Econometrics Lecture Notes (I)

Montserrat Farell

Universitat Autònoma de Barcelona

21st November 2003

Contents

General Information

Prof. M. Farell

Office: B3-192

Tel. 581-2932

Web page

There is a class web page which will serve as the distribution point for the class notes and materials for the problem sets. To access it, point your favorite browser at ***pareto.uab.es/mfarell/econometria***. The Facultat has a computer room in Aules 21-22-23 where you can find plenty of computers connected to the internet.

1 Economic and econometric models

Economic theory postulates a mathematical relashionship between some economic variables

$$y = f(X),$$

say x is a $kx1$vector ($k = 2$ for the two variable model) . It is a parametric relationship which describes the economic agents' behaviour, and when applying economic analysis to policy questions one can get estimations of the parameters of interest using econometric technics. For example, we can be interested on:

- the input elasticities of substitution in a production or a cost function
- the direct eleasticity quantity-price in a demand function
- the returns to scale a technology exhibits in a production or a cost function
- the marginal propensity to consume in a consumtion function, among others.

Economic theory tells us some characteristics that f should have.

Economic theory tells us which are the vaiables in x.

The econometrician will try to "guess" f and, the parameters of the function from the data gathered of y and x.

The econometric model will be the representation of the economic model. We will consider first that this relation is linear (or that we can make it linear).

$y = f(X, \beta) + \varepsilon$

- y is a $(nx1)$ vector
- x is (nxk)vector
- β is a $(kx1)$ vector

- ε is a $(nx1)$ vector. It is the error term and is interpreted as the sum of several independent effects (variables) not controled by the econometrician.

For the economist the most interesting cases will be the ones where y and X are highly quality data and $y = f(X)$ has a solid theoretical base. Examples can come from two big pieces in the economic theory literature:

a) the theory of the consumer

b) the theory of the firm

a) From utility maximization we obtain the demand functions $x_j^* = (p, M)$, the demand function should satisfy some regularity conditions, such as homogeneity of degree zero on p and M, symetry effects etc.

b) From cost minimization we obtain the indirect cost function $C_i = f(p, Q)$, it is a representation of the technology, from duality theory it recovers all the parameters of the production function. The cost function also should satisfy some conditions, such as homogeneity of degree one on input prices, p, non-decreasing on p, continuous with respect to p etc.

When doing econometric modelling: $y = f(X, \beta) + \varepsilon$ the econometrician needs to specify:

- Which variables X belong to the model

  economic theory

  common sense

- Which is the functional form f

  economic theory gives us the regularity conditions but there are a lot of functional forms that satisfy those conditions

- Which is the error specification $\varepsilon \sim$

  analysis of the context

Origin of the word regression

The linear model presented nowadays, what is understood by the *Gauss linear model* or *the linear regression model* has its main contributions on the work of Halley (1656-1742), De Moivre (1667-1754), Bernoulli (1700-1782), Bayes (1702-1761), Lagrange (1736-1813), Laplace (1749-1827), Legendre (1752-1833), Gauss (1789-1857). Afterwards, with Galton (1822-1911), Edgeworth (1845-1926), Pearson (1857-1936) and Yule (1871-1951) the convergence of descriptuve statistics and the calculus of probabilities in the context of the Gauss linear model was a reality (Spanos 1986; Statistical foundations of Econometric Modelling. *A brief historical overview*. Cambridge University Press).

But the use of the word regression in this context has its origins in the experiments of Galton, when trying to see the relationship between height of parents and height of their children. It was clear that tall parents had tall children and short parents had short children. Nevertheless, he discovered that the height of children from unusually tall or unusually short parents would have the tendency to go to the average of the polpulation height. He did the experiment fitting the line through the data points that would minimize the sum of the squared of the distances between the points and the line. He called the phenomenon explained above ”regression to the mean” in his words ”regression to mediocrity” (1886) that is the unconditional mean.

2 The classical linear model

The econometric model is a representation of the economic model.

$$y = f(X,\beta) + \varepsilon$$

- the variable under study is the left-hand side, *the dependent variable y*.

- this variable is explained by (related to) several other variables, the right-hand side, *the independent, explanatory, variables, the regressors X.*

- y_i is the $i-th$ value of the dependent variable.

- $x_i = (x_{i1}, x_{i2})$ is the $i-th$ observation of the k regressors. Where $x_{i1} = 1$.

Data in economics is not experimental so, y and x can be both treated as random variables or y is random and x are fixed values.

