Universidad del País Vasco · Euskal Herriko Unibertsitatea
CAMPUS OF INTERNATIONAL EXCELLENCE

Lesson 5

The Multiple Regression Model

Estimation

Pilar González and Susan Orbe

Dpt. Applied Econometrics III (Econometrics and Statistics)

Learning objectives

- To estimate a linear regression model by Ordinary Least Squares (OLS) using available data.
- To interpret the estimated coefficients of a linear regression model.
- To derive the properties of the OLS estimator.
- To derive the properties of the sample regression function.
- To calculate a measure of Goodness-of-fit.
- To estimate the linear regression model using non sample information by Restricted Least Squares.

Contents

Contents

Estimation of the Linear Regression Model.

Objective.

To estimate the unknown parameters of the linear regression model:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i \quad i = 1, 2, \ldots, N. \quad u_i \sim NID(0, \sigma^2)$$

using the available sample.

- Y: dependent variable, endogenous variable.
- $X_j, j = 2, \ldots, k$: regressors.
- β: unknown coefficients.
- u: non observable error term.
- i: index used with cross-section data (t in case of time series data).
- N: sample size (T in case of time series data).

Estimation of the Linear Regression Model.

General linear regression model.

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i \quad i = 1, 2, \ldots, N. \quad u_i \sim NID(0, \sigma^2)$$

Some remarks:

- This is the expression that will be used for the GLRM throughout the theoretical exposition of the estimation methodology.
- This expression includes all the regression models linear in the coefficients discussed in Lesson 4.
- $X_j, j = 2, \ldots, k,$: regressors that can be quantitative variables, dummy variables or other terms, such as product of variables, quadratic or logarithmic transformations, ...
- The interpretation of the coefficients depends on what X_j stands for.

Estimation of the Linear Regression Model.

Consider the general linear regression model:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i \qquad i = 1, 2, \ldots, N$$

Written for each observation:

$$\begin{cases} Y_1 = \beta_1 + \beta_2 X_{21} + \beta_3 X_{31} + \ldots + \beta_k X_{k1} + u_1 & i = 1 \\ Y_2 = \beta_1 + \beta_2 X_{22} + \beta_3 X_{32} + \ldots + \beta_k X_{k2} + u_2 & i = 2 \\ \vdots & \vdots \\ Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i & i = i \\ \vdots & \vdots \\ Y_N = \beta_1 + \beta_2 X_{2N} + \beta_3 X_{3N} + \ldots + \beta_k X_{kN} + u_N & i = N \end{cases}$$

Estimation of the Linear Regression Model.

The linear regression model can be written as well in matrix form. This expression will be used for some proofs and to derive some complex expressions.

The GLRM in matrix form.

$$Y = X\beta + u$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_i \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} 1 & X_{21} & X_{31} & \cdots & X_{k1} \\ 1 & X_{22} & X_{32} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{2i} & X_{3i} & \cdots & X_{ki} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{2N} & X_{3N} & \cdots & X_{kN} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_i \\ \vdots \\ u_N \end{bmatrix}$$

The OLS estimator.

Objective: to estimate the unknown coefficients of the **Population Regression Function**:

$$E(Y_i|X) = \beta_1 + \beta_2\, X_{2i} + \ldots + \beta_k\, X_{ki}$$

Sample Regression Function (SRF): obtained substituting the estimated coefficients:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2\, X_{2i} + \ldots + \hat{\beta}_k\, X_{ki}$$

where $\hat{Y}_i$ are called the fitted or adjusted values.

The estimation error is referred as **residual**:

$$\hat{u}_i = Y_i - \hat{Y}_i = \hat{u}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2\, X_{2i} - \ldots - \hat{\beta}_k\, X_{ki}$$

This residual depends on the error term and on the error due to the estimation of the coefficients:

$$\hat{u}_i = u_i + (\beta_1 - \hat{\beta}_1) + (\beta_2 - \hat{\beta}_2)\, X_{2i} + \ldots + (\beta_k - \hat{\beta}_k)\, X_{ki}$$

The OLS estimator.

The OLS criterion.

The ordinary least squares estimator (OLS) is obtained minimizing the objective function:

$$\min_{\hat{\beta}_1,\ldots,\hat{\beta}_k} \sum_{i=1}^{N} \hat{u}_i^2 \quad = \quad \min_{\hat{\beta}_1,\ldots,\hat{\beta}_k} \sum_{i=1}^{N} (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \ldots - \hat{\beta}_k X_{ki})^2$$

First order conditions.

Given by the first derivatives of the objective function equal to zero:

$$\left.\frac{\partial \sum_{i=1}^{N} \hat{u}_i^2}{\partial \hat{\beta}_1}\right|_{\hat{\beta}_1 = \hat{\beta}_1^{OLS}} = 0, \quad \ldots, \quad \left.\frac{\partial \sum_{i=1}^{N} \hat{u}_i^2}{\partial \hat{\beta}_k}\right|_{\hat{\beta}_k = \hat{\beta}_k^{OLS}} = 0$$

The OLS estimator.

