**Chapter one**

Regression Analysis with Qualitative Information: Binary/ **Dummy Variables**

## Describing Qualitative Information

In regression analysis the dependent variable, or regressand, is frequently influenced not only by ratio scale variables (e.g., income, output, prices, costs, height, and temperature) but also by variables that are essentially qualitative or nominal scale, in nature, such as sex, race, color, religion, nationality, geographical region, political upheavals, and party affiliation. For example, holding all other factors constant, female workers are found to earn less than their male counterparts.

Qualitative factors often come in the form of binary information: a person is female or male; a person does or does not own a personal computer; a firm offers a certain kind of employee pension plan or it does not; a state administers capital punishment or it does not.In all of these examples, the relevant information can be captured by defining a **binary variable** or a **zero-one variable**. In econometrics, binary variables are most commonly called **dummy variables.**

In defining a dummy variable, we must decide which event is assigned the value one and which is assigned the value zero. For example, in a study of individual wage determination, we might define *female* to be a binary variable taking on the value one for females and the value zero for males. In defining a dummy variable, we must decide which event is assigned the value one and which is assigned the value zero. For example, in a study of individual wage determination, we might define *female* to be a binary variable taking on the value one for females and the value zero for males.

## Dummy as Independent Variables

How do we incorporate binary information into regression models? In the simplest case, with only a single dummy explanatory variable, we just add it as an independent variable in the equation. For example, consider the following simple model of hourly wage determination:

We use *α0* as the parameter on *female* in order to highlight the interpretation of the parameters multiplying dummy variables; later, we will use whatever notation is most convenient.

In the above model, only two observed factors affect wage: gender and education. Because *female* = 1 when the person is female and *female* = 0 when the person is male, the parameter *α0* has the following interpretation: *α*0 is the difference in hourly wage between females and males, *given* the same amount of education (and the same error term *u*). Thus, the coefficient *α0* determines whether there is discrimination against women: if *α0*< 0, then, for the same level of other factors, women earn less than men on average.

In terms of expectations, if we assume the zero conditional mean assumption E (*u***/***female*, *educ*) = 0, then

Because *female* = 1 corresponds to females and *female* = 0 corresponds to males, we canwrite this more simply as

The key here is that the level of education is the same in both expectations; the difference,*α0*, is due to gender only.

## Dummy as Dependent Variable

In the previous section, we studied how, through the use of binary independent variables, we can incorporate qualitative information as explanatory variablesin a multiple regression model. In all of the models up until now, the dependent variable *y* has had *quantitative* meaning (for example, *y* is a dollar amount, a test score, a percentage,or the logs of these). What happens if we want to use regression to *explain* aqualitative event?

In a model where *Y* is quantitative, our objective is to estimate its expected,or mean, value given the values of the regressors.

 where the *X*’s are regressors, both quantitative and qualitative

In models where *Y* is qualitative, our objectiveis to find the probability of something happening, such as having a high school education, or owning a house, or belonging to a union, or participatingin a sport etc. Hence, qualitative response regression models are often known as *probability models.*

There are three approaches to developing a probability model for a binary response variable: The **linear probability model (LPM),** the **logit model, t**he **probit model.**

### The Linear Probability Model (LPM)

In the simplest case, and one that often arises in practice, the event we would like to explain is a binary outcome. In other words, our dependent variable, *y*, takes on only two values: zero and one. For example, *y* can be defined to indicate whether an adult has a high school education; *y* can indicate whether a college student used illegal drugs during a given school year; or *y* can indicate whether a firm was taken over by another firm during a given year. In each of these examples, we can let *y* = 1 denote one of the outcomes and *y* = 0 the other outcome.

What does it mean to write down a multiple regression model, such as:

When *y* is a binary variable, because *y* can take on only two values, *βi* cannot be interpreted as the change in *y* given a one-unit increase in *xi*, holding all other factors fixed: *y* either changes from zero to one or from one to zero (or does not change). Nevertheless, the *βi* still have useful interpretations. The justification of the name LPM for models like the above can be seen this way: if we assume that the zero conditional mean assumption holds, that is, E (*u****/****x*1, …, *xk*) = 0, then we have, as always,

 **x** is shorthand for all of the explanatory variables.

Now, if *p* = probability that *y* = 1 (that is, the event occurs), and (1 − *p*) = probability that *y* = 0 (that is, that the event does not occur), the variable *y* has the following (probability) distribution.

***Yi* Probability**

0 1 −*p*

1 *p*

That is, *y* follows the **Bernoulli probability distribution.**

Now, by the definition of mathematical expectation, we obtain:

Thus; that is, the conditional expectation of the model can, in fact, be interpreted as the conditional probability of *y*. The key point is that when *y* is a binary variable taking on the values zero and one,it is always true that the probability of “success” that is, the probability that *y* = 1 is the same as the expected value of *y*. Thus, we have the important equation

which says that the probability of success, say, *p*(**S**) = P(*y* = 1**/x**), is a linear function of the *x*. this equation is an example of a binary response model, and P(*y* = 1**/x**) is also called the **response probability**. Because probabilities must sum to one, P(*y* = 0**/x**) = 1 - P(*y* = 1**/x**) is also a linear functionof the *x*.

This kind of model with a binary dependent variable is called the **linear probability model** (**LPM**) because the response probability is linear in the parameters *βi*. In the LPM, *βi*measures the change in the probability of success when *xi* changes, holding other factors fixed:

From the preceding discussion it would seem that OLS can be easily extended to binary dependent variable regression models that the mechanics of OLS are the same as before. So, perhaps there is nothing new here. Unfortunately, this is not the case, for the LPM poses several problems, which are as follows:

**(a) Non-Normality of the Disturbances *u***

Although OLS does not require the disturbances (*u*) to be normally distributed, we assumed them to be normally distributed for the purpose of statistical inference. But the assumption of normality for the error termis not tenable for the LPMs because, like *y*, the disturbances *u* also takes only two values; that is, they also follow the Bernoulli distribution. This can be seen clearly if we write the model as

The probability distribution of *u* is: ***u* Probability**

 **When *y* =1** 1 -*β0 - β1x1- … - βkxk p*

**When *y* = 0** - *β0- β1x1- … - βkxk* (1 − *p*)

Obviously, *u* cannot be assumed to be normally distributed; they follow the Bernoulli distribution.

**(b) Heteroscedastic Variances of the Disturbances**

Even if *E* (*u*) = 0, it can no longer be maintained that in the LPM the disturbances are homoscedastic. This is, however, not surprising. As statistical theory shows, for a Bernoulli distribution the theoretical mean and variance are, respectively, *p* and *p*(1 −*p*), where *p* is the probability of success (i.e., something happening), showing that the variance is a function of the mean. The mean in return is a function of x. Hence the error variance is heteroscedastic.

For the distribution of the error term given above, applying the definitionof variance, the reader should verify that

That is, the variance of the error term in the LPM is heteroscedastic.

Since *p*=, the variance of *u* ultimately depends on the values of *x* and hence is not homoscedastic.

**(C) Non-fulfillment of 0 ≤ *E* (*y*| *x*) ≤ 1**

Since *E* (*y*| *x*) in the linear probability models measures the conditional probability of the event *y* occurring given *x*, it must necessarily lie between0 and 1. Although this is true a priori, there is no guarantee that *ŷ*, the estimatorsof *E* (*y*| *x*), will necessarily fulfill this restriction, *and this is the realproblem with the OLS estimation of the LPM*.

The logit and probit models will guarantee that the estimatedprobabilities will indeed lie between the logical limits 0 and 1.

### The Logit and Probit Models

The fundamental problem with the LPM is that it is not logically a very attractive model because it assumes that *p = E(y = 1 | x)* increases linearly with *x*, that is, the marginal or incremental effect of *x* remains constant throughout. For instance, consider *y* = home ownership (1 if the person owns a house and 0 otherwise) and *x* = income in thousands, in the following example;

We find that as *x* increases by a unit ($1000), the probability of owning a house increases by the same constant amount of 0.10. This is so whether the income level is $800, $1,000, $10,000, or $22,000. This seems patently unrealistic. In reality one would expect that *p* is nonlinearly related to *x*: at very low income a family will not own a house but at a sufficiently high level of income, say, *x*\*, it most likely will own a house. Any increase in income beyond *x*\* will have little effect on the probability of owning a house.Thus, at both ends of the income distribution, the probability of owning ahouse will be virtually unaffected by a small increase in *x*.