- the model is a linear function of the parameter vector β :

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i,$$

$$i = 1, 2, ..., n$$

- $\beta_1 x_{i1} + \beta_2 x_{i2}$ is the regression line, the $\beta' s$ are the regression coefficients

- $\beta_j = \frac{\partial y_i}{\partial x_{ij}}$; $j = 1, 2$ because of linearity the marginal effect does not depend on the level of the regressors.

- the error term is the part unexplained by the regressors. It is the sum of the variables not under the control of the econometrician.

Example : The consumption function

$$con_i = \beta_1 + \beta_2 DI_i + \varepsilon_i$$

consumption is a linear function of disposable income. In the general notation we used $x_{i1} = 1$for every i. Question: why is it important to introduce a column of ones?

Example 2: A semi-log or log-linear wage equation

$$wage_i = e^{\beta_1} e^{\beta_2 S_i} e^{\varepsilon_i}$$

- $wage_i$ is the wage rate for individual i
- S_i education in years for individual i

Taking logs in both sides leads to the semi-log model:

$$log(wage_i) = \beta_1 + \beta_2 S_i + \varepsilon_i$$

the coefficients in this model are percentatge changes, in a general two variable case we have:

$$log(y_i) = \beta_1 + \beta_2 x_{i2} + \varepsilon_i$$

$\beta_2 = \frac{\partial log(y_i)}{\partial x_{i2}} * \frac{1}{y_i} = \frac{\frac{\partial y_i}{y_i}}{\partial x_{i2}}$ =*relative change in* y_i */ absolute change in* x_{i2}

Example 3: A Cobb-Douglas cost function

By duality a cost function is a dual of a production function and it represents the technology. The indirect cost function gives us, for every level of output and a set of prices, the minimum cost you can produce that output with the represented technology.

$$c = f(p_k, p_l, Q)$$

the Cobb-Douglas econometric model representation is:

$$c = A * p_k^{\beta_k} * p_l^{\beta_l} * Q^{\beta_Q} * e^{\varepsilon}$$

for a set of firms $i = 1, 2..., n$ and taking logarithms in both sides we obtain the following log. model

$$log(c_i) = logA + \beta_k log(p_{ik}) + \beta_l log(p_{il}) + \beta_Q log(Q_i) + \varepsilon_i$$

in a general two varaible case model

$$log(y_i) = \beta_1 + \beta_2 log(x_{i2}) + \varepsilon_i$$

$\beta_2 = \frac{\partial log(y_i)}{\partial log(x_{i2})} = \frac{\frac{\partial y_i}{y_i}}{\frac{\partial x_{i2}}{x_{i2}}} = \frac{\partial y_i}{\partial x_{i2}} * \frac{x_{i2}}{y_i} = \xi$ an elasticity.

2.1 The model in Matrix Notation:

$$y = X\beta + \varepsilon;$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix}; \quad X = \begin{bmatrix} 1 & x_{12} \\ 1 & x_{22} \\ . & . \\ . & . \\ . & . \\ 1 & x_{n2} \end{bmatrix}; \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}; \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{bmatrix};$$

Assumptions:

1. ε_i is normally distributed
2. $E(\varepsilon_i) = 0$; (zero mean)
3. $V(\varepsilon_i) = \sigma^2; \forall i$; (homoscedasticity)
4. $E(\varepsilon_i \varepsilon_j) = 0$; (no-correlation)
5. X is fixed, non random

In matrix notation:

1. $\varepsilon \sim N$
2. $E(\varepsilon) = \begin{bmatrix} E(\varepsilon_1) \\ E(\varepsilon_2) \\ . \\ . \\ . \\ E(\varepsilon_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ . \\ . \\ . \\ 0 \end{bmatrix};$
3. and 4. $E(\varepsilon\varepsilon') = \sigma^2 I_n;$

$$\varepsilon\varepsilon' = \begin{bmatrix} \varepsilon_1 & \varepsilon_2 & . & . & \varepsilon_n \end{bmatrix} \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ \varepsilon_n \end{bmatrix}$$

$$= \begin{bmatrix} \varepsilon_1^2 & \varepsilon_1\varepsilon_2 & . & . & \varepsilon_1\varepsilon_n \\ \varepsilon_2\varepsilon_1 & \varepsilon_2^2 & . & . & \varepsilon_2\varepsilon_n \\ . & . & . & . & . \\ . & . & . & . & . \\ \varepsilon_n\varepsilon_1 & \varepsilon_n\varepsilon_2 & . & . & \varepsilon_n^2 \end{bmatrix}$$

$$E(\varepsilon\varepsilon') = \begin{bmatrix} \sigma^2 & 0 & . & . & 0 \\ 0 & \sigma^2 & 0 & . & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & 0 & 0 & \sigma^2 \end{bmatrix} = \sigma^2 I_n;$$

therefore in matrix notation we can write:

