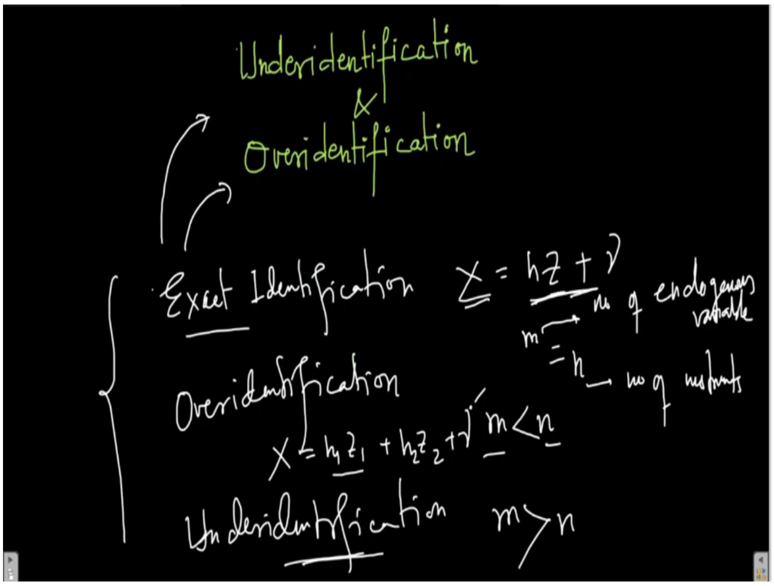
Applied Econometrics
Prof. Tutan Ahmed
Vinod Gupta School of Management
Indian Institute of Technology – Kharagpur

Lecture – 105
Overidentification Underidentification Exact Identification – Instrumental Variable

Hello and welcome back to the lecture on applied econometry.

(Refer Slide Time: 00:28)



We have been talking about instrumental variable. In the previous lectures what we have seen is the problem of endogenity. We talked in detail about the simultaneous equation bias problem. We have seen how instrumental variable can come to rescue in all these different cases of endogenity and then we actually talked about different types of instrumental variable strategy that we might have to face.

And we might have to get rid of situations like bad I V or weak I V and we would rather prefer to have a good idea. Now, in this lecture we are going to talk about another property of instrumental variable and that is called identification property. Now, identification could be of three types. It could be over identification, under identification and exact identification. Now, what I mean by all of this? Let us begin with the exact identification.

Now, exact identification is something that you have already seen. So, let us say we have one indogenous variable exact identification. Let us say we have one indogenous variable and that is let us say our X and our instrument is z. And X is, let us say, I have some h z and some error term. So, I have one endogenous variable and I have one instrumental variable. So, let us say, if my number of endogenous variable is m.

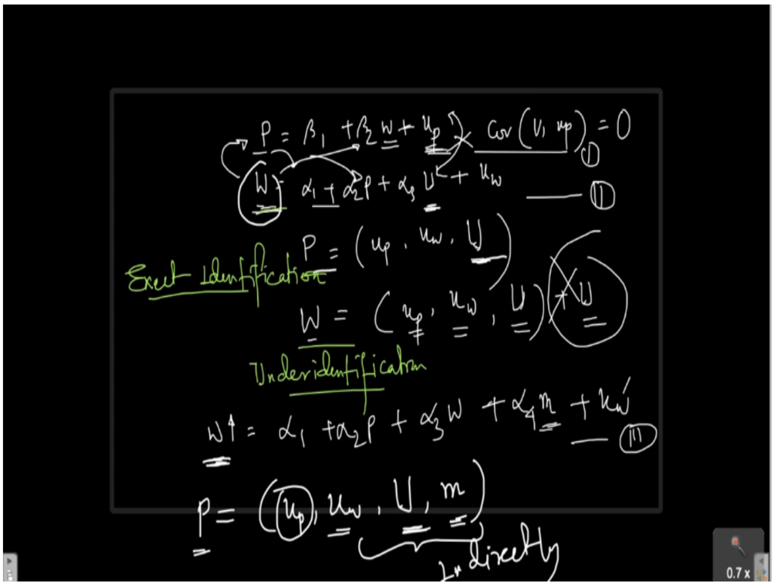
Let us say this is number of endogenous variable. And if my number of instrument is equal to let us say n. So then in case of exact rancidification, what I am going to have m = n, where n is number of instruments. Now, in this equation we can see I have one endogenous variable that is X and I have one instrumental variable that is z and you know basically my number of endogenous variable is equal to number of instruments.