Therefore, what we need is a (probability) model that has these two features :( 1) as *x* increases, *p*= *E*(*y*= 1 | *x*) increases but never steps outside the 0–1 interval, and (2) the relationship between *p* and *X* is nonlinear, that is, “one which approaches zero at slower and slower rates as *x* gets small and approaches one at slower and slower rates as *x*gets verylarge.’’***P***

Geometrically, the model we want -- -------------------------------**1---------------------------------**

would look something like this figure. CDF

Notice in this model that the probability

lies between 0 and 1 and that it

varies nonlinearly with *x*. (Figure 1)

1. ***X***

The reader will realize that the sigmoid, or S-shaped, curve in the figure every much resembles the **cumulative distribution function** (CDF) of a random variable. Therefore, one can easily use the CDF to model regressions where the response variable is dichotomous, taking 0–1 values. The practical question now is, which CDF? For although all CDFs are S shaped, for each random variable there is a unique CDF. For historical as well as practical reasons, the CDFs commonly chosen to represent the 0–1 response models are (1) the logistic and (2) the normal, the former giving rise to the **logit** model and the latter to the **probit** (or **normit**) model.

In a binary response model, interest lies primarily in the **response probability**

where we use ***x*** to denote the full set of explanatory variables. For example, when *y* is an employment indicator, ***x*** might contain various individual characteristics such as education, age, marital status, and other factors that affect employment status, including abinary indicator variable for participation in a recent job training program.

**Specifying Logit and Probit Models**

1. **The Logit Model**

The Logit Model is based on the assumption that the probability function follows a logistic (cumulative) distribution as follows,





Where *z = β1 + β2x*. This equation represents what is known as the (cumulative) **logistic distribution function** which is between zero and one for all real numbers *z*. This is the cumulative distribution function for a standard logistic random variable. If we let *p* be the probability of owning a house, then (1 −*p*), the probability of not owning a house is:



Thus, 

Now *p***/** (1 −*p*) is simply the **odds ratio** in favor of owning a house—the ratio of the probability that a family will own a house to the probability that it will not own a house. Thus, if *p*= 0*.*8, it means that odds are 4 to 1 infavor of the family owning a house.

Now if we take the natural logarithm, we obtain



That is, *L*, the log of the odds ratio, is not only linear in *x*, but also (from the estimation view point) linear in the parameters. *L* is called the **logit,** and hence the name **logit model.**

Notice the following important features of the logit model.

**1.** As *p* goes from 0 to 1 (i.e., as *Z* varies from −∞ to +∞), the logit *L* goes from −∞ to +∞. That is, although the probabilities (of necessity) lie between 0 and 1, the logits are not so bounded.

**2.** Although *L* is linear in *x*, the probabilities themselves are not. This property is in contrast with the LPM model where the probabilities increase linearly with *x*.

**3.** If *L*, the logit, is positive, it means that when the value of the regressor increases, the odds that (*y =* 1) (meaning some event of interest happens) increases. If *L* is negative, the odds that (*y =* 1) decreases as the value of *X* increases.

**4.** Although we have included only a single *x* variable, or regressor, in the preceding model, one can add as many regressors as may be dictated by the underlying theory.

 **(b) The Probit Model**

As we have noted, to explain the behavior of a dichotomous dependent variable, we will have to use a suitably chosen CDF. The logit model uses the cumulative logistic function. In the **probit model**, it is assumed that the probability function follows the standard normal cumulative distribution function (CDF), which is expressed as an integral:

**

This choice of *F* again ensures that the equation is strictly between zero and one for all values of the parameters and the *xi*.

The probability functions in logit and normit (probit) models are both increasing functions. Each increases most quickly at *z* = 0;*F*(*z*) → 0 as *z* →; and *F*(*z*) → 1 as *z* →∞. The logistic functionis plotted in the figure on page 4. The standard normal CDF has a shape very similar to that of the logistic CDF. In fact they are the same except that the logit probability distribution function is a bit flatter.

![C:\Users\user\Downloads\Screenshot_2018-11-03 An Introduction to Logistic and Probit Regression Models - Fall2013_Moore_Logistic_Probit_Regression [...].png]()

2

* + 1. **Interpreting the Probit and Logit Model Estimates**

In general, you cannot interpret the coefficients from the output of a probit regression (not in any standard way, at least). You need to interpret the marginal effects of the regressors, that is, how much the (conditional) probability of the outcome variable changes when you change the value of a regressor, holding all other regressors constant at some values. This is different from the linear regression case where you are directly interpreting the estimated coefficients. This is so because in the linear regression case, the regression coefficients are the marginal effects.

In the probit regression, there is an additional step of computation required to get the marginal effects once you have computed the probit regression fit.

* **Comparing Linear and probit regression models**

**Probit regression:** Recall that in the probit model, you are modelling the (conditional) probability of a "successful" outcome, that is, *yi=1*,

where F (⋅) is the cumulative distribution function of the standard normal distribution. This basically says that, conditional on the regressors, the probability that the outcome variable, *yi* is 1 is a certain function of a linear combination of the regressors.

**Linear regression**: Compare this to the linear regression model, where the (conditional) mean of the outcome is a linear combination of the regressors.

* **Marginal effects**

Other than in the linear regression model, coefficients rarely have any direct interpretation. We are typically interested in the *ceteris paribus* effects of changes in the regressors affecting the features of the outcome variable. This is the notion that marginal effects measure.

* **Linear regression**: I would now like to know how much the *mean* of the outcome variable moves when I move one of the regressors



But this is just the regression coeffcient, which means that the marginal effect of a change in the *k*th regressor is just the regression coefficient.

* **Probit regression**: However, it is easy to see that this is not the case for the probit regression



which is *not* the same as the regression coefficient. These are the *marginal effects* for the probit model, and the quantity we are after. In particular, this depends on the values of all the other regressors, and the regression coefficients. Here *F*(⋅) is the standard normal probability density function.

How do you compute this quantity and what are are the choices of the other regressors that should enter this formula? Thankfully, Stata easily provides this computation after a probit regression.

The following stata result is the probit regression with the dependent variable being 1 if graduate and 0 if not graduate.

 ***Coefficient*  [Marginal Effects]**

 ***School-M* –0.008 –0.001**

 ***School-F* 0.006 0.000**

 ***MALE* 0.063 0.004**

*Assuming that they are all statistically significant; we can interpret the result as follows:*

* Every extra year of schooling of the mother (***School-M***) decreases the probability of graduating by 0.1 percent (–0.001\*100).
* Father's schooling (***School-F***) has no discernible effect (0.000\*100).
* **Male**s have 0.4 percent higher probability of graduating than females (0.004\*100).

Similarly, the coefficients of logit results cannot be directly interpreted. The slope coefficients measure the change in *L (log of odds ratio)* for a unit change in *x*, in favor of *y=1*. The following example illustrates Logit model result of stata for Rural household involvement in non-farm (NF) activities in some woreda, where *y=*1 if the person is involved in non-farm activities, and *y=*0, otherwise.

* **Independent variable parameter odd ratio Marginal effect(dy/dx)**
	+ **Sex(F=1) 2.96 19.47 0.49**
	+ **HH head age -1.31 0.27 -0.27**
	+ **Head education 5.45 234.95 0.78**
	+ **Family size(FZ) -1.27 0.28 -0.26**

*Assuming that they are all statistically significant, we can interpret the result as follows:*

1. **Sex(F=1)**
	* As the coefficient of sex is significantly different from zero with positive sign, it puts forward that female are more likely to involve in NF activities as compared to male
	* Female , likelihood to involve in NF activities is increasing by 19.47 odd ratio as one move from male to female
	* Female involvement rate increase by the marginal effect of 49% (0.49\*100) holding all other variable constant
2. **HH head age**
	* As age of household head increase by one year, the probability of involvement of the household head in NF activities decreases the odd ratio by 0.27 from previous year
	* As age of household head increase by one year the probability of involvement of the household head in NF activities decrease 27% times likely from previous age
	* As age of household head increase by one year the probability of involvement of household head in NF activities decrease by 27% of marginal effect, holding other variable constant
3. **Head education**
	* one additional year of education will increase the likelihood of an individual to involve in NF by 78% marginal effect
4. **Family size(FZ)**
	* Odd ratio: For additional one more member to a household, the head’s probability to involve in NF activities will decrease with odd ratio 0.28
	* Marginal effect: one additional family to a household will decrease the probability of a hh to participate in NF activities by 26 %

**Chapter Two:**

**Introduction to Basic Regression Analysis with Time Series Data**

Now that we have a solid understanding of how to use the multiple regression models for cross-sectional applications, we can turn to the econometric analysis of time series data. Since we will rely heavily on the method of ordinary least squares, most of the work concerning mechanics and inference has already been done. However, as you might have noted, time series data have certain characteristics that cross-sectional data do not, and these can require special attention when applying OLS to time series data.

