ECON5103 Panel Data

Basic linear panel data

Introduction

Let's start with an example:

- 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 fatality rate.
- 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?

$$y_{it} = \beta_0 + \beta_1 x_{it} + v_{it}$$

- y_{it} traffic fatality rate in state i in year t, x_{it} is real tax on beer in state i in year t

Example: the effect of the beer tax on the traffic fatality rate

Data set in wide form:

	state	bee~1982	fat~1982	bee~1988	fat~1988
1.	AL	1.539379	2.12836	1.501444	2.49391
2.	AZ	.2147971	2.49914	.346487	2.70565
3.	AR	.650358	2.38405	.5245429	2.54697
4.	CA	.1073986	1.86194	.0866218	1.90365
5.	CO	.2147971	2.17448	.1732435	1.5056
6.	CT	.2243437	1.64695	.2172185	1.49706
7.	DE	.173031	2.03333	.1395573	2.42424
8.	FL	1.073986	2.53197	1.039461	2.49534
9.	GA	2.720764	2.17484	2.194418	2.60643
10.	ID	.4027447	2.61759	.3248316	2.56231

Example: the effect of the beer tax on the traffic fatality rate

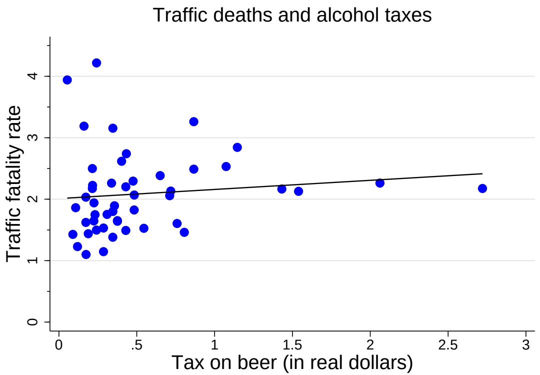
To analyze data, reshape to long form:

```
1 . reshape long fatalityrate beertax, i(state) j(year)
  (note: j = 1982 1988)

  Data                               wide   ->   long
  -----------------------------------------------------------------------------
  Number of obs.                       48   ->      96
  Number of variables                   5   ->       4
  j variable (2 values)                     ->   year
  xij variables:
       fatalityrate1982 fatalityrate1988    ->   fatalityrate
                 beertax1982 beertax1988    ->   beertax
  -----------------------------------------------------------------------------

2 . list, sep(2)
```

	state	year	beertax	fatali~e
1.	AL	1982	1.539379	2.12836
2.	AL	1988	1.501444	2.49391
3.	AZ	1982	.2147971	2.49914
4.	AZ	1988	.346487	2.70565
5.	AR	1982	.650358	2.38405
6.	AR	1988	.5245429	2.54697
7.	CA	1982	.1073986	1.86194
8.	CA	1988	.0866218	1.90365

Example: the effect of the beer tax on the traffic fatality rate

We can analyze data as if it is cross section data:




```
1 . regress fatalityrate beertax, noheader
```

fatalityrate	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
beertax	**.2684732**	**.1258346**	**2.13**	**0.035**	**.0186257**	**.5183207**
_cons	**1.943759**	**.0872035**	**22.29**	**0.000**	**1.770614**	**2.116904**

Example: the effect of the beer tax on the traffic fatality rate

$$y_{it} = \beta_0 + \beta_1 x_{it} + v_{it}$$

- OLS provides consistent estimate of β if $E[v_{it}|x_{it}] = 0$ for all $t = 1, ..., T$
- Assumption fails if v_{it} contains state specific characteristics that are correlated with the beer tax.

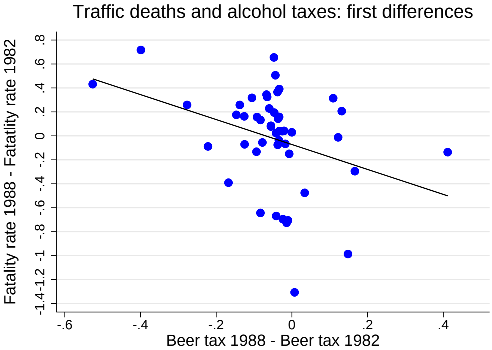
$$y_{it} = \beta_0 + \beta_1 x_{it} + c_i + u_{it}, \quad E[c_i|x_{it}] \neq 0$$

- If the (unobserved) state characteristics that are correlated with x_{it} are constant over time we can use first differencing to eliminate c_i

$$y_{it} - y_{it-1} = \beta(x_{it} - x_{it-1}) + (u_{it} - u_{it-1})$$

- OLS provides consistent estimate of β if $E[(u_{it} - u_{it-1})|(x_{it} - x_{it-1})] = 0$ (strict exogeneity assumption)
- Is this assumption credible? (more on this later)

Example: the effect of the beer tax on the traffic fatality rate




```
1 . regress D.fatalityrate D.beertax, noheader
```

D. fatalityrate	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
beertax D1.	**-1.040973**	**.4172279**	**-2.49**	**0.016**	**-1.880809**	**-.2011364**
_cons	**-.0720371**	**.060644**	**-1.19**	**0.241**	**-.1941072**	**.050033**

Motivation for using panel data

Panel data follows agents over several time-periods, outcomes and characteristics of individuals are observed at multiple points in time.

Advantages of panel data compared to cross-section data:

1. More observations ($N \times T$)
   - Improves the precision of the estimators. However, observations of the same individual are very likely to be correlated over time.
2. Learn about dynamics
3. Robust to certain types of omitted variable bias
4. Additional source of variation
   - time vs cross section

Balanced versus unbalanced panels

We consider *balanced* panels: all units i are observed for all T time periods

Unbalanced panels arise because of

- Attrition
  - track of some individual units is lost at some point (e.g. individuals die or retire, firms go bankrupt, traders leave a market. . .)
- Entry
  - new individuals enter the panel at some point (e.g. individuals turning 16 years old enter household surveys).
- Exogenous attrition (exit)
  - Independent of the dependent variable $\rightarrow$ balanced and unbalanced panels share the same properties
- Endogenous attrition
  - Needs modeling

The unobserved effects model

The standard panel data model is

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + \underbrace{c_i + u_{it}}_{v_{it}} = X_{it}\beta + \underbrace{c_i + u_{it}}_{v_{it}}$$

for $i = 1, \ldots, N$ individuals over $t = 1, \ldots, T$ time periods.

Note: we assume that T is fixed and $N \longrightarrow \infty$

Here $\underset{1x(K+1)}{X_{it}} = (1 \; x_{it1} \; \cdots \; x_{itK})$, and $\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{pmatrix}$.

Model can also be written as

$$Y_i = X_i\beta + \underbrace{c_i J_T + U_i}_{V_i}$$

with $\underset{Tx1}{Y_i} = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix}$, $\underset{Tx(K+1)}{X_i} = \begin{pmatrix} X_{i1} \\ \vdots \\ X_{iT} \end{pmatrix}$, $\underset{Tx1}{J_T} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$, $\underset{Tx1}{U_i} = \begin{pmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{pmatrix}$

The unobserved effects model

$$y_{it} = X_{it}\beta + \underbrace{c_i + u_{it}}_{v_{it}}$$

How to estimate the above equation depends mainly on whether X_{it} is correlated with c_i or not.

Random effects framework: $E[c_i | X_{i1}, \ldots, X_{iT}] = E[c_i] \, (= 0)$

- We assume that regressors are uncorrelated with the unobserved component c_i
- Estimate by pooled OLS or GLS (random effects analysis)

Fixed effects framework: $E[c_i | X_{i1}, \ldots, X_{iT}] \neq E[c_i]$

- We allow regressors to be correlated with the unobserved component c_i
- Estimate by Least Squares with Dummy Variables (LSDV), within estimation, or first differences

Pooled OLS

$$y_{it} = X_{it}\beta + v_{it}, \quad v_{it} = c_i + u_{it}$$

Estimating the equation by OLS gives consistent estimate of β if:

Pooled OLS assumption: $E[X_{it}c_i] = 0$ and $E[X_{it}u_{it}] = 0$

The effect of the beer tax on traffic fatalities ($N = 48$ states, $T = 7$ years):

```
1 . regress fatalityrate beertax, cluster(state)

Linear regression                               Number of obs =        336
                                                   F(  1,    47) =       9.28
                                                   Prob > F      =     0.0038
                                                   R-squared     =     0.0934
                                                   Root MSE      =     .54374

                               (Std. Err. adjusted for    48 clusters in state)
------------------------------------------------------------------------------
             |               Robust
fatalityrate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |   .3646054   .1196856     3.05   0.004     .1238291    .6053818
       _cons |   1.853308   .1185192    15.64   0.000     1.614878    2.091738
------------------------------------------------------------------------------
```

Use clustered standard errors to account for the serial correlation in v_{it}.

