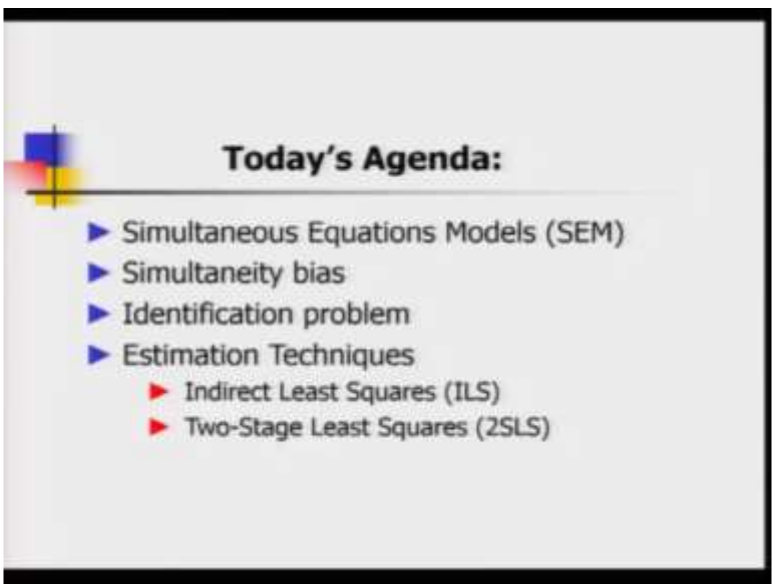
Applied Statistics and Econometrics
Professor Deep Mukherjee
Department of Economic Sciences
Indian Institute of Technology, Kanpur
Lecture 37
Simultaneous Equations Model

Hello friends. Welcome back to the lecture series on Applied Statistics and Econometrics. So, we are in the second last week of the course and this week also, I am going to dedicate three lectures to three different topics in applied econometrics. So of course you can understand that we are not going to get into a lot of details or derivation of econometric methods here.

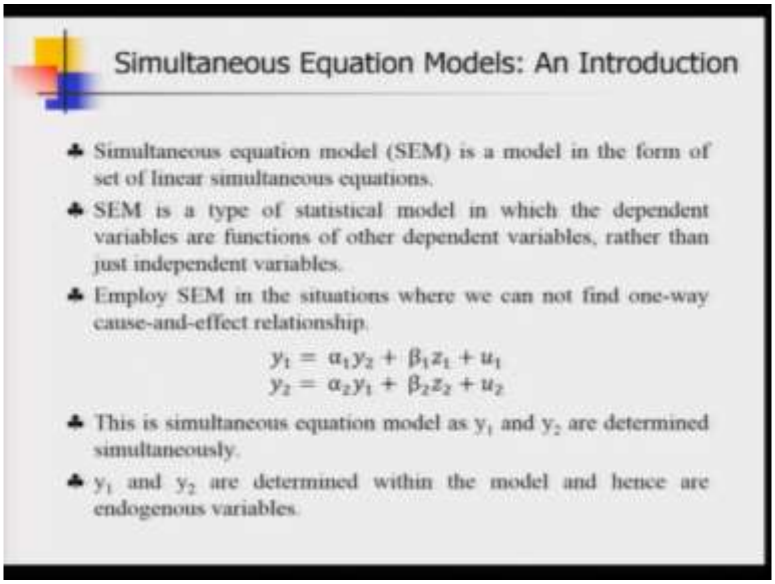
Because needless to say in 50-55 minutes lecture I cannot do justice to this huge fields that I am going to introduce to you but my idea is to show you that whatever we have learnt so far in this course, how they can be slightly extended to address very interesting problems that we find around us. So today's lecture we are going to talk about such an interesting advanced topic and it is called Simultaneous Equation Systems.

(Refer Slide Time: 1:16)



So let us first look at today's agenda items. So we are going to briefly define what Simultaneous Equations Models are all about. Then I am going to talk about the related concepts like simultaneity bias and identification problem and then I am going to talk about two estimation techniques which are useful to solve simultaneous equations models because in this type of special models our good old friend OLS does not help us.

(Refer Slide Time: 01:49)



So we are going to start with the formal definition and Simultaneous Equation Model which is also abbreviated as ACM is a kind of model in the form of a set of linear simultaneous equations. Now what do I mean by linear simultaneous equations? So the second bullet point is going to be on formal definition again but from a different perspective. So let me know at this point of explain what do I mean by linear simultaneous equation.

So linear equations of course that we are writing equations where parameters and variables are all expressed in linear manner and simultaneous means that there are variables which are appearing in almost all equations. So that is why there is kind of a relationship between all the variables in the equation system or the model.

So, if we look at the definition from another alternative angle, then we can define ACM as a type of statistical model in which the dependent variables are functions of other dependent variables in the model rather than just independent variables. So you see the second definition actually is pointing to this first definition where I told you that we are talking about linear simultaneous equations, basically simultaneity comes because the dependent variable in one equation becomes independent variable in the other equation.

