ECONOMETRIC MODEL WITH CROSS-SECTIONAL, TIME SERIES, AND PANEL DATA

Tomescu Dumitrescu Cornelia

Universitatea „Constantin Brâncuşi" from Tg-Jiu, Street Victoria, Nr. 24, Tg-Jiu, Gorj county

In this paper is performed the ways to estimate parameters of a linear regression model for that models which use different type of data sets: cross-sectional data, time series data, and panel data.

Keywords: regression model, intercept, slope, marginal effect.

Estimation with straightforward OLS

Data sets may be of three types: cross-sectional data, time series data, and panel data.

Cross-sectional data sets are generated at one moment in time and the observations generally relate to households, individuals, enterprises, or geographical areas. It is usually very desirable that the sample should be drawn from a well-defined population using a statistically respectable sampling scheme, so that one may generalize from the results.

Time series data sets consist of repeated observations on a set of variables over an interval of time. Generally the interval between the observations is fixed, often being a year, a quarter, or a month, but in some cases, such as analysis using stock market prices, the frequency may be much greater.

Panel data sets have both cross-sectional and time series dimensions, being repeated observations over an interval of time on the same cross-section sample.

Consider the model :

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_2 X_3 + u \tag{1}$$

This is linear model in parameters and it may be fitted using straightforward OLS, provided that the regression model assumptions are satisfied. However, the fact that it is nonlinear in variables has implications for the interpretation of the parameters. In the multiple regressions the slope coefficients may represent the separate, individual marginal effects of the variables on Y, holding the other variables constant. In this model, such an interpretation is not possible. In particular, it is not possible to interpret β_2 as the effect of X_2 on Y, holding X_3 and X_2X_3 constant, because it is not possible to hold both X_3 and X_2X_3 constant if X_2 changes.

To give a proper interpretation to the coefficients, we can rewrite the model as:

$$Y = \beta_1 + (\beta_2 + \beta_4 X_3) X_2 + \beta_3 X_3 + u \tag{2}$$

The coefficient of X_2, $(\beta_2 + \beta_4 X_3)$, can now be interpreted as the marginal effect of X_2 on Y, holding X_3 constant. This representation makes explicit the fact that the marginal effect of X_2 depends on the value of X_3. The interpretation of β_2 now becomes the marginal effect of X_2 on Y, when X_3 is equal to zero.

One may equally well rewrite the model as:

$$Y = \beta_1 + \beta_2 X_2 + X_3(\beta_3 + \beta_4 X_2) + u \tag{3}$$

From this it may be seen that the marginal effect of X_3 on Y, holding X_2 constant, is $(\beta_3 + \beta_4 X_2)$ and that β_3 may be interpreted as the marginal effect of X_3 on Y, when X_2 is equal to zero.

If $X_3 = 0$ is a long way outside its range in the sample, the interpretation of the estimate of β_2 as an estimate of the marginal effect of X_2 when $X_3 = 0$ should be treated with caution. Sometimes the estimate will be completely implausible, in the same way as the estimate of the intercept in a regression is often implausible if given a literal interpretation. This can make it difficult to compare the estimates of the effects of X_2 and

X_3 on Y in models excluding and including the interactive term. One way of mitigating the problem is to rescale X_2 and X_3 so that they are measured from their sample means:

$$\begin{aligned} X_2^* &= X_2 - \overline{X}_2 \\ X_3^* &= X_3 - \overline{X}_3 \end{aligned} \tag{4}$$

We will obtain:

$$\begin{aligned} X_2 &= X_2^* + \overline{X}_2 \\ X_3 &= X_3^* + \overline{X}_3 \end{aligned} \tag{5}$$

Substituting for X_2 and X_3, the model becomes:

$$Y = \beta_1 + \beta_2\left(X_2^* + \overline{X}_2\right) + \beta_3\left(X_3^* + \overline{X}_3\right) + \beta_4\left(X_2^* + \overline{X}_2\right)\left(X_3^* + \overline{X}_3\right) + u$$

$$Y = \beta_1 + \beta_2 X_2^* + \beta_2 \overline{X}_2 + \beta_3 X_3^* + \beta_3 \overline{X}_3 + \beta_4 X_2^* X_3^* + \beta_4 X_2^* \overline{X}_3 + \beta_4 \overline{X}_2 X_3^* + \beta_4 \overline{X}_2 \overline{X}_3 + u \tag{6}$$

If we make follow notations:

- $\alpha_1 = \beta_1 + \beta_2 \overline{X}_2 + \beta_3 \overline{X}_3 + \beta_4 \overline{X}_2 \overline{X}_3$
- $\alpha_2 = \beta_2 + \beta_4 \overline{X}_3$
- $\alpha_3 = \beta_3 + \beta_4 \overline{X}_2$

we will obtain:

$$Y = \alpha_1 + \alpha_2 X_2^* + \alpha_3 X_3^* + \beta_4 X_2^* X_3^* + u \tag{7}$$

The point of doing this is that the coefficients of X_2 and X_3 now give the marginal effects of the variables when the other variable is held at its sample mean, which is to some extent a representative value. For example, rewriting the new equation as:

$$Y = \alpha_1 + X_2^*\left(\alpha_2 + \beta_4 X_3^*\right) + \alpha_3 X_3^* + u \tag{8}$$

it can be seen that α_2 gives the marginal effect of X_2^*, and hence X_2, when $X_3^* = 0$, that is, when $X_3 = 0$ is at its sample mean. α_3 has a similar interpretation.