So, this is a case of exact identification. Now, from this definition we can clearly make sense what could be a case of under identification and what could be a case of over identification. So, when we say over identification. So, in that case, what I am going to have? I will have more instrument than my endogenous variable. So, I am going to have my m less than n. I have more instrument for my endogenous variables.

And under identification quite obviously is going to be this reverse under identification. It is going to be m greater than n. So, an example of over identification. Let us say I have a variable X and I have let us say 2 instruments which one h 1 z 1 + h 2 z 2 and let us say some gamma prime whereas for under identification I do not have, let us say any of this instrument. So, if I have an endogenous variable that does not have an instrument that is there is a case of under identification.

So, let us elaborate with all these different cases with an example. An example that we have already seen.

(Refer Slide Time: 03:38)



Let us think of the example that I have previously done where we talked about the price rate is equal to beta 1 + beta 2 wage rate and then I have an error term u p. And then we also wrote a wage rate equation and which rate is equal to alpha 1 + alpha 2 price rate plus alpha 3 unemployment rate and we had a u w. So, we explained this P and W they are basically endogenous variable in these two equations.

And if I want to get an equation for P. So, what I can do I can actually substitute W here from this equation. And then what I can get? I can get in the equation of P, I can get a U. So, P could be expressed in terms of u p u w and U. So, this u w and U are coming from this w which is an endogenous regressor. And u p is already present in this regression equation. So, from here since U is an exogenous variable and since U is not related with the u p term.

So, covariance of U and u p is going to be 0. So, basically I can use U as an instrument. And in this equation I have W and U. They are correlated of course. So, basically covariance of U and W is not going to be 0. So that would satisfy the properties of I V. Now, that is all good but, if I think of this equation. Let us say this is equation 2 and the first one is equation 1, if I think about the equation 2.

So, for here, if I actually substitute with this P from this equation. So, what I will get W in terms of I will get u p and u w. And of course I will have my u which is already present in my W equation. So, what I am getting here is u p I am getting instead of P because that I got from this equation. And I do not have any exogenous variable here. I only have one exogenous variable which is U.

But that exogenous variable is already present in the model in this equation of W. So, I cannot use U again as an instrument because W is present, U is present sorry not W, U is present in this equation on its own right. So, if I again use another U. So, what is going to happen is, we are going to see the problem of exact collinearity. So, we do not want that if same variable is repeated twice in the regression equation we are going to have the case of exact collinearity.

So, we cannot use U as an instrument when U is already present in the regression equation. So, this is a case of under identification whereas the first one here is the case of exact identification and this is the case of exact identification. Now, from here we can actually make sense what could be the case of over identification. So, for that let me actually write down another equation. And this is going to be a little change, little variation in equation 2.

I will rewrite the equation 2. Let us say I write W wage rate is equal to alpha 1 + alpha 2 P alpha 3 W. Let us say now, I get a alpha 4 and now, I use a money supply rate m. And now, my error term let us say u w prime. Now, what is m? m I claim is an exogenous variable because m is basically the money supply rate. So, when I am basically talking about the money supply rate what happens to wage rate?

So, money supply rate is something that is exogenous because it is basically controlled by reserve bank of India in the context of India. And money when money is supplied in the economy. So, let us say money is channeled through different organizations and a part of the money goes in the form of wages. So, when money is pouring into the economy, there will be an upward push to the wage rate.

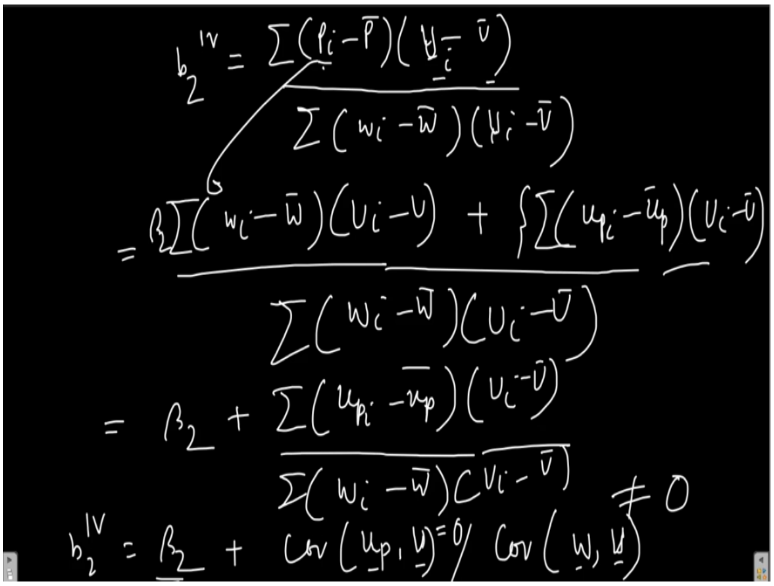
So then it makes sense to have money supply rate in this equation because money supply rate is going to have some impact on the wage rate. So, on that note, we can actually include the money supply rate. So, let us name this equation as equation 3 and let us also reduce the screen size a little bit. Now, if I try to now, express P in terms of all these error terms and exogenous variable terms. So, what I am going to get here?

