JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **21**: 543–547 (2006)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jae.861

REPLICATION SECTION

ESTIMATING AN ECONOMIC MODEL OF CRIME USING PANEL DATA FROM NORTH CAROLINA

BADI H. BALTAGI*

Department of Economics and Center for Policy Research, 426 Eggers Hall, Syracuse University, Syracuse, New York, 13244-1020

SUMMARY

This paper replicates the Cornwell and Trumbull (1994) estimation of a crime model using panel data on 90 counties in North Carolina over the period 1981–1987. While the Between and Within estimates are replicated, the fixed effects 2SLS as well as the 2SLS estimates are not. In fact, the fixed effects 2SLS estimates turn out to be insignificant for all important deterrent variables as well as legal opportunity variables. We argue that the usual Hausman test, based on the difference between fixed effects and random effects, may lead to misleading inference when endogenous variables of the conventional simultaneous equation type are among the regressors. We estimate the model using random effects 2SLS and perform a Hausman test based on the difference between fixed effects 2SLS and random effects 2SLS. We cannot reject the consistency of the random effects 2SLS estimator and this estimator yields plausible and significant estimates of the crime model. This result should be tempered by the legitimacy of the chosen instruments.

1. REPLICATION RESULTS

Cornwell and Trumbull (1994), hereafter (CT), estimated an economic model of crime using panel data on 90 counties in North Carolina over the period 1981–1987. Table I replicates the Between and fixed effects estimates of CT using Stata. The empirical model follows Becker (1968) and Ehrlich (1973) among others, and relates the crime rate (which is an FBI index measuring the number of crimes divided by the county population) to a set of explanatory variables which include deterrent variables as well as variables measuring returns to legal opportunities. All variables are in logs except for the regional and time dummies. The explanatory variables consist of the probability of arrest P_A (which is measured by the ratio of arrests to offences), probability of conviction given arrest P_C (which is measured by the ratio of convictions to arrests), probability of a prison sentence given a conviction P_P (measured by the proportion of total convictions resulting in prison sentences); average prison sentence in days (S) as a proxy for sanction severity; the number of

* Correspondence to: Professor Badi H. Baltagi, Department of Economics and Center for Policy Research, 426 Eggers Hall, Syracuse University, Syracuse, New York, 13244-1020. E-mail: bbaltagi@maxwell.syr.edu

Received 26 May 2004
Revised 9 November 2004

police per capita as a measure of the county's ability to detect crime (Police); the population density, which is the county population divided by county land area (Density); a dummy variable (Urban) indicating whether the county is in the SMSA with population larger than 50 000; percent minority, which is the proportion of the county's population that is minority or non-white; percent young male, which is the proportion of the county's population that is male and between the ages of 15 and 24; regional dummies for western and central counties. Opportunities in the legal sector are captured by the average weekly wage in the county by industry. These industries are: construction; transportation, utilities and communication; wholesale and retail trade; finance, insurance and real estate; services; manufacturing; and federal, state and local government.

Fixed effects results show that the probability of arrest, the probability of conviction given arrest and the probability of imprisonment given conviction all have a negative and significant effect on the crime rate with estimated elasticities of -0.355, -0.282 and -0.173, respectively. The sentence severity has a negative but insignificant effect on the crime rate. The greater the number of police per capita, the greater the number of reported crimes per capita. The estimated elasticity is 0.413 and it is significant. This could be explained by the fact that the larger the police force, the larger the reported crime. Alternatively, this could be an endogeneity problem with more crime resulting in the hiring of more police. The higher the density of the population the higher the crime rate, but this is insignificant. Returns to legal activity are insignificant except for wages in the manufacturing sector and wages in the transportation, utilities and communication sector. The first wage has a negative and significant effect on crime with an estimated elasticity of -0.36, while the second wage has a positive and significant effect on crime with an estimated elasticity of 0.046. Percent young male is insignificant at the 5% level.