One of the basic points we make in econometrics is that the properties of the estimators and their usefulness for point estimation and hypothesis testing depends on how the data behave. For instance, in a linear regression model where errors are correlated with regressors, least squares won't be consistent and consequently it should not be used for either estimation or subsequent testing. In this chapter, we begin to study the properties of OLS for estimating linear regression models using time series data.

While considering the standard regression model, we did not pay attention to the timing of the explanatory variable(s) on the dependent variable. The standard linear regression implies that change in one of the explanatory variables causes a change in the dependent variable during the same time period and during that period alone. But in economics, such specification is scarcely found. In economic phenomenon, generally, a cause often produces its effect only after a lapse of time; this lapse of time (between cause and its effect) is called a lag. Therefore, realistic formulations of economic relations often require the insertion of lapped values of the explanatory or insertion of lagged dependent variables.

**2.1. The nature of Time Series Data**

More often than not, economists study time series data. For example, economists might study import-export time trends in Ethiopian GDP, consumption, investment, unemployment, inflation, interest rates and so on. Time series data are often the only window we have into important economic processes. Many data are collected or analyzed only at the national level.Unfortunately, time series data hold their own challenges.

***Trends***: persistent upward or downward movements of variables over time. It can be very difficult to disentangle trends over time. Trends can threaten the consistency and asymptotic normality of OLS. Many macroeconomic variables have long-term trends: Real GDP per capita, Real consumption per capita, Real investment per capita, Inflation (the CPI).

***Figure 2.1. GDP, Consumption, Investment, and the consumer price index (CPI), 1948–1998***



When we talk about trends, there are two common types of trends:

*Deterministic Trends*: *E(yt) – E(yt-1) = a*. The trending variable changes by a constant amount each period

*Stochastic Trends*: *E(yt) – E(yt-1) = b + vt*. The trending variable changes by a random amount each period (*vt*).

We will discuss these two trends in detail on the coming sections.

An obvious characteristic of time series data that distinguishes them from cross-sectional data is temporal ordering. A time series *yt*is a process observed in sequence over time, *t = 1, … , T*. To indicate the dependence on time, we adopt new notation, and use the subscript *t* to denote the individual observation, and *T* to denote the number of observations. Because of the sequential nature of time series, we expect that *yt* and *yt-1*to be not independent. So, classical assumptions are not valid.

For analyzing time series data, we must recognize that the past can affect the future, but not vice versa. To emphasize the proper ordering of time series data, Table 2.1 gives a partial listing of the data on U.S. inflation and unemployment rates from various editions of the *Economic Report ofthe President*.

|  |  |  |
| --- | --- | --- |
| Year | Inflation | Unemployment |
| 1998 | **1.6** | **4.5** |
| 1999 | **2.2** | **4.2** |
| 2000 | **3.4** | **4.0** |
| 2001 | **2.8** | **4.7** |
| 2002 | **1.6** | **5.8** |
| 2003 | **2.3** | **6.0** |

***Table 2.1: U.S. Inflation and Unemployment Rates, 1998 - 2003***

Another difference between cross-sectional and time series data is more subtle. In Econometrics I, we studied statistical properties of the OLS estimators based on the notion that samples were randomly drawn from the appropriate population. Understanding why cross-sectional data should be viewed as random outcomes is fairly straightforward: a different sample, drawn from the population, will generally yield different values of the independent and dependent variables (such as education, experience, wage, and so on).Therefore, the OLS estimates computed from different random samples will generally differ, and this is why we consider the OLS estimators to be random variables.

How should we think about randomness in time series data? Certainly, economic time series satisfy the intuitive requirements for being outcomes of random variables. For example, today we do not know what the Real Estate Industrial Average will be at the close of the next trading day. We do not know what the annual growth in output will be in Ethiopia during the coming year. Since the outcomes of these variables are not foreknown, they should clearly be viewed as random variables.

Formally, a sequence of random variables indexed by time is called a **stochastic Process** or a **time series process**. (“Stochastic” is a synonym for random.) A random or stochastic process is a collection of random variables ordered in time. When we collect a time series data set, we obtain one possible outcome, or *realization*, of the stochastic process.

We can only see a single realization, because we cannot go back in time and start the process all over again. (This is analogous to cross-sectional analysis where we can collect only one random sample.) However, if certain conditions in history had been different, we would generally obtain a different realization for the stochastic process, and this is why we think of time series data as the outcome of random variables. The set of all possible realizations of a time series process plays the role of the population in cross-sectional analysis. The sample size for a time series data set is the number of time periods over which we observe the variables of interest.

**2.2. Stationary and non-stationary stochastic Processes**

***Stationary Stochastic Processes***

A type of stochastic process that has received a great deal of attention and scrutiny by time series analysts is the so-called **stationary stochasticprocess.** Broadly speaking, a stochastic process is said to be stationary if itsmean and variance are constant over time and the value of the covariancebetween the two time periods depends only on the distance or gap or lag betweenthe two time periods and not the actual time at which the covariance is computed.

In the time series literature, such a stochastic process is known as a **weakly stationary,** or **covariance stationary,** or **second-order stationary, stochastic process.** To explain weak stationarity, let *Yt*be a stochastic time series with these properties:

*Mean: E*(*Yt*) = *μ* **(2.1)**

*Variance:* var (*Yt*) = *E*(*Yt*− *μ*)2 = *σ*2**(2.2)**

*Covariance: γk*= *E*[(*Yt*− *μ*)(*Yt*+*k*− *μ*)] **(2.3)**

where*γk*, the covariance (or autocovariance) at lag *k*, is the covariance between the values of *Yt*and *Yt*+*k*, that is, between two *Y* values *k* periods apart. If *k* = 0, we obtain *γ*0, which is simply the variance of *Y* (= *σ*2); if *k* = 1, *γ*1 is the covariance between two adjacent values of *Y*.

If a time series is stationary, its mean, variance, and autocovariance (at various lags) remain the same no matter at what point we measure them; that is, they are time invariant.Such a time series will tendto return to its mean (called **mean reversion**) and fluctuations around thismean (measured by its variance) will have a broadly constant amplitude.

If a time series is not stationary in the sense just defined, it is called a **nonstationary time series** (keep in mind we are talking only about weak stationarity). In other words, a nonstationary time series will have a time varying mean or a time-varying variance or both.

Why are stationary time series so important? Because if a time series is nonstationary, we can study its behavior only for the time period under consideration. Each set of time series data will therefore be for a particular episode. As a consequence, it is not possible to generalize it to other time periods. Therefore, for the purpose of forecasting, such (nonstationary) time series may be of little practical value.

A stationary time-series' statistical properties like mean & variance will be constant over time. They can (and will) move around but revert to the mean over time.

For example, **Price to Earning ration** of a stock market index, say The Standard & Poor's 500 (often abbreviated as S&P 500 which is an American stock market index) is likely to be stationary (see figure 2.2).



**Figure 2.2: S&P 500 P/E Ratio**

**Finite Sample Properties of Ordinary Least Squares Estimators**

In analysing time series data, we need to alter some of our assumptions in the standard OLS regression to take into account the fact that we no longer have the usual random sample of individual items.

***I. Linear in parameters***: the stochastic process *xt1,xt2, …, xtk, yt:* follows the linear model

*yt = β0 + β1xt1 + β2xt2+ …+ βnxtk + ut*

where t=1, 2, …, n; and n=the number of observations (number of time periods)

***II. Zero conditional mean***: for each t, the expected value of the error term (*ut*), given the explanatory variables for all time periods, is zero.

*E(ut / xti) = E(ut****/****xt1, xt2, …,xtk) = 0;* where *t=1, 2, …, n*

This assumption implies that the error term at time t is uncorrelated with each explanatory variable in every time period. If *ut i*s independent of *x’s* and *E(ut) = 0*, this assumption automatically holds.

***III. No Perfect collinearity***: like in cross-sectional regression, in the sample (in the underlying stochastic process), no independent variable is a perfect linear combination of another independent variable.

