“Reassessment of the Economic Model of Crime”

Research Project

By

SHUKHRAT KHUSEYNOV

Department of Economics
Bilkent University

In Partial Fulfilment of the Requirements for Econ 426

January 2019

I. Introduction

The consequences of local criminal laws on the crime rate have always been of concern to the legal authorities of different countries. The question of whether the severe penalties, such as capital punishment, truly discourage people from committing crime was raised. In order to measure those effects empirically, an economic model of crime was introduced by Gary Becker in 1968. Later on, there was a huge wave of literature estimating the economics of crime with various interpretations, which had a widespread influence on the legislature of many states. In the general model of crime, it is expected that an increase in the number of arrests, convictions, or severity of punishment, which are considered as deterring factors, yields the decrease of crime in the population, yet it has to be clarified by the empirical studies. In their paper, Cornwell and Trumbull (1994) argue that the vast majority of previous studies on the same issue do not account for the individual heterogeneity and the conventional simultaneity causing the endogeneity problem and, therefore, have overstated results. Econometrically, both heterogeneity and endogeneity violate Gauss-Markov assumptions, resulting in the distortions to the original ordinary least squares model (OLS). Hence, in this replication of Cornwell and Trumbull's "Estimating the Economics Model of Crime with Panel Data" paper, many different models and tests should be applied to obtain both unbiased and consistent results. In the light of improved software of the econometric techniques within the past decades, it would interesting to compare the estimates and shed light on the new view and the contemporary interpretation of the economic crime model based on the data from North Carolina.

II. Literature Review

There was a great number of works published on this issue after the introduction of the economics crime model in 1968. Many of them studied the effect of the percentage changes in the severity of punishment indicator and the probabilities of arrest, conviction, and imprisonment, i. e. their elasticities, on the percentage change in the crime rate. For instance, among the most noticeable ones was Ehrlich's paper (1975) concluding that the severity of punishment has a substantial deterring effect, which influenced the Supreme Court's decision on the constitutionality of death penalty. Similarly, these papers formed a consensus in the empirical literature of those times that any legislative change that increases the severity or the probability of punishment by one percent leads to a diminution in the crime rate by an interval between 0.3 and 1.1 percent on the average, having a predisposition towards one percent (Hirsch, 1988). Around the half of the conducted studies use OLS method and almost the rest of them use two (2SLS) or three (3SLS) stages least squares as their primary technique. The results of these studies demonstrate the negative elasticities of deterring factors' probabilities, which are fully consistent with the theoretical idea of the economic model of crime, agreeing also with the common expectations at the time of publications. However almost none of these papers account for the unobserved individual differences, i. e. heterogeneity, which is an essential omission, greatly affecting the estimated coefficients. Additionally, many papers ignore the endogeneity issue and those, who do not, have problematic instrumental variables with a possibility of the conventional simultaneity. Cornwell and Trumbull (1994) criticized their usage of the econometric cross-section techniques and, hence, employed a panel data to estimate a model with the resolved heterogeneity. They also selected appropriate and robust instrumental variables to account for the endogeneity in their 2SLS specifications, which were later criticized for having several wrong assumptions and insignificant estimates by Baltagi (2006). To clarify the case, a thorough estimation with remedies for the heterogeneity and the endogeneity has to be made.

III. Data

The panel data used in the paper was formed from different sources, such as the Federal Bureau of Investigation (FBI), North Carolina Department of Correction, North Carolina Employment Security Commission, and Census data, on 90 counties of North Carolina within the time frame between 1981 and 1987, having 630 observations in total.