The first order conditions are:

$$\left.\frac{\partial \sum_{i=1}^{N} \hat{u}_i^2}{\partial \hat{\beta}_1}\right|_{\hat{\beta}_1=\hat{\beta}_1^{OLS}} = -2\sum_{i=1}^{N}(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \ldots - \hat{\beta}_k X_{ki}) = 0$$

$$\left.\frac{\partial \sum_{i=1}^{N} \hat{u}_i^2}{\partial \hat{\beta}_2}\right|_{\hat{\beta}_2=\hat{\beta}_2^{OLS}} = -2\sum_{i=1}^{N}(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \ldots - \hat{\beta}_k X_{ki})X_{2i} = 0$$

$$\vdots$$

$$\left.\frac{\partial \sum_{i=1}^{N} \hat{u}_i^2}{\partial \hat{\beta}_k}\right|_{\hat{\beta}_k=\hat{\beta}_k^{OLS}} = -2\sum_{i=1}^{N}(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \ldots - \hat{\beta}_k X_{ki})X_{ki} = 0$$

The OLS estimator.

Normal equations.

$$
\begin{aligned}
\sum Y_i &= N\hat{\beta}_1 + \hat{\beta}_2 \sum X_{2i} + \ldots + \hat{\beta}_k \sum X_{ki} \\
\sum X_{2i} Y_i &= \hat{\beta}_1 \sum X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 + \ldots + \hat{\beta}_k \sum X_{2i} X_{ki} \\
&\vdots \\
\sum X_{ki} Y_i &= \hat{\beta}_1 \sum X_{ki} + \hat{\beta}_2 \sum X_{ki} X_{2i} + \ldots + \hat{\beta}_k \sum X_{ki}^2 .
\end{aligned}
$$

The first order conditions provide a system of k linear independent equations with k unknown parameters, which is called normal equations. The OLS estimator is obtained by solving this system of equations.

Given that the second order conditions of minimization are always satisfied, we may affirm that the solution of the normal equations system is a minimum.

The OLS estimator.

The OLS estimator in matrix form.

Objective function: $\min_{\hat\beta} \hat{u}'\hat{u} = \min_{\hat\beta} (Y - X\hat\beta)'(Y - X\hat\beta)$.

First order conditions:

$$\left.\frac{\partial \hat{u}'\hat{u}}{\partial \hat\beta}\right|_{\hat\beta = \hat\beta^{OLS}} = 0 \quad \Rightarrow \quad -2X'(Y - X\hat\beta^{OLS}) = 0 \quad \Rightarrow \quad X'\hat{u}^{OLS} = 0.$$

Normal equations: $X'Y = X'X\hat\beta^{OLS}$.

Second order conditions:

$$\left.\frac{\partial \hat{u}'\hat{u}}{\partial \hat\beta \partial \hat\beta'}\right|_{\hat\beta = \hat\beta^{OLS}} = 2X'X$$ It is always a semidefinite positive matrix.

The OLS estimator.

OLS estimator.

Solving this system of k linear independent equations, we get the OLS estimators:

$$\hat{\beta}^{OLS} = (X'X)^{-1}X'Y$$

$$X'X = \begin{bmatrix} N & \sum X_{2i} & \sum X_{3i} & \cdots & \sum X_{ki} \\ \sum X_{2i} & \sum X_{2i}^2 & \sum X_{2i}X_{3i} & \cdots & \sum X_{2i}X_{ki} \\ \sum X_{3i} & \sum X_{3i}X_{2i} & \sum X_{3i}^2 & \cdots & \sum X_{3i}X_{ki} \\ \vdots & \vdots & \vdots & & \vdots \\ \sum X_{ki} & \sum X_{ki}X_{2i} & \sum X_{ki}X_{3i} & \cdots & \sum X_{ki}^2 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} \sum Y_i \\ \sum X_{2i}Y_i \\ \sum X_{3i}Y_i \\ \vdots \\ \sum X_{ki}Y_i \end{bmatrix} \qquad \hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}$$

See **Examples 5.1 and 5.2** for applications.

The OLS estimator.

Some equivalences

Observation i

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_K X_{ki} + u_i$$

$$E(Y_i|X) = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}$$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \ldots + \hat{\beta}_K X_{ki}$$

$$u_i = Y_i - \beta_1 - \beta_2 X_{2i} - \ldots - \beta_k X_{ki}$$

$$\hat{u}_i = Y_i - \hat{Y}_i$$

$$\hat{u}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \ldots - \hat{\beta}_K X_{ki}$$

$$\hat{u}_i = u_i + (\beta_1 - \hat{\beta}_1) + \ldots + (\beta_k - \hat{\beta}_k) X_{ki}$$

In matrix form

$$Y = X\beta + u$$

$$E(Y|X) = X\beta$$

$$\hat{Y} = X\hat{\beta}$$

$$U = Y - X\beta$$

$$\hat{U} = Y - \hat{Y}$$

$$\hat{U} = Y - X\hat{\beta}$$

$$\hat{U} = U + X(\beta - \hat{\beta})$$

Contents

GLRM estimation. The Sample Regression Function (SRF).

