Environment Department

Advanced Environmental Economics

Spring Term 2002

Preliminary Notes on Estimating an EKC Using Panel Data in LimDep

Lecturer: Giovanni Baiocchi

Course Web Page:
http://www-users.york.ac.uk/~jb35/aee/aee.htm

- Aims
  - The aim of these lectures to enable students to reproduce the results of a paper published on a top environmental economics journal using LimDep.
- Learning Outcomes
  - Understand how to read and transform data in LimDep .
  - Use panel data regression techniques in LimDep.
  - Choose the appropriate method.
  - Know how quantities needed for policy analysis can be computed from regression results in LimDep.

Contents

1 Panel Data File Structure

We will reproduce the results from the following paper:

- Stern D. I. and Mick S. Common (2001), Is there an environmental Kuznets curve for sulfur?, *Journal of Environmental Economics and Management*, 40(2).

For the dataset used by Common and Stern the first column is the time index (years from 1960 to 1990) whereas the second column contains the individual country index (there are 74 countries in the sample). The correspondence between codes and countries is provided in Table 1. The third column contains the populations, the fourth SO_2 emissions. The fifth is the GDP in real 1990 international dollars. The sixth column contains the SO_2 concentration per capita, and the last column a OECD/non-OECD dummy. The emission data comes from ASL and Associates; GDP and population is taken from the Penn World Table.

1960	54	17910	1099.72	7258	0.0614	1
1961	54	18270	1076.06	7261	0.0589	1
1962	54	18614	1073.68	7605	0.05768	1
1963	54	18963	1087.53	7876	0.05735	1
1964	54	19326	1142.22	8244	0.0591	1
1965	54	19678	1206.56	8664	0.06132	1
1966	54	20049	1174.17	9093	0.05857	1
1967	54	20411	1304.04	9231	0.06389	1
1968	54	20744	1328.19	9582	0.06403	1
1969	54	21028	1239.66	9975	0.05895	1
1970	54	21324	1356.98	10124	0.06364	1
1971	54	21592	1416.06	10599	0.06558	1
1972	54	21822	1478.4	11125	0.06775	1
1973	54	22072	1540.07	11854	0.06977	1
1974	54	22364	1553.33	12225	0.06946	1
1975	54	22697	1505.36	12287	0.06632	1
1976	54	22993	1433.46	12929	0.06234	1
1977	54	23273	1531.62	13184	0.06581	1
1978	54	23517	1340.82	13631	0.05701	1
1979	54	23747	1424.14	14114	0.05997	1
1980	54	24043	1512.53	14133	0.06291	1
1981	54	24342	1439.02	14555	0.05912	1
1982	54	24583	1267.14	13740	0.05155	1
1983	54	24787	1213.11	14105	0.04894	1
1984	54	24978	1363.25	14954	0.05458	1
1985	54	25165	1304.08	15589	0.05182	1
1986	54	25353	1238.74	16029	0.04886	1
1987	54	25625	1388.12	16602	0.05417	1

```
1988    54  25950   1417.18 17258   0.05461 1
1989    54  26219   1334.33 17524   0.05089 1
1990    54  26522   1366.41 17173   0.05152 1
1960    72  180673  9479.18 9895    0.05247 1
1961    72  183687  9246.24 9946    0.05034 1
1962    72  186537  9492.94 10358   0.05089 1
...
```

2 Reading Data

```
READ; file=M:\stern.dat; nvar=7; nobs=2294
   ; names=year, country, pop, so, gdpc, sopc, oe$
```

3 Descriptive Statistics

We can now summarise the data by obtaining some descriptive statistics (dstat).

```
DSTAT; Rhs=pop, so, gdpc, sopc, oe$
```

```
                             Descriptive Statistics
               All results based on nonmissing observations.
Variable        Mean          Std.Dev.        Minimum         Maximum      Cases
--------------------------------------------------------------------------------
POP       47466.1081      130444.260      231.000000      1133683.00        2294
SO        703.026687      1991.16822      .100000000E-01  14213.8900        2294
GDPC      5359.90817      6244.16765      303.000000      80830.7551        2294
SOPC      .215023403E-01  .366821025E-01  .890000000E-06  .465551720        2294
OE        .310810811      .462926418      .000000000      1.00000000        2294
```

4 Creating and Transforming Variables

For the whole sample create squared and log terms.

```
CREATE ; lgdp = log(gdpc)
       ; lgdpsq=lgdp*lgdp
       ; lsopc=log(sopc)$
```

