Interpreting Regression Results

Carlo Favero

Interpreting Regression Results

Interpreting regression results is not a simple exercise. We propose to split these procedure in three steps.

- First, introduce a measure of sampling variability and evaluate again what you know taking into account that parameters are estimated and there is uncertainty surrounding your point estimates.
- Second, understand the relevance of our regression independently from inference on the parameters. There is an easy way to do this: suppose all parameters in the model are known and identical to the estimated values and learn how to read these.
- Third, remember that each regression is run after a reduction process has been, explicitly or implicitly implemented. The relevant question is what happens if something went wrong in the reduction process? What are the consequences of omitting relevant information or of including irrelevant ones in your specification?

Statistical Significance and Relevance

- Relevance of a regression is different form statistical significance of the estimated parameters.
- In fact, confusing statistical significance of the estimated parameter describing the effect of a regressor on the dependent variable with practical relevance of that effect is a rather common mistake in the use of the linear model.
- Statistical inference is a tool for estimating parameters in a probability model and assessing the amount of sampling variability. Statistics gives us indication on what we can say about the values of the parameters in the model on the basis of our sample.
- The relevance of a regression is determined by the share of the unconditional variance of $\mathbf{y}$ that is explained by the variance of $E(\mathbf{y} \mid \mathbf{X})$. Measuring how large is the share of the unconditional variance of $\mathbf{y}$ explained by the regression function is the fundamental role of R^2.

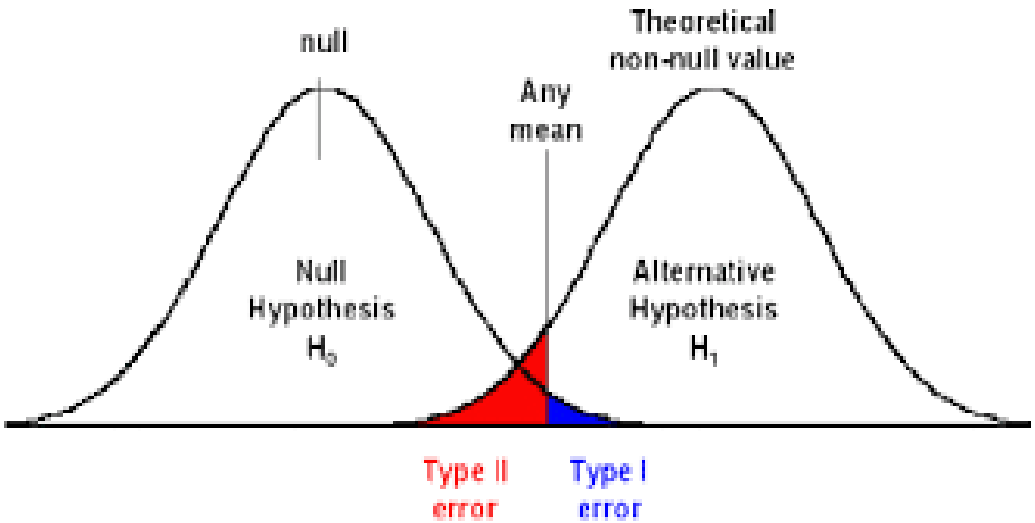
Statistical Significance of regression coefficients

- Estimate the coefficients in a regression, specify a null hypothesis of interest (for example, the coefficient on a given factor is zero).
- Derive a statistic (i.e. a quantity function of the regression coefficients) whose distribution is known under the null hypothesis, compute the observed value of the statistics
- Compute p as be the probability (under the null) of getting the value you have observed for the statistics
- p is called the p-value. Adopt a decision rule about p, call it p* and reject the null if the observed value of your statistic is smaller than p*. For example, if you take p=0.05 you reject the null everytime your observed statistics is smaller than 0.05. In this case you make the call that the observation of an event that has very low probability under the null is an indication that the null is rejected.

