Applied Microeconometrics - Assignments

483670, 630516

September 4, 2022

1

Explain why first differencing the equation does not solve the endogeneity problem of lagged consumption.

In a dynamic panel data model with a lagged (endogenous) variable, the issue arises that by construction the lagged variable, here the consumption, is correlated with η_i, the individual specific effect. Through first differencing, we can cancel out η_i:

$$log(consumption_{i,t}) - log(consumption_{i,t-1}) = \beta_1(log(prices_{i,t}) - log(prices_{i,t-1}))$$
$$+\beta_2(log(income_{i,t}) - log(income_{i,t-1})) + \beta_3(log(illegalopium_{i,t}) - log(illegalopium_{i,t-1}))$$
$$+\beta_4 + \beta_5(log(consumption_{i,t-1}) - log(consumption_{i,t-2})) + U_{i,t} - U_{i,t-1}$$

However, with a lagged endogenous regressor, the first-difference estimator stays biased and inconsistent as

$$E[U_{i,t} - U_{i,t-1}|consumption_{i,t-1} - consumption_{i,t-2}] \neq 0$$

Even if $T \to \infty$, the inconsistency does not disappear.

2

Anderson & Hsiao propose a specific instrumental variable procedure for the model. Write down and perform the associated first stage regression. Comment on its outcomes.

Assuming that $U_{i,t}$ is not serially correlated, we can use $log(consumption_{i,t-2})$ as an instrumental variable in the first-difference specification, formulating a

2SLS:

$$log(consumption_{i,t}) - log(consumption_{i,t-1}) = \beta_1(log(prices_{i,t}) - log(prices_{i,t-1}))$$
$$+\beta_2(log(income_{i,t}) - log(income_{i,t-1})) + \beta_3(log(illegalopium_{i,t}) - log(illegalopium_{i,t-1}))$$
$$+\beta_4 + \beta_5(log(consumption_{i,t-1}) - log(consumption_{i,t-2})) + U_{i,t} - U_{i,t-1}$$

$$log(consumption_{i,t-1}) - log(consumption_{i,t-2}) = \alpha_1 log(consumption_{i,t-2})$$
$$+\alpha_2(log(prices_{i,t}) - log(prices_{i,t-1})) + \alpha_3(log(income_{i,t}) - log(income_{i,t-1}))$$
$$+\alpha_4(log(illegalopium_{i,t}) - log(illegalopium_{i,t-1})) + \alpha_5 + V_{i,t}$$

$log(consumption_{i,t-2})$ can considered as an instrumental variable as it is correlated with the lagged first difference of consumption (validity), $log(consumption_{i,t-1}) - log(consumption_{i,t-2})$, but at the same time it is not correlated with the first difference of the error term (exogeneity/relevance), $U_{i,t} - U_{i,t-1}$. Note also that all exogenous variables have to be included in the first state regression when we perform a 2SLS!

Table 1

First-stage G2SLS regression

Number of obs	=	308
Wald chi(4)	=	61
Prob > chi2	=	0.0000

LD. logquantity	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
logprice D1.	-.6168866	.0894669	-6.90	0.000	-.7922386	-.4415346
logincome D1.	-.835989	.2192149	-3.81	0.000	-1.265642	-.4063357
logillegal D1.	-.0047149	.0128956	-0.37	0.715	-.0299898	.02056
logquantity L2.	-.0150557	.0091625	-1.64	0.100	-.0330138	.0029024
_cons	.0863313	.0605107	1.43	0.154	-.0322676	.2049302

The results of the first stage regression show that the percentage change in price and income is statistically significant at 1% while the percentage change in consumption is only statistically significant at 10%. The percentage change in intercepted illegal opium is not statistically significant, as well as the constant that corresponds to β_4.

When consumption from two periods before $(t-2)$ increases by 1%, consumption drops from period $t-2$ to period $t-1$ by around 0.015%. Considering that the t-test of lagged consumption is rather low, it may be a weak instrument. On the other hand, when we add **small** to **xtivreg**, we get an F-test of

20.27 at the first stage. While there are different opinions on how high a F test needs to be to consider an instrument strong, the general understanding is that a F test above 10 is sufficient which holds in our case.

If the price increases by 1%, the amount of opium that is consumed between period $t-2$ and $t-1$ drops by around 0.62%. If income increases by 1%, the demand drops by approximately 0.84%. However, we would like to point out that it is difficult to interpret these results as the change in price and income is basically in the future, taking place between $t-1$ and t, in comparison to the change in consumption that takes place between period $t-2$ and $t-1$.