Table 1. Summary Statistics

Variables	Labels	Mean	Standard Dev.	Minimum	Maximum
CRMRTE	crimes committed per person	0.0316	0.0181	0.00181	0.164
PRBARR (P_A)	'probability' of arrest	0.307	0.171	0.0588	2.750
PRBCONV (P_C)	'probability' of conviction	0.689	1.690	0.0684	37
PRBPRIS (P_P)	'probability' of prison sentenc	0.426	0.0872	0.149	0.679
AVGSEN (S)	avg. sentence, days	8.955	2.658	4.220	25.83
POLPC	police per capita	0.00192	0.00273	0.000459	0.0356
DENSITY	people per sq. mile	1.386	1.440	0.198	8.828
PCTYMLE	percent young male	0.0890	0.0243	0.0622	0.274
WCON	weekly wage, construction	245.7	122.0	65.62	2,325
WTUC	wkly wge, trns, util, commun	406.1	266.5	28.86	3,042
WTRD	wkly wge, whlesle, retail trade	192.8	88.41	16.87	2,243
WFIR	wkly wge, fin, ins, real est	272.1	55.77	3.516	509.5
WSER	wkly wge, service industry	224.7	104.9	1.844	2,177
WMFG	wkly wge, manufacturing	285.2	82.37	101.8	646.8
WFED	wkly wge, fed employees	403.9	63.07	255.4	598.0
WSTA	wkly wge, state employees	296.9	53.43	173.0	548
WLOC	wkly wge, local gov emps	258.0	41.36	163.6	388.1
WEST	=1 if in western N.C.	0.233	0.423	0	1
CENTRAL	=1 if in central N.C.	0.378	0.485	0	1
URBAN	=1 if in SMSA	0.0889	0.285	0	1
PCTMIN80	perc. minority, 1980	25.71	16.90	1.284	64.35
COUNTY (i)	county identifier	100.6	58.04	1	197
YEAR (t)	81 to 87	84	2.002	81	87

In the data, COUNTY and YEAR represent the cross-section (i) and the time (t) variables, respectively, CRMRTE stands for the crime rate of a county at a particular time period, PRBARR, PRBCONV, and PRBPRIS are the probabilities of arrest (P_A), conviction (P_C), and imprisonment (P_P), accordingly, AVGSEN is the severity indicator (S) measured by the average number of sentence in days, POLPC stands for the number of police force per capita, DENSITY represents the amount of people per square mile, i. e. population density, and PCTYMLE is the proportion of the male between the ages of 15 and 24 in the population. Furthermore, weekly wages by various industries were included in the data, where WCON stands for the weekly wages in the construction sector, WTUC is for the transportation, utilities, and communications sector, WTRD represents the wholesale and retail trade industry, WFIR is for the finance, insurance, and real estate industry, WSER is for the services industry, WMFG represents the manufacturing sector, WFED, WSTA, and WLOC are the wages for the federal, state, and local governments, respectively. Lastly, several time invariant variables were also deployed, where WEST is a dummy variable for location with western counties of North Carolina being equal to 1, CENTRAL is another location binary variable with central counties of North Carolina being equal to 1, URBAN stands for the dummy variable being equal to 1 for the cities with population over 50000 people, and PCTMIN80 is a proportion of minorities or nonwhite in the population of counties taken from the Census data of 1980s. Since the percentage differences, the elasticities of most variables, are used in the estimation, their logarithms were taken and named as LCRMRTE, LPRBARR, LPOLPC, LWCON, LPCTMIN, etc.

As seen in the Table 1, the average crime rate in the counties of North Carolina of those times is 3.16 percent with a standard deviation of 1.81 percent, which is not a small number compared to the average. The expected probability of arrest (P_A) is 30.7 percent, while the probability of conviction (P_C) given P_A has an expected value of 68.9 percent but with a high variability, and the average probability of imprisonment (P_P) conditional on P_C is 42.6 percent with a low standard deviation. Moreover, the average measure of severity (S) is around 9 days of imprisonment. The highest expected wages are seen in the sectors of transportations, utilities, and communication with the federal government positions, while the lowest are in the wholesale and retail trade sector. In the population of North Carolina counties, the average percentage of young males is 8.9 percent and the minorities constitute 25.71 percent of the population. Lastly, 8.89 percent of these counties are considered as urban area of these counties on the average (see Table 1). Some of the variables in the panel data are more believed to be the reason for crime than the others, yet it has to be clarified in the estimation part.

