Econometrics

Properties of OLS and the FWL Theorem

J. Eduardo Vera-Valdés
eduardo@math.aau.dk

Department of Mathematical Sciences
Aalborg University
Denmark

Outline

J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

Outline

J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

Formalities

Course: 10 lectures followed by exercise sessions and 2 self-studies. (1 extra "in case" lecture).

Evaluation: Oral exam without preparation.

Literature: "Econometric Theory and Methods" (ETM) by Russell Davidson and James G. MacKinnon.

Alternatively, "Econometric Analysis" (EA) by William H. Greene for a more analytical perspective, and "A Guide to Modern Econometrics" (AGME) by Marno Verbeek for a more intuitive approach.

Software: **R** via RStudio.

Data: The "Ecdat" package in R, but more importantly, data from your projects if appropriate.

J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

Curriculum

Knowledge

- understanding of the most common estimation methods in econometrics.
- knowledge about finite sample and asymptotic properties of the ordinary least squares estimator.
- knowledge about method of moments estimators, e.g., the instrumental variables estimator.
- knowledge about models for panel data, e.g., random effects and fixed effects estimators.
- knowledge about models for discrete and limited dependent variables, e.g., censored data and sample selection.
- knowledge about estimation of multivariate models and systems of equations.
- knowledge about simulation methods for inference, e.g., Monte Carlo and bootstrapped tests.

J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

Curriculum

Skills

- are able to argue for the importance of using econometric/statistical methods in the analysis of a given economic problem.
- are able to build econometric models and judge their applicability.

J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

Curriculum

Competencies

- are able to demonstrate deep understanding of the theory of econometric models and know how to reason within the models.
- are able to communicate the results of an econometric analysis to non-specialists.
- are able to analyse economic data using the available software.

J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

Schedule (subject to change)

Lectures on Wednesdays from 8.15 to 12.00 (two exceptions noted below)

Week	Subject
1	Properties of OLS and FWL Theorem
2	Misspecification
3	GLS and FGLS
4	Instrumental Variables
5	Panel Data Models
6	Self-study
7	Seemingly Unrelated Regressions
8	Simultaneous Equations Models
9	GARCH Models*
10	Cointegration*
11	Self-study
12	Censored and Truncated Data

J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

Outline

J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

Ordinary Least Squares

Assumptions for OLS:

- Correct specification
- Linear independence
- Exogeneity
- Homoskedasticity
- No autocorrelation
- Normality

J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

Orthogonal Projections

When we estimate a regression model, we map Y into a vector of fitted values $X\hat{\beta}$ and a vector of residuals $\hat{U}$.

OLS uses the two projection matrices:

$$P_X = X(X^T X)^{-1} X^T, \qquad \text{and} \qquad M_X = I - P_X.$$

And make,

$$X\hat{\beta} = P_X Y, \qquad \text{and} \qquad \hat{U} = M_X Y.$$

It is easy to show that P_X and M_X are orthogonal projection matrices, and that $P_X X = X$, and $M_X X = 0$.

Furthermore, they are **complementary projections** since $P_X M_X = 0$ (which you will show).

J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

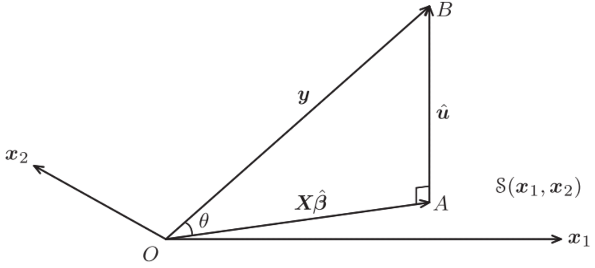
Orthogonal Projections

OLS decomposes Y in

$$Y = P_X Y + M_X Y,$$

where $X\hat{\beta} = P_X Y$ are the fitted values, and $\hat{U} = M_X Y$ are the residuals.

This decomposition can be represented by a right-angled triangle.



Orthogonal Projections

By Pythagora's Theorem,

$$||Y||^2 = ||P_X Y||^2 + ||M_X Y||^2,$$

so that

$$||P_X Y||^2 \leq ||Y||^2, \text{ and } ||\hat{U}||^2 \leq ||U||^2.$$

Moreover, note that

$$X^T \hat{U} = X^T M_X Y = 0,$$

so that **OLS residuals are orthogonal to all the regressors.**

In particular, if the regressors include a constant, then the residuals sum to zero, $\sum_{t=1}^{N} \hat{U}_t = 0$.

