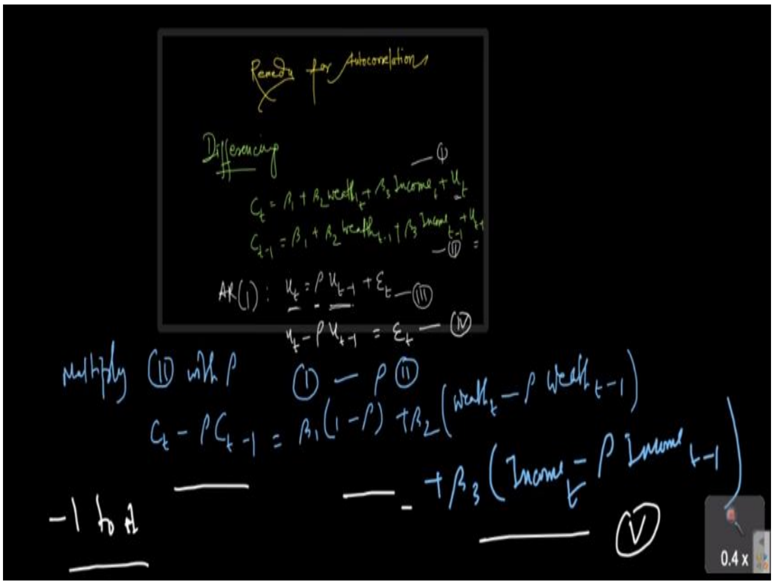
Applied Econometrics
Prof. Tutan Ahmed
Vinod Gupta School of Management
Indian Institute of Technology - Kharagpur

Lecture – 82
Remedy for Autocorrelation

(Refer Slide Time: 00:26)



Hello and welcome back to the lecture on applied econometrics. We have been talking about autocorrelation and we have been talking about different tests to identify if there is autocorrelation present in my data. Now, just knowing is not enough, we have to find a remedy to the problem. And in this lecture, we are actually going to talk about the remedy of autocorrelation.

And as I told you previously that one very important remedy, remedial measures that we take to sort of get rid of autocorrelation is differencing. So, we will see what is differencing and how we actually perform. So let me write down, let me use a different color, differencing technique we will talk about. Now, let us again go back to the previous example that we had used when we sort of used consumption data in United States.

And we wrote it as let us say my consumption, Ct consumption in time period t is equal to let us say beta 1 + beta 2 wealth beta 3 and let us say there is some disturbing stuff u t. So this is wealth t, income t and it is time series data. Now, if I take this model at t–1th time, so then it will be beta 1 + beta 2 wealth t–1 + beta 3 income t–1 + u t–1. Now, let us remember that; at least remember the AR 1 process a way to sort of understand the relationship among the error terms.

So if I write down an AR 1 process and in that AR 1 process, I wrote down that u t is related to the previous period error term and I can write u t is equal to some rho into u t–1 + some innovation term E t which is an IID. And here what is happening since I have autocorrelation, I am sort of assuming that in this AR 1 process, my th period error term is only related to my t–1th period error term, which is u t–1 this correlation coefficient between these two is basically rho.

So now, look at the previous two equations that I have just written, this is 1, this is 2 and let us this is 3. Now, if I try to see that actually if we can remove this u t or u t–1, so then we can actually possibly get rid of the sort of the pattern of the error. So in fact, if I write down this AR 1 processes u t – rho u t–1 minus one, all I will be left with is that innovation term which is IID and this is absolutely random and then I do not have any problem in having an OLS model.

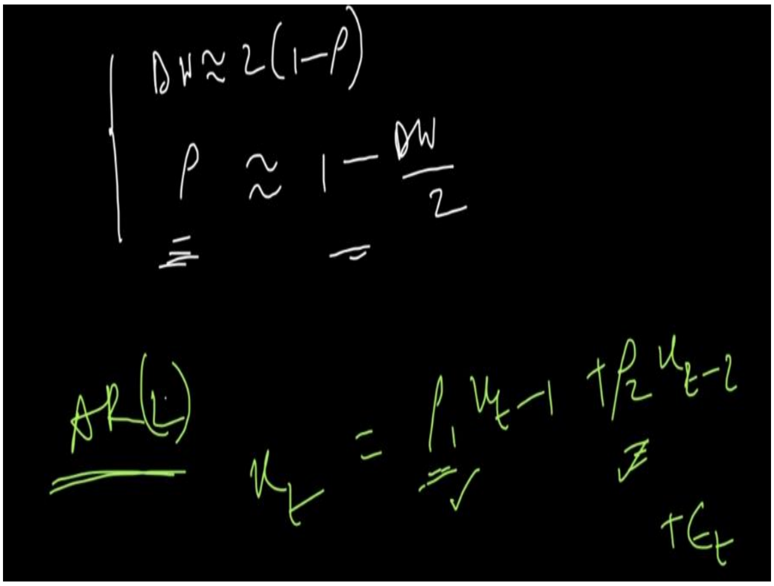
So basically then I will be able to get rid of the autocorrelation present in the data. Now, if I want to achieve that what I have to do is essentially I have to basically multiply equation two with rho. So if I further reduce the font, so let us say; let me use a different color multiplying multiplying 2 with rho and then subtracting it from 1. So let us say 1 – 2 rho into, 2 if I do this what I get is C t – rho C t–1 = beta 1 1 – rho beta 2 into I will have wealth at t period t – rho into wealth at t–1.