The strict exogeneity assumption

- Apart from pooled OLS all methods to analyze panel data rely on (a form of) the strict exogeneity assumption.

Strict exogeneity assumption: $E\left[u_{it}|c_i,\ X_{i1},\ \ldots\ ,\ X_{iT}\right] = 0, \qquad t = 1, \ldots, T$

- Conditional on the unobserved effect the explanatory variables in each time period are uncorrelated with the idiosyncratic error in each time period, $E[X_{it} u_{is}] = 0$ for all s, t.

This rules out:

- Lagged dependent variables
- Feedback (x_{it} depends on $y_{i,t-1}$),
- and other types of endogenous regressors

The strict exogeneity assumption

Failure of strict exogeneity example 1:

$$y_{it} = \beta_0 + \beta_1 \cdot prog_{i,t} + c_i + u_{it}$$

where $X_{it} = prog_{i,t}$. The strict exogeneity assumption implies:

$$E[u_{it} | c_i, prog_{i1}, \ldots, prog_{iT}] = 0, \qquad t = 1, \ldots, T$$

This assumption is violated if:

- program participation has a lasting effect. This can be solved by including lagged values ($prog_{it-1},\ prog_{it-2}$)
- program participation in the next period ($prog_{it+1}$) depends on shocks to the outcome in this period (u_{it})

The strict exogeneity assumption

Failure of strict exogeneity example 2:

$$fatalityrate_{it} = \beta_0 + \beta_1 \cdot beertax_{it} + c_i + u_{it}$$

The strict exogeneity assumption implies:

$$E[u_{it}|c_i, beertax_{i1}, \ldots, beertax_{iT}] = 0, \qquad t = 1, \ldots, T$$

This assumption is violated if:

- a sudden increase in the number of alcohol related traffic deaths induces policy makers to increase the tax on beer in the next period (feedback).

The strict exogeneity assumption

Failure of strict exogeneity example 3:

$$y_{it} = \beta_0 + \beta_1 \cdot y_{i,t-1} + c_i + u_{it}$$

where $X_{it} = y_{i,t-1}$. The strict exogeneity assumption implies:

$$E[u_{it}|c_i, X_{i1}, \ldots, X_{iT}] = 0$$

then this implies that

$$E[y_{it} u_{it}] = E[y_{i,t-1} u_{it}]\beta_1 + E[c_i u_{it}] + E[u_{it}^2] = E[u_{it}^2] > 0$$

and since $E[y_{it} u_{it}] \equiv E[X_{i,t+1} u_{it}] \Rightarrow E[u_{it}|X_{i,t+1}, \ldots, X_{iT}] \neq 0$.

- Strict exogeneity never holds in unobserved effects models with lagged dependent variables!
- Note that since $y_{i,t-1}$ is correlated with c_i the exogeneity assumption required for OLS also fails

Random effects analysis

If the random effects framework holds Pooled OLS is consistent but not efficient.

$$y_{it} = X_{it}\beta + \underbrace{c_i + u_{it}}_{v_{it}}$$

The random effects approach exploits the serial correlation in the composite error, $v_{it} = c_i + u_{it}$, in a GLS framework.

Random Effects Assumption 1:

- **Strict exogeneity:** $E\left[u_{it}|c_i,\ X_{i1},\ \ldots\ ,\ X_{iT}\right] = 0, \qquad t = 1, \ldots, T$

Random effects analysis

Random Effects Assumption 2:

- **Orthogonality:** $E\left[c_i | X_{i1}, \ldots, X_{iT}\right] = E\left[c_i\right] = 0$

In the standard random effects analysis we further assume:

Random Effects Assumption 3: $E\left[U_i U_i' | X_i, c_i\right] = \sigma_u^2 I_T$ and $E[c_i^2 | X_i] = \sigma_c^2$

This implies homoskedasticity, and no serial correlation in u_{it}

Random effects analysis

We can write the model as

$$Y_i = X_i\beta + V_i, \quad V_i = c_i J_T + U_i$$

Under the random effects assumptions we have that:

$$\underset{T \times T}{\Omega} = E[V_i V_i'] = \begin{pmatrix} \sigma_c^2 + \sigma_u^2 & \sigma_c^2 & \cdots & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2 + \sigma_u^2 & & \vdots \\ \vdots & & \ddots & \sigma_c^2 \\ \sigma_c^2 & \cdots & \cdots & \sigma_c^2 + \sigma_u^2 \end{pmatrix} = \sigma_u^2 I_T + \sigma_c^2 J_T J_T'$$

If σ_c^2 and σ_u^2 are known, then the *Generalized Least Squares* (GLS) estimator for β is the *Best Linear Unbiased Estimator* (BLUE)

$$\hat{\beta}_{GLS} = \sum_i \left(X_i' \Omega^{-1} X_i \right)^{-1} \sum_i X_i \Omega^{-1} Y_i$$

However, usually Ω is unknown and needs to be estimated $\rightarrow$ Feasible GLS

The RE FGLS Estimator

We need a consistent estimate of $\Omega = E[V_i V_i']$

1. Estimate $y_{it} = X_{it}\beta + v_{it}$ by pooled OLS and obtain the residuals $\hat{v}_{it}$

   Under the random effects assumptions the following are consistent estimators:

$$\hat{\sigma}_v^2 = \frac{1}{(NT-K)} \sum_{i=1}^N \sum_{t=1}^T \hat{v}_{it}^2$$

$$\hat{\sigma}_c^2 = \frac{1}{[NT(T-1)/2-K]} \sum_{i=1}^N \sum_{t=1}^{T-1} \sum_{s=t+1}^T \hat{v}_{it}\hat{v}_{is}$$

$$\hat{\sigma}_u^2 = \hat{\sigma}_v^2 - \hat{\sigma}_c^2$$

2. Obtain the RE Feasible Generalized Least Squares Estimator:

$$\hat{\beta}_{RE} = \left(\sum_{i=1}^N X_i'\hat{\Omega}^{-1}X_i\right)^{-1}\left(\sum_{i=1}^N X_i'\hat{\Omega}^{-1}Y_i\right)$$

   with

$$\hat{\Omega} = \hat{\sigma}_u^2 I_T + \hat{\sigma}_c^2 J_T J_T'$$

The RE FGLS Estimator

- The RE FGLS estimator is consistent under the first and second random effect assumption and the rank condition: $rank\ E\left[X_i'\hat{\Omega}^{-1}X_i\right] = K$.

- The RE FGLS estimator is $\sqrt{N}$-efficient if in addition the third random effects assumption holds

- Note: $\hat{\sigma}_c^2 = \frac{1}{[NT(T-1)/2-K]}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\hat{v}_{it}\hat{v}_{is}$ need not be positive

- A negative value for $\hat{\sigma}_c^2$ indicates serial correlation in u_{it} which means that the third random effects assumption is violated.