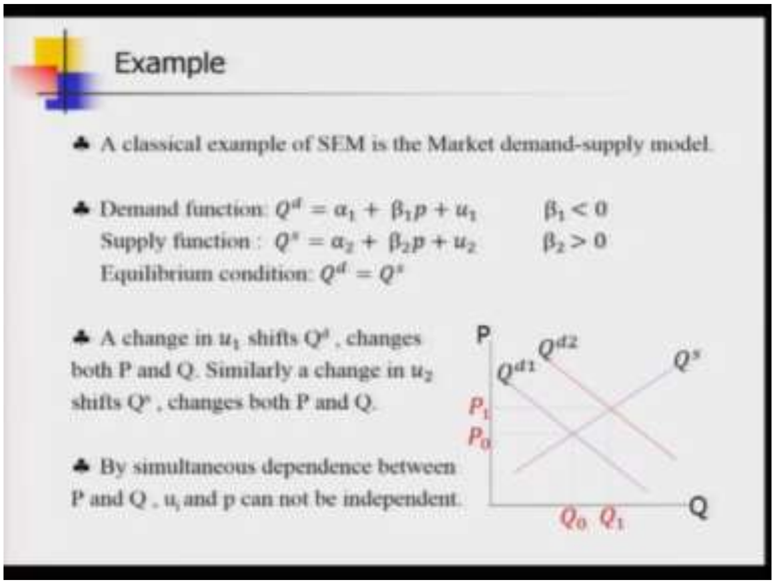
So basically the variables change their roles from one equation to the other, so in some equation a particular variable may be dependent variable but the same variable could be independent variable in another equation. So, we have to apply this simultaneous equation models in the situations where we cannot find one way cause and effect relationship.

So this point about simultaneous equation model becomes much more clear when we look at the equations at the bottom, in fact, these two equations are going to help us to understand the first two bullet points in this slide as well. So the first equation y1 is defined over another variable y2 and z1 and of course there is an error term u1. Now note that generally we have preserved the symbol y to denote the dependent variable but now this time we see for the first time this y symbol is being used in both sides of the regression equation.

So what is the point? The point is going to be clear as we look into the second equation. So you see the second equation says that y2 now is a function of the variable y1 and z2 and of course it is a linear equation and there is an error component u2. Now you see that y2 that I had as the independent explanatory variable in equation number 1 has become the dependent variable of equation number 2 in my system of equations.

So, basically what happens that it changes its role so that is why I am denoting y1 and y2 to denote some variables which can change their roles across equations. So here in this model you see the truly exogenous variables are also there in the equation and they are actually z1 and z2. So this above simultaneous equation model has y1 and y2 as two endogenous variables. Why they are called endogenous? Because they are going to be determined jointly or simultaneously in this system of equations framework.

(Refer Slide Time: 6:01)



So, here I give you a classic example of simultaneous equations modeling and this time I am going to borrow the example again from microeconomic literature. So those of you who are from economics background have heard of the market mechanism and we all know how the perfectly competitive market works, right. So here in this slide I am showing you the nuts and bolts of a simple market mechanism.

So we have a demand function and here I say that quantity demanded is a function of price of the commodity and quantity demanded is a linear function of the price in the market and then we have supply function that is denoted by Qs and that is also a linear equation and quantity of supply actually is affected or determined by the market price of the commodity for that time period.

So this is the way you have a demand and supply function and the most important part is the equilibrium part and how in a market or to be very specific in a perfectly competitive market, price is determined. How equilibrium is obtained? So to know that actually we have to go back to the year 1776, when a professor from university of Glasgow named Adam Smith has published a book called you know Wealth of Nations and that is considered to be the starting point of this field Modern Economics what we call Modern Economics today and Adam Smith in that book actually talked about in that book he talked about a mechanism for market solution to set price for a particular commodity.

So he said that in a perfectly competitive market, we have infinite number of buyers and infinite number of sellers, so actually we may not have infinite number of buyers and sellers but what he meant was that you have a very large pool of buyers and you have a large pool of sellers. So not an individual has enough power to or in that case actually he said that in its bargaining power, so it is not possible for an individual to have a significant bargaining power so that he or she can individually tweak the market price or influence the market price.

So that is what we mean by infinite number of buyers and sellers. So if that condition holds in market then we can call that it is a perfectly competitive market, so you also assume that there is no intervention from the government side so no tax no subsidy. So of course way back in 1776 the economies were much more simple compared to as of today.

Now Adam Smith said that in the market, there are two invisible hands are interacting with each other and these are the invisible hands of the market and if demand is the left hand, then supply is the right hand. So at the intersection point of these two functions demand functions and supply functions, the market equilibrium is obtained where there is one equilibrium price obtained in the market and at this equilibrium market price market actually clears.

So what do I mean by market clear? So market clears means that at that particular equilibrium price p star, you can call it in the quantity demanded will be equal to quantity supplied. So, here we can see that that Adam Smith's famous invisible hand principle for market mechanism is represented through simple diagram where we have Q as the quantity bought and sold in the market measured along the horizontal axis and then we have market price measured along the vertical axis and then you see that suppose Qd1 is basically the demand function for time period 1 and Qs is my supply function, so you see they intersect at a point for which you get the price P not and quantity Q not.

So you can say that P not Q not are basically the equilibrium market price and equilibrium quantity which will be bought and sold in the market because P not is that price where the market demand is exactly equal to the market supply. So basically you see this is the crux of the market model. Now we are going to come to econometric, that was micro whatever we have discussed so far. Now suppose you have got data on a particular commodity’s purchase in real markets and you also know at what price this good was or this particular commodity was sold.

So suppose you have a time series data, okay and suppose you observe data for 20-30 time periods, they could be weeks they could be months. So you have that data with you and suppose you are interested to find out the demand function for your firm because if you are a manager of a firm you would like to have some idea about the demand function because supply is in your control right, so if you wish tomorrow if you want to produce 10 units more, then probably you can do it but demand actually is not in your control.

So you have to get a fair idea about the demand you can look for demand projection etc. so for various purposes the form of demand function should be known to the firm or a company which is supplying goods to the market. So you collected data on the market prices and the sales of a particular commodity and you plotted a scattered diagram and suppose you get a scatter plot which is showing some kind of a negative correlation between these two variables.

