

Problem 1

Consider the fixed effect model

$$y_{it} = \alpha_i + u_{it}$$

The disturbances are supposed to be independent and heteroscedastic: $V(u_{it}) = \sigma_i^2$. Data are unbalanced, with different number of observations for each industry.

(a) Show that OLS and GLS estimates of α_i are identical.

(b) Let $\sigma^2 = \sum_{i=1}^N T_i \frac{\sigma_i^2}{n}$, $n = \sum_{i=1}^N T_i$ be the average disturbance variance. Show that the OLS estimator of σ^2 is biased. Also show that that bias disappears if the data are balanced or the variances are homoscedastic.

Solution

(a) OLS: $\sum_{i=1}^N \sum_{t=1}^{T_i} (y_{it} - \alpha_i)^2 \xrightarrow{\alpha_i} \min$ hence, $\hat{\alpha}_{i,OLS} = \bar{y}_i$.

GLS: $\sum_{i=1}^N \sum_{t=1}^{T_i} \frac{1}{\sigma_i^2} (y_{it} - \alpha_i)^2 \xrightarrow{\alpha_i} \min$, hence $\hat{\alpha}_{i,GLS} = \bar{y}_i$.

(b) OLS: $\hat{\sigma}_{OLS}^2 = \frac{1}{n-N} \sum_{i=1}^N \sum_{t=1}^{T_i} (y_{it} - \bar{y}_i)^2$; $E(\hat{\sigma}_{OLS}^2) = \frac{1}{n-N} \sum_{i=1}^N \sum_{t=1}^{T_i} E(y_{it} - \bar{y}_i)^2 = \frac{1}{n-N} \sum_{i=1}^N (T_i - 1) \sigma_i^2 =$
 $= \frac{n}{n-N} \sum_{i=1}^N \frac{(T_i - 1)}{T_i} T_i \frac{\sigma_i^2}{n} \neq \sum_{i=1}^N T_i \frac{\sigma_i^2}{n}$.

If $T_i = T$ then $\sigma^2 = \sum_{i=1}^N T \frac{\sigma_i^2}{n} = \frac{T}{n} \sum_{i=1}^N \sigma_i^2 = \frac{1}{N} \sum_{i=1}^N \sigma_i^2$. And

$$E(\hat{\sigma}_{OLS}^2) = \frac{n}{n-N} \sum_{i=1}^N (T-1) \frac{\sigma_i^2}{n} = \frac{NT}{NT-N} \sum_{i=1}^N (T-1) \frac{\sigma_i^2}{NT} = \frac{NT}{NT-N} \frac{T-1}{NT} \sum_{i=1}^N \sigma_i^2 = \frac{1}{N} \sum_{i=1}^N \sigma_i^2 = \sigma^2.$$

Problem 2**Estimating the Economic Model of Crime with Panel Data**

Becker (1968) introduced an economic model explaining the number of crimes. The main implication of this model is that the number of crimes depends negatively on the probability to be arrested, the probability to be convicted conditional on being arrested, the probability to be imprisoned conditional on being convicted, and the average length of the imprisonment sentence. Since 1968, many empirical studies have tested the empirical implications of Becker's model, usually with cross-section data. Cornwell and Trumbull (1994) use panel data and their results suggest that the cross-section based estimates can be misleading. In this set of exercises, we will use some standard panel data models for an analysis similar to that of Cornwell and Trumbull. The Cornwell and Trumbull study is also briefly described as an example in Wooldridge's textbook (Wooldridge, 2000, p. 432-433), and the data set is taken from the web site of this textbook.

Data

The data set CRIME4.DTA contains data on 90 counties in North Carolina covering the years 1981 till 1987. The data are stored in ASCII format. The variables are as follows, stored in the order given below:

1. county	county identifier
2. year	81 to 87
3. crmrte	crimes committed per person
4. prbarr	'probability' of arrest
5. prbconv	'probability' of conviction
6. prbpris	'probability' of prison sentence
7. avgsen	avg. sentence, days
8. polpc	police per capita
9. density	people per sq. mile
10. taxp	tax revenue per capita
11. west	=1 if in western N.C.
12. central	=1 if in central N.C.

