HW-06

Task 1.

The model is

$$y_{it} = x'_{it}\beta + c_i + u_{it}, i = 1, ..., n, t = 1, 2.$$

The OLS estimator in

$$y_{i2} - \bar{y}_i = (x_{it} - \bar{x}_i)'\beta + (u_{i2} - \bar{u}_i)$$

is FE estimator, where

$$\begin{split} y_{i2} - \bar{y}_i &= y_{i2} - 1/2(y_{i1} + y_{i2}) = 1/2(y_{i2} - y_{i1}) = 1/2\Delta y_{i2}, \\ x_{i2} - \bar{x}_i &= x_{i2} - 1/2(x_{i1} + x_{i2}) = 1/2(x_{i2} - x_{i1}) = 1/2\Delta x_{i2}, \\ u_{i2} - \bar{u}_i &= u_{i2} - 1/2(u_{i1} + u_{i2}) = 1/2(u_{i2} - u_{i1}) = 1/2\Delta u_{i2}. \end{split}$$

Therefore,

$$y_{i2} - \bar{y}_i = (x_{it} - \bar{x}_i)'\beta + (u_{i2} - \bar{u}_i)$$

can be rewritten as

$$1/2\Delta y_{i2} = 1/2\Delta x'_{i2}\beta + 1/2\Delta u_{i2},$$

$$\Delta y_{i2} = \Delta x_{i2}'\beta + \Delta u_{i2}\text{,}$$

which is an equation for FD estimator, i.e. FE and FD estimators are numerically identical in this case.

Task 2.

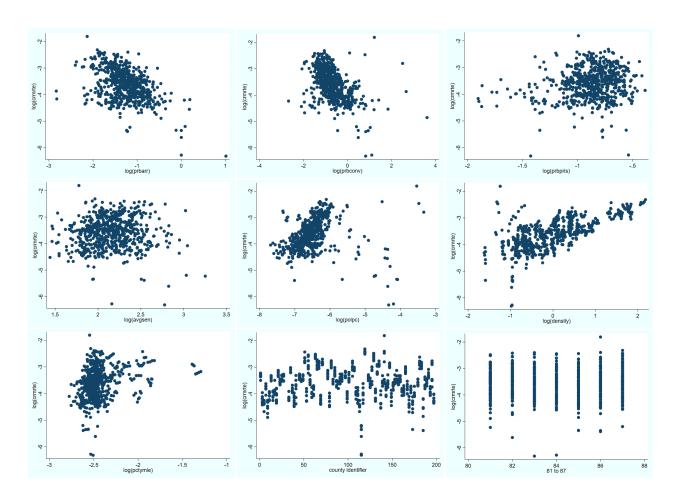
(1)

Descriptive statistics on the variables of main interest:

Variable	Obs	Mean	Std. Dev.	Min	Max
crmrte	630	.0315876	.0181209	.0018116	.163835
prbarr	630	.3073682	.1712047	.0588235	2.75
prbconv	630	.6886176	1.690345	.0683761	37
prbpris	630	.4255184	.0872452	.148936	.678571
avgsen	630	8.95454	2.658082	4.22	25.83
polpc	630	.0019168	.0027349	.0004585	.0355781
density	630	1.386062	1.439703	.1977186	8.827652
taxpc	630	30.23919	11.4547	14.30256	119.7615
pctymle	630	.0889739	.0243493	.0621577	.2743584
wser	630	224.6705	104.8667	1.843794	2177.068

However the data is heterogeneous; therefore, scatter plots for the values of interest are plotted on a logarithmic scale:

```
scatter lcrmrte lprbarr
scatter lcrmrte lprbconv
scatter lcrmrte lprbpris
scatter lcrmrte lavgsen
scatter lcrmrte lpolpc
scatter lcrmrte ldensity
scatter lcrmrte lpctymle
scatter lcrmrte county
scatter lcrmrte year
```



corr lcrmrte lprbarr lprbconv lprbpris lavgsen lpolpc

	lcrmrte	lprbarr	lprbconv	lprbpris	lavgsen	lpolpc
lcrmrte	1.0000					
lprbarr	-0.4498	1.0000				
lprbconv	-0.4569	-0.0973	1.0000			
lprbpris	0.1805	-0.0212	-0.1581	1.0000		
lavgsen	0.0214	-0.0483	-0.0269	0.0029	1.0000	
lpolpc	0.1395	0.0531	0.2730	-0.0560	0.0336	1.0000