- If the third random effects assumption is violated we can
  - use general FGLS
  - compute a robust variance matrix

General FGLS

- If you suspect serial correlation in u_{it}, or a variance of u_{it} that is not constant over time:

1. obtain the Pooled OLS residuals $\hat{V}_i$ and obtain

$$\hat{\Omega} = \frac{1}{N} \sum_{i=1}^{N} \hat{V}_i \hat{V}_i'$$

2. obtain the general FGLS estimator

$$\hat{\beta}_{GFGLS} = \left(\sum_{i=1}^{N} X_i' \hat{\Omega}^{-1} X_i \right)^{-1} \left(\sum_{i=1}^{N} X_i' \hat{\Omega}^{-1} Y_i \right)$$

General FGLS

- No efficiency loss when $N \to \infty$
- General FGLS is asymptotically more efficient if $E\left[V_i V_i' | X_i\right] = E\left[V_i V_i'\right]$ but Ω does not have the random effects form.
- But if N is not several times larger than T, the general FGLS estimator can have poor finite sample properties
- Note that now we estimate $T(T+1)/2$ parameters instead of 2 with the standard RE covariance structure

Robust variance covariance matrix

- If the third random effects assumption is violated the RE FGLS estimator of β is consistent
- but the standard errors will be incorrect and statistical inference using these incorrect standard errors will be invalid
- It is always possible to obtain a variance covariance matrix that is robust to any type of serial correlation and heteroskedasticity
- Obtain the random effects residuals: $\hat{V}_i = Y_i - X_i\hat{\beta}_{RE}$
- and compute the robust variance covariance matrix

$$\hat{V} = \left(\sum_i X_i'\hat{\Omega}^{-1}X_i\right)^{-1}\left(\sum_i X_i'\hat{\Omega}^{-1}\hat{V}_i\hat{V}_i'\hat{\Omega}^{-1}X_i\right)\left(\sum_i X_i'\hat{\Omega}^{-1}X_i\right)^{-1}$$

The effect of the beer tax on traffic fatalities
Random effects analysis

Standard random effects analysis:

```
1 . xtreg fatalityrate beertax, re theta

  Random-effects GLS regression                    Number of obs      =       336
  Group variable:  state                              Number of groups   =        48

  R-sq:  within  =  0.0407                            Obs per group: min =         7
         between =  0.1101                                           avg =       7.0
         overall =  0.0934                                           max =         7

                                                      Wald chi2( 1)        =      0.18
  corr(u_i, X)    =  0 (assumed)                    Prob > chi2        =    0.6753
  theta           =  .86220102

  ------------------------------------------------------------------------------------
  fatalityrate |     Coef.    Std. Err.       z     P>|z|      [95% Conf. Interval]
  -------------+----------------------------------------------------------------------
       beertax |  -.0520158   .1241758    -0.42    0.675     -.2953959     .1913643
         _cons |   2.067141   .0999715    20.68    0.000      1.871201     2.263082
  -------------+----------------------------------------------------------------------
       sigma_u |   .5157915
       sigma_e |  .18985942
           rho |  .88067496   (fraction of variance due to u_i)
  ------------------------------------------------------------------------------------
```

The effect of the beer tax on traffic fatalities
Random effects analysis

Random effects analysis with cluster-robust standard errors

```
1 . xtreg fatalityrate beertax, re cluster(state)

  Random-effects GLS regression                   Number of obs      =       336
  Group variable:  state                            Number of groups   =        48

  R-sq:  within   =  0.0407                         Obs per group: min =         7
         between =  0.1101                                         avg =       7.0
         overall =  0.0934                                         max =         7

                                                    Wald chi2( 1)        =      0.22
  corr(u_i, X)    =  0 (assumed)                  Prob > chi2          =    0.6373

                                  (Std. Err. adjusted for    48 clusters in state)
  ------------------------------------------------------------------------------
               |               Robust
  fatalityrate |      Coef.   Std. Err.       z     P>|z|     [95% Conf. Interval]
  -------------+----------------------------------------------------------------
       beertax |  -.0520158   .1103327    -0.47    0.637    -.2682638    .1642323
         _cons |   2.067141   .1212281    17.05    0.000     1.829539    2.304744
  -------------+----------------------------------------------------------------
       sigma_u |   .5157915
       sigma_e |  .18985942
           rho |  .88067496   (fraction of variance due to u_i)
  ------------------------------------------------------------------------------
```

Quasi-differencing method

With some algebra it can be shown that the RE FGLS estimator can also be obtained using the following quasi-differencing method

$$y_{it} - \hat{\theta}\bar{y}_i = (X_{it} - \hat{\theta}\bar{X}_i)\beta + (u_{it} - \hat{\theta}\overline{u}_i)$$

where

$$\hat{\theta} = 1 - \frac{\hat{\sigma}_u}{\sqrt{\hat{\sigma}_u^2 + T\hat{\sigma}_c^2}}$$

This transformed equation can be estimated by OLS and now it is especially easy to perform cluster-robust inference

- $\hat{\theta} = 0$: RE=Pooled OLS
- $\hat{\theta} = 1$: RE=Within (FE)

This relies on consistent estimates of σ_u^2 and σ_c^2

Quasi-differencing method

After *xtreg, re* Stata saves $\hat{\theta}$ as *e(theta)*

```
1 . bys state: egen Mfatality=mean(fatality)

2 . bys state: egen Mbeertax=mean(beertax)

3 . gen yquasi= fatalityrate-e(theta)* Mfatality

4 . gen xquasi= beertax-e(theta)*Mbeertax

5 . regress yquasi xquasi, noheader
```

yquasi	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
xquasi	-.0520158	.1241758	-0.42	0.676	-.296281	.1922495
_cons	.2848499	.013776	20.68	0.000	.2577513	.3119485

```
6 . regress yquasi xquasi, cluster(state) noheader
                              (Std. Err. adjusted for    48 clusters in state)
```

yquasi	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
xquasi	-.0520158	.1103327	-0.47	0.640	-.2739764	.1699449
_cons	.2848499	.0167051	17.05	0.000	.2512436	.3184563

Random effects with maximum likelihood

- If we are willing to assume normality of the errors, we can also estimate the standard random effects model by maximum likelihood
- We maximize the likelihood function with respect to β, σ_c^2 and σ_u^2
- For given σ_c^2 and σ_u^2 the maximum likelihood estimator of β is the same as the GLS estimator
- But MLE gives estimators, $\tilde{\sigma}_c^2$ and $\tilde{\sigma}_u^2$ that differ from those used by Feasible GLS (shown on slide 19)
- Asymptotically the MLE and FGLS estimators of the standard random effects model are equivalent but they differ in finite samples.
- The MLE estimate of β can also be obtained by the quasi differencing method using the alternative estimate of θ

$$\widetilde{\theta} = 1 - \frac{\tilde{\sigma}_u}{\sqrt{\tilde{\sigma}_u^2 + T\tilde{\sigma}_c^2}}$$

Random effects with maximum likelihood

```
1 . xtreg fatality beertax, re mle

  Fitting constant-only model:
  Iteration 0:   log likelihood =  -21.873518
  Iteration 1:   log likelihood =  -20.933238
  Iteration 2:   log likelihood =   -20.91122
  Iteration 3:   log likelihood =  -20.911211

  Fitting full model:
  Iteration 0:   log likelihood =  -26.399609
  Iteration 1:   log likelihood =  -21.063077
  Iteration 2:   log likelihood =  -20.771776
  Iteration 3:   log likelihood =  -20.765275
  Iteration 4:   log likelihood =  -20.765269

  Random-effects ML regression                   Number of obs      =       336
  Group variable:  state                           Number of groups   =        48

  Random effects u_i ~  Gaussian                  Obs per group: min =         7
                                                               avg =       7.0
                                                               max =         7

                                                 LR chi2(   1)         =      0.29
  Log likelihood   =  -20.765269                   Prob > chi2          =    0.5890
```

fatalityrate	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
beertax	-.0752753	.1409581	-0.53	0.593	-.3515482	.2009975
_cons	2.079079	.1077579	19.29	0.000	1.867878	2.290281
/sigma_u	.548473	.06273			.4383304	.6862919
/sigma_e	.1926579	.0081676			.1772968	.20935
rho	.8901665	.0244921			.8344365	.9309605