Outline

J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

Assume we broke up the regressors into two groups $X = [X_1 \ X_2]$ so that

$$Y = X_1\beta_1 + X_2\beta_2 + U,$$

we are interested in analysing the effect of this partition on the estimators.

First, suppose that X_1 is orthogonal to X_2, then we obtain the same estimate for β_1 above than the one using the regression

$$Y = X_1\beta_1 + V.$$

Analogously, we obtain the same estimate for β_2 if we remove X_1 from the regression.

Econometrics
J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

J. Eduardo Vera-Valdés
eduardo@math.aau.

In the general case, the two regressions

$$Y = X_1\beta_1 + M_{X_1}X_2\beta_2 + U,$$

and

$$Y = M_{X_1}X_2\beta_2 + V,$$

yield identical estimates for β_2.

Nonetheless, they do not yield the same residuals.

Department of Mathematical Sciences
Aalborg University
Denmark

The Frisch-Waugh-Lovell Theorem

J. Eduardo Vera-Valdés
eduardo@math.aau.

The FWL Theorem

The OLS estimates of β_2 in the regressions

$$Y = X_1\beta_1 + X_2\beta_2 + U,$$

and

$$M_{X_1}Y = M_{X_1}X_2\beta_2 + U,$$

are numerically identical.

Moreover, the residuals in both regressions are numerically identical.

Department of Mathematical Sciences
Aalborg University
Denmark

Assume the regression includes a constant

$$Y = \iota\beta_0 + X\beta_1 + U,$$

where ι is a vector of ones.

Then the FWL theorem states that the estimator from β_1 is the same if we instead run the regression

$$M_\iota Y = M_\iota X\beta_1 + U.$$

Econometrics
J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

Econometrics
J. Eduardo Vera-Valdés
eduardo@math.aau.

Many economic activities are strongly affected by the season, for obvious reasons like Christmas shopping, or summer holidays.

It is thus common to model this seasonality using seasonal dummy variables.

$$Y = \alpha_1 s_1 + \alpha_2 s_2 + \alpha_3 s_3 + \alpha_4 s_4 + X\beta + U,$$

where are the s_i seasonal dummy variables.

The FWL theorem tells us that we can estimate β using the raw data and the above regression or by a simpler regression without the dummy variables using deseasonalised (by regression) data.

Department of Mathematical Sciences
Aalborg University
Denmark

Another set of variables that are often encountered are time trends.

We consider a regression like

$$Y = \alpha_0 \iota + \alpha_1 T + X\beta + U,$$

where $T^T = [1, 2, \cdots, N]$.

An increasing time trend might be appropriate, for instance, in a model of a production function where technical progress is taking place.

The FWL Theorem shows that it is possible to use detrended variables.

J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

Recall that OLS decomposes Y in

$$Y = P_X Y + M_X Y,$$

where $P_X Y$ is the part that X is able to explain from Y.

We define the coefficient of determination or (uncentered) R^2 by

$$R^2 = \frac{||P_X Y||^2}{||Y||^2},$$

which fall between 0 and 1 given what we previously discussed.

It is invariant under nonsingular linear transformations of the regressors and, because it is defined as a ratio, it is invariant to changes in the scale of Y.

Econometrics

J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

Nonetheless, the (uncentered) R^2 is not invariant to translations

$$\tilde{Y} := Y + \alpha\iota = X\beta + U.$$

If X includes a constant, we have that

$$\tilde{Y} = P_X(\tilde{Y}) + M_X(\tilde{Y}) = P_X Y + \alpha\iota + M_X Y,$$

so that

$$R^2 = \frac{||P_X Y + \alpha\iota||^2}{||Y + \alpha\iota||^2}.$$

By choosing α, we can make the R^2 as close to 1 as we want.

J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

The last result is misleading because the fit is accounted almost exclusively by the constant.

To avoid this problem, for regressions that **include a constant**, the FWL theorem tells us that we can demean the series without changing the estimates or residuals.

This gives raise to the (centered) R^2 defined as

$$R^2 = \frac{||P_X M_\iota Y||^2}{||M_\iota Y||^2},$$

which is unaffected by translations.

Econometrics
J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

Outline

J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences Aalborg University Denmark

OLS is BLUE

J. Eduardo Vera-Valdés
eduardo@math.aau.

The Gauss-Markov Theorem

In a regression under correct specification, exogenous regressors, homoskedastic and no-autocorrelated errors, then the OLS estimator is more efficient than any other linear unbiased estimator.

In this sense, the OLS estimator is the Best Linear Unbiased Estimator, or BLUE.

