

The dataset CORNWELL.dta (also available in .xls) was used by C. Cornwell and W. Trumbull (1994) in their paper “Estimating the Economic Model of Crime with Panel Data” (Review of Economics and Statistics, 76, pp. 360-366).

The paper aimed at providing further empirical evidence to the economic model of crime already present in the literature. The dataset collects information on a set of variables relating to a sample of 90 counties of North Carolina for the period 1981-1987.

Using panel data techniques, estimate the following equation model:

$$lcrmrte_{it} = \beta_0 + \beta_1 lprbarr_{it} + \beta_2 lprbconv_{it} + \beta_3 lprbpris_{it} + \beta_4 lavgsen_{it} + \beta_5 lpolpc_{it} + \beta_6 ldensity_{it} + \beta_7 lpctymle_{it} + \beta_8 lpctm_{it}$$

$$i=1, \dots, N; t=1, \dots, T \quad (1)$$

where the variables are described in the following table:

county	county identifier
year	81 to 87
crmrte	crimes committed per person
prbarr	'probability' of arrest
prbconv	'probability' of conviction
prbpris	'probability' of prison sentence
avgsen	avg. sentence, days
polpc	police per capita
density	people per sq. mile
west	=1 if in western N.C.
central	=1 if in central N.C.
urban	=1 if in SMSA
pctmin80	perc. minority, 1980
wcon	weekly wage, construction
wtuc	wkly wge, trns, util, commun
wtrd	wkly wge, whlesle, retail trade
wfir	wkly wge, fin, ins, real est
wser	wkly wge, service industry
wmfg	wkly wge, manufacturing
wfed	wkly wge, fed employees
wsta	wkly wge, state employees
wloc	wkly wge, local gov emps
pctymle	percent young male

(Notice that you need to transform the variables in logarithm).

(1) Describe the dataset, providing the main descriptive statistics for panel data and also replying the following questions:

**1.1. How many observations are there in the dataset?**

**Answer:**

**1.2. Determine how many different counties there are in the dataset.**

**Answer:**

**1.3. Determine how many years there are in the dataset.**

**Answer:**

**1.4 Is the panel balanced or unbalanced? Why?**

**Answer:**

**1.5. What are time invariant variables? What are time-variant variables?**

**Answer:**

**1.6. Describes the variables according to their mean and their variability (compare the within, between, and overall variability)**

**Answer:**

**(2) Estimate the equation by pooled OLS and report the results.**

(2.1) What are the statistical significant variables? (at the 10% significant level)

**Answer:**

(2.2) What are the estimated coefficient of  $lprbarr$  ( $\beta_1$ ) and  $lpolpc$  ( $\beta_5$ ). If they are statistically significant, interpret the results.

**Answer:**

**(3) Now estimate the equation using fixed effects.**

(3.1) What are the statistical significant variables? (at the 10% significant level)

**Answer:**

(3.2) What are the estimated coefficient of  $lprbarr$  ( $\beta_1$ ) and  $lpolpc$  ( $\beta_5$ ). If they are statistically significant, interpret the results.

**Answer:**

**(4) Now estimate the equation using random effects. What is the RE estimate of  $\beta_5$  and of  $\beta_1$ .**

(4.1) What are the statistical significant variables? (at the 10% significant level)

**Answer:**

(4.2) What are the estimated coefficient of  $lprbarr$  ( $\beta_1$ ) and  $lpolpc$  ( $\beta_5$ ). If they are statistically significant, interpret the results.

**Answer:**

**(5) After providing the results of the Breush-Pagan Test and the Hausman test, identify the best model to describe our data. Explain.**

**Answer:**