$y = X\beta + \varepsilon;$

$\varepsilon \sim N(0, \sigma^2 I_n);$

X non-stochastic;

2.2 Estimation by least squares

Definitions:

- β the vector of unknown parameters

- $\tilde{\beta}$ any value for the parameters

- $\hat{\beta}$ the OLS estimate for β

- $\tilde{\varepsilon}_i = y_i - x_i'\tilde{\beta}$ residual for observation i

$$\begin{aligned} \hat{\beta} &= \arg\min SSR(\tilde{\beta}) = \sum_{i=1}^{n} \left(y_i - x_i'\tilde{\beta}\right)^2 \\ s(\tilde{\beta}) = SSR(\tilde{\beta}) &= \left(y - X\tilde{\beta}\right)' \left(y - X\tilde{\beta}\right) \\ &= y'y - 2y'X\tilde{\beta} + \beta' X' X\tilde{\beta} \end{aligned}$$

To minimize the criterion $SSR(\tilde{\beta})$, take the f.o.n.c. and set them to zero:

$$D_{\tilde{\beta}} s(\hat{\beta}) = \frac{\partial SSR(\hat{\beta})}{\partial \tilde{\beta}} \quad = -2X'y + 2X'X\hat{\beta} \quad = 0$$

properties used

$$i)\ \frac{\partial a'x}{\partial x} \quad = \quad a$$

$$ii)\ \frac{\partial x'Ax}{\partial x} = 2Ax;$$

if A symetric,

so

$$\hat{\beta} = (X'X)^{-1}X'y.$$

To verify that this is a minimum, check the s.o.s.c.:

$$D^2_{\tilde{\beta}} s(\widehat{\beta}) = \frac{\partial^2_{\tilde{\beta}} SSR(\widehat{\beta})}{\partial \widehat{\beta} \partial \widehat{\beta}'} = 2X'X$$

Since $\rho(X) = k$, this matrix is positive definite, since it's a quadratic form in a p.d. matrix (identity matrix of order n), so $\hat{\beta}$ is in fact a minimizer.

- The *fitted values* are in the vector $\hat{y} = X\hat{\beta}$.
- The *residuals* are in the vector $\hat{\varepsilon} = y - X\hat{\beta}$

Note that

$$\begin{aligned} y &= X\beta + \varepsilon \\ &= X\hat{\beta} + \hat{\varepsilon} \end{aligned}$$

2.3 Properties of the OLS estimator in small samples:

2.3.1 Unbiasedness

Under assumptions 1-5:

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'y \\ &= (X'X)^{-1}X'(X\beta + \varepsilon) \\ &= \beta + (X'X)^{-1}X'\varepsilon \\ \mathcal{E}(\hat{\beta}) &= \beta. \end{aligned}$$

2.3.2 Efficiency (Gauss-Markov theorem)

The variance-covariance matrix of the OLS estimator

$$V(\widehat{\beta}) = E((\widehat{\beta} - \beta)(\widehat{\beta} - \beta)')$$

$$
\begin{aligned}
V(\widehat{\beta}) &= E((\widehat{\beta}-\beta)(\widehat{\beta}-\beta)') \\
&= E((X'X)^{-1}X'\varepsilon)(X'X)^{-1}X'\varepsilon)') \\
&= E((X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}) \\
&= (X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1} \\
&= (X'X)^{-1}X'\sigma^2 I_n X(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}
\end{aligned}
$$

therefore the distribution of $\widehat{\beta}$ is $N(\beta, \sigma^2(X'X)^{-1})$.

Asside: A usufull way to derive the distributions of vectors of random variables which are linear combinations of a normaly distributed random vector is aplying the following proposition:

If x is a $nx1$random vector where

$$
\begin{aligned}
x &\sim N(\mu, \Sigma) \\
\theta = a + Ax &\sim N(a + A\mu, A\Sigma A')
\end{aligned}
$$

Definition:

We say that $\hat{\beta}$ is efficient for β if $\hat{\beta}$ is unbiassed for β and

$$
V\left(\hat{\beta} \leq \tilde{\beta}\right)
$$

where $\tilde{\beta}$ is any other unbiassed estimator of β.