***Theorem 1:*** under assumptions I, II, III, in other words, if these three assumptions are satisfied, the Ordinary Least Squares Estimators (OLSEs) are unbiased: i.e. *E(i) = βi ; for all i=0, 1, …, k*

***IV.Homoscedasticity***: conditional on *x’s*, the variance of *ut*is the samefor all *t*.

*Var(ut /x)= Var(ut) =* σ*2 ;* where *t=1, 2, …, n*

If this does not hold true, the errors are heteroskedastic.

***V. No serial correlation***: conditional on *x’s*, the errors in two time periods are uncorrelated.

*Corr(utus /x) = Corr(utus) = 0; for all t ≠ s*

***Guass-Markov Theorem***: given the asumptions I through V, the OLSEs are BLUE.

***Hypothesis Testing***

In order to use the usual OLS standard errors in hypothesis testing, t-statistics and F-statistics, we need the normality assumption.

***VI. Normality of the error term***: the error *ut* is independently and identically distributed as normal with is zero mean and constant variance (*σ2).*

*ut = IIDN(0,* σ*2)*

When assumptions I through VI hold true, everything that applies to estimation and inference for cross-sectional regression applies directly to time series regressions. t-statistic tests the statistical significance of individual explanatory variables; whereas F-statistic tests joint significance.

***Nonstationary Stochastic Processes***

A **non-stationary** time series’ statistical properties like mean, variance etc will not be constant over time An example of a nonstationary time series is a series with a trend - something that grows over time, for instance. The sample mean and variance of such a series will grow as you increase the size of the sample.

Many economic and financial variables are nonstationary. Nominal GDP is one such. Below (in figure 2.3) is UK's GDP over the years. There is a trend as you can see.



**Figure 2.3: Annual GDP, in billion pounds**

Although our interest is in stationary time series, one often encounters nonstationary time series, the classic example being the **random walk model** (RWM). It is often said that asset prices, such as stock prices or exchange rates, follow a random walk; that is, they are nonstationary. We distinguish two types of random walks: (1) random walk without drift (i.e., no constant or intercept term) and (2) random walk with drift (i.e., a constant term is present).

**Random Walk without Drift:** Suppose *ut*is a white noise error term with mean 0 and variance *σ*2*.* Then the series *Yt*is said to be a random walk if

*Yt*= *Yt*−1 + *ut*

In the above random walk model, the value of *Y* at time *t* is equal to its value at time (*t* − 1) plus a random shock. We can think of random walk without a drift as a regression of *Y* at time *t* on its value lagged one period.

Now from *Yt*= *Yt*−1 + *ut*, we can write

*Y*1 = *Y*0 + *u*1

*Y*2 = *Y*1 + *u*2 = *Y*0 + *u*1 + *u*2

*Y*3 = *Y*2 + *u*3 = *Y*0 + *u*1 + *u*2 + *u*3

In general, if the process started at some time *0* with a value of *Y*0, we have

*Yt*= *Y*0 + t

*E(Yt) = E(Y0 + t)*

*= E(Y0) + E( t)*

*= Y0,* since*t) = 0*

When you calculate the variance, you will find that

*Var*(*Yt*) = *E(Yt – E(Yt))2*

*= E(Y0 +* t *– Y0)2*

*= E(t)2*

*= E(u1 + u2 + u3 + … + ut)2*

*= E[(u1 + u2 + u3 + … + ut)( u1 + u2 + u3 + … + ut)]*

*= E[(u1)2 + (u2)2 + (u3)2 + … + (ut)2 + (u1u2) + (u2u1) + (u1u3) +…+( uiuj)*

*= E(u1)2 + E(u2)2 + E(u3)2 +…+ E(ut)2 + E(iuj))*

*= σ2 + σ2 + σ2 +…+ σ2; since E(iuj) = 0 for all i ≠ j*

*= n2*

*= tσ2*

As the preceding expression shows, the mean of *Y* is equal to its initial or starting value, which is constant, but as *t* increases, its variance increases indefinitely, thus violating a condition of stationarity. In short, the RWM without drift is a nonstationary stochastic process. In practice *Y*0 is often set at zero, in which case *E*(*Yt*) = *0*.

An interesting feature of RWM is the *persistence of random shocks* (i.e., random errors). *Yt*is the sum of initial *Y*0 plus the sum of random shocks. As a result, the impact of a particular shock does not vanish. For example, if *u*2 = 2 rather than *u*2 = 0, then all *Yt*’s from *Y*2 onward will be 2 units higher and the effect of this shock never dies out. That is why random walk is said to have an *infinite memory.*

If we take the first difference of a random walk without a drift, we get

*(Yt - Yt-1) = ΔYt*

*= (Yt+ ut) -Y*t

=*ut*

where*Δ* is the first difference operator. It is easy to show that, while *Yt*is nonstationary, its first difference is stationary. In other words, the first differences of a random walk time series are stationary.

**Random Walk with Drift:** Let us modify the random walk without a drift a little bit as follows:

*Yt = α + Yt−1 + ut*

where*α* is known as the **drift parameter.** The name drift comes from the fact that if we write the preceding equation as

*Yt − Yt−1 = ΔYt = α + ut*

It shows that *Yt*drifts upward or downward, depending on *α* being positive or negative. Following the procedure discussed for random walk without drift, it can be shown that for the random walk with drift model:

*Y*1 = *α + Y*0 + *u*1

*Y*2 = *α + Y*1 + *u*2 = *α + α + Y*0 + *u*1 + *u*2

*Y*3 = *α + Y*2 + *u*3 = *α + α + α + Y*0 + *u*1 + *u*2 + *u*3

*Yt* = *t* ·*α + Y0+* t

*E*(*Yt*) = E(*t* ·*α + Y0 + t*)

= *Y*0 + *t* ·*α*

And the variance will be:

*Var*(*Yt*) = *E(Yt – E(Yt))2*

*= E(t* ·*α +Y0 +* t *– (t* ·*α +Y0))2 = E(*t*)2*

*= tσ2*

As you can see, for random walk model (RWM) with a drift, the mean as well as the variance increases over time. Again it violates the conditions of (weak) stationarity. In short, RWM, with or without drift, is a nonstationary stochastic process. Figure 2.4 may be an illustration of RWM with and without a drift. In the graph, RWM with a drift is slightly above the RWM without a drift which shows the drift is positive.



***Figure 2.4: Random Walk Models***

***Deterministic trend:*** if the stochastic trend is expressed as:

*yt = β0 + β1t + ut*

This is called a Trend Stationary Process (TSP).

*E(yt) = E(β0 + β1t + ut)*

*= E(β0) + E(β1t) + E(ut)*

*= β0 + β1t*

Although the mean of *yt* is not constant, its variance, indeed, is.

*Var(yt) = E(yt – μ)2 =*

*= E[(β0 + β1t + ut) –( β0 + β1t)]2 = E(ut)2 =* σ*2*

Once the values of *β0& β1*are known, the mean can be forecast perfectly since *μ = β0 + β1t*. Therefore, if we subtract the mean of *yt*from *yt,* the resulting series will be stationary. That is why, it is called trend stationary. This procedure of removing the trend is called ***detrending***.

## 2.3. Trend Stationary and Difference Stationary Stochastic Processes

Non-stationary data, as a rule, are unpredictable and cannot be modeled or forecasted. The results obtained by using non-stationary time series may be spurious in that they may indicate a relationship between two variables where one does not exist. In order to receive consistent, reliable results, the non-stationary data needs to be somehow transformed into stationary data. In contrast to the non-stationary process that has a variable variance and a mean that does not remain near, or returns to a long-run mean over time, the stationary process reverts around a constant long-term mean and has a constant variance independent of time.

The stationary stochastic process is a building block of many econometric time series models. Many observed time series, however, have empirical features that are inconsistent with the assumptions of stationarity.For example, the following plot shows quarterly U.S. GDP measured from 1947 to 2005. There is a very obvious upward trend in this series that one should incorporate into any model for the process.

****

**Figure 2.5: Quarterly U.S. GDP, 1947-2005**

Before we get to the point of transformation for the non-stationary financial time series data, we should distinguish between the different types of the non-stationary processes discussed above. This will provide us with a better understanding of the processes and allow us to apply the correct transformation. A trending mean is a common violation of stationarity. Examples of nonstationary processes are **stochastic trends** being either random walk with or without a drift (a slow steady change) and **deterministic trends** (trends that are constant, positive or negative, independent of time for the whole life of the series).

* *Trend stationary*: The mean trend is deterministic. Once the trend is estimated and removed from the data, the residual series is a stationary stochastic process.
* *Difference stationary*: The mean trend is stochastic. Differencing the series *d* times yields a stationary stochastic process.

