Econ 312

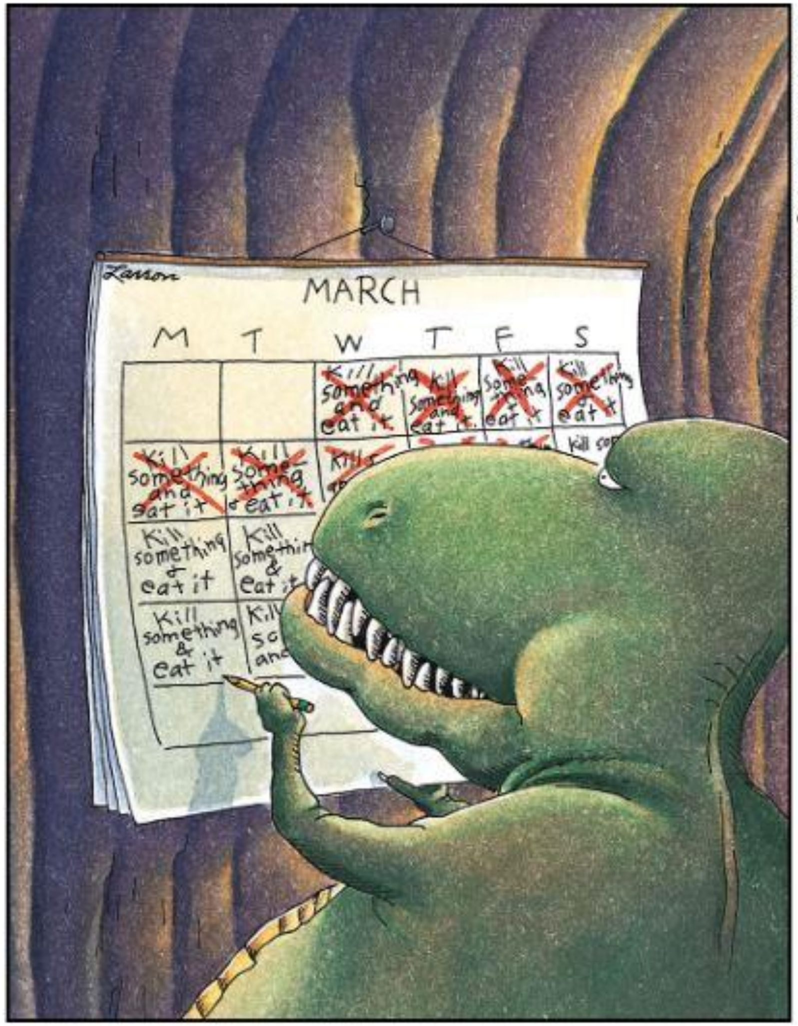
Friday, April 10

Random-Effects Model

Reading: Wooldridge, Section 14.2

Class notes: 129 - 132

Todav’s Far Side offering



Jurassic calendars

My calendar has a lot less on it these days, but it’s not *quite* this boring!

Context and overview

- We presented the **fixed-effects model** for panel-data estimation in the last class (April 8)
- We now consider the **random-effects model** as an alternative
- It has some **advantages** over fixed effects:
    - More degrees of freedom
    - Allows regressors that do not vary across time
- It also a big **disadvantage**: It is often inconsistent
    - Hausman test can test for this
- We also do a detailed econometric **example** using panel-data estimators

Setup of random-effects (RE) model

- As before, $y_{it} = \beta_{0i} + \beta_1 x_{1it} + \beta_2 x_{2it} + u_{it}$
  - Intercept term differs across cross-sectional units
- In fixed-effects model, we think of β_{0i} as n constants to be estimated
- In random-effects model, differences in intercept across i are random:

$$\beta_{0i} = \overline{\beta}_0 + a_i$$

$$E(a_i) = 0$$

$$\operatorname{cov}(a_i, a_j) = 0,\ i \neq j$$

$$\operatorname{var}(a_i) = \sigma_a^2$$

Regression in RE model

$$
\begin{aligned}
y_{it} &= \left(\bar{\beta}_0 + a_i\right) + \beta_1 x_{1it} + \beta_2 x_{2it} + u_{it} \\
&= \bar{\beta}_0 + \beta_1 x_{1it} + \beta_2 x_{2it} + \left(a_i + u_{it}\right) \\
&= \bar{\beta}_0 + \beta_1 x_{1it} + \beta_2 x_{2it} + v_{it}
\end{aligned}
$$