P I am going to get u p, u w and I am going to have capital U and I am going to have m. Now, you see I have two exogenous variables. Now, instead of one exogenous variable that I had here. Now which one I am going to use as an instrument for my endogenous variable W. So, there both U and m are indirectly influencing P along with u w, these three variables are indirectly influencing P because all these three variables are coming through this W root.

So, now, out of these three of course these two are exogenous variable U and m and both I can actually ideally use as instrument. Now, we have seen that both U is related with W and m is also related with W. And because U and m are exogenous they are not related with the error term u p. So, basically, it is both of these variables are going to satisfy my condition of instrument. So which one I am going to use as I V?

So that is the question we have. Let us see, let us actually write down the beta 2 I V, if I use each of these instruments.

(Refer Slide Time: 09:49)



So, beta 2 I V is going to be summation so, this is equation of price so, P i – P bar and let us say my first instrument is going to be u which I have been using U i – U bar and the denominator my X is w. So, W i – W bar and U i – U bar. So, we already are quite familiar with the expansion of this equation. So, it is going to be, there will be two beta 1. So, we cancel the beta 1.

So, I have let us say beta 2 and I am going to have W i – W bar into U i – U bar plus I am going to have u p i – u p bar and U i – U bar and in both the cases I am going to have this denominator let me have some single line here and the denominator is going to be summation of W i – W bar U i – U bar. So, here just note that I have actually substituted the equation for P i and I basically cancel out beta 1 and then I have taken beta 2 out of this summation.

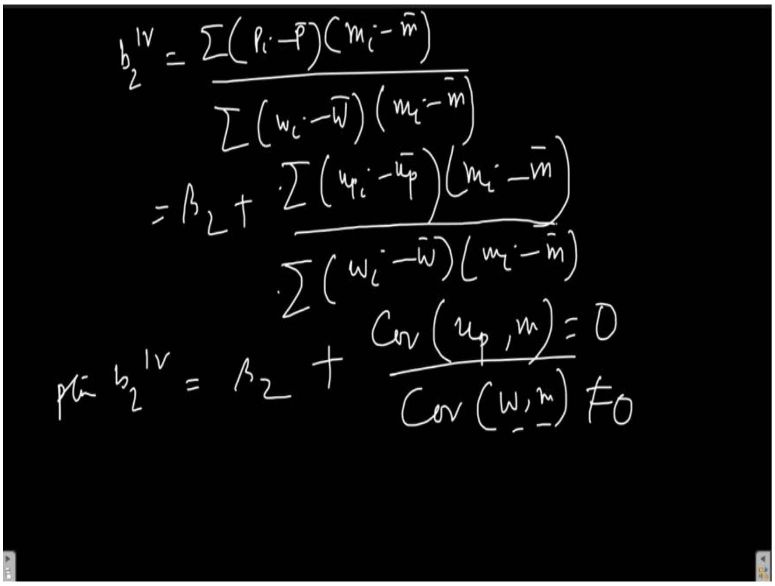
So, if I, you know further basically make it simple. So, it is going to be beta 2 plus this term here and we already know where we are heading. So, it is going to be u p i – u p bar and U i – U bar and then in my denominator, it is going to be summation of W i – W bar into U i minus U bar. So, we again take a P lim and if we take a P lim on both the sides. So, it is going to be beta 2 ols ah beta 2 I V = beta 2, P lim of beta 2 is beta 2.

And then, if I basically multiply 1 by n on both numerator and denominator. So, essentially I am going to get covariance of u p and U divided by covariance of W and U here. So, the denominator W and U of course we have seen in the previous equation that U is influencing W there in the regression equation, U is regressor so that is not going to be equal to 0. So, I am fine with this.

Denominator is not equal to 0 whereas the numerator we have seen u p and U they are not related. So, this term is going to be 0. So that means beta 2 I V, if we see, it is unbiased and at the same time when n tends to infinity, I am going to get this you know, if I use a P lim rule so, beta 2 is going to converge beta 2 I V is going to converge to beta 2. b 2 I V is going to converge to beta 2.