If all the observation are pooled together no relation can be seen between lcrmrte and lprbpris, lcrmrte and lavgsen, negative relation lcrmrte and lprbarr, lcrmrte and lprbconv. That is if probability to be arrested or to be convicted is low, then crime is higher. However looking on pairplots is not enough, e.g. probability to be convicted might be correlated to police per capita. Almost no difference can be observed between the lcrmrte grouped by years, however, between the counties the difference can be seen.

(2)

reg lcrmrte lprbarr lprbconv lprbpris lavgsen lpolpc

Source	ss	df	MS	Number of ob	_	630
Model Residual	116.778368 89.6019767	5 624	23.3556736 .143592911		= = = ed =	162.65 0.0000 0.5658 0.5624
Total	206.380345	629	.328108656		=	.37894
lcrmrte	Coef.	Std. Err.	t	P> t [95%	Conf.	Interval]
lprbarr lprbconv lprbpris lavgsen lpolpc _cons	7215113 5492767 .2379716 0652007 .3625234 -2.206729	.0367089 .0262701 .0664302 .0553516 .0299608 .2386927	-19.65 -20.91 3.58 -1.18 12.10 -9.25	0.0007933 0.0006007 0.000 .1073 0.2391738 0.000 .3036 0.000 -2.673	8652 5178 8987 5873	6494234 4976882 .3684254 .0434972 .4213596 -1.73799

If a pooled regression model is chosen to explain lcrmrte, than it is assumed that there are no unique attributes of individuals within the measurement set, and no universal effects across time. Though, these results can't be realistic, because probably there are some common factor for each of the counties.

(3)

Estimate the same model for each separate cross-section.

reg lcrmrte lprbarr lprbconv lprbpris lavgsen lpolpc if year==81 reg lcrmrte lprbarr lprbconv lprbpris lavgsen lpolpc if year==82 reg lcrmrte lprbarr lprbconv lprbpris lavgsen lpolpc if year==83 reg lcrmrte lprbarr lprbconv lprbpris lavgsen lpolpc if year==84 reg lcrmrte lprbarr lprbconv lprbpris lavgsen lpolpc if year==85 reg lcrmrte lprbarr lprbconv lprbpris lavgsen lpolpc if year==86 reg lcrmrte lprbarr lprbconv lprbpris lavgsen lpolpc if year==87

Рисунок	1	 1981
1 110 , 11010	-	1,01

Рисунок 2 — 1982

Source	SS	df	MS	Number	er of obs =	50	Source	SS	df	MS	Number of obs	-	90 21.66
Model Residual	14.4530184 11.449483	5 84	2.8906036 .1363033	8 Prob 7 R-sq		0.0000 0.5580	Model Residual	14.8694631 11.5350724	5 84	2.97389261 .137322291	Prob > F	=	0.0000 0.5631 0.5371
Total	25.9025014	89	.29103934	-	-		Total	26.4045355	89	.296680174		=	.37057
lcrmrte	Coef.	Std. Err.	t	P> t	[95% Conf.	Interval]	lcrmrte	Coef.	Std. Err.	t	P> t [95% Con	f.	Interval]
lprbarr lprbconv lprbpris lavgsen lpolpc _cons	6389785 5742229 .0908728 0389218 .2980306 -2.688402	.0991133 .0678576 .1889373 .1312269 .0885235 .6721831	-6.45 -8.46 0.48 -0.30 3.37 -4.00	0.000 0.000 0.632 0.768 0.001 0.000	8360763 7091653 2848497 299881 .1219918 -4.025112	4418808 4392806 .4665953 .2220374 .4740693 -1.351692	lprbarr lprbconv lprbpris lavgsen lpolpc _cons	6833121 5972914 .2425427 0879718 .3039994 -2.479068	.0994148 .0801288 .1640054 .1524012 .0848964 .6633709	-7.45 1.48 -0.58 3.58	0.0008810094 0.0007566362 0.1430836001 0.5653910384 0.001 .1351736 0.000 -3.798253		4856148 4379466 .5686855 .2150948 .4728252 -1.159882