Random effects with maximum likelihood, quasi differencing

```
. gen thetaB=1-(e(sigma_e)/(7*e(sigma_u)^2+e(sigma_e)^2)^(1/2))

. bys state: egen Mfatality=mean(fatality)

. bys state: egen Mbeertax=mean(beertax)

. gen yquasiB= fatalityrate-thetaB* Mfatality

. gen xquasiB= beertax-thetaB*Mbeertax

. regress yquasiB xquasiB

      Source |       SS           df       MS      Number of obs   =       336
-------------+----------------------------------   F(1, 334)       =      0.35
       Model |  .013195566         1  .013195566   Prob > F        =    0.5526
    Residual |  12.4713386       334  .037339337   R-squared       =    0.0011
-------------+----------------------------------   Adj R-squared   =   -0.0019
       Total |  12.4845342       335  .037267266   Root MSE        =    .19323

------------------------------------------------------------------------------
     yquasiB |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     xquasiB |  -.0752753   .1266257    -0.59   0.553    -.3243598    .1738091
       _cons |   .2736274   .0135754    20.16   0.000     .2469233    .3003314
------------------------------------------------------------------------------
```

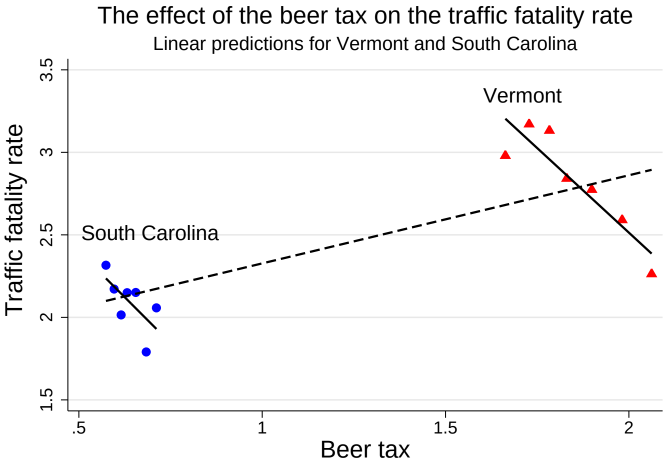
Fixed effects framework

- In most economic applications the explanatory variables of interest are **unlikely** to be **uncorrelated** with the unobserved effect
- This brings us to the fixed effects framework

Fixed effects framework: $E\left[c_i | X_{i1}, \ldots, X_{iT}\right] \neq E\left[c_i\right]$

- We allow regressors to be correlated with the unobserved component c_i
- Estimate by Least Squares with Dummy Variables (LSDV), within estimation, or first differences

The effect of the beer tax on the traffic fatality rate



Fixed effects analysis

Fixed effects assumption 1:

- **Strict exogeneity:** $E\left[u_{it}|c_i,\ X_{i1},\ \ldots\ ,\ X_{iT}\right] = 0, \qquad t = 1, \ldots, T$

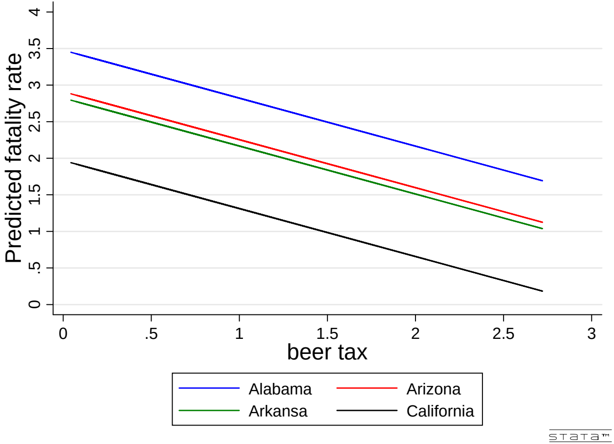
Fixed effects assumption 2: $E\left[U_i U_i' |X_i, c_i\right] = \sigma_u^2 I_T$

- Under fixed effects assumption 1 (and a rank condition) we can use Least Squares with Dummy Variables , within estimation or first differences to get a consistent estimate of β
- If in addition fixed effects assumption 2 hold LSDV and within estimation are efficient.
- NOTE: we can only obtain consistent estimates of time varying regressors!

State specific intercepts, the beer tax and the traffic fatality rate

$$y_{it} = X_{it}\beta + c_i + u_{it}$$

In the fixed effects framework c_i can be interpreted as unit specific intercepts.



Least Squares With Dummy Variables

$$y_{it} = X_{it}\beta + c_i + u_{it}$$

- One way to estimate this equation is to create *N* dummy variables

$$d1_i, \cdots, dN_i \quad with \quad d1_i = 1 \quad if \quad i = 1, \quad etc$$

- Include *N* dummy variables, exclude constant term and estimate by OLS

$$y_{it} = X_{it}\beta + \alpha_1 d1_i + \alpha_2 d2_i + ... + \alpha_N dN_i + u_{it}$$

Least Squares With Dummy Variables

- Or $N-1$ dummy variables, include constant term and estimate by OLS

$$y_{it} = \alpha + X_{it}\beta + \alpha_2 d2_i + \ldots + \alpha_N dN_i + u_{it}$$

- Under fixed effects assumption 1 LSDV gives a consistent estimate of β for T fixed and $N \longrightarrow \infty$
- But the estimates of $\alpha_1, \cdots, \alpha_N$ are only consistent for $T \longrightarrow \infty$
- Not problematic if interest is in estimating causal effect of X_{it} but it is problematic for forecasting.

Least Squares With Dummy Variables

```
1 . qui tab state, gen(State)

2 . regress fatalityrate beertax State*, noconstant noheader
```

fatalityrate	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
beertax	-.6558736	.18785	-3.49	0.001	-1.025612	-.2861352
State1	3.47763	.3133568	11.10	0.000	2.860861	4.094399
State2	2.909903	.0925389	31.45	0.000	2.727762	3.092044
State3	2.822678	.1321253	21.36	0.000	2.562621	3.082736
State4	1.968161	.0740068	26.59	0.000	1.822496	2.113826
State5	1.99335	.0803709	24.80	0.000	1.835159	2.151541
State6	1.615373	.083913	19.25	0.000	1.45021	1.780536
State7	2.170028	.0774569	28.02	0.000	2.017572	2.322484
State8	3.2095	.2215135	14.49	0.000	2.773503	3.645497
State9	4.002233	.4640315	8.62	0.000	3.088896	4.915569
State10	2.808608	.0987666	28.44	0.000	2.614209	3.003006
State11	1.516008	.0784782	19.32	0.000	1.361542	1.670473
State12	2.016088	.0886722	22.74	0.000	1.841558	2.190619
State13	1.933698	.1022168	18.92	0.000	1.732508	2.134888
State14	2.254414	.1086317	20.75	0.000	2.040598	2.46823
State15	2.260113	.0804616	28.09	0.000	2.101743	2.418483
	⋮	⋮	⋮			
State46	2.580876	.1076679	23.97	0.000	2.368957	2.792795
State47	1.718364	.0774569	22.18	0.000	1.565908	1.870819
State48	3.249126	.0723283	44.92	0.000	3.106765	3.391488

Least Squares With Dummy Variables

```
1 . drop State1

2 . regress fatalityrate beertax State*, noheader
```

fatalityrate	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
beertax	-.6558736	.18785	-3.49	0.001	-1.025612	-.2861352
State2	-.5677268	.2666662	-2.13	0.034	-1.092596	-.0428573
State3	-.6549515	.2190203	-2.99	0.003	-1.086041	-.2238616
State4	-1.509469	.3043508	-4.96	0.000	-2.108512	-.9104259
State5	-1.48428	.2873532	-5.17	0.000	-2.049867	-.9186933
State6	-1.862257	.2805333	-6.64	0.000	-2.414421	-1.310094
State7	-1.307602	.2939478	-4.45	0.000	-1.886169	-.729035
State8	-.2681302	.1393267	-1.92	0.055	-.5423619	.0061016
State9	.5246029	.1839474	2.85	0.005	.1625457	.88666
State10	-.6690224	.2579674	-2.59	0.010	-1.17677	-.1612745
State11	-1.961622	.291496	-6.73	0.000	-2.535363	-1.387881
State12	-1.461542	.2725398	-5.36	0.000	-1.997972	-.9251112
State13	-1.543932	.2534422	-6.09	0.000	-2.042773	-1.045091
State14	-1.223216	.2454374	-4.98	0.000	-1.706302	-.7401302
State15	-1.217517	.2871651	-4.24	0.000	-1.782734	-.6523001
	⋮	⋮	⋮			
State46	-.8967539	.246611	-3.64	0.000	-1.38215	-.4113583
State47	-1.759266	.2939478	-5.98	0.000	-2.337833	-1.180699
State48	-.2285036	.3128959	-0.73	0.466	-.8443654	.3873581
_cons	3.47763	.3133568	11.10	0.000	2.860861	4.094399