13. urban	=1 if in SMSA
14. pctmin80	perc. minority, 1980
15. wcon	weekly wage, construction
16. wtuc	wkly wge, trns, util, commun
17. wtrd	wkly wge, whlesle, retail trade
18. wfir	wkly wge, fin, ins, real est
19. wser	wkly wge, service industry
20. wmfgr	wkly wge, manufacturing
21. wfed	wkly wge, fed employees
22. wsta	wkly wge, state employees
23. wloc	wkly wge, local gov emps
24. mix	offense mix: face-to-face/other
25. pctymle	percent young male
26. d82	=1 if year == 82
27. d83	=1 if year == 83
28. d84	=1 if year == 84
29. d85	=1 if year == 85
30. d86	=1 if year == 86
31. d87	=1 if year == 87
32. lcrmte	log(crmte)
33. lprbarr	log(prbarr)
34. lprbconv	log(prbconv)
35. lprbpris	log(prbpris)
36. lavgsen	log(avgsen)
37. lpolpc	log(polpc)
38. ldensity	log(density)
39. ltaxpc	log(taxpc)
40. lwcon	log(wcon)
41. lwtuc	log(wtuc)
42. lwtrd	log(wtrd)
43. lwfir	log(wfir)
44. lwser	log(wser)
45. lwmfgr	log(wmfgr)
46. lwfed	log(wfed)
47. lwsta	log(wsta)
48. lwloc	log(wloc)
49. lmix	log(mix)
50. lpctymle	log(pctymle)
51. lpctmin	log(pctmin)
52. clcrmte	lcrmte - lcrmte[t-1]
53. clprbarr	lprbarr - lprbarr[t-1]
54. clprbconv	lprbconv - lprbconv[t-1]
55. clprbpri	lprbpri - lprbpri[t-1]
56. clavgsen	lavgsen - lavgsen[t-1]
57. clpolpc	lpolpc - lpolpc[t-1]
58. cltaxpc	ltaxpc - ltaxpc[t-1]
59. clmix	lmix - lmix[t-1]

The probability of arrest is an estimated probability, obtained as the ratio of the number of arrests and the reported number of crimes. Similarly, the probabilities of conviction is estimated by the ratio of the number of convictions and the number of arrests, and the probability of prison sentence is estimated as the ratio of the number of people sent to prison and the number of convictions.