IV. Empirical Results

The estimation part of the paper consists of numerous regression methods and hypotheses testing. In the general model, the logarithm of the crime rate (C) is taken as a dependent variable that is indicated as LCRMRTE in the data. The estimated equation is:

$$\log C_{it} = \log P'_{it} + X'_{it} + \alpha_i + \varepsilon_{it}\,, \quad \text{where} \quad i = 1, \ldots, N; \quad t = 1, \ldots, T.$$

In this equation, log P' is a set of the main independent variables' elasticities that represent that deterring effect, which are log P_A, log P_C, log P_P, and log S indicators, and X' is a set of the external factor variables, most of which are calculated as logarithms to reveal the elasticities, such as logarithms of weekly wages, and several dummy variables,

Table 2. Estimation Results (not accounting for the endogeneity)

Variables	(1) OLS	(2) Between	(3) Random Effects	(4) Fixed Effects
LPRBARR (log P_A)	-0.537***	-0.648***	-0.387***	-0.355***
	(0.0296)	(0.0878)	(0.0301)	(0.0322)
LPRBCONV (log P_C)	-0.431***	-0.528***	-0.306***	-0.282***
	(0.0215)	(0.0667)	(0.0200)	(0.0211)
LPRBPRIS (log P_P)	-0.116**	0.297	-0.179***	-0.173***
	(0.0485)	(0.231)	(0.0322)	(0.0323)
LAVGSEN (log S)	-0.0907**	-0.236	-0.0104	-0.00245
	(0.0401)	(0.174)	(0.0264)	(0.0261)
LPOLPC	0.360***	0.364***	0.410***	0.413***
	(0.0223)	(0.0601)	(0.0244)	(0.0266)
LDENSITY	0.299***	0.168**	0.437***	0.414
	(0.0284)	(0.0774)	(0.0512)	(0.283)
LPCTYMLE	-0.145**	-0.0950	-0.0792	0.627*
	(0.0632)	(0.158)	(0.125)	(0.364)
LWCON	0.0853	0.195	-0.0112	-0.0378
	(0.0545)	(0.210)	(0.0390)	(0.0391)
LWTUC	0.0216	-0.196	0.0463**	0.0455**
	(0.0312)	(0.170)	(0.0193)	(0.0190)
LWTRD	0.0431	0.129	-0.00939	-0.0205
	(0.0623)	(0.278)	(0.0409)	(0.0405)
LWFIR	-0.00188	0.113	-0.00331	-0.00390
	(0.0447)	(0.220)	(0.0286)	(0.0283)
LWSER	-0.0265	-0.106	0.00520	0.00888
	(0.0312)	(0.163)	(0.0194)	(0.0191)
LWMFG	-0.0791	-0.0249	-0.213***	-0.360***
	(0.0529)	(0.134)	(0.0814)	(0.112)
LWFED	0.0843	0.156	-0.157	-0.309*
	(0.114)	(0.287)	(0.149)	(0.176)
LWSTA	-0.202**	-0.284	-0.0406	0.0529
	(0.0930)	(0.256)	(0.103)	(0.114)
LWLOC	0.0554	0.0103	0.157	0.182
	(0.140)	(0.463)	(0.115)	(0.118)
WEST	-0.229***	-0.230**	-0.225**	-0.167***
	(0.0459)	(0.108)	(0.105)	(0.0553)
CENTRAL	-0.178***	-0.164**	-0.194***	-0.102***
	(0.0276)	(0.0645)	(0.0625)	(0.0317)
URBAN	-0.136**	-0.0346	-0.215*	-0.176***
	(0.0537)	(0.132)	(0.114)	(0.0454)
LPCTMIN	0.179***	0.148***	0.187***	0.193***
	(0.0195)	(0.0485)	(0.0428)	(0.0223)
Constant	-2.721***	-2.097	-1.002	2.393
	(0.966)	(2.822)	(1.194)	(1.678)

Observations	630	630	630	630
Number of years (T)	7	7	7	7
Number of counties (N)	90	90	90	90
Heteroscedasticity Correction	No	No	No	No
Cross Section Fixed Effects	No	No	No	Yes
Time Series Fixed Effects	Yes	Yes	Yes	Yes
R-square [within]	-	0.0932	0.453	0.463
R-square [between]	-	0.880	0.814	0.554
R-square [overall]	0.814	0.757	0.783	0.545

Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1)

which are all time invariant. Additionally, parameter alpha that represents the effect of unobserved individual characteristics being constant over time and the error term are on the right-hand side of the equation. Both N and T stand for the number of counties and the number of years, respectively. In general, among the main econometric problems of panel models are heterogeneity and endogeneity. To remedy the former, different specifications of panel data techniques were applied. The latter, endogeneity, on the other hand, was cured by 2SLS models using instrumental variables method. In order to choose the best specification with the most unbiased and consistent results, various models were tested and compared.

