Linear Models in Econometrics

Nicky Grant

At the most fundamental level econometrics is the development of statistical techniques suited primarily to answering economic questions and testing economic theories. For example to answer the question 'what is the return to an extra years education?' or 'what will inflation be next quarter given past economic fundamentals?' Alternatively we may wish to test the implications of some economic theory or idea, for example to test the Efficient Market Hypothesis or whether consumers make decisions that satisfy the classical axioms of choice.

The most used and studied statistical technique is the Linear Model, in particular Ordinary Least Squares (OLS) which this note will recap. The linearity assumption (defined below) is largely made for simplicity than reality, as establishing the properties of estimators of linear models is more tractable and has a clear visual interpretation.

Commonly we wish to look at the relationship between a a set of variables x_i (known as the *Dependent Variable*) and y_i (known as the *Independent Variable*). For example we may wish to estimate the effect of an extra years education on wages, the effect of an increase in taxation of number of cigarettes consumed.

The Linear Model assumes that

$$\text{Linear in Parameters(LIP)} \qquad y_i = x_i'\beta_0 + u_i \tag{1}$$

for every observation in the population where the subscript i may refer to an individual, a country or such like.[1] Here y_i is the scalar dependent variable, x_i is a $p \times 1$ vector of dependent variables and β_0 a $p \times 1$ vector known as the 'true parameter' where u_i is a scalar unobserved variable that generates y_i over and above x_i. Here $x_i = (x_{i1}, .., x_{ip})'$ and $\beta_0 = (\beta_{01}, .., \beta_{0p})'$ where $x_i'\beta_0 = \sum_{j=1}^{p} x_{ij}\beta_{0j}$.

[1] The discussion here is not solely linked to Cross Sectional Data. The same argument made here holds for Time Series Data if the assumptions here hold for the time series data considered. Issues arise for Time Series Data which don't often occur in most cross section settings and as such are studied separately. To focus discussion we will consider Cross Section Data where Time Series Data will be studied in ECON61001. Panel Data will not be covered in ECON61001.

We firstly provide discussion on the Ordinary Least Squares Estimator which is one of the most popular estimation techniques in econometrics. OLS (defined below) is an estimator based on the assumption that u_i is unrelated with x_i. More formally when

$$\text{Zero Conditional Mean (ZCM)} \qquad E[u_i|x_i] = 0 \tag{2}$$

i.e that u_i is mean independent of x_i. An implication of ZCM is that u_i and x_i are statistically orthogonal, i.e

$$\text{Exogeneity} \qquad E[u_i x_i] = 0 \tag{3}$$

where $E[\cdot]$ is the expectation taken over the distribution of (u_i, x_i). Intuitively this is the average of $u_i x_i$, across the whole of the population and is sometimes referred to as the 'population average'. This assumption is known as the exogeneity condition and intuitively says that x_i are u_i do not move together on average. In the case x_i includes a constant (as is usually the case in practise) then $E[u_i x_i] = 0$ is equivalent to $Corr(x_{ij}, u_i) = 0$ for all $j = \{1, .., p\}$ [see Exercise Sheet on OLS]. In essence Exogeneity says the correlation between the regressors and the unobservable is 0.[2]

Exogeneity intuitively says that all the unobserved variables (or variables we observe but have not been included) are uncorrelated with the dependent variables x_i. In practise we must use economic intuition or otherwise to justify whether this assumption is likely to hold. In many cases it may not (see the discussion on wage-schooling regressions below). Assuming LIP then taking conditional expectations on both sides of (1)

$$E[y_i|x_i] = x_i'\beta_0 + E[u_i|x_i] \tag{4}$$

where under ZCM $E[u_i|x_i] = 0$ and hence $E[y_i|x_i] = x_i'\beta_0$. An equivalent way of thinking about the ZCM and LIP assumption (the crucial assumptions underpinning OLS) is that the conditional mean of y_i given a particular value of x_i is $x_i'\beta_0$. This says the average value of y_i for an individual with a particular x_i is $x_i'\beta_0$. So for example in the wage-schooling equation in (9) if $e_i = 10$ then $E[y_i|x_i] = \beta_{01} + 10\beta_{02} - 10^2\beta_{03}$. Once we know the parameters β_0 we know the average wages of people with different levels of education. OLS is a method to estimate these parameters which provides valid inference under a host