Exercises

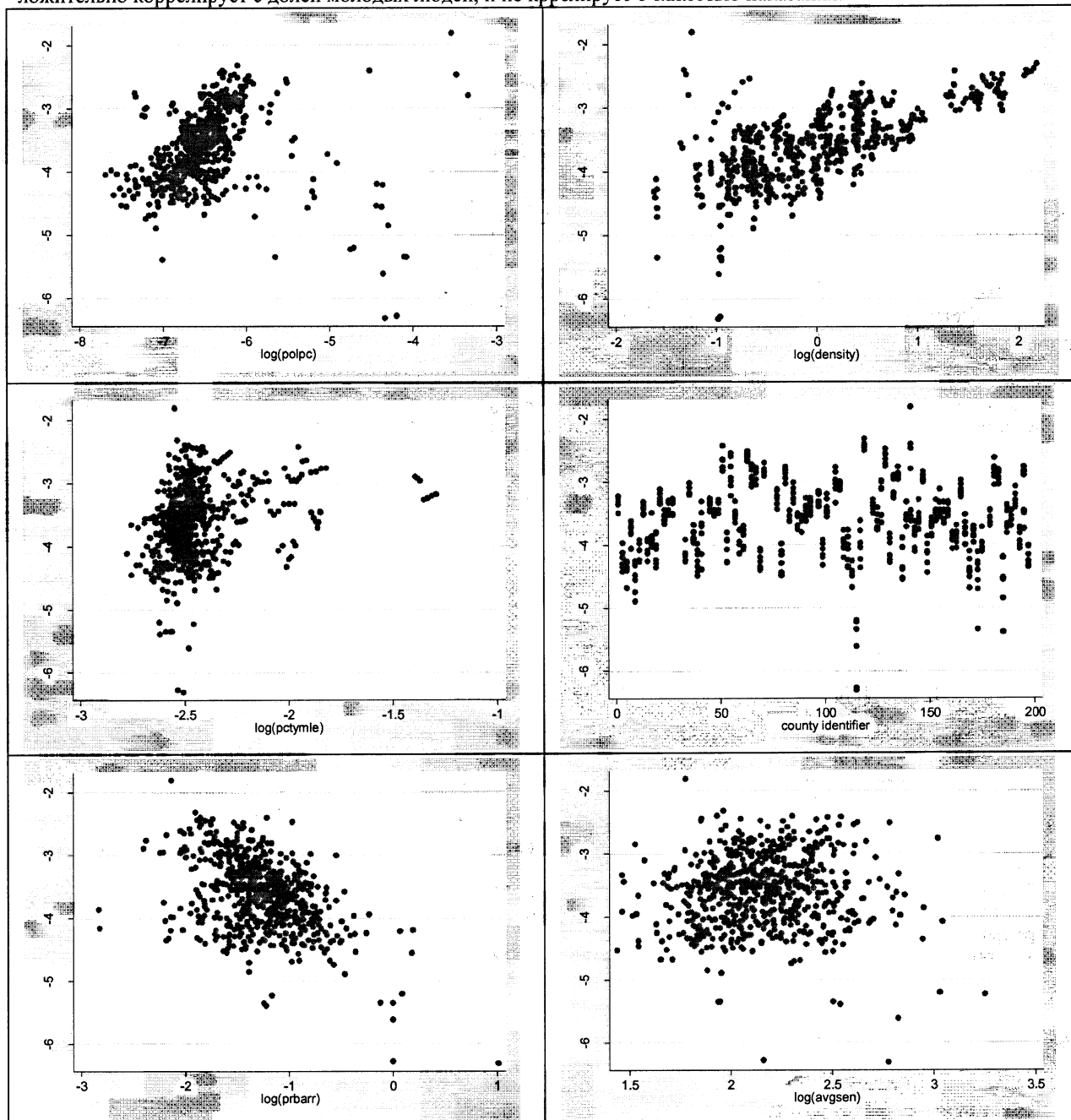
1. Read the data and present some descriptive statistics on the variables of main interest. Discuss some findings that you think are interesting.
2. Estimate a pooled regression model explaining lcrmte from an intercept and the variables lprbarr, lprbconv, lprbpris, lavgsen and lpolpc. Briefly discuss the results. Which assumptions have you made? Do you think these assumptions are realistic?
3. Estimate the same model for each separate cross-section. Compare the results with those of the previous exercise.
4. Construct the means over time for each county of the dependent and independent variables in the regression in the previous exercise. Carry out the “between groups” regression on these individual means. Discuss the results and compare them with those of the previous exercises.
5. Estimate a fixed effects model with the same dependent and independent variables as in the previous model. Discuss the results and compare them to those of the previous exercises. Explain why Cornwell and Trumbull concluded that cross-section estimates are misleading.
6. Test the fixed effects model against the pooled regression model.
7. Regress the estimated county specific effects on the means of the regressors included in the previous exercises and on the (time invariant) variables WEST, CENTRAL, URBAN, and PCTMIN80. Discuss the results.

8. Test whether it is useful to add time dummies to the fixed effects model.
9. Go through the other (time varying) variables available in the data set and select some that in your opinion could help to explain the number of crimes. Include them in the fixed effects model and test whether they are significant. In this way, select your favorite model and discuss the policy implications of the results.
10. Estimate a random effects model assuming that the random effects are independent of the regressors. Discuss the results.
11. Test the random effects model in the previous exercise against the fixed effects model and against the pooled regression model.

Solution

1. `gen lpolpc1=lpolpc- clpolpc`

Из графиков можно заключить, что количество преступлений положительно коррелирует с плотностью населения, количеством полиции (?), отрицательно с вероятностью ареста, сильно варьирует в зависимости от района, слабо положительно коррелирует с долей молодых людей, и не коррелирует с тяжестью наказания.