- **Error-components model**: v has pattern of autocorrelation between observations on same i

$$
\operatorname{corr}\left(v_{it} v_{is}\right) = \frac{\operatorname{cov}\left(v_{it}, v_{is}\right)}{\sqrt{\operatorname{var}\left(v_{it}\right)\operatorname{var}\left(v_{is}\right)}} = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_u^2} = \rho
$$

- Estimate ρ with correlation of residuals within units, then do **feasible GLS**

Assumptions of RE model

$$E(v_{it}) = 0 \qquad \operatorname{cov}(v_{it}, v_{js}) = 0,\ i \neq j$$

$$\operatorname{var}(v_{it}) = \sigma_u^2 + \sigma_a^2 \qquad \operatorname{cov}(u_{ij}, x_{kij}) = 0,\ k = 1, 2, \ldots, K$$

$$\operatorname{cov}(v_{it}, v_{is}) = \sigma_a^2,\ t \neq s \qquad \operatorname{cov}(a_i, x_{kit}) = 0,\ k = 1, 2, \ldots, K$$

- **Last assumption** is most likely to be problematic
  - Is random effect really uncorrelated with regressors?
  - Random effect is unobserved variables that have values specific to i
  - We would often expect to be correlated with x values specific to i
- RE estimators are **biased and inconsistent** if last assumption fails

Testing for RE vs. OLS: Is var(a_i) = 0?

- If a doesn't vary across i, then no correlations in v
- Positive a_i is positive expected value of v_{it} for that i
- Estimate with $\hat{a}_i = \frac{1}{T}\sum_{t=1}^{T}\hat{v}_{it}$
- Then estimate $\hat{\sigma}_a^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{a}_i^2$
- Test $H_0 : \sigma_a^2 = 0$ (one-tailed) with $LM = \sqrt{\frac{nT}{2(T-1)}}\left[\frac{\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\hat{v}_{it}\right)^2}{\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{v}_{it}^2} - 1\right] \sim N(0,1)$
- Reject → OLS is inefficient

Feasible GLS estimator for RE model

- Use OLS residuals to estimate σ_a^2 and σ_u^2
- Calculate $\theta = 1 - \dfrac{\sigma_u}{\sqrt{T\sigma_a^2 + \sigma_u^2}}$

- **Quasi-de-mean model** to get v* error term that satisfies OLS assumptions

$$y_{it}^* \equiv y_{it} - \theta \bar{y}_i$$

$$x_{0it}^* \equiv 1 - \theta$$

$$x_{kit}^* \equiv x_{kit} - \theta \bar{x}_{ki},\ k = 1, 2, \ldots, K$$

$$v_{it}^* \equiv v_{ij} - \theta \bar{v}_i$$

- Random effects = fixed effects if θ = 1 (as $T \rightarrow \infty$)

Problem with random effects

- Unit-specific error term a_i represents **omitted variables** that are constant over time within each cross-sectional unit i
- Can we assume that these are **independent** of included regressors?
  - Probably not in many/most cases
  - Random-effects estimator is biased and inconsistent if not independent
- **Hausman test** compares results of RE and FE estimator
  - Null hypothesis is that RE is similar and valid
  - Reject if results of FE and RE differ significantly
- Common practice: Use RE unless Hausman test rejects validity

Sample application: Seat belts

Variable Definitions

Variable	Definition
fatalityrate	Number of fatalities per million of traffic miles
sb_usage	Seat belt usage rate
speed65	Binary variable for 65 mile per hour speed limit
speed70	Binary variable for 70 or higher mile per hour speed limit
ba08	Binary variable for blood alcohol limit $\leq .08\%$
drinkage21	Binary variable for age 21 drinking age
income	Per capita income
age	Mean age
primary	Binary variable for primary enforcement of seat belt laws
secondary	Binary variable for secondary enforcement of seat belt laws
vmt	Millions of traffic miles per year. (Note: Number of fatalities = $fatalityrate \times vmt$)
state	State
year	Year
fips	State ID Code

