

ECONOMETRICS TOPICS

Chapter 1: An Overview of Regression Analysis

What is econometrics?

Econometrics: economic measurement

Uses of econometrics:

1. Describing economic reality
2. Testing hypothesis about economic theory
3. Forecasting future economic activity

Alternative economic approaches

Steps necessary for any kind of quantitative research:

1. Specifying the models or relationships to be studied
2. Collecting the data needed to quantify the models
3. Quantifying the models with the data

Single-equation linear regression analysis is one particular economic approach that is the focus of this book.

What is regression analysis?

Dependent variables, independent variables, and causality

Regression analysis: a statistical technique that attempts to explain movements in one variable, the dependent variable, as a function of movements in a set of other variables, called the independent (or explanatory) variables, through the quantification of a single equation.

A regression result, no matter how statistically significant, cannot prove causality. All regression analysis can do is test whether a significant quantitative relationship exists.

Single-equation linear models

Betas: the coefficients that determine the coordinates of the straight line at any point.

Beta-null: the constant or intercept term; it indicates the value of Y when X equals zero.

Beta-one: the slope coefficient; it indicates the amount that Y will change when X increases by one unit.

An equation is linear in the variables if plotting the function in terms of X and Y generates a straight line.

An equation is linear in the coefficients only if the coefficients appear in the simplest form – they are not raised to any powers are not multiplied or divided by other coefficients, and do not themselves include some sort of function.

The stochastic error term

Stochastic error term: a term that is added to a regression equation to introduce all of the variation in Y that cannot be explained by the included X's.

The deterministic component: $B_0 + B_1X$; can be thought of as the expected value of Y given X , the mean value of the Y s associated with a particular value of X .

The stochastic error term must be present in a regression equation because there are at least four sources of variation in Y other than the variation in the included X s:

1. Many minor influences on Y are omitted from the equation (for example, because data are unavailable).
2. It is virtually impossible to avoid some sort of measurement error in at least one of the equation's variables.
3. The underlying theoretical equation might have a different functional form than the one chosen for the regression. For example, the underlying equation might be nonlinear in the variables for a linear regression.
4. All attempts to generalize human behavior must contain at least some amount of unpredictable or purely random variation.

Extending the notation

The meaning of the regression coefficient beta-one: the impact of a one unit increase in X -one on the dependent variable Y , holding constant the other included independent variables.

Multivariate regression coefficients: serve to isolate the impact on Y of a change in one variable from the impact on y of the changes in the other variables.

The estimated regression equation

Estimated regression equation: a quantified version of the theoretical regression equation

Estimated regression coefficients: empirical best guesses of the true regression coefficients and are obtained from a sample of the X s and Y s; denoted by beta-hats

A simple example of regression analysis

Using regression to explain housing prices

Chapter 2: Ordinary Least Squares

Estimating single-independent-variable models with OLS

Ordinary least squares (OLS): a regression estimation technique that calculates the beta-hats so as to minimize the sum of the squared residuals.

Why use ordinary least squares?

1. OLS is relatively easy to use.
2. The goal of minimizing the sum of the squared residuals is quite appropriate from a theoretical point of view.
3. OLS estimates have a number of useful characteristics:

- a. The estimated regression line goes through the means of Y and X. That is, if you substitute \bar{Y} and \bar{X} into the equation it holds exactly.
- b. The sum of the residuals is exactly zero.
- c. OLS can be shown to be the best estimator possible under a set of fairly restrictive assumptions.

Estimator: a mathematical technique that is applied to a sample of data to produce real-world numerical estimates of the true population regression coefficients (or other parameters). OLS is an estimator.

How does OLS work?

Regression model equation

Estimate equations of beta-one and beta-null

Total, unexplained, and residual sum of squares

Total sum of squares (TSS): the squared variations of Y around its mean as a measure of the amount of variation to be explained by the regression.

Explained sum of squares (ESS): measures the amount of the squared deviation of Y_i from its mean that is explained by the regression line.

Residual sum of squares (RSS): the unexplained portion of the total sum of squares.

OLS minimizes the RSS and therefore maximizes the ESS.

An illustration of OLS estimation

Estimating multivariate regression models with OLS

The meaning of multivariate regression coefficients

Multivariate regression coefficient: indicates the change in the dependent variable associated with a one-unit increase in the independent variable in question holding constant the other independent variables in the equation.