So, of course, you are very happy so you now apply your OLS to get a fitted line where you claim that well my sum of square residuals are minimized when I am fitting in a straight line in this scatter diagram and I get downward sloping straight line and if you conduct hypothesis testing suppose your t test also tells you that your slope coefficient is significantly negative and also is there any reason to be very happy that you have identified or discovered the market demand function for your commodity? The answer is no.

Why? Because this price and quantity data that you are observing in that scatter diagram and ultimately you are trying to fit a line through it, they are actually the equilibrium prices and the equilibrium quantities at which the market got cleared at a particular time period. So if you remember that diagram that I showed you where this invisible hands of Adam Smith demand and supply functions intersect each other at price P not and quantity Q not, so that was the basically price at which market cleared for one particular time point.

Now if there are 30 or 40 time points on which you have got the data, so what you observe is basically 30-40 points which are this equilibrium point. So each time when market opened, market closed with an equilibrium price and you observe that price and at that price what is the amount of purchase and sale in a particular time period.

So basically you are observing the intersection points of the demand function and the supply functions at different time period you are not actually getting the data from the demand function or the supply function, so you see the problem here. So you are just lucky that your scatter gifted you scattered a pattern between this p star, q star numbers and you got a downward sloping straight line and you were very happy that I identified my demand function but not really, you haven't identified your demand function because these individual points are basically the equilibrium market prices and the equilibrium quantity at which market got cleared.

So how do you now extract demand function from this scatter plot or the data set that you have? Can we do something or it is just beyond our caliber or our capacity? No there is a solution to this particular problem, if this problem is called identification problem there is a solution to it and we are going to discuss the solution of this kind of problem in today's lecture.

So let us go back to that diagram again, so you see there could be a change in the random variable u1 in a different time period that can shift my quantity demanded function from time period one to time period two. So now you see there is a parallel shift in the purple color downward sloping demand function Qd1 to red color straight line downward sloping demand function Qd2 and if that happens then you see that as supply function did not change, there is another new equilibrium restored in the market and this time the new equilibrium price is P1 and new quantity bottom sold in the market is Q1.

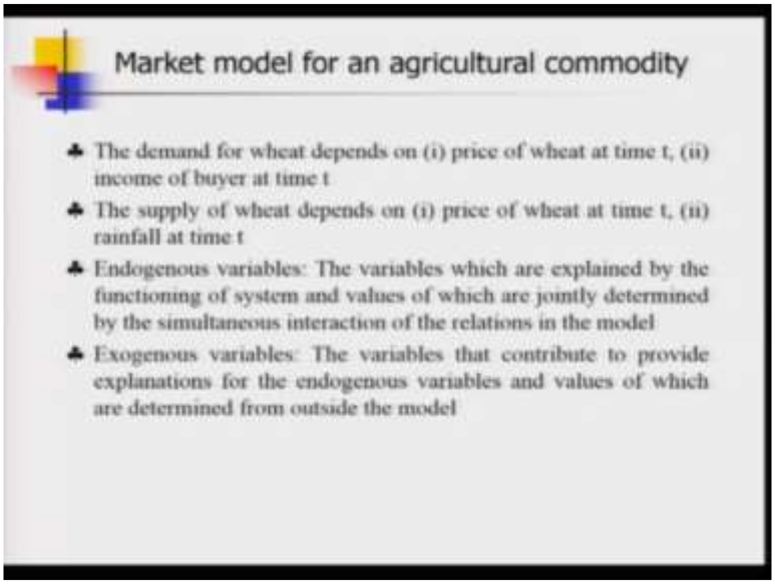
So you see as there is a shift in the demand function, then basically you got another equilibrium point such that now if you compare the first equilibrium point and the second equilibrium point, if you can assume that by this time the supply function did not change anywhere, so then you actually got two data points on the supply function. So basically if you now join these two points by a straight line and get an equation then that equation actually will give you the supply function.

So the idea is simple, so in the market system the price and the quantity these are simultaneously determined through a simultaneous equation system where individual functions are demand functions and the supply functions or individual equations are the demand equation and the supply equations and if you want to recover one particular equation from the data or graph whatever you call, then you have to shift the other equation in the question or in the model so that you can identify at least one equation.

So the crux or in a nutshell, the main point is that by simultaneous dependence between price and quantity for a commodity this random errors ui and the price p cannot remain independent. So you see if there is a correlation between the market price which is the explanatory variable in your regression equation and the error term ui, then the exogenetic condition breaks down.

That is one of the major assumptions before you start thinking about running OLS and this is an assumption that you must make that your explanatory variable is exogenous, so there is no correlation between it and the error term but here you see as there is simultaneity actually this exogenetic condition breaks down. So our good old friend OLS is not of any use here, so we have to now come up with an alternative estimation method.

(Refer Slide Time: 19:36)



So now in this slide, we are going to develop further using that market model story that I introduced in the last slide only. So here let us talk about an agricultural commodity, to be very specific. So here we assume that the demand for wheat depends on price of wheat at time period t and income of buyer at time period t. And we can also assume that supply of wheat depends on price of wheat at time period t and rainfall at time period t.

So here in this context let us in look at the endogenous variable and the exogenous variable, so what is an endogenous variable? I already defined this concept before but if you have forgotten, let me repeat it will not take even a minute. So these are the variables which are explained by the functioning of the system and values of these variables are jointly determined by the simultaneous interaction of the relations in the model and what is endogenous variable, so these are the variables that contribute to provide explanations for the indigenous variables and values of these type of variables are determined outside the model.

