ECON4150 - Introductory Econometrics

Lecture 14: Panel data

Monique de Haan
(moniqued@econ.uio.no)

Stock and Watson Chapter 10

OLS: The Least Squares Assumptions

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

Assumption 1: conditional mean zero assumption: $E[u_i|X_i] = 0$

Assumption 2: (X_i, Y_i) are i.i.d. draws from joint distribution

Assumption 3: Large outliers are unlikely

- Under these three assumption the OLS estimators are unbiased, consistent and normally distributed in large samples.

- Last week we discussed threats to internal validity

- In this lecture we discuss a method we can use in case of omitted variables
  - Omitted variable is a determinant of the outcome Y_i
  - Omitted variable is correlated with regressor of interest X_i

Omitted variables

- Multiple regression model was introduced to mitigate omitted variables problem of simple regression

$$Y_i = \beta_0 + \beta_1 X1_i + \beta_2 X2_i + \beta_3 X3_i + ... + \beta_k Xk_i + u_i$$

- Even with multiple regression there is threat of omitted variables:
    - some factors are difficult to measure
    - sometimes we are simply ignorant about relevant factors
- Multiple regression based on panel data may mitigate detrimental effect of omitted variables *without actually observing them*.

Panel data

Cross-sectional data:

A sample of individuals observed in 1 time period

2010

Panel data: same sample of individuals observed in multiple time periods

2010

2011

2012

Panel data; notation

- Panel data consist of observations on n entities (cross-sectional units) and T time periods
- Particular observation denoted with two subscripts (i and t)

$$Y_{it} = \beta_0 + \beta_1 X_{it} + u_{it}$$

- Y_{it} outcome variable for individual i in year t
- For balanced panel this results in nT observations

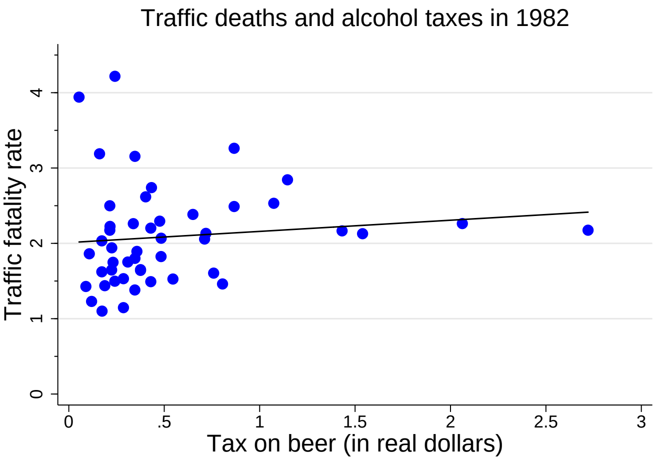
Advantages of panel data

- More control over omitted variables.
- More observations.
- Many research questions typically involve a time component.

The effect of alcohol taxes on traffic deaths

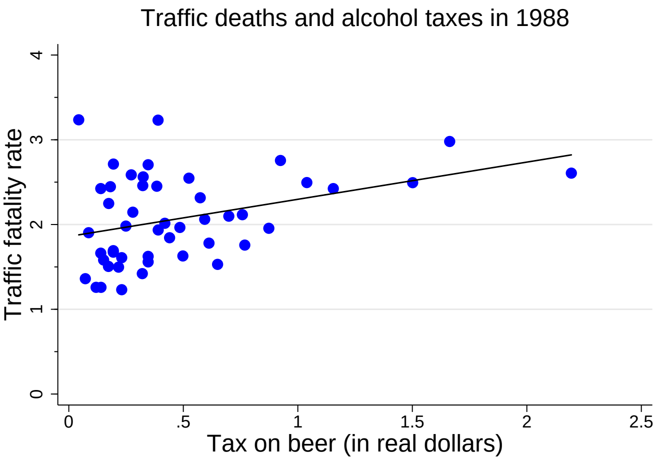
- About 40,000 traffic fatalities each year in the U.S.
- Approximately 25% of fatal crashes involve driver who drunk alcohol.
- Government wants to reduce traffic fatalities.
- One potential policy: increase the tax on alcoholic beverages.
- We have data on traffic fatality rate and tax on beer for 48 U.S. states in 1982 and 1988.
- What is the effect of increasing the tax on beer on the traffic fatality rate?

