\widehat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}. \tag{4.3}$$

4.3 The OLS estimator of σ^2

The error variance σ^2 that appears in formula (4.3) is itself unknown and so in practice it needs to be estimated. We now show that the estimator

$$\widehat{\sigma}^2 = \frac{\mathbf{e}'\mathbf{e}}{n-k} \tag{4.4}$$

is an *unbiased* estimator of σ^2.

Note that from (3.1)

$$\mathbf{e}'\mathbf{e} = \mathbf{u}'\mathbf{M}'\mathbf{M}\mathbf{u} = \mathbf{u}'\mathbf{M}\mathbf{u}$$

by the idempotency and symmetry of $\mathbf{M}$.

Firstly, note that

$$\mathrm{E}(\mathbf{e}'\mathbf{e}) = \mathrm{E}(\mathrm{tr}(\mathbf{e}'\mathbf{e})) = \mathrm{tr}(\mathrm{E}(\mathbf{e}'\mathbf{e}))$$

since $\mathbf{e}'\mathbf{e}$ is a scalar expression and so is trivially equal to its trace. The second equality holds because both the expectations and trace operators are linear so that their order can be swapped and $\mathrm{E}(\mathrm{tr}(\mathbf{A})) = \mathrm{tr}(\mathrm{E}(\mathbf{A}))$. Hence

$$\begin{aligned} \mathrm{E}(\mathrm{tr}(\mathbf{e}'\mathbf{e})) &= \mathrm{E}(\mathrm{tr}(\mathbf{u}'\mathbf{M}\mathbf{u})) \\ &= \mathrm{E}(\mathrm{tr}(\mathbf{M}\mathbf{u}\mathbf{u}')) = \mathrm{tr}(\mathrm{E}(\mathbf{M}\mathbf{u}\mathbf{u}')) \end{aligned}$$

using the result that $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$.

But by assumption (A4), $\mathbf{M}$ is fixed and can be taken out of the expectation so that

$$\mathrm{tr}(\mathrm{E}(\mathbf{Muu}')) = \mathrm{tr}(\mathbf{M}\,\mathrm{E}(\mathbf{uu}')) = \mathrm{tr}(\sigma^2\mathbf{M})$$

by assumption (A2). Finally, since we have shown that $\mathrm{tr}(\mathbf{M}) = n - k$, it follows that

$$\mathrm{E}(\mathbf{e}'\mathbf{e}) = \sigma^2(n-k)$$

or that

$$\mathrm{E}\left(\frac{\mathbf{e}'\mathbf{e}}{n-k}\right) = \sigma^2$$

which proves that $\widehat{\sigma}^2$ is an unbiased estimator of the error variance σ^2.

5 The Gauss-Markov Theorem

Definition 5.1. *A linear estimator is one that can be written in the form*

$$\widetilde{\boldsymbol{\beta}} = \mathbf{Cy}$$

where $\mathbf{C}$ *is a* $k \times n$ *matrix of fixed constants.*

Note that the OLS estimator $\widehat{\boldsymbol{\beta}}$ is a linear estimator with

$$\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'.$$

Theorem 5.1. *The* OLS *estimator* $\widehat{\boldsymbol{\beta}}$ *is the* Best Linear Unbiased Estimator *(*BLUE*) of the classical regresssion model. By* ***best*** *we mean the estimator in the class that achieves* ***minimum variance****.*

Proof. Taking expectations

$$\mathrm{E}(\widetilde{\boldsymbol{\beta}}) = \mathbf{C}\,\mathrm{E}(\mathbf{y}) = \mathbf{C}\,\mathrm{E}(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) = \mathbf{CX}\boldsymbol{\beta} + \mathbf{C}\,\mathrm{E}(\mathbf{u})$$

so the condition for unbiasedness of $\widetilde{\boldsymbol{\beta}}$ is that

$$\mathbf{CX} = \mathbf{I}_k$$

The variance of the estimator $\widetilde{\boldsymbol{\beta}}$ is given by

$$\begin{aligned} \mathrm{var}(\widetilde{\boldsymbol{\beta}}) &= \mathrm{var}(\mathbf{CX}\boldsymbol{\beta} + \mathbf{Cu}) = \mathrm{var}(\mathbf{Cu}) \\ &= \mathrm{E}(\mathbf{Cuu}'\mathbf{C}') = \sigma^2\mathbf{CC}' \end{aligned}$$

by assumption (A2).