OLS estimate of multivariate regression models

Estimate equations for multivariate regression coefficients

An example of a multivariate regression model

Evaluating the quality of a regression equation

1. Is the equation supported by sound theory?
2. How well does the estimated regression as a whole fit the data?
3. Is the data set reasonably large and accurate?
4. Is OLS the best estimator to be used for this equation?
5. How well do the estimated coefficients correspond to the expectations developed by the researcher before the data were collected?
6. Are all the obviously important variables included in the equation?
7. Has the most theoretically logical functional form been used?
8. Does the regression appear to be free of major econometric problems?

Describing the overall fit of the estimated model

R-squared, the coefficient of determination

Coefficient of determination: the ratio of the ESS to the TSS

Equation for R-squared

R-bar-squared, the adjusted R-squared

Degrees of freedom: the excess of the number of observations over the number of coefficients estimated

R-bar-squared: R-squared adjusted for degrees of freedom

Equation for R-bar-squared

An example of the misuse of the adjusted R^2 , the R-bar-squared

Do not use R-bar-squared as the sole measure of the quality of an equation at the expense of economic theory or statistical significance.

Chapter 3: Learning to Use Regression Analysis

Steps in applied regression analysis

1. Review the literature and develop the theoretical model.
2. Specify the model: select the independent variables and the functional form.

The following components should be specified:

- a. The independent variables and how they should be measured.
- b. The functional (mathematical) form of the variables.
- c. The type of stochastic error term.

A mistake in any of the three elements results in a specification error.

Dummy variable: takes on the values of one or zero depending on whether a specified condition holds.

3. Hypothesize the expected signs of the coefficients.
4. Collect the data.
5. Estimate and evaluate the equation.
 - a. Are there errors in the variables?
6. Document the results.
 - a. Recognize the difference between the number of digits computed and the number of significant figures.

Using regression analysis to pick restaurant locations

Chapter 4: The Classical Model

The classical assumptions

The classical assumptions must be met in order for OLS estimators to be the best available.

The Classical Assumptions:

1. The regression model is linear in the coefficients, is correctly specified, and has an additive error term.
2. The error term has a zero population mean.
3. All explanatory variables are uncorrelated with the error term.
4. Observations of the error term are uncorrelated with each other (no serial correlation or autocorrelation).
5. The error term has a constant variance (no heteroskedasticity).

6. No explanatory variable is a perfect linear function of any other explanatory variable(s) (no perfect multicollinearity).
 - a. Perfect collinearity between two independent variables implies that they are really the same variable, or that one is a multiple of the other, and/or that a constant has been added to one of the variables.
 - b. Multicollinearity occurs when more than two independent variables are involved.
7. The error term is normally distributed (this assumption is optional but usually is invoked).
 - a. Its major use is in hypothesis testing, which uses the estimated regression statistics to accept or reject hypothesis about economic behavior.

An error term satisfying assumptions 1-5 is called a classical error term and if assumption 7 is added, the error term is called a classical normal error term.

The normal distribution of the error term

1. The error term can be thought of as the composite of a number of minor influences or errors. As the number of these minor influences gets larger, the distribution of the error term tends to approach the normal distribution. This tendency is called the Central Limit Theorem.
2. The t-statistic and the F-statistic are not truly applicable unless the error term is normally distributed.

The normal distribution

Standard normal distribution: a normal distribution with a mean equal to zero and a variance equal to one: $N(0,1)$

The Central Limit Theorem and the normality of the error term

The Central Limit Theorem: the mean (or sum) of a number of independent, identically distributed random variables will tend to be normally distributed, regardless of their distribution, if the number of different random variables is large enough.

The sampling distribution of beta-hat

The probability distribution of the beta-hats is called a sampling distribution because it is based on a number of sample drawings of the error term.

Sampling distributions of estimators

Properties of the mean

A desirable property of a distribution of estimates is that its mean equals the true mean of the variable being estimated. An estimator that yields such estimates is called an unbiased estimator. An unbiased estimator is an estimator whose sampling distribution has as its expected value the true value of beta.

If an estimator produces beta-hats that are not centered around the true beta, the estimator is referred to as a biased estimator.

Properties of variance

The variance that is the most narrow is the most desirable. The variance can be decreased by increasing the size of the sample.