So they are truly given to you, they are not going to be determined by the model that you are looking at. So now in this market story for an agricultural commodity, what could be the endogenous and what could be the exogenous variables? So needless to say from the previous lecture you can anyway say that price of wheat and quantity bought and sold in the market, they are going to be endogenous variables yes and do not forget the demand and supply also. So what could be the exogenous variables in this model?

So if you look at the assumption for demand function and supply function, we say that there are variables like income and rainfall. Now we know it is kind of irrational to assume that values of rainfall is in our control or income of the customer or the buyer in the market is also within our control. So basically these are purely exogenous variables in this simultaneous equation systems model. Now if I want to identify either demand or supply function in the market, can I do it given the variables that I have?

Look here I have two equations and there are endogenous variables and exogenous variables but note that here rainfall is going to be a shifter in the supply function because of course if there is very good rainfall we know we have very monsoon by god’s grace, then there will be bumper crops. So of course you can have a lot of wheat produced in the market at the same price level.

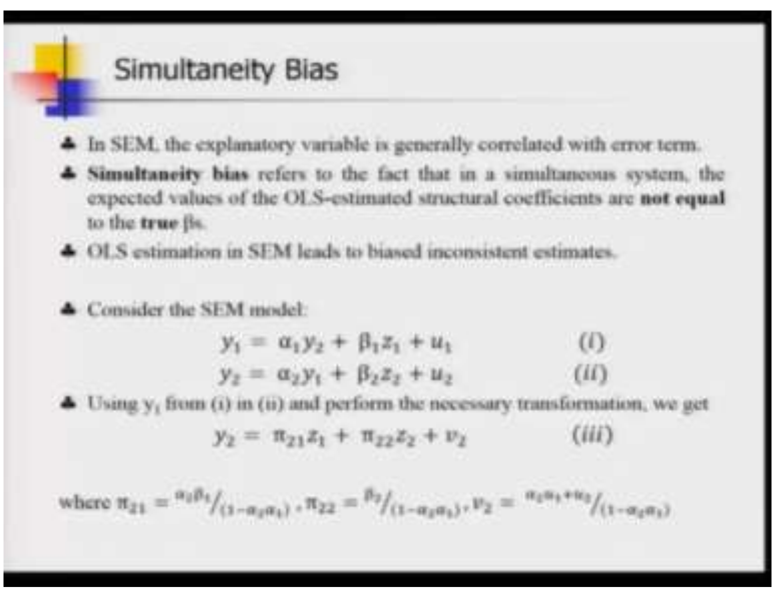
So supply will be higher, so rainfall works as a supply shifter and if there is an increase in the income, so there will be more money in buyers pocket so at a given price the buyer actually it is expected that he will demand for more commodity. So income will work as a demand shifter. Now the question is that here we have two shifters working simultaneously in the data set, so the problem remains the same.

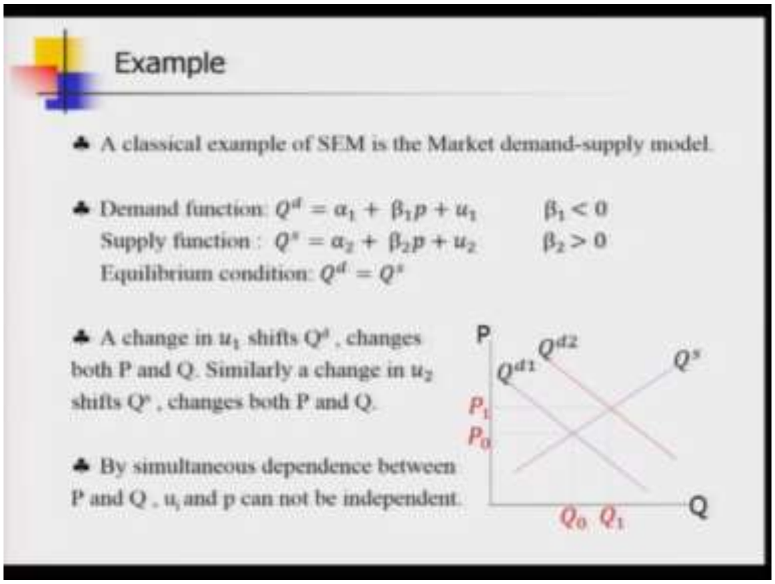
So we cannot identify whether we are estimating a demand function or a supply function because if there are changes and if we see that supply curve and demand curve are both shifting, then actually we cannot identify one particular curve. So we have now figured out at least one variable which will appear in one equation as an independent variable but it will not appear in the other equation. So what could be a variable?

Now from microeconomic theory, those who have done mathematical economics little bit probably have heard of a model Cobb wave model and there it is assumed that supply of an agricultural commodity or any commodity in a particular time period t actually depends somewhat on the price that was observed in the previous period. So basically I say that p t minus 1 actually can play a role in determining the quantity supplied in time period t.

So here if we include this variable in my model, it can be truly exogenous because when we are looking at the demand and supply equations for time period t and we have data on price of time period t minus 1, the past period's price has already been determined. So it is exogenous in time period t so now you can use that as an exogenous shifter in your model and it will shift which function? It will shift the supply function because it is going to enter the supply function it is not going to enter the demand function. So if you have this extra variable, you can actually identify the demand function by shifting the supply function for different time periods.

(Refer Slide Time: 25:21)





So stories are over, so now we are going to talk about simultaneity bias and some estimation methods but please note that this simultaneous equation systems is very complicated field, it is an advanced econometric topic and when I know I have only 50 minutes in a lecture I cannot do a proper justice if I start talking about a lot of things.