Statistical Significance of regression coefficients

- Of course by using the criterion adopted you run the risk of rejecting an hypothesis when that hypothesis is true. This is called the Probability of Type I error or the size of your test.
- There is another risk that you run: the probability of type II error, that is the probability of not rejecting a null when it is false. Think about an alternative hypothesis on the coefficients, you can compute the probability with which your statistics will be smaller than the cutoff point to which you associate a probability p*. That is the probability of type II error. The power of the test is 1- Pr (type II error).
- Note that the p-value can be computed in two ways i) by deriving the relevant distribution under the null ii) by simulating via Monte-Carlo or bootstrap the relevant distribution under the null. Using simulation makes easy to calculate the power of your test against given alternatives.

Statistical Significance of regression coefficients



Inference in the Linear Regression Model

- Inference in the Linear Regression Model is about design the appropriate statistics to test the hypothesis of interest on the coefficients in a linear model. We shall address this process in two steps
  - how to formalize the relevant hypothesis
  - how to build the statistics.

Intuitive explanation

What we are interested into is to test hypothesis on the vector $\boldsymbol{\beta}$, the idea is to construct on the basis of $\hat{\boldsymbol{\beta}}$:

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\epsilon}$$

therefore, conditional upon $\mathbf{X}$, the distribution of $\hat{\boldsymbol{\beta}}$ around $\boldsymbol{\beta}$ is fully determined by the distribution of $\boldsymbol{\epsilon}$ conditional upon $\mathbf{X}$.
So if

$$\epsilon \mid \mathbf{X} \sim \mathbf{N}\left(\mathbf{0}, \sigma^2 I\right).$$

or, equivalently

$$y \mid \mathbf{X} \sim \mathbf{N}\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2 I\right),$$

Intuitive explanation

the distribution of $\left(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right)$ which, being a linear combination of a normal distribution, is also normal:

$$\left(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right) \sim \mathbf{N}\left(\boldsymbol{\beta}, \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right). \tag{1}$$

Now specifying a null hypothesis implies to impose some restrictions on $\boldsymbol{\beta}$.Once you have imposed restrictions on $\boldsymbol{\beta}$ you would know the distribution of $\left(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right)$ if the variance was known.

Intuitive explanation

- But here you have two problems: the variance is unknown and sometimes you want to specify null hypothesis only for some elements of β, leaving the others free. (think of a situation in which you want to test that the coefficient on one of the four factors is equal to zero, leaving the other three unrestricted) .
- Fortunately we have a solution: we have an estimate $\hat{\sigma}^2$ for σ^2 and we can derive a statistics for the general null hypothesis of interest based on $\widehat{\beta}, \hat{\sigma}^2 \; and \; \mathbf{X}$.Fortunately we have a solution: we have an estimate $\hat{\sigma}^2$ for σ^2 and we can derive a statistics for the general null hypothesis of interest based on $\widehat{\beta}, \hat{\sigma}^2 \; and \; \mathbf{X}$.

How to formalize the relevant hypothesis

Given our linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

Our general case of interest is that of r restrictions on the vector of parameters with $r < k$. If we limit our interest to the class of linear restrictions on coefficients, we can express them as

$$H_0 = \mathbf{R}\boldsymbol{\beta} = \mathbf{r},$$

where $\mathbf{R}$ is an $(r \times k)$ matrix of parameters with rank k and $\mathbf{r}$ is an $(r \times 1)$ vector of parameters.

How to formalize the relevant hypothesis

To illustrate how **R** and **r** are constructed, we consider the baseline case of the four factor model:

$$\begin{aligned} W_{it} &= \beta_0 + \beta_1 F1_{it} + \beta_2 F2_{it} + \beta_3 F3_{it} + \beta_4 F4_{it} + \epsilon_{it} \\ v_{it} &\sim N.I.D\left(0, \sigma^2\right) \end{aligned}$$

The restrictions $\beta_1 = 0, \beta_2 = \beta_3,$ can be represented as follows:

$$\mathbf{R}\boldsymbol{\beta} = \mathbf{r},$$

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

The Statistics

If $\left(\widehat{\boldsymbol{\beta}} \mid \mathbf{X}\right) \sim \mathbf{N}\left(\boldsymbol{\beta}, \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)$, then:

$$\left(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r} \mid \mathbf{X}\right) \sim \mathbf{N}\left(\mathbf{R}\boldsymbol{\beta} - \mathbf{r}, \sigma^2 \mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\right). \tag{2}$$

Unfortunately, using the normal distribution would require the knowledge of σ^2, which in general is not known. Fortunately, a statistics can be built based on the OLS estimate for σ^2.
Fortunately, a statistics can be built based on the OLS estimate for σ^2.
In fact, it can be shown that

$$\frac{\left(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r}\right)' \left(\mathbf{R}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{R}'\right)^{-1} \left(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r}\right)}{s^2} \sim \mathbf{rF}\left(r, T-k\right), \qquad \text{under } H_0,$$

that can be used to test the relevant hypothesis.