Table 1: Country Codes

1	ALGERIA	95	JAPAN
14	EGYPT	97	KOREA,
18	GHANA	98	KUWAIT
22	KENYA	100	MALAYSIA
25	MADAGASCAR	102	MYANMAR
30	MOROCCO	106	PHILIPPINES
31	MOZAMBIQUE	108	SAUDI ARABIA
32	NAMIBIA	109	SINGAPORE
34	NIGERIA	110	SRI LANKA
41	SAFRICA	111	SYRIA
44	TANZANIA	112	TAIWAN
46	TUNISIA	113	THAILAND
48	ZAIRE	116	AUSTRIA
49	ZAMBIA	117	BELGIUM
50	ZIMBABWE	119	CYPRUS
52	BARBADOS	120	CZECHOSLOVAKIA
54	CANADA	121	DENMARK
60	GUATEMALA	122	FINLAND
62	HONDURAS	123	FRANCE
64	MEXICO	125	WGERMANY
65	NICARAGUA	126	GREECE
71	TRINIDAD&TOBAGO	129	IRELAND
72	U.S.A.	130	ITALY
73	ARGENTINA	131	LUXEMBOURG
74	BOLIVIA	133	NETHERLANDS
75	BRAZIL	134	NORWAY
76	CHILE	136	PORTUGAL
77	COLOMBIA	137	ROMANIA
81	PERU	138	SPAIN
83	URUGUAY	139	SWEDEN
84	VENEZUELA	140	SWITZERLAND
88	CHINA	141	TURKEY
89	HONG KONG	142	U.K.
90	INDIA	143	USSR
91	INDONESIA	144	YUGOSLAVIA
92	IRAN	145	AUSTRALIA
94	ISRAEL	147	NZ

5 Plotting Data in LimDep

```
PLOT; Lhs=lgdp; Rhs=lsopc$
```

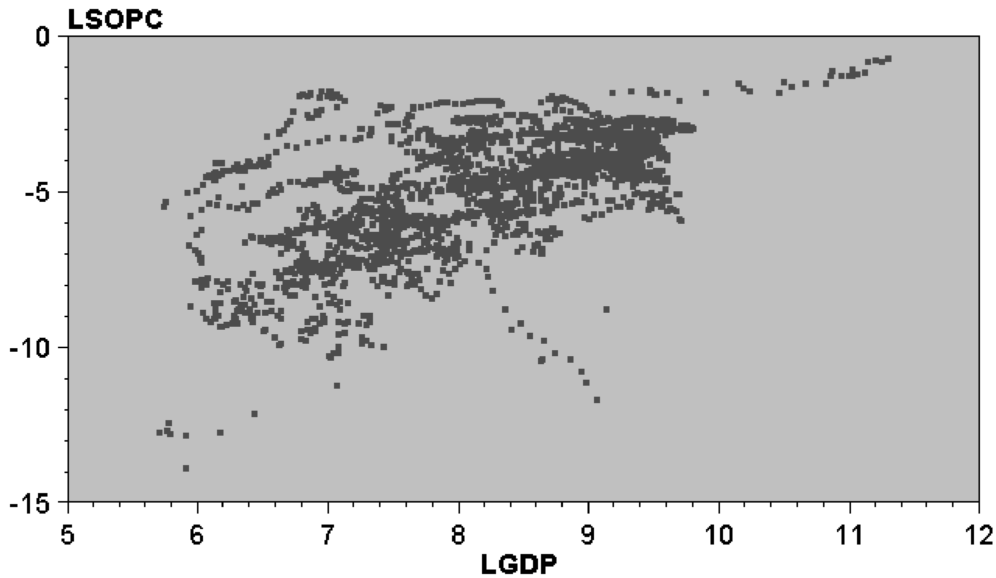

Figure 1: Economics Program Group

6 Running the EKC Panel Regression in LimDep

```
REGRESS  ;Lhs=lsopc
         ;Rhs=ONE,lgdp,lgdpsq
         ;Str=country
         ;Panel$

REGRESS         ;Lhs=lsopc
                ;Rhs=ONE,lgdp,lgdpsq
                ;Str=country
                ;period=year
                ;Panel$
```

The output is divided into four parts.

1. The first part gives the OLS (OLS Without Group Dummy Variables) estimates,
2. the second the fixed effect estimates (Least Squares with Group Dummy Variables),
3. the third the random effect estimates, and
4. the fourth the fixed individuals and time effects estimates.

