PSET 3 STATS II RIPS

Vaibhavi Sharma Pathak, Marcelo Piemonte Ribeiro, Fredrik Wallin

5/7/2022

Question 1

i)

Estimate the effect of job training grant (*grant*) on the hours of job training per employee (*hrsemp*). This simple relation can be summarized by the equation $hrsemp = \beta_0 + \beta_1 grant + u$. The OLS results are reported below.

Table 1: Cross section

	Dependent variable:
	hrsemp
grant	
Constant	8.887***
	(1.532)
Observations	129
R^2	0.000
Adjusted R^2	0.000
Residual Std. Error	17.403 (df = 128)
Note:	*p<0.1; **p<0.05; ***p<0.01

The initial results present no effect of the *grant*. This happens because because no firms received grants in 1987 - Wooldridge, J. M. (2019). Introductory econometrics: A modern approach. Cengage learning, Ch. 13, p.445.

ii)

In a panel context, the previous relation can be summarized by the following equation: $hrsemp_{it} = \beta_0 + \beta_1 grant_{it} + a_i + u_{it}$, t=1987, 1988, where a_i is term for unobserved heterogeneity.

iii)

The simple regression performed in *i)* likely suffers from omitted variable issues. This happens especially if the latter does not contain all possible control variables, which is very often the case. The use of panel data allow us to overcome such issue without the need of additional variables. This is possible because the unobserved non-time varying effects present in the error term and affecting the dependent variable can be accounted for in a panel setting.

iv)

The first difference equation is characterized by $\Delta hrsemp_i = \beta_0 + \beta_1 \Delta grant_i + \Delta u_i$

Table 2: First differences

	Dependent variable:
	hrsemp
grant	27.878***
	(3.129)
Constant	0.509
	(1.558)
Observations	125
R^2	0.392
Adjusted R^2	0.387
F Statistic	79.369*** (df = 1; 123)
Note:	*p<0.1; **p<0.05; ***p<0.01

Having a grant signficantly increases the hours of trainning per employee. An unit increase in the grant reflected around 28 more hours of trainning per employee.

v)

The fixed effect equation could be summarized as folows: $\overline{hrsemp_{i,t}} = \beta_1 \overline{grant_{it}} + \overline{a_i} + \overline{u_{it}}$, t=1987, 1988. We should expect the same results between FE and FD estimations because in this case T=2, in other words only two years are being considered. While the FD estimation performs the difference between grant in the year of 1988 and in 1987, the FE estimation performs the difference between each observation and the mean. Because T=2 the FD and FE equations will be equivalents

Table 3: Fixed effects

	Dependent variable:
	hrsemp
grant	28.387***
	(2.704)
Observations	256
R^2	0.471
Adjusted R^2	−0.089
F Statistic	110.229*** (df = 1; 124)
Note:	*p<0.1; **p<0.05; ***p<0.01

However, the estimates are slightly different. This could be a signal for the violation of the strict exogeneity.

vi)

The table below presents the FE and FD estimates. Although both estimates are significant, FE preseted a higher magnitude.

Table 4: First-differences and Fixed-effects full sample

	Dependent variable:	
	hrsemp	
	First-differences	Fixed-effects
	(1)	(2)
Constant	2.4856 (0.7709)**	
grant	30.7782 (2.8327)***	34.6301 (3.5529)***
Observations	255	390
R^2	0.4643	0.4709
Adjusted R^2	0.4622	0.1896
F Statistic	219.2749*** (df = 1; 253)	226.0308*** (df = 1; 254)
Note:	*p<0.1; **p<0.05; ***p<0.01 datasets::freeny lm() function vcovHC(type = 'HC1')-Robust SE	

vii)

Strict exogeneity implies that u_{it} should not correlate with the independent variables of all time periods. A serial correlation test allows to identify if such assumption holds. The results below reject the null of serial correlation for the FD model but not for the FE model. This corroborates the results found in *v)*.

```
Wooldridge's first-difference test for serial correlation in panels
```

data: fd_full F = 15.338, df1 = 1, df2 = 122, p-value = 0.0001486 alternative hypothesis: serial correlation in differenced errors

```
Wooldridge's test for serial correlation in FE panels
```

data: fe_within_full F = 0.84479, df1 = 1, df2 = 253, p-value = 0.3589 alternative hypothesis: serial correlation

viii)

The model to estimate either using FE or FD then becomes $hrsemp_{it} = \beta_0 + \beta_1 grant_{it} + \beta_1 union_{it} + a_i + u_{it}$, t=1987, 1988, 1989.

