

INTRODUCTORY ECONOMETRICS' GLOSSARY

JFM

A

Adjusted R-Squared: A goodness-of-fit measure in multiple regression analysis that penalises additional explanatory variables by using a degrees of freedom adjustment in estimating the error variance.

Alternative Hypothesis: The hypothesis against which the null hypothesis is tested.

AR(1) Serial Correlation: The errors in a time series regression model follow an AR(1) model.

Attenuation Bias: Bias in an estimator that is always toward zero; thus, the expected value of an estimator with attenuation bias is less in magnitude than the absolute value of the parameter.

Autocorrelation: See serial correlation.

Autoregressive Process of Order One [AR(1)]: A time series model whose current value depends linearly on its most recent value plus an unpredictable disturbance.

Auxiliary Regression: A regression used to compute a test statistic-such as the test statistics for heteroskedasticity and serial correlation or any other regression that does not estimate the model of primary interest.

Average: The sum of n numbers divided by n .

B

Base Group: The group represented by the overall intercept in a multiple regression model that includes dummy explanatory variables.

Benchmark Group: See base group.

Bernoulli Random Variable: A random variable that takes on the values zero or one.

Best Linear Unbiased Estimator (BLUE): Among all linear unbiased estimators, the estimator with the smallest variance. OLS is BLUE, conditional on the sample values of the explanatory variables, under the Gauss-Markov assumptions.

Beta Coefficients: See standardised coefficients.

Bias: The difference between the expected value of an estimator and the population value that the estimator is supposed to be estimating.

Biased Estimator: An estimator whose expectation, or sampling mean, is different from the population value it is supposed to be estimating.

Biased Towards Zero: A description of an estimator whose expectation in absolute value is less than the absolute value of the population parameter.

Binary Response Model: A model for a binary (dummy) dependent variable.

Binary Variable: See dummy variable.

Binomial Distribution: The probability distribution of the number of successes out of n independent Bernoulli trials, where each trial has the same probability of success.

Bivariate Regression Model: See simple linear regression model.

BLUE: See best linear unbiased estimator.

C

Causal Effect: A ceteris paribus change in one variable has an effect on another variable.

Ceteris Paribus: All other relevant factors are held fixed.

Chi-Square Distribution: A probability distribution obtained by adding the squares of independent standard normal random variables. The number of terms in the sum equals the degrees of freedom in the distribution.

Classical Errors-in-Variables (CEV): A measurement error model where the observed measure equals the actual variable plus an independent, or at least an uncorrelated, measurement error.

Classical Linear Model: The multiple linear regression model under the full set of classical linear model assumptions.

Date: © UPV/EHU 2005.

Excerpted from Wooldridge, J.M., (2003), *Introductory Econometrics, 2nd ed.*, Thomson.

Classical Linear Model (CLM) Assumptions: The ideal set of assumptions for multiple regression analysis. The assumptions include linearity in the parameters, no perfect collinearity, the zero conditional mean assumption, homoskedasticity, no serial correlation, and normality of the errors.

Coefficient of Determination: See R-squared.

Conditional Distribution: The probability distribution of one random variable, given the values of one or more other random variables.

Conditional Expectation: The expected or average value of one random variable, called the dependent or explained variable, that depends on the values of one or more other variables, called the independent or explanatory variables.

Conditional Forecast: A forecast that assumes the future values of some explanatory variables are known with certainty.

Conditional Variance: The variance of one random variable, given one or more other random variables.

Confidence Interval (CI): A rule used to construct a random interval so that a certain percentage of all data sets, determined by the confidence level, yields an interval that contains the population value.

Confidence Level: The percentage of samples in which we want our confidence interval to contain the population value; 95% is the most common confidence level, but 90% and 99% are also used.

Consistent Estimator: An estimator that converges in probability to the population parameter as the sample size grows without bound.

Consistent Test: A test where, under the alternative hypothesis, the probability of rejecting the null hypothesis converges to one as the sample size grows without bound.

Constant Elasticity Model: A model where the elasticity of the dependent variable, with respect to an explanatory variable, is constant; in multiple regression, both variables appear in logarithmic form.

Continuous Random Variable: A random variable that takes on any particular value with probability zero.

Control Variable: See explanatory variable.