First of all, Pooled OLS method was implemented to compare other models with its output (see Table 2). All of the observed coefficients, except for the most wage elasticities, are significant, showing great negative impact of the deterring variables on the crime rate, as it was expected by the theory and found in the previous studies. The proportion of the police forces, the population density, and the minorities seem to have an increasing effect on the crime rate. Apparently, there is a backward causation regarding the police forces' elasticity. In other words, for the counties where the crime rate is expected to be higher, more police officers are engaged. Although LPCTYMLE was expected to affect positively on the crime rate, in this specification percentage of the young male are observed to have a decreasing effect. Among the wage elasticities, the only significant coefficient is LWSTA, demonstrating that the higher wage of state government workers decreases the crime rate. After that, to adjust for the heterogeneity issue, different models, such as the Between-Group Estimator model, the Random Effects model (GLS), and also the Fixed Effects (Mean deviated model) were analyzed. In the Between-Group Estimator model, as depicted in Table 2, fewer coefficients are significant. Among the significant deterring variables, the coefficients of log P_A and log P_C are much higher in absolute values than their OLS counterparts. It implies that an increase in elasticities of P_A and P_C by 1 percent would decrease the log C by 0.648 percent and 0.528 percent on average, respectively, according to this specification. The external variables have more or less similar coefficients, except for the fact that there are more insignificant ones. The Random Effects model has both many significant and insignificant coefficients in the estimated regression (see Table 2). The estimated deterring variables are considerably lower in absolute values than their equivalents in previous specifications. Interestingly, the Random Effects model suggests that LPOLPC and LDENSITY increase the crime rate's elasticity with large significant coefficients. Among the wage elasticities, LWTUC is expected to influence positively and LWMFG negatively on the crime rate's elasticity, according to this model. Lastly, in the Mean Deviated Fixed Effects model, although four time invariant variables were dropped to avoid the multicollinearity, they were estimated as a second stage by regressing the residuals on the dropped variables (see Table 2). There are both many significant and insignificant coefficients here. The significant deterring variables are even less in absolute

Table 3. Estimation Results (with instrumental variables)