The Gauss-Markov Theorem:

The OLS estimator is a *linear estimator*, which means that it is a linear function of the dependent variable, y.

$$\begin{aligned}\hat{\beta} &= \left[(X'X)^{-1}X'\right] y \\ &= Cy\end{aligned}$$

It is also *unbiased*, as we proved above. One could consider other weights W in place of the OLS weights. We'll still insist upon unbiasedness. Consider $\tilde{\beta} = Wy$. If the estimator is unbiased

$$\begin{aligned}\mathcal{E}(Wy) &= \mathcal{E}(WX\beta + W\varepsilon) \\ &= WX\beta \\ &= \beta \\ &\Rightarrow \\ WX &= I_K\end{aligned}$$

The variance of $\tilde{\beta}$ is

$$V(\tilde{\beta}) = WW'\sigma^2.$$

Define

$$D = W - (X'X)^{-1}X'$$

so

$$W = D + (X'X)^{-1}X'$$

Since $WX = I_K$, $DX = 0$, so

$$\begin{aligned}V(\tilde{\beta}) &= \left(D + (X'X)^{-1}X'\right)\left(D + (X'X)^{-1}X'\right)'\sigma^2 \\ &= \left(DD' + (X'X)^{-1}\right)\sigma^2\end{aligned}$$

So

$$V(\tilde{\beta}) \geq V(\hat{\beta}).$$

This is a proof of the Gauss-Markov Theorem.

Theorem 1 (Gauss-Markov) *Under the classical assumptions, the variance of any linear unbiased estimator minus the variance of the OLS estimator is a positive semidefinite matrix.*

- It is worth noting that we have not used the normality assumption in any way to prove the Gauss-Markov theorem, so it is valid if the errors are not normally distributed, as long as the other assumptions hold.

Estimation of σ^2 :

For $\hat{\sigma}^2$ we have

$$
\begin{aligned}
\widehat{\sigma^2} &= \frac{1}{n-K}\hat{\varepsilon}'\hat{\varepsilon} \\
\hat{\varepsilon} &= y - X\hat{\beta} \\
&= y - X(X'X)^{-1}X'y \\
&= M_X y \\
&= M_X(X\beta + \varepsilon) \\
&= M_X\varepsilon \\
\mathcal{E}(\widehat{\sigma^2}) &= \frac{1}{n-K}\mathcal{E}(\varepsilon' M\varepsilon) \\
&= \frac{1}{n-K}\mathcal{E}(TrM\varepsilon\varepsilon') \\
&= \frac{1}{n-K}Tr\mathcal{E}(M\varepsilon\varepsilon') \\
&= \frac{1}{n-K}\sigma^2 TrMI_n \\
&= \frac{1}{n-K}\sigma^2\left(n - TrX(X'X)^{-1}X'\right) \\
&= \frac{1}{n-K}\sigma^2\left(n - Tr(X'X)^{-1}X'X\right) \\
&= \sigma^2
\end{aligned}
$$

$\hat{\sigma}^2$ is an unbiassed estimator for σ^2.

2.4 Goodness of fit

We are trying to find an statistic that would measure how the estimated regression line fits the data.

From the first order conditions we have that

$$
\begin{aligned}
x'x\widehat{\beta} &= x'y \\
x'y - x'x\widehat{\beta} &= \overrightarrow{0} \\
x'(y - x'\widehat{\beta}) &= \overrightarrow{0} \\
x'\widehat{\varepsilon} &= \overrightarrow{0}
\end{aligned}
$$

from this we obtain:

$$
\begin{aligned}
\Sigma\widehat{\varepsilon_i} &= \overrightarrow{0} \\
\Sigma x_{i2}\widehat{\varepsilon_i} &= \overrightarrow{0}
\end{aligned}
$$

we will write the model in deviations from the mean:

$$
\begin{aligned}
y_i^* &= y_i - \overline{y} \\
x_{i2}^* &= x_{i2} - \overline{x_2}
\end{aligned}
$$

the model can be written as:

$$
y_i^* = \widehat{\beta_2} x_{i2} + \widehat{\varepsilon_i}
$$

we square the equation and we sum over i:

$$
\begin{aligned}
y_i^{*2} &= (\widehat{\beta_2} x_{i2}^* + \widehat{\varepsilon_i})^2 \\
y_i^{*2} &= \widehat{\beta_2}^2 x_{i2}^{*2} + \widehat{\varepsilon_i}^2 + 2\widehat{\beta_2} x_{i2}^* \widehat{\varepsilon_i} \\
\Sigma y_i^{*2} &= \Sigma\widehat{\beta_2}^2 x_{i2}^{*2} + \Sigma\widehat{\varepsilon_i}^2 + \Sigma 2\widehat{\beta_2} x_{i2}^* \widehat{\varepsilon_i}
\end{aligned}
$$

last term is iqual to zero from f.o.c.:

$$\begin{aligned}\Sigma(x_{i2}-x_2)\widehat{\varepsilon}_i &= \Sigma x_{i2}\widehat{\varepsilon}_i - \Sigma\overline{x_2}\widehat{\varepsilon}_i \\ &= 0\end{aligned}$$