Department of Mathematical Sciences
Aalborg University
Denmark

Bias

To show that $E[(X^TX)^{-1}X^TU] = 0$, and thus that $\hat{\beta}$ is unbiased, there are two possibilities.

(a) X is **nonstochastic** and U has mean zero.

(b) X is **exogenous**, i.e., $E[U|X] = 0$.

J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

Bias

The nonstochastic assumption may be sensible when we are able to control the inputs, like in an experimental setting.

Nonetheless, it is often not a reasonable assumption to make in applied econometric work.

The exogeneity assumption is reasonable in the context of some types of data like cross-sectional data in which each observation might correspond to an individual household or person.

On the other hand, for time-series data, in which each observation might correspond to a year, quarter, or day, the exogeneity assumption is a very strong one.

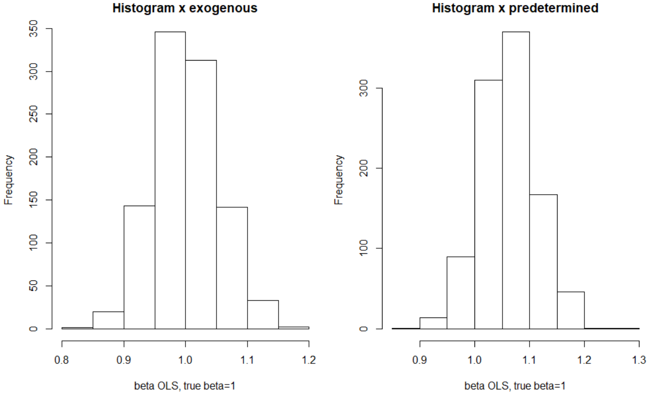
J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

Bias

The exogeneity assumption imposes that all errors are not related to past and future values of the regressors.

For instance, $\hat{\beta}$ is biased under the weaker assumption of **predetermined** regressors, $E(U_t|X_t) = 0$.



Consistency

Another statistical property we may want from our estimators is consistency; that is, we want the estimator to asymptotically achieve the true value.

Thus, we are interested in whether

$$\underset{n\to\infty}{\text{plim}}\,\hat{\beta} = \beta$$

Recalling that we have assumed a correct specification, and the more general stochastic regressors assumptions, consistency requires that

$$\underset{n\to\infty}{\text{plim}}(X^TX)^{-1}X^TU = 0$$

J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

Consistency

Hence, consistency requires:

(a) $\operatorname{plim}_{n\to\infty}(X^TX)^{-1}$ exists, in the sense of having a finite limit, possibly after some standardization.

(b) $\operatorname{plim}_{n\to\infty} X^TU = 0$, possibly after some standardization.

J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

Consistency

Given that we are interested in the limit as the sample size n increases, we would expect the standardization terms to depend on n.

Note that

$$(X^T X)^{-1} X^T U = \left(\frac{1}{n} X^T X\right)^{-1} \left(\frac{1}{n} X^T U\right).$$

So that we look for conditions so that $\text{plim}_{n\to\infty} \left(\frac{1}{n} X^T X\right)^{-1} = S_{XX} < \infty$ and $\text{plim}_{n\to\infty} \frac{1}{n} X^T U = 0$.

J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

Consistency

Using the law of large numbers, we can show that **predeterminedness**, $E(U_t|X_t)$, implies

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n} X^T U = 0.$$

Formally, the proof requires the conditions for a law of large numbers to apply.

Thus, the errors cannot be too correlated or have unbounded variances.

J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

Consistency

Another use of the law of large numbers can be used to show that

$$\operatorname*{plim}_{n\to\infty}\left(\frac{1}{n}X^TX\right)^{-1} = S_{XX} < \infty.$$

So that we are implicitly assuming that the regressors are not too correlated, and the variances do not increase without bound.

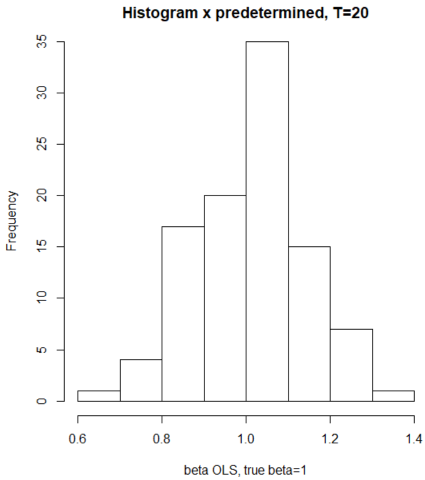
Typically, these conditions are satisfied by most type of regressors.

