Regression #2

Econ 671

Purdue University

Estimation

- In this lecture, we address estimation of the linear regression model.
- There are many objective functions that can be employed to obtain an estimator; here we discuss the most common one that delivers the familiar *OLS estimator.*
- We then discuss issues of parameter interpretation, prediction and review details associated with R-squared.

Estimation

$$y_i = x_i\beta + \epsilon_i.$$

The most widely employed approach seeks to minimize the contribution of the error term ϵ_i by minimizing the sum of squared residuals:

$$\min_{\tilde{\beta}} \sum_i (y_i - x_i\tilde{\beta})^2 = \min_{\tilde{\beta}_1, \tilde{\beta}_2, \ldots, \tilde{\beta}_k} \sum_i (y_i - \tilde{\beta}_1 - \tilde{\beta}_2 x_{i2} - \cdots - \tilde{\beta}_k x_{ik})^2.$$

Unlike the simple regression case, where we consider $k = 2$ specifically, and derive an estimator for that particular case, we seek to obtain an estimator when k is an arbitrary number.

Estimation

A "representative" first-order condition from this objective function (differentiating with respect to $\tilde{\beta}_j$) yields an equation of the form:

-

This implies that, for the intercept parameter:

-

The complete vector $\hat{\beta}$ is obtained as the solution of this set of k equations in k unknowns.

Estimation

We can assemble these k equations together in vector / matrix form as:

$$\begin{bmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{nk} \end{bmatrix} \left(\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} \right) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

or, compactly in terms of our regression notation,

-

Under the assumptions of our regression model, then,

-

Estimation

Of course, arriving at $\hat{\beta} = (X'X)^{-1}X'y$ is easier and more direct if we simply apply rules governing vector differentiation. (See, Appendix A). That is, we seek to minimize:

$$\min_{\tilde{\beta}} (y - X\tilde{\beta})'(y - X\tilde{\beta})$$

or

$$\min_{\tilde{\beta}} \left(y'y - \tilde{\beta}'X'y - y'X\tilde{\beta} + \tilde{\beta}'X'X\tilde{\beta} \right).$$

Differentiating with respect to the vector $\tilde{\beta}$ and setting the result to zero gives:

$$-2X'y + 2X'X\hat{\beta} = 0$$

or

$$\hat{\beta} = (X'X)^{-1}X'y.$$

Estimation

- Thus we have a simple-to-calculate, closed form solution for the estimated coefficient vector.
- Given this estimated coefficient vector, fitted (predicted) values are easily obtained:

$$\hat{y} = X\hat{\beta}$$

- as are the residuals:

$$\hat{\epsilon} = y - X\hat{\beta}.$$

Interpretation

- As you all know, multiple regression is advantageous in that it allows the researcher to "control" for other factors when determining the effect of a particular x_j on y.
- Indeed, the language "*After controlling for the influence of other factors, the marginal effect of x_j on y is $\hat{\beta}_j$*" is commonly used.
- In the following subsection, we justify this interpretation in a more formal way.

Interpretation

Consider the regression model:

$$y = X_1\beta_1 + x_2\beta_2 + \epsilon.$$

- Here, X_1 represents a set of covariates that are important to account for, but are not necessarily the objects of interest.
- x_2 is regarded as a vector (for simplicity and without loss of generality), so that β_1 is a $(k-1) \times 1$ vector while β_2 is a scalar.
- Some questions: *How can we get $\hat{\beta}_2$ directly*? *Does this provide any insight behind the interpretation of multiple regression coefficients*?