Summary statistics

```
. summ
```

Variable	Obs	Mean	Std. Dev.	Min	Max
state	0				
year	765	1990	4.32332	1983	1997
fips	765	28.96078	15.68709	1	56
vmt	765	41447.73	43961.99	3099	285612
fatalityrate	765	.0214895	.0061713	.0083273	.0454701
sb_useage	556	.5288518	.1701859	.06	.87
speed65	765	.6457516	.4785978	0	1
speed70	765	.0705882	.2563034	0	1
drinkage21	765	.8849673	.3192701	0	1
ba08	765	.1163399	.3208418	0	1
income	765	17992.59	4811.459	8372	35863
age	765	35.13719	1.698131	28.23497	39.16958
primary	765	.1215686	.3270007	0	1
secondary	765	.4954248	.5003062	0	1

Glance at data

	state	year	fips	vmt	fatalityrate	sb_useage	speed65	speed70	drinkage21	ba08	income	age
1	AL	1983	1	31032	.0299691	.	0	0	0	0	9505	34.20065
2	AL	1984	1	32961	.0282758	.13	0	0	0	0	10417	34.42163
3	AL	1985	1	35091	.0251346	.17	0	0	1	0	11133	34.60257
4	AL	1986	1	34003	.0317913	.29	0	0	1	0	11736	34.76067
5	AL	1987	1	37426	.0296852	.21	1	0	1	0	12394	34.95566
6	AL	1988	1	39684	.0258039	.29	1	0	1	0	13288	35.13696
7	AL	1989	1	40765	.0252422	.38	1	0	1	0	14266	35.32786
8	AL	1990	1	42347	.0264718	.44	1	0	1	0	15213	35.62292
9	AL	1991	1	42924	.0259994	.47	1	0	1	0	15895	35.68874
10	AL	1992	1	45762	.0225296	.58	1	0	1	0	16817	35.83262
11	AL	1993	1	47337	.0220546	.55	1	0	1	0	17398	35.94825
12	AL	1994	1	48956	.0221219	.55	1	0	1	0	18163	36.09201
13	AL	1995	1	50628	.0220036	.52	1	1	1	1	19041	36.20575
14	AL	1996	1	51433	.0222814	.54	1	1	1	1	19838	36.38513
15	AL	1997	1	53458	.0222979	.52	1	1	1	1	20672	36.50072
16	AK	1983	2	3358	.0446694	.	0	0	1	0	17973	28.23497
17	AK	1984	2	3589	.0373363	.	0	0	1	0	18093	28.34354
18	AK	1985	2	3840	.0330729	.	0	0	1	0	18925	28.37282
19	AK	1986	2	4008	.0251996	.	0	0	1	0	18466	28.39665
20	AK	1987	2	3900	.0194872	.	0	0	1	0	18021	28.45325
21	AK	1988	2	3841	.0252538	.	0	0	1	0	18447	28.85142
22	AK	1989	2	3887	.0216105	.	0	0	1	0	19970	29.14895
23	AK	1990	2	3979	.0246293	.45	0	0	1	0	21073	29.58628
24	AK	1991	2	4021	.0251181	.66	0	0	1	0	21496	29.82771
25	AK	1992	2	3841	.0281177	.66	1	0	1	0	22073	30.2107
26	AK	1993	2	3918	.0301174	.69	1	0	1	0	22711	30.46439
27	AK	1994	2	4150	.0204819	.69	1	0	1	0	23417	30.75657

Missing data

- Some states are missing data on seat-belt usage

- This is our key regressor

- Are these data missing at random?
  - Probably not, but we don’t yet have a solution
  - We’ll proceed with caution

Housekeeping

- Define data as panel: xtset fips year
  - Note two state identifiers, state and fips
  - State is alphabetical and cannot be used in xtset
  - FIPS codes are standardized, numerical codes for states and territories
- Generate lnincome variable
- Spell “usage” correctly: rename sb_useage sb_usage
- Basic regression:
  - regress fatalityrate sb_usage speed65 speed70 drinkage21 ba08 lnincome age