Cornwell and Trumbull (1994) argue that the Between estimates do not control for county effects and yield much higher deterrent effects than the fixed effects estimates. These Between estimates as well as the random effects estimates are rejected as inconsistent by a Hausman (1978) test. CT worried about the endogeneity of police per capita and the probability of arrest. They used as instruments two additional variables. Offence mix, which is the ratio of crimes involving face-to-face contact (such as robbery, assault and rape) to those that do not. The rationale for using this variable is that arrest is facilitated by positive identification of the offender. The second instrument is per capita tax revenue. This is justified on the basis that counties with preferences for law enforcement will vote for higher taxes to fund a larger police force. The fixed effects 2SLS estimates are reported in Table I. These results do not replicate those in table 3 of Cornwell and Trumbull (1994).[1] All deterrent variables had insignificant t-statistics. These include the probability of arrest, the probability of conviction given arrest as well as the probability of imprisonment given conviction. Also insignificant were sentence severity and police per capita. Manufacturing wage, which was significant using the fixed effects 2SLS estimates of Cornwell and Trumbull (1994), turn out to be insignificant in our replication. In fact, CT find that all variables were insignificant using fixed effects 2SLS, except for the percent young male and the manufacturing wage. CT also performed a Hausman test based on the difference between FE and FE-2SLS. Because the test did not reject the null, CT emphasize their FE estimates. Note that the FE estimator is not consistent if the endogeneity of police per capita and the probability of arrest is of the conventional type. In

[1] The Cornwell and Trumbull (1994) fixed effects 2SLS was replicated only after replacing the right-hand side endogenous variables, i.e., the probability of arrest and police per capita by their predicted values from a first stage fixed effects regression which *ignores* the time dummies. The second stage regression runs a fixed effects regression *including* the time dummies.

Table I. Economics of Crime Estimates for North Carolina, 1981–1987 (standard errors in parentheses)

	Between	Fixed effects	FE2SLS	BE2SLS	EC2SLS
P_A	−0.648	−0.355	−0.576	−0.503	−0.413
	(0.088)	(0.032)	(0.802)	(0.241)	(0.097)
P_C	−0.528	−0.282	−0.423	−0.525	−0.323
	(0.067)	(0.021)	(0.502)	(0.100)	(0.054)
P_P	0.297	−0.173	−0.250	0.187	−0.186
	(0.231)	(0.032)	(0.279)	(0.318)	(0.042)
S	−0.236	−0.002	0.009	−0.227	−0.010
	(0.174)	(0.026)	(0.049)	(0.179)	(0.027)
Police	0.364	0.413	0.658	0.408	0.435
	(0.060)	(0.027)	(0.847)	(0.193)	(0.090)
Density	0.168	0.414	0.139	0.226	0.429
	(0.077)	(0.283)	(1.021)	(0.102)	(0.055)
wcon	0.195	−0.038	−0.029	0.314	−0.007
	(0.210)	(0.039)	(0.054)	(0.259)	(0.040)
wtuc	−0.196	0.046	0.039	−0.199	0.045
	(0.170)	(0.019)	(0.031)	(0.197)	(0.020)
wtrd	0.129	−0.021	−0.018	0.054	−0.008
	(0.278)	(0.040)	(0.045)	(0.296)	(0.041)
wfir	0.113	−0.004	−0.009	0.042	−0.004
	(0.220)	(0.028)	(0.037)	(0.306)	(0.029)
wser	−0.106	0.009	0.019	−0.135	0.006
	(0.163)	(0.019)	(0.039)	(0.174)	(0.020)
wmfg	−0.025	−0.360	−0.243	−0.042	−0.204
	(0.134)	(0.112)	(0.420)	(0.156)	(0.080)
wfed	0.156	−0.309	−0.451	0.148	−0.164
	(0.287)	(0.176)	(0.527)	(0.326)	(0.159)
wsta	−0.284	0.053	−0.019	−0.203	−0.054
	(0.256)	(0.114)	(0.281)	(0.298)	(0.106)
wloc	0.010	0.182	0.263	0.044	0.163
	(0.463)	(0.118)	(0.312)	(0.494)	(0.120)
Percent young male	−0.095	0.627	0.351	−0.095	−0.108
	(0.158)	(0.364)	(1.011)	(0.192)	(0.140)
Percent minority	0.148	—	—	0.169	0.189
	(0.049)			(0.053)	(0.041)
west	−0.230	—	—	−0.205	−0.227
	(0.108)			(0.114)	(0.100)
central	−0.164	—	—	−0.173	−0.194
	(0.064)			(0.067)	(0.060)
urban	−0.035	—	—	−0.080	−0.225
	(0.132)			(0.144)	(0.116)
_cons	−2.097	—	—	−1.977	−0.954
	(2.822)			(4.001)	(1.284)

Time dummies were included except for Between and BE2SLS. The number of observations is 630. The F-statistics for significance of county dummies in fixed effects is F(89,518) = 36.38. The corresponding F-statistic using FE2SLS is 29.66. Both are significant. Hausman's test for (fixed effects—random effects) is $\chi^2(22) = 49.4$ with p-value of 0.0007. The corresponding Hausman test for (FE2SLS—EC2SLS) is $\chi^2(22) = 19.5$ with a p-value of 0.614.

fact, the legitimacy of these two instruments cannot be tested since the model is just-identified. This fact was recognized by Cornwell and Trumbull (1994, p. 364).[2]

[2] CT also report 2SLS estimates ignoring the heterogeneity in the county effects for comparison. However, they warn against using biased and inconsistent estimates that ignore county effects. In fact, county effects were always significant, see the F-statistics reported in Table I.