The Statistics

Notice that, as we know that in the case r=1, $t_{t-k} = \sqrt{F(1, T-k)}$,if we are interested in testing hypothesis on a single coefficients (say β_1) we can use the following statistic:

$$\frac{\hat{\beta}_1 - \beta_1}{\left(Var\left(\hat{\beta}_1\right)\right)^{1/2}} \sim t(T-k) \text{ under } H_0.$$

The Relevance of a Regression

- The relevance of a regression is determined by the share of the unconditional variance of $\mathbf{y}$ that is explained by the variance of $E(\mathbf{y} \mid \mathbf{X})$. Measuring how large is the share of the unconditional variance of $\mathbf{y}$ explained by the regression function is the fundamental role of R^2.
- a variable can be very significant in explaining the variance of $E(\mathbf{y} \mid \mathbf{X})$, but little of the unconditional variance of $\mathbf{y}$ can be explained by the variance of $E(\mathbf{y} \mid \mathbf{X})$.
- statistical significance does not imply relevance.

Relevance of regression coefficients

- Estimate the coefficients in a regression and keep them fixed at their point estimate
- Run an experiment by changing the conditional mean of the dependent variable via a shock to the regressors
- Assess how relevant is the shock to the regressor(s) (say, one of the four factor) to determine the dependent variables (say,team performance)

How do I interpret a regression coefficient?

The Frisch-Waugh Theorem tells us than any given regression coefficient in the model $E(y \mid \mathbf{X}) = \mathbf{X}\boldsymbol{\beta}$ can be computed in two different but exactly equivalent ways:
1) by regressing y on all the columns of $\mathbf{X}$,
2) by first regressing the j-th column of $\mathbf{X}$ on all the other columns of $\mathbf{X}$, computing the residuals of this regression and then by regressing y on these residuals.
This result is relevant in that it clarifies that the relationships pinned down by the estimated parameters in a linear model do not describe the connections between the regressand and each regressor but the connection between the part of each regressor that is not explained by the other ones and the regressand.

What if analysis

- The relevant question in this case becomes "how much shall y change if I change $\mathbf{X}_i$?"
- The estimation of a single equation linear model does not allow to answer that question, for a number of reasons.
- First, estimated parameters in a linear model can only answer the question how much shall $E(y \mid \mathbf{X})$ if I change $\mathbf{X}$? We have seen that the two questions are very different if the R^2 of the regression is low, in this case a change in $E(y \mid \mathbf{X})$ may not effect any visible and relevant effect on y.
- Second, a regression model is a conditional expected value GIVEN $\mathbf{X}$. In this sense there is no space for "changing" the value of any element in $\mathbf{X}$.

What if analysis

- Any statement involving such a change requires some assumption on how the conditional expectation of y changes if $\mathbf{X}$ changes and a correct analysis of this requires an assumption on the joint distribution of y and $\mathbf{X}$.
- Simulation might require the use of the multivariate joint model even when valid estimation can be performed concentrating only on the conditional model.

What if analysis

Think of a linear model with known parameters

$$y = \beta_1 x_1 + \beta_2 x_2$$

What is in this model the effect of on y of changing x_1 by one unit while keeping x_2 constant ? Easy β_1.
Now think of the estimated linear model:

$$y = \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{u}$$

Now y is different from $E(y \mid \mathbf{X})$ and the question "what is in this model the effect of on $E(y \mid \mathbf{X})$ of changing x_1 by one unit while keeping x_2 constant ?" does not in general make sense.