So then basically my properties of I V is being satisfied. Now, so that is basically for U as an instrument. Now, what will happen, if I take m as an instrument? Let us see that.

(Refer Slide Time: 13:02)



Beta 2 I V. Now, this time I am going to estimate beta to I V with m as an instrument. So, it is going to be summation of P i – P bar this is the price equation and instead of U, I am going to write m i – m bar and the denominator I am going to write my W is the X variable. So, W i – W bar and then I am going to I will multiply m i – m bar. Now, we have seen this previous equation. So, I am not going to repeat the step 2.

I am straight away going to step 3 which is basically beta 2 plus summation of, I am going to have u p i – u p bar into m i – m bar and in the denominator I am going to have the same terms as it is W i – W bar into m i – m bar. Now, if I take a limit on both the sides. So, basically what I am and basically multiply 1 by n on both numerator and denominator. So, basically I am going to have P lim beta 2 I V.

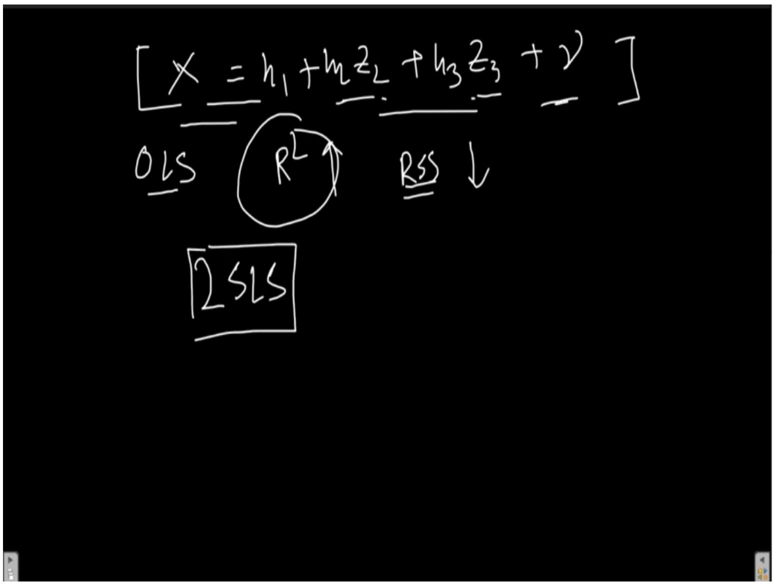
So, this is basically going to be beta 2 and I multiply both 1 by n numerator and denominator. So, I basically get the covariance of u p and m. And in the denominator I am going to get covariance of W and m. So, covariance of W and m is not equal to 0 because wage rate is related to the money supply. So, I am fine and here the u p and m, they are not related. So, I am going to. So, m is coming from the third equation and u p is coming from the first equation. So, they are not related.

So, I can write this equal to 0. So, essentially what is happening is b 2 I V is basically an unbiased estimator of beta 2. And for large n, b 2 I V is converging to beta 2. So, in both the cases what I am going to see is that they are basically satisfying the properties of instrumental variables. So, I can use either m or U for the determination for basically for our first regression where I will use them for w, instead of w. Now which one to do that which one to use?

So, one criteria could be we can simply see which one is giving minimum variance and we can use that. But there is another way to sort of identify you know like how this both the variables can be used in fact the strategies to use both the variables. So, what I do instead of using one you know instrumental variable what we do is, we actually use both the instrumental variable.

So, what I do there is, here what we have done we in the I V regression, we simply took one I V. But here what I am going to do is, I will use two I V.

(Refer Slide Time: 15:59)



So, X is going to be linear combination of let us say h 1 h 2 z 2 + h 3 z 3 and let us do some error term. So, why I do that? So, essentially this is simply going to be an ols regression. And I do that because actually this is a better combination instead of just using one I V. And what you, all we need to do is to get a good R square. We have to ensure that we get a good R square and we also try to see other properties of regression satisfies.

For example, we want to get a less, you know minimum sort of will reduce the residual sum of square and so, forth. So that is how we can actually combine both the instrumental variable when we regress them on this regress for this endogenous variable. So, this is the strategy. And we will see with this we will end this lecture here. So, we talked about both the you know over identification and under identification problem.

So, for over identification problem what do we do? We use both the I Vs in the first phase regression. And under identification problem we basically do not have a good I V. So, we perhaps, it is a good strategy to find a good I V and when you have exact identification, we are fine. So, with this we end this lecture here and in the next lecture we are going to talk about the regression strategy.

When you use, when you do this fast stage regression and the reduced form regression and we will say we will see this 2 stage least square method in the next lecture. So, with this we end this lecture here. Thank you.