the decomposition of the variability of y is then:

$$\begin{aligned}\Sigma(y_i-\overline{y})^2 &= \Sigma(\widehat{y}_i-\overline{y})^2+\Sigma(y_i-\widehat{y}_i)^2 \\ TSS &= ESS+RSS \\ 1 &= \frac{ESS}{TSS}+\frac{RSS}{TSS}\end{aligned}$$

the equivalence $\Sigma\widehat{\beta_2}^2 x_{i2}^{*2} = \Sigma(\widehat{y}_i-\overline{y})^2$is easy to see graphically. See notes from class.

$$\begin{aligned}R^2 &= \frac{ESS}{TSS}\ or \\ &= 1-\frac{RSS}{TSS} \\ &= 1-\frac{\widehat{\varepsilon}'\widehat{\varepsilon}}{y^{*\prime}y^*}\end{aligned}$$

we have then a measure for the goodness of fit R^2,

$$0 \leq R^2 \leq 1$$

3 Restrictions

Restrictions are statements about the values of the parameters of our model or functions of them. In this course we will only consider statements about the values of our parameters or linear functions of them. We will be interested on testing those statements or imposing them in our model. Therefore, we will be talking about the unrestricted model (that is, the model before any restriction is considered) and the restricted model (the model that satisfies the restriction)

3.1 Linear Hypothesis testing under normality

We have now, one or more linear equations on the parameters of the model. These equations are often motivated by the same economic theory on which the model is based or they are based on statements about the explanatory capacity of the variables in the model.

- the restriction to be tested is called the null hypothesis, H_0.
- the model is "the mantained hypothesis", a set of assumptions.
- the model together with the null produces some test-statistic with a known distribution.
- too large of a value of the test statistic is interpreted as a failure of the null. This interpretation is only valid if the model is correctly specified.
- the test statistic may not have the supposed distribution when the null is true but the model is false.

3.1.1 Exact linear restrictions

In many cases, economic theory suggests restrictions on the parameters of a model. For example, a demand function is supposed to be homogeneous of degree zero in prices and income. If we have a Cobb-Douglas (log-linear) model,

$$\ln q = \beta_0 + \beta_1 \ln p_1 + \beta_2 \ln p_2 + \beta_3 \ln m + \varepsilon,$$

then we need that

$$k^0 \ln q = \beta_0 + \beta_1 \ln k p_1 + \beta_2 \ln k p_2 + \beta_3 \ln k m + \varepsilon,$$

so

$$\begin{aligned}
\beta_1 \ln p_1 + \beta_2 \ln p_2 + \beta_3 \ln m &= \beta_1 \ln k p_1 + \beta_2 \ln k p_2 + \beta_3 \ln k m \\
&= (\ln k)(\beta_1 + \beta_2 + \beta_3) + \beta_1 \ln p_1 + \beta_2 \ln p_2 + \beta_3 \ln m.
\end{aligned}$$

The only way to guarantee this for arbitrary k is to set

$$\beta_1 + \beta_2 + \beta_3 = 0,$$

which is a *parameter restriction.* In particular, this is a linear equality restriction, which is probably the most commonly encountered case.

3.1.2 Imposition

The general formulation of linear equality restrictions is the model

$$\begin{aligned}
y &= X\beta + \varepsilon \\
R\beta &= r
\end{aligned}$$

where R is a $Q \times K$ matrix, $Q < K$ and r is a $Q \times 1$ vector of constants.

- We assume R is of rank Q, so that there are no redundant restrictions.
- We also assume that $\exists\beta$ that satisfies the restrictions: they aren't infeasible.

  Draw a picture for two var. model with $\beta = 1$ *as a restriction to motivate the relevant distances for an statistic to test the restrictions.*

Let's consider how to estimate β subject to the restrictions $R\beta = r.$ The most obvious approach is to set up the Lagrangean

$$\min_{\tilde{\beta}} s(\tilde{\beta}) = \left(y - X\tilde{\beta}\right)' \left(y - X\tilde{\beta}\right) + 2\lambda'(R\tilde{\beta} - r).$$

The Lagrange multipliers are scaled by 2, which makes thing less messy. The f.o.n.c. are

$$\begin{aligned} D_{\tilde{\beta}} s(\hat{\beta}, \hat{\lambda}) &= -2X'y + 2X'X\hat{\beta}_R + 2R'\hat{\lambda} \equiv 0 \\ D_{\lambda} s(\hat{\beta}, \hat{\lambda}) &= R\hat{\beta}_R - r \equiv 0, \end{aligned}$$

which can be written as

$$\begin{bmatrix} X'X & R' \\ R & 0 \end{bmatrix} \begin{bmatrix} \hat{\beta}_R \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} X'y \\ r \end{bmatrix}.$$

We get

$$\begin{bmatrix} \hat{\beta}_R \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} X'X & R' \\ R & 0 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ r \end{bmatrix}.$$

The next section is not for the undergaraduate course but it is left here to show how we would solve this problem. After all the math, we have the restricted estimates from the unrestricted ones, which means that we only have to estimate the unrestricted model to calculate the restricted estimates afterwards.