The first example is the econometric model of the consumption function:

$$C = a + bY + u \tag{9}$$

If we now have a data sample over time - *this is called a time series* - of data on C and Y, we would write this as:

$$Ct = a + bYt + ut; \quad t=1,2,\ldots,n \tag{10}$$

Interpreting the intercept and the slope of a simple regression

The estimate of the intercept is usually of little interest. If the intercept was not in the model then the regression line would be forced to pass through the origin which might not be appropriate. Sometimes the intercept is interpreted as the value of the dependent variable when the explanatory is zero. Often this counterfactual does not make much sense. Slightly unusually, in this model the intercept does have an interpretation in relation to the average propensity to consume.

The estimate of the slope is more interesting. Precisely what it means depends on the model. The slope parameter in this model is the marginal propensity to consume.

As mentioned above the standard errors of the estimated parameters are estimates of the standard deviation of these estimates and are thus an indicator of their precision. The square of the standard errors gives an estimate of the variance of the estimated parameters.

It is important to remember that estimates in statistics are realizations of random variables. Random variables generally have a certain distribution and any good estimation method provides estimates with a known distribution.

Are these regression results any good?

It appears that these regression results entirely confirm Keynes' theory that consumption is a linear function of disposable income and that the slope parameter (the mpc) lies between zero and one. So far so good.

But are these regression results reliable? If they are, this is important information for policy makers and other agents in the economy. If not, they are likely to be misleading. It is crucial to ask these questions when considering regression results and a large part of Econometrics is devoted to answering these kinds of questions. At the moment, two ways of tackling this problem are suggested, both of them graphical.

> (i) The model ASSUMES that the relationship between consumption and disposable income is linear. Is this true? A scatter diagram of Ct against Yt provides one way of answering this question.
>
> (ii) Although we do not observe the random disturbances we can observe the regression residuals. These can be graphed over time. In the linear regression model, the disturbances are often assumed to be independent of each other.

Are the regression residuals also independent?

Production Function Example

An example of an econometric model of production using the Cobb-Douglas production function has the forme:

$$y = A + \alpha_1 k + \alpha_2 l + u \tag{11}$$

Where y = log(Y), A = log(α_0), k = log(K), l = log(L) and the random disturbance is now called *u*.

Equation (11) is an example of some estimates of a multiple regression. It is called multiple because there is more than one explanatory variable. As before we will not pay much attention to the estimate of the intercept. It is the estimates of the slopes which are much more important. Are these estimates the sort of estimates which we would expect?

To answer this question we have to consider the model which we are estimating. Recall that this is the Cobb-Douglas production function. For this production function, if the industry (or industries) concerned are competitive and pay both capital and labor inputs their marginal products, then the slope parameters (□1,□2) have the interpretation that they are the shares of total output paid to each factor. Thus we expect the slopes to be positive and less than one. Both the point estimates satisfy these requirements.

In both the examples given above, we began with point estimates and then used the standard errors of the estimated parameters and knowledge of their distribution to construct confidence intervals. We can also use the same information to test hypotheses about the parameters of interest.

In the consumption function case we might wish to test the hypothesis that b = 1. Thus:

$$H0 : b \geq 1$$

$$H1 : b < 1$$

In the production function example if we wished to test the hypothesis that $\alpha_1 = 1$,we could do so by testing the null against a two-sided alternative:

$$H0 : \alpha_1 = 1$$

$$H1 : \alpha_1 \neq 1$$

References:

1. ANSION, G. Économétrie pour l'entreprise, Paris, EYROLLES, 1988
2. APOSTOL, T. M. (1974): Mathematical Analyşis. Addison-Wesley, Rading, Massachusetts.
3. CUTHBERTSON, K., TAYLOR, M. P., Applied Econometric Techniques, Exeter, BPCC
4. DURLAUF, S. N., MACCINI, L. J., Measuring Noise in Inventory Models, Journal of Monetary Economics, 1995
5. DUSSAIX, A. M., INDJEHAGOPIAN, J. P., Méthodes statistiques appliquées á la gestion, Les éditions d'organisation, Paris, 1979
6. HANSEN, L. P., SINGLETON, K. J., Efficient Estimation of Liniar Asset-Pricing Models with Moving Average Errors, Journal of Business and Economic Statistics, 1996
7. MASAO, O., KYUNGHO, J., HYOUNG-SEOK, L., YOUNGSOO, B., Structural Macroeconometrics, 2007
8. PECICAN, E. Macroeconometrie. Politici economice guvernamentale şi econometrie, Bucureşti, Editura Economică, 1996
9. TOMESCU DUMITRESCU, C., Econometrie generală şi financiară, Bucureşti, Editura Didactică şi Pedagogică, 2008