So I will just talk about the tips of the iceberg, some major points some terms we know which you better know as you are in this econometrics course. So this is more or less you can say a very preliminary kind of a discussion or very intermediate kind of discussion on this particular topic. So those who are interested please consult I know a proper econometrics textbook. So what is simultaneity bias?

Simultaneity bias refers to the fact that in a simultaneous system the expected values of the OLS estimated structural coefficients are not equal to the true betas. So I just have introduced a fresh term and that is called structural coefficient. So what do I mean by structural coefficient? It is also called behavioral coefficient.

So basically what happens that if you go back to that discussion on the demand function and the supply function so you can actually start with two linear equations right and that is what we have done. So if you go back to previous slides, so here yes so please look at the demand function equation here. So here you see alpha 1 and beta 1 as the coefficients in the demand function and in the supply function you have alpha 2 plus and beta 2 as the regression coefficients.

So these are basically the structural coefficients that I am talking about. So why are they called structural? Because they are coming from behavioral equations as they are portraying the structure of an economic model or portraying the behavior of an economic agent and then here in this market story, the economic agents are consumers and producers needless to say. So these coefficients that are appearing in the demand function and the supply function they are called the structural parameters or coefficients.

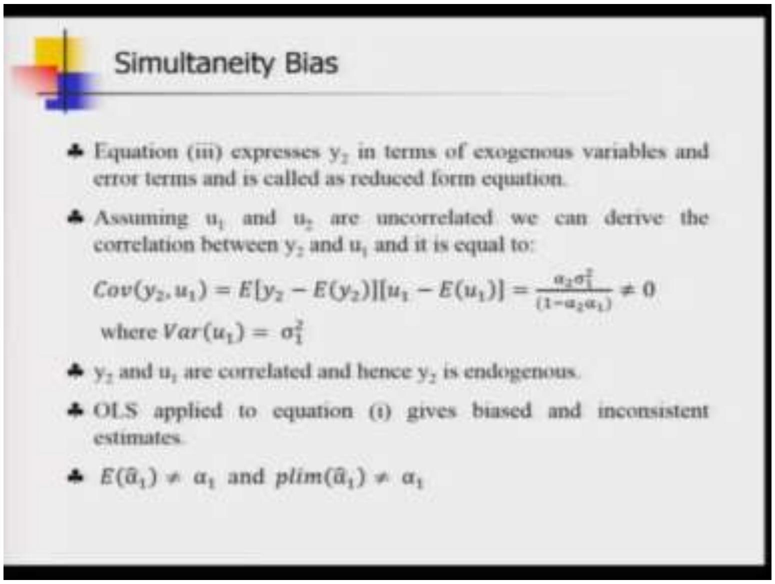
Now what after that? There is another concept called reduced form coefficient, so what happens in the reduced form coefficient? So first of all you have to get the reduced form equation, so a reduced form equation is one that expresses an endogenous variable solely in terms of the predetermined variables and the stochastic disturbances.

So now in this lecture we are going to actually look at how the reduced form equations are going to be found so now we know let us concentrate on these two equations that we are seeing in this particular slide. So we have two endogenous variables y1 and y2 and two exogenous variables z1 and z2 and these are equation numbers 1 and 2.

So now you use y1 from equation 1 and you plug that in equation number 2 and perform some transformations to get a third equation which is numbered 3 and here you see that it is written as y2 equals to pi 21 times z1 plus pi 22 times z2 plus v2. So here pi 21 and pi 22 are basically the simplified forms because the original form is very complicated and messy, so here at the bottom of the slide I have actually given you the full expression for pi 21 and pi 22 and also see this new error term v2 is also complicated because it involves in a lot of terms, so the final expression for v2 is also given at the bottom of the slide.

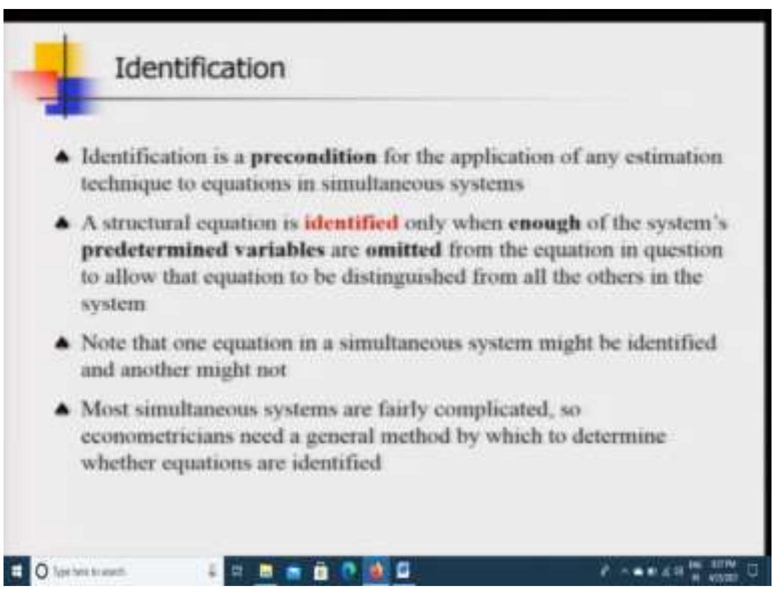
Now note that in this equation 3, you have an endogenous variable expressed fully in terms of two exogenous variables z1 and z2, so this is the reduced form equation where we actually express an endogenous variable in terms of exogenous variables and these two newly derived coefficients pi 21 and pi 22, these are called the reduced form coefficients.

(Refer Slide Time: 30:31)



So now you assume that u1 and u2 are uncorrelated and then we can derive the correlation between y2 and u1 and the expression is given here and this is just in a mathematical investigation, then you assume that y2 and u1 are correlated.