Матрица корреляций:

corr	lcrmte	lprbarr	lprbconv	lprbpris	lavgsen	lpolpc
	lcrmte	lprbarr	lprbconv	lprbpris	lavgsen	lpolpc
lcrmte	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

```

2.

Неотвратимость наказания (lprbarr, lprbconv) снижает преступность, тяжесть наказания (lavgsen) не значима, другой показатель неотвратимости наказания (lprbpris) положительно коррелирует с преступностью, полиция повышает (?) преступность (тут видимо есть эндогенность – при более высокой преступности возрастают расходы местного бюджета на полицию). Предполагается отсутствие индивидуального эффекта, что видимо, не вполне корректно из-за неоднородности районов.

```

reg lcrmte lprbarr lprbconv lprbpris lavgsen lpolpc
Prob > F = 0.0000
R-squared = 0.5658
Adj R-squared = 0.5624

```

lcrmte	Coef.	Std. Err.	t	P> t
lprbarr	-.7215113	.0367089	-19.65	0.000
lprbconv	-.5492767	.0262701	-20.91	0.000
lprbpris	.2379716	.0664302	3.58	0.000
lavgsen	-.0652007	.0553516	-1.18	0.239
lpolpc	.3625234	.0299608	12.10	0.000
_cons	-2.206729	.2386927	-9.25	0.000

3.

Есть некоторые вариации коэффициентов по годам по годам, но не существенные
Lavgsen и часто lprbpris — незначимы

1981

lcrmte	Coef.	Std. Err.	t	P> t
lprbarr	-.6389785	.0991133	-6.45	0.000
lprbconv	-.5742229	.0678576	-8.46	0.000
lprbpris	.0908728	.1889373	0.48	0.632
lavgsen	-.0389218	.1312269	-0.30	0.768
lpolpc	.2980306	.0885235	3.37	0.001
_cons	-2.688402	.6721831	-4.00	0.000

1982

lcrmte	Coef.	Std. Err.	t	P> t
lprbarr	-.6833121	.0994148	-6.87	0.000
lprbconv	-.5972914	.0801288	-7.45	0.000
lprbpris	.2425427	.1640054	1.48	0.143
lavgsen	-.0879718	.1524012	-0.58	0.565
lpolpc	.3039994	.0848964	3.58	0.001
_cons	-2.479068	.6633709	-3.74	0.000

1983

lcrmte	Coef.	Std. Err.	t	P> t
lprbarr	-.7654428	.0835683	-9.16	0.000
lprbconv	-.4664764	.0688177	-6.78	0.000
lprbpris	.4981606	.1396416	3.57	0.001
lavgsen	-.1795763	.1517595	-1.18	0.240
lpolpc	.4176287	.0760655	5.49	0.000
_cons	-1.370831	.6181525	-2.22	0.029

1984

lcrmte	Coef.	Std. Err.	t	P> t
lprbarr	-.6800545	.1099577	-6.18	0.000
lprbconv	-.5770381	.0842731	-6.85	0.000
lprbpris	.1064558	.243549	0.44	0.663
lavgsen	-.0254459	.1926989	-0.13	0.895
lpolpc	.3632997	.0753655	4.82	0.000
_cons	-2.426434	.6166852	-3.93	0.000

1985

lcrmte	Coef.	Std. Err.	t	P> t
--------	-------	-----------	---	------

lprbarr	-.842586	.1130809	-7.45	0.000
lprbconv	-.5112046	.0660793	-7.74	0.000
lprbpris	.1702136	.2158408	0.79	0.433
lavgsen	-.0375874	.1749804	-0.21	0.830
lpolpc	.3760042	.0810506	4.64	0.000
_cons	-2.399902	.6398518	-3.75	0.000

1986

lcrmte	Coef.	Std. Err.	t	P> t
lprbarr	-.6944479	.0985755	-7.04	0.000
lprbconv	-.647814	.0675528	-9.59	0.000
lprbpris	.2807411	.1641958	1.71	0.091
lavgsen	-.1259598	.1490572	-0.85	0.400
lpolpc	.4759127	.0724661	6.57	0.000
_cons	-1.33185	.6372348	-2.09	0.040

1987

lcrmte	Coef.	Std. Err.	t	P> t
lprbarr	-.7339084	.1093422	-6.71	0.000
lprbconv	-.4340935	.0796492	-5.45	0.000
lprbpris	.1307631	.1958738	0.67	0.506
lavgsen	-.1456424	.1692714	-0.86	0.392
lpolpc	.4141207	.1271017	3.26	0.002
_cons	-1.708116	1.052529	-1.62	0.108

Однако в 1984 уровень преступности был несколько ниже чем в 1981 (на 10%-ном уровне), остальные года эквивалентны 1981-му.

lcrmte	Coef.	Std. Err.	t	P> t
lprbarr	-.7195033	.0367657	-19.57	0.000
lprbconv	-.5456588	.0263683	-20.69	0.000
lprbpris	.2475521	.0672268	3.68	0.000
lavgsen	-.0867575	.0579205	-1.50	0.135
lpolpc	.3659887	.0300252	12.19	0.000
d82	.0051371	.057931	0.09	0.929
d83	-.043503	.0576243	-0.75	0.451
d84	-.1087542	.057923	-1.88	0.061
d85	-.0780453	.0583244	-1.34	0.181
d86	-.0420791	.0578218	-0.73	0.467
d87	-.0270426	.056899	-0.48	0.635
_cons	-2.082293	.2516253	-8.28	0.000

4.

xtreg lcrmte lprbarr lprbconv lprbpris lavgsen lpolpc, be

lcrmte	Coef.	Std. Err.	t	P> t
lprbarr	-.8128525	.0925787	-8.78	0.000
lprbconv	-.5920724	.070137	-8.44	0.000
lprbpris	1.16067	.237592	4.89	0.000
lavgsen	-.1312699	.2085175	-0.63	0.531
lpolpc	.3446985	.0716332	4.81	0.000
_cons	-1.515489	.7063476	-2.15	0.035

Lprbpris — значима, положительный коэффициент, больше, чем в других регрессиях

5.

xtreg lcrmte lprbarr lprbconv lprbpris lavgsen lpolpc, fe

lcrmte	Coef.	Std. Err.	t	P> t
lprbarr	-.3835369	.0334672	-11.46	0.000
lprbconv	-.3059757	.0218578	-14.00	0.000
lprbpris	-.1954515	.0333637	-5.86	0.000
lavgsen	.0356643	.0261247	1.37	0.173
lpolpc	.4137712	.0274687	15.06	0.000
_cons	-1.872858	.1729314	-10.83	0.000