Aside:**Partition inverse matrices**

Consider the following partitioned matrix: $\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$

Let's define pivot elements the elements in the main diagonal:

i) Construct the first tableau, pivot element is 1,1 (a):

- element a becomes a^{-1}

- in pivot row do:$(pivot-element)^{-1}*(pivot-row-element)$: $\begin{bmatrix} a^{-1} & a^{-1}b & a^{-1}c \end{bmatrix}$

-in pivot column:$-(pivot-column-element)*(pivot-element)^{-1}$:

$$\begin{bmatrix} a^{-1} & a^{-1}b & a^{-1}c \\ -da^{-1} & & \\ -ga^{-1} & & \end{bmatrix}$$

-off elements: $(off-elements)-[(element-in-same-pivot-row)*(pivot-element)^{-1}$

$*(element-same-pivot-col)]$:

$$\begin{bmatrix} a^{-1} & a^{-1}b & a^{-1}c \\ -da^{-1} & e-da^{-1}b & f-da^{-1c} \\ -ga^{-1} & h-ga^{-1b} & i-ga^{-1}c \end{bmatrix}$$

ii) Construct the next tableau with element 2,2 (e) as the pivot element.

iii) Construct the last tableau with element 3,3 (i) as the pivot element. This will be the final partitioned inverse.

Another way to do it is the following:

Stepwise Inversion:

note that

$$
\begin{aligned}
\begin{bmatrix} (X'X)^{-1} & 0 \\ -R(X'X)^{-1} & I_Q \end{bmatrix} \begin{bmatrix} X'X & R' \\ R & 0 \end{bmatrix} \quad &\equiv \quad AB \\
&= \quad \begin{bmatrix} I_K & (X'X)^{-1}R' \\ 0 & -R(X'X)^{-1}R' \end{bmatrix} \\
&\equiv \quad \begin{bmatrix} I_K & (X'X)^{-1}R' \\ 0 & -P \end{bmatrix} \\
&\equiv C,
\end{aligned}
$$

and

$$
\begin{aligned}
\begin{bmatrix} I_K & (X'X)^{-1}R'P^{-1} \\ 0 & -P^{-1} \end{bmatrix} \begin{bmatrix} I_K & (X'X)^{-1}R' \\ 0 & -P \end{bmatrix} \quad &\equiv \quad DC \\
&= \quad I_{K+Q},
\end{aligned}
$$

so

$$
\begin{aligned}
DAB \quad &= \quad I_{K+Q} \\
DA \quad &= \quad B^{-1} \\
B^{-1} \quad &= \quad \begin{bmatrix} I_K & (X'X)^{-1}R'P^{-1} \\ 0 & -P^{-1} \end{bmatrix} \begin{bmatrix} (X'X)^{-1} & 0 \\ -R(X'X)^{-1} & I_Q \end{bmatrix} \\
&= \quad \begin{bmatrix} (X'X)^{-1} - (X'X)^{-1}R'P^{-1}R(X'X)^{-1} & (X'X)^{-1}R'P^{-1} \\ P^{-1}R(X'X)^{-1} & -P^{-1} \end{bmatrix},
\end{aligned}
$$

so

$$
\begin{aligned}
\begin{bmatrix} \hat{\beta}_R \\ \hat{\lambda} \end{bmatrix} &= \begin{bmatrix} (X'X)^{-1} - (X'X)^{-1}R'P^{-1}R(X'X)^{-1} & (X'X)^{-1}R'P^{-1} \\ P^{-1}R(X'X)^{-1} & -P^{-1} \end{bmatrix} \begin{bmatrix} X'y \\ r \end{bmatrix} \\
&= \begin{bmatrix} \hat{\beta} - (X'X)^{-1}R'P^{-1}\left(R\hat{\beta} - r\right) \\ P^{-1}\left(R\hat{\beta} - r\right) \end{bmatrix} \\
&= \begin{bmatrix} \left(I_K - (X'X)^{-1}R'P^{-1}R\right) \\ P^{-1}R \end{bmatrix} \hat{\beta} + \begin{bmatrix} (X'X)^{-1}R'P^{-1}r \\ -P^{-1}r \end{bmatrix}
\end{aligned}
$$

The fact that $\hat{\beta}_R$ and $\hat{\lambda}$ are linear functions of $\hat{\beta}$ makes it easy to determine their distributions, since the distribution of $\hat{\beta}$ is already known. Recall that for x a random vector, and for A and b a matrix and vector of constants, respectively, $Var(Ax+b) = AVar(x)A'$.