[2]Often these two assumptions are taken to mean the same thing. Though not technically the case, in that ZCM implies exogeneity but not the reverse and hence is a stronger assumption, the difference between the two assumptions in practise is often irrelevant. Discussion of these finer points will be made in ECON61001

of conditions- the crucial ones being LIP and Exogeneity. The LIP assumption is crucial in determining the properties in OLS (defined below) with which to base inference and is somewhat restrictive. LIP does not rule out non linearities in the relationship between the independent and dependent variable, just that the parameters β_0 enter linearly and also the unobservable u_i impacts y_i separate to x_i (known as additive separability). For example LIP rules out the case where

$$y_i = \frac{1}{1 + x_i'\beta_0} + u_i \tag{5}$$

or where the unobserved variables impact y_i in a non-superable way (i.e separate to the effects of x_i), for example if

$$y_i = x_i'\beta_0 u_i \tag{6}$$

or more generally where

$$y_i = f(x_i, u_i) \tag{7}$$

for some function $f(\cdot)$ where $f(x_i, u_i) \neq x_i'\beta_0 + u_i$ for any β_0. In essence OLS requires $E[y_i|x_i] = x_i'\beta_0$ which can be restrictive.

For example if we are interested in the wage return (w_i) to an extra year of education (e_i) we may assume

$$w_i = \beta_{01} + \beta_{02}e_i + u_i \tag{8}$$

so that $y_i = w_i$ and $x_i = (1, e_i)'$ so that β_{02} is the return to an extra year of education (all other things held constant). We may believe that the effect of an extra year of education on wages may not be linear. Namely the marginal effect of an extra year of education is likely to decrease at higher levels of education. It may be that

$$w_i = \beta_{01} + \beta_{02}e_i + \beta_{03}e_i^2 + u_i \tag{9}$$

where $\beta_{02} > 0$, $\beta_{03} < 0$ so that returns to education increases in e_i up to a point then begin to decrease. LIP allows this case as the parameters $\beta_{01}, \beta_{02}, \beta_{03}$ all enter linearly.

LIP is a strong assumption which provides a lot of structure with which to estimate the paramaters of interest. However without further information on the distribution of u_i and its relationship to x_i we have no way of identifying what is β_0. Many different forms of estimators exist which provide valid inference on β_0 under different assumptions on u_i

(esp. in relation to the dependence with x_i). In practise we try to choose an estimator which will work given what we assume is true about the distribution of u_i.

We now outline the OLS estimator. Define $u_i(\beta) = y_i - x_i'\beta$ so that $y_i = x_i'\beta + u_i(\beta)$ (which can always be performed for any β, breaking up y_i in to $x_i'\beta$ and the residual $u_i(\beta)$). Given a dataset of (y_i, x_i) for $i = \{1, .., n\}$ (i.e sample of size n) the Sum Of Squared Residuals (SSR) for a particular β is

$$SSR(\beta) = \frac{1}{n}\sum_{i=1}^{n} u_i(\beta)^2 \tag{10}$$

and is the average of the deviation of y_i from $x_i'\beta$ squared. The OLS estimator of β_0 (defined as $\hat{\beta}_{OLS}$) is the minimiser of the sum of squared residuals which we can show is equal to

$$\hat{\beta}_{OLS} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} y_i x_i. \tag{11}$$

You will likely have seen the formula for the OLS estimator in the case of a Simple Linear Regression (sometimes known as a Univariate Regression) where x_i includes a constant and one variable. The OLS Exercise Sheets asks you to show the general OLS formula is equivalent to the OLS formula in the simple linear regression.

OLS may be viewed as the 'line of best fit' to a set of data points. The closer $x_i'\beta$ is to y_i across $i = \{1, .., n\}$ the smaller is $SSR(\beta)$. In essence OLS chooses β to make $x_i'\beta$ as close as possible to y_i. This method also performs the decomposition of y_i in to a function of $x_i'\hat{\beta}_{OLS}$ and the residual $u_i(\hat{\beta}_{OLS})$ which are uncorrelated in the sample. Hence they mimic the exogeneity condition where $u_i = u_i(\beta_0)$ is uncorrelated with $x_i'\beta_0$.

Another way to view OLS is that it enforces the moment condition $E[u_i x_i] = 0$ in the sample. Namely under LIP and Exogeneity $E[x_i u_i(\beta)] = 0$ at $\beta = \beta_0$ with solution

$$\beta_0 = E[x_i x_i']^{-1} E[y_i x_i] \tag{12}$$

when $E[x_i x_i']$ is full rank.[3] OLS solves the sample version of this moment condition, namely it solves $\frac{1}{n}\sum_{i=1}^{n} x_i u_i(\beta) = 0$ at $\beta = \hat{\beta}_{OLS}$. This sample moment condition is the first order condition of the minimisation of the sum of squared residuals and the two interpretations are equivalent. The ECON61001 goes in to these issues in more details.