An alternative to dealing with the endogeneity problem is to run a random effects 2SLS estimator that allows for the endogeneity of police per capita and the probability of arrest. This estimator is a matrix weighted average of Between 2SLS and fixed effects 2SLS and was derived by Baltagi (1981). This was denoted by error components 2SLS or EC2SLS. A Hausman test based on the difference between fixed effects 2SLS and random effects 2SLS, suggested by Baltagi (2004), yields a $\chi^2(22) = 19.50$, which is insignificant with a p-value of 0.614. This does not reject the null hypothesis that EC2SLS yields a consistent estimator. Recall that the random effects estimator was rejected by Cornwell and Trumbull (1994) based on the standard Hausman (1978) test. The latter was based on the contrast between fixed effects and random effects assuming that the endogeneity comes entirely from the correlation between the county effects and the explanatory variables. This does not account for the endogeneity of the conventional simultaneous equation type between police per capita and the probability of arrest and the crime rate. An alternative Hausman test based on the contrast between fixed effects 2SLS and EC2SLS failed to reject the null hypothesis. This means that we cannot reject the consistency of EC2SLS. Again, this result depends upon the legitimacy of the instruments chosen by CT. For the EC2SLS results in Table I, all the deterrent variables are significant with slightly higher elasticities than fixed effects. The sentence severity variable is still insignificant and police per capita is still positive and significant. Manufacturing wages are negative and significant and percent minority is positive and significant. Obtaining an estimate of the last coefficient is an advantage of EC2SLS over the fixed effects estimators, because it allows us to recapture estimates of variables that were invariant across time and wiped out by the fixed effects transformation, see also Hausman and Taylor (1981). We also ran the first stage regressions to check for weak instruments. For the probability of arrest, the F-statistic of the fixed effects first stage regression was 15.6 as compared to 4.62 for the Between first stage regression. Similarly, for the police per capita, the F-statistic of the fixed effects first stage regression was 9.27 as compared to 2.60 for the Between first stage regression. This indicates that these instruments may be weaker in the Between first stage regressions (for Between 2SLS) than in the fixed effects first stage regressions (for fixed effects 2SLS).

2. CONCLUSION

This paper confirms the Cornwell and Trumbull (1994) conclusion that county effects cannot be ignored in estimating an economic model of crime using panel data from North Carolina. This paper also shows that the usual Hausman test based on the difference between fixed effects and random effects may lead to misleading inference if there are endogenous regressors of the conventional simultaneous equation type. We suggested an alternative Hausman test based on the difference between fixed effects 2SLS and random effects 2SLS. This test did not reject the consistency of random effects 2SLS, an estimator which yields plausible estimates of the crime equation. This result should be tempered by the fact that FE2SLS estimates for this crime equation are imprecise and the fact that they are consistent only if the instruments are legitimate.

ACKNOWLEDGEMENTS

I would like to thank In-Sik Min for his helpful research assistance, Chris Cornwell, two referees and the editor Hashem Pesaran for their comments and suggestions.

REFERENCES

Baltagi BH. 1981. Simultaneous equations with error components. *Journal of Econometrics* **17**: 189–200.

Baltagi BH. 2004. A Hausman test based on the difference between fixed effects two-stage least squares and error components two-stage least squares, problem 04.1.1. *Econometric Theory* **20**: 223–224.

Becker GS. 1968. Crime and punishment: an economic approach. *Journal of Political Economy* **76**: 169–217.

Cornwell C, Trumbull WN. 1994. Estimating the economic model of crime with panel data. *Review of Economics and Statistics* **76**: 360–366.

Ehrlich I. 1973. Participation in illegitimate activities: a theoretical and empirical investigation. *Journal of Political Economy* **81**: 521–567.

Hausman JA. 1978. Specification tests in econometrics. *Econometrica* **46**: 1251–1271.

Hausman JA, Taylor WE. 1981. Panel data and unobservable individual effects. *Econometrica* **49**: 1377–1398.