3.2 Testing

In many cases, one wishes to test economic theories. If theory suggests parameter restrictions, as in the above homogeneity example, one can test theory by testing parameter restrictions. A number of tests are available.

3.2.1 t-test

Suppose one has the model

$$y = X\beta + \varepsilon$$

and one wishes to test the *single restriction* $H_0 : R\beta = r$ vs. $H_A : R\beta \neq r$. Under H_0, with normality of the errors,

$$R\hat{\beta} - r \sim N\left(0, R(X'X)^{-1}R'\sigma^2\right)$$

so

$$\frac{R\hat{\beta} - r}{\sqrt{R(X'X)^{-1}R'\sigma^2}} = \frac{R\hat{\beta} - r}{\sigma^2\sqrt{R(X'X)^{-1}R'}} \sim N(0,1).$$

The problem is that σ^2 is unknown. One could use the consistent estimator $\widehat{\sigma^2}$ in place of σ^2, but the test would only be valid asymptotically in this case.

Consider the random variable

$$\begin{aligned}
\frac{\hat{\varepsilon}'\hat{\varepsilon}}{\sigma^2} &= \frac{\varepsilon' M_X \varepsilon}{\sigma^2} \\
&= \left(\frac{\varepsilon}{\sigma}\right)' M_X \left(\frac{\varepsilon}{\sigma}\right) \\
&\sim \chi^2(n-K)
\end{aligned}$$

Now consider (remember that we have only one restriction in this case)

$$\frac{\frac{R\hat{\beta}-r}{\sigma\sqrt{R(X'X)^{-1}R'}}}{\sqrt{\frac{\hat{\varepsilon}'\hat{\varepsilon}}{(n-K)\sigma^2}}} = \frac{R\hat{\beta} - r}{\hat{\sigma}\sqrt{R(X'X)^{-1}R'}}$$

This will have the $t(n-K)$ distribution.

so

$$\frac{R\hat{\beta} - r}{\hat{\sigma}\sqrt{R(X'X)^{-1}R'}} = \frac{R\hat{\beta} - r}{\hat{\sigma}_{R\hat{\beta}}} \sim t(n-K)$$

In particular, for the commonly encountered *test of significance* of an individual coefficient, for which $H_0 : \beta_j = 0$ vs. $H_0 : \beta_j \neq 0$, the test statistic is

$$\frac{\hat{\beta}_j}{\hat{\sigma}_{\hat{\beta}j}} \sim t(n-K)$$

3.2.2 p-value (as a decision rule)

Instead of finding the t-statistic you can calculate

$$\begin{aligned} p-value &= Pr(t > |t_j|) * 2 \\ Pr(-|t_j| < t < |t_j| \quad) &= 1-p \end{aligned}$$

3.2.3 *F* test

The F test allows testing multiple restrictions jointly.

Now that we have a menu of test statistics, we need to know how to use them.

Proposition 2 *If $x \sim \chi^2(r)$ and $y \sim \chi^2(s)$, then*

$$\frac{x/r}{y/s} \sim F(r,s) \tag{1}$$

provided that x and y are independent.

Proposition 3 *If the random vector (of dimension n) $x \sim N(0,I)$, then $x'Ax$ and $x'Bx$ are independent if $AB = 0$.*

Using these results, and previous results on the χ^2 distribution, it is simple to show that the following statistic has the F distribution:

$$F = \frac{\left(R\hat{\beta} - r\right)' \left(R(X'X)^{-1}R'\right)^{-1} \left(R\hat{\beta} - r\right)}{q\hat{\sigma}^2} \sim F(q, n-K).$$

