ARE 256B Applied Econometrics
Home Assignment 4

ZHIRAN QIN
Student ID: 917910940
Cooperate with: Yuhan, Heyin
March 13, 2020

1. Gun Control and Violence: Pooled OLS, Random Effects or Fixed Effects

(a)

I choose to include *density*, *avginc*, *pop* and *pm1029* as control variables. Suspects may prefer population dense areas since there are more people for them to choose as victims; the reason to include population is similar. For average income, lower average income areas may have more crimes. For young male people, they may be more aggressive than others and have a higher tendency to commit a crime.

I don't want to include *incarc_rate*, since it is simultaneously determined with crime rates. Including it would cause a problem of endogeneity. I don't include *pw1064* and *pb1064* because differences between racial groups could be explained by other variables, and races have no direct influence on crime rates.

The regression equation is

$$\begin{aligned} vio_{it} &= \beta_0 + \beta_1 shall_{it} + \beta_2 year_{it} + \beta_3 density_{it} + \beta_4 avginc_{it} + \beta_5 pop_{it} + \beta_6 pm1029_{it} + \epsilon_{it} \\ rob_{it} &= \beta_0 + \beta_1 shall_{it} + \beta_2 year_{it} + \beta_3 density_{it} + \beta_4 avginc_{it} + \beta_5 pop_{it} + \beta_6 pm1029_{it} + \epsilon_{it} \\ mur_{it} &= \beta_0 + \beta_1 shall_{it} + \beta_2 year_{it} + \beta_3 density_{it} + \beta_4 avginc_{it} + \beta_5 pop_{it} + \beta_6 pm1029_{it} + \epsilon_{it} \end{aligned}$$

, and the regression outcomes are shown in Table 1.

(b)

The regression equation is

$$\begin{aligned} vio_{it} &= \beta_0 + \beta_1 shall_{it} + \beta_2 year_{it} + \beta_3 density_{it} + \beta_4 avginc_{it} + \beta_5 pop_{it} + \beta_6 pm1029_{it} + \alpha_i + \epsilon_{it} \\ rob_{it} &= \beta_0 + \beta_1 shall_{it} + \beta_2 year_{it} + \beta_3 density_{it} + \beta_4 avginc_{it} + \beta_5 pop_{it} + \beta_6 pm1029_{it} + \alpha_i + \epsilon_{it} \\ mur_{it} &= \beta_0 + \beta_1 shall_{it} + \beta_2 year_{it} + \beta_3 density_{it} + \beta_4 avginc_{it} + \beta_5 pop_{it} + \beta_6 pm1029_{it} + \alpha_i + \epsilon_{it} \end{aligned}$$

, and the regression outcomes are shown in Table 2.

Table 1: Pooled OLS Regression

	(1) vio	(2) rob	(3) mur
shall	-112.5***	-21.22***	-1.685***
	(16.48)	(4.922)	(0.238)
year	8.071***	-3.607***	0.238***
	(2.173)	(0.767)	(0.0395)
density	164.0***	95.52***	4.445***
	(10.64)	(5.791)	(0.503)
avginc	4.553	7.342***	-0.315***
	(3.317)	(1.347)	(0.0811)
pop	21.52***	11.25***	0.256***
	(1.356)	(0.706)	(0.0233)
pm1029	4.388	-8.088**	0.790***
	(7.488)	(2.486)	(0.160)
_cons	-474.2	425.9***	-24.01***
	(289.7)	(97.01)	(5.694)
N	1173	1173	1173
adj. R^2	0.608	0.768	0.605

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Random-effects Regression

	(1)	(2)	(3)
	vio	rob	mur
shall	-8.581	8.761	-0.277
	(12.19)	(5.809)	(0.378)
year	-1.822	-1.077	-0.339***
	(1.929)	(0.904)	(0.0583)
density	52.12**	94.43***	0.965**
	(17.20)	(6.450)	(0.370)
avginc	6.975	-4.268*	1.060***
	(3.869)	(1.803)	(0.116)
pop	12.61***	7.936***	-0.161
	(3.714)	(1.484)	(0.0871)
pm1029	-27.39***	-4.711	-0.454*
	(6.726)	(3.180)	(0.206)
_cons	931.1***	317.4**	30.77***
	(255.5)	(120.5)	(7.799)
N	1173	1173	1173
adj. R^2			