To prove that *OLS* is the *best* in the class of unbiased estimators it is necessary to show that the matrix

$$\text{var}(\widetilde{\boldsymbol{\beta}}) - \text{var}(\widehat{\boldsymbol{\beta}})$$

is *positive semi-definite*. We use the result that for any matrix $\mathbf{A}$, the matrix products $\mathbf{A}'\mathbf{A}$ and $\mathbf{A}\mathbf{A}'$ are both positive semi-definite. This is easy to show since the quadratic form $\mathbf{x}'\mathbf{A}'\mathbf{A}\mathbf{x}$ can be written as $\mathbf{z}'\mathbf{z}$ where $\mathbf{z} = \mathbf{A}\mathbf{x}$ and $\mathbf{z}'\mathbf{z} = \sum z_i^2 \geq 0$ for all $\mathbf{x}$. Similarly, $\mathbf{x}'\mathbf{A}\mathbf{A}'\mathbf{x}$ can be written as $\mathbf{w}'\mathbf{w}$ where $\mathbf{w} = \mathbf{A}'\mathbf{x}$ and $\mathbf{w}'\mathbf{w} = \sum w_i^2 \geq 0$ for all $\mathbf{x}$.

We use a trick and write

$$\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{D}$$

where $\mathbf{D}$ is the difference between an arbitrary estimator and the OLS estimator. Note that

$$\mathbf{C}\mathbf{X} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} + \mathbf{D}\mathbf{X} = \mathbf{I}_k + \mathbf{D}\mathbf{X}$$

but since, for unbiasedness $\mathbf{C}\mathbf{X} = \mathbf{I}_k$, it follows that $\mathbf{D}\mathbf{X} = \mathbf{0}$. Consequently,

$$\begin{aligned}
\mathbf{C}\mathbf{C}' &= [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + \mathbf{D}][\mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}} + \mathbf{D}'] \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}} + \mathbf{D}\mathbf{D}' \\
&= (\mathbf{X}'\mathbf{X})^{-\mathbf{1}} + \mathbf{D}\mathbf{D}'
\end{aligned}$$

so that

$$\begin{aligned}
\text{var}(\widetilde{\boldsymbol{\beta}}) &= \sigma^2\mathbf{C}\mathbf{C}' \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-\mathbf{1}} + \sigma^{\mathbf{2}}\mathbf{D}\mathbf{D}' \\
&= \text{var}(\widehat{\boldsymbol{\beta}}) + \sigma^{\mathbf{2}}\mathbf{D}\mathbf{D}'
\end{aligned}$$

or

$$\text{var}(\widetilde{\boldsymbol{\beta}}) - \text{var}(\widehat{\boldsymbol{\beta}}) = \sigma^{\mathbf{2}}\mathbf{D}\mathbf{D}'$$

which proves the result since $\mathbf{D}\mathbf{D}'$ must be positive semi-definite. □

6 Hypothesis Testing

6.1 The Joint Normal Distribution

A $p \times 1$ vector of random variables $\mathbf{x}$ that follows the distribution function

$$f(\mathbf{x}) = (2\pi)^{-\frac{p}{2}} \left|\boldsymbol{\Sigma}\right|^{-\frac{1}{2}} \exp(-\frac{1}{2}(\mathbf{x} - \mu)'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \mu))$$

is said to be jointly normally distributed with mean $\boldsymbol{\mu}$ and variance covariance matrix $\boldsymbol{\Sigma}$, written as

$$\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

In the case that $p = 1$ this reduces to the one-dimensional normal density function. Note that any linear function of $\mathbf{x}$, $\mathbf{Ax}$, will also be jointly normally distributed with distribution given by

$$\mathbf{Ax} \sim N(\mathbf{A}\mu, \mathbf{A\Sigma A}').$$

6.2 The Chi-squared distribution

Suppose that $\mathbf{x}$ is a $p \times 1$ vector of random variables, jointly normally distributed with

$$\mathbf{x} \sim N(\mathbf{0}, \mathbf{I}_p).$$

Then the scalar sum of squares $\mathbf{x}'\mathbf{x}$ is distributed with a chi-squared distribution with p degrees of freedom or algebraically

$$\mathbf{x}'\mathbf{x} \sim \chi^2_p.$$

It follows that, more generally, if $\mathbf{x} \sim N(\mathbf{0}, q^2\mathbf{I}_p)$, then $(\mathbf{x}'\mathbf{x}/q^2 \sim \chi^2_p$.