Data from 1982



$$\widehat{FatalityRate}_{i,1982} = \underset{(0.14)}{2.01} + \underset{(0.18)}{0.15}\ BeerTax_{i,1982}$$

Data from 1988



$$\widehat{FatalityRate}_{i,1988} = \underset{(0.11)}{1.86} + \underset{(0.16)}{0.44} \; BeerTax_{i,1988}$$

Panel data: before-after analysis

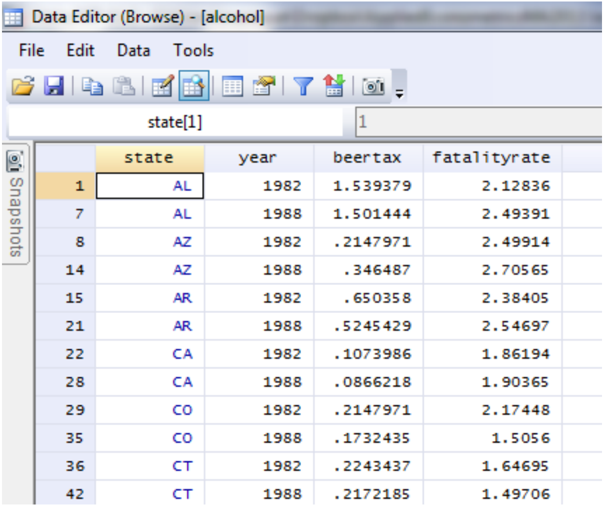
- Both regression using data from 1982 & 1988 likely suffer from omitted variable bias
- We can use data from 1982 and 1988 together as panel data
- Panel data with $T = 2$
- Observed are Y_{i1}, Y_{i2} and X_{i1}, X_{i2}
- Suppose model is

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_i + u_{it}$$

and we assume $E(u_{it}|X_{i1}, X_{i2}, Z_i) = 0$

- Z_i are (unobserved) variables that vary between states but not over time
    - (such as local cultural attitude towards drinking and driving)
- Parameter of interest is β_1

Panel data



	state	year	beertax	fatalityrate
1	AL	1982	1.539379	2.12836
7	AL	1988	1.501444	2.49391
8	AZ	1982	.2147971	2.49914
14	AZ	1988	.346487	2.70565
15	AR	1982	.650358	2.38405
21	AR	1988	.5245429	2.54697
22	CA	1982	.1073986	1.86194
28	CA	1988	.0866218	1.90365
29	CO	1982	.2147971	2.17448
35	CO	1988	.1732435	1.5056
36	CT	1982	.2243437	1.64695
42	CT	1988	.2172185	1.49706

Panel data: before

- Consider cross-sectional regression for first period ($t = 1$):

$$Y_{i1} = \beta_0 + \beta_1 X_{i1} + \beta_2 Z_i + u_{i1} \quad E\left[u_i | X_{i1}, Z_i\right] = 0$$

- Z_i observed: multiple regression of Y_{i1} on constant, X_{i1} and Z_i leads to unbiased and consistent estimator of β_1
- Z_i not observed: regression of Y_{i1} on constant and X_{i1} only results in unbiased estimator of β_1 when $Cov(X_{i1}, Z_i) = 0$
- What can we do if we don't observe Z_i?

Panel data: after

- We also observe Y_{i2} and X_{i2}, hence model for second period is:

$$Y_{i2} = \beta_0 + \beta_1 X_{i2} + \beta_2 Z_i + u_{i2}$$

- Similar to argument before cross-sectional analysis for period 2 might fail
- Problem is again the unobserved heterogeneity embodied in Z_i

Before-after analysis (first differences)

- We have

$$Y_{i1} = \beta_0 + \beta_1 X_{i1} + \beta_2 Z_i + u_{i1}$$

  and

$$Y_{i2} = \beta_0 + \beta_1 X_{i2} + \beta_2 Z_i + u_{i2}$$

- Subtracting period 1 from period 2 gives

$$Y_{i2} - Y_{i1} = (\beta_0 + \beta_1 X_{i2} + \beta_2 Z_i + u_{i2}) - (\beta_0 + \beta_1 X_{i1} + \beta_2 Z_i + u_{i1})$$

- Applying OLS to:

$$Y_{i2} - Y_{i1} = \beta_1 (X_{i2} - X_{i1}) + (u_{i2} - u_{i1})$$

  will produce an unbiased and consistent estimator of β_1

- Advantage of this regression is that we do not need data on Z
- By analyzing changes in dependent variable we automatically control for time-invariant unobserved factors