Some algebraic properties of the sample regression function:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \ldots + \hat{\beta}_k X_{ki} \qquad i = 1, 2, \ldots, N$$

1. The sum of the residuals is zero: $\displaystyle\sum_{i=1}^{N} \hat{u}_i = 0$.
2. The residuals are orthogonal to the regressors: $\displaystyle\sum_{i=1}^{N} \hat{u}_i X_{ji} = 0$

Proof. From the normal equations system:

$$X'(Y - X\hat{\beta}) = X'\hat{u} = 0 \Leftrightarrow \begin{bmatrix} \sum_1^N \hat{u}_i \\ \sum_1^N X_{2i}\hat{u}_i \\ \sum_1^N X_{3i}\hat{u}_i \\ \vdots \\ \sum_1^N X_{ki}\hat{u}_i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \Rightarrow \begin{cases} \sum_{i=1}^N \hat{u}_i = 0 \\ \sum_{i=1}^N \hat{u}_i X_{ji} = 0, \ \forall j \end{cases}$$

The Sample Regression Function.

3. The residuals are orthogonal to the fitted values: $\hat{Y}'\hat{u} = 0$.

Proof:

$$\hat{Y}'\hat{u} = (X\hat{\beta})'\hat{u} = \hat{\beta}' \underbrace{X'\hat{u}}_{=0} = 0$$

4. The sample means of Y and $\hat{Y}$ are equal: $\bar{Y} = \bar{\hat{Y}}$.

Proof:

$$\hat{u}_i = Y_i - \hat{Y}_i \quad \Longleftrightarrow \quad Y_i = \hat{Y}_i + \hat{u}_i$$

$$\sum Y_i = \sum \hat{Y}_i + \underbrace{\sum \hat{u}_i}_{=0}$$

$$\frac{1}{N}\sum Y_i = \frac{1}{N}\sum \hat{Y}_i \quad \Rightarrow \quad \bar{Y} = \bar{\hat{Y}}$$

The Sample Regression Function.

5. The SRF lays on the vector of sample means $(\bar{Y}, \bar{X}_2, \ldots, \bar{X}_k)$.

Proof:

$$\sum_{i=1}^{N} \hat{u}_i = 0 \quad \Leftrightarrow \quad \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \ldots - \hat{\beta}_k X_{ki}) = 0$$

$$\sum Y_i - N\hat{\beta}_1 - \hat{\beta}_2 \sum X_{2i} - \ldots - \hat{\beta}_k \sum X_{ki} = 0$$

$$\sum Y_i = N\hat{\beta}_1 + \hat{\beta}_2 \sum X_{2i} + \ldots + \hat{\beta}_k \sum X_{ki}$$

$$\frac{1}{N} \sum Y_i = \hat{\beta}_1 + \hat{\beta}_2 \frac{1}{N} \sum X_{2i} + \ldots + \hat{\beta}_k \frac{1}{N} \sum X_{ki}$$

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \ldots + \hat{\beta}_k \bar{X}_k$$

Contents

Properties of the OLS estimator: $\hat{\beta} = (X'X)^{-1}X'Y$

1. **Linearity**. An estimator is linear if and only if it can be expressed as a linear function of the dependent variable, conditional on X.

   Note that, conditional on X, $\hat{\beta}$ can be written as a linear function of the error term as well.

Proof. Under assumptions A1 and A2:

$$\hat{\beta} \quad = \quad (X'X)^{-1}X'Y = (X'X)^{-1}X'(X\beta + u) = \beta + (X'X)^{-1}X'u$$

2. **Unbiasness**. $\hat{\beta}$ is unbiased, that is, the expected value of the OLS estimator is equal to the population value of the coefficients of the model.

Proof. Under assumptions A1 through A3:

$$E(\hat{\beta}|X) \quad = \quad E[(\beta + (X'X)^{-1}X'u)|X] = \beta + (X'X)^{-1}X'E(u|X) = \beta$$

Properties of the OLS estimator.

3. **Variance**: $V(\hat{\beta}|X) = \sigma^2(X'X)^{-1}$

Proof. Under assumptions A1 through A5:

$$\begin{aligned} V(\hat{\beta}|X) &= E[(\hat{\beta} - E(\hat{\beta}|X))(\hat{\beta} - E(\hat{\beta}|X))'|X] = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'|X] = \\ &= E\left[\left[(X'X)^{-1}X'u\right]\left[(X'X)^{-1}X'u\right]'|X\right] = \\ &= (X'X)^{-1}X'\sigma^2 I_T X(X'X)^{-1} = \\ &= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1} \end{aligned}$$

For simplicity, this covariance matrix will be denoted by $V(\hat{\beta})$.

$$V(\hat{\beta}) = \sigma^2(X'X)^{-1} = \begin{bmatrix} V(\hat{\beta}_1) & Cov(\hat{\beta}_1, \hat{\beta}_2) & \cdots & Cov(\hat{\beta}_1, \hat{\beta}_k) \\ Cov(\hat{\beta}_2, \hat{\beta}_1) & V(\hat{\beta}_2) & \cdots & Cov(\hat{\beta}_2, \hat{\beta}_k) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\hat{\beta}_k, \hat{\beta}_1) & Cov(\hat{\beta}_k, \hat{\beta}_2) & \cdots & V(\hat{\beta}_k) \end{bmatrix}$$

Properties of the OLS estimator.