Рисунок 3 — 1983

Рисунок 4 — 1984

_	Model Residual	19.3942555 10.2648559	5 84	3.87885111 .122200665		= = :ed =	0.6539	Model Residual	16.9919863 13.842251	5 84	3.3983972	3 R-sc) > F quared R-squared	= = =	0.0000 0.5511 0.5244
	Total	29.6591114	89	.333248443	Root MSE	-	.34957	Total	30.8342373	89	.34645210	5 Root	MSE	=	.40594
_															
	lcrmrte	Coef.	Std. Err.	t	P> t [95	Conf.	Interval]	lcrmrte	Coef.	Std. Err.	t	P> t	[95% Conf	. I	nterval]
	lprbarr lprbconv	7654428 4664764	.0835683		0.000931 0.000603	16275 33279	5992581 329625	lprbarr lprbconv	6800545 5770381	.1099577	-6.18 -6.85	0.000	8987175 7446244		. 4613915
	lprbpris	.4981606	.1396416			14678	.7758533	lprbpris	.1064558	.243549	0.44	0.663	3778681		5907798
	lavgsen	1795763	.1517595	-1.18	0.24048	13666	.1222141	lavgsen	0254459	.1926989	-0.13	0.895	4086487		.357757
	lpolpc	.4176287	.0760655	5.49	0.000 .266	3641	. 5688934	lpolpc	.3632997	.0753655	4.82	0.000	.2134272		5131723
	_cons	-1.370831	.6181525	-2.22	0.029 -2.60	00095	1415673	_cons	-2.426434	.6166852	-3.93	0.000	-3.65278	-1	L.200088

Рисунок 5 — 1985

Рисунок 6 — 1986

Model Residual Total	17.3961672 14.0305151 31.4266824	5 84 89	3.47923345 .167029942 .353108791	R-sq Adj) > F = quared = R-squared =	= 0.0000 = 0.5535 = 0.5270 = .40869	Model Residual Total	22.4219479 11.0642858 33.4862337	5 84 89	4.4843895 .13171768	8 R-squared - Adj R-squar	ed	= 34.05 = 0.0000 = 0.6696 = 0.6499 = .36293
lcrmrte	Coef.	Std. Err.	t	P> t	[95% Conf.	. Interval]	lcrmrte	Coef.	Std. Err.	t	P> t [95	Conf	. Interval]
lprbarr lprbconv lprbpris lavgsen lpolpc _cons	842586 5112046 .1702136 0375874 .3760042 -2.399902	.1130809 .0660793 .2158408 .1749804 .0810506 .6398518	-7.74 0.79 -0.21 4.64	0.000 0.000 0.433 0.830 0.000	-1.06746 6426105 2590095 38555552 .2148262 -3.672318	6177123 3797988 .5994366 .3103804 .5371821 -1.127487	lprbarr lprbconv lprbpris lavgsen lpolpc _cons	6944479 647814 .2807411 1259598 .4759127 -1.33185	.0985755 .0675528 .1641958 .1490572 .0724661 .6372348	-7.04 -9.59 1.71 -0.85 6.57 -2.09	0.000890 0.000782 0.091043 0.400422 0.000 .333 0.040 -2.59	21501 57802 23763 18059	4984196 5134779 .6072625 .1704568 .6200195 0646389