Data from 1982 and 1988

$$\widehat{Fatality_{i,1988} - Fatality_{i,1982}} = \underset{(0.06)}{-0.07} - \underset{(0.42)}{1.04} \left(BeerTax_{i,1988} - BeerTax_{i,1982} \right)$$

Panel data with more than 2 time periods

Data Editor (Browse) - [alcohol]

File Edit View Data Tools

var25[24]

	state	year	fatalityrate	beertax
1	AL	1982	2.12836	1.539379
2	AL	1983	2.34848	1.788991
3	AL	1984	2.33643	1.714286
4	AL	1985	2.19348	1.652542
5	AL	1986	2.66914	1.609907
6	AL	1987	2.71859	1.56
7	AL	1988	2.49391	1.501444
8	AZ	1982	2.49914	.2147971
9	AZ	1983	2.26738	.206422
10	AZ	1984	2.82878	.2967033
11	AZ	1985	2.80201	.3813559
12	AZ	1986	3.07106	.371517
13	AZ	1987	2.76728	.36
14	AZ	1988	2.70565	.346487
15	AR	1982	2.38405	.650358
16	AR	1983	2.3957	.6754587
17	AR	1984	2.23785	.5989011
18	AR	1985	2.26367	.5773305
19	AR	1986	2.54323	.5624355
20	AR	1987	2.67588	.545
21	AR	1988	2.54697	.5245429

Panel data with more than 2 time periods

- Panel data with $T \geq 2$

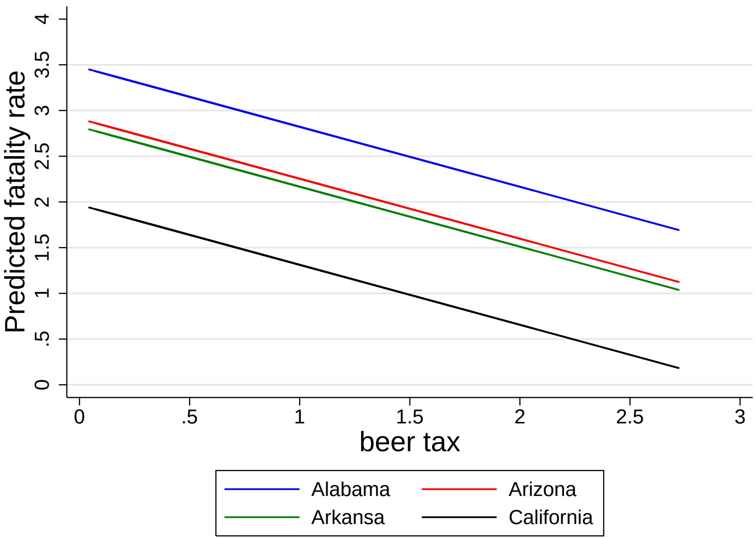
$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 Z_i + u_{it}, \qquad i = 1, ..., n; \quad t = 1, ..., T$$

- Y_{it} is dependent variable; X_{it} is explanatory variable; Z_i are state specific, time invariant variables
- Equation can be interpreted as model with n specific intercepts (one for each state)

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it}, \qquad \textit{with} \quad \alpha_i = \beta_0 + \beta_2 Z_i$$

- α_i, $i = 1, ..., n$ are called entity fixed effects
- α_i models impact of omitted time-invariant variables on Y_{it}

State specific intercepts



Fixed effects regression model

Least squares with dummy variables

Having data on Y_{it} and X_{it} how to determine β_1?

- Population regression model: $Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it}$
- In order to estimate the model we have to quantify α_i
- Solution: create n dummy variables $D1_i, ..., Dn_i$
    - with $D1_i = 1$ if $i = 1$ and 0 otherwise,
    - with $D2_i = 1$ if $i = 2$ and 0 otherwise,....
- Population regression model can be written as:

$$Y_{it} = \beta_1 X_{it} + \alpha_1 D1_i + \alpha_2 D2_i + ... + \alpha_n Dn_i + u_{it}$$