Properties of the standard error

The standard error of the estimated coefficient is the square root of the estimated variance of the beta-hats and it is affected by the size of the sample and other factors.

Equation for the standard error of an estimated coefficient

Impact of changes in the terms of the standard error equation

A demonstration that the beta-hats are normally distributed

1. The distribution of beta-hats appears to be a symmetrical, bell-shaped distribution that is approaching a continuous normal distribution as the number of samples of beta-hats increases.
2. The distribution of the beta-hats is unbiased but shows surprising variations.

Monte Carlo experiments

1. Assume a “true” model with specific coefficient values and an error term distribution.
2. Select values for the independent variables.
3. Select an estimating technique (usually OLS).
4. Create various samples of values of the dependent variable, using the assumed model, by randomly generating error terms from the assumed distribution.
5. Compute the estimates of the betas from the various samples using the estimating technique.
6. Evaluate the results.
7. Return to step 1 and choose other values for the coefficients, independent variables, or error term variance; compare these results with the first set. (This step is optional and is called sensitivity analysis.)

The Gauss-Markov theorem and the properties of OLS estimators

The Gauss-Markov Theorem states that given Classical Assumptions 1-6 (assumption 7, normality, is not needed for this theorem), the ordinary least squares estimator of β_k is the minimum variance estimator from among the set of all linear unbiased estimators of β_k for $k = 0, 1, 2, \dots, k$.

An easy way to remember this is by stating that OLS is BLUE where BLUE stands for best (meaning minimum variance), linear, unbiased, estimator.

Efficient: an unbiased estimator with the smallest variance; that estimator is said to have the property of efficiency.

Given all seven classical assumptions, the OLS coefficient estimators can be shown to have the following properties:

1. They are unbiased.
2. They are minimum variance.
3. They are consistent.
4. They are normally distributed.

Standard econometric notation

Chapter 5: Hypothesis Testing

Bayesian statistics: an alternative to hypothesis testing; adds prior information to the sample to draw statistical inferences.

What is hypothesis testing?

Three topics central to the application of hypothesis testing to regression analysis:

1. The specification of the hypothesis to be tested.
2. The decision rule to use in deciding whether to reject the hypothesis in question.
3. The kinds of errors that might be encountered if the application of the decision rule to the appropriate statistics yields an incorrect inference.

Classical null and alternative hypothesis

Null hypothesis: a statement of the range of values of the regression coefficient that would be expected to occur if the researcher's theory were not correct.

Alternative hypothesis: used to specify the range of values of the coefficient that would be expected to occur if the researcher's theory were correct.

Two-sided test (two-tailed test): when the alternative hypothesis has values on both sides of the null hypothesis

One-sided test: the alternative hypothesis is only on one side of the null hypothesis

Type I and type II errors

Type I: rejecting a true null hypothesis

Type II: not rejecting a false null hypothesis

Decision rules of hypothesis testing

Decision rule: the testing of a hypothesis by comparing the magnitude of the sample statistic with a preselected critical value

Critical value: a value that divides the acceptance region from the rejection region when testing a null hypothesis.

The t-test

The t-statistic

Equation for the t-statistic

The critical t-value and the t-test decision rule

Choosing a level of significance

The level of significance indicates the probability of observing an estimated t-value greater than the critical t-value if the null hypothesis were correct. It measures the amount of Type I error implied by a particular critical t-value.

Confidence Intervals

Confidence interval: a range within which the true value of an item is likely to fall a specified percentage of the time; this percentage is the level

of confidence associated with the level of significance used to choose the critical t-value in the interval

Examples of the t-test

Examples of one-sided t-tests

Four steps to use when working with t-tests:

1. Set up the null and alternative hypotheses.
2. Choose a level of significance and therefore a critical t-value.
3. Run the regression and obtain an estimated t-value (or t-score).
4. Apply the decision rule by comparing the calculated t-value with the critical t-value in order to reject or “accept” the null hypothesis.

Examples of two-sided t-tests

The kinds of circumstances that call for a two-sided test fall into two categories:

1. Two-sided tests of whether an estimated coefficient is significantly different than zero.
2. Two-sided tests of whether an estimated coefficient is significantly different from a specific nonzero value.

The t-test of the simple correlation coefficient, r

The simple correlation coefficient, r : a measure of the strength and direction of the linear relationship between two variables.