The distinction between a deterministic and stochastic trend has important implications for the long-term behavior of a process:

* Time series with a deterministic trend always revert to the trend in the long run (the effects of shocks are eventually eliminated). Forecast intervals have constant width.
* Time series with a stochastic trend never recover from shocks to the system (the effects of shocks are permanent). Forecast intervals grow over time.

Unit root tests are a tool for assessing the presence of a stochastic trend in an observed series.

The distinction between stationary and nonstationary stochastic processes (or time series) has a crucial bearing on whether the trend is **deterministic** or **stochastic.** Broadly speaking, if the trend in a time series is completely predictable and not variable, we call it a deterministic trend, whereas if it is not predictable, we call it a stochastic trend.

***Difference Stationary***

A random walk with or without a drift can be transformed to a stationary process by differencing (subtracting *Yt-1*from *Yt*, taking the difference *Yt - Yt-1*) correspondingly to *Yt - Yt-1* = *εt*or *Yt - Yt-1*= *α + εt* and then the process becomes difference-stationary. The disadvantage of differencing is that the process loses one observation each time the difference is taken.



**Figure 2.6: Differencing**

***Trend Stationary***

|  |
| --- |
| A non-stationary process with a deterministic trend becomes stationary after removing the trend, or detrending. For example, *Yt* = *α + βt + ut* is transformed into a stationary process by subtracting the trend *βt*: *Yt - βt* = *α + ut*, as shown in the Figure below. No observation is lost when detrending is used to transform a non-stationary process to a stationary one. |



**Figure 2.7: Detrending**

* 1. **Integrated Stochastic Process**

Time series that can be made stationary by differencing is called integrated stochastic process. Recall that the RWM without a drift is nonstationary but its first difference is stationary. Thus we call RWM without a drift integrated of order 1, denoted as *yt ~ I (1).*

Similarly, if a time series has to be differenced twice to make it stationary, such a time series is called integrated of order 2, denoted as *yt ~ I (2).* In general, if a nonstationary time series has to be differenced d times to make it stationary, that time series is said to be integrated of order d, *yt ~ I (d).*

If a time series *yt* is stationary from the start, it is called integrated of order 0, *yt ~ I (0).* We often use the terms *‘stationary time series’* and *‘time series integrated of order zero’* to say the same thing.

***Properties of integrated series***

Let *xt, yt, &zt* be three time series:

1. If *xt ~ I(0)* and *yt ~ I(1)*, then *zt = (xt + yt)*is *I(1)*.

The sum of stationary and nonstationary time series is nonstationary.

1. If *xt ~ I(d)*, then *yt = (a + bxt) ~ I(d)*; where a and b are constants.

The linear combination of I(d) series is also I(d).

1. If *xt ~ I(d1)* and *yt ~ I(d2)*, then *zt = (axt + byt) ~ I(d1)*, where d1 > d2.
2. If *xt ~ I(d)* and*yt ~ I(d)*, then *zt = (axt + byt) ~ I(d’), where d’= d*, but sometimes *d’< d*.
	1. **Tests of Stationarity: The Unit Root Test**

Recall that stationary time series is what we most care about mainly because nonstationary time series gives spurious results. So the question is ‘*how do we know whether a given time series is stationary or not*?’ To find out the stationarity of a time series, it is always important and advisable to plot the time series under study graphically as a starting point of more formal tests of stationarity. There are several tests of stationarity. But we will focus on a test which has become popular in recent past, which is *the unit root test*.

***The Unit Root Test***

The starting point to unit root test is the following autoregressive process.

*yt = ρyt-1 + ut*

When *ρ=1*, we have a unit root, and thereby a RWM without a drift. The general essence behind the unit root test of stationarity is, therefore, to find out if the estimated rho(*ρ*) is statistically equal to one. In principle, we can run this regression (*yt= ρyt-1 + ut*) and see if *ρ=1*, but we cannot estimate model regressing the series on its lagged value to find out if the estimated *ρ=1* because in the presence of a unit root, the t-statistic for rho(*ρ*) coefficient is severely biased.

Therefore, we manipulate this equation (*yt = ρyt-1 + ut*) as follows:

*yt – yt-1 = ρyt-1 – yt-1 + ut*

*Δyt = (ρ-1)yt-1 + ut*

If we let (*ρ-1*) = *δ*, then

*Δyt = δyt-1 + ut*

Now it is a matter of testing if *δ* is zero or less than zero.

* If *δ=0, ρ-1=0* 🡺*ρ = 1; implying a unit root (nonstationarity)*
* *δ < 0, ρ-1<0* 🡺*ρ < 1; implying stationarity*
* we exclude a situation *δ > 0, ρ>1*

But the problem is that we cannot rely on the usual t-test on the significance of . It is because the null hypothesis is *δ=0 (i.e. ρ=1)*, and the t-value of the estimated coefficient of *yt-1*does not follow the t-distribution. The alternative is the Dicky-Fuller (DF) test.

Dicky and Fuller have shown that under the null hypothesis that *δ=0*, the standard t-value of the coefficient of *yt-1* follows *Ʈ (tau)* statistic. These authors have computed the critical values of the Ʈ-statistics. In principle, three specifications can be tried, depending on whether or not the series show a trend.

1. *Δyt = δyt-1+ ut*🡪 random walk (no drift, no trend)……………………..*eq. 1*
2. *Δyt = β0 + δyt-1+ ut*🡪 random walk (with drift, no trend) …………….*eq. 2*
3. *Δyt = β0 + β1t + δyt-1+ ut*🡪 random walk (with drift, with trend)…….*eq. 3*

Note that for each case, *H0: δ=0* (i.e. there is a unit root, and the series is nonstationary or it has a stochastic trend) against *H1: δ<0* (i.e. there is no unit root and the series is stationary, possibly around a deterministic trend).

If the null hypothesis is rejected, it means that *yt*is a stationary time series with zero mean in the case of *eq. 1*; that *yt*is stationary with a nonzero mean *[= β1****/*** *(1 −ρ)]* in the case of *eq. 2*; and that *yt*is stationary around a deterministic trend in *eq. 3*.

It is extremely important to note that the critical values of the tau test to test the hypothesis that *δ = 0*, are different for each of the preceding three specifications of the DF test. Moreover, if, say, specification in *eq. 2* is correct, but we estimate*eq. 1*, we will be committing a specification error. The same is true if we estimate *eq. 3* rather than the true *eq. 2*. Of course, there is no way of knowing whichspecification is correct to begin with. Some trial and error is inevitable, datamining, nonetheless. The actual estimation procedure is as follows: Estimate *eq. 1*, or *eq. 2*, or *eq. 2* by OLS; divide the estimated coefficient of *yt-1* in each case by its standard error to compute the (*τ*) tau statistic; and refer to the DF tables (or any statistical package).

***Stata Commands for DF-test***

1. dfuller*y*, noconstant regress
2. dfuller*y*, regress
3. dfuller*y*, trend regress

The results report a Mackinnon p-value.

* If the p-value is less than the significance level, reject the null hypothesis (*δ=0*).
* If the p-value is greater than the significance level, there is a unit root.

Or alternatively, check the tau-statistic of the lagged *yt-1* or its coefficient.

* If **|**computed Ʈ-statistic**|**>**|**critical Ʈ-value**|**, reject the null, (*δ=0*), hypothesis which implies the time series is stationary
* If the reverse is true, do not reject the null hypothesis which implies the time series is nonstationary.

**Critical/ table value of Ʈ-statistic in the three cases**

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1% | 5% | 10% |
| 1. *Δyt = δyt-1+ ut*
 | -2.59 | -1.94 | -1.62 |
| 1. *Δyt = β0 + δyt-1+ ut*
 | -3.51 | -2.89 | -2.58 |
| 1. *Δyt = β0 + β1t + δyt-1+ ut*
 | -4.07 | -3.46 | -3.16 |

**Example:** the DF-test results of time series variable, *yt*, is given as follows.

*Δŷt= -0.5yt-1 ………………………………………………… Eq. 4*

 *(-6.03)*

*Δŷt= 15 + 0.8yt-1……………………………………………. Eq. 5*

 *(0.09) (0.97)*

*Δŷt= 27 – 0.6yt-1 + 2.317t………………………………… Eq. 6*

*(15.03) (-4.78) (0.876)*

Our primary interest here is in the *τ* value of the *yt-1* coefficient. The critical, as given above, 1 percent, 5 percent, and 10 percent *τ* values for model (*Eq. 4)* are −2.59, −1.94, and −1.62, respectively; and are −3.51, −2.89, and −2.58 for model (*Eq. 5)*, respectively;and −4.07, −3.46, and −3.16 for model (*Eq. 6),* respectively.