Рисунок 7 — 1987

Source	ss	df	MS		er of obs	; = =	90 15.62
Model Residual	12.9102805 13.8894211	5 84	2.5820561 .16535025	l Prok L R-so	84) > F quared	=	0.0000
Total	26.7997016	89	.30112024	-	R-squared : MSE	l = =	0.4509 .40663
lcrmrte	Coef.	Std. Err.	t	P> t	[95% C	onf.	Interval]
lprbarr lprbconv lprbpris lavgsen lpolpc _cons	7339084 4340935 .1307631 1456424 .4141207 -1.708116	.1093422 .0796492 .1958738 .1692714 .1271017	-6.71 -5.45 0.67 -0.86 3.26 -1.62	0.000 0.000 0.506 0.392 0.002 0.108	95134 59248 25875 48225 .16136 -3.8011	i47 i33 i72 i52	5164696 2757023 .5202796 .1909723 .6668763

In 1981, 1982, 1984, 1986, 1987 lprbpris is insignificant (on 5% level), though in pooled regression it has a significant negative effect.

(4)

xtreg lcrmrte lprbarr lprbconv lprbpris lavgsen lpolpc, be

Between regres	eeion (raarae	ion on arou	n maanal	Number	of obs	=	630
1		oron on grou	p mcano,				
Group variable	e: county			Number	of groups	=	90
R-sq:				Obs per	group:		
within =	0.0405				mir	1 =	7
between =	0.7099				avo	1 =	7.0
overall =	- 0.4888				max	. =	7
				F(5,84)		=	41.11
sd(u i + avg(e	e i.))= .3048	3712		Prob >	F	=	0.0000
	94	00 d - 7		Polos	ross a		
lcrmrte	Coei.	Std. Err.	t	P> t	[95% Co	oni.	Interval]
lprbarr	8128525	.0925787	-8.78	0.000	996955	54	6287496
lprbconv	5920724	.070137	-8.44	0.000	731547	76	4525973
lprbpris	1.16067	.237592	4.89	0.000	.688192	26	1.633148
lavgsen	1312699	.2085175	-0.63	0.531	545929	98	. 28339
1	l	.0716332			0000		407440
lpolpc	.3446985	.0/10332	4.81	0.000	. 20224	ŀВ	.487149
_cons	.3446985 -1.515489	.7063476	-2.15	0.000	-2.92013		

In "between" regression in comparison to pooled lprbpris has much stronger effect, and in comparison to most of the regressions on cross-section data this effect is significant. Other significant coefficient estimates are roughly the same.

(5)

xtreg lcrmrte lprbarr lprbconv lprbpris lavgsen lpolpc, fe

Number o	f groups = group:	90
Obs per (group:	
	min =	7
	avg =	7.0
	max =	7
F (5,535)	=	59.92
Prob > F	=	0.0000
P> t	[95% Conf.	Interval]
0.000	4492801	3177938
0.000	3489133	2630381
0.000	2609915	1299116
0.173	0156552	.0869838
0.000	.3598113	.467731
0.000	-2.212566	-1.53315
	Prob > F P> t 0.000 0.000 0.000 0.173 0.000	max = F(5,535) = Prob > F = P> t [95% Conf. 0.0004492801 0.0003489133 0.0002609915 0.1730156552 0.000 .3598113

In fixed effects model in comparison to between-estimates the estimates of lprbarr and lprbconv are still significant, but are almost twice smaller by absolute value. The estimate of lprbpris is significant, but in comparison to between-estimates it becomes negative. Cornwell and Trumbull concluded that cross-section estimates do not account for an individual effect, therefore, are misleading.

(6)

To test the fixed effects model against the pooled regression model, F-test is used. The test is significant, i.e., hypothesis that all counties' effects are equal to 0 is rejected, therefore, the fixed effects model is preferred over pooled regression.

(7)

```
xtreg lcrmrte lprbarr lprbconv lprbpris lavgsen lpolpc, fe
predict uu, u
local list "lcrmrte lprbarr lprbconv lprbpris lavgsen lpolpc"

foreach xx of local list {
   display "`xx'"
   xtreg `xx', fe
   predict mm, xb
   predict u1, u
   gen tm_`xx' = u1+mm
   drop u1 mm
}
```

reg uu west central urban pctmin80 tm_lprbarr tm_lprbconv tm_lprbpris
tm_lavgsen tm_lpol