We now prove the result that, if $\mathbf{A}$ is a *symmetric idempotent* matrix, then the quadratic form $\mathbf{x}'\mathbf{Ax}/q^2 \sim \chi^2_r$ where r is the rank of the matrix $\mathbf{A}$. Consider the eigenvalue decomposition of $\mathbf{A}$

$$\mathbf{A} = \mathbf{V\Lambda V}'$$

where $\mathbf{V}$ is the $p \times p$ matrix of eigenvectors of $\mathbf{A}$ satisfying $\mathbf{V}'\mathbf{V} = \mathbf{I}_p$, and $\boldsymbol{\Lambda}$ is the $p \times p$ diagonal matrix of associated eigenvalues. Then the quadratic form

$$\mathbf{x}'\mathbf{Ax} = \mathbf{x}'\mathbf{V\Lambda V}'\mathbf{x} = \mathbf{z}'\boldsymbol{\Lambda}\mathbf{z}$$

where $\mathbf{z} = \mathbf{V}'\mathbf{x}$ is a $p \times 1$ vector with normal distribution given by

$$\mathbf{z} = \mathbf{V}'\mathbf{x} \sim N(\mathbf{0}, q^2\mathbf{V}'\mathbf{V}) = N(\mathbf{0}, q^2\mathbf{I}_p).$$

However, since A is symmetric idempotent, all its eigenvalues are either equal to zero or one, with the number of unit eigenvalues being equal to the rank of $\mathbf{A}$, which is r. Thus

$$\mathbf{z}'\boldsymbol{\Lambda}\mathbf{z} = \sum_{i=1}^{p} \lambda_i z_i^2 = \sum_{i=1}^{r} z_i^2 = \mathbf{z}_1'\mathbf{z}_1$$

where the eigenvalues have been ordered so that the first r are equal to one and the last $p-r$ equal to zero, and $\mathbf{z}_1$ is an $r \times 1$ vector comprising the first r elements of $\mathbf{z}$. It follows that

$$\frac{\mathbf{x}'\mathbf{Ax}}{q^2} = \frac{\mathbf{z}_1'\mathbf{z}_1}{q^2} \sim \chi^2_r.$$

6.3 The F Distribution

Two chi-squared distributions of idempotent quadratic forms $\mathbf{x}'\mathbf{A}\mathbf{x}$ and $\mathbf{x}'\mathbf{B}\mathbf{x}$ are defined to be *independent* if

$$\mathbf{AB} = \mathbf{BA} = \mathbf{0}.$$

Consider the ratio of two *independent* chi-squared variates with m and n degrees of freedom, respectively. It can be shown that

$$\frac{\frac{1}{m}\chi_m^2}{\frac{1}{n}\chi_n^2} \sim F_{m,n}$$

where $F_{m,n}$ is *Fisher's F distribution* with m and n degrees of freedom.

Note that if x is distributed as $F_{1,n}$ then $\sqrt{x}$ is distributed as t_n where t_n is the *Student t distribution* with n degrees of freedom.

6.4 Distributions of the *OLS* Parameter Estimates

In order to make statistical inferences on the parameter estimates $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}^2$we must add a further assumption to the classical regression model:

$$\mathbf{u} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n) \tag{A5}$$

which is that the error vector $\mathbf{u}$ is distributed jointly normally. It follows that, since $\mathbf{y}$ and $\widehat{\boldsymbol{\beta}}$ are both linear combinations of $\mathbf{u}$, they are both also distributed normally with

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)$$

and

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \sim N(\mathbf{0}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}). \tag{6.1}$$

To make this practical we need to replace the unknown parameter σ^2 with the estimator $\widehat{\sigma}^2$ defined in (4.4) which was shown above to be equal to

$$\widehat{\sigma}^2 = \frac{\mathbf{e}'\mathbf{e}}{n-k} = \frac{\mathbf{u}'\mathbf{M}\mathbf{u}}{n-k}$$

where the matrix $\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is symmetric and idempotent with rank $n-k$. It follows from assumption (A5) and the result just proved on idempotent quadratic forms in normal variables that

$$(n-k)\frac{\widehat{\sigma}^2}{\sigma^2} \sim \chi_{n-k}^2. \tag{6.2}$$

where χ_{n-k}^2 is the Chi-squared distribution with $n-k$ degrees of freedom.