Correlation Coefficient: A measure of linear dependence between two random variables that does not depend on units of measurement and is bounded between -1 and 1 .

Count Variable: A variable that takes on nonnegative integer values.

Covariance: A measure of linear dependence between two random variables.

Covariate: See explanatory variable.

Critical Value: In hypothesis testing, the value against which a test statistic is compared to determine whether or not the null hypothesis is rejected.

Cross-Sectional Data Set: A data set collected from a population at a given point in time.

Cumulative Distribution Function (cdf): A function that gives the probability of a random variable being less than or equal to any specified real number.

D

Data Frequency: The interval at which time series data are collected. Yearly, quarterly, and monthly are the most common data frequencies.

Degrees of Freedom (df): In multiple regression analysis, the number of observations minus the number of estimated parameters.

Denominator Degrees of Freedom: In an F test, the degrees of freedom in the unrestricted model.

Dependent Variable: The variable to be explained in a multiple regression model (and a variety of other models).

Descriptive Statistic: A statistic used to summarise a set of numbers; the sample average, sample median, and sample standard deviation are the most common.

Deseasonalizing: The removing of the seasonal components from a monthly or quarterly time series.

Detrending: The practice of removing the trend from a time series.

Difference in Slopes: A description of a model where some slope parameters may differ by group or time period.

Discrete Random Variable: A random variable that takes on at most a finite or countably infinite number of values.

Distributed Lag Model: A time series model that relates the dependent variable to current and past values of an explanatory variable.

Disturbance: See error term.

Downward Bias: The expected value of an estimator is below the population value of the parameter.

Dummy Dependent Variable: See binary response model.

Dummy Variable: A variable that takes on the value zero or one.

Dummy Variable Regression: In a panel data setting, the regression that includes a dummy variable for each cross-sectional unit, along with the remaining explanatory variables. It produces the fixed effects estimator.

Dummy Variable Trap: The mistake of including too many dummy variables among the independent variables; it occurs when an overall intercept is in the model and a dummy variable is included for each group.

Durbin-Watson (DW) Statistic: A statistic used to test for first order serial correlation in the errors of a time series regression model under the classical linear model assumptions.

E

Econometric Model: An equation relating the dependent variable to a set of explanatory variables and unobserved disturbances, where unknown population parameters determine the *ceteris paribus* effect of each explanatory variable.

Economic Model: A relationship derived from economic theory or less formal economic reasoning.

Economic Significance: See practical significance.

Elasticity: The percent change in one variable given a 1% *ceteris paribus* increase in another variable.

Empirical Analysis: A study that uses data in a formal econometric analysis to test a theory, estimate a relationship, or determine the effectiveness of a policy.

Endogeneity: A term used to describe the presence of an endogenous explanatory variable.

Endogenous Explanatory Variable: An explanatory variable in a multiple regression model that is correlated with the error term, either because of an omitted variable, measurement error, or simultaneity.

Endogenous Variables: In simultaneous equations models, variables that are determined by the equations in the system.

Error Term: The variable in a simple or multiple regression equation that contains unobserved factors that affect the dependent variable. The error term may also include measurement errors in the observed dependent or independent variables.

Error Variance: The variance of the error term in a multiple regression model.

Errors-in-Variables: A situation where either the dependent variable or some independent variables are measured with error.

Estimate: The numerical value taken on by an estimator for a particular sample of data.

Estimator: A rule for combining data to produce a numerical value for a population parameter; the form of the rule does not depend on the particular sample obtained.

Event Study: An econometric analysis of the effects of an event, such as a change in government regulation or economic policy, on an outcome variable.

Excluding a Relevant Variable: In multiple regression analysis, leaving out a variable that has a nonzero partial effect on the dependent variable.

Exclusion Restrictions: Restrictions which state that certain variables are excluded from the model (or have zero population coefficients).

Exogenous Explanatory Variable: An explanatory variable that is uncorrelated with the error term.

Exogenous Variable: Any variable that is unconnected with the error term in the model of interest.

Expected Value: A measure of central tendency in the distribution of a random variable, including an estimator.

Experiment: In probability, a general term used to denote an event whose outcome is uncertain. In econometric analysis, it denotes a situation where data are collected by randomly assigning individuals to control and treatment groups.

Experimental Data: Data that have been obtained by running a controlled experiment.