Source	SS	df	MS	Number of ol	os =	630
				- F(9, 620)	=	208.40
Model	85.5175906	9	9.50195451	1 Prob > F	=	0.0000
Residual	28.2688545	620	.045594927	7 R-squared	=	0.7516
				- Adj R-squar	ed =	0.7480
Total	113.786445	629	.180900549	9 Root MSE	=	.21353
	'					
uu	Coef.	Std. Err.	t	P> t [95%	Conf.	Interval]
west	3159483	.0341209	-9.26	0.000382	9549	2489418
central	0567069	.0227318	-2.49	0.0131013		0120662
	.1332446	.0369903	3.60		0603	.2058861
urban						
pctmin80	.0062172	.0008046	7.73	0.000 .004		.0077974
tm_lprbarr	4301418	.0269604	-15.95	0.000483	0866	3771969
tm_lprbconv	3082678	.0206731	-14.91	0.000348	3656	26767
tm lprbpris	.8765429	.0677316	12.94	0.000 .743	5319	1.009554
tm lavgsen	2785488	.0576916	-4.83	0.000391	8434	1652543
tm_lpolpc	0551223	.0199678	-2.76	0.00609	1335	0159096
_cons	.173909	.1985676	0.88	0.381216	0375	. 5638555

In western and central North Carolina crime rate is lower, in urban counties and counties with higher minority percentage – higher.

(8)

xtreg lcrmrte lprbarr lprbconv lprbpris lavgsen lpolpc d8*, fe test d82 d83 d84 d85 d86 d87

By F-test we conclude that dummy variables are significant, therefore, it can be useful to account for time dummies.

(9)

xtreg lcrmrte lprbarr lprbconv lprbpris lavgsen lpolpc d8*
ldensity lwcon lwtuc lwtrd lwfir lwser lwmfg lwfed lwsta lwloc
lpctymle ltaxpc lmix, fe

Fixed-effects (within) regression Group variable: county	Number of obs Number of groups	=	630 90
R-sq:	Obs per group:		
within = 0.4635	min	L =	7
between = 0.5534	avo	r =	7.0
overall = 0.5444	max	=	7
	F(24,516)	=	18.57
corr(u_i, Xb) = -0.2233	Prob > F	=	0.0000

lcrmrte	Coef.	Std. Err.	t	P> t	[95% Conf.	. Interval]
lprbarr	3553352	.0331411	-10.72	0.000	4204433	2902273
lprbconv	28167	.0212032	-13.28	0.000	3233252	2400149
lprbpris	1725937	.0324023	-5.33	0.000	2362504	108937
lavgsen	0022518	.0261755	-0.09	0.931	0536756	.0491719
lpolpc	.4125651	.0269519	15.31	0.000	.3596162	.4655139
d82	.0220252	.0261164	0.84	0.399	0292824	.0733327
d83	041761	.0359918	-1.16	0.246	1124694	.0289474
d84	0457774	.0471339	-0.97	0.332	1383753	.0468203
d85	0204236	.0646682	-0.32	0.752	1474689	.1066217
d86	.0300543	.0795825	0.38	0.706	1262912	.1863999
d87	.0939942	.0948951	0.99	0.322	092434	. 2804225
ldensity	.4039363	.2851002	1.42	0.157	1561635	.9640361
lwcon	0378999	.0391845	-0.97	0.334	1148807	. 039081
lwtuc	.0457869	.0191931	2.39	0.017	.0080806	.0834932
lwtrd	0206108	.0405906	-0.51	0.612	100354	.0591324
lwfir	0038305	.0283148	-0.14	0.892	0594569	.051796
lwser	.0086382	.0192507	0.45	0.654	0291811	.0464573
lwmfg	3587417	.112218	-3.20	0.001	579202	1382814
lwfed	3082912	.1768173	-1.74	0.082	6556614	.0390791
lwsta	.0494109	.1144341	0.43	0.666	1754031	.2742249
lwloc	.1822251	.1180615	1.54	0.123	0497152	.4141655
lpctymle	.6235668	.3647114	1.71	0.088	0929351	1.340069
ltaxpc	.0144755	.0448261	0.32	0.747	0735885	.1025395
lmix	.0005681	.0151308	0.04	0.970	0291575	.0302936
_cons	2.340509	1.689701	1.39	0.167	9790297	5.660047
sigma u	.37754187					
sigma e	.13677145					
rho	.88398703	(fraction	of varia	nce due t	oui)	

F test that all $u_i=0$: F(89, 516) = 32.82

Prob > F = 0.0000

With all the time-varying regressors lprbpris is again significant and negative.