Then I can further write beta 3 + let us say beta 3 income – rho into, sorry income at tth period rho into income t–1th period. So that is what we get and if we can do that essentially I will be able to get rid of the u t terms and then I will essentially have only my random error present in my model, which I basically want. Now, in order to achieve this; so this is simply the technique of differencing, now in order to achieve this so there is there is a challenge in terms of finding the rho.

How would you find the right rho? And we know the value of rho actually varies between, we all know it is between –1 to +1. And then you can do one thing, you can actually sort of substitute different values of rho and see where basically we can plot this data and you can see or you can actually adopt other tests like Durbin-Watson and Breusch–Godfrey and see if you still have autocorrelation present when you actually get this new equation.

Now, that is one way of doing but in this case, what do you need to do is you need to sort of do a lot of trial and error, so this is equation 5. So you need to do a lot of trial and error to identify the right rho but this is a doable process, but there is a little bit of hard work you need to do. The other way you can actually do is basically you can run a regression between u t and u t–1 and identify the the value of rho. You can actually identify the value of rho.

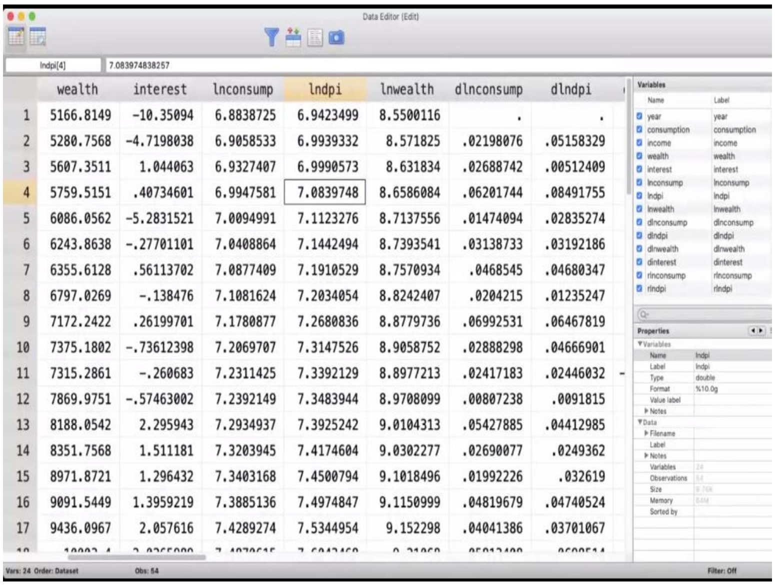
(Refer Slide Time: 06:54)



There are other possibilities also you can actually have that if it is an AR 1 process you know that the Durbin Watson statistic is nothing but or let me write following the same convention DW is almost equal to 2 into 1 – rho. Now, if you get the value of Durbin Watson statistic what you can do you can basically get the value of rho. So, rho is essentially nothing but rho is going to be Durbin Watson; so basically it will be 1 – Durbin Watson statistic by 2.

So, that is the value of rho you are going to get. So, you can use the value of rho to sort of get your equation 5. Now, let me actually show you the data. We will be using the same data that we used previously. So, let me take you to the data editor. Let me show you that data editor.

(Refer Slide Time: 08:26)