(Refer Slide Time: 30:52)

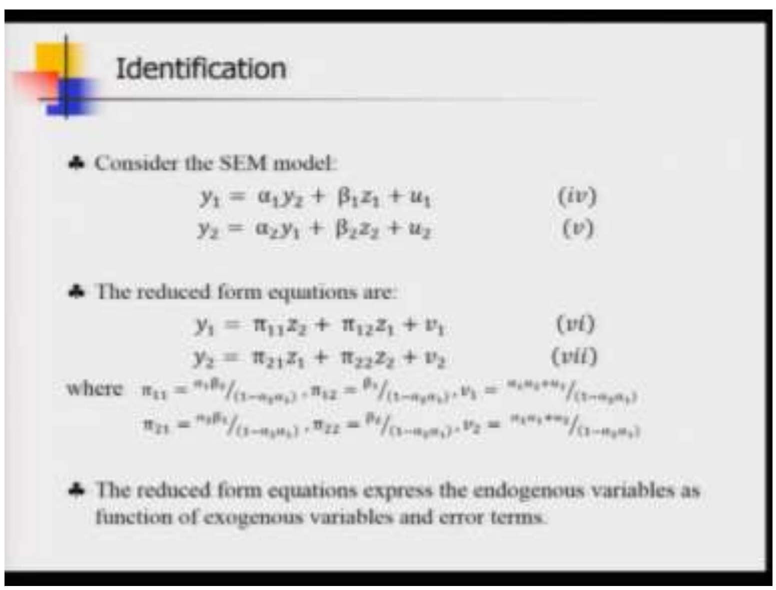


Now we are going to we know talk about identification problem again we have expressed this concept identification problem in terms of simple market models but let us now talk in terms of econometrics jargon. So identification is a precondition for the application of any estimation technique to the equations in a simultaneous system.

So a structural equation is identified only when enough of the systems predetermine variables are omitted from the equation in question, to allow that equation to be distinguished from all the others in the system. So it is looking very text bookish kind of definition but you do not have to worry if you do not follow each and every word in this particular definition.

Think about that market model discussion that we had when I said how we know shifter in one particular equation is going to identify the other. So you just remember that and that will be good enough to have an idea about what do we mean by identification. So simultaneous systems are fairly complicated, so econometricians need a general method by which they determine whether equations are identified or not and for that, we have some methods called in order condition and rank condition and that is what we are going to study next.

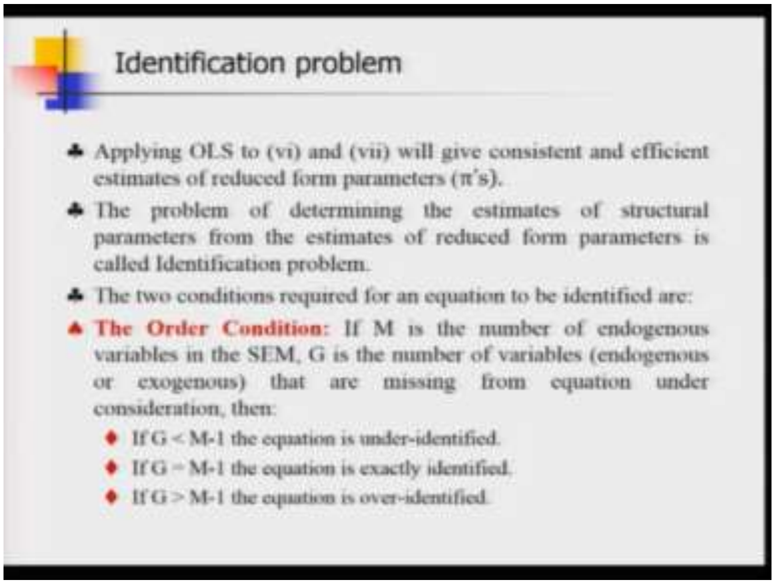
(Refer Slide Time: 32:30)



So now we know we again go back to you know the previous equation systems, where we have two endogenous variables y1 and y2 so we get two reduced form equations because we need to express each endogenous variable in terms of the exogenous variables only, so as there are two endogenous variables we required two reduced form equations.

So from two structural equations 4 and 5, we have to reduce from equations ah 6 and 7 and now you see we get 4 reduced form coefficients and errors are also shown here. Now these reduced form equations express the endogenous variables as a function of exogenous variables that we all know.

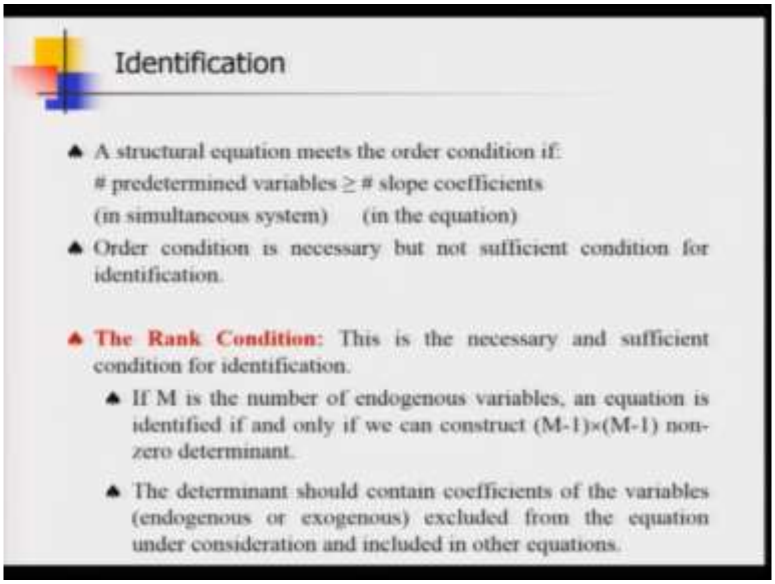
(Refer Slide Time: 33:24)



Now you can apply OLS to these reduced form equations number 6 and 7 and if you do so then you will get consistent and efficient estimates of the reduced form parameters. So basically you are going to get the true values of this reduced form coefficient pi. Now the problem of determining the estimates of the structural parameters still remains and how to derive the values of unknown structural coefficients from the reduced coefficient that is basically I know my identification problem is all about in econometric terms.