Fixed effects regression model

Least squares with dummy variables

- Alternatively, population regression model can be written as:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + ... + \gamma_n Dn_i + u_{it}$$

with $\beta_0 = \alpha_1$ and $\gamma_i = \alpha_i - \beta_0$ for $i > 1$

- Interpretation of β_1 identical for both representations

- Ordinary Least Squares (OLS): choose $\hat{\beta}_0, \hat{\beta}_1, \hat{\gamma}_2 ..., \hat{\gamma}_n$ to minimize squared prediction mistakes (*SSR*):

$$\sum_{i=1}^{n} \sum_{t=1}^{T} \left(Y_{it} - \hat{\beta}_0 - \hat{\beta}_1 X_{it} - \hat{\gamma}_2 D2_i - ... - \hat{\gamma}_n Dn_i \right)^2$$

- *SSR* is function of $\hat{\beta}_0, \hat{\beta}_1, \hat{\gamma}_2 ..., \hat{\gamma}_n$

Fixed effect regression model

Least squares with dummy variables

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\left(Y_{it}-\hat{\beta}_0-\hat{\beta}_1 X_{it}-\hat{\gamma}_2 D2_i-\ldots-\hat{\gamma}_n Dn_i\right)^2$$

OLS procedure:

- Take partial derivatives of *SSR* w.r.t. $\hat{\beta}_0, \hat{\beta}_1, \hat{\gamma}_2 \ldots, \hat{\gamma}_n$
- Equal partial derivatives to zero resulting in $n+1$ equations with $n+1$ unknown coefficients
- Solutions are the OLS estimators $\hat{\beta}_0, \hat{\beta}_1, \hat{\gamma}_2 \ldots, \hat{\gamma}_n$

Fixed effect regression model

Least squares with dummy variables

- Analytical formulas require matrix algebra
- Algebraic properties OLS estimators (normal equations, linearity) same as for simple regression model
- Extension to multiple X's straightforward: $n + k$ normal equations
- OLS procedure is also labeled Least Squares Dummy Variables (LSDV) method
- Dummy variable trap: Never include all n dummy variables and the constant term!

Fixed effect regression model
Within estimation

- Typically n is large in panel data applications
- With large n computer will face numerical problem when solving system of $n + 1$ equations
- OLS estimator can be calculated in two steps
- First step: demean Y_{it} and X_{it}
- Second step: use OLS on demeaned variables

Fixed effect regression model

Within estimation

- We have

$$
\begin{aligned}
Y_{it} &= \beta_1 X_{it} + \alpha_i + u_{it} \\
\bar{Y}_i &= \beta_1 \bar{X}_i + \alpha_i + \bar{u}_i
\end{aligned}
$$

- $\bar{Y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}$, etc. is entity mean
- Subtracting both expressions leads to

$$
Y_{it} - \bar{Y}_i = (\beta_1 X_{it} + \alpha_i + u_{it}) - (\beta_1 \bar{X}_i + \alpha_i + \bar{u}_i)
$$

$$
\tilde{Y}_{it} = \beta_1 \tilde{X}_{it} + \tilde{u}_{it}
$$

- $\tilde{Y}_{it} = Y_{it} - \bar{Y}_i$, etc. is entity demeaned variable
- α_i has disappeared; OLS on demeaned variables involves solving one normal equation only!

Fixed effect regression model

Within estimation

Data Editor (Browse) - [alcohol]

File Edit View Data Tools

state[1] 1

Snapshots

	state	year	fatalityrate	Meanfatality	DmeanFatal	beertax	Meanbeertax	DmeanBeertax
1	AL	1982	2.12836	2.412627	-.2842672	1.539379	1.623793	-.0844132
2	AL	1983	2.34848	2.412627	-.0641472	1.788991	1.623793	.1651981
3	AL	1984	2.33643	2.412627	-.0761971	1.714286	1.623793	.090493
4	AL	1985	2.19348	2.412627	-.2191472	1.652542	1.623793	.0287497
5	AL	1986	2.66914	2.412627	.2565126	1.609907	1.623793	-.0138856
6	AL	1987	2.71859	2.412627	.3059628	1.56	1.623793	-.0637927
7	AL	1988	2.49391	2.412627	.0812829	1.501444	1.623793	-.122349
8	AZ	1982	2.49914	2.7059	-.2067599	.2147971	.3110403	-.0962432
9	AZ	1983	2.26738	2.7059	-.43852	.206422	.3110403	-.1046183
10	AZ	1984	2.82878	2.7059	.12288	.2967033	.3110403	-.014337
11	AZ	1985	2.80201	2.7059	.0961101	.3813559	.3110403	.0703156
12	AZ	1986	3.07106	2.7059	.36516	.371517	.3110403	.0604767
13	AZ	1987	2.76728	2.7059	.0613801	.36	.3110403	.0489597
14	AZ	1988	2.70565	2.7059	-.0002499	.346487	.3110403	.0354467
15	AR	1982	2.38405	2.435336	-.0512855	.650358	.5905753	.0597827
16	AR	1983	2.3957	2.435336	-.0396357	.6754587	.5905753	.0848835
17	AR	1984	2.23785	2.435336	-.1974857	.5989011	.5905753	.0083258
18	AR	1985	2.26367	2.435336	-.1716657	.5773305	.5905753	-.0132447
19	AR	1986	2.54323	2.435336	.1078944	.5624355	.5905753	-.0281398
20	AR	1987	2.67588	2.435336	.2405446	.545	.5905753	-.0455753
21	AR	1988	2.54697	2.435336	.1116343	.5245429	.5905753	-.0660324

