Correlated Random Effects Methods for Panel Data Models with Heterogeneous Time Effects

2020 UK Stata Conference
September 10-11 2020

Jeff Wooldridge
Department of Economics
Michigan State University

1. Introduction

- Microeconometric setting with small T, large N.
- Standard unobserved effects model for random draw from the population:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{d}_t\boldsymbol{\gamma} + c_i + u_{it},\ t = 1,\ldots,T$$

- c_i are unobserved random variables (heterogeneity).
- Time period dummies:

$$\mathbf{d}_t = (d2_t,\ldots,dT_t)$$

  - Used to flexibly control for aggregate factors.

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{d}_t\boldsymbol{\gamma} + c_i + u_{it},\ t = 1,\ldots,T$$

- $\mathbf{x}_{it}$ only includes variables that have variation across i and t.
- $\boldsymbol{\beta}$ is of interest.
- Use fixed effects estimation to remove c_i.
- Sometimes called "two-way fixed effects," but $\boldsymbol{\gamma}$ are parameters, c_i are not.

- Limitation of the model: If $\mathbf{d}_t$ represents flexible trends, all units i follow the same trend.
- Allow heterogeneous time effects:

$$\begin{aligned} y_{it} &= \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{d}_t\mathbf{g}_i + c_i + u_{it} \\ &= \mathbf{x}_{it}\boldsymbol{\beta} + d2_t g_{i2} + d3_t g_{i3} + \cdots + dT_t g_{iT} + c_i + u_{it} \end{aligned}$$

- Each unit has its own intercept, and these also vary across t.

• Now write

$$\mathbf{g}_i \equiv \boldsymbol{\gamma} + \mathbf{h}_i, \; E(\mathbf{h}_i) = \mathbf{0}$$

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{d}_t\boldsymbol{\gamma} + c_i + \mathbf{d}_t\mathbf{h}_i + u_{it}$$

• Cannot treat $(c_i, \mathbf{h}_i)$ as parameters to estimate without assuming $\{u_{it} : t = 1, \ldots, T\}$ are IID. [Ahn, Lee, Schmidt (1993, Journal of Econometrics)].

• Should (but will not) carefully compare with small-T factor literature, especially ALS (2013, J of E).

▸ Current approach allows more heterogeneity.

▸ Current approach is computationally much simpler.

Remainder of Talk

- When are (generalized) FE estimators that ignore $\mathbf{d}_t\mathbf{h}_i$ consistent?
- Modeling $E(\mathbf{h}_i|\mathbf{x}_i)$ in

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{d}_t\boldsymbol{\gamma} + c_i + \mathbf{d}_t\mathbf{h}_i + u_{it}$$

use CRE and equivalance of different estimators.

- Empirical example.
- Extensions to unit-specific trends.

2. Robustness of Standard Fixed Effects Estimators

- Suppose the true model is

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{d}_t\boldsymbol{\gamma} + c_i + \mathbf{d}_t\mathbf{h}_i + u_{it}$$

$$E(u_{it}|\mathbf{x}_i, c_i, \mathbf{h}_i) = 0$$

but we ignore $\mathbf{d}_t\mathbf{h}_i$.

- Apply standard FE to remove c_i.
- Can apply an extension of Wooldridge (2005, REStat).

$$\bar{\mathbf{x}}_i = T^{-1} \sum_{t=1}^{T} \mathbf{x}_{it}$$

$$\ddot{\mathbf{x}}_{it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_i$$

- Sufficient is

$$E(\ddot{\mathbf{x}}_{it} \otimes \mathbf{h}_i) = \mathbf{0},\ t = 1, \ldots, T.$$

- Holds if

$$\mathbf{x}_{it} = \mathbf{q}_i + \mathbf{e}_{it}$$

$$E(\mathbf{e}_{it} \otimes \mathbf{h}_i) = \mathbf{0}, t = 1, \ldots, T$$

$$Cov(\mathbf{q}_i, \mathbf{h}_i) \text{ unrestricted}$$

- If $T > 2$, can remove more heterogeneity from $\mathbf{x}_{it}$, such as unit-specific linear trends.
- Now obtain $\ddot{\mathbf{x}}_{it}$ – detrended covariates – as residuals from

$$\mathbf{x}_{it} \text{ on } 1, t, \quad t = 1, \ldots, T$$

- This allows for unit-specific trends:

$$\mathbf{x}_{it} = \mathbf{q}_i + \mathbf{m}_i t + \mathbf{e}_{it}$$

$$Cov(\mathbf{q}_i, \mathbf{h}_i),\ Cov(\mathbf{m}_i, \mathbf{h}_i) \text{ unrestricted}$$

- But this representation for $\mathbf{x}_{it}$ is still special.