As noted before, these critical values are different for the three models. Before we examine the results, we have to decide which of the three models may be appropriate. We should rule out model (*Eq. 5)* because the coefficient of *yt-1*, which is equal to *δ*is positive. Because *δ* = (*ρ* − 1), a positive *δ* would imply that *ρ >*1. Although it is a theoretical possibility, we rule this case out because in this case the time series *y* would be explosive. More technically, the so-called stability condition requires that |*ρ*| *<*1*.*

That leaves us with models (*Eq. 4*) and (*Eq. 6*). In both cases the estimated *δ* coefficient is negative, implying that the estimated *ρ* is less than 1. For these two models, the estimated *ρ* values are 0.5 (1 – 0.5) and 0.4 (1 – 0.6), respectively. The only question now is if these values are statistically significantly below 1 for us to declare that the *y* time series is stationary.

For model *(Eq. 4)* the estimated *τ* value is −6.03, which in absolute value is above even the 1 percent critical value of −2.59. Since, in absolute terms, the former is larger than the latter, our conclusion is that the *y* time series is stationary. The story is the same for model (*Eq. 6*). The computed *τ* value of −*4.78* is greater than even the 1 percent critical *τ* value of −*4.07* in absolute terms. Therefore, on the basis of the Dickey–Fuller test, the conclusion is that the given *y* time series does not contain a unit root.

# Introduction to Simultaneous Equation models

## The Nature of Simultaneous Equation Models

So far we were concerned exclusively with single equation models. In those models, we had a single dependent variable and one or more independent variables. The emphasis, in those kinds of models, has been on estimating and predicting the mean values of the dependent variable conditional upon the fixed values of the independent variables.

However, in many situations, we may not find such unidirectional cause-effect relationships meaningful. This is so if *y* is determined by the *x’s*, and some of the *x’s* are, in turn, determined by *y*. there is a two-way or simultaneous relationship between *y* and the *x’s*. In such models, there is more than one equation. Unlike the single equation models, one should not estimate the parameters of a single equation without considering the information provided by other equations in the system.

In particular, when a relationship is a part of a system, some of the explanatory variables will be stochastic and are correlated with disturbances. So the basic assumption of a linear regression model that the explanatory variable and the disturbance are uncorrelated or explanatory variables are fixed/ nonstochastic are violated. Consequently, OLS estimator becomes inconsistent as a result of endogeneity problem. In econometrics endogeneity problem arises when an explanatory variable is correlated with the error term. 🡺*Cov(xi, ui) ≠ 0*. Endogeneity may arise as a result of:

* Measurement error
* Autoregression with autocorrelated errors
* Simultaneous causality
* Omitted variables

Variables in simultaneous equation models are classified as endogenous and exogenous variables.***Endogenous variables*** are jointly determined variables explained by the functioning of the system. The values of endogenous variables are determined jointly by the simultaneous interaction of the relations in the system. On the other hand, ***exogenous variables*** are predetermined variables that contribute to explanations for the endogenous variables. The values of the exogenous variables are determined from outside the model.

***Example:*** consider the Keynesian model of how income is determined

*Yt = Ct + It + Gt*

*Ct = α0 + α1Yt + et*

Where *Yt& Ct*are endogenous variables; and *It + Gt* are exogenous variables.

What happens if one tries to estimate *α0& α1* by applying OLS without taking into account the information provided by the first equation? A change in *et* leads to a change in *Ct*. And since *Ct*affects *Yt*, the change in *Ct* means there is also a change in *Yt* in the first equation. That change in *Yt*, in turn, leads back into the second equation as *Yt* is a predictor for *Ct* causing *Ct* to change. In the end, there is a correlation between *Yt* (the explanatory variable) and *et* (the error term). Mathematically, *Cov(Yt, et) ≠ 0*.

* 1. **Simultaneity bias**

Simultaneity bias is a term for biased results of OLSE that happen when the explanatory variable is correlated with the error term because of simultaneity. Simultaneity is one type of endogeneity problem in which one or more explanatory variables are jointly determined with the dependent variable.

*Consider a structural model:*

*y1 = a1 + b1y2 + c1z1 + e1*

*y2 = a2 + b2y1 + c2z2 + e2*

While *z1*&*z2*are exogenous =>*Cov(z1,e1)=0*  and *Cov(z2,e2 )=0; y1*&*y2* are endogenous.

If we insert the first equation into the second equation, we get a **reduced form** for *y2*.

*y2 = a2 + b2(a1 + b1y2 + c1z1 + e1) + c2z2 + e2*

*=a2 + b2a1 + b2b1y2 + b2 c1z1 + b2e1 + c2z2 + e2*

*y2 – b1b2y2= a2 + a1b2 + b2 c1z1 + b2e1 + c2z2 + e2*

*(1 – b1b2)y2= (a2 + a1b2) + b2 c1z1 + c2z2+ (b2e1+ e2)*



Let then we obtain

*y2 = 1 + 2z1+ 3z2 + u2*

Because*u2*is a linear function of *e1& e2*, *u1& u2* are uncorrelated with *z1& z2*. Therefore, OLS estimators are consistent for the **reduced forms** of *y1& y2*. However, if we were to estimate the structural equations by OLS, the estimation would suffer from simultaneity bias.

Simultaneity bias is simply a term for the biased and inconsistent results of OLS estimators which occur when the explanatory variable is correlated with the error term because of simultaneity. Thus, we can see that estimating structural equation in a simultaneous equations system by OLS results in biased and inconsistent estimators. We can solve this problem by using ***Indirect Least Squares (ILS)*** estimator or ***Two-Stages Least Squares (2SLS)*** estimator. As we specify the structural equation for each endogenous variable, we can immediately see if there are sufficient instrumental variables in the equation. We call this process an **identification problem.**

* 1. **Order and rank conditions of identification (without proof)**

By the identification problem, we mean whether or not numerical estimates of the parameters of a structural equation can be obtained from the estimated *reduced-form* parameters. If this can be done, we say that particular equation is identified. If not, the structural equation is unidentified or alternatively, under-identified.

An identified equation may be either **exactly/just identified** or **over-identified**. It is said to be just identified if unique numerical values of the structural equation could be obtainedfrom the estimated reduced-form coefficients.It is said to be over-identified if more than one numerical value can be obtained for some of the parametersof the structural equations.

We have two ways that help us decide whether or not a given equation in a system is identified; ***Order condition & Rank condition.*** Let us use the following notations:

* M = the number of endogenous variables in the model
* m = the number of endogenous variables in a given equation
* K = the number of exogenous variables in the model
* k = the number of exogenous variables in a given equation
1. ***The order condition***: it is considered to bea necessary but not sufficient, condition for identification.

*Definition:*in a model of M simultaneous equations, in order for an equation to be identified, the equation must exclude at least M-1 variables (endogenous as well as exogenous) appearing in the model. If it excludes exactly M-1 variables, it is just identified. If it excludes more than M-1 variables, it is over-identified.

*Definition:* or alternatively, in a model of *M* simultaneous equations, in order for an equation to be identified, the numberof predetermined (exogenous) variables excluded from the equation must not be less than the number ofendogenous variables included in that equation less 1, that is,*K* − *k* ≥ *m*− 1.

If *K* − *k* = *m* − 1, the equation is just identified, but if *K* − *k > m* − 1, it is over-identified.

***Examples:***in the following examples, price (P) and quantity (Q) are considered to be endogenous variables.

1. *Demand function: Qt*= *α*0+ *α*1*Pt*+ *u*1*t*

*Supply function: Qt*= *β*0+ *β*1*Pt* + *u*2*t*

This model has two endogenous variables *P* and *Q* and no predetermined variables. To beidentified, each of these equations must exclude at least *M* − 1 = 1 variable. Since this is notthe case, neither equation is identified.

1. *Demand function: Qt = α0+ α1Pt + α2It + u1t*

*Supply function: Qt = β0+ β1Pt + u2t*

In this model *Q* and *P* are endogenous and *I*is exogenous. In the order condition given above, we see that the demand function is unidentified. On the other hand, the supplyfunction is just identified because it excludes exactly *M − 1 = 1* variable i.e. *It*.