What if analysis

- Changing x_1 keeping x_2 unaltered implies that there is zero correlation among this variables.
- But the estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ are obtained by using data in which in general there is some correlation between x_1 and x_2.
- Data in which fluctuations in x_1 do not have any effect on x_2 would have most likely generated different estimates from those obtained in the estimation sample.
- The only valid question that can be answered using the coefficients in linear regression is "What is the effect on $E(y \mid \mathbf{X})$ of changing the part of each regressors that is orthogonal to the other ones".
- "What if" analysis requires simulation and in most cases a lower level of reduction than that used for regression analysis.

The semi-partial R-squared

- When the columns of **X** are orthogonal to each other the total R^2 can be exactly decomposed in the sum of the partial R^2 due to each regressor x_i (the partial R^2 of a regressor i is defined as the R^2 of the regression of y on x_i).
- This is in general not the case in applications with non experimental data: columns of **X** are correlated and a (often large) part of the overall R^2 does depend on the joint behaviour of the columns of **X**.
- However, it is always possible to compute the marginal contribution to the overall R^2 due to each regressor x_i, defined as the difference between the overall R^2 and the R^2 ot the regression that includes all columns **X** except x_i. This is called the semi-partial R^2.

The semi-partial R-squared

Interestingly, the the semi-partial R^2 is a simple tranformation of the t-ratio:

$$spR_i^2 = \frac{t_{\beta_i}^2 \left(1 - R^2\right)}{(T - k)}$$

This result has two interesting implications.

- First, a quantity which we considered as just a measure of statistical reliability, can lead to a measure of relevance when combined with the overall R^2 of the regression.
- Second, we can re-iterate the difference between statistical significance and relevance. Suppose you have a sample size of 10000 and you have 10 columns in $\mathbf{X}$ and the t-ratio on a coefficient β_i is of about 4 with an associate P-value of the order .01: "very" statistical significant! The derivation of the semi-partial R^2 tells us that the contribution of this variable to the overall R2 is at most approximately 16/(10000-10) that is: less than two thousands.

Model Mis-specification

- Each model specification can be interpreted of the result of a reduction process, what happens if the reduction process that has generated $E(y \mid \mathbf{X})$ omits some relevant information?
- There are three general cases of mis-specification.
  - Mis-specification related to the choice of variables included in the regressions,
  - Mis-specification related to ignoring the existence on constraints on the estimated parameters
  - Misspecification related to wrong assumptions on the properties of the error terms.

Choice of variables

- under-parameterization (the estimated model omits variables included in the DGP)
  - under parameterization will casue the estimated coefficients to caputure the effects of included and omitted variables unless the omitted variables are orthogonal to the included ones
- over-parameterization (the estimated model includes more variables than the DGP).
  - over-parametrization leads to consistent but less efficient estimators

Under-parameterization

Given the DGP:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}, \tag{3}$$

for which the usual hypotheses hold, the following model is estimated:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\nu}. \tag{4}$$

The OLS estimates are given by the following expression:

$$\widehat{\boldsymbol{\beta}}_1^{up} = \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{y}, \tag{5}$$

Under-parameterization

The estimates in (5) are biased unless $\mathbf{X}_1$and $\mathbf{X}_2$ are uncorrelated. To show this, substitute for $\mathbf{y}$ from the DGP :

$$\widehat{\boldsymbol{\beta}}_1 = \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\left(\mathbf{X}_1'\mathbf{y} - \mathbf{X}_1'\mathbf{X}_2\widehat{\boldsymbol{\beta}}_2\right) \tag{6}$$

$$= \widehat{\boldsymbol{\beta}}_1^{up} - \widehat{\mathbf{D}}\widehat{\boldsymbol{\beta}}_2, \tag{7}$$

where $\widehat{\mathbf{D}}$ is the vector of coefficients in the regression of $\mathbf{X}_2$ on $\mathbf{X}_1$ and $\widehat{\boldsymbol{\beta}}_2$ is the OLS estimator obtained by fitting the DGP.