Table 5: First-differences and Fixed-effects full sample and union

	Dependent variable:	
	hrsemp	
	First-differences	Fixed-effects
	(1)	(2)
Constant	2.4856 (0.7709)**	
grant	30.7782 (2.8327)***	34.6301 (3.5529)***
Observations	255	390
R^2	0.4643	0.4709
Adjusted R^2	0.4622	0.1896
F Statistic	219.2749*** (df = 1; 253)	226.0308*** (df = 1; 254)
Note:	*p<0.1; **p<0.05; ***p<0.01	
	datasets::freeny	
	lm() function	
	vcovHC(type = 'HC1')-Robust SE	

Both results are the same as the previous estimation because both relies on within variation, but *union* does not vary across individuals. In other words, there is no within variation in *union*, individuals associated to an union were linked to the latter in all three years.

Question 2

i)

The IV used in the paper "is an indicator variable for whether or not the country is a former colony of the EU Council presidency in the second 6 months of the year t-2, when the budget is determined", Carneige and Marinov (2017), p.677. The authors employed such strategy because the original relation they aimed to estimate contained an endogenous variable, logged net EU official development assistance (ODA). This happened because the previous variable is not ramdonly assigned as "aid disbursements are made in ways that are systematically related to the recipient countries' human rights" p.677, where recipient countries' human rights is the dependent variable. This strategy was necessary to respond such estimation, otherwise the inital results could mask reverse causality.

ii)

First, they assume that $Colony_{i(t-2)2}$ only affects DV_{it} through the explanatory variable ($log(ODA)i(t-1)$), which is the exclusion restriction (p. 677). They therefore need two statistical assumptions to be met: Random assignment of colony status; and a significant effect of the $Colonyi(t-2)2$ on $log(ODA)_{i(t-1)}$. They show that the first is met through a quasi-random assignment of the presidency of the EU Council. As it rotates due to a mechanism decided upon long before, the randomness seems given and valid. The second assumption holds in the paper due to a significant effect of the instrument in the first stage regression (which we later also observe).

iii)

Equation for the first stage:

$$log(ODA)_{i(t-i)} = \theta_0 + \theta_1 Colony_{i(t-2)2} + \sum_{k \in K} \theta_k I(i = k) + \sum_{j \in J} \theta_j I(t = j) + e_{it}$$

Equation for the second stage:

$$DV_{it'} = \beta_0 + \beta_1 log(ODA)_{i(t-i)} + \sum_{k \in K} \theta_k I(i = k) + \sum_{j \in J} \theta_j I(t = j) + u_{it}$$

Note that in the data set:

$$log(ODA)_{i(t-i)} = EV$$

$$Colony_{i(t-2)2} = l2CPcol2$$

and

$$DV\_it' = new\_empinxavg$$

iv)

When running the first stage regression, we find an estimated relationship of $Colony_{i(t-2)2}$ on $\widehat{log(ODA)}_{i(t-i)}$ is 0.154 (SE = 0.073, p < 0.05). Even though we get a significant result in this regression, when conducting a t-test (regressing EV on *l2CPcol2*, without control variables), the result is not significant (-0.003, p =0.96). Consequently, it is difficult to conclude that the instrument is strong. Moreover, we cannot report the standard errors from this analysis.

v)

The result for the `ivreg` command was quite different to either the manual calculations or the calculation of the panel instrumental variable (via `plm()`-command). While the latter two - which are depicted as the first two columns in the table above - are more or less consistent, the result of the `ivreg`-command is neither significant, nor positive (for EV). We guess that this stems from the fact that in this command, the panel data is not taken into account, whereby the fixed effects are then also not accounted for.

vi)

Again, there is a difference. When choosing the iv-approach, the result for EV is not significant, while it is for the panel OLS regression. This would then indicate that an approach via instrumental variable is not necessary, because it also entails an inherent loss of precision.

vii)

When running an endogeneity tests with the residuals of the first stage regression (`resid(tsls1_first)`), we find that we can reject the null hypothesis (H_0) at $p < 0.1$. Based on this, it appears the explanatory variable is indeed not exogenous.