1. *Demand function: Qt = α0+ α1Pt + α2It + u1t*

*Supply function: Qt= β0+ β1Pt + β2Pt−1 + u2t*

Given that *Pt* and *Qt* are endogenous and *It*and *Pt−1*are predetermined, the demand equation excludes exactly one variable *Pt−1*and the supply equation also excludes exactly one variable *It*.Hence each equation is identified by the order condition. Therefore, the model as a whole isidentified.

1. ***The rank condition***:

In a model containing M equations in M endogenous variables, an equation is identified if and only if at least one nonzero determinant of order *(M − 1)(M − 1)* can be constructed from the coefficients of the variables (both endogenous and predetermined) excluded from that particularequation but included in the other equations of the model.

***Example:*** consider the following system of equations.

*y1 = a0 + a1y3 + a2x1 + a3x3 + e1*

*y2 = b0 + b1y1 + b2x2 + b3x3 + e2*

*y3 = c0 + c1y2 + c2x1 + c3x2 + e3*

The system contains three endogenous variables (*y1*, *y2*, *y3*) and three exogenous variables (*x1, x2, x3*).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | *y1* | *y2* | *y3* | *x1* | *x2* | *x3* |
| Equation 1 | *1* | *0* | *1* | *1* | *0* | *1* |
| Equation 2 | *1* | ***1*** | *0* | *0* | ***1*** | *1* |
| Equation 3 | *0* | ***1*** | *1* | *1* | ***1*** | *0* |

In order to check the rank condition for the first equation;

* Delete the first row
* Collect the columns for those variables of the first equation that are marked with zero (*y2&x2*)
* So we have left a matrix of *2x2, (M-1)(M-1),*

Since the determinant of the 2 by 2 matrix is non-vanishing (nonzero), the first equation is identified. Note that number 1 in the matrix is a symbol for nonzero coefficients in the model.Because if so, the determinant would have been zero. If we follow the same strategy, we will find that equations 2 and 3are also identified.

Generally, to test the rank condition, we may follow the steps as follows.

1. List all the endogenous and exogenous variables in a tabular form. And for each equation, give a code of 1 for included and code of 0 for excluded variables in the equation under consideration.
2. Delete the row (equation) under consideration
3. Strike out the columns corresponding to the 1’s of the equation under consideration
4. Now, the entries left in the table are only the coefficients of the variables included in the system but not in the equation under consideration.
5. If the remaining *(M-1) by (M-1)* matrix has a full rank, then the equation is identified.

***Example:***consider the following system. (*y1, y2, y3, y4*) are endogenous & (*x1, x2, x3*) are exogenous variables.

*y1 = a1 + b1y2 + c1y3 + d1x1 + e1*

*y2 = a2 + b2y3 + c2x1 + d2x2 + e2*

*y3 = a3 + b3y1 + c3x1 + d3x2 + e3*

*y4 = a4 + b4y1 + c4y2 + d4x3 + e4*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Step 1:** |  | *y1* | *y2* | *y3,* | *y4* | *x1* | *x2* | *x3* |
|  | *Equation 1* | *1* | *1* | *1* | *0* | *1* | *0* | *0* |
|  | *Equation 2* | *0* | *1* | *1* | *0* | *1* | *1* | *0* |
|  | *Equation 3* | *1* | *0* | *1* | *0* | *1* | *1* | *0* |
|  | *Equation 4* | *1* | *1* | *0* | *1* | *0* | *0* | *1* |

**Step 2:**if we want to identify the first equation, we strike out the first row.

**Step 3;**we strike out the columns if they appear to be included in the first equation (1’s)

**Step 4:**now we are left with

**Step 5:** we have *(M-1) by (M-1)* i.e. *3 by 3* matrix. Unfortunately, the determinant of this matrix is vanishing or zero. *Why?* Thus, we conclude that the first equation is unidentified.

We can follow the same procedure to find out whether or not the other equations in the system are identified.

***Summary of Order and Rank Conditions***

1. If *K – k > m – 1*, and the rank of the matrix is *M-1*, the equation is over-identified.
2. If *K – k = m – 1*, and the rank of the matrix is *M-1*, the equation is exactly identified.
3. If *K – k* ≥*m – 1*, and the rank of the matrix is less than*M-1*, the equation is not identified.
4. If *K – k < m – 1*, the equation is not identified.
	1. **Indirect squares and 2SLS estimation of structural equations**
	2. ***The Method of Indirect Least Squares (ILS)***

It is a method of deriving the estimates of the structural parameters from the OLS estimates of the reduced-form parameters. It applies for exactly identified equations. ILS is done in three steps.

*Step 1*: form the reduced form equation(s)

Reduced form equations are derived from structural equations in such a way that, in each equation, the dependent variable (on the left-hand side) is the only endogenous variable in the equation.

*Step 2*: estimate the coefficients of the reduced form equation(s) by applying OLS

*Step 3*: use the estimated coefficients of the reduced form to derive the structural coefficients

**Example:** consider the following model where *Yt*&*Ct* are endogenous and *It* is exogenous variables.

*Yt= Ct + It*

*Ct= α1 + α2Yt +et*

We would like to estimate *α1*and *α2*. The consumption function is just identified. *Why?*

*Step 1*: form the reduced form equation(s)

*Ct = α1 + α2(Ct + It) +et*

*Ct = α1 + α2Ct + α2It +et*

(1 - *α2)Ct = α1+α2It +et*

*Ct = β1 + β2It+ut ; where β1 = α1****/***(1 - *α2) &β2 = α2* ***/***(1 - *α2) &ut = et* ***/***(1 - *α2)*

*Step 2*: estimate the coefficients of the reduced form equation(s) by applying OLS

Assume the following regression result. *Ĉt = -631.8 +4It*

*t-statistics(-6.34) (3.72)*

*Step 3*: use the estimated coefficients of the reduced form to derive the structural coefficients



* 1. ***The Method of Two Stages Least Squares (2SLS)***

The 2SLS method is specifically designed for an over-identified equation, even though it can also be applied to just identified equations. The idea behind this method is to replace the endogenous explanatory variable by a proxy variable that satisfies the following two properties.

1. The new proxy variable is strongly correlated with the endogenous explanatory variable
2. The new proxy variable is not correlated with the error term in the equation

Such a proxy variable is known as the *instrumental variable*. The 2SLS estimation is done this way. First find the instrumental variable, and then find the estimators. We do OLS regression twice.

***Stage 1***: Regress the endogenous explanatory variable on all exogenous variables which are included in the system not only in the equation. We obtain the predicted values of the endogenous explanatory variable.

***Stage 2***: We apply OLS on the structural equation by replacing the estimated endogenous explanatory variable in the place of the endogenous explanatory variable.

**Example**:

*y1 = a0 + a1y2 + a2x1 + a3x2 + e1*

*y2 = b0 + b1y1 + b2x3 + e2*

If we are interested only in the second equation, you can see that it is over-identified. *Why?* So we should use 2SLS to estimate *b0, b1& b2*. Since *y1*is an endogenous explanatory variable in the second equation, we first have to find an instrument for *y1*.

***Stage 1***: we regress *y1*on *x1, x2, x3*by using OLS, and we get the predicted values of *y1, i.e. ŷ1*

*ŷ1 = ĉ0 + ĉ1x1 + ĉ2x2 + ĉ3x3*

***Stage 2***: replace *y1,* by*ŷ1*, and estimate *b0, b1& b2* by regressing *y2* on *ŷ1*&*x3*as in the original model

**Note** the following points in ILS and 2SLS.

* + 1. 2SLS can be applied to an individual equation without considering other equations in the model
		2. Unlike ILS method, 2SLS provides unique solution for each parameter for over-identified equation
		3. 2SLS can be applied to just identified equations as well
		4. Both are easy to apply
		5. The standard errors of 2SLS (and ILS) are larger as compared to OLS => inefficiency
		6. 2SLS or ILS are not unbiased in small samples. But OLS is inconsistent in simultaneous equation models

# Introduction to Panel Data Regression Models

## Introduction

Panel data, also known as cross-sectional time series data, is a data set in which the behaviour of entities is observed over time. These entities/units could be people, companies, countries, schools, regions, etc.

**Cross-sectional data Time series data Panel data**

 **N > 1, T = 1 N = 1, T > 1 N > 1, T > 1**

Where N is the number of cross-sectional units/entities and T is the number of time periods over which the data set is collected.