```

sigma_u | .42736725
sigma_e | .14681793
rho | .89443836 (fraction of variance due to u_i)

```

```

-----
F test that all u_i=0:      F(89, 535) =      40.69      Prob > F = 0.0000

```

6.

F-тест показывает наличие индивидуального эффекта. При его учете коэффициент при `lprbpris` становится отрицательным, как и при других показателях неотвратимости наказания.

7.

```

xtreg lcrmte lprbarr lprbconv lprbpris lavgsen lpolpc, fe
predict uu, u
local list "lcrmte 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_lpolpc

```

uu	Coef.	Std. Err.	t	P> t
west	-.3159483	.0341209	-9.26	0.000
central	-.0567069	.0227318	-2.49	0.013
urban	.1332446	.0369903	3.60	0.000
pctmin80	.0062172	.0008046	7.73	0.000
tm_lprbarr	-.4301418	.0269604	-15.95	0.000
tm_lprbconv	-.3082678	.0206731	-14.91	0.000
tm_lprbpris	.8765429	.0677316	12.94	0.000
tm_lavgsen	-.2785488	.0576916	-4.83	0.000
tm_lpolpc	-.0551223	.0199678	-2.76	0.006
_cons	.173909	.1985676	0.88	0.381

На Западе и в Центре преступность ниже, в городах и районах с высокой долей меньшинств – выше.

8.