Explained Sum of Squares (ESS): The total sample variation of the fitted values in a multiple regression model.

Explained Variable: See dependent variable.

Explanatory Variable: In regression analysis, a variable that is used to explain variation in the dependent variable.

Exponential Function: A mathematical function defined for all values that has an increasing slope but a constant proportionate change.

F

F Distribution: The probability distribution obtained by forming the ratio of two independent chi-square random variables, where each has been divided by its degrees of freedom.

F Statistic: A statistic used to test multiple hypotheses about the parameters in a multiple regression model.

First Difference: A transformation on a time series constructed by taking the difference of adjacent time periods, where the earlier time period is subtracted from the later time period.

First Order Autocorrelation: For a time series process ordered chronologically, the correlation coefficient between pairs of adjacent observations.

First Order Conditions: The set of linear equations used to solve for the OLS estimates.

Fitted Values: The estimated values of the dependent variable when the values of the independent variables for each observation are plugged into the OLS regression line.

Forecast Error: The difference between the actual outcome and the forecast of the outcome.

Forecast Interval: In forecasting, a confidence interval for a yet unrealised future value of a time series variable. (See also prediction interval.)

Functional Form Misspecification: A problem that occurs when a model has omitted functions of the explanatory variables (such as quadratics) or uses the wrong functions of either the dependent variable or some explanatory variables.

G

Gauss-Markov Assumptions: The set of assumptions under which OLS is BLUE.

Gauss-Markov Theorem: The theorem which states that, under the five Gauss-Markov assumptions (for cross-sectional or time series models), the OLS estimator is BLUE (conditional on the sample values of the explanatory variables).

General Linear Regression (GLR) Model: A model linear in its parameters, where the dependent variable is a function of independent variables plus an error term.

Goodness-of-Fit Measure: A statistic that summaries how well a set of explanatory variables explains a dependent or response variable.

Growth Rate: The proportionate change in a time series from the previous period. It may be approximated as the difference in logs or reported in percentage form.

H

Heteroskedasticity: The variance of the error term, given the explanatory variables, is not constant.

Homoskedasticity: The errors in a regression model have constant variance, conditional on the explanatory variables.

Hypothesis Test: A statistical test of the null, or maintained, hypothesis against an alternative hypothesis.

I

Impact Elasticity: In a distributed lag model, the immediate percentage change in the dependent variable given a 1% increase in the independent variable.

Impact Multiplier: See impact propensity.

Impact Propensity: In a distributed lag model, the immediate change in the dependent variable given a one-unit increase in the independent variable.

Inclusion of an Irrelevant Variable: The including of an explanatory variable in a regression model that has a zero population parameter in estimating an equation by OLS.

Inconsistency: The difference between the probability limit of an estimator and the parameter value.

Independent Random Variables: Random variables whose joint distribution is the product of the marginal distributions.

Independent Variable: See explanatory variable.

Index Number: A statistic that aggregates information on economic activity, such as production or prices.

Infinite Distributed Lag (IDL) Model: A distributed lag model where a change in the explanatory variable can have an impact on the dependent variable into the indefinite future.

Influential Observations: See outliers.

Information Set: In forecasting, the set of variables that we can observe prior to forming our forecast.

In-Sample Criteria: Criteria for choosing forecasting models that are based on goodness-of-fit within the sample used to obtain the parameter estimates.

Interaction Effect: In multiple regression, the partial effect of one explanatory variable depends on the value of a different explanatory variable.

Interaction Term: An independent variable in a regression model that is the product of two explanatory variables.

Intercept Parameter: The parameter in a multiple linear regression model that gives the expected value of the dependent variable when all the independent variables equal zero.

Intercept Shift: The intercept in a regression model differs by group or time period.

Interval Estimator: A rule that uses data to obtain lower and upper bounds for a population parameter. (See also confidence interval.)

J

Joint Distribution: The probability distribution determining the probabilities of outcomes involving two or more random variables.

Joint Hypothesis Test: A test involving more than one restriction on the parameters in a model.

Jointly Statistically Significant: The null hypothesis that two or more explanatory variables have zero population coefficients is rejected at the chosen significance level.

L

Lag Distribution: In a finite or infinite distributed lag model, the lag coefficients graphed as a function of the lag length.

Lagged Dependent Variable: An explanatory variable that is equal to the dependent variable from an earlier time period.