Illustration

Given the DGP:

$$\mathbf{y} = \mathbf{X}_1\mathbf{0.5} + \mathbf{X}_2\mathbf{0.5} + \boldsymbol{\epsilon}_1, \tag{8}$$

$$\mathbf{X}_2 = 0.8\mathbf{X}_1 + \boldsymbol{\epsilon}_2 \tag{9}$$

the following model is estimated by OLS

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\nu}. \tag{10}$$

(a) The OLS estimate of β_1 will be 0.5
(b) The OLS estimate of β_1 will be 0
(c) The OLS estimate of β_1 will be 0.9
(d) The OLS estimate of β_1 will have a mean of 0.9

Under-parameterization

Note that if

$$\begin{aligned} E(\mathbf{y} \mid \mathbf{X}_1, \mathbf{X}_2) &= \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2, \\ E(\mathbf{X}_1 \mid \mathbf{X}_2) &= \mathbf{X}_1\mathbf{D}, \end{aligned}$$

then,

$$E(\mathbf{y} \mid \mathbf{X}_1) = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_1\mathbf{D}\boldsymbol{\beta}_2 = \mathbf{X}_1\boldsymbol{\alpha}.$$

Therefore the OLS estimator in the under-parameterized model is a biased estimator of $\boldsymbol{\beta}_1$, but an unbiased estimator of $\boldsymbol{\alpha}$.

- Then, if the objective of the model is forecasting and $\mathbf{X}_1$ is more easily observed than $\mathbf{X}_2$, the under-parameterized model can be safely used.
- On the other hand, if the objective of the model is to test specific predictions on parameters, the use of the under-parameterized model delivers biased results.

Over-parameterization

Given the DGP,

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}, \tag{11}$$

for which the usual hypotheses hold, the following model is estimated:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{v}. \tag{12}$$

It can be shown that OLS estimator of the over-parameterized model is

$$\widehat{\boldsymbol{\beta}}_1^{op} = \left(\mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{M}_2\mathbf{y}, \tag{13}$$

$$\mathbf{M}_2 = \left(\mathbf{I} - \mathbf{X}_2\left(\mathbf{X}_2'\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\right) \tag{14}$$

Over-parameterization

By substituting $\mathbf{y}$ from the DGP, one finds that estimators are unbiased and the difference is now made by the variance which is higher. In fact we have:

$$var\left(\widehat{\boldsymbol{\beta}}_1^{op} \mid \mathbf{X}_1, \mathbf{X}_2\right) = \sigma^2 \left(\mathbf{X}_1' \mathbf{M}_2 \mathbf{X}_1\right)^{-1}, \tag{15}$$

$$var\left(\widehat{\boldsymbol{\beta}}_1 \mid \mathbf{X}_1, \mathbf{X}_2\right) = \sigma^2 \left(\mathbf{X}_1' \mathbf{X}_1\right)^{-1}. \tag{16}$$

and it can be shown that $\mathbf{X}_1'\mathbf{X}_1 - \mathbf{X}_1'\mathbf{M}_2\mathbf{X}_1$ is a positive semidefinite matrix

Linear constraints

An analogous effect to that of over-parameterization is obtained when there are linear constraints on the coefficients that are not imposed. think for eample of the case of estimating:

$$\begin{aligned} W_{it} &= \beta_0 + \beta_1 EFG_{it} + \beta_2 OEFG_{it} + \epsilon_{it} \\ \epsilon_{it} &\sim N.I.D\left(0, \sigma^2\right) \end{aligned}$$

when the true model is:

$$\begin{aligned} W_{it} &= \beta_0 + \beta_1 \left(EFG_{it} - OEFG_{it}\right) + \epsilon_{it} \\ \epsilon_{it} &\sim N.I.D\left(0, \sigma^2\right) \end{aligned}$$

and therefore $\beta_1 = -\beta_2$.
In this case the first model will deliver less precise estimates of the coefficients.

Heteroscedasticity, Autocorrelation, and the GLS estimator

Let us reconsider the single equation model and generalize it to the case in which the hypotheses of diagonality and constancy of the conditional variances-covariance matrix of the residuals do not hold:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \\ \boldsymbol{\epsilon} &\sim n.d.\left(\mathbf{0}, \sigma^2\Omega\right), \end{aligned} \tag{17}$$

where Ω is a $(T \times T)$ symmetric and positive definite matrix. When the OLS method is applied to model (17), it delivers estimators which are consistent but not efficient; moreover, the traditional formula for the variance-covariance matrix of the OLS estimators, $\sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}$, is wrong and leads to an incorrect inference. Different type of estimators, known as GLS, should be applied in this case to obtain valid inference.