Within estimation

- With 48 states it is possible to include a dummy for each state, but suppose you have panel data for 10000 individuals.......
- We can also perform fixed effects analysis by using within estimation:

Step 1: obtain the time averages: $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, $\bar{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{it}$, $\bar{c}_i = \frac{1}{T}\sum_{t=1}^{T} c_i = c_i$ and $\bar{u}_i = \frac{1}{T}\sum_{t=1}^{T} u_{it}$

Step 2: subtract the time averaged equation from the original equation:

$$y_{it} - \bar{y}_i = (X_{it} - \bar{X}_i)\beta + c_i - c_i + u_{it} - \bar{u}$$

Step 3: estimate the following equation by OLS

$$\ddot{y}_{it} = \ddot{X}_{it}\beta + \ddot{u}_{it}, \quad \ddot{y}_{it} = y_{it} - \bar{y}_i, \quad \ddot{X}_{it} = X_{it} - \bar{X}_i, \quad \ddot{u}_{it} = u_{it} - \bar{u}_i$$

- c_i drops out, but also time invariant regressors drop out!
- $\widehat{\beta}$ consistent if $E[\ddot{X}_{it}\ddot{u}_{it}] = E[(X_{it} - \bar{X}_i)(u_{it} - \bar{u}_i)] = 0 \quad t = 1, \ldots, T$ (strict exogeneity!)

Within estimation

$$\hat{\beta}_{within} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{X}_{it}' \ddot{X}_{it} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{X}_{it}' \ddot{y}_{it} \right)$$

$$\widehat{Avar\left(\hat{\beta}_{within}\right)} = \hat{\sigma}_u^2 \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{X}_{it}' \ddot{X}_{it} \right)^{-1}$$

- Under the fixed effects assumptions the following is a consistent estimate

$$\hat{\sigma}_u^2 = \frac{1}{NT - N - K} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\ddot{u}}_{it}^2 \right)$$

- but OLS regression of transformed model gives

$$\frac{1}{NT - K} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\ddot{u}}_{it}^2 \right)$$

- This implies that the standard errors are incorrect and should be multiplied by a factor $\sqrt{(NT - K)/(NT - N - K)}$
- Because of time-demeaning we loose N degrees of freedom

Within estimation

```
1 . bys state: egen Mfatality=mean(fatality)
2 . bys state: egen Mbeertax=mean(beertax)
3 . gen DMfatality=fatality-Mfatality
4 . gen DMbeertax=beertax-Mbeertax
5 . regress DMfatality DMbeertax, noheader noconstant
```

DMfatality	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
DMbeertax	**-.6558736**	**.173872**	**-3.77**	**0.000**	**-.9978922**	**-.3138551**

xtreg, fe command gives standard errors with the correct degrees of freedom adjustment

Within estimation

```
1 . xtreg fatality beertax, fe i(state)

  Fixed-effects (within) regression               Number of obs      =       336
  Group variable:  state                             Number of groups   =        48

  R-sq:  within  =  0.0407                           Obs per group: min =         7
         between =  0.1101                                          avg =       7.0
         overall =  0.0934                                          max =         7

                                                     F(1,287)           =     12.19
  corr(u_i, Xb)  =  -0.6885                          Prob > F           =    0.0006

------------------------------------------------------------------------------
fatalityrate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -.6558736     .18785    -3.49   0.001    -1.025612   -.2861352
       _cons |   2.377075   .0969699    24.51   0.000     2.186212    2.567937
-------------+----------------------------------------------------------------
     sigma_u |   .7147146
     sigma_e |  .18985942
         rho |  .93408484   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(  47,  287) =    52.18              Prob > F =  0.0000
```

Inference on the within estimator

Within estimation is efficient if

$$E[U_i U_i' | X_i, c_i] = \sigma_u^2 I_T$$

which can be interpreted as consisting of two parts

$$E[U_i U_i' | X_i, c_i] \stackrel{(1)}{=} E[U_i U_i'] \stackrel{(2)}{=} \sigma_u^2 I_T$$

(1) assumes homoskedasticity, and (2) rules out serial correlation

this mirrors the assumptions we made with Random effects analysis

Inference on the within estimator

- The Within estimator is consistent under the first fixed effect assumption and the rank condition: $rank\ E\left[\ddot{X}_i'\ddot{X}_i\right] = K$.

- If X_{it} contains time invariant explanatory variables, $\ddot{X}_i$ contains a column of zeros for all i and the rank condition is violated

- If the second fixed effects assumption is violated standard errors are incorrect and inference based on these standard errors is invalid

- If FE assumption 2 is violated we can:
  - use general FGLS
  - compute a robust variance matrix

Fixed effects FGLS estimator

- If $E[U_i U_i' | X_i, c_i] = E[U_i U_i']$ but $E[U_i U_i'] \neq \sigma_u^2 I_T$
- it is possible to use a fixed effects FGLS approach:

Step 1: Estimate β by within estimation and obtain the within residuals $\hat{\ddot{u}}_{it} = \ddot{y}_{it} - \ddot{X}_{it}\hat{\beta}_{within}$

Step 2: For each i drop the last time period (otherwise variance matrix cannot be inverted) and obtain

$$\hat{\Omega} = \frac{1}{N} \sum_{i=1}^{N} \hat{\ddot{U}}_i \hat{\ddot{U}}_i'$$

Step 3: Obtain the fixed effects FGLS estimator:

$$\hat{\beta}_{FEGLS} = \left(\sum_{i=1}^{N} \ddot{X}_i' \hat{\Omega}^{-1} \ddot{X}_i \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{X}_i' \hat{\Omega}^{-1} \ddot{Y}_i \right)$$

Robust variance covariance matrix

- Under FE assumptions but with $E[U_i U_i'] \neq \sigma_u^2 I_T$ FEGLS more efficient
- But if N is not several times larger than T, the FEGLS estimator can have poor finite sample properties
- FEGLS estimator not used much in practice
- More common to compute variance covariance matrix that is robust to any type of heteroskedasticity or serial correlation:

$$\hat{V} = \left(\sum_i \ddot{X}_i' \ddot{X}_i \right)^{-1} \left(\sum_i \ddot{X}_i' \hat{\ddot{U}}_i \hat{\ddot{U}}_i' \ddot{X}_i \right) \left(\sum_i \ddot{X}_i' \ddot{X}_i \right)^{-1}$$

Robust variance covariance matrix

```
1 . xtreg fatality beertax, fe i(state) cluster(state)

  Fixed-effects (within) regression               Number of obs      =       336
  Group variable:  state                             Number of groups   =        48

  R-sq:  within  =  0.0407                           Obs per group: min =         7
         between =  0.1101                                          avg =       7.0
         overall =  0.0934                                          max =         7

                                                     F(1,47)            =      5.05
  corr(u_i, Xb)  =  -0.6885                          Prob > F           =    0.0294

                                    (Std. Err. adjusted for    48 clusters in state)
  ------------------------------------------------------------------------------------
               |              Robust
  fatalityrate |     Coef.   Std. Err.       t    P>|t|      [95% Conf. Interval]
  -------------+----------------------------------------------------------------------
       beertax | -.6558736   .2918556     -2.25   0.029     -1.243011   -.0687358
         _cons |  2.377075   .1497966     15.87   0.000      2.075723    2.678427
  -------------+----------------------------------------------------------------------
       sigma_u |  .7147146
       sigma_e | .18985942
           rho | .93408484   (fraction of variance due to u_i)
  ------------------------------------------------------------------------------------
```

First differencing

- Next to LSDV and Within estimation we can also estimate the fixed effects model using first differencing.

$$y_{it} - y_{it-1} = (X_{it} - X_{it-1})\beta + c_i - c_i + u_{it} - u_{it-1}$$

$$\triangle y_{it} = \triangle X_{it}\beta + \triangle u_{it}$$

- In differencing we lose the first observation for each i and we cannot identify the constant term.
- The first differences estimator, $\hat{\beta}_{FD}$, is the pooled OLS estimator from the regression of Δy_{it} on ΔX_{it}, $i = 1, \ldots, N,\ t = 2, \ldots, T$.