So for that we need to talk about two conditions and the first one in the list is called the order condition. So I am going to state the order condition quite mechanically by following in econometrics textbooks. So if m is the number of endogenous variables in simultaneous equations systems model and g is the number of weight total number of variables be it endogenous or exogenous that are missing from the equation under consideration, then if g is less than m minus 1 then the equation is under identified and if g equal to if g equals to m minus 1 then the equation is exactly identified and if g is greater than m minus 1 then the equation is over identified.

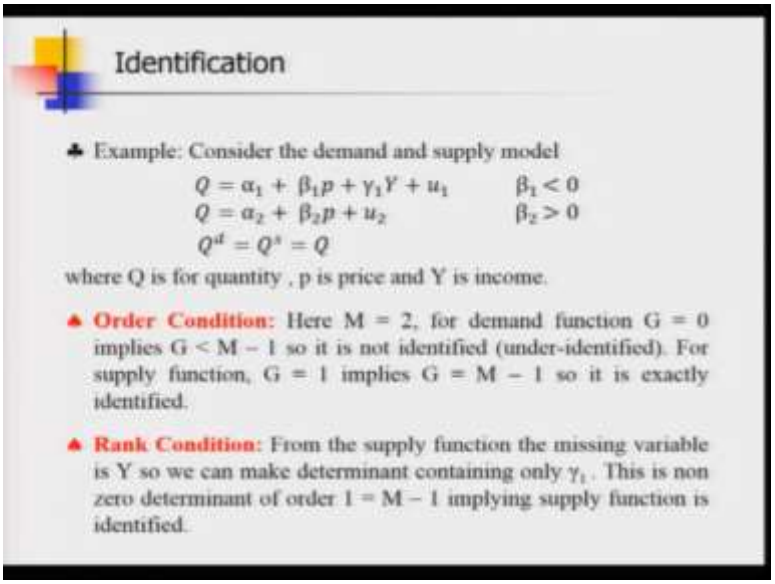
(Refer Slide Time: 34:55)



So a structural equation meets the order condition if the number of predetermined variables in the simultaneous systems is greater than or equal to the number of slope coefficients in the particular equation on which we have our interest and order condition is a necessary condition but it is not a sufficient condition. So what is the sufficient condition?

So there is another concept called rank condition which is both necessary and sufficient condition for identification purpose. So it says that if m is the number of endogenous variables, an equation is identified if and only if we can construct m minus 1 times m minus 1 non-zero determinant and the determinant should contain coefficients of the variables endogenous or exogenous excluded from the equation under consideration and included in other equations. So these things will be clear through an example that we are going to discuss at the end.

(Refer Slide Time: 35:57)

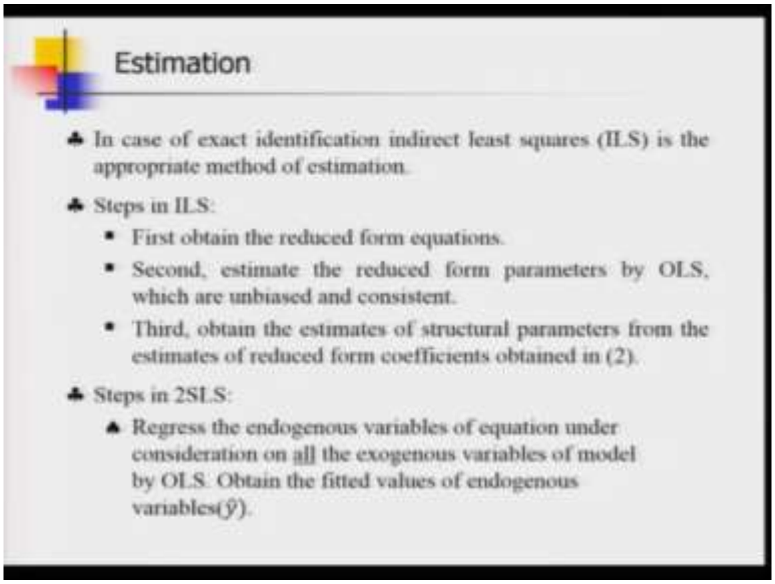


So now we are going to discuss this example, so let us go back to the market model. So it is a demand supply model you have the supply function and here you see that I have demand function which is defined over the market price and consumers income and then my supply function is defined over the market price only and thirdly I have this market equilibrium condition which is Qd equal to Qs equal to Q.

So now we know let us revisit the order condition and the rank condition by looking at this system of equations. There are three equations you know in this simultaneous equation systems model and you see here m is equal to 2 and for demand function the g is equal to 0. So basically we see here g is less than m minus one so it is not identified so it is actually under identified okay but for the supply function, g is equal to one so we see that g is equal to m minus one. So it is exactly identified.