Fixed effect regression model

Within estimation

- Entity demeaning is often called the Within transformation
- Within transformation is generalization of "before-after" analysis to more than $T = 2$ periods
- Before-after: $Y_{i2} - Y_{i1} = \beta_1(X_{i2} - X_{i1}) + (u_{i2} - u_{i1})$
- Within: $Y_{it} - \bar{Y}_i = \beta_1(X_{it} - \bar{X}_i) + (u_{it} - \bar{u}_i)$
- LSDV and Within estimators are identical:

$$\widehat{FatalityRate}_{it} = \underset{(0.19)}{-0.66}\ BeerTax_{it} + State\ dummies$$

$$(\widehat{FatalityRate_{it} - \overline{FatalityRate}}) = \underset{(0.19)}{-0.66}\ (BeerTax_{it} - \overline{BeerTax})$$

Fixed effects regression model

time fixed effects

- In addition to entity effects we can also include time effects in the model
- Time effects control for omitted variables that are common to all entities but vary over time
- Typical example of time effects: macroeconomic conditions or federal policy measures are common to all entities (e.g. states) but vary over time
- Panel data model with entity and time effects:

$$Y_{it} = \beta_1 X_{it} + \alpha_i + \lambda_t + u_{it}$$

Fixed effects regression model

time fixed effects

- OLS estimation straightforward extension of LSDV/Within estimators of model with only entity fixed effects
- LSDV: create T dummy variables $B1_t....BT_t$

$$
\begin{aligned}
Y_{it} = \quad & \beta_0 + \beta_1 X_{it} + \gamma_2 D2_i + ... + \gamma_n Dn_i \\
& + \delta_2 B2_t + \delta_3 B3_t + ... + \delta_T BT_t + u_{it}
\end{aligned}
$$

- Within estimation: Deviating Y_{it} and X_{it} from their entity *and* time-period means
- The effect of the tax on beer on the traffic fatality rate:

$$
\widehat{FatalityRate_{it}} = \underset{(0.20)}{-0.64} \quad BeerTax_{it} \quad + \quad State\ dummies + Time\ dummies
$$

Fixed effects regression model
statistical properties OLS

$$Y_{it} = \beta_1 X_{it} + \alpha_i + \lambda_t + u_{it}$$

statistical assumptions are:

ASS #1: $E\left(u_{it} | X_{i1}, ..., X_{iT}, \alpha_i, \lambda_t\right) = 0$

ASS #2: $(X_{i1}, ..., X_{iT}, Y_{i1}, ..., Y_{iT})$ are i.i.d. over the cross-section

ASS #3: large outliers are unlikely

ASS #4: no perfect multicollinearity

ASS #5: $cov\left(u_{it}, u_{is} | X_{i1}, ..., X_{iT}, \alpha_i, \lambda_t\right) = 0$ for $t \neq s$

Fixed effects regression model

statistical properties OLS

ASS #1 to ASS #5 imply that:

- OLS estimator $\hat{\beta}_1$ is *unbiased* and *consistent* estimator of β_1
- OLS estimators approximately have a normal distribution

remarks:

- ASS #1 is most important
- extension to multiple X's straightforward

$$Y_{it} = \beta_1 X1_{it} + \beta_2 X2_{it} + \ldots + \beta_k Xk_{it} + \alpha_i + \lambda_t + u_{it}$$

- additional assumption ASS #5 implies that error terms are uncorrelated over time (no autocorrelation)