First differencing

First Differences assumption 1:

$$E\left[\triangle X_{it} \triangle u_{it}\right] = E\left[\left(X_{it} - X_{it-1}\right)\left(u_{it} - u_{it-1}\right)\right] = 0$$

- This assumption is implied by, but weaker than the strict exogeneity assumption
- Assumption fails if shock to outcome last period u_{it-1} affects value of regressor(s) X_{it} today (feedback)

First differencing

First Differences assumption 2:

$$u_{it} = u_{it-1} + e_{it} \quad \textit{with} \quad E\left[e_i e_i^{'} | X_{i1}, \ldots, X_{iT}, c_i\right] = \sigma_e^2 I_{T-1}$$

- where e_i is the $(T{-}1) \times 1$ vector containing e_{it}, $t = 2, \ldots, T$ and $I_{T\text{-}1}$ is the identity matrix of dimension $T{-}1$.
- First differences is consistent under the first assumption and the rank condition $Rank\left(\sum_{t=2}^{T} E\left(\triangle X_{it} \triangle X_{it}^{'}\right)\right) = K$
- First differences is efficient if in addition the second assumption hold (u_{it} follows a random walk)

First differences: Robust variance covariance matrix

- Under the FD assumptions the asymptotic variance can be estimated by:

$$\widehat{Avar(\hat{\beta}_{FD})} = \hat{\sigma}_e^2(\Delta X' \Delta X)^{-1}$$

- If u_{it} are i.i.d. with $E[u_{it}^2|X_i] = \sigma_u^2$ then

$$E[\Delta u_{it} \Delta u_{is}] = \begin{cases} 2\sigma_u^2 & |s-t| = 0 \\ -\sigma_u^2 & |s-t| = 1 \\ 0 & |s-t| > 1 \end{cases}$$

- If u_{it} is a random walk then $\Delta u_{it} = e_{it}$ is i.i.d and FD is efficient
- if u_{it} is not a random walk or if there is heteroskedasticity use the following robust variance matrix

$$\widehat{Var(\hat{\beta}_{FD})} = (\sum_i \Delta X_i' \Delta X_i)^{-1} (\sum_i \Delta X_i' \widehat{\triangle U_i} \widehat{\triangle U_i}' \Delta X_i)(\sum_i \Delta X_i' \Delta X_i)^{-1}$$

First differencing

```
1 . xtset state year
        panel variable:   state (strongly balanced)
         time variable:   year, 1982 to 1988
                 delta:   1 unit

2 . gen Lfatality=L.fatality
  (48 missing values generated)

3 . gen Dfatality=D.fatality
  (48 missing values generated)

4 . gen Lbeertax=L.beertax
  (48 missing values generated)

5 . gen Dbeertax=D.beertax
  (48 missing values generated)

6 . list state year fatalityrate Lfatality Dfatality beertax Lbeertax Dbeertax, sep(7)
```

	state	year	fatali~e	Lfatality	Dfatality	beertax	Lbeertax	Dbeertax
1.	AL	1982	2.12836	.	.	1.539379	.	.
2.	AL	1983	2.34848	2.12836	.22011995	1.788991	1.5393795	.24961126
3.	AL	1984	2.33643	2.34848	-.01204991	1.714286	1.7889907	-.07470512
4.	AL	1985	2.19348	2.3364301	-.14295006	1.652542	1.7142856	-.06174326
5.	AL	1986	2.66914	2.19348	.47565985	1.609907	1.6525424	-.04263532
6.	AL	1987	2.71859	2.6691399	.04945016	1.56	1.609907	-.04990709
7.	AL	1988	2.49391	2.71859	-.22467995	1.501444	1.5599999	-.05855632
8.	AZ	1982	2.49914	.	.	.2147971	.	.
9.	AZ	1983	2.26738	2.49914	-.23176003	.206422	.21479714	-.00837511
10.	AZ	1984	2.82878	2.26738	.56139994	.2967033	.20642203	.09028128
11.	AZ	1985	2.80201	2.8287799	-.02676988	.3813559	.29670331	.08465263
12.	AZ	1986	3.07106	2.8020101	.26904988	.371517	.38135594	-.00983891
13.	AZ	1987	2.76728	3.0710599	-.30377984	.36	.37151703	-.01151702
14.	AZ	1988	2.70565	2.7672801	-.06163001	.346487	.36000001	-.013513

First differencing

Don’t do this in Stata:

```
1 . gen LfatalityB=fatality[_n-1]
  (1 missing value generated)

2 . gen DfatalityB=fatality-LfatalityB
  (1 missing value generated)

3 . gen LbeertaxB=beertax[_n-1]
  (1 missing value generated)

4 . gen DbeertaxB=beertax-LbeertaxB
  (1 missing value generated)

5 . list state year fatalityrate LfatalityB DfatalityB beertax LbeertaxB DbeertaxB, sep(7)
```

	state	year	fatali~e	Lfatali~B	DfatalityB	beertax	LbeertaxB	DbeertaxB
1.	AL	1982	2.12836	.	.	1.539379	.	.
2.	AL	1983	2.34848	2.12836	.22011995	1.788991	1.5393795	.24961126
3.	AL	1984	2.33643	2.34848	-.01204991	1.714286	1.7889907	-.07470512
4.	AL	1985	2.19348	2.3364301	-.14295006	1.652542	1.7142856	-.06174326
5.	AL	1986	2.66914	2.19348	.47565985	1.609907	1.6525424	-.04263532
6.	AL	1987	2.71859	2.6691399	.04945016	1.56	1.609907	-.04990709
7.	AL	1988	2.49391	2.71859	-.22467995	1.501444	1.5599999	-.05855632
8.	AZ	1982	2.49914	2.4939101	.00522995	.2147971	1.5014436	-1.2866465
9.	AZ	1983	2.26738	2.49914	-.23176003	.206422	.21479714	-.00837511
10.	AZ	1984	2.82878	2.26738	.56139994	.2967033	.20642203	.09028128
11.	AZ	1985	2.80201	2.8287799	-.02676988	.3813559	.29670331	.08465263
12.	AZ	1986	3.07106	2.8020101	.26904988	.371517	.38135594	-.00983891
13.	AZ	1987	2.76728	3.0710599	-.30377984	.36	.37151703	-.01151702
14.	AZ	1988	2.70565	2.7672801	-.06163001	.346487	.36000001	-.013513

First differencing, the beer tax and the traffic fatality rate

```
1 . regress D.fatality D.beertax, noconstant

      Source |       SS       df       MS              Number of obs =     288
-------------+------------------------------           F(  1,   287) =    0.01
       Model |  .000417008     1  .000417008           Prob > F      =  0.9178
    Residual |  11.2154889   287  .039078358           R-squared     =  0.0000
-------------+------------------------------           Adj R-squared = -0.0034
       Total |  11.2159059   288  .038944118           Root MSE      =  .19768

------------------------------------------------------------------------------
D.           |
fatalityrate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |
         D1. |   .0288161   .2789533     0.10   0.918    -.5202376    .5778698
------------------------------------------------------------------------------
```

- Note: no constant term included
- Note: only 288 observations are used.

FD, the beer tax and the traffic fatality rate, clustered se's

- In most cases there is no reason to believe that u_{it} follows a random walk
- Therefore better to use clustered se's

```
1 . regress D.fatality D.beertax, noconstant cluster(state)

Linear regression                                 Number of obs =         288
                                                     F(   1,     47) =      0.01
                                                     Prob > F        =    0.9130
                                                     R-squared       =    0.0000
                                                     Root MSE        =    .19768

                              (Std. Err. adjusted for     48 clusters in state)
```

D. fatalityrate	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
beertax D1.	.0288161	.2623217	0.11	0.913	-.4989071	.5565394

Comparing estimators

The effect of the beer tax on the traffic fatality rate

	Pooled OLS	RE	LSDV	WE	FD
Beer tax	**0.365**	**-0.052**	**-0.656**	**-0.656**	**0.029**
	(0.062)	(0.124)	(0.188)	(0.188)	(0.279)
	[0.120]	[0.110]	[0.292]	[0.292]	[0.262]

- standard se's in parentheses
- cluster-robust se's in brackets
- NOT: WE and FD different!