To study the statistical properties of OLS for large sample sizes we must make further assumptions on the distribution of (y_i, x_i, u_i) for $i = \{1, .., n\}$. If the data are *independent*

[3] This is known as the no perfect multicollinearity assumption and says we do not include redundant linear combinations of variables. For example including the same dependent variable twice.

and identically distributed (i.i.d) then as n increases $\frac{1}{n}\sum_{i=1}^{n} x_i x_i'$ converges in probability to $E[x_i x_i']$ and similarly $\frac{1}{n}\sum_{i=1}^{n} y_i x_i$ converges in probability to $E[y_i x_i]$.[4] Intuitively this says as the sample size increases the sample average becomes closer and closer to the population moment with increasing probability. These concepts will be explained formally and in more detail in the Econometric Methods course (ECON61001).

The OLS estimator will provide a 'good' estimator of β_0 under the assumptions above, where the crucial and most stringent assumptions are the LIP and Exogeneity (or the stronger condition ZCM) such that β_0 satisfies (12). By a good estimator we mean one which satisfies a set of favourable statistical conditions for example that the OLS estimator is consistent (i.e $\hat{\beta}_{OLS}$ is 'close' to β_0 for n large).[5]

The Exogeneity Assumption, much like the LIP assumption may not hold in practise and is largely made for simplicity. For example in the wage-education regressions above we have not controlled for natural ability. As such natural ability (along with a host of other factors) are in u_i (all other variables driving wages over and above education). It is likely that more able people earn more and are also more likely to have higher levels of education. Hence if e_i increases on average it is likely that u_i increases (as people with more education tend to be more able on average) and as such the OLS estimator will likely overestimate the true return to education.

This is the classic example of Endogeneity, where $E[u_i x_i] \neq 0$ (i.e the exogeneity condition does not hold). The issue of endogeneity is one of the most commonly faced problems in applied econometric research such that OLS is often not a 'good' estimator of β_0 since the OLS estimator is built to work well under the exogeneity assumption. In this case alternative (often more general) estimators are required that provide valid inference in more general settings than OLS often under a further set of assumptions. The classic example is the Instrumental Variables (IV) Estimator which you should have covered in your undergraduate notes. The IV (and more general the 2 Stage Least Squares estimator)

[4]The i.i.d assumption is a sufficient condition for this result. A similar result can be found under different conditions to allow for example dependent data. Commonly the i.i.d assumption as with the other assumptions is made as a starting point for simplicity to simplify proofs of the statistical properties of OLS. The ECON61001 course will start making the simplest assumptions known as the Classical OLS Conditions and then weaken these assumptions to allow for example dependent data, non linear models and violations of exogeneity. It is crucial that you understand the basics of OLS under the simplest assumptions and revise your undergraduate econometrics notes before starting the course.

[5]Specifically ECON61001 will provide conditions and (sketches of) proofs that the OLS estimator is Consistent, Unbiased and Asymptotically Normally Distributed. See Greene reference below.

will be covered in detail in ECON61001 and it is also advisable you refresh yourself with the basic ideas of these techniques.

Note that ECON61001 is predominantly a technical course with strong emphasis on derivations of results. Some of the more advanced proofs are omitted and emphasis is placed on providing an intuition and a sketch of the results which requires a good ability for technical and abstract thought. There are some applied examples in the course (for example applications of the Instrument Variables Estimator) though these mainly serve to highlight the theoretical results and are not the main focus of the course.

As a broad overview of the main topics covered in ECON61001 (in order)

- OLS under the Classical Assumptions (4 Lectures, properties of multivariate OLS studied likely more formally than you have seen at Undergrad **writing the problem in matrix form**. Covering consistency, unbiasedness, asymptotic normality and inference (the t and F-test covered more formally, not just memorizing how to perform the test in practise).
- Robust OLS (1 lecture, relaxes classical assumptions to allow heteroskedasticity and serially correlated errors. Introduces the Weighted Least Squares Estimator)
- Instrumental Variables (2 lectures, covers IV/2SLS, also covers the intuition of the weak instrument problem and endogeneity).
- Maximum Likelihood (1 Lecture outlines the Maximum Likelihood estimator with applications to non-linear models (specifically binary choice models- Logit/Probit)
- Time Series (1 Lecture- measures of dependence, AR/MA/ARMA modelling).

Preparation for ECON61001

The above is a broad sketch of the intuition of OLS. The first half of ECON61001 course will derive the statistical properties of OLS under a set of conditions (many of them listed above).

For those that must take ECON61001 it is highly advised that you recap and get up to speed with intermediate econometrics at the undergraduate level, especially the basic properties and ideas behind OLS. A good undergrad textbook is Introductory Econometrics by Wooldridge. The textbook used in ECON61001 is Greene, Econometric Analysis.

It may be useful to read through the first few chapters of this book on linear models and OLS. The more prepared you are before starting the course the more likely you are to perform well in the final exams/coursework. ECON61001 is quite a challenging fast paced course and is taught based on the assumption of a solid understanding at a minimum of Intermediate Econometrics with basic statistics and linear algebra. If you do not possess this knowledge it greatly increases your chance of falling behind with the work.