OLS results

```
. regress fatalityrate sb_usage speed65 speed70 drinkage21 ba08 lnincome age

      Source |       SS           df       MS      Number of obs   =       556
-------------+----------------------------------   F(7, 548)       =     95.41
       Model |  .007711649         7  .001101664   Prob > F        =    0.0000
    Residual |  .006327757       548  .000011547   R-squared       =    0.5493
-------------+----------------------------------   Adj R-squared   =    0.5435
       Total |  .014039406       555  .000025296   Root MSE        =     .0034

------------------------------------------------------------------------------
fatalityrate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    sb_usage |   .0040684   .0012158     3.35   0.001     .0016803    .0064565
     speed65 |   .0001479   .0004029     0.37   0.714    -.0006436    .0009394
     speed70 |   .0024045   .0005112     4.70   0.000     .0014003    .0034086
  drinkage21 |   .0000799   .0008756     0.09   0.927    -.0016401    .0017998
        ba08 |  -.0019246   .0004447    -4.33   0.000    -.0027982    -.001051
    lnincome |  -.0181444   .0009311   -19.49   0.000    -.0199733   -.0163155
         age |  -7.22e-06   .0001089    -0.07   0.947    -.0002212    .0002067
       _cons |   .1965469   .0082232    23.90   0.000     .1803941    .2126998
------------------------------------------------------------------------------
```

Anomaly

- Higher seat-belt usage → higher fatality rate
- Why is this happening?
  - Are there omitted variables?
- Could some of these omitted variables vary mostly across states?
  - If so, then fixed effects might proxy for them and eliminate bias
  - Use robust option in FE to get clustered standard errors

Fixed effects results

```
. xtreg fatalityrate sb_usage speed65 speed70 drinkage21 ba08 lnincome age , fe robust

Fixed-effects (within) regression               Number of obs      =       556
Group variable: fips                            Number of groups   =        51

R-sq:                                           Obs per group:
     within  = 0.6868                                         min =          8
     between = 0.1957                                         avg =       10.9
     overall = 0.3896                                         max =         15

                                                F(7,50)            =     96.72
corr(u_i, Xb)  = -0.1332                        Prob > F           =    0.0000

                                  (Std. Err. adjusted for 51 clusters in fips)
------------------------------------------------------------------------------
             |               Robust
fatalityrate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    sb_usage |  -.0057748   .0016693    -3.46   0.001    -.0091276    -.002422
     speed65 |   -.000425   .0004555    -0.93   0.355    -.0013399    .0004898
     speed70 |   .0012333   .0003483     3.54   0.001     .0005337    .0019329
  drinkage21 |   .0007453   .0007184     1.04   0.305    -.0006976    .0021883
        ba08 |  -.0013775   .0003751    -3.67   0.001    -.0021308   -.0006241
    lnincome |  -.0135144   .0023849    -5.67   0.000    -.0183047   -.0087241
         age |   .0009787   .0007461     1.31   0.196    -.0005199    .0024772
       _cons |   .1209958   .0184235     6.57   0.000     .0839912    .1580005
-------------+----------------------------------------------------------------
     sigma_u |  .00383103
     sigma_e |   .0017871
         rho |  .82128567   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Would time effects make sense?

- Might control for changes over time
  - Air bags
  - Other safety features
  - Changes in highway system
  - Changes in traffic congestion
- This would be differences-in-differences estimator that ignores variation that is
  - Purely across states
  - Purely over time

Differences-in-differences results (1)

```
. xtreg fatalityrate sb_usage speed65 speed70 drinkage21 ba08 lnincome age i.year , fe robus
> t

Fixed-effects (within) regression               Number of obs     =        556
Group variable: fips                            Number of groups  =         51

R-sq:                                           Obs per group:
     within  = 0.7506                                         min =          8
     between = 0.1139                                         avg =       10.9
     overall = 0.0338                                         max =         15

                                                F(21,50)          =      52.30
corr(u_i, Xb)  = -0.5086                        Prob > F          =     0.0000