Lagged Endogenous Variable: In a simultaneous equations model, a lagged value of one of the endogenous variables.

Least Absolute Deviations: A method for estimating the parameters of a multiple regression model based on minimising the sum of the absolute values of the residuals.

Level-Level Model: A regression model where the dependent variable and the independent variables are in level (or original) form.

Level-Log Model: A regression model where the dependent variable is in level form and (at least some of) the independent variables are in logarithmic form.

Linear Function: A function where the change in the dependent variable, given it one-unit change in an independent variable, is constant.

Linear Unbiased Estimator: In multiple regression analysis, an unbiased estimator that is a linear function of the outcomes on the dependent variable.

Logarithmic Function: A mathematical function defined for positive arguments that has a positive, but diminishing, slope.

Log-Level Model: A regression model where the dependent variable is in logarithmic form and the independent variables are in level (or original) form.

Log-Log Model: A regression model where the dependent variable and (at least some of) the explanatory variables are in logarithmic form.

Long-Run Elasticity: The long-run propensity in a distributed lag model with the dependent and independent variables in logarithmic form; thus, the long-run elasticity is the eventual percentage increase in the explained variable, given a permanent 1% increase in the explanatory variable.

Long-Run Multiplier: See long-run propensity.

Long-Run Propensity: In a distributed lag model, the eventual change in the dependent variable given a permanent, one-unit increase in the independent variable.

Longitudinal Data: See panel data.

M

Marginal Effect: The effect on the dependent variable that results from changing an independent variable by a small amount.

Matrix: An array of numbers.

Matrix Notation: A convenient mathematical notation, grounded in matrix algebra, for expressing and manipulating the multiple regression model.

Mean: See expected value.

Mean Absolute Error (MAE): A performance measure in forecasting, computed as the average of the absolute values of the forecast errors.

Mean Squared Error: The expected squared distance that an estimator is from the population value; it equals the variance plus the square of any bias.

Measurement Error: The difference between an observed variable and the variable that belongs in a multiple regression equation.

Median: In a probability distribution, it is the value where there is a 50% chance of being below the value and a 50% chance of being above it. In a sample of numbers, it is the middle value after the numbers have been ordered.

Method of Moments Estimator: An estimator obtained by using the sample analog of population moments; ordinary least squares and two stage least squares are both method of moments estimators.

Minimum Variance Unbiased Estimator: An estimator with the smallest variance in the class of all unbiased estimators.

Missing Data: A data problem that occurs when we do not observe values on some variables for certain observations (individuals, cities, time periods, and so on) in the sample.

Misspecification Analysis: The process of determining likely biases that can arise from omitted variables, measurement error, simultaneity, and other kinds of model misspecification.

Multicollinearity: A term that refers to correlation among the independent variables in a multiple regression model; it is usually invoked when some correlations are "large," but an actual magnitude is not well-defined.

Multiple Hypothesis Test: A test of a null hypothesis involving more than one restriction on the parameters.

Multiple Linear Regression (MLR) Model: See general linear regression model.

Multiple Regression Analysis: A type of analysis that is used to describe estimation of and inference in the multiple linear regression model.

Multiple Restrictions: More than one restriction on the parameters in an econometric model.

Multiple Step-Ahead Forecast: A time series forecast of more than one period into the future.

Multiplicative Measurement Error: Measurement error where the observed variable is the product of the true unobserved variable and a positive measurement error.

N

Natural Logarithm: See logarithmic function.

Nominal Variable: A variable measured in nominal or current euros.

Nonexperimental Data: Data that have not been obtained through a controlled experiment.

Nonlinear Function: A function whose slope is not constant.

Normal Distribution: A probability distribution commonly used in statistics and econometrics for modelling a population. Its probability distribution function has a bell shape.

Normality Assumption: The classical linear model assumption which states that the error (or dependent variable) has a normal distribution, conditional on the explanatory variables.

Null Hypothesis: In classical hypothesis testing, we take this hypothesis as true and require the data to provide substantial evidence against it. Numerator Degrees of Freedom: In an F test, the number of restrictions being tested.

O

Observational Data: See nonexperimental data.

OLS: See ordinary least squares.

OLS Intercept Estimate: The intercept in an OLS regression line.