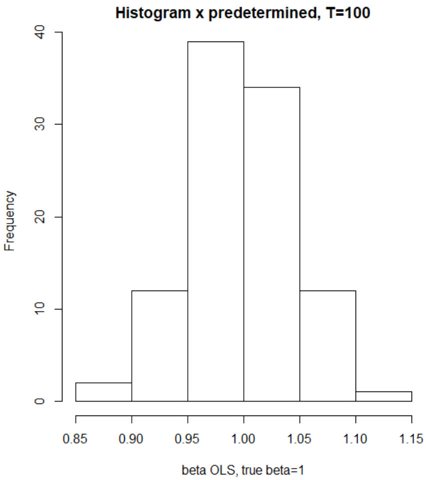
Yet, these are some regressors where this is not satisfied.

J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

Consistency





J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

Precision

Turning our attention now to analyse the second moment of the OLS estimator, we measure the precision of the OLS estimator by its covariance matrix.

In order to derive some properties of the covariance matrix of the OLS estimator, we must make assumptions regarding the second moments of the error term.

The usual assumptions about the errors' second moment of **homoskedasticity** and **no autocorrelation** can be written as

$$Var(U|X) = E(UU^T|X) = \sigma^2\mathbb{I},$$

where $\mathbb{I}$ is a $n \times n$ identity matrix.

Precision

Under the assumptions on the error term, and **exogeneity** of the regressors, the variance of the OLS estimator is

$$Var(\hat{\beta}) = \sigma^2(X^TX)^{-1}.$$

Thus, the precision of the OLS estimator can be shown to depend on three things:

- The variance of the error term
- The sample size
- The relation between the regressors

Precision

The dependence of the variance of the OLS estimator on the variance of the **homoskedastic** error term is straightforward.

The dependence on the sample size can be seen if we write

$$Var(\hat{\beta}) = \sigma^2(X^T X)^{-1} = \left(\frac{1}{n}\sigma^2\right)\left(\frac{1}{n}X^T X\right)^{-1},$$

and assuming, as before, that $\text{plim}_{n\to\infty}\left(\frac{1}{n}X^T X\right)^{-1} = S_{XX}$.

J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

Precision

To show the dependence on the relation between regressors consider

$$y = x_1\beta_1 + X_2\beta_2 + U,$$

where $X = [x_1,\ X_2]$, and x_1 is a column vector.

From the FWL theorem, $\hat{\beta}_1$ is the same as the one from

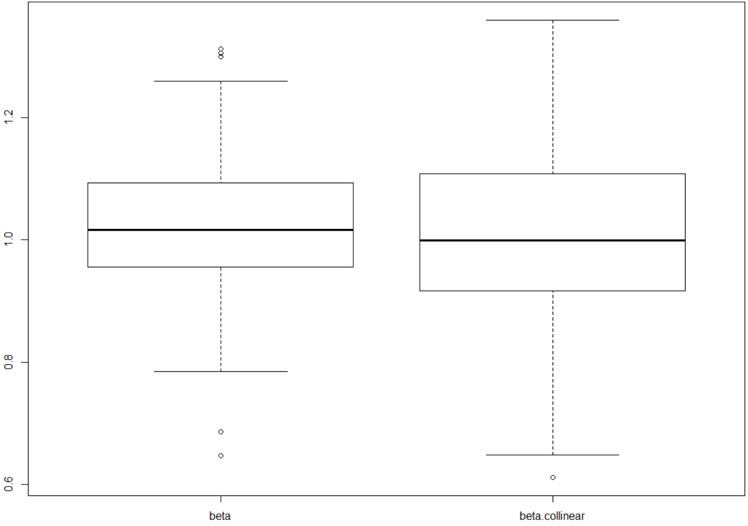
$$M_{X_2}y = M_{X_2}x_1\beta_1 + V.$$

Thus,

$$Var(\hat{\beta}_1) = \sigma^2/(x_1^T M_{X_2}x_1),$$

which shows its dependence on the relation between the regressors.

Precision



J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

Outline

J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark

Summing Up

- We have seen that OLS uses orthogonal projections to decompose the regressand in fitted values and residuals.
- The residuals are orthogonal to all the regressors.
- If the regression includes a constant, the residuals have mean zero.
- We study the FWL theorem and applications to demeaned series, seasonality, and trends.
- Showed numerical properties of the R^2.
- Under correct specification and exogenous regressors, OLS is an unbiased estimator.
- Under just predetermined regressors, OLS are consistent (with "well-behaved" regressors).

J. Eduardo Vera-Valdés
eduardo@math.aau.

Department of Mathematical Sciences
Aalborg University
Denmark