This **variance is minimum** in the class of all linear and unbiased estimators.

Gauss-Markov Theorem

Under assumptions A1 through A5, in the class of linear unbiased estimators, OLS has the smallest variance, conditional on X.

> *Under assumptions A1 through A5, $\hat{\beta}^{OLS}$ is the best linear unbiased estimator (BLUE) of $\hat{\beta}$.*

This theorem justifies the use of the OLS method instead of other competing estimators.

Note. All the analysis is conditional on X, even though it is not going to be explicitely written from now onwards in order to simplify the notation.

OLS residuals.

The OLS residuals can be written in terms of the error term as follows:

$$\hat{u} = Mu$$

where M is a squared matrix of order N, symmetric ($M = M'$), idempotent ($MM = M$), with rank rg(M)= tr(M)= $N - K$ and orthogonal to X ($MX = 0$).

Proof:

$$\begin{aligned}
\hat{u} &= Y - \hat{Y} = Y - X\hat{\beta} = Y - X(X'X)^{-1}X'Y = \\
&= [I_N - X(X'X)^{-1}X']Y = [I_N - X(X'X)^{-1}X'](X\beta + u) = \\
&= X\beta - X(X'X)^{-1}X'X\beta + [I_N - X(X'X)^{-1}X']u = \\
&= \underbrace{[I_N - X(X'X)^{-1}X']}_{=M} u = Mu
\end{aligned}$$

Estimation.

Properties of the OLS residuals.

1. Expected value: $E(\hat{u}) = 0$.

Proof:

$$E(\hat{u}) = ME(u) = M\ 0 = 0$$

2. Variance: Even in the case of homoskedastic and uncorrelated disturbances ($u \sim (0, \sigma^2 I)$), the OLS residuals are **NOT** homoskedastic and they **ARE** correlated.

Proof:

$$\begin{aligned} V(\hat{u}) &= E(\hat{u}\hat{u}') = E(Muu'M') = \\ &= M\sigma^2 I_N M = \sigma^2 M \quad \rightarrow \quad M \neq I \end{aligned}$$

Estimation.

Estimator of the variance of the error term, σ^2.

$$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{N-k} = \frac{SSR}{N-k} = \frac{\sum \hat{u}_i^2}{N-k}$$

Proof:

This estimator of the variance of the error term is unbiased:

$$E(\hat{\sigma}^2) = \frac{E(\hat{u}'\hat{u})}{N-k} = \frac{\sigma^2(N-k)}{N-k} = \sigma^2$$

given that:

$$\begin{aligned} E(\hat{u}'\hat{u}) &= E(u'Mu) = E(tr(u'Mu)) = E(tr(Muu')) = \\ &= tr(E(Muu')) = tr(M\sigma^2 I_N) = \\ &= \sigma^2 tr(M) = \sigma^2(N-k) \end{aligned}$$

Estimation.

Estimator of the covariance matrix of $\hat{\beta}$.

$$\widehat{V}(\hat{\beta}) = \begin{bmatrix} \widehat{V}(\hat{\beta}_1) & \widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2) & \cdots & \widehat{Cov}(\hat{\beta}_1, \hat{\beta}_k) \\ \widehat{Cov}(\hat{\beta}_2, \hat{\beta}_1) & \widehat{V}(\hat{\beta}_2) & \cdots & \widehat{Cov}(\hat{\beta}_2, \hat{\beta}_k) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{Cov}(\hat{\beta}_k, \hat{\beta}_1) & \widehat{Cov}(\hat{\beta}_k, \hat{\beta}_2) & \cdots & \widehat{V}(\hat{\beta}_k) \end{bmatrix} = \widehat{\sigma}^2 (X'X)^{-1}$$

Contents

GLRM estimation. Goodness-of-fit.

Sum of squares decomposition.

$$SST = SSE + SSR$$

- **Total Sum of Squares**: $SST = \sum_{t=1}^{T}(Y_t - \bar{Y})^2$

  It is a measure of the total sample variation in Y.

- **Explained Sum of Squares**: $SSE = \sum_{t=1}^{T}(\hat{Y}_t - \bar{Y})^2$

  It is a measure of the sample variation in $\hat{Y}$.

- **Residual Sum of Squares**: $SSR = \sum_{t=1}^{T}(Y_t - \hat{Y}_t)^2 = \sum_{t=1}^{T}\hat{u}_t^2$

  It is the sample variation in the OLS residuals.

Goodness-of-fit.

Regression model: $Y_t = \hat{Y}_t + \hat{u}_t$

$\Longrightarrow$ Y_t is decomposed into two parts: fitted value and residual.

Estimating by OLS, the following decomposition holds:

$$\boxed{\text{SST} = \text{SSE} + \text{SSR}}$$

$$\sum_{t=1}^{N}(Y_t - \bar{Y})^2 = \sum_{t=1}^{N}(\hat{Y}_t - \bar{Y})^2 + \sum_{t=1}^{N}\hat{u}_t^2$$

Therefore, it can be proved that:

The total sample variation in Y can be expressed as the sum of

the variation in the fitted values of Y, $\hat{Y}$
*(the variation in Y **explained** by the regression)*

$+$ *the variation in the OLS residuals, $\hat{u}$*
*(the variation in Y **UNexplained** by the regression)*

Goodness-of-fit.

