A.IV. Linear regression

Olivier Scaillet

University of Geneva and Swiss Finance Institute

Outline

1. Introduction
2. The OLS Estimator
3. Goodness-of-fit Measures
4. Statistical Inference & Tests

Introduction

A *linear regression* is defined by

$$\begin{aligned} Y_t &= X_t\beta + \varepsilon_t \\ &= \sum_{k=1}^{K} X_{k,t}\beta_k + \varepsilon_t \end{aligned}$$

where

ε_t = noise with mean zero, i.e., innovation or error term,

$$\begin{aligned} X_t\beta &= E\left[Y_{i,t}\,|X_t\right] \\ &= \text{conditional mean of } Y_t \\ &= \text{linear function of } X_t \end{aligned}$$

β = parameter (estimation by OLS)

OLS Estimator

The *OLS estimator* is the value of β which minimizes the sum of squared residuals

$$\hat{\beta} = \arg\min \sum_{t=1}^{T} (Y_t - X_t\beta)^2$$
$$= \left[\sum_{t=1}^{T} X_t' X_t\right]^{-1} \sum_{t=1}^{T} X_t' Y_t$$

The *OLS residuals* are given by

$$\hat{\varepsilon}_t = Y_t - X_t\hat{\beta}$$

OLS Estimator

Matrix notation:

$$\underset{(T\times 1)}{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_T \end{pmatrix}, \underset{(T\times K)}{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_T \end{pmatrix}, \underset{(T\times 1)}{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix}$$

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

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

Interpretation

Interpretation:
OLS consists in decomposing Y into two *orthogonal projections*, one on the space spanned by the columns of X, and the other on the space of innovations.

Let $P_X = X(X'X)^{-1}X'$ be the orthogonal projection matrix associated with the regressors, and

$$M_X = Id - P_X,$$

then

$$\begin{aligned} Y &= P_X Y + M_X Y \\ &= X(X'X)^{-1}X'Y + (Id - X(X'X)^{-1}X')Y \\ &= X\hat{\beta} + (Y - X\hat{\beta}) \\ &= X\hat{\beta} + \hat{\varepsilon} \end{aligned}$$

Goodness-of-fit Measures

When there is a constant in the model (X_1 is a vector of ones), a measure of goodness-of-fit is

$$R^2 = \frac{\sum_{t=1}^{T}\left(\hat{Y}_t - \bar{Y}\right)^2}{\sum_{t=1}^{T}\left(Y_t - \bar{Y}\right)^2} = \frac{\text{EXPLAINED VARIANCE}}{\text{TOTAL VARIANCE}}.$$

It satisfies $0 < R^2 < 1$.

Statistical Inference

Statistical inference:

H1: a) X_t is deterministic

b) ε_t is i.i.d. with mean 0 and variance σ^2

c) ε_t is Gaussian

$\hat{\beta}$ is normally distributed with mean β and covariance matrix $\sigma^2(X'X)^{-1}$

Gauss-Markov Theorem

Gauss Markov theorem:

OLS estimator

$$= \text{Best Linear Unbiased Estimator (BLUE)}$$

OLS estimator has the smallest variance (largest precision) among all unbiased estimators linear in Y

t-Tests

t-Tests:

$$\begin{cases} H_0 : \beta_k = b \\ H_1 : \beta_k \neq b \end{cases}$$

For example, take $b = 0$ to obtain the test of the significance of the presence of the k-th regressor in explaining Y.

t-Statistic

t-Statistic: $t = \frac{\hat{\beta}_k - b}{\hat{\sigma}_{\hat{\beta}_k}}$, where

$$\hat{\sigma}_{\hat{\beta}_k} = \sqrt{s^2 \varsigma_{ii}},$$

with

$$s^2 = \frac{\hat{\varepsilon}' \hat{\varepsilon}}{T - K}$$
$$\varsigma_{ii} = \text{i-th diagonal element of } (X'X)^{-1}$$

Reject the null hypothesis H_0 at level α if

$$|t| > t_{1-\alpha/2}$$

where $t_{1-\alpha/2}$ is the 1 $- \alpha/2$ quantile of a student distribution with $T - K$ degrees of freedom

F tests

F tests:

$$\begin{cases} H_0 : R\beta = r \\ H_1 : R\beta \neq r \end{cases}$$

Test of m linear restrictions on β

$R = (m \times K)$ matrix, $r = (m \times 1)$ vector

F statistic:

$$F = \frac{(R\hat{\beta} - r)'(R(X'X)^{-1}R')(R\hat{\beta} - r)}{s^2 m},$$

Reject the null hypothesis H_0 at level α if

$F > F_{1-\alpha}$

where $F_{1-\alpha}$ is the quantile of level $1-\alpha$ of a Fisher variable with m and $T-K$ degrees of freedom.

Other Sets of Assumptions

Other sets of assumptions:

H2: a) X_t is stochastic
and independent of ε_s, $\forall s$

b) ε_t is i.i.d. with mean 0
and variance σ^2

c) ε_t is Gaussian

$\hat{\beta}$ is normally distributed with mean β and covariance matrix $\sigma^2(X'X)^{-1}$,but conditionally to X (when X is treated as fixed).

Unconditionally, $\hat{\beta}$ is no longer normally distributed.

However the distributions for t and F remain valid.

Remark

Remark:

For large T (asymptotic theory), in H2,

we may drop the assumption of Gaussian innovations,

but the distributions for t and F become

a) standard normal for t

b) chi-square with m degrees of freedom for F