	wealth	interest	lnconsump	lndpi	lnwealth	dlnconsump	dlndpi
1	5166.8149	-10.35094	6.8838725	6.9423499	8.5500116	.	.
2	5280.7568	-4.7198038	6.9058533	6.9939332	8.571825	.02198076	.05158329
3	5607.3511	1.044063	6.9327407	6.9990573	8.631834	.02688742	.00512409
4	5759.5151	.40734601	6.9947581	7.0839748	8.6586084	.06201744	.08491755
5	6086.0562	-5.2831521	7.0094991	7.1123276	8.7137556	.01474094	.02835274
6	6243.8638	-.27701101	7.0408864	7.1442494	8.7393541	.03138733	.03192186
7	6355.6128	.56113702	7.0877409	7.1910529	8.7570934	.0468545	.04680347
8	6797.0269	-.138476	7.1081624	7.2034054	8.8242407	.0204215	.01235247
9	7172.2422	.26199701	7.1780877	7.2680836	8.8779736	.06992531	.06467819
10	7375.1802	-.73612398	7.2069707	7.3147526	8.9058752	.02888298	.04666901
11	7315.2861	-.260683	7.2311425	7.3392129	8.8977213	.02417183	.02446032
12	7869.9751	-.57463002	7.2392149	7.3483944	8.9708099	.00807238	.0091815
13	8188.0542	2.295943	7.2934937	7.3925242	9.0104313	.05427885	.04412985
14	8351.7568	1.511181	7.3203945	7.4174604	9.0302277	.02690077	.0249362
15	8971.8721	1.296432	7.3403168	7.4500794	9.1018496	.01992226	.032619
16	9091.5449	1.3959219	7.3885136	7.4974847	9.1150999	.04819679	.04740524
17	9436.0967	2.057616	7.4289274	7.5344954	9.152298	.04041386	.03701067

So, if I go to the data editor you can see now we have data for; wait a second see the data editor, for some reason it is not showing, so give me a moment. Here we have the standard data editor. And in this data editor, we can see actually we have this consumption, income, wealth all these variables as this interest and then we have taken the log of it, log of consumption, log of disposable personal income DPIs is actually the same log of income.

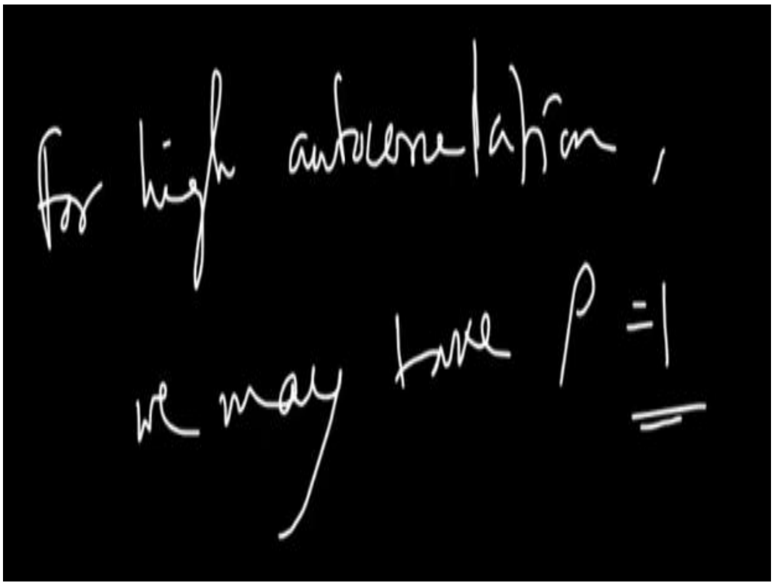
And then we have log of wealth, but then what I have done is I have actually subtracted log of consumption from the previous period. So essentially log of consumption in this period is 6.88, here it is 6.905. So, if I take a difference it is going to be 0.021. Similarly, for the next period what I have done it is 6.93, it is 6.90. So, if I take a difference it is going to be something like 0.026.

Similarly, for the next period is 6.99 and 6.66. If I take a difference it is going to be 0.06. So, that way I have kept on taking difference between two periods exactly the way we have done. Similarly for other variables also I have taken the same difference. And note here, I just come to this point, so note here, the value of rho is 1 here. So essentially, I am just differencing one period value from the past period value.

So, if we do that for log of disposable personal income, so it is going to be ln consumption of this period ln consumption of that period. So essentially, if I take a deep disposable personal income this period 6.94 and 6.99 and if I take a difference it is going to be 5, here again it is 6.999, here 6.993, if I take a difference it is going to be 0.005. The next period is 7.083 and it is 6.999. So, if I take a difference it is going to be this 0.08 because 7.08 – 6.99 is going to be 0.08 and so forth.

So, basically you take the difference between two periods t and t–1th period exactly the way we have seen in our equation here. So, here we have taken tth period and t–1th period and we have multiplied the value of rho. Now, what I sort of mentioned is that the rho in the previous data set what are you seeing in the data set the value of rho is basically 1. Now, why do I take the value of rho = 1?