Note that:

- you can suppress one of the panel models this is useful to compute turning points of a specific model) by including the subcommands **; Fixed** or **; Random**
- it is not necessary to include **ONE** amongst the regressors
- to obtain a list of the fixed effects use the **;Output = 2** subcommand

6.1 OLS

```
+----------------------------------------------------------------------+
| OLS Without Group Dummy Variables                                    |
| Ordinary    least squares regression    Weighting variable = none    |
| Dep. var. = LSOPC    Mean=  -5.033948624    , S.D.=   1.880254548    |
| Model size: Observations =    2294, Parameters =   3, Deg.Fr.=   2291 |
| Residuals:  Sum of squares= 5087.108121    , Std.Dev.=        1.49013 |
| Fit:        R-squared=  .372471, Adjusted R-squared =          .37192 |
| Model test: F[  2,   2291] =  679.91,    Prob value =          .00000 |
| Diagnostic: Log-L =  -4168.5302, Restricted(b=0) Log-L =   -4702.9930 |
|             LogAmemiyaPrCrt.=     .799, Akaike Info. Crt.=      3.637 |
| Panel Data Analysis of LSOPC      [ONE way]                          |
|           Unconditional ANOVA (No regressors)                        |
| Source      Variation       Deg. Free.     Mean Square               |
| Between       7219.07              73.         98.8914               |
| Residual      887.504            2220.         .399776               |
| Total         8106.57            2293.         3.53536               |
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 LGDP      1.820929092      .43802340        4.157   .0000  8.1226219
 LGDPSQ   -.4175485084E-01  .27100655E-01   -1.541   .1234  66.974400
 Constant -17.02816116      1.7524821       -9.717   .0000
```

6.2 Fixed Country Effects

```
+----------------------------------------------------------------------------+
| Least Squares with Group Dummy Variables                                   |
| Ordinary    least squares regression    Weighting variable = none          |
| Dep. var. = LSOPC    Mean=  -5.033948624    , S.D.=   1.880254548          |
| Model size: Observations =    2294, Parameters =  76, Deg.Fr.=   2218      |
| Residuals:  Sum of squares= 756.0200229    , Std.Dev.=         .58383      |
| Fit:        R-squared=  .906740, Adjusted R-squared =          .90359      |
| Model test: F[ 75,   2218] =  287.53,    Prob value =          .00000      |
| Diagnostic: Log-L =  -1981.8929, Restricted(b=0) Log-L =   -4702.9930      |
|             LogAmemiyaPrCrt.=   -1.044, Akaike Info. Crt.=      1.794      |
| Estd. Autocorrelation of e(i,t)     .824207                                |
+----------------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 LGDP      3.253033758      .35333572        9.207   .0000  8.1226219
 LGDPSQ   -.1525441984      .21296400E-01   -7.163   .0000  66.974400
```

6.3 Random Country Effects

```
+------------------------------------------------------------------------+
|                Test Statistics for the Classical Model                 |
|                                                                        |
|        Model            Log-Likelihood    Sum of Squares    R-squared  |
| (1)  Constant term only    -4702.99293   .8106573979D+04     .0000000  |
| (2)  Group effects only    -2165.80845   .8875037643D+03     .8905205  |
| (3)  X - variables only    -4168.53014   .5087108121D+04     .3724713  |
| (4)  X and group effects   -1981.89285   .7560200229D+03     .9067399  |
|                                                                        |
|                                Hypothesis Tests                        |
|              Likelihood Ratio Test               F Tests               |
|           Chi-squared   d.f.  Prob.         F    num. denom. Prob value|
| (2) vs (1)  5074.369     73     .00000   247.367   73  2220     .00000 |
| (3) vs (1)  1068.926      2     .00000   679.914    2  2291     .00000 |
| (4) vs (1)  5442.200     75     .00000   287.533   75  2218     .00000 |
| (4) vs (2)   367.831      2     .00000   192.872    2  2218     .00000 |
| (4) vs (3)  4373.275     73     .00000   174.061   73  2218     .00000 |
+------------------------------------------------------------------------+
```

```
            +-----------------------------------------------------+
            | Random Effects Model: v(i,t) = e(i,t) + u(i)        |
            | Estimates:  Var[e]              =   .340857D+00     |
            |             Var[u]              =   .193366D+01     |
            |             Corr[v(i,t),v(i,s)] =   .850141         |
            | Lagrange Multiplier Test vs. Model (3) =24162.52    |
            | ( 1 df, prob value =  .000000)                      |
            | (High values of LM favor FEM/REM over CR model.)    |
            | Fixed vs. Random Effects (Hausman)     =    5.94    |
            | ( 2 df, prob value =  .051175)                      |
            | (High (low) values of H favor FEM (REM).)           |
            | Reestimated using GLS coefficients:                 |
            | Estimates:  Var[e]              =   .340899D+00     |
            |             Var[u]              =   .208904D+01     |
            |             Sum of Squares          .537828D+04     |
            |             R-squared               .336553D+00     |
            +-----------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 LGDP      3.266327171      .34946779        9.347   .0000  8.1226219
 LGDPSQ   -.1520906198      .21094792E-01   -7.210   .0000  66.974400
 Constant -21.37891105      1.4525931      -14.718   .0000
```