Equation for the simple correlation coefficient

Equation for the t-statistic using r

Limitations of the t-test

The t-test does not test theoretical validity

The t-test does not test “importance”

The t-test is not intended for tests of the entire population, as the standard error will approach zero as the sample size approaches infinity so the t-score will eventually be equal to infinity.

The F-test of overall significance

The F-test is a method of testing a null hypothesis that includes more than one coefficient; it works by determining whether the overall fit of an equation is significantly reduced by constraining the equation to conform to the null hypothesis.

Equation for the F-test

Chapter 6: Specification – Choosing the Independent Variables

A specification error results when choices, such as choosing the correct independent variables, the correct functional form, and the correct form of the stochastic error term, are made incorrectly.

Omitted variables

Omitted variable: an important explanatory variable that has been left out of a regression equation.

Omitted variable bias: the bias caused by leaving a variable out of an equation (or, more generally, specification bias)

The consequences of an omitted variable

The expected value of the coefficients does not equal the true value

Equation for the amount of bias for the slope coefficients

Bias exists unless:

1. The true coefficient equals zero.
2. The included and omitted variables are uncorrelated.

An example of specification bias

Correcting for an omitted variable

Expected bias: the likely bias that omitting a particular variable would have caused in the estimated coefficient of one of the included variables.

Uses of the expected bias equation

Irrelevant variables

Irrelevant variables: the converse of omitted variables; a variable that is included in an equation that doesn't belong there

Impact of irrelevant variables

Impact on the error term

Impact on the variance of the coefficients

An example of an irrelevant variable

Four important specification criteria

1. Theory: Is the variable's place in the equation unambiguous and theoretically sound?
2. t-test: Is the variable's estimated coefficient significant in the expected direction?
3. R-bar-squared: Does the overall fit of the equation (adjusted for degrees of freedom) improve when the variable is added to the equation?
4. Bias: Do other variables' coefficients change significantly when the variable is added to the equation?

An illustration of the misuse of specification criteria

Specification searches

Data mining

Stepwise regression procedures

Stepwise regression: involves the use of a computer program to choose the independent variables to be included in the estimation of a particular equation.

Sequential specification searches

Sequential specification search technique: allows a researcher to estimate an undisclosed number of regressions and then present a final choice (which is based upon an unspecified set of expectations about the signs

and significance of the coefficients) as if it were the only specification estimated.

Such a method misstates the statistical validity of the regression results for two reasons:

1. The statistical significance of the results is overestimated because the estimations of the previous regressions are ignored.
2. The set of expectations used by the researcher to choose between various regression results is rarely if ever disclosed. This the reader has no way of knowing whether or not all the other regression results had opposite signs or insignificant coefficients for the important variables.

Bias caused by relying on the t-test to choose variables

Scanning and sensitivity analysis

Scanning: involves analyzing a data set not for the purpose of testing a hypothesis but for the purpose of developing a testable theory or hypothesis

Sensitivity analysis: consists of purposely running a number of alternative specifications to determine whether particular results are robust (not statistical flukes)

Lagged independent variables

Lag: the length of time between cause and effect

Lagged independent variables

Distributed lags

An example of choosing independent variables

Appendix: additional specification criteria

Chapter 7: Specification – Choosing a Functional Form

The use and interpretation of the constant term

Do not suppress the constant term

Do not rely on estimates of the constant term

Alternative functional forms

Linear form

Elasticity of Y with respect to X: the percentage change in the dependent variable caused by a one percent increase in the independent variable, holding the other variables in the equation constant

In the linear form the slope is constant so the elasticity is not constant

Double-log form

In a double-log functional form, the natural log of Y is the dependent variable and the natural log of X is the independent variable.

Exponential functional form: the proper form given the assumption of constant elasticity

In a double-log equation, an individual regression coefficient can be interpreted as an elasticity.

Semilog form

Semilog functional form: a variant of the double-log equation in which some but not all of the variables (dependent and independent) are expressed in terms of their natural logs.

Polynomial form

Polynomial functional form: expresses Y as a function of independent variables, some of which are raised to powers other than one.

Inverse form

Inverse functional form: expresses Y as a function of the reciprocal (or inverse) of one or more of the independent variables

Problems with incorrect functional forms

R-bar-squareds are difficult to compare when the dependent variable is transformed

Incorrect functional forms outside the range of the sample

An incorrect functional form may provide a reasonable fit within the sample but have the potential to make large forecast errors when used outside the range of the sample.