6.5 Jointly testing all the coefficients

Consider a test of the hypothesis that

$$H_0 : \boldsymbol{\beta} - \boldsymbol{\beta}_0 = \mathbf{0} \tag{6.3}$$

for some set constants $\boldsymbol{\beta}_0$. This hypothesis involves all of the coefficients in the model. A joint test on all the components of the estimated parameter vector $\widehat{\boldsymbol{\beta}}$ can be based on the expression

$$\begin{aligned} & (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\ = \; & \mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \\ = \; & \mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} = \mathbf{u}'(\mathbf{I} - \mathbf{M})\mathbf{u}. \end{aligned}$$

This is a symmetric idempotent quadratic form in the normally distributed error vector $\mathbf{u}$. It follows that

$$\frac{(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{\sigma^2} \sim \chi^2_k. \tag{6.4}$$

This expression still involves the unknown error variance σ^2 and so is not operational. However, the chi-squared statistic (6.4) is independent of the chi-squared statistic (6.2) since the idempotent matrices $\mathbf{M}$ and $\mathbf{I} - \mathbf{M}$ satisfy

$$\mathbf{M}(\mathbf{I} - \mathbf{M}) = \mathbf{0}.$$

Consequently, the ratio of the two chi-squared statistics (6.4) and (6.2), divided by their degrees of freedom

$$\begin{aligned} & \frac{(\widehat{\beta} - \beta)'\mathbf{X}'\mathbf{X}(\widehat{\beta} - \beta)/\sigma^2}{(n-k)\widehat{\sigma}^2/\sigma^2}\frac{n-k}{k} \\ = \; & (\widehat{\beta} - \beta)'\frac{1}{\widehat{\sigma}^2}\mathbf{X}'\mathbf{X}(\widehat{\beta} - \beta)/k \sim F_{k,n-k} \end{aligned}$$

has an *F distribution* with k and $n - k$ degrees of freedom. This statistic is operational since the unknown error variance cancels from both numerator and denominator. Note that the expression $\mathbf{X}'\mathbf{X}/\widehat{\sigma}^2$ in the centre of this quadratic form is the inverse of the estimated parameter variance covariance matrix

$$\widehat{\text{var}}(\widehat{\beta}) = \widehat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}. \tag{6.5}$$

hence, on the null hypothesis (6.3)

$$\frac{1}{k}(\widehat{\beta} - \beta_0)'(\widehat{\text{var}}(\widehat{\beta}))^{-1}(\widehat{\beta} - \beta_0) \sim F_{k,n-k}.$$

6.6 Testing a linear combination of the coefficients

Now consider tests involving a linear combination of the coefficients

$$H_0 : \mathbf{w}'\beta = w_0 \tag{6.6}$$

where $\mathbf{w}$ is a $k \times 1$ vector of constants and w_0 is a scalar. We know that

$$\mathbf{w}'(\widehat{\beta} - \beta) \sim N(0, \sigma^2 \mathbf{w}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{w})$$

so that

$$\frac{\mathbf{w}'\widehat{\beta} - \mathbf{w}'\beta}{\sigma\sqrt{\mathbf{w}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{w}}} \sim N(0,1).$$

From (6.2) it follows that the ratio

$$\begin{aligned} & \frac{\frac{\mathbf{w}'\widehat{\beta} - \mathbf{w}'\beta}{\sigma\sqrt{\mathbf{w}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{w}}}\sqrt{n-k}}{\sqrt{(n-k)\widehat{\sigma}^2/\sigma^2}} \\ = & \frac{\mathbf{w}'\widehat{\beta} - \mathbf{w}'\beta}{\widehat{\sigma}\sqrt{\mathbf{w}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{w}}} \sim t_{n-k} \end{aligned}$$

follows a Student t distribution with $n - k$ degrees of freedom.

Hence on the null hypothesis (6.6) it follows that

$$= \frac{\mathbf{w}'\widehat{\beta} - w_0}{\widehat{\sigma}\sqrt{\mathbf{w}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{w}}} \sim t_{n-k} \tag{6.7}$$

A particular example of this test is a test of the hypothesis that $\beta_j = 0$. In this case the vector $\mathbf{w}$ consists of zeros except for a one in the jth position, and the scalar w_0 is equal to zero. The denominator in (6.7) in this case will pick out the square root of the jjth diagonal element in the estimated variance covariance matrix (6.5) which is the *estimated standard error* of the coefficient β_j. Thus the test statistic is the familiar t-ratio

$$\frac{\widehat{\beta}_j}{s\,\mathrm{E}(\widehat{\beta}_j)} \sim t_{n-k}.$$