Fixed effects regression model

Clustered standard errors

- Violation of assumption #5: error terms are correlated over time: $(Cov(u_{it}, u_{is}) \neq 0)$
- u_{it} contains time-varying factors that affect the traffic fatality rate (but that are uncorrelated with the beer tax)
- These omitted factors might for a given entity be correlated over time
- Examples: downturn in local economy, road improvement project
- Not correcting for autocorrelation leads to standard errors which are often too low

Fixed effects regression model

Clustered standard errors

- Solution: compute HAC-standard errors (clustered se’s)
  - robust to arbitrary correlation within clusters (entities)
  - robust to heteroskedasticity
  - assume no correlation across entities
- Clustered standard errors valid whether or not there is heteroskedasticity and/or autocorrelation
- Use of clustered standard errors problematic when number of entities is below 50 (or 42)
- In stata: **command, cluster(**entity**)**

The effect of a tax on beer on traffic fatalities

Dependent variable: traffic fatality rate (number of deaths per 10 000)					
Beer tax	0.36***	-0.66***	-0.64***	-0.59***	-0.59*
	(0.06)	(0.19)	(0.20)	(0.18)	(0.33)
State fixed effects	-	yes	yes	yes	yes
Time fixed effects	-	-	yes	yes	yes
Additional control variables	-	-	-	yes	yes
Clustered standard errors	-	-	-	-	yes
N	336	336	336	336	336

Note: * significant at 10% level, ** significant at 5% level, *** significant at 1% level. Control variables: Unemployment rate, per capita income, minimum legal drinking age.

Panel data: an example

returns to schooling

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it}$$

- Y_{it} is logarithm of individual earnings; X_{it} is years of completed education
- α_i unobserved ability
- Likely to be cross-sectional correlation between X_{it} and α_i, hence standard cross-sectional analysis with OLS fails
- However, in this case panel data does not solve the problem because X_{it} typically lacks time series variation ($X_{it} = X_i$)
- We have to resort to cross-sectional methods (instrumental variables) to identify returns to schooling

Panel data: Cigarette taxes and smoking

- Is there an effect of cigarette taxes on smoking behavior?

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it}$$

- Y_{it} number of packages per capita in state i in year t, X_{it} is real tax on cigarettes in state i in year t
- α_i is a state specific effect which includes state characteristics which are constant over time
- Data for 48 U.S. states in 2 time periods: 1985 and 1995

Panel data: Cigarette taxes and smoking

Lpackpc = log number of packages per capita in state *i* in year *t*

rtax = real avr cigarette specific tax during fiscal year in state *i*

Lperinc = log per capita real income

```
. regress lpackpc rtax lperinc

      Source |       SS       df       MS              Number of obs =      96
-------------+------------------------------           F(  2,    93) =   21.25
       Model |  1.76908655     2  .884543277           Prob > F      =  0.0000
    Residual |  3.87049389    93  .041618214           R-squared     =  0.3137
-------------+------------------------------           Adj R-squared =  0.2989
       Total |  5.63958045    95  .059364005           Root MSE      =  .20401

------------------------------------------------------------------------------
     lpackpc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rtax |  -.0156393   .0027975    -5.59   0.000    -.0211946   -.0100839
     lperinc |  -.0139092    .158696    -0.09   0.930    -.3290481    .3012296
       _cons |   5.206614   .3781071    13.77   0.000     4.455769     5.95746
------------------------------------------------------------------------------
```

Panel data: Cigarette taxes and smoking

Before-After estimation

```
. gen diff_rtax= rtax1995- rtax1985

. gen diff_lpackpc= lpackpc1995- lpackpc1985

. gen diff_lperinc= lperinc1995- lperinc1985

. regress  diff_lpackpc diff_rtax diff_lperinc, nocons

      Source |       SS       df       MS              Number of obs =      48
-------------+------------------------------           F(  2,    46) =  145.66
       Model |  3.33475011     2  1.66737506           Prob > F      =  0.0000
    Residual |  .526571782    46  .011447213           R-squared     =  0.8636
-------------+------------------------------           Adj R-squared =  0.8577
       Total |  3.86132189    48  .080444206           Root MSE      =  .10699

------------------------------------------------------------------------------
diff_lpackpc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   diff_rtax |  -.0169369   .0020119    -8.42   0.000    -.0209865   -.0128872
diff_lperinc |  -1.011625   .1325691    -7.63   0.000    -1.278473   -.7447771
```