Proof.

Writting the GLRM as follows: $\hat{u}_t = Y_t - \hat{Y}_t \quad \Rightarrow \quad Y_t = \hat{Y}_t + \hat{u}_t$

The total sample variation in Y is: $\text{SST} = \sum_{t=1}^{T}(Y_t - \bar{Y})^2$

$$\begin{aligned}
\sum_{t=1}^{T}(Y_t - \bar{Y})^2 &= \sum_{t=1}^{T}(\hat{Y}_t + \hat{u}_t - \bar{Y})^2 = \sum_{t=1}^{T}[(\hat{Y}_t - \bar{Y}) + \hat{u}_t]^2 = \\
&= \sum_{t=1}^{T}(\hat{Y}_t - \bar{Y})^2 + \sum_{t=1}^{T}\hat{u}_t^2 + 2\sum_{t=1}^{T}(\hat{Y}_t - \bar{Y})\hat{u}_t = \\
&= \sum_{t=1}^{T}(\hat{Y}_t - \bar{Y})^2 + \sum_{t=1}^{T}\hat{u}_t^2 = \sum_{t=1}^{T}(\hat{Y}_t - \bar{\hat{Y}})^2 + \sum_{t=1}^{T}\hat{u}_t^2
\end{aligned}$$

$$\boxed{\text{SST} = \text{SSE} + \text{SSR}}$$

Goodness-of-fit.

Goodness-of-fit measure.

Coefficient of determination:

$$R^2 = \frac{SSE}{SST}$$

R^2 is the fraction of the sample variation in Y that is explained by the variation in the explanatory variables, X.

If the model contains an intercept, the coefficient of determination may be calculated as follows:

$$R^2 = 1 - \frac{SSR}{SST} \in (0, 1)$$

Gretl computes the coefficient of determination using the formula in terms of the Sum of Squares Residuals!

Goodness-of-fit.

Some performance criteria.

These criteria are useful to compare the performance of nested regression models.

Rationale: to adjust the sum of squared residuals by the degrees of freedom.

Given that they are based on the SSR, the smaller the value of the criterion the better the performance of the model except for the adjusted R^2.

Adjusted R^2: $\bar{R}^2 = 1 - \frac{N-1}{N-k} R^2$

Log-likelihood: $\ell(\hat{\theta}) = \frac{N}{2}(1 + \ln 2\pi - \ln N) - \frac{N}{2} \ln SSR$

Akaike criterion: $AIC = -2\ell(\hat{\theta}) + 2k$

Schwarz criterion: $BIC = -2\ell(\hat{\theta}) + k \ln N$

Hannan-Quinn criterion: $HQC = -2\ell(\hat{\theta}) + 2k \ln \ln N$

Contents

Restricted Least Squares estimator.

Objective.

To estimate a linear regression model:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i \quad i = 1, 2, \ldots, N. \quad u_i \sim NID(0, \sigma_u^2)$$

including non sample information on the coefficients.

The Restricted Least Squares estimator (RLS) is

$$\hat{\beta}^{RLS} = (X_R' X_R)^{-1} X_R' Y_R$$

where X_R y Y_R are the observation matrix and the dependent variable of the restricted model, that is, of the model that results from including the non sample information in the regression model.

See **Example 5.3** for applications.

Restricted Least Squares estimator.

Properties.

$$\hat{\beta}^{RLS} = (X_R' X_R)^{-1} X_R' Y_R$$

Conditioning on X, the RLS estimator is:

- Linear in u.
- Unbiased if the non sample information included in the estimation is true.
- The variance of the RLS estimator is always smaller than the variance of the OLS estimator, even if the non sample information included is false.

Therefore, if the non sample information available is true, the RLS estimator should be used because it is unbiased and has a smaller variance than the OLS estimator.

Restricted Least Squares estimator.

Example 1.

Consider the regression model:

$$pizza_i = \beta_1 + \beta_2\, income_i + \beta_3\, age_i \, + \, u_i \qquad i = 1, 2, \ldots, N$$

subject to: $\quad \beta_2 = -\beta_3$

Restricted model:

$$pizza_i = \beta_1 + \beta_2 (income_i - age_i) + u_i$$

$$X_R = \begin{pmatrix} 1 & income_1 - age_1 \\ 1 & income_2 - age_2 \\ \vdots & \vdots \\ 1 & income_N - age_N \end{pmatrix} \quad Y_R = \begin{pmatrix} pizza_1 \\ pizza_2 \\ \vdots \\ pizza_N \end{pmatrix}$$

$$\hat{\beta}^{RLS} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} \qquad \hat{\beta}_3^{RLS} = -\hat{\beta}_2^{RLS}$$

Restricted Least Squares estimator.

Example 2.