6.4 Fixed Country and Time Effects

Unlike the individual stratification dummy variable, the time dummy variable must be a sequence of integers starting at 1, $(1, 2, \ldots, T)$. As noted earlier, it is not necessary for every group to have data in every period; there may be gaps. With a balanced panel, it is possible to create the needed sequence using the CREATE Trn function. Also, knowing the starting date, one can subtract the starting year minus one to obtain the sequence. In our case, having 31 years, starting in 1960,

```
CREATE  ;  Time  =  Trn(-31,0) $
? or
CREATE  ; Tt = year - 1959 $
```

LimDep will not print the time effects on the screen. To obtain the effects the output has to be saved in a file.

```
? Create file to save output
OPEN;Output=M:panel.out$
? panel with time effects
REGRESS  ;Lhs=lsopc
         ;Rhs=lgdp,lgdpsq
         ;Str=country
         ;period=Time
         ;Output=2
         ;Panel$
```

```
+----------------------------------------------------------------------+
| Least Squares with Group Dummy Variables and Period Effects          |
| Ordinary    least squares regression    Weighting variable = none    |
| Dep. var. = LSOPC    Mean=  -5.033948624    , S.D.=   1.880254548    |
| Model size: Observations =    2294, Parameters = 107, Deg.Fr.=   2187 |
| Residuals:  Sum of squares= 727.7615281    , Std.Dev.=         .57686 |
| Fit:        R-squared=  .910185, Adjusted R-squared =          .90583 |
| Model test: F[106,   2187] =  209.08,    Prob value =          .00000 |
| Diagnostic: Log-L =  -1938.1986, Restricted(b=0) Log-L =   -4702.9930 |
|             LogAmemiyaPrCrt.=   -1.055, Akaike Info. Crt.=      1.783 |
| Estd. Autocorrelation of e(i,t)     .811309                          |
+----------------------------------------------------------------------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 LGDP      3.846522878      .36525929       10.531   .0000  8.1226219
 LGDPSQ   -.1706012228      .21473304E-01   -7.945   .0000  66.974400
 Constant -24.85188530      1.5617902      -15.912   .0000
```

Now the file `panelie.out` will contain the Time effects. Figure 2 shows the time effects with the two standard errors bands.

```
        Estimated Fixed Effects
        Period       Coefficient        Standard Error       t-ratio
             1            .16217                .07128       2.27508
             2            .15322                .07080       2.16421
             3            .07063                .07010       1.00749
             4            .15818                .06958       2.27334
             5            .16781                .06896       2.43346
             6            .17013                .06848       2.48439
             7            .19815                .06816       2.90692
             8            .20700                .06791       3.04816
             9            .15860                .06755       2.34785
            10            .10664                .06715       1.58794
            11            .15045                .06691       2.24847
            12            .12153                .06680       1.81925
            13            .06982                .06678       1.04561
            14            .06100                .06683        .91276
            15            .05629                .06693        .84101
            16           -.01857                .06694       -.27743
            17           -.05375                .06704       -.80177
            18           -.03203                .06719       -.47673
            19           -.10770                .06738      -1.59833
            20           -.10525                .06753      -1.55860
            21           -.06862                .06757      -1.01567
            22           -.06066                .06764       -.89681
            23           -.09541                .06756      -1.41226
            24           -.15720                .06749      -2.32904
            25           -.15758                .06757      -2.33230
            26           -.15921                .06762      -2.35438
            27           -.16811                .06774      -2.48181
            28           -.17548                .06791      -2.58407
            29           -.18165                .06803      -2.67004
            30           -.21141                .06822      -3.09910
            31           -.25898                .06829      -3.79209
```

Figure 2: Time Effects for World Plus or Minus Two sd.