Panel data: Cigarette taxes and smoking

Least squares with dummy variables (no constant term)

```
. regress lpackpc rtax lperinc stateB*, nocons

      Source |       SS       df       MS              Number of obs =      96
-------------+------------------------------           F( 50,    46) = 7317.61
       Model |  2094.15728    50  41.8831457           Prob > F      =  0.0000
    Residual |  .263285891    46  .005723606           R-squared     =  0.9999
-------------+------------------------------           Adj R-squared =  0.9997
       Total |  2094.42057    96  21.8168809           Root MSE      =  .07565

------------------------------------------------------------------------------
     lpackpc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rtax |  -.0169369   .0020119    -8.42   0.000    -.0209865   -.0128872
     lperinc |  -1.011625   .1325691    -7.63   0.000    -1.278473   -.7447771
     stateB1 |   7.663688   .3037711    25.23   0.000     7.052229    8.275148
     stateB2 |   7.834448   .2926539    26.77   0.000     7.245367     8.42353
     stateB3 |   7.678433   .3121525    24.60   0.000     7.050103    8.306763
     stateB4 |    7.66627   .3392221    22.60   0.000     6.983451    8.349088
                                     :       :       :
    stateB45 |   7.844359   .3193189    24.57   0.000     7.201603    8.487114
    stateB46 |    7.92666   .3154175    25.13   0.000     7.291758    8.561563
    stateB47 |   7.644741   .2936826    26.03   0.000     7.053589    8.235894
    stateB48 |   7.825943   .3275694    23.89   0.000      7.16658    8.485306
```

Panel data: Cigarette taxes and smoking

Least squares with dummy variables with constant term

```
. regress lpackpc rtax lperinc stateB*

      Source |       SS       df       MS              Number of obs =      96
-------------+------------------------------           F( 49,    46) =   19.17
       Model |  5.37629455    49  .109720297           Prob > F      =  0.0000
    Residual |  .263285891    46  .005723606           R-squared     =  0.9533
-------------+------------------------------           Adj R-squared =  0.9036
       Total |  5.63958045    95  .059364005           Root MSE      =  .07565

------------------------------------------------------------------------------
     lpackpc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rtax |  -.0169369   .0020119    -8.42   0.000    -.0209865   -.0128872
     lperinc |  -1.011625   .1325691    -7.63   0.000    -1.278473   -.7447771
     stateB1 |  -.1530275   .0900694    -1.70   0.096    -.3343279    .0282728
     stateB2 |   .0177322   .1005272     0.18   0.861    -.1846185     .220083

                                      :        :        :
           .
    stateB42 |   -.771239   .0918679    -8.40   0.000    -.9561594   -.5863186
    stateB43 |  (dropped)
    stateB44 |   .1757536   .0854144     2.06   0.045     .0038233     .347684
    stateB45 |   .0276429   .0948094     0.29   0.772    -.1631985    .2184843
    stateB46 |   .1099444   .0918156     1.20   0.237    -.0748708    .2947597
    stateB47 |  -.1719747   .0959042    -1.79   0.080    -.3650198    .0210705
    stateB48 |   .0092272   .0787188     0.12   0.907    -.1492255      .16768
       _cons |   7.816716   .3458507    22.60   0.000     7.120554    8.512877
------------------------------------------------------------------------------
```

Panel data: Cigarette taxes and smoking

Within estimation

```
. xtreg lpackpc rtax lperinc, fe i(STATE)

Fixed-effects (within) regression               Number of obs      =        96
Group variable: STATE                           Number of groups   =        48

R-sq:  within  = 0.8636                         Obs per group: min =         2
       between = 0.0896                                        avg =       2.0
       overall = 0.2354                                        max =         2

                                                F(2,46)            =    145.66
corr(u_i, Xb)  = -0.5687                        Prob > F           =    0.0000

------------------------------------------------------------------------------
     lpackpc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        rtax |  -.0169369   .0020119    -8.42   0.000    -.0209865   -.0128872
     lperinc |  -1.011625   .1325691    -7.63   0.000    -1.278473   -.7447771
       _cons |   7.856714   .3150362    24.94   0.000     7.222579    8.490849
-------------+----------------------------------------------------------------
     sigma_u |  .25232518
     sigma_e |  .07565452
         rho |  .91751731   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(47, 46) =    13.41             Prob > F = 0.0000
```