OLS Regression Line: The equation relating the predicted value of the dependent variable to the independent variables, where the parameter estimates have been obtained by OLS.

OLS Slope Estimate: A slope in an OLS regression line.

Omitted Variable Bias: The bias that arises in the OLS estimators when a relevant variable is omitted from the regression.

Omitted Variables: One or more variables, which we would like to control for, have been omitted in estimating a regression model.

One-Sided Alternative: An alternative hypothesis which states that the parameter is greater than (or less than) the value hypothesised under the null.

One-Step-Ahead Forecast: A time series forecast one period into the future.

One-Tailed Test: A hypothesis test against a one sided alternative.

Ordinal Variable: A variable where the ordering of the values conveys information but the magnitude of the values does not.

Ordinary Least Squares (OLS): A method for estimating the parameters of a multiple linear regression model. The ordinary least squares estimates are obtained by minimising the sum of squared residuals.

Outliers: Observations in a data set that are substantially different from the bulk of the data, perhaps because of errors or because some data are generated by a different model than most of the other data.

Out-of-Sample Criteria: Criteria used for choosing forecasting models that are based on a part of the sample that was not used in obtaining parameter estimates.

Overall Significance of a Regression: A test of the joint significance of all explanatory variables appearing in a multiple regression equation.

Overspecifying a Model: See inclusion of an irrelevant variable.

P

p-value: The smallest significance level at which the null hypothesis can be rejected. Equivalently, the largest significance level at which the null hypothesis cannot be rejected.

Pairwise Uncorrelated Random Variables: A set of two or more random variables where each pair is uncorrelated.

Panel Data: A data set constructed from repeated cross sections over time. With a balanced panel, the same units appear in each time period. With an unbalanced panel, some units do not appear in each time period, often due to attrition.

Parameter: An unknown value that describes a population relationship.

Parsimonious Model: A model with as few parameters as possible for capturing any desired features.

Partial Effect: The effect of an explanatory variable on the dependent variable, holding other factors in the regression model fixed.

Percentage Change: The proportionate change in a variable, multiplied by 100.

Percentage Point Change: The change in a variable that is measured as a percent.

Perfect Collinearity: In multiple regression, one independent variable is an exact linear function of one or more other independent variables.

Plug-In Solution to the Omitted Variables Problem: A proxy variable is substituted for an unobserved omitted variable in an OLS regression.

Point Forecast: The forecasted value of a future outcome.

Policy Analysis: An empirical analysis that uses econometric methods to evaluate the effects of a certain policy.

Pooled Cross Section: A data configuration where independent cross sections, usually collected at different points in time, are combined to produce a single data set.

Population: A well-defined group (of people, firms, cities, and so on) that is the focus of a statistical or econometric analysis.

Population Model: A model, especially a multiple linear regression model, that describes a population.

Population R-Squared: In the population, the fraction of the variation in the dependent variable that is explained by the explanatory variables.

Population Regression Function: See conditional expectation.

Power of a Test: The probability of rejecting the null hypothesis when it is false; the power depends on the values of the population parameters under the alternative.

Practical Significance: The practical or economic importance of an estimate, which is measured by its sign and magnitude, as opposed to its statistical significance.

Predicted Variable: See dependent variable.

Prediction: The estimate of an outcome obtained by plugging specific values of the explanatory variables into an estimated model, usually a multiple regression model.

Prediction Error: The difference between the actual outcome and a prediction of that outcome.

Prediction Error Variance: The variance in the error that arises when predicting a future value of the dependent variable based on an estimated multiple regression equation.

Prediction Interval: A confidence interval for an unknown outcome on a dependent variable in a multiple regression model.

Predictor Variable: See explanatory variable.

Probability Density Function (pdf): A function that, for discrete random variables, gives the probability that the random variable takes on each value; for continuous random variables, the area under the pdf gives the probability of various events.

Probability Limit: The value to which an estimator converges as the sample size grows without bound.

Program Evaluation: An analysis of a particular private or public program using econometric methods to obtain the causal effect of the program.

Proportionate Change: The change in a variable relative to its initial value; mathematically, the change divided by the initial value.

Proxy Variable: An observed variable that is related but not identical to an unobserved explanatory variable in multiple regression analysis.

Q

Quadratic Functions: Functions that contain squares of one or more explanatory variables; they capture diminishing or increasing effects on the dependent variable.