Using dummy variables

Intercept dummy: a dummy variable that changes the constant or intercept term, depending on whether the qualitative condition is met

The omitted condition: the event not explicitly represented by a dummy variable; forms the basis against which the included conditions are compared

Slope dummy variables

Interaction term: an independent variable in a regression equation that is the multiple of two or more other independent variables.

Slope dummy variable: allows the slope of the relationship between the dependent and variable and an independent variable to be different depending on whether the condition specified by a dummy variable is met.

Appendix: more uses for the F-test

Chapter 8: Multicollinearity

For violations of the classical assumptions, the following questions are addressed:

1. What is the nature of the problem?
2. What are the consequences of the problem?
3. How is the problem diagnosed?
4. What remedies for the problem are available?

Perfect versus imperfect multicollinearity

Perfect multicollinearity

Perfect multicollinearity: the violation of the assumption that no independent variable is a perfect linear function of one or more other independent variables (classical assumption 6).

Equation for perfectly linearly related variables

Imperfect multicollinearity

Imperfect multicollinearity: a linear functional relationship between two or more independent variables that is so strong that it can significantly effect the estimation of the coefficients of the variables.

Equation for imperfectly linearly related variables

The consequences of multicollinearity

What are the consequences of multicollinearity?

1. Estimates will remain unbiased.
2. The variances and standard errors of the estimates will increase.
3. The computed t-scores will fall.
4. Estimates will become very sensitive to changes in specification.
5. The overall fit of the equation and the estimation of nonmulticollinear variables will be largely unaffected.

Two examples of the consequences of multicollinearity

The detection of multicollinearity

High simple correlation coefficients

High variance inflation factors (VIFs)

Variance inflation factor (VIF): a method of detecting the severity of multicollinearity by looking at the extent to which a given explanatory variable can be explained by all the other explanatory variables in the equation; an estimate of how much multicollinearity has increased the variance of the estimated coefficient.

Calculating the VIF for a given X_i involves three steps:

1. Run an OLS regression that has X_i as a function of all the other explanatory variables in the equation.
2. Calculate the variance inflation factor for $\beta_{\hat{i}}$
 - a. Equation for the VIF of $\beta_{\hat{i}}$
3. Analyze the degree of multicollinearity by evaluating the size of the VIF of $\beta_{\hat{i}}$.
 - a. A VIF greater than 5 is most commonly used to determine that multicollinearity is severe, but there no table of formal critical VIF values.

Remedies for multicollinearity

Do nothing

Drop a redundant variable

Redundant variable: only one variable is needed to represent the effect on the dependent variable that all of the multicollinear variables currently represent.

Transform the multicollinear variables

Two most common transformations:

1. Form a combination of the multicollinear variables.
 - a. The technique of forming a combination of two or more of the multicollinear variables consists of creating a new variable that is a function of the multicollinear variables and using the new variable to replace the old ones in the regression equation.
 - b. General equation for a combination transformation
2. Transform the equation into first differences.
 - a. A first difference is nothing more than the change in a variable from the previous time period to the current time period.
 - b. General equation for a first difference transformation

Increase the size of the sample

Choosing the proper remedy

Why multicollinearity often should be left unadjusted

A more complete example of dealing with multicollinearity

Appendix: the SAT interactive regression learning exercise

Chapter 16: Statistical Principles

Describing data

Median

Median: the middle value when the data are arranged in numerical order from the smallest value to the largest value

Mean

Mean: the simple arithmetic average value of the data

Outlier: a value very different from the other observations

Histogram: a diagram in which the relative frequency of the observations in each interval is shown by the height of the bar spanning the interval (for equal intervals)

Variance and standard deviation

Variance: the average squared deviation of the observations about their mean

Standard deviation: the square root of variance

Probability distributions

Probability

A random variable X: a variable whose numerical value is determined by chance, the outcome of a random phenomenon

Discrete random variable: has a countable number of possible values

Continuous random variables: can take on any value in an interval, such as time and distance

A probability distribution for a discrete random variable X assigns probabilities to the possible values of X

Mean, variance, and standard deviation

Mean: the expected value of a discrete random variable X

Equation for the expected value of the random variable X

Finding the expected value of X :

1. Determine the possible outcomes (the possible values of X).
2. Determine the probability of each possible outcome.
3. The expected value is equal to the sum of the possible outcomes multiplied by their respective probabilities.