Fixed effects versus random effects

Fixed effects

- Based on weaker assumptions, allows for correlation between X_{it} and c_i
- Only possible to estimate effect of time-varying regressors
- Based only on within-variation and therefore estimators can be imprecise
- prediction/ forecasting complicated

Random effects

- Based on the assumption of no correlation between X_{it} and c_i
    - often too strong!
- possible to estimate effects of time-invariant regressors
- based on within and between variation
- prediction/ forecasting not complicated

Variance decomposition

$$\sum_i \sum_t (x_{it} - \bar{x})^2 = \sum_i \sum_t (x_{it} - \bar{x}_i)^2 + \sum_i \sum_t (\bar{x}_i - \bar{x})^2$$

or

$$(NT - 1)\sigma^2_{\text{total}} = N(T - 1)\sigma^2_{\text{within}} + (N - 1)T\sigma^2_{\text{between}}$$

where

$$\sigma^2_{\text{within}} = \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)^2$$

$$\sigma^2_{\text{between}} = \frac{1}{N-1} \sum_{i=1}^{N} (\bar{x}_i - \bar{x})^2$$

Variance decomposition: beer tax and the traffic fatality rate

```
1 . xtsum fatalityrate beertax
```

Variable		Mean	Std. Dev.	Min	Max	Observations
fatali~e	overall	**2.040444**	**.5701938**	**.82121**	**4.21784**	N = 336
	between		**.5461407**	**1.110077**	**3.653197**	n = 48
	within		**.1794253**	**1.45556**	**2.962664**	T = 7
beertax	overall	**.513256**	**.4778442**	**.0433109**	**2.720764**	N = 336
	between		**.4789513**	**.0481679**	**2.440507**	n = 48
	within		**.0552203**	**.1415352**	**.7935126**	T = 7

Hausman test

Does the second random effects assumption hold?

- **Orthogonality:** $E\left[c_i | X_{i1}, \dots, X_{iT}\right] = E\left[c_i\right] = 0$

Hausman (Ectra, 1978) proposed the following test

- $H_0 : E\left[c_i | X_{i1}, \dots, X_{iT}\right] = 0$, both RE and FE estimators are consistent (but the RE estimator is more efficient)
- $H_1 : E\left[c_i | X_{i1}, \dots, X_{iT}\right] \neq 0$, only the FE estimator is consistent.

This implies that under the null $plim \hat{\beta}_{FE} = plim \hat{\beta}_{RE}$

Hausman test

We can test $H_0: \ plim\hat{\beta}_{FE} = plim\hat{\beta}_{RE}$ using the following statistic

$$H = (\hat{\beta}_{RE} - \hat{\beta}_{FE}) \left(\widehat{Avar(\hat{\beta}_{FE})} - \widehat{Avar(\hat{\beta}_{RE})} \right)^{-1} (\hat{\beta}_{RE} - \hat{\beta}_{FE}) \sim \chi^2_K$$

Strict exogeneity is maintained under the null and the alternative

Test statistic only correct if $E\left[U_i U_i' | X_i, c_i\right] = \sigma_u^2 I_T$ *and* $E[c_i^2 | X_i] = \sigma_c^2$
(Random Effects Assumption 3)

- Note: best to use same estimator of σ_u^2 for $\widehat{Avar(\hat{\beta}_{RE})}$ and $\widehat{Avar(\hat{\beta}_{FE})}$.

Hausman test

```
1 . hausman FE RE

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       FE           RE         Difference          S.E.
-------------+----------------------------------------------------------------
     beertax |   -.6558736    -.0520158       -.6038579        .1409539
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(  1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                            =       18.35
                Prob>chi2 =      0.0000
```

Mundlak

Mundlak (Ectra, 1978) proposed a model on pooling time-series and cross-section data.

$$y_{it} = X_{it}'\beta + \bar{X}_i'\gamma + \omega_i + u_{it}$$

where ω_i is a RE uncorrelated with x_{it}. This model should be estimated using FGLS.

Mundlak showed that the random effect estimator for β in this specification is identical to the within estimator.

The individual specific effect equals

$$c_i = \bar{X}_i\gamma + \omega_i$$

and $\mathcal{H}_0 : \gamma = 0$ (using a Wald test) is therefore a test for whether the c_i are correlated with X_{it}.

Note: Asymptotically the Hausman and Mundlak test are identical

Mundlak: beer tax and traffic fatality

```
1 . bys state : egen Mbeertax=mean(beertax)

2 . xtreg fatalityrate beertax Mbeertax, re

Random-effects GLS regression                 Number of obs      =       336
Group variable: state                            Number of groups   =        48

R-sq:  within  =  0.0407                         Obs per group: min =         7
       between =  0.1101                                        avg =       7.0
       overall =  0.1033                                        max =         7

                                                 Wald chi2( 2)      =     17.88
corr(u_i, X)   =  0 (assumed)                  Prob > chi2        =    0.0001

------------------------------------------------------------------------------
fatalityrate |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     beertax |  -.6558736     .18785    -3.49   0.000    -1.024053   -.2876944
    Mbeertax |   1.034291   .2458472     4.21   0.000     .5524397    1.516143
       _cons |   1.846219   .1107969    16.66   0.000     1.629061    2.063377
-------------+----------------------------------------------------------------
     sigma_u |   .5157915
     sigma_e |  .18985942
         rho |  .88067496   (fraction of variance due to u_i)
------------------------------------------------------------------------------

3 . test Mbeertax

 ( 1)  Mbeertax = 0

           chi2(  1) =     17.70
         Prob > chi2 =      0.0000
```

Robust Mundlak: beer tax and traffic fatality

Easy to relax the third random effects assumption and to obtain a robust test statistic:

```
1 . xtreg fatalityrate beertax Mbeertax, re cluster(state)

  Random-effects GLS regression                   Number of obs      =       336
  Group variable:  state                            Number of groups   =        48

  R-sq:  within   =  0.0407                         Obs per group: min =         7
         between =  0.1101                                         avg =       7.0
         overall =  0.1033                                         max =         7

                                                    Wald chi2( 2)       =     13.34
  corr(u_i, X)    =  0 (assumed)                  Prob > chi2         =    0.0013

                                   (Std. Err. adjusted for    48 clusters in state)
  ------------------------------------------------------------------------------------
               |               Robust
  fatalityrate |      Coef.   Std. Err.       z     P>|z|      [95% Conf. Interval]
  -------------+----------------------------------------------------------------------
       beertax |  -.6558736   .2922935     -2.24    0.025     -1.228758   -.0829889
      Mbeertax |   1.034291   .3279115      3.15    0.002      .3915967    1.676986
         _cons |   1.846219   .1193465     15.47    0.000      1.612304    2.080133
  -------------+----------------------------------------------------------------------
       sigma_u |   .5157915
       sigma_e |  .18985942
           rho |  .88067496   (fraction of variance due to u_i)
  ------------------------------------------------------------------------------------

2 . test Mbeertax

   ( 1)  Mbeertax = 0

            chi2(  1) =       9.95
          Prob > chi2 =     0.0016
```

Within Estimator (WE) or First differences (FD)?

When should one use the within estimator (WE) and when the first difference estimator (FD)?