3. CRE Approach to Heterogeneous Time Effects

- Now explicitly recognize the heterogeneity terms $\mathbf{d}_t\mathbf{h}_i$ in

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{d}_t\boldsymbol{\gamma} + c_i + \mathbf{d}_t\mathbf{h}_i + u_{it}$$

$$E(u_{it}|\mathbf{x}_i, c_i, \mathbf{h}_i) = 0$$

- Remove c_i by within-unit demeaning:

$$\ddot{y}_{it} = \ddot{\mathbf{x}}_{it}\boldsymbol{\beta} + \ddot{\mathbf{d}}_t\boldsymbol{\gamma} + \ddot{\mathbf{d}}_t\mathbf{h}_i + \ddot{u}_{it}$$

- Make a Mundlak (1978) CRE assumption on $\mathbf{h}_i$:

$$E(h_{ir}|\mathbf{x}_i) = (\bar{\mathbf{x}}_i - \boldsymbol{\mu}_{\bar{\mathbf{x}}})\boldsymbol{\xi}_r, \quad r = 2, \ldots, T$$

$$\boldsymbol{\mu}_{\bar{\mathbf{x}}} \equiv E(\bar{\mathbf{x}}_i)$$

- Leads to

$$\ddot{y}_{it} = \ddot{\mathbf{x}}_{it}\boldsymbol{\beta} + \ddot{\mathbf{d}}_t\boldsymbol{\gamma} + \ddot{d}2_t(\bar{\mathbf{x}}_i - \boldsymbol{\mu}_{\bar{\mathbf{x}}})\boldsymbol{\xi}_2 + \cdots + \ddot{d}T_t(\bar{\mathbf{x}}_i - \boldsymbol{\mu}_{\bar{\mathbf{x}}})\boldsymbol{\xi}_T + \ddot{e}_{it}$$

$$E(\ddot{e}_{it}|\mathbf{x}_i) = 0$$

  - Replace $\boldsymbol{\mu}_{\bar{\mathbf{x}}}$ with the overall sample average, $\bar{\mathbf{x}}$.
  - Can estimate all parameters consistently by POLS.

- Numerically identical to applying usual FE to the equation

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{d}_t\boldsymbol{\gamma} + d2_t(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})\boldsymbol{\xi}_2 + \cdots + dT_t(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})\boldsymbol{\xi}_T + f_i + e_{it}$$

- Call the estimates $\hat{\boldsymbol{\beta}}_{FEH}, \hat{\boldsymbol{\gamma}}_{FEH}$.
- Use cluster-robust standard errors.
- Can use robust Wald test of

$$H_0 : \boldsymbol{\xi}_2 = \boldsymbol{\xi}_3 = \cdots = \boldsymbol{\xi}_T = \mathbf{0}$$

Comments

- It is possible to strongly reject

$$H_0 : \boldsymbol{\xi}_2 = \boldsymbol{\xi}_3 = \cdots = \boldsymbol{\xi}_T = \mathbf{0}$$

and have $\hat{\boldsymbol{\beta}}_{FEH} \approx \hat{\boldsymbol{\beta}}_{FE}$.

  - Provides a robustness check on usual FE estimates.

- All estimates can be obtained from the Mundlak equation:

$$y_{it} = \alpha + \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{d}_t\boldsymbol{\gamma} + \bar{\mathbf{x}}_i\boldsymbol{\psi} + d2_t(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})\boldsymbol{\xi}_2 + \cdots + dT_t(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})\boldsymbol{\xi}_T + r_{it}$$

  - Pooled OLS or RE both produce $\hat{\boldsymbol{\beta}}_{FEH}$, $\hat{\boldsymbol{\gamma}}_{FEH}$.

- Can easily include time-constant covariates, say $\mathbf{z}_i$.
  - Just adding $\mathbf{z}_i$ to the Mundlak equation changes none of the estimates (except intercept): Wooldridge (2019, Journal of Econometrics).
- Can add

$$ds_t(\mathbf{z}_i - \bar{\mathbf{z}})$$

to allow more heterogeneity in time effects.

  - This will change the estimates of $\boldsymbol{\beta}$ (and $\boldsymbol{\gamma}$).