Consider the regression model:

$$pizza_i = \beta_1 + \beta_2\, income_i + \beta_3\, (age_i\ \times\ income_i) + u_i \quad i = 1, ..., N$$

subject to: $\quad \beta_2 = 5$

Restricted model:

$$pizza_i - 5\, income_i = \beta_1 + \beta_3\, (age_i\ \times\ income_i) + u_i$$

$$X_R = \begin{pmatrix} 1 & age_1 \times income_1 \\ 1 & age_2 \times income_2 \\ \vdots & \vdots \\ 1 & age_N \times income_N \end{pmatrix} \quad Y_R = \begin{pmatrix} pizza_1 - 5\, income_1 \\ pizza_2 - 5\, income_2 \\ \vdots \\ pizza_N - 5\, income_N \end{pmatrix}$$

$$\hat{\beta}^{RLS} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_3 \end{pmatrix} \qquad \hat{\beta}_2^{RLS} = 5$$

Restricted Least Squares estimator.

Example 3.

Consider the regression model:

$$Y_t = \beta_1 + \beta_2\, X2_t + \beta_3\, X3_t + \beta_4\, X4_t \, + \, u_t \qquad t = 1, 2, \ldots, T$$

subject to: $\beta_3 + \beta_4 = 1$

Restricted model: $Y_t - X4_t = \beta_1 + \beta_2\, X2_t + \beta_3\,(X3_t - X4_t) \, + \, u_t$

$$X_R = \begin{pmatrix} 1 & X2_1 & X3_1 - X4_1 \\ 1 & X2_2 & X3_2 - X4_2 \\ \vdots & \vdots & \\ 1 & X2_T & X3_T - X4_T \end{pmatrix} \quad Y_R = \begin{pmatrix} Y_1 - X4_1 \\ Y_2 - X4_2 \\ \vdots \\ Y_T - X4_T \end{pmatrix}$$

$$\hat{\beta}^{RLS} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} \qquad \hat{\beta}_4^{RLS} = 1 - \hat{\beta}_3^{RLS}$$

Restricted Least Squares estimator.

Example 4.

Consider the regression model:

$$Y_t = \beta_1 + \beta_2 \, X3_t + \beta_3 \, X3_t^2 + \beta_4 \, time \, + u_t \qquad t = 1, 2, \ldots, T$$

subject to: $\quad \beta_2 = \beta_3 = 0$

Restricted model: $\quad Y_t = \beta_1 + \beta_4 \, time + u_t$

$$X_R = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & T \end{pmatrix} \quad Y_R = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}$$

$$\hat{\beta}^{RLS} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_4 \end{pmatrix} \qquad \hat{\beta}_2^{RLS} = \hat{\beta}_3^{RLS} = 0$$

GLRM estimation. Omitting a relevant variable.

Omitting a relevant variable.

Let's assume that this regression model satisfies all the assumptions:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i \qquad i = 1, 2, \ldots, N.$$

To omit X_2 in this model $\equiv$ to include a false restriction $(\beta_2 = 0)$

The restricted model (omitting X_2) will be:

$$Y_i = \beta_1 + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i \qquad i = 1, 2, \ldots, N$$

The Restricted Least Squares Estimator, $\hat{\beta}_{RLS} = (X_R' X_R)^{-1} X_R' Y_R$, will have the following properties, conditional on X:

- Linear in u.
- BIASED because the restriction included is NOT true.
- Its variance is smaller than the variance of the OLS estimator.

GLRM estimation. Including an irrelevant variable.

Including an irrelevant variable.

Consider the linear regression model:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i \qquad i = 1, 2, \ldots, N. \qquad (1)$$

Assume that it is known that $\beta_2 = 0$, that is, that the regressor X_2 is irrelevant. In this case, it can be proved that the OLS estimator in model (1), conditional on X, is:

- Linear in u.
- Unbiased because, even though an irrelevant variable is included, $E(u|X) = 0$.
- The variance of the OLS estimator is the smallest in the class of linear and unbiased estimators.

BUT it is possible to estimate the coefficients of model (1) with a smaller variance:

$\Longrightarrow$ including in the model the true restriction $\beta_2 = 0$ and estimating by Restricted Least Squares.

Contents

Estimation results.

Example. Pizza consumption.

The file `pizza.gdt` contains information from N individuals on annual pizza consumption (in dollars) and some characteristics of the individuals, such as:

- Annual income (in thousands of dollars).
- Age of the consumer (in years).
- Gender.
- Highest level of studies (basic, high-school, college, postgraduated)

The factors that determine pizza consumption will be analysed in detail in the **Example 5.1**. The objective of this section is to estimate a very simple model for pizza consumption in order to explain the estimation output produced by Gretl.

Estimation results.

Consider the linear regression model:

$$pizza_i = \beta_1 + \beta_2\, income_i + \beta_3\, age_i + u_i \quad i = 1, ..., N$$

where:

- $pizza$: dependent variable (pizza consumption).
- $income$: quantitative explanatory variable.
- age: quantitative explanatory variable.
- $\beta_1, \beta_2, \beta_3$: unknown coefficients.
- u: error term.
- N: sample size.
- i: index.

Estimation results.

Gretl output.

