Econometrics notes round 1.

Chapter 1: here is where everything starts.

[Important notation]:

represents the independent variable for observation t. In a sample of size *n*, there are *n* ’s.

represents the vector for the independent variable for all observations. In a sample of size n,

represents the matrix which contains multiple independent for observation t.

represents the matrix which contains multiple independent variables for all observations.In a sample of size n,

We start from the simple linear regression model. e.g.,

This model can also be written as:

Or equivalently:

In this model, we tend to use the independent variable to (partially) ‘explain’ .

The subscript *t* indicates an observation (e.g., the observation ‘*t*’) in a sample (of size *n*). In this example, observation *t* contains a dependent variable of , and an independent variable of . These are what we observe in a sample. The observation *t* also contains the error term which represents the effect on by influencing factors other than . represents our ignorance to the economic mechanism in the reality. We do not observe (otherwise we would include them as part of ).

We assume follows a distribution with an expected value of zero, i.e., or , as we assume that ‘the factors which we ignore in this model are just as likely to make bigger than it would have been if those factors were absent as they are to make smaller, for observation *t*. Thus, on average, the effects of the neglected influencing factors tend to cancel out. We also assume that the expected value of conditional on equals to the expected value of . e.g., . is mean independent of . Since , we have . By definition, is a variable depending on the value of . We make this assumption of *predeterminedness* – this indicates that that, the expected value of the error term does not depend on anything in the same observation *t*.

**1.3 we assume that the regression model is correctly specified**

We have the following regression model:

(1.01)

Since we are usually interested in the expected value of y, we can write the model specification as:

When we estimate the parameters using MM, we make two assumptions which then lead to:

In this case, we implicitly assume that the model (1.01) is **correctly specified**, e.g., the model embeds the DGP. However, sometimes this may not be true, e.g., suppose the DGP is generated by:

(1.18)

where

If (1.18) is true, then the error term in our model (1.01) would be , and

Therefore, the assumption of which we made for the model (1.01) will not hold, and this would lead to consequence which we discuss later.

**Error terms**

We assume that all the error terms (e.g., , including , , …, ) come from the same distribution which has a mean of 0, and all the error terms are mutually independent (this suggests that the observations are collected randomly):

is equivalent to

That is, , , …, all follow the same distribution of zero mean value. The variance-covariance matrix for is an identify matrix with . The covariance between any and ( is zero. The variance of , , …, are all equal (e.g., ).

**Linear and non-linear models and the interpretation of log function**

**1.5 method of moments (MM) estimator**

In econometrics, we need to develop the model (which is assumed to be correctly specified) and we collect a sample of data – we then estimate the parameters which describe the relationship between the variables. When we estimate the parameters, we need to calculate the population moments (e.g., mean, variance). We do not observe the population. Thus, we use the sample moments to represent the population moments, which is the idea of MM.

e.g., suppose we have the following model (assumed to be correctly specified):

(1.01)

We previously assumed that the mean of the error term is zero, e.g., . However, we do not observe the whole population but only have data for a sample. Thus, we represent the population mean of the error using the sample mean of the error term:

(1.40)

Where *n* is the size of the sample.

We have another assumption that the mean of the error term conditional on the value of the explanatory variable is equal to the mean of the error term (which equals to zero). e.g., . According to the Law of Iterated Expectation, we have: . Again, we represent the population error mean using the sample error mean:

(1.42)

We can rewrite (1.40) and (1.42) in the matrix notation, and we have:

(1.45)

In (1.45), . is a vector of 1’s, is a vector with typical elements of (e.g., , , …, ), and is a vector of errors (e.g., , … ). Thus, where the upper element of the matrix represents (1.40), and the lower element represents (1.42).

(1.45) can be rewritten as:

Thus, we have:

The MM estimator also applies when there are more explanatory variables in the regression model.

**Linear and nonlinear regression models**

We may have more sophisticated models compared to the simple regression model, e.g.,

In the model, by definition

This is still a linear regression model as the dependent variable is a linear function of the parameters (e.g., it is easy to estimate the parameters).

It is reasonable that we have non-linear regression models, e.g.,

(1.23)

In this model, the independent variables affect the dependent variable in a multiplicative way (sometimes this is more reasonable). We may still assume that . However, (1.23) is rarely used because the (1.23) assumes that the independent variables affect the dependent variable in a multiplicative way, but the unobserved influencing factors (e.g., the error term) affect the dependent variable in an additive way – which is not very reasonable. Therefore, we may twist (1.23) as:

(1.24)

Where is a dimensionless number (e.g., a ratio), and we assume .

However, (1.24) is difficult to estimate.

If the model (1.24) is a good model, we expect to be small (e.g., <0.05), thus we can use an approximation, e.g.,

Thus, we can rewrite (1.24) as:

(1.26)

Thus, we have a loglinear regression model (1.26) which is easy to estimate than (1.24). However, we also pay the price:

(1.24) tells us the expected value of conditional on .

(1.26) tells us the expected value of conditional on .

Often, we are interested in the effect of on and for simplicity we use (1.26) (e.g., we may find to be positive or negative with statistical significance, and we conclude that has a positive or negative effect on ). The reason that we could use (1.26) instead of (1.24) is that the logarithm transformation is monotonous (e.g., if has a positive/negative effect on , it also has a positive/negative effect on , and *vice versa*).

Using (1.26) as a forecasting model will be subject to transformation bias if we simply take an exponential form of the forecast value, e.g., say, we have a forecast value and we believe it is , and we take its exponential value and we think we have. However, in fact, the forecast we generate is not, but , and . There are procedures to mitigate the bias.

Using loglinear regression models such as (1.26) has two benefits, e.g., for the model:

We can take the partial derivative of, say, , on both side:

Thus, is the percentage change of given the percentage change of , which is interpreted as the elasticity of on .

The second benefit of (1.26) is that, in practice, the (in)dependent variables could have long tail in their distributions, by taking logs we can transform/twist them (and more importantly, the error term) to make them closer to a normal (or at least, symmetric) distribution – this leads to many desirable properties which, e.g., makes later statistical inference more reliable.