Now we look at the rank condition from the supply function what do we learn? So there is a missing variable y in the supply function so we can make determinant containing only the gamma coefficient. So this is the non-zero determinant of order one so m minus one that implies that supply function is identified. So even if you forget this complex rank condition and order conditions, if you still recollect that shifter story that I spoke about you can easily identify here from this model that its basically the supply function which is identified and demand function is not.

(Refer Slide Time: 37:59)

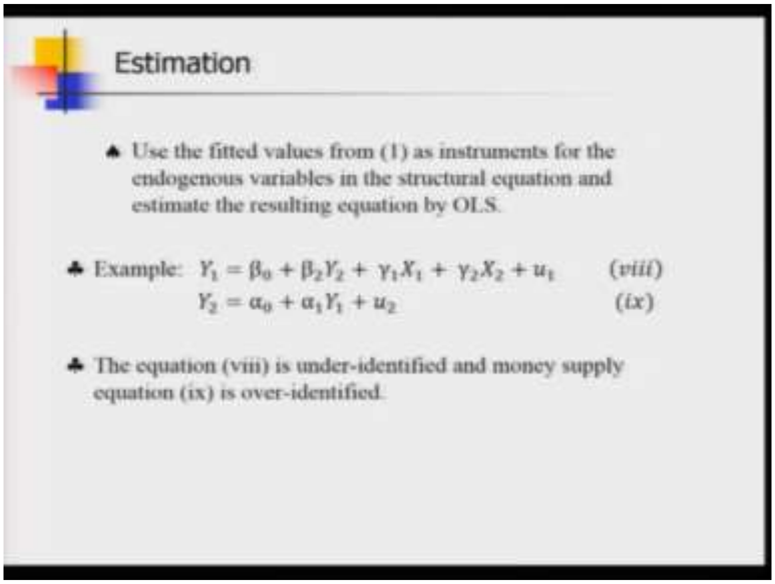


Suppose we are working with the simultaneous equation systems and then we know for sure that OLS is not going to be of any use, so then what estimation technique we are going to adopt? So econometricians come up with some solutions to handle this kind of simultaneous equation systems models and we are going to talk about two such solutions very briefly I will just give you like all the steps in a cook book kind of manner and we are going to discuss the iterated least squares and the two stage least squares as the estimation techniques.

So in the iterated least squares what are the steps? So in the first step you have to obtain the reduced form equations, then in the second step you estimate the reduced form parameters by OLS which are unbiased and consistent and then in the third and last stage you obtain the estimates of the structural parameters from the estimates of the reduced form coefficients obtained in the second stage.

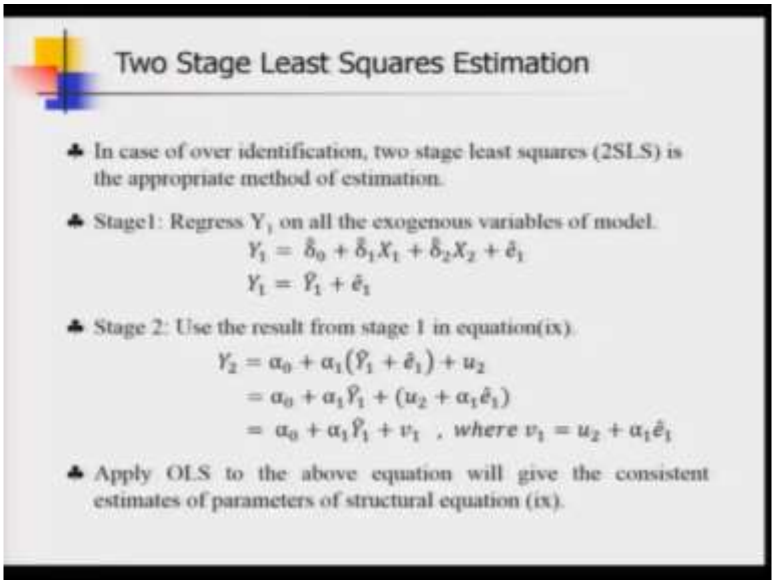
Now what are the steps in the two stage least squares procedure? Here you start by regressing the endogenous variables of equation under consideration on all the exogenous variables of the simultaneous equation systems model by running OLS, often the fitted values of this endogenous variables these are y hats.

(Refer Slide Time: 39:32)



Now in the second step you use this fitted values from step 1 as instruments for the endogenous variables in the structural equation and you estimate the resulting equation by OLS. So how do we adopt this method 2 SLS. So if by looking at the rank order conditions you figure out that the simultaneous equation systems that you have and if you are interested in one particular equation, so that equation is just identified, then basically you do not have to apply two SLS. Then you know you can you know use the ILS method that I just spoke about but if you have you know evidence ah in favor of over identification then actually two SLS is the only you know method that you can you know apply here.

(Refer Slide Time: 40:28)



So here you know ah in stage one, we regress ah you know y1 on all the exogenous variables in the model, so y1 is basically the first endogenous variable ah if you remember and you know we have two exogenous variables say x1 and x2. So you run OLS and you know you get the ah you know fitted values y hat 1. Now you take that y hat 1 value and you know plug that in the equation for the second endogenous variable because you know second endogenous variable y2 is a function of y1.

So you are basically replacing y1 by the fitted value plus the fitted residual expression, that is what you know we are doing here and then basically you we know ah you know do some algebra and then you get you know a latest ah you know revised form of equation which says that your endogenous variable y2 must be regressed on the fitted values of the first endogenous variable y1. So you apply OLS to this you know above equation and that will give you the consistent estimates of the parameters of the structural equation 9.

So we end our discussion here on simultaneous equation systems. So in the next lecture I am going to you know ah talk about another interesting ah you know advanced econometrics topic and that is the topic of panel data. So see you then. Thank you.