Model 1: OLS, using observations 1–40
Dependent variable: pizza

	Coefficient	std. error	t-ratio	p-value
const	342.885	72.3434	4.7397	0.0000
income	1.83248	0.464301	3.9467	0.0003
age	−7.57556	2.31699	−3.2696	0.0023

Mean dependent var	191.5500	S.D. dependent var	155.8806
Sum squared resid	635636.7	S.E. of regression	131.0701
$R-$squared	0.329251	Adjusted $R-$squared	0.292994
$F(2, 37)$	9.081100	p-value (F)	0.000619
Log-likelihood	−250.2276	Akaike criterion	506.4552
Schwarz criterion	511.5218	Hannan–Quinn	508.2871

Estimation results.

Gretl output.

Model 1: OLS, using observations 1–40
Dependent variable: pizza

The information shown in the estimation output heading is:

Model 1 $\rightarrow$ this is the first model estimated since Gretl was started.

OLS $\rightarrow$ the estimation method used.

using observations 1–40 $\rightarrow$ the sample size is 40.

Dependent variable: pizza $\rightarrow$ the dependent variable of the estimated regression model is $pizza$.

Estimation results.

Gretl output.

Model 1: OLS, using observations 1–40
Dependent variable: pizza

	Coefficient	std. error	t-ratio	p-value
const	342.885	72.3434	4.7397	0.0000
income	1.83248	0.464301	3.9467	0.0003
age	−7.57556	2.31699	−3.2696	0.0023

The first column shows the regressors included in the estimated regression model. These regressors may be quantitative variables, dummy variables, products or transformations of variables, ... That is, they represent the columns of the observation matrix X: the constant (column of 1s) and the regressors income and age.

Estimation results.

Gretl output.

Model 1: OLS, using observations 1–40
Dependent variable: pizza

	Coefficient	std. error	t-ratio	p-value
const	342.885	72.3434	4.7397	0.0000
income	1.83248	0.464301	3.9467	0.0003
age	−7.57556	2.31699	−3.2696	0.0023

The second column shows the estimates of the coefficients of the model obtained using the OLS estimator: $\hat{\beta}^{OLS} = (X'X)^{-1}X'Y$.

The third column shows the standard errors of the OLS estimators of the coefficients. These standard errors are the squared root of the elements in the diagonal of the covariance matrix: $\widehat{V}(\hat{\beta}) = \hat{\sigma}^2(X'X)^{-1}$, where the estimator of the variance of the error term is given by $\hat{\sigma}^2 = SSR/(N-k)$.

Estimation results.

Gretl output.

Model 1: OLS, using observations 1–40
Dependent variable: pizza

	Coefficient	std. error	t-ratio	p-value
const	342.885	72.3434	4.7397	0.0000
income	1.83248	0.464301	3.9467	0.0003
age	−7.57556	2.31699	−3.2696	0.0023

The two last columns contain the t-ratios and the p-values. This information is useful to test the individual significance of the regressors (see Lesson 6).

Estimation results.

Gretl output.

Mean dependent var	191.5500	S.D. dependent var	155.8806
Sum squared resid	635636.7	S.E. of regression	131.0701
$R-$squared	0.329251	Adjusted $R-$squared	0.292994
$F(2, 37)$	9.081100	p-value (F)	0.000619
Log-likelihood	-250.2276	Akaike criterion	506.4552
Schwarz criterion	511.5218	Hannan–Quinn	508.2871

At the bottom of the estimation output appears some information of interest: the sum of squared residuals, coefficient of determination and the performance criteria explained in this lesson (adjusted R^2, Log-likelihood, Akaike criterion, Schwarz criterion and Hannan–Quinn criterion).

Estimation results.

Gretl output.

Mean dependent var	191.5500	S.D. dependent var	155.8806
Sum squared resid	635636.7	S.E. of regression	131.0701
$R-$squared	0.329251	Adjusted $R-$squared	0.292994
$F(2, 37)$	9.081100	p-value (F)	0.000619
Log-likelihood	-250.2276	Akaike criterion	506.4552
Schwarz criterion	511.5218	Hannan–Quinn	508.2871

Mean dependent var = 191.5500 $\rightarrow \overline{pizza} = \frac{1}{N} \sum_{i=1}^{N} pizza_i$

S.D. dependent var = 155.8806 $\rightarrow SD_{pizza} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (pizza_i - \overline{pizza})^2}$

S.E. regression = 131.0701 $\rightarrow \hat{\sigma} = \sqrt{\frac{SSR}{N-k}}$

Estimation results.

Gretl output.

Mean dependent var	191.5500	S.D. dependent var	155.8806
Sum squared resid	635636.7	S.E. of regression	131.0701
$R-$squared	0.329251	Adjusted $R-$squared	0.292994
$F(2, 37)$	9.081100	p-value (F)	0.000619
Log-likelihood	-250.2276	Akaike criterion	506.4552
Schwarz criterion	511.5218	Hannan–Quinn	508.2871

The information given by $F(2, 37) = 9.081100$ and the p-value $(F) = 0.000619$ are useful to test the joint significance of the regressors (see Lesson 6).

Estimation results.

Gretl output.

The file `chicken.gdt` contains annual time series data from 1990 to 2012 on:

- Y: annual chicken consumption (in kilograms)
- $X2$: per capital real disposable income (in euros).
- $X3$: price of chicken (in euros/kilogram).
- $X4$: price of pork (in euros/kilogram)
- $X5$: price of beef (in euros/kilogram)
- Avian flue epidemic (1999-2003)

The factors that determine the evolution of chicken consumption will be analysed in detail in the **Example 5.2**. In this section, a very simple model to determine chicken consumption will be estimated in order to explain the Gretl estimation output, focusing on the specific results for time series data.