Table 6: 2SLS manual estimates

	Dependent variable:	
	EV First-differences	new_empinxavg Fixed-effects
	(1)	(2)
l2CPcol2	0.1543	
fitted(tsls1_first)		1.7054
covihme_ayem	0.0807	−0.1484
covwdi_exp	−0.1276	0.0669
covwdi_fdi	−0.0016	−0.0067
covwdi_imp	0.1501	−0.1510
covwvs_rel	−0.0500	0.0527
coviNY_GDP_PETR_RT_ZS	0.0087	−0.0093
covdemregion	−0.4595	0.7562
covloggdp	0.0872	−0.4175
covloggdpC	−0.1521	−0.0118
covihme_ayemF	8.3538	−17.0816
covwdi_expF	2.1947	−8.8106
covwdi_fdiF	−0.3980	0.3320
covwvs_relF	−5.0692	5.3605
coviNY_GDP_PETR_RT_ZSF	1.8458	−2.6265
covdemregionF	−46.1410	79.4564
covloggdpF	−6.4346	−52.4687
X_Iyear_1987	0.3070	−0.3514
X_Iyear_1989	−0.5117	1.0817
X_Iyear_1990	−0.3964	1.0402
X_Iyear_1991	−0.3635	1.0554
X_Iyear_1992	−0.2824	1.0221
X_Iyear_1993	−0.0472	0.8054
X_Iyear_1994	−0.1682	0.9741
X_Iyear_1995	−0.0257	0.9531
X_Iyear_1996	0.0784	0.8293
X_Iyear_1997	0.0171	0.8353
X_Iyear_1998	−0.0850	0.9260
X_Iyear_1999	−0.1200	0.8121
X_Iyear_2000	−0.2762	0.9993
X_Iyear_2001	−0.5875	1.4509
X_Iyear_2002	−0.3256	0.8417
X_Iyear_2003	−0.3419	0.9623
X_Iyear_2004	−0.1932	0.7471
X_Iyear_2005	0.0081	0.1366
Observations	1,792	1,792
R^2	0.1428	0.1295
Adjusted R^2	0.0650	0.0505
F Statistic (df = 35; 1642)	7.8155***	6.9800***
Note:	*p<0.1; **p<0.05; ***p<0.01 datasets::freeny lm() function vcovHC(type = 'HC1')-Robust SE	

Table 7: 2SLS ivreg

	Dependent variable:		
	new_empinxavg		
	panel linear		*instrumental variable*
	First-differences	Fixed-effects	
	(1)	(2)	(3)
fitted(tsls1_first)	1.7054		
Constant			10.7215 (7.3470)
EV		1.7054	−0.2503 (3.3964)
covihme_ayem	−0.1484	−0.1484	−0.3276 (0.1872)*
covwdi_exp	0.0669	0.0669	0.3878 (0.4610)
covwdi_fdi	−0.0067	−0.0067	−0.0135 (0.0248)
covwdi_imp	−0.1510	−0.1510	−0.2102 (0.4019)
covwvs_rel	0.0527	0.0527	0.1476 (0.8949)
coviNY_GDP_PETR_RT_ZS	−0.0093	−0.0093	−0.0586 (0.0408)
covdemregion	0.7562	0.7562	7.7776 (0.4871)***
covloggdp	−0.4175	−0.4175	−0.2819 (0.9289)
covloggdpC	−0.0118	−0.0118	0.5180 (2.1471)
covihme_ayemF	−17.0816	−17.0816	−35.9327 (21.2971)*
covwdi_expF	−8.8106	−8.8106	17.0116 (18.5364)
covwdi_fdiF	0.3320	0.3320	−4.7893 (3.7242)
covwdi_impF			
covwvs_relF	5.3605	5.3605	13.8112 (90.3680)
coviNY_GDP_PETR_RT_ZSF	−2.6265	−2.6265	−5.5209 (1.7083)***
covdemregionF	79.4564	79.4564	784.3217 (47.9654)***
covloggdpF	−52.4687	−52.4687	18.9632 (120.5712)
covloggdpCF			
X_Iyear_1987	−0.3514	−0.3514	−5.7402 (3.2909)*
X_Iyear_1988			−4.7691 (2.5763)*
X_Iyear_1989	1.0817	1.0817	0.9118 (2.3808)
X_Iyear_1990	1.0402	1.0402	0.8631 (2.0863)
X_Iyear_1991	1.0554	1.0554	1.2162 (1.6933)
X_Iyear_1992	1.0221	1.0221	0.9898 (1.4739)
X_Iyear_1993	0.8054	0.8054	0.7571 (0.8274)
X_Iyear_1994	0.9741	0.9741	0.7761 (1.0364)
X_Iyear_1995	0.9531	0.9531	1.3416 (0.7285)*
X_Iyear_1996	0.8293	0.8293	1.3560 (0.4645)***
X_Iyear_1997	0.8353	0.8353	1.1558 (0.4975)**
X_Iyear_1998	0.9260	0.9260	0.6665 (0.7870)
X_Iyear_1999	0.8121	0.8121	0.7686 (0.9045)
X_Iyear_2000	0.9993	0.9993	0.6259 (1.4375)
X_Iyear_2001	1.4509	1.4509	0.6376 (2.2924)
X_Iyear_2002	0.8417	0.8417	0.5392 (1.1514)
X_Iyear_2003	0.9623	0.9623	0.4629 (1.4090)
X_Iyear_2004	0.7471	0.7471	0.5678 (0.9378)
X_Iyear_2005	0.1366	0.1366	0.2179 (0.3978)
X_Iyear_2006			
Observations	1,792	1,792	1,792
R^2	0.1295	0.0504	0.4626
Adjusted R^2	0.0505	−0.0357	0.4515
Residual Std. Error			2.7883 (df = 1755)
F Statistic	6.9800*** (df = 35; 1642)	138.0321***	
Note:	*p<0.1; **p<0.05; ***p<0.01 datasets::freeny lm() function vcovHC(type = 'HC1')-Robust SE		