Qualitative Variable: A variable describing a nonquantitative feature of an individual, a firm, a city, and so on.

R

R-Bar Squared: See adjusted R-squared.

R-Squared: In a multiple regression model, the proportion of the total sample variation in the dependent variable that is explained by the independent variable.

R-Squared Form of the F Statistic: The F statistic for testing exclusion restrictions expressed in terms of the R-squareds from the restricted and unrestricted models.

Random Sampling: A sampling scheme whereby each observation is drawn at random from the population. In particular, no unit is more likely to be selected than any other unit, and each draw is independent of all other draws.

Random Variable: A variable whose outcome is uncertain.

Random Walk: A time series process where next period's value is obtained as this period's value, plus an independent (or at least an uncorrelated) error term.

Random Walk with Drift: A random walk that has a constant (or drift) added in each period.

Real Variable: A monetary value measured in terms of a base period.

Regressand: See dependent variable.

Regression Through the Origin: Regression analysis where the intercept is set to zero; the slopes are obtained by minimising the sum of squared residuals, as usual.

Regressor: See explanatory variable.

Rejection Region: The set of values of a test statistic that leads to rejecting the null hypothesis.

Rejection Rule: In hypothesis testing, the rule that determines when the null hypothesis is rejected in favour of the alternative hypothesis.

Relative Change: See proportionate change.

Residual: The difference between the actual value and the fitted (or predicted) value; there is a residual for each observation in the sample used to obtain an OLS regression line.

Residual Analysis: A type of analysis that studies the sign and size of residuals for particular observations after a multiple regression model has been estimated.

Residual Sum of Squares (RSS): In multiple regression analysis, the sum of the squared OLS residuals across all observations.

Response Variable: See dependent variable.

Restricted Model: In hypothesis testing, the model obtained after imposing all of the restrictions required under the null.

Root Mean Squared Error (RMSE): Another name for the standard error of the regression in multiple regression analysis.

S

Sample Average: The sum of n numbers divided by n ; a measure of central tendency.

Sample Correlation: For outcomes on two random variables, the sample covariance divided by the product of the sample standard deviations.

Sample Covariance: An unbiased estimator of the population covariance between two random variables.

Sample Regression Function: See OLS regression line.

Sample Standard Deviation: A consistent estimator of the population standard deviation.

Sample Variance: An unbiased, consistent estimator of the population variance.

Sampling Distribution: The probability distribution of an estimator over all possible sample outcomes.

Sampling Variance: The variance in the sampling distribution of an estimator; it measures the spread in the sampling distribution.

Seasonal Dummy Variables: A set of dummy variables used to denote the quarters or months of the year.

Seasonality: A feature of monthly or quarterly time series where the average value differs systematically by season of the year.

Seasonally Adjusted: Monthly or quarterly time series data where some statistical procedure possibly regression on seasonal dummy variables-has been used to remove the seasonal component.

Semi-Elasticity: The percentage change in the dependent variable given a one-unit increase in an independent variable.

Sensitivity Analysis: The process of checking whether the estimated effects and statistical significance of key explanatory variables are sensitive to inclusion of other explanatory variables, functional form, dropping of potentially outlying observations, or different methods of estimation.

Serial Correlation: In a time series or panel data model, correlation between the errors in different time periods.

Serially Uncorrelated: The errors in a time series or panel data model are pairwise uncorrelated across time.

Short-Run Elasticity: The impact propensity in a distributed lag model when the dependent and independent variables are in logarithmic form.

Significance Level: The probability of Type I error in hypothesis testing.

Single (Simple) Linear Regression Model: A model where the dependent variable is a linear function of a single independent variable, plus an error term.

Simultaneous Equations Model (SEM): A model that jointly determines two or more endogenous variables, where each endogenous variable can be a function of other endogenous variables as well as of exogenous variables and an error term.

Slope Parameter: The coefficient on an independent variable in a multiple regression model.

Spreadsheet: Computer software used for entering and manipulating data.

Spurious Correlation: A correlation between two variables that is not due to causality, but perhaps to the dependence of the two variables on another unobserved factor.

Spurious Regression Problem: A problem that arises when regression analysis indicates a relationship between two or more unrelated time series processes simply because each has a trend, is an integrated time series (such as a random walk), or both.