A numerically equivalent expression is

$$\frac{(RSS_R - RSS_U)/q}{RSS_U/(n-K)} \sim F(q, n-K).$$

3.2.4 Confidence intervals

Confidence intervals for single coefficients are generated in the normal manner. Given the t statistic

$$t - statistic = \frac{\hat{\beta}_j - a}{\widehat{\sigma_{\hat{\beta}_j}}}$$

a $100(1-\alpha)\%$ confidence interval for β is defined by the bounds of the set of a such that the $t - statistic$ does not reject $H_0 : \beta_j = a$, using a α significance level:

$$\begin{aligned} C(\alpha) &= \{\beta_j : -t_{\alpha/2,(n-k)} < \frac{\hat{\beta}_j - a}{\widehat{\sigma_{\hat{\beta}_j}}} < t_{\alpha/2,(n-k)}\} \\ Pr(-t_{\alpha/2,(n-k)} < \frac{\hat{\beta}_j - a}{\widehat{\sigma_{\hat{\beta}_j}}} < t_{\alpha/2,(n-k)}) &= 1-\alpha \end{aligned}$$

inside the expression of the probability, multiply by $\widehat{\sigma_{\hat{\beta}_j}}$, subtract $\hat{\beta}_j$, multiply by -1 and reorder.

The set of such β_j is the interval

$$\hat{\beta}_j \pm \widehat{\sigma_{\hat{\beta}_j}} t_{\alpha/2}$$

It is not the same to test two coefficients individually that to test them jointly: A confidence ellipse for two coefficients jointly would be, analogously, the set of $\{\beta_1, \beta_2\}$ such that the F (or some other test statistic) doesn't reject at the specified critical value. This generates an ellipse, if the estimators are correlated. *Draw a picture here.*

- The region is an ellipse, since the CI for an individual coefficient defines a (infinitely long) rectangle with total prob. mass $1-\alpha$, since the other coefficient is marginalized (e.g., can take on any value). Since the ellipse is bounded in both dimensions but also contains mass $1-\alpha$, it must extend beyond the bounds of the individual CI.

- From the pictue we can see that:
  - Rejection of hypotheses individually does not imply that the joint test will reject.
  - Joint rejection does not imply individal tests will reject.

Example: (Hayashi pg 45) assume $k=2$ and consider:

$$\begin{aligned} H_0:\beta_1 &= 1 \\ \beta_2 &= 0 \end{aligned}$$

this can be written as a linear hypothesis $R\beta = r$ for $R = I_2$ and $r = (1,0)\prime$ so the $F\ test$ should be used. It is tempting, however, to conduct the $t-test$ separately for each individual coefficient of the hypothesis. We might accept H_0if both restrictions $\beta_1 = 1$ *and* $\beta_2 = 0$ pass the $t-test$. This amounts to using the confidence region of:

$$\begin{aligned} CI[(\beta_1,\beta_2) &: (\hat{\beta}_1 - \widehat{\sigma_{\hat{\beta}_1}} t_{\alpha/2,(n-k)} < \beta_1 < (\hat{\beta}_1 + \widehat{\sigma_{\hat{\beta}_1}} t_{\alpha/2,(n-k)}) \\ &: (\hat{\beta}_2 - \widehat{\sigma_{\hat{\beta}_2}} t_{\alpha/2,(n-k)} < \beta_2 < (\hat{\beta}_2 + \widehat{\sigma_{\hat{\beta}_2}} t_{\alpha/2,(n-k)})] \end{aligned}$$

which is a rectangular region in the (β_1,β_2) plane.

On the other hand, the confidence region for the $F-test$ is

$$[(\beta_1,\beta_2):(\hat{\beta}_1-\beta_1,\hat{\beta}_2-\beta_2)(\widehat{V(\hat{\beta})})^{-1}\begin{pmatrix}\hat{\beta}_1-\beta_1\\ \hat{\beta}_2-\beta_2\end{pmatrix}]<2F_{\alpha(q,n-k)}$$

the acceptance region is an elipse in the (β_1,β_2) plane.

3.2.5 Examples

When considering a set of restrictions we have to be careful about not stating redundant or inconsistent equations. Let's consider the wage equation in the example of the introduction

$$log(wage_i)=\beta_1+\beta_2 S_i+\beta_3 TEN_i+\beta_4 EXP_4+\varepsilon_i$$

We might wish to test that education and tenure have equal impact on the wage rate and that there is no experience effect:

$$\begin{aligned} H_0:\beta_2 &= \beta_3 \Rightarrow \beta_2-\beta_3=0 \\ \beta_4 &= 0 \end{aligned}$$

since the two rows are linearly independent the rank condition is satisfied.

Suppose now that additionally we write $\beta_2-\beta_3=\beta_4$ If you construct the R matrix you will see that is a $3x4$matrix but the rank is 2. This are called *redundant restrictions*.

You can have also *inconsistent restrictions* That happen when there is no β that can satisfy those restrictions example

$$\begin{aligned} H_0 : \beta_2 &- \beta_3 = 0 \\ \beta_4 &= 0 \\ \beta_4 &= 0.5 \end{aligned}$$