- $T = 2$, doesn't matter they are identical
- $T > 2$, FD$\neq$WE
  - u_{it} is i.i.d.: FD less efficient than WE
  - u_{it} follows a random walk: FD is more efficient than WE
- FD is more sensitive to violations of strict exogeneity
- If WE and FD differ then this suggests that strict exogeneity does not hold

FD and WE T=2: beer tax and traffic fatality

```
1 . keep if year==1982 | year==1983
  (240 observations deleted)

2 . xtreg fatalityrate beertax, fe

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

  R-sq:  within   =  0.0001                         Obs per group: min =         2
         between =  0.0339                                         avg =       2.0
         overall =  0.0324                                         max =         2

                                                    F(1,47)             =      0.01
  corr(u_i, Xb)   =  -0.2193                        Prob > F            =    0.9363

  ------------------------------------------------------------------------------
  fatalityrate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
  -------------+----------------------------------------------------------------
       beertax |  -.0452044   .5623282    -0.08   0.936    -1.176463    1.086054
         _cons |   2.072495   .2993334     6.92   0.000     1.470314    2.674676
  -------------+----------------------------------------------------------------
       sigma_u |   .6310782
       sigma_e |  .17766724
           rho |  .92656169   (fraction of variance due to u_i)
  ------------------------------------------------------------------------------
  F test that all u_i=0:     F(  47,  47) =     24.02            Prob > F = 0.0000

3 . regress D.fatality D.beertax, nocons noheader

  ------------------------------------------------------------------------------
  D.           |
  fatalityrate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
  -------------+----------------------------------------------------------------
       beertax |
           D1. |  -.0452044   .5623282    -0.08   0.936    -1.176463    1.086054
  ------------------------------------------------------------------------------
```

FD and WE T=7: beer tax and traffic fatality

```
1 . xtreg fatalityrate beertax, fe cluster(state)

  Fixed-effects (within) regression               Number of obs      =       336
  Group variable:  state                             Number of groups   =        48

  R-sq:  within  =  0.0407                           Obs per group: min =         7
         between =  0.1101                                          avg =       7.0
         overall =  0.0934                                          max =         7

                                                     F(1,47)            =      5.05
  corr(u_i, Xb)  =  -0.6885                          Prob > F           =    0.0294

                                  (Std. Err. adjusted for    48 clusters in state)
```

fatalityrate	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
beertax	-.6558736	.2918556	-2.25	0.029	-1.243011	-.0687358
_cons	2.377075	.1497966	15.87	0.000	2.075723	2.678427
sigma_u	.7147146					
sigma_e	.18985942					
rho	.93408484	(fraction of variance due to u_i)				

```
2 . regress D.fatality D.beertax, noheader nocons cluster(state)
                                  (Std. Err. adjusted for    48 clusters in state)
```

D. fatalityrate	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
beertax D1.	.0288161	.2623217	0.11	0.913	-.4989071	.5565394

Testing strict exogeneity

- Fixed effects and Within estimators differ, this can indicate violation of strict exogeneity
- We can use Hausman test to test $H_0 : \hat{\beta}_{FD} - \hat{\beta}_{WE} = 0$ but complicated to obtain $\hat{Var}(\hat{\beta}_{FD} - \hat{\beta}_{WE})$
- Wooldridge proposes number of regression-based test which can be made robust for heteroskedasticity and serial correlation:

Testing strict exogeneity

Within estimation: include W_{it+1} which is (subset of) X_{it+1} in fixed effects equation

$$y_{it} = \beta X_{it} + \gamma W_{it+1} + c_i + \epsilon_{it}$$

estimate by within estimation and test $H_0 : \gamma = 0$

First differences: include W_{it} which is (subset of) X_{it} in first differenced equation

$$\triangle y_{it} = \triangle X_{it}\beta + \delta W_{it} + \triangle\epsilon_{it}$$

and test $H_0 : \delta = 0$

NOTE: if H_0 is not rejected we cannot conclude that strict exogeneity holds!

Testing strict exogeneity: within estimation

```
1 . bys state: gen beertax_next=beertax[_n+1]
  (48 missing values generated)

2 . xtreg fatality beertax beertax_next, fe cluster(state)

  Fixed-effects (within) regression               Number of obs      =       288
  Group variable:  state                              Number of groups   =        48

  R-sq:  within  =  0.0533                            Obs per group: min =         6
         between =  0.1058                                           avg =       6.0
         overall =  0.0912                                           max =         6

                                                      F(2,47)            =      2.98
  corr(u_i, Xb)  =  -0.7666                           Prob > F           =    0.0603

                                   (Std. Err. adjusted for    48 clusters in state)
  ------------------------------------------------------------------------------
               |               Robust
  fatalityrate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
  -------------+----------------------------------------------------------------
       beertax |  -.1449376    .363917    -0.40   0.692    -.8770442     .587169
  beertax_next |    -.80744   .4234033    -1.91   0.063    -1.659218    .0443376
         _cons |   2.522917   .2059797    12.25   0.000     2.108539    2.937294
  -------------+----------------------------------------------------------------
       sigma_u |  .82326153
       sigma_e |   .1893186
           rho |  .94977372   (fraction of variance due to u_i)
  ------------------------------------------------------------------------------

3 . test beertax_next

   ( 1)  beertax_next = 0

         F(  1,    47) =     3.64
              Prob > F =     0.0626
```

Testing strict exogeneity: first differences

```
1 . regress D.fatality D.beertax beertax, cluster(state) nocons

  Linear regression                                     Number of obs =        288
                                                           F(  2,     47) =       0.38
                                                           Prob > F       =     0.6889
                                                           R-squared      =     0.0026
                                                           Root  MSE      =     .19777

                                            (Std. Err. adjusted for     48 clusters in state)
```

D. fatalityrate	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
beertax						
D1.	.1162343	.2767848	0.42	0.676	-.4405849	.6730535
--.	.0153319	.0176901	0.87	0.391	-.020256	.0509198

```
2 . test beertax

   ( 1)   beertax = 0

         F(  1,     47) =       0.75
              Prob > F =       0.3905
```

Beer tax and the traffic fatality rate: time effects

- Eventhough tests don't reject H_0 strict exogeneity might be violated
- For example because of federal policy measures that affect the fatality rate and that are correlated with the beer tax
- Possible solution: include time fixed effects λ_t

$$y_{it} = X_{it}\beta + c_i + \lambda_t + \varepsilon_{it}$$

- λ_t capture all (unobserved) variables that vary over time but that do not vary between states.

Beer tax and the traffic fatality rate: time effects

```
Fixed-effects (within) regression               Number of obs      =       336
Group variable: state                              Number of groups   =        48

R-sq:  within  =  0.0803                           Obs per group: min =         7
       between =  0.1101                                          avg =       7.0
       overall =  0.0876                                          max =         7

                                                   F(7,47)            =      4.36
corr(u_i, Xb)  =  -0.6781                          Prob > F           =    0.0009

                                 (Std. Err. adjusted for    48 clusters in state)
```

fatalityrate	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
beertax	-.6399799	.3570783	-1.79	0.080	-1.358329	.0783691
Year2	-.0799029	.0350861	-2.28	0.027	-.1504869	-.0093188
Year3	-.0724206	.0438809	-1.65	0.106	-.1606975	.0158564
Year4	-.1239763	.0460559	-2.69	0.010	-.2166288	-.0313238
Year5	-.0378645	.0570604	-0.66	0.510	-.1526552	.0769262
Year6	-.0509021	.0636084	-0.80	0.428	-.1788656	.0770615
Year7	-.0518038	.0644023	-0.80	0.425	-.1813645	.0777568
_cons	2.42847	.2016885	12.04	0.000	2.022725	2.834215
sigma_u	.70945965					
sigma_e	.18788295					
rho	.93446372	(fraction of variance due to u_i)				

Beer tax and the traffic fatality rate: control variables

Dependent variable: traffic fatality rate				
	(1)	(2)	(3)	(4)
Beer tax	**-0.656****	**-0.640***	**-0.680***	**-0.548***
	(0.292)	(0.357)	(0.346)	(0.320)
min. legal drinking age			0.020	0.001
			(0.032)	(0.022)
mandatory jail time			-0.016	0.024
			(0.018)	(0.016)
mandatory community service			0.134	0.025
			(0.142)	(0.135)
unemployment rate				-0.077***
				(0.013)
per capita income				0.000*
				(0.000)
Year fixed effects	no	yes	yes	yes

Note: Standard errors (between parentheses) are clustered at the state level; ***significant at 1%, **significant at 5%, *significant at 10%.