The command implemented in Stata is:
xtset region year
xtreg L.d.logquantity L2.logquantity d.logprice d.logincome d.logillegal
OR
xtivreg d.logquantity d.logprice d.logincome d.logillegal (L.d.logquantity = L2.logquantity), first

3

Estimate the specification above using the Anderson & Hsiao approach. Comment on the underlying assumptions, tabulate the results and comment on the outcomes.

As already elaborated in the previous question, a central assumption of the Anderson & Hsiao approach is that $U_{i,t}$ is not serially correlated. To test for autocorrelation, we conduct an Arellano-Bond test, where the Null-hypothesis is that there is no autocorrelation.The results show that we can reject the null-hypothesis which is why have evidence that there is autocorrelation.

Figure 1

```
Arellano-Bond test for AR(1): z =  -0.62  Pr > z = 0.5370
Arellano-Bond test for AR(2): z =  -2.34  Pr > z = 0.0192
```

Further, the instruments need to be strong and relevant, $cov(Y_{i,t-2}, Y_{i,t-1} - Y_{i,t-2}) \neq 0$ as well as valid, $cov(Y_{i,t-2}, U_{i,t} - U_{i,t-1}) = 0$. The first-stage result has shown that the t-test of the instrument is rather low but at the same time the F-test is above 10 which is why we can consider the instrument as being sufficiently strong. With regards to validity, we would need to apply the Sargan test to test for it. However, as the system is not overidentified, having only endogenous variable and one instrument (the other exogenous variables do not count), we cannot apply the test.

Table 2

```
G2SLS random-effects IV regression              Number of obs     =        308
Group variable: region                          Number of groups  =         22

R-squared:                                      Obs per group:
     Within  = 0.2646                                         min =         14
     Between = 0.9260                                         avg =       14.0
     Overall = 0.2847                                         max =         14

                                                Wald chi2(4)      =      76.60
corr(u_i, X) = 0 (assumed)                      Prob > chi2       =     0.0000
```

__000004	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
logquantity						
LD.	1.470323	.8507647	1.73	0.084	-.1971454	3.137791
logprice						
D1.	.0217456	.5564101	0.04	0.969	-1.068798	1.112289
logincome						
D1.	1.878002	.762169	2.46	0.014	.3841782	3.371826
logillegal						
D1.	-.0289777	.0184598	-1.57	0.116	-.0651582	.0072028
_cons	-.0020396	.0225047	-0.09	0.928	-.046148	.0420688
sigma_u	0					
sigma_e	1.5485187					
rho	0	(fraction of variance due to u_i)				

```
Instrumented: LD.logquantity
 Instruments: D.logprice D.logincome D.logillegal L2.logquantity
```

The results show that only the lagged percentage change in consumption (p-value = 0.084) and the percentage change in income (p-value = 0.014)is significant.

If consumption increases by 1% in the previous period, the demand increases by approximately 1.47%. If the income increases by 1%, the demand increases by around 1.89%. It does make sense that the demand increases, when consumption increases in the previous period as opium is an addictive drug that often results in the drug addict consuming more and more (Van Ours 1995). As the income increases, more opium is consumed which indicates that opium is a normal good, a good whose demand increases with the income.

The command implemented in Stata is:
ivregress 2sls d.logquantity d.logprice d.logincome d.logillegal (L.d.logquantity = L2.logquantity)
abar, lags(2)
OR
xtivreg d.logquantity d.logprice d.logincome d.logillegal (L.d.logquantity = L2.logquantity), first

4

Describe the Arellano & Bond GMM estimator for this model.

The Arellano & Bond (AB) estimator also uses instruments to circumvent the endogeneity problem but uses many more moments in its estimation. As each additional instrument allows for more moments, more observations feed into the estimation. The instruments are the past lags of quantity. However, there are downsides to including too many instruments, as more weak instruments threaten small sample bias.

Under the assumption that the instrument is exogenous, this gives us all the moments we can use

$$\mathrm{E}\left[Z_{it-k}\Delta U_{it}\right] = 0 \tag{1}$$

where Z_{it-k} is the $(t-k)$th lag of the instrument and $\Delta U_i it$ is the first differenced error term (to summarise the regression equation given in part 1). In this case, we are instrumenting using the absolute values of past lagged quantities.

5

Estimate the model parameters using the Arellano & Bond estimator, tabulate the results and discuss the parameter estimates.