                                    (Std. Err. adjusted for 51 clusters in fips)
------------------------------------------------------------------------------
             |               Robust
fatalityrate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    sb_usage |  -.0037186   .0014515    -2.56   0.013    -.0066339   -.0008032
     speed65 |  -.0007833   .0005801    -1.35   0.183    -.0019484    .0003818
     speed70 |   .0008042   .0004572     1.76   0.085    -.0001142    .0017225
  drinkage21 |  -.0011337   .0006221    -1.82   0.074    -.0023831    .0001158
        ba08 |  -.0008225   .0004433    -1.86   0.069    -.0017128    .0000678
    lnincome |   .0062643   .0066992     0.94   0.354    -.0071913      .01972
         age |    .001318   .0006937     1.90   0.063    -.0000753    .0027114
```

Differences-in-differences results (2)

year						
1984	-.0004319	.001378	-0.31	0.755	-.0031998	.002336
1985	-.0010707	.0017641	-0.61	0.547	-.004614	.0024726
1986	-.0005777	.0020078	-0.29	0.775	-.0046106	.0034551
1987	-.0008722	.0024939	-0.35	0.728	-.0058813	.0041368
1988	-.001885	.002877	-0.66	0.515	-.0076636	.0038936
1989	-.0041766	.0032564	-1.28	0.206	-.0107172	.0023641
1990	-.005266	.0035402	-1.49	0.143	-.0123767	.0018448
1991	-.0066622	.0037593	-1.77	0.082	-.0142131	.0008886
1992	-.008518	.0039855	-2.14	0.037	-.0165232	-.0005128
1993	-.0089399	.004199	-2.13	0.038	-.0173738	-.000506
1994	-.0096297	.0045934	-2.10	0.041	-.0188559	-.0004035
1995	-.0101123	.0048961	-2.07	0.044	-.0199464	-.0002782
1996	-.0110766	.0052089	-2.13	0.038	-.0215389	-.0006142
1997	-.0116075	.0055341	-2.10	0.041	-.0227231	-.0004919
_cons	-.0779904	.0663611	-1.18	0.245	-.2112805	.0552998
sigma_u	.00575371					
sigma_e	.00161752					
rho	.92675648	(fraction of variance due to u_i)				

Random effects results

```
. xtreg fatalityrate sb_usage speed65 speed70 drinkage21 ba08 lnincome age

Random-effects GLS regression                   Number of obs      =       556
Group variable: fips                            Number of groups   =        51

R-sq:                                           Obs per group:
     within  = 0.6834                                         min =         8
     between = 0.3426                                         avg =      10.9
     overall = 0.4805                                         max =        15

                                                Wald chi2(7)       =   1086.83
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
fatalityrate |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    sb_usage |  -.004504   .0011238    -4.01   0.000    -.0067066   -.0023014
     speed65 | -.0003406   .0003276    -1.04   0.298    -.0009827    .0003015
     speed70 |  .0013351   .0003287     4.06   0.000     .0006909    .0019793
  drinkage21 |   .000767   .0005097     1.50   0.132     -.000232     .001766
        ba08 | -.0013643    .000367    -3.72   0.000    -.0020836    -.000645
    lnincome | -.0126154   .0011453   -11.01   0.000    -.0148602   -.0103707
         age |  .0002318   .0002394     0.97   0.333    -.0002373     .000701
       _cons |  .1379473    .008919    15.47   0.000     .1204664    .1554282
-------------+----------------------------------------------------------------
     sigma_u | .00301581
     sigma_e |  .0017871
         rho | .74011226   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Review and summary

- **Random-effects estimator** saves degrees of freedom and allows estimation of effects of pure cross-section regressors if it is valid
- It is **inconsistent** if unit effects are correlated with regressors
- RE model is estimated by **feasible GLS** using a "quasi-de-meaning" process based on correlations of OLS residuals
- Hausman test can be used to examine whether RE estimator is consistent

Know your professor?

Which one of the following statements is true?

a. I played in a rock band in high school.
b. I have webbed toes.
c. My wife was 14 when we started dating.
d. I have performed in both New York's Macy's Thanksgiving Parade and Pasadena's Tournament of Roses Parade.
e. My first car had 3 cylinders.
f. All of the above are true.

What's next?

- We have now completed our brief examination of panel data models
- Next topic (April 13 and 16) is one of the very most important of the semester: **instrumental-variables** (IV) models for dealing with **endogenous regressors**
- We'll derive an estimator using the method of moments
- Standard IV method is **two-stage least squares**