Regression models based on panel data are called *panel data regression models*. Panel data could either be balanced or unbalanced. In balanced panel data, each cross-sectional unit/entity is observed in each time period. In unbalanced panel data, one or more entities are not observed in one or more time periods. While total observation in balanced panel is *N\*T*, it is less than *N\*T* in unbalanced panel data. It is because, in unbalanced panel data, one or more units are missing in one or more time periodsfor different reasons. Unbalanced panel data may be due to:

* Bankruptcies if the entities under study are companies or investors
* Moving from one place to another
* Death if the study is about people
* Refusal of respondents for further cooperation, etc

***Advantages of Panel Data***

1. Explicit modelling of the effect of time and heterogeneity in cross-sectional units. Panel data estimation techniques solve these problems by allowing time dummies, αt, and unit specific effect variables, αi, into the model explicitly.
2. Potential missing variable bias is minimized as the model includes unit specific effects, αi.
3. Inclusion of time series and cross-sectional elements in panel data generates more informative data and more variability which results in smaller effects of multicollinearity, more degrees of freedom and more efficiency of estimators.
4. It allows for analysing the dynamics of change for each unit on individual level.
5. It enables us to study more complex models that can grasp the effects of abstract variables like technological progress.
6. It controls unobserved variables. For example, difference in business practices across companies.

However, it does not mean panel data analysis is always advantageous. It has its own drawbacks such as data collection issues (like sampling design) and non-response.

***Example of panel data***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Entity***  | ***Year***  | ***Y*** | ***X1*** | ***X2*** |
| 1 | 1999 | 10.5 | 20 | 2.5 |
| 1 | 1999 | 9 | 18 | 3.2 |
| 1 | 1999 | 7 | 25 | 4.1 |
| 2 | 2000 | 9.6 | 16 | 6.5 |
| 2 | 2000 | 12 | 19 | 5.5 |
| 2 | 2000 | 9 | 22.5 | 4.2 |
| 3 | 2001 | 6.5 | 19.5 | 3.1 |
| 3 | 2001 | 8.7 | 18 | 3.5 |
| 3 | 2001 | 9.5 | 21 | 4 |
| . | . | . | . | . |
| . | . | . | . | . |

Generally, a matrix of balanced panel data observation for variable, say Y, would be as follows

Where N is cross-sectional entities; and T is time periods over which entities are observed.

Suppose

* *y* is investment *x* is profit
* N companies observed in T time periods
* A simple econometric model that says investment depends on profit will look like

*yit = a1 + a2xit + uit*

*Wherei = 1, 2, …, N; t = 1, 2, …, T; and uit*is a random error term i.e. *E(uit) ~ (0, σ2)*

Estimation of parameters in panel data, as in the above, depends on the assumption we make about *a1, a2, and uit.* Several possible assumptions could be made to estimate the above equation, each assumption resulting in different levels of biasedness, consistency and efficiency.

1. Assume that the intercept (*a1*) and slope coefficient (*a2*)are constant acrosstime and space (across individual firms) and the error term captures differences over time andspace.
2. The slope coefficient (*a2*) is constant but the intercept (*a1*) varies across space (across firms), but not over time.
3. The slope coefficient (*a2*) is constant but the intercept(*a1*) varies over individual firmsandtime.
4. All parameters (the intercept as well as slope coefficient) vary over space, but not over time.
5. All the parameters (the intercept as well as slope coefficient) vary over individuals andtime.

As you can see, each of these cases introduces increasing complexity (and perhaps more reality) in estimating panel data regression models, such as the models given above. Of course, the complexity will increase if we add more regressors to the model because of the possibility of collinearity among the regressors.

 **All Coefficients Constant across Time and Entities**

The simplest approach is to disregard the space andtime dimensions of the pooled data and just estimate the usual OLS regression. Suppose there are twenty companies in four time periods. Stack the 20 observations for each company one on top of theother, thus giving in all 80 observations for each of the variables in themodel. If you examine the results of the **pooled regression**, the estimated model assumes that the intercept value for all companies is the same. It also assumesthat the slope coefficients of*x* variables are all identical for allthe four firms. *yŷit = â1 + â2xit (****pooled OLS regression****)*

Obviously, these are highly restricted assumptions. The figure suggests that a better way to model the data is to allow each company to have its own intercept.Therefore,despite its simplicity, a pooled *ŷ1t = â11 + â12xit* regression may distort thetrue picture of the *ŷ2t = â21 + â22xit*relationship between *y*and the *x*’s across the four companies.What we need to do is find some *ŷ3t = â31 + â32xit*way to take into account the specificnature of the *ŷ4t = â41 + â42xit*four companies. *x*

Thus, rather than considering a model of pooled regression (*yit = a1 + a2xit + eit*), we may allow each entity, each company in the previous example to have its own intercept and model (*yit = a1i + a2xit + eit*). This method of estimation is known as *Fixed Effects Model (FEM).*

* 1. ***Estimation of Panel Data Regression Model: The Fixed Effects Approach***

**Slope Coefficients Constant but the Intercept Varies acrossIndividuals:**

*yit = a1i + a2xit + eit*

Notice that we have put the subscript *i* on the intercept term to suggest thatthe intercepts of the four firms may be different; the differences may be dueto special features of each company, such as managerial style or managerialphilosophy.

In the literature, such a model is known as the **fixed effects model** (**FEM**). The fixed effects model is sometimes known as Least Squares Dummy Variable (LSDV) model. The term “fixed effects” is due to the fact that, although theintercept may differ across individuals (here across the four companies), each individual’sintercept does not vary over time; that is, it is *time invariant*. Noticethat if we were to write the intercept as *a1it*, it will suggest that the interceptof each company or individual is *time variant.* It may be noted that the FEM given above assumes that the slope coefficient, *a2*, of the regressors do notvary across individuals or over time.

FEM is used when you are interested only in analyzing the impact of independent variables that vary over time. FEM explores the relationship between the outcome variable and its predictor within entity.Each entity has its own individual characteristics that may or may not influence the predictors (independent variables).

When using fixed model, we assume that time-invariant variables characteristics within the entity may bias the effect of the predictor on the outcome variable and we need to control for this. This is the rationale behind the fundamental assumption of the correlation between entity’s error term and predictors. Fixed effects model removes the effect of those time-invariant characteristics so we can assess the net effect of the predictors on the outcome variable.

Another important assumption of FEM is that those time-invariant characteristics are unique to the individual entities and should not be correlated with other entity’s characteristics. Each entity is different, therefore, the entity’s error term and the constant, which captures individual characteristics, should not be correlated with other entities.

**Note** that the key insight in FEM is that if the unobserved variables do not change over time, then any change in the dependent variable must be due to influences other than the fixed characteristics. Thus, the interpretation of coefficients would be “for a given entity, as the independent variable changes by one unit, the dependent variable increases/decreases by the coefficient amount”.

FEM has two major drawbacks. It includes enormous amount of dummy variables one for each entity. As a result, if there are one hundred units to be studied, we lose ninety nine degrees of freedom. The second one is that the transformation process wipes out all time invariant explanatory variables from the model. For example, for variables like gender, location, etc, we cannot estimate slope coefficients.

* 1. ***Estimation of Panel Data Regression Model: The Random Effects Approach***

The rationale behind Random Effects Model (REM) is that unlike FEM, the variation across entities is assumed to be random and uncorrelated with the explanatory variables included in the model.

Let us start from fixed effects model

*yit = a1i + a2xit + eit*

Instead of assuming *a1i*as fixed for each entity, we treat it as a random variable with a mean of *a1*. And therefore,

*a1i = a1 + λi ; i = 1, 2, …, N;* where *λi* is a random error term (*λi~ IIDN(0,σ2))*

Therefore, our random effects model becomes

*yit = a1 + a2xit + uit*where *uit = eit + λi*

* *eit –* the combined time-series and cross-sectional error term
* *λi–*the cross-section or unit-specific error component

The advantages of REM are:

1. We can include time-invariant variables. Remember, in FEM, these variables are absorbed by the intercept.
2. By saving more degrees of freedom, REM produces more efficient estimators than FEM.

Unlike FEM, unit specific effects should not be correlated with the predictors.

***Fixed effects model versus random effects model***

* FEM is often preferred for the simple reason that correlation between the unit specific effect, *ai*, and the predictors is common
* FEM cannot be used if none of the predictors/covariates/explanatory variables varies over time within entities. *Why?*
* REM is preferred if there are small number of observations for each entity as fixed effects consumes many degrees of freedom
* Generally, if N is large and T is small, REM is preferred; and if N is small and T is large, then FE model is preferred in many situations.