Interpretation

$$y = X_1\beta_1 + x_2\beta_2 + \epsilon.$$

We can write this as

$$y = [X_1 \;\; x_2] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \epsilon.$$

To calculate

$$\hat{\beta} = (X'X)^{-1}X'y,$$

we then note that

$$X'X = \begin{bmatrix} X_1' \\ x_2' \end{bmatrix} [X_1 \;\; x_2] = \begin{bmatrix} X_1'X_1 & X_1'x_2 \\ x_2'X_1 & x_2'x_2 \end{bmatrix}.$$

Interpretation

$$y = X_1\beta_1 + x_2\beta_2 + \epsilon.$$

Likewise,

$$X'y = \begin{bmatrix} X_1' \\ x_2' \end{bmatrix} y = \begin{bmatrix} X_1'y \\ x_2'y \end{bmatrix}.$$

Putting these two equations together, we then obtain:

$$\begin{bmatrix} X_1'X_1 & X_1'x_2 \\ x_2'X_1 & x_2'x_2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} X_1'y \\ x_2'y \end{bmatrix}.$$

This produces two "equations:" the first, a vector-valued equation for $\hat{\beta}_1$ and the second a scalar equation for $\hat{\beta}_2$.

Interpretation

The first of these equations gives:

-

We can rearrange this to get:

-

The second of these equations also gives:

-

Interpretation

The last slide gave two equations in two unknowns. We can substitute the first of these into the second to get an expression that only involves $\hat{\beta}_2$:

-

Regrouping terms gives:

-

Let us define

-

so that

-

This provides a "direct" way to estimate the *short regression* coefficient β_2.

Interpretation

The matrix M_1 has some nice properties:

- M_1 is *symmetric*:

- M_1 is *idempotent*:

Interpretation

With this in mind, we provide an alternate way to obtain $\hat{\beta}_2$ which helps to clarify its interpretation:

Theorem

$\hat{\beta}_2$ *can be obtained from the following two-step procedure:*

1.

2.

Interpretation

Proof.

The proof of this result is straightforward. To see this, consider the regression in step 1:

$$x_2 = X_1\theta + u$$

The estimated residuals are obtained as:

□

Interpretation

Proof.

In the second step, we fit the regression:

$$y = \gamma \hat{u} + \eta.$$

Therefore,

□

Interpretation

- So, $\hat{\beta}_2$ can be obtained from this two-step procedure.
- The procedure is rather intuitive: The way that we estimate β_2 is to first find the part of x_2 that cannot be (linearly) explained by X_1. (These are the residuals, $\hat{u}$). We then regress y on this "leftover" part of x_2 to get $\hat{\beta}_2$.
- This procedure is often referred to as *residual regression*. It also clearly justifies our common interpretation of the coefficient as "after controlling for the effect of X_1."

Coefficient of Determination: R^2

- Undoubtedly, most (all?) of you are familiar with the *coefficient of determination* or R^2.
- R^2 is often a useful diagnostic, measuring the proportion of variation in the data that is explainable by variation among the explanatory variables in the model.
- In what follows, we briefly review its derivation and discuss some of its properties.

Coefficient of Determination: R^2

To begin, define the *total sum of squares (TSS)* as follows:

-

Likewise, we can define the *explained sum of squares (ESS)* as follows:

-

and, again, in vector / matrix form, this can be written as:

-

and the *residual sum of squares*:

-

Coefficient of Determination: R^2

We now introduce two results, both rather obvious, but will be needed when constructing R^2:

$$\iota'\hat{\epsilon} = 0$$

Proof.

□

Our second result states:

$$\left(\hat{y} - \iota\overline{y}\right)'\hat{\epsilon} = 0.$$

Coefficient of Determination: R^2

Proof.

$\square$

The second line follows from the result we have just proved. The last line again follows from the fact that $X'\hat{\epsilon} = 0$, by construction of the OLS estimator.

Coefficient of Determination: R^2

With these result in hand, we can now derive R^2. To this end, note:

-

which implies

-

Since this holds for each element of the above vectors, the sum of squared elements must also be equal:

-

Using our previous result, the right hand side reduces to:

-

or

-

Coefficient of Determination: R^2

The coefficient of determination, R^2 is then defined as:

$$R^2 = \frac{ESS}{TSS} \quad \text{or} \quad R^2 = 1 - \frac{RSS}{TSS}.$$

Some properties:

-
-
-
-