Variables	(5) 2SLS	(6) BE2SLS	(7) RE2SLS	(8) FE2SLS
<u>*LPRBARR*</u> (log P_A)	-0.379***	-0.503**	-0.414*	-0.576
	(0.0827)	(0.241)	(0.221)	(0.802)
<u>*LPOLPC*</u>	0.372***	0.408**	0.505**	0.658
	(0.0817)	(0.193)	(0.228)	(0.847)
LPRBCONV (log P_C)	-0.398***	-0.525***	-0.343***	-0.423
	(0.0431)	(0.0999)	(0.132)	(0.502)
LPRBPRIS (log P_P)	-0.116**	0.187	-0.190***	-0.250
	(0.0498)	(0.318)	(0.0733)	(0.279)
LAVGSEN (log S)	-0.0881**	-0.227	-0.00644	0.00910
	(0.0413)	(0.179)	(0.0289)	(0.0490)
LDENSITY	0.346***	0.226**	0.434***	0.139
	(0.0337)	(0.102)	(0.0711)	(1.021)
LPCTYMLE	-0.0878	-0.0947	-0.146	0.351
	(0.0859)	(0.192)	(0.227)	(1.011)
LWCON	0.124**	0.314	-0.00430	-0.0287
	(0.0608)	(0.259)	(0.0414)	(0.0535)
LWTUC	0.0295	-0.199	0.0445**	0.0391
	(0.0328)	(0.197)	(0.0215)	(0.0309)
LWTRD	0.0216	0.0536	-0.00856	-0.0178
	(0.0643)	(0.296)	(0.0420)	(0.0453)
LWFIR	-0.0210	0.0417	-0.00403	-0.00934
	(0.0475)	(0.306)	(0.0295)	(0.0366)
LWSER	-0.0241	-0.135	0.0106	0.0186
	(0.0321)	(0.174)	(0.0216)	(0.0388)
LWMFG	-0.103*	-0.0420	-0.202**	-0.243
	(0.0581)	(0.156)	(0.0839)	(0.420)
LWFED	0.0876	0.148	-0.213	-0.451
	(0.132)	(0.326)	(0.215)	(0.527)
LWSTA	-0.146	-0.203	-0.0601	-0.0187
	(0.0999)	(0.298)	(0.120)	(0.281)
LWLOC	0.0667	0.0444	0.184	0.263
	(0.144)	(0.494)	(0.140)	(0.312)
WEST	-0.218***	-0.205*	-0.228**	-0.282***
	(0.0473)	(0.114)	(0.101)	(0.0611)
CENTRAL	-0.189***	-0.173***	-0.199***	-0.0327
	(0.0286)	(0.0667)	(0.0607)	(0.0351)
URBAN	-0.153**	-0.0805	-0.260*	0.0713
	(0.0646)	(0.144)	(0.150)	(0.0503)
LPCTMIN	0.184***	0.169***	0.195***	0.168***
	(0.0211)	(0.0527)	(0.0459)	(0.0247)
Constant	-2.597**	-1.977	-0.454	2.943
	(1.204)	(4.001)	(1.703)	(2.694)

Observations	630	630	630	630
Number of years (T)	7	7	7	7
Number of counties (N)	90	90	90	90
Heteroscedasticity Correction	No	No	No	No
Cross Section Fixed Effects	No	No	No	Yes
Time Series Fixed Effects	Yes	Yes	Yes	Yes
R-square [within]	-	0.106	0.452	0.359
R-square [between]	-	0.874	0.804	0.444
R-square [overall]	0.804	0.759	0.772	0.443

Standard errors in parentheses (*** $p<0.01$, ** $p<0.05$, * $p<0.1$)

values than in the Random Effects model. The external factor variables are very similar to those of the previous specification with addition of LPCTYMLE having great positive coefficient, which was initially expected, however it is only significant with the acceptable error amount of 10 percent. To choose the best model, several hypotheses tests have to be conducted.

Since there is a potential endogeneity issue in the model, 2SLS specifications, using the instrumental variables method, were also utilized. As it was also stated by Cornwell and Trumbull (1994), the conventional simultaneity, which is a source of endogeneity, exists between the independent variable log C and dependent log P_A with LPOLPC, which are both assumed as endogenous variables in the 2SLS model. This replication uses the same instrumental variables, the exogenous variables for log P_A and LPOLPC, which are the elasticities of the mix of different offense types (LMIX) and the tax revenue per capita (LTAXPC). All of the procedures used in the first part of the estimation were repeated with the instrumental variables: 2SLS model, the model of 2SLS with the Between-Group Estimator (BE2SLS), the model of 2SLS with the Random Effects (RE2SLS), and 2SLS with the Fixed Effects model (FE2SLS), which can observed in Table 3. Newly estimated specifications seem to be less significant in general. The FE2SLS is mostly insignificant, except for two omitted coefficients estimated as a second stage. Other specifications are less daunting on this issue. Among significant deterring variables, in absolute values the greatest coefficients belong to BE2SLS and RE2SLS specifications. Those of 2SLS are closer by their magnitude to Random and Fixed Effects models. Apparently, all LPOLPC coefficients are very large in numbers compared to the models not accounting for the endogeneity. Other external factor variables are not very different from initial estimations (see Table 3). In order to compare the specifications, the hypotheses testing results should be analyzed.