OLS Excercise

Consider the Linear Model

$$y_i = x_i'\beta_0 + u_i \tag{13}$$

where y_i is the scalar dependent variable, x_i is a $p \times 1$ vector of dependent variables and β_0 a $p \times 1$ vector known as the 'true parameter' where u_i is a scalar unobserved variable that generates y_i over and above x_i. Here $x_i = (x_{i1}, .., x_{ip})'$ and $\beta_0 = (\beta_{01}, .., \beta_{0p})'$. Define the residual $u_i(\beta)$ for a particular β as $u_i(\beta) = y_i - x_i'\beta$ and $SSR(\beta) = \frac{1}{n}\sum_{i=1}^n u_i(\beta)^2$. Define $\hat{\beta}_{OLS} = \left(\frac{1}{n}\sum_{i=1}^n x_i x_i'\right)\frac{1}{n}\sum_{i=1}^n y_i x_i$.

1. If $x_i = (1, w_i)'$ where w_i is a scalar and $\hat{\beta}_{OLS} = (\hat{\beta}_1, \hat{\beta}_2)'$ show that $\hat{\beta}_1 = \bar{y} - \hat{\beta}_1\bar{w}$, $\hat{\beta}_2 = \frac{\frac{1}{n}\sum_{i=1}^n (w_i - \bar{w})(y_i - \bar{y})}{\frac{1}{n}\sum_{i=1}^n (w_i - \bar{w})^2}$ where $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i$, $\bar{w} = \frac{1}{n}\sum_{i=1}^n w_i$.
2. Show if $x_i = (1, w_i)'$ for some $(p-1) \times 1$ vector (i.e x_i includes a constant) that $E[u_i x_i] = 0$ implies $Corr(x_{ij}, u_i) = 0$ for all $j = \{1, .., p\}$.

Solutions to Excercise

1. Note that $x_i = (1, w_i)'$ hence $x_i x_i' = \begin{pmatrix} 1 & w_i \\ w_i & w_i^2 \end{pmatrix}$ so that

$$\frac{1}{n}\sum_{i=1}^n x_i x_i' = \begin{pmatrix} n & \bar{w} \\ \bar{w} & \frac{1}{n}\sum_{i=1}^n w_i^2 \end{pmatrix} \tag{14}$$

Also $x_i y_i = \begin{pmatrix} y_i \\ w_i y_i \end{pmatrix}$ so that

$$\frac{1}{n}\sum_{i=1}^n x_i y_i = \begin{pmatrix} \bar{y} \\ \frac{1}{n}\sum_{i=1}^n w_i y_i \end{pmatrix} \tag{15}$$

Using (14), (15) in $\hat{\beta}_{OLS}$ then

$$\hat{\beta}_{OLS} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} y_i x_i \tag{16}$$

$$= \begin{pmatrix} n & \bar{w} \\ \bar{w} & \frac{1}{n}\sum_{i=1}^{n} w_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \bar{y} \\ \frac{1}{n}\sum_{i=1}^{n} w_i y_i \end{pmatrix} \tag{17}$$

$$= \frac{1}{\sum_{i=1}^{n} w_i^2 - \bar{w}^2} \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} w_i^2 & -\bar{w} \\ -\bar{w} & n \end{pmatrix} \begin{pmatrix} \bar{y} \\ \frac{1}{n}\sum_{i=1}^{n} w_i y_i \end{pmatrix} \tag{18}$$

$$= \frac{1}{\sum_{i=1}^{n} w_i^2 - \bar{w}^2} \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} w_i^2 \bar{y} - \bar{w}\frac{1}{n}\sum_{i=1}^{n} w_i y_i \\ \sum_{i=1}^{n} w_i y_i - \bar{y}\bar{w} \end{pmatrix} \tag{19}$$

where (18) holds using the inverse of a 2-2 matrix namely $\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ and (19) holds by matrix multiplication. By further algebraic manipulation of the first and second element of (19) the result follows.

2. Since $x_i = (1, w_i)'$ then $x_i u_i = (u_i, w_i u_i)$ and $E[x_i u_i] = (E[u_i], E[u_i w_i]) = 0$ implies $E[u_i] = 0$ and $E[w_i u_i] = 0$. Then by definition $Cov(x_{ij}, u_i) = E[x_{ij} u_i] - E[x_{ij}]E[u_i]$. Since $E[w_i u_i] = 0$ and $E[u_i] = 0$ then for any x_{ij} (i.e any element of w_i) then $Cov(x_{ij}, u_i) = E[x_{ij} u_i] - E[x_{ij}]E[u_i] = 0$ establishing the result.