Standard Deviation: A common measure of spread in the distribution of a random variable.

Standard Deviation of $\hat{\beta}_k$: A common measure of spread in the sampling distribution of β_k .

Standard Error of $\hat{\beta}_k$: An estimate of the standard deviation in the sampling distribution of β_k .

Standard Error of the Estimate: See standard error of the regression.

Standard Error of the Regression (SER): In multiple regression analysis, the estimate of the standard deviation of the population error, obtained as the square root of the sum of squared residuals over the degrees of freedom.

Standard Normal Distribution: The normal distribution with mean zero and variance one.

Standardised Random Variable: A random variable transformed by subtracting off its expected value and dividing the result by its standard deviation; the new random variable has mean zero and standard deviation one.

Static Model: A time series model where only contemporaneous explanatory variables affect the dependent variable.

Statistical Inference: The act of testing hypotheses about population parameters.

Statistically Different from Zero: See statistically significant.

Statistically Insignificant: Failure to reject the null hypothesis that a population parameter is equal to zero, at the chosen significance level.

Statistically Significant: Rejecting the null hypothesis that a parameter is equal to zero against the specified alternative, at the chosen significance level.

Stratified Sampling: A nonrandom sampling scheme whereby the population is first divided into several nonoverlapping, exhaustive strata, and then random samples are taken from within each stratum.

Strict Exogeneity: An assumption that holds in a time series or panel data model when the explanatory variables are strictly exogenous.

Strictly Exogenous: A feature of explanatory variables in a time series or panel data model where the error term at any time period has zero expectation, conditional on the explanatory variables in all time periods; a less restrictive version is stated in terms of zero correlations.

Sum of Squared Residuals: See residual sum of squares (RSS).

Summation Operator: A notation, denoted by Σ , used to define the summing of a set of numbers.

Symmetric Distribution: A probability distribution characterised by a probability density function that is symmetric around its median value, which must also be the mean value (whenever the mean exists).

T

t Distribution: The distribution of the ratio of a standard normal random variable and the square root of an independent chi-square random variable, where the chi-square random variable is first divided by its df.

t Ratio: See t statistic.

t Statistic: The statistic used to test a single hypothesis about the parameters in an econometric model.

Test Statistic: A rule used for testing hypotheses where each sample outcome produces a numerical value.

Text Editor: Computer software that can be used to edit text files.

Text(ASCII) File: A universal file format that can be transported across numerous computer platforms.

Time-Demeaned Data: Panel data where, for each cross-sectional unit, the average over time is subtracted from the data in each time period.

Time Series Data: Data collected over time on one or more variables.

Time Trend: A function of time that is the expected value of a trending time series process.

Total Sum of Squares (TSS): The total sample variation in a dependent variable about its sample average.

True Model: The actual population model relating the dependent variable to the relevant independent variables, plus a disturbance, where the zero conditional mean assumption holds.

Two-Sided Alternative: An alternative where the population parameter can be either less than or greater than the value stated under the null hypothesis.

Two-Tailed Test: A test against a two-sided alternative.

Type I Error: A rejection of the null hypothesis when it is true.

Type II Error: The failure to reject the null hypothesis when it is false.

U

Unbiased Estimator: An estimator whose expected value (or mean of its sampling distribution) equals the population value (regardless of the population value).

Unconditional Forecast: A forecast that does not rely on knowing, or assuming values for, future explanatory variables.

Uncorrelated Random Variables: Random variables that are not linearly related.

Underspecifying a Model: See excluding a relevant variable.

Unrestricted Model: In hypothesis testing, the model that has no restrictions placed on its parameters.

Upward Bias: The expected value of an estimator is greater than the population parameter value.

V

Variance: A measure of spread in the distribution of a random variable.

Variance of the Prediction Error: See prediction error variance.

W

Weighted Least Squares (WLS) Estimator: An estimator used to adjust for a known form of heteroskedasticity, where each squared residual is weighted by the inverse of the (estimated) variance of the error.

Y

Year Dummy Variables: For data sets with a time series component, dummy (binary) variables equal to one in the relevant year and zero in all other years.

Z

Zero Conditional Mean Assumption: A key assumption used in multiple regression analysis which states that, given any values of the explanatory variables, the expected value of the error equals zero.