7 Model specification: OLS Vs. Individual Effects, Fixed Vs. Random Effects

First, the (Breusch and Pagan) LM-test is used to test the hypothesis that individual effects are significant. If these tests do not provide any evidence for individual effects, then the model can simply be estimated by ordinary least squares (OLS).

Second, the Hausman test is used to decide whether the regressors are correlated with the individual effect.

The large value of the LM statistic argues in favor of the of a panel data model, the small Hausman statistic argues in favor of the random effect model. A small Hausman statistic argues in favor of the random effect model, a large in favor of a fixed effects one.

Note that:

- you can suppress one of the panel models by including the subcommands **Fixed** or **Random**
- it is not necessary to include **ONE** amongst the regressors
- to obtain a list of the fixed effects use the **;Output = 2** subcommand

8 Selecting a Subsample

```
INCLUDE;oe=1;new$
?Or exclude the observations such that oe=0.
?REJECT; oe=0$
REGRESS;Lhs=lsopc;Rhs=ONE,lgdp,lgdpsq; Str=country; period=year;
    Panel$
```

9 EKC Computations

When you estimate a model, the estimation results are displayed on the screen and in the output file if one is open. In addition, each model produces a number of results which are saved automatically and can be used in subsequent procedures and commands.

```
+----------------------------------------------------------------------+
| Least Squares with Group Dummy Variables and Period Effects          |
| Ordinary    least squares regression    Weighting variable = none    |
| Dep. var. = LSOPC    Mean=  -5.033948624    , S.D.=   1.880254548    |
| Model size: Observations =    2294, Parameters = 107, Deg.Fr.=   2187 |
| Residuals:  Sum of squares= 727.7615281    , Std.Dev.=         .57686 |
| Fit:        R-squared=  .910185, Adjusted R-squared =          .90583 |
| Model test: F[106,   2187] =  209.08,    Prob value =          .00000 |
| Diagnostic: Log-L =  -1938.1986, Restricted(b=0) Log-L =   -4702.9930 |
|             LogAmemiyaPrCrt.=   -1.055, Akaike Info. Crt.=      1.783 |
| Estd. Autocorrelation of e(i,t)     .811309                          |
+----------------------------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 LGDP      3.846522878      .36525929       10.531   .0000  8.1226219
 LGDPSQ   -.1706012228      .21473304E-01   -7.945   .0000  66.974400
 Constant -24.85188530      1.5617902      -15.912   .0000

Matrix ; list ; B $

Matrix Result   has  3 rows and  1 columns.
              1
        +--------------
       1|  .3846523D+01
       2| -.1706012D+00
       3| -.2485189D+02

--> CALC; LIST; TP= -B(1)/(2*B(2)) $
    TP      =  .11273432910213310D+02
--> CALC; LIST; exp(TP)$
    Result  =  .78702713292846510D+05

--> CALC; LIST; VARTP=1/B(2)^2 *  (
            VARB(1,1)   + (B(1)/B(2))^2* VARB(2,2)
                        - 2*(B(1)/B(2))* VARB(1,2)
```

```
                        ) $
    VARTP   =  .63112066734721470D+00
```

```
Matrix ; list ; Varb $

Matrix Result   has  3 rows and  3 columns.
               1             2             3
        +-----------------------------------------
       1|  .1334143D+00 -.7749469D-02  .0000000D+00
       2| -.7749469D-02  .4611028D-03  .0000000D+00
       3|  .0000000D+00  .0000000D+00  .2439189D+01
```

10 Exercise (Due by noon on Friday the 16th of February 2002)

Following Common and Stern's (2001) paper and data provided on the course web page,

- Estimate the EKC for the world data.
- Discuss the economic and statistical significance of the coefficients for GDP in the world equation.
- Based on the Hausman statistic decide which model, fixed or random, should be used.
- A high Hausman statistic implies that there is correlation between country effects and income variables. What could be the most likely cause of this problem?
- Compute the relevant turning points of the estimated curves for the world, OECD and non-OECD regressions.
- How could trade account for the differences in turning point estimates when estimating the EKC for only OECD and only non-OECD countries? Explain.
- Obtain, plot and interpret the fixed time effects for the OECD sample.