Table 4. Model Comparisons

	(1) OLS	(2) BE	(3) RE	(4) FE	(5) 2SLS	(6) BE2SLS	(7) RE2SLS	(8) FE2SLS
Wald test (F stat) or Breusch Pagan LM			vs. OLS (LM test)	vs. OLS (F test)				vs. 2SLS (F test)
Comparing models with least squares			712.29 Reject H_0	36.38 Reject H_0				13.93 Reject H_0
Hausman test - χ^2		vs. FE	vs. FE			vs. FE2SLS	vs. FE2SLS	vs. FE
Comparing models with fixed effects		21.18 Fail to reject H_0	49.39 Reject H_0			54.30 Reject H_0	16.45 Fail to reject H_0	0.08 Fail to reject H_0

Finally, the specification models were compared by using several testing methodologies. As reported in Table 4, when Pooled OLS and 2SLS are compared with the Fixed Effects model and FE2SLS, respectively, by using F test, the null hypothesis of having no observable difference between the two in both hypotheses tests is rejected with 5 percent acceptable error amount. In other words, there is some unobserved heterogeneity in this framework and the models with no adjustment for the individual characteristics are inferior to those with fixed effects. Moreover, the Random Effects model was also tested against OLS by using Breusch-Pagan Lagrangian Multiplier test (see Table 4). The test concluded that the Random Effects model is far better than the Pooled OLS. To compare the rest models, Hausman specification tests were conducted.

While comparing the Fixed Effects model with Between-Group and Random Effects model in separate tests, the coefficients of Between-Group are assumed to be same with the Fixed Effects model, but the null of Random Effects model being equivalent was rejected. Therefore, Random Effects model is not consistent. Even though the Between-Group Estimator model is assumed to be same with the Fixed Effects, the latter is preferred due to higher efficiency. Similarly, when the equivalent 2SLS models are compared, the best among them seems to be RE2SLS, which has similar coefficients with FE2SLS but is more efficient. Lastly, when the Fixed Effects model is compared with FE2SLS, the test fails to reject the null hypothesis and, therefore, both of the specifications are assumed to be similar (see Table 4). Therefore, both Fixed Effects model and RE2SLS are the best specifications from the first and the second parts of estimation, respectively, but the former is preferred on efficiency grounds.

V. Conclusion

In a nutshell, the data and tests conclude that the panel data estimations with accounting for the heterogeneity and endogeneity may lead to different results. Although there are some differences in the results of specifications and hypotheses testing, the paper agrees with Cornwell and Trumbull (1994) that omission of these essential techniques was a flaw in previous studies. After resolving the econometric problems and comparing different specifications, the Fixed Effects model was chosen as the most consistent and efficient. It estimates that one percent increase in the elasticities of either P_A, P_C, or P_P leads to 35.5 percent, 28.2 percent, or 17.3 percent decrease in the elasticity of the crime rate. These estimates agree with the general idea of the crime economics but they are far less than the predictions of previous empirical studies, which were overstated.

VI. References

Baltagi, Badi H. "Estimating an Economic Model of Crime Using Panel Data from North Carolina." *Journal of Applied Econometrics*, vol. 21, no. 4, 2006, pp. 543–547. *JSTOR*, JSTOR, www.jstor.org/stable/25146443.

Becker, Gary S. "Crime and Punishment: An Economic Approach." *Journal of Political Economy*, vol. 76, no. 2, 1968, pp. 169–217., doi:10.1086/259394.

Cornwell, Christopher, and William N. Trumbull. "Estimating the Economic Model of Crime with Panel Data." *The Review of Economics and Statistics*, vol. 76, no. 2, 1994, pp. 360–366. *JSTOR*, JSTOR, www.jstor.org/stable/2109893.

Ehrlich, Isaac. "Participation in Illegitimate Activities: A Theoretical and Empirical Investigation." *Journal of Political Economy*, vol. 81, no. 3, 1973, pp. 521–567., doi:10.1086/260058.

Ehrlich, Isaac. "The Deterrent Effect of Capital Punishment: A Question of Life and Death." *American Economic Review*, vol. 65, June 1975, pp. 397–417.

Hirsch, Werner Z. *Law and Economics*. 2nd ed., Academic Press, 1988.