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Fixed-effects Regression

	(1)	(2)	(3)
	vio	rob	mur
shall	-4.349	10.25	-0.00670
	(20.60)	(7.508)	(0.322)
year	1.425	0.681	-0.218*
	(2.739)	(1.175)	(0.106)
density	-220.3***	51.58***	-14.97***
	(16.55)	(11.75)	(0.933)
avginc	-2.686	-8.311*	0.765
	(6.902)	(3.494)	(0.452)
pop	10.59	-0.349	-0.446
	(6.990)	(3.179)	(0.300)
pm1029	-20.65*	-1.956	-0.156
	(10.02)	(4.324)	(0.194)
_cons	774.3*	228.4	26.31***
	(351.1)	(149.4)	(7.252)
N	1173	1173	1173
adj. R^2	0.189	0.035	0.292

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(c)

The regression equation is

$$\begin{aligned}
vio_{it} &= \beta_0 + \beta_1 shall_{it} + \beta_2 year_{it} + \beta_3 density_{it} + \beta_4 avginc_{it} + \beta_5 pop_{it} + \beta_6 pm1029_{it} + \alpha_i + \epsilon_{it} \\
rob_{it} &= \beta_0 + \beta_1 shall_{it} + \beta_2 year_{it} + \beta_3 density_{it} + \beta_4 avginc_{it} + \beta_5 pop_{it} + \beta_6 pm1029_{it} + \alpha_i + \epsilon_{it} \\
mur_{it} &= \beta_0 + \beta_1 shall_{it} + \beta_2 year_{it} + \beta_3 density_{it} + \beta_4 avginc_{it} + \beta_5 pop_{it} + \beta_6 pm1029_{it} + \alpha_i + \epsilon_{it}
\end{aligned}$$

, and the regression outcomes are shown in Table 3.

(d)

Pooled OLS regression gives us very significant coefficients. The outcomes seem exciting, but there are systematical differences we haven't take into account among cross-section units. Omitted variable problem could cause inconsistent coefficients.

Both Random-effects regression and fixed-effects regression shows no significant coefficients of *shall*. For control variables, there are also quite different outcomes from 3 models. Random-effects regression gives us fewer significant coefficients than pooled regression. Among them, for violent crime rates, coefficients of *density* and *pop* become smaller; for robbery rates, the coefficients of average income and becomes negative; for murder rates, the coefficients of time trend and *pm1029* become negative, that of *density* becomes much smaller, that of average income becomes positive. These big changes indicates that at least one of pooled OLS regression and random-effects regression gives us inconsistent estimators.

Fixed-effects regression relaxes the random effects assumption so that it could give us more robust estimators. Here, coefficients of *shall* don't change too much and keep insignificant. It also gives us even fewer significant coefficients in other control variables than random-effects regression. For violent crime rates and murder rates, coefficients of *density* become negative. There could be chances that random effects affection don't hold, and we need further test.

(e)

Figure 1 shows the Hausman test for Violent Crime Rate; figure 2 shows the Hausman test for robbery rate; figure 3 shows the Hausman test for murder rate. All of them are rejected.

I cannot use the regression I performed in (c), since Hausman test requires homoscedasticity but the command "vce(robust)" assumes heteroscedasticity. Hausman test is only valid under the homoskedasticity of α_i and u_{it} as well as the serial uncorrelatedness of u_{it}.

	Coefficients (b) bfe	(B) bre	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
shall	-4.349308	-8.581308	4.232	.
year	1.42478	-1.822223	3.247003	.
density	-220.3029	52.11817	-272.421	25.54238
avginc	-2.686343	6.97511	-9.661453	.5748024
pop	10.59238	12.61003	-2.01765	3.295978
pm1029	-20.65137	-27.38912	6.737753	.

```
                  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(6) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                      =      118.64
            Prob>chi2 =      0.0000
            (V_b-V_B is not positive definite)
```

Figure 1: Hausman Test for Violent Crime Rate

	Coefficients (b) bfe	(B) bre	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
shall	10.25076	8.761407	1.48935	.2403327
year	.6811256	-1.077104	1.758229	.3033258
density	51.58466	94.43171	-42.84705	13.82835
avginc	-8.310881	-4.26825	-4.042631	.7116059
pop	-.349054	7.93578	-8.284834	1.96251
pm1029	-1.955505	-4.711253	2.755747	.6995338

```
                  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(6) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                      =       40.01
            Prob>chi2 =      0.0000
            (V_b-V_B is not positive definite)
```