Estimation results.

Consider the regression model

$$Y_t = \beta_1 + \beta_2\, X2_t + \beta_3\, X3_t + \beta_4\, X4_t + u_t \quad t = 1, ...T$$

where:

- Y: dependent variable (chicken consumption).
- $X2$: quantitative explanatory variable.
- $X3$: quantitative explanatory variable.
- $X4$: quantitative explanatory variable.
- $\beta_1, \beta_2, \beta_3, \beta_4$: unknown coefficients.
- u: error term.
- T: sample size.
- t: index.

Estimation results.

Gretl output.

Model 1: OLS, using observations 1990–2012 ($T = 23$)
Dependent variable: Y

	Coefficient	Std. Error	t-ratio	p-value
const	38.7207	3.61477	10.7118	0.0000
X4	4.37920	1.55013	2.8250	0.0108
X2	0.0109351	0.00236651	4.6208	0.0002
X3	−13.6515	3.91865	−3.4837	0.0025

Mean dependent var	39.66957	S.D. dependent var	7.372950
Sum squared resid	74.77556	S.E. of regression	1.983824
R^2	0.937475	Adjusted R^2	0.927603
$F(3, 19)$	94.95932	P-value(F)	1.28e–11
Log-likelihood	−46.19405	Akaike criterion	100.3881
Schwarz criterion	104.9301	Hannan–Quinn	101.5304
$\hat{\rho}$	0.563845	Durbin–Watson	0.882646

Estimation results.

Gretl output.

Model 1: OLS, using observations 1990–2012 ($T = 23$)
Dependent variable: Y

If the sample consists of time series data, the heading of the estimation output shows the sample frequency:

1990–2012 (T = 23) $\rightarrow$ 23 annual observations

For quarterly data: 1950:1-1955:3 (T = 23).

For monthly data: 1980:01-1981:11 (T = 23).

For daily data: 1950-01-01:1950-06-04 (T = 23).

Estimation results.

Gretl output.

Model 1: OLS, using observations 1990–2012 ($T = 23$)
Dependent variable: Y

	Coefficient	Std. Error	t-ratio	p-value
const	38.7207	3.61477	10.7118	0.0000
X4	4.37920	1.55013	2.8250	0.0108
X2	0.0109351	0.00236651	4.6208	0.0002
X3	−13.6515	3.91865	−3.4837	0.0025

Note that while the regression model is written as:

$$Y_t = \beta_1 + \beta_2 \, X_{2t} + \beta_3 \, X_{3t} + \beta_4 \, X_{4t} + u_t$$

the list of regressors in the estimation table is: const, $X4$, $X2$, $X3$. This is because the regressors appear in the table in the same order we introduce them in the Gretl specification window (see Example 5.1.1). This fact has to be taken into account to read the results, that is, $\hat{\beta}_2 = 0.0109351$ and $\hat{\beta}_4 = 4.3792$.

Estimation results.

Gretl output.

Mean dependent var	39.66957	S.D. dependent var	7.372950
Sum squared resid	74.77556	S.E. of regression	1.983824
R^2	0.937475	Adjusted R^2	0.927603
$F(3, 19)$	94.95932	P-value(F)	1.28e–11
Log-likelihood	−46.19405	Akaike criterion	100.3881
Schwarz criterion	104.9301	Hannan–Quinn	101.5304
$\hat{\rho}$	0.563845	Durbin–Watson	0.882646

$$\hat{\rho} = \frac{\sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=2}^{T} \hat{u}_{t-1}^2} \qquad \text{Durbin-Watson: } \frac{\sum_{t=2}^{T} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{T} \hat{u}_t^2}$$

These results appear only when the data set structure is time series. The Durbin-Watson autocorrelation test will be studied in Lesson 7.

Contents

Presentation of results.

The estimation results are usually summarized writing down the Sample Regression Function along with the standard error of the estimators, the coefficient of determination and the Sum of Squares Residuals.

Summary.

$$\widehat{pizza}_i = \underset{(72.343)}{342.885} + \underset{(0.46430)}{1.83248}\, income_i - \underset{(2.3170)}{7.57556}\, age_i \qquad i = 1, \ldots, 40$$

$$R^2 = 0.329 \qquad SSR = 635636.7$$

(standard error in parentheses)

Presentation of results.

If we are working with time series data, the same information is presented but adding the value of the Durbin-Watson statistic (see Lesson 7).

Summary.

$$\widehat{\mathrm{Y}}_t = \underset{(3.614476)}{38.7207} + \underset{(1.5501)}{4.37920}\ \mathsf{X4}_t + \underset{(0.0023665)}{0.0109351}\ \mathsf{X2}_t - \underset{(3.91865)}{13.6515}\ \mathsf{X3}_t$$

$$t = 1990, \ldots, 2012 \quad R^2 = 0.937475 \quad SCR = 74.77556 \quad DW = 0.8832646$$

(standard error in parentheses)

Contents

Contents