Coefficient of lpctymle is positive and significant, therefore, if crime rate is positively dependent on percentage of young male, the government should work on programs to lower unemployment rate for this group, provide them with education programs, etc.

Higher wages for tranport, utility and communal servives correspond to higher crime rates (this may be caused by the fact that this type of job doen't require a university degree but specific skills are needed, so if wages in this sector are higher the competition is also higher; therefore, some groups, especially the ones in need, can't get the job and might commit crime), higher wages for manufacturing lowers crime rate (this job is quite hard and in most cases not payed as well, so in counties where workers are underpayed more crimes are committed).

(10)

xtreg lcrmrte lprbarr lprbconv lprbpris lavgsen lpolpc, re

Random-effects GLS regression				Number	of obs	=	630
Group variable: county				Number	of groups	=	90
R-sq:				Obs per	group:		
within =	0.3538				min	. =	7
between = 0.4730				avg =			7.0
overall = 0.4572					max	=	7
				Wald ch	(12 (5)	=	374.30
corr(u i V)	= 0 (assumed	15		Prob >		=	0.0000
corr(u_i, X)	- u (assumed	.,		FLOD >	CIIIZ	_	0.0000
	·						
lcrmrte	Coef.	Std. Err.	z	P> z	[95% Co	nf.	Interval]
lprbarr	4485975	.0326419	-13.74	0.000	512574	5	3846205
lprbconv	3469171	.0214454	-16.18	0.000	388949	3	3048848
lprbpris	1876919	.0348083	-5.39	0.000	255914	8	1194689
lavgsen	.0276295	.0274935	1.00	0.315	026256	8	.0815157
lpolpc	.4184813	.0269885	15.51	0.000	. 365584	7	.4713779
_cons	-1.92944	.177319	-10.88	0.000	-2.27697	9	-1.581901
adama u	. 2997784						
sigma_u	.14681793						
sigma_e		/ 					
rho	.80654279	(fraction	of varia	nce due t	o u_1)		

Random effects model assuming that the unique, time constant attributes of individuals that are not correlated with the individual regressors. The same regressors have significant estimates of their coefficients, with the same signs as in FE-model.

(11)

xttest0

To test the random effects model in comparison to pooled regression model Breusch-Pagan LM-test is used, where the null is is that the variance of the unobserved fixed effects is zero. The null is rejected therefore RE should be prefered over pooled regression.

```
xtreg lcrmrte lprbarr lprbconv lprbpris lavgsen lpolpc, re
est store RE
xtreg lcrmrte lprbarr lprbconv lprbpris lavgsen lpolpc, fe
est store FE
hausman FE RE, sigmamore
```

	Coeffi	cients ——						
	(b)	(B) (b-B)		sqrt(diag(V_b-V_B))				
	FE	RE	Difference	S.E.				
lprbarr	3835369	4485975	. 0650606	.0139921				
lprbconv	3059757	3469171	.0409414	.0088373				
lprbpris	1954515	1876919	0077597	.006471				
lavgsen	.0356643	.0276295	.0080348	.003558				
lpolpc	.4137712	.4184813	0047101	.011013				
В	b = consistent under Ho and Ha; obtained from xtreg B = inconsistent under Ha, efficient under Ho; obtained from xtreg							
Test: Ho:	Test: Ho: difference in coefficients not systematic							
chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B) = 74.31								
Prob>chi2 = 0.0000								

Test the random effects model in comparison to fixed effects model Hausman test is used. The unique attributes of individuals may or may not be correlated with the individual dependent variables. Under null hypothesis individual effects and the regressors are uncorrelated. In this case it is rejected; therefore, RE-estimates are inconsistent and FE is prefered over RE.