Table 3

	(1) GMM	(2) System
logprice	-0.422***	-0.363***
	(-9.70)	(-9.30)
logincome	1.662***	1.604***
	(8.06)	(8.56)
logillegal	-0.0240**	0.0258***
	(-2.27)	(3.28)
year	-0.0182***	0.00458
	(-3.29)	(1.01)
L.logquantity	0.681***	0.910***
	(22.45)	(86.60)
_cons		-15.05*
		(-1.85)
N	308	330

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

There is risk of auto-correlation in the residuals as the AR(1) test is rejected at the 10% level. Hence, we cannot use the Sargan test for the joint validity of the instruments. We turn our attention to the Hansen test, which is not rejected. This means that we can assume the instruments are jointly valid. However, the auto-correlation also poses a problem as an assumption for the GMM estimator is no auto-correlation of the error terms. This means the results are unreliable. The persistence of the quantity is also high at around 0.7. This means that we may already worry about the results of a GMM estimation, which is shown to be unreliable with high persistence. The system GMM approach may be better at handling this.

The results in Table 3 column 1 are the GMM estimation using the AB estimator. All the coefficients are statistically significant. The independent variables are measured in their natural logarithms. This means that they can be interpreted as percentage changes and not just in absolute terms. The coefficient of -0.422 for log prices means that a 1% increase in the price is associated with a demand drop of 0.4%. If income increases by 1%, the quantity demanded also increases by 1.6%. If the number of illegal drugs found increases, this is associated with a decrease in the demand for legal production. This also makes sense as the two can be expected to be substitutes. There also seems to be a small downward trend over time. As the year variable only increases by around 0.0005% each year (1/1924) this is economically insignificant. Finally, there is a high persistence in the demand. An increase (decrease) last year by 1 percent increases (decreases) current demand by 0.68%.

The command implemented in Stata are:
xtset region year xtabond2 logquantity logprice logincome logillegal year L(1).logquantity, gmmstyle(L(1).logquantity) ivstyle(logprice logincome logillegal year) nolevel robust

6

What is in your estimate for the short-run and the long-run price elasticity of opium?

The short-run estimate for the elasticity is the effect a change in the price this period has on the current demand. Table 3 shows this to be around -0.422. This means that if the price increases by 1% the demand decreases by 0.422%.

The long-run estimate is the extent to which demand in the future changes in response to a change in the current price. This is calculated as: $\frac{1}{1-\beta_5}\beta_1$. Using the numbers estimated in Table 1, the long-run effect is around 1.323. This means that an increase (decrease) in the price today by 1% is expected to decrease (increase) demand in the future by 1.323%.

7

Now estimate the model parameters using the system estimator (Blundell & Bond). Tabulate results, compute the elasticities (as in 6.).

Tests for autocorrelation of the error terms are still rejected at the 10% level. As no autocorrelation of errors is again an assumption, we are concerned about the reliability of the results. The high persistence worried about in part 5 is even higher here. Fortunately, the systems GMM estimator works well even with persistence.

The results of the system GMM are reported in column 2 of Table 3. The short-run price elasticity is -0.363. Using the same formula for the long-run price elasticities as part 6, it is calculated as -4.03. If the price increases by 1 percent, the demand decreases in the long run by around 4 percent. As for the other variables, the direction of effect for the measure of illegal produce switches compared to part 5. There is now a positive coefficient, suggesting that with more illegal trade the demand increases. Otherwise, the year trend is not statistically insignificant.

The commands used were:
xtabond2 logquantity logprice logincome logillegal year L(1).logquantity, gmmstyle(L(1).logquantity) ivstyle(logprice logincome logillegal year) robust

8

Which parameter estimates do you prefer? Explain why. Are there remaining problems with your preferred estimates?

The preferred method of estimation is the system estimate. The results of the 2SLS model have shown that the instrument is not completely weak but at the same time the instrument is not very strong. Further, there is evidence that the assumption of no autocorrelation does not hold as we rejected the nullhypothesis of no autocorrelation. We also cannot make any inferences on the validity of the instrument as have a just-identified system. Looking at the estimates logically, it is also surprising that an increase in the price has a positive effect on consumption amount of opium as well as that the effect of the price is highly insignificant.

Th system estimate is preferred over the general GMM because of the risk of persistence in the demand over time. The system estimator is reliable even then, whereas the general one is not. The autocorrelation in the error terms however still poses a risk to the conclusions drawn. The tests for this are rejected at the 10% sig. level meaning it could be problematic.

References

Van Ours, J.C. (1995), The price elasticity of hard drugs: the case of opium in the Dutch East Indies, 1923–1938, Journal of Political Economy 103, 261–279.