```

xtreg lcrmte lprbarr lprbconv lprbpris lavgsen lpolpc d8*, fe

```

lcrmte	Coef.	Std. Err.	t	P> t
lprbarr	-.3597945	.0324192	-11.10	0.000
lprbconv	-.2858734	.0212173	-13.47	0.000
lprbpris	-.1827813	.0324611	-5.63	0.000
lavgsen	-.0044879	.0264471	-0.17	0.865
lpolpc	.4241144	.0263661	16.09	0.000
d82	.0125802	.0215416	0.58	0.559
d83	-.0792812	.0213399	-3.72	0.000
d84	-.1177281	.0216145	-5.45	0.000
d85	-.1119561	.0218459	-5.12	0.000
d86	-.0818268	.0214266	-3.82	0.000
d87	-.0404705	.0210392	-1.92	0.055
_cons	-1.604135	.1685739	-9.52	0.000

sigma_u	.43487414
sigma_e	.13871214
rho	.90765322 (fraction of variance due to u_i)

```

-----
F test that all u_i=0:      F(89, 529) =      45.87      Prob > F = 0.0000

```

```

test d82 d83 d84 d85 d86 d87

```

- (1) d82 = 0
- (2) d83 = 0
- (3) d84 = 0
- (4) d85 = 0
- (5) d86 = 0
- (6) d87 = 0

```

F( 6, 529) = 11.73
Prob > F = 0.0000

```

Фиктивные переменные по времени значимы.

9.

```
xtreg lcrmte lprbarr lprbconv lprbpris lavgsen lpolpc d8* ldensity lwcon lwtuc lwtrd lwfir
      lwser l wmf g lwfed lwsta lwloc lpctymle ltaxpc lmix, fe
```

lcrmte	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lprbarr	-.3553352	.0331411	-10.72	0.000	-.4204433	-.2902272
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	.0468205
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	.0464575
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 variance due to u_i)				

F test that all u_i=0: F(89, 516) = 32.82 Prob > F = 0.0000

Значимы зарплаты

wtuc (wkly wge, trns, util, commun) –увеличивает и wmf g (wkly wge, manufacturing) –уменьшает преступность

10.11

```
xtreg lcrmte lprbarr lprbconv lprbpris lavgsen lpolpc, re
```

lcrmte	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lprbarr	-.4485975	.0326419	-13.74	0.000	-.5125745	-.3846205
lprbconv	-.3469171	.0214454	-16.18	0.000	-.3889493	-.3048848
lprbpris	-.1876919	.0348083	-5.39	0.000	-.2559148	-.1194689
lavgsen	.0276295	.0274935	1.00	0.315	-.0262568	.0815157
lpolpc	.4184813	.0269885	15.51	0.000	.3655847	.4713779
_cons	-1.92944	.177319	-10.88	0.000	-2.276979	-1.581901
sigma_u	.2997784					
sigma_e	.14681793					
rho	.80654279	(fraction of variance due to u_i)				

xttest0

Breusch and Pagan Lagrangian multiplier test for random effects

$$lcrmte[country,t] = Xb + u[country] + e[country,t]$$

Estimated results:

	Var	sd = sqrt(Var)
lcrmte	.3281087	.5728077
e	.0215555	.1468179
u	.0898671	.2997784

Test: Var(u) = 0

chibar2(01) = 933.68
Prob > chibar2 = 0.0000

Тест Бреуш-Пагана отвергает Pooled в пользу RE

```
xtreg lcrmte lprbarr lprbconv lprbpris lavgsen lpolpc, re
est store RE
xtreg lcrmte lprbarr lprbconv lprbpris lavgsen lpolpc, fe
```

```
est store FE
hausman FE RE, sigmamore
```

---- Coefficients ----				
	(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

b = consistent under Ho and Ha; obtained from xtreg
 B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B)
 = 74.31
 Prob>chi2 = 0.0000

Тест Бреуш-Пагана отвергает RE в пользу FE