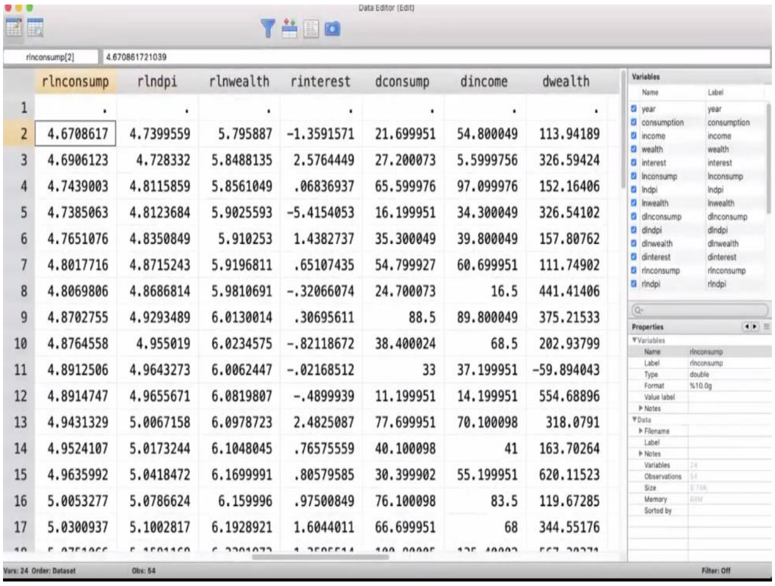
(Refer Slide Time: 11:09)



Now, there is a good reason why we take value of rho = 1 and that is because when we think that there is a high correlation present, autocorrelation is high, for high autocorrelation we may take rho = 1 but again it is just experiment basis you can take rho = 1 and you see the results and you can plot the data and you can actually do the tests and see where we stand in terms of removing autocorrelation.

But on the other hand, you can also use these different tests like Durbin Watson test or you can actually run the regression on u t and u t–1 and you can actually find out the value of rho and you can substitute the value of rho and see what is happening.

(Refer Slide Time: 12:01)



	rlnconsump	rlndpi	rlnwealth	rinterest	dconsump	dincome	dwealth
1	.	.	.	.	.	.	.
2	4.6708617	4.7399559	5.795887	-1.3591571	21.699951	54.800049	113.94189
3	4.6906123	4.728332	5.8488135	2.5764449	27.200073	5.5999756	326.59424
4	4.7439003	4.8115859	5.8561049	.06836937	65.599976	97.099976	152.16406
5	4.7385063	4.8123684	5.9025593	-5.4154053	16.199951	34.300049	326.54102
6	4.7651076	4.8350849	5.910253	1.4382737	35.300049	39.800049	157.80762
7	4.8017716	4.8715243	5.9196811	.65107435	54.799927	60.699951	111.74902
8	4.8069806	4.8686814	5.9810691	-.32066074	24.700073	16.5	441.41406
9	4.8702755	4.9293489	6.0130014	.30695611	88.5	89.800049	375.21533
10	4.8764558	4.955019	6.0234575	-.82118672	38.400024	68.5	202.93799
11	4.8912506	4.9643273	6.0062447	-.02168512	33	37.199951	-59.894043
12	4.8914747	4.9655671	6.0819807	-.4899939	11.199951	14.199951	554.68896
13	4.9431329	5.0067158	6.0978723	2.4825087	77.699951	70.100098	318.0791
14	4.9524107	5.0173244	6.1048045	.76575559	40.100098	41	163.70264
15	4.9635992	5.0418472	6.1699991	.80579585	30.399902	55.199951	620.11523
16	5.0053277	5.0786624	6.159996	.97500849	76.100098	83.5	119.67285
17	5.0300937	5.1002817	6.1928921	1.6044011	66.699951	68	344.55176

So, for example in our data set actually show you in our data set we also have identified the value of r or rho and if I see I have actually taken this r, this r or rho and so I have after obtaining the value of r what I have done is I have multiplied r with ln consumption or ln DPI or ln wealth and once I do that, I have basically sort of do the subtraction again from the original value. So, this is how we sort of remove the autocorrelation present in our model.

Now, one more point that I need to tell you; actually go back to the whiteboard, is that you can talk about if we can sort of make sense when we say that we subtract the value of the past period value multiplied by rho and we get sort of our model error free, but then again we have to kind of think that we are talking about, in this case we are only talking about AR 1 process where my my error is only related to the past period, but it could be; let me use different color for AR 2 process.

Where my u t is actually basically nothing but rho 1 u t–1 + rho 2, here I do not have a constant rho, I have rho 1 and rho 2, u t–2, or some error term. Now, in this case the simple formula may not work, so we have to do a little bit more iteration. But usually when you use software, you usually can mention AR 1 process or AR 2 process and software is smart enough to actually identify some value of rho 1 and rho 2 to see where you can get the model almost error free .

Where it can actually plot the data or it can actually do the tests to sort of give you the right value of rho 1 and rho 2. I will not cover more details on this, so for more details you can actually follow Damodar Gujarati’s book in Econometrics by Example. There are some more tests provided, you can look into that, but this is essentially the concept how you can get rid of autocorrelation problem using the differencing technique. So with this, we end this lecture here. Thank you.