Equations for the variance and standard deviation of a discrete random variable X

Finding the variance and standard deviation of a discrete random variable X :

1. Determine the expected value of X .
2. For each possible value of X , determine the size of the squared deviation from the expected value of the true population mean.
3. The variance is equal to the sum of the squared deviations of X_i from the true population mean, multiplied by their respective probabilities.
4. The standard deviation is the square root of the variance.

Continuous random variables

Continuous probability density curve: the probability that the outcome is in a specified interval is given by the corresponding area under the curve

Standardized variables

Equation for the z -score

No matter what the initial units of X , the standardized random variable Z has a mean of zero and a standard deviation of one.

The normal distribution

The central limit theorem: if Z is a standardized sum of n independent, identically distributed (discrete or continuous) random variables with a finite, nonzero standard deviation, then the probability distribution of Z approaches the normal distribution as n increases.

Sampling

Selection bias

Selection bias: occurs when the selection of the sample systematically excludes or underrepresents certain groups.

Convenience sample: consists of data that are easily collected

Survivor bias

Retrospective study: looking at past data for a contemporaneously selected sample

Prospective study: selects a sample and then tracks the members over time

Survivor bias: when a sample is chosen from a current population in order to draw inferences about a past population, it necessarily excludes

members of the past population who are no longer around; only looking at survivors

Nonresponse bias

Nonresponse bias: the systematic refusal of some groups to participate in an experiment or to respond to a poll

The power of random selection

Estimation

Parameter: the unknown true values that describe the population

Estimator: a sample statistic that is used to estimate the value of a population parameter

Estimate: the specific value of the estimator that is obtained in a particular sample

Sampling distributions

Sampling error: the difference between the value of one particular sample mean and the average of the means of all possible samples of this size

Systematic error or bias: cause the sample means to differ, on average, from the population parameter being estimated

Sampling distribution of a statistic: the probability distribution or density curve that describes the population of all possible values of this statistic

The mean of the sampling distribution

Unbiased estimator: a sampling statistic is an unbiased estimator of a population parameter if the mean of the sampling distribution of this statistic is equal to the value of the population parameter.

The standard deviation of the sampling distribution

The t-distribution

Equation for the standard error

Equation for the z-statistic with n number of observations and a known standard deviation

Equation for the t-statistic

Equation for the degrees of freedom

Confidence intervals

Equation for confidence intervals

Confidence level

General procedure for determining a confidence interval for a population mean:

1. Calculate the sample mean.
2. Calculate the standard error of the sample mean by dividing the sample standard deviation s by the square root of the sample size n .
3. Select a confidence level and look up the t-value t^* that corresponds to its probability.
4. A confidence interval for the population mean is equal to the sample mean plus or minus t^* standard errors of the sample mean.

Sampling from finite populations

Hypothesis tests

A general framework for hypothesis tests

The null and alternative hypotheses

Null hypothesis: used for a proof by statistical contradiction

The test statistic and statistical significance

Test statistic: the estimator that is used to test the null hypothesis

Statistically significant: the probability that the value of the test statistic would be so far from its expected value were the null hypothesis true is less than this specified significance level

P-values

P-value: for a test of a null hypothesis about the population mean is the probability, if the null hypothesis is in fact true, that a random sample of this size would yield a sample mean that is this far (or further) from the value of the population mean assumed by the null hypothesis; a small P-value casts doubt on the null hypothesis

Using confidence intervals

Is it important?

An overview of hypothesis testing

The general procedure can be summarized as follows:

1. Specify the null hypothesis and whether the alternative hypothesis is one sided or two sided.
2. Use the sample data to estimate the value of the population parameter whose value is specified by the null hypothesis.
3. If this sample statistic is approximately normally distributed, calculate the t-value, which measure how many standard errors the estimate is from the null hypothesis. For testing a null hypothesis about the value of the population mean, the sample mean is used as the test statistic and the t-value is calculated.
4. Determine the critical values of t corresponding to the test's selected significance level, and see if the t-value for this sample is outside this range.
5. Report a confidence interval in order to assess the practical importance of the results.