Figure 2: Hausman Test for Robbery Rate

```
                    ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |      bfe          bre         Difference          S.E.
-------------+----------------------------------------------------------------
       shall |   -.0066966    -.2766634        .2699668               .
        year |   -.2183879    -.3392672        .1208793               .
     density |   -14.97313     .9648073       -15.93793        .7723211
      avginc |    .7653751     1.060399       -.2950237               .
         pop |   -.4461263    -.1612998       -.2848265        .1071126
      pm1029 |   -.1559347    -.4539376         .298003               .
------------------------------------------------------------------------------
                          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(6) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =      402.11
                Prob>chi2 =      0.0000
                (V_b-V_B is not positive definite)
```

Figure 3: Hausman Test for Murder Rate

2. Seat Belt Usage: First-Difference or Fixed-Effects?

(a)

The regression outcomes are shown in Table 4. (1) (2) and (3) correspond to the following regression equations:

$$fatalityrate_{it} = \beta_0 + \beta_1 sb\_useage_{it} + \beta_2 drinkage21_{it} + \beta_3 drinkage21\_speed70_{it} + \alpha_i + \epsilon_{it}$$
$$fatalityrate_{it} = \beta_0 + \beta_1 sb\_useage_{it} + \beta_2 drinkage21_{it} + \beta_3 drinkage21\_speed70_{it} + \alpha_i + \lambda_t + \epsilon_{it}$$
$$fatalityrate_{it} = \beta_0 + \beta_1 sb\_useage_{it} + \beta_2 drinkage21_{it} + \beta_3 drinkage21\_speed70_{it} + \alpha_i + \lambda_t + \eta_i t + \epsilon_{it}$$

(b)

To interpret whether lower speed limits reduce fatality risk of young drivers, we can focus on β_3 from the above regression equations. It measures whether the subgroup of sample who are above 21 and live in states with low speed limits have different fatality rate from others ceteris paribus.

Since we find that β_3 is not significant in all of the three models and are very close to 0, we could say that lower speed limits don't reduce fatality risk of young drivers.

(c)

The regression outcomes are shown in Table 5. (1) (2) and (3) correspond to the following regression equations:

$$\Delta fatalityrate_{it} = \beta_1 \Delta sb\_useage_{it} + \beta_2 \Delta drinkage21_{it} + \beta_3 \Delta drinkage21\_speed70_{it} + \epsilon_{it}$$
$$\Delta fatalityrate_{it} = \beta_1 \Delta sb\_useage_{it} + \beta_2 \Delta drinkage21_{it} + \beta_3 \Delta drinkage21\_speed70_{it} + \Delta\lambda_t + \epsilon_{it}$$
$$\Delta fatalityrate_{it} = \beta_1 \Delta sb\_useage_{it} + \beta_2 \Delta drinkage21_{it} + \beta_3 \Delta drinkage21\_speed70_{it} + \Delta\lambda_t + \eta_i + \epsilon_{it}$$

Table 4: Fixed-effects Regression

	(1) fatalityrate	(2) fatalityrate	(3) fatalityrate
sb_useage	-0.0173***	-0.00358*	-0.00281*
	(0.000831)	(0.00146)	(0.00132)
drinkage21	0.000648	-0.000684	-0.000965
	(0.000817)	(0.000661)	(0.000584)
drinkage21_speed70	-0.000234	0.000476	0.000579
	(0.000475)	(0.000492)	(0.000435)
_cons	0.0283***	0.0255***	0.780***
	(0.000768)	(0.000938)	(0.143)
N	556	556	556
adj. R^2	0.573	0.728	0.818

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: First-difference Regression

	(1) D.fatalityrate	(2) D.fatalityrate	(3) D.fatalityrate
D.sb_useage	-0.00551***	-0.00244	-0.00234
	(0.00101)	(0.00135)	(0.00146)
D.drinkage21	-0.000352	-0.000881	-0.000955
	(0.000507)	(0.000640)	(0.000719)
D.drinkage21_speed70	0.000121	-0.000101	-0.000154
	(0.000296)	(0.000379)	(0.000492)
N	497	497	497
adj. R^2	0.053	0.196	0.136

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(d)

If the coefficients in FE and FD are significantly different, the strict exogeneity assumption is violated. Here, we find that in the FD model, the coefficient of *sb_usage* for model (1) keeps significant, but becomes half of that in FE model, so there could be a significant change, and strict exogeneity doesn't hold.

On contrast, that coefficients in model (2) and (3) become not significant in the FD model, but change by not so much. We may think that after controlling the time trend and fixed effects, that strong exogeneity could hold.

A formal test is still needed to verify the above hypothesises.