Table 8: IV and OLS

	Dependent variable:	
	new_empinxavg	
	instrumental variable	*panel linear*
	First-differences	Fixed-effects
	(1)	(2)
Constant	10.7215 (0.7709)	
EV	−0.2503	0.1903
covihme_ayem	−0.3276	−0.0384
covwdi_exp	0.3878	−0.1313
covwdi_fdi	−0.0135	−0.0092
covwdi_imp	−0.2102	0.0863
covwvs_rel	0.1476	−0.0165
coviNY_GDP_PETR_RT_ZS	−0.0586	0.0042
covdemregion	7.7776	0.0955
covloggdp	−0.2819	−0.2768
covloggdpC	0.5180	−0.2435
covihme_ayemF	−35.9327	−5.7156
covwdi_expF	17.0116	−4.9929
covwdi_fdiF	−4.7893	−0.2773
covwdi_impF		
covwvs_relF	13.8112	−1.6483
coviNY_GDP_PETR_RT_ZSF	−5.5209	0.2158
covdemregionF	784.3217	13.1817
covloggdpF	18.9632	−61.2952
covloggdpCF		
X_Iyear_1987	−5.7402	0.0080
X_Iyear_1988	−4.7691	
X_Iyear_1989	0.9118	0.2927
X_Iyear_1990	0.8631	0.4266
X_Iyear_1991	1.2162	0.5571
X_Iyear_1992	0.9898	0.5848
X_Iyear_1993	0.7571	0.7239
X_Iyear_1994	0.7761	0.8077
X_Iyear_1995	1.3416	0.9069
X_Iyear_1996	1.3560	0.9385
X_Iyear_1997	1.1558	0.9024
X_Iyear_1998	0.6665	0.7886
X_Iyear_1999	0.7686	0.6231
X_Iyear_2000	0.6259	0.5750
X_Iyear_2001	0.6376	0.5561
X_Iyear_2002	0.5392	0.4077
X_Iyear_2003	0.4629	0.4468
X_Iyear_2004	0.5678	0.4512
X_Iyear_2005	0.2179	0.1474
X_Iyear_2006		
Observations	1,792	1,792
R^2	0.4626	0.1379
Adjusted R^2	0.4515	0.0596
Residual Std. Error	2.7883 (df = 1755)	
F Statistic		7.5022^{***} (df = 35; 1642)
Note:	*p<0.1; **p<0.05; ***p<0.01	
	datasets::freeny	
	lm() function	
	vcovHC(type = 'HC1')-Robust SE	

Table 9: Endogeneity test with residuals from the 1st-stage

	Dependent variable:
	new_empinxavg First-differences
EV	1.7054
covihme_ayem	−0.1484
covwdi_exp	0.0669
covwdi_fdi	−0.0067
covwdi_imp	−0.1510
covwvs_rel	0.0527
coviNY_GDP_PETR_RT_ZS	−0.0093
covdemregion	0.7562
covloggdp	−0.4175
covloggdpC	−0.0118
covihme_ayemF	−17.0816
covwdi_expF	−8.8106
covwdi_fdiF	0.3320
covwvs_relF	5.3605
coviNY_GDP_PETR_RT_ZSF	−2.6265
covdemregionF	79.4564
covloggdpF	−52.4687
X_Iyear_1987	−0.3514
X_Iyear_1989	1.0817
X_Iyear_1990	1.0402
X_Iyear_1991	1.0554
X_Iyear_1992	1.0221
X_Iyear_1993	0.8054
X_Iyear_1994	0.9741
X_Iyear_1995	0.9531
X_Iyear_1996	0.8293
X_Iyear_1997	0.8353
X_Iyear_1998	0.9260
X_Iyear_1999	0.8121
X_Iyear_2000	0.9993
X_Iyear_2001	1.4509
X_Iyear_2002	0.8417
X_Iyear_2003	0.9623
X_Iyear_2004	0.7471
X_Iyear_2005	0.1366
resid(tsls1_first)	−1.5193
Observations	1,792
R^2	0.1397
Adjusted R^2	0.0611
F Statistic	7.4039^{***} (df = 36; 1641)
Note:	*p<0.1; **p<0.05; ***p<0.01 datasets::freeny lm() function vcovHC(type = 'HC1')-Robust SE