4. Empirical Illustration

- Data on $N = 1,149$ U.S. air routes. $T = 4$.
- $lfare_{it}$ is log of average airfare. $concen_{it}$ is a market power measure.

```
. egen double concenbar = mean(concen), by(id)

. sum concenbar

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
   concenbar |     4,596    .6101149    .1888741      .1862      .9997

. gen double concenbar_dm = concenbar - r(mean)

. sum ldist

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       ldist |     4,596    6.696482    .6593177   4.553877   7.909857

. gen double ldist_dm = ldist - r(mean)
```

```
. xtreg lfare concen y98 y99 y00, fe vce(cluster id)

Fixed-effects (within) regression               Number of obs     =      4,596
Group variable: id                              Number of groups  =      1,149

R-sq:                                           Obs per group:
     within  = 0.1352                                         min =          4
     between = 0.0576                                         avg =        4.0
     overall = 0.0083                                         max =          4

                                                F(4,1148)         =     120.06
corr(u_i, Xb)  = -0.2033                        Prob > F          =     0.0000

                                 (Std. Err. adjusted for 1,149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |    .168859   .0494587     3.41   0.001     .0718194    .2658985
         y98 |   .0228328    .004163     5.48   0.000     .0146649    .0310007
         y99 |   .0363819   .0051275     7.10   0.000     .0263215    .0464422
         y00 |   .0977717   .0055054    17.76   0.000     .0869698    .1085735
       _cons |   4.953331   .0296765   166.91   0.000     4.895104    5.011557
-------------+----------------------------------------------------------------
     sigma_u |  .43389176
     sigma_e |  .10651186
         rho |  .94316439   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

```
. * Usual Mundlak regression, with time constant variables added:

. reg lfare concen concenbar c.ldist c.ldist_dm#c.ldist_dm y98 y99 y00 , vce(cluster id)

Linear regression                               Number of obs     =      4,596
                                                F(7, 1148)        =     181.88
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4068
                                                Root MSE          =     .33637

                                  (Std. Err. adjusted for 1,149 clusters in id)
-------------------------------------------------------------------------------
                      |               Robust
                lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------------+--------------------------------------------------------
               concen |    .168859   .0494749     3.41   0.001     .0717877    .2659303
            concenbar |   .2136346   .0816403     2.62   0.009     .0534538    .3738155
                ldist |   .4818306   .0178697    26.96   0.000     .4467698    .5168914
                      |
c.ldist_dm#c.ldist_dm |   .1038426   .0201911     5.14   0.000      .064227    .1434582
                      |
                  y98 |   .0228328   .0041643     5.48   0.000     .0146622    .0310033
                  y99 |   .0363819   .0051292     7.09   0.000     .0263183    .0464455
                  y00 |   .0977717   .0055072    17.75   0.000     .0869663     .108577
                _cons |   1.551289   .1473768    10.53   0.000     1.262131    1.840447
-------------------------------------------------------------------------------
```

```
. * Heterogeneous time effects using Mundlak:
.
. reg lfare concen concenbar c.ldist c.ldist_dm#c.ldist_dm y98 y99 y00 ///
>         c.y98#c.concenbar_dm c.y99#c.concenbar_dm c.y00#c.concenbar_dm, ///
>         vce(cluster id)

Linear regression                               Number of obs     =      4,596
                                                F(10, 1148)       =     136.24
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4078
                                                Root MSE          =     .33619

                                  (Std. Err. adjusted for 1,149 clusters in id)
--------------------------------------------------------------------------------------
                      |               Robust
                lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------------+---------------------------------------------------------------
               concen |    .168456   .0490432     3.43   0.001     .0722316    .2646805
            concenbar |    .116914    .083664     1.40   0.163    -.0472374    .2810655
                ldist |   .4818306   .0178755    26.95   0.000     .4467583    .5169029
                      |
c.ldist_dm#c.ldist_dm |   .1038426   .0201977     5.14   0.000     .0642141    .1434711
                      |
                  y98 |   .0228364   .0041561     5.49   0.000     .0146819    .0309908
                  y99 |   .0363788   .0050715     7.17   0.000     .0264284    .0463291
                  y00 |   .0977672   .0053859    18.15   0.000     .0871999    .1083346
```

```
                               |
        c.y98#c.concenbar_dm |   .0616642   .0232143     2.66   0.008     .0161169    .1072114
                               |
        c.y99#c.concenbar_dm |   .1307868   .0285472     4.58   0.000     .0747762    .1867974
                               |
        c.y00#c.concenbar_dm |   .1960431   .0318187     6.16   0.000     .1336138    .2584724
                               |
                       _cons |   1.610546   .1486973    10.83   0.000     1.318797    1.902295
------------------------------------------------------------------------------------------
```

```
. * FE gives same estimates:
.
. xtreg lfare concen y98 y99 y00 ///
>         c.y98#c.concenbar_dm c.y99#c.concenbar_dm c.y00#c.concenbar_dm, fe ///
>         vce(cluster id)

                                 (Std. Err. adjusted for 1,149 clusters in id)
------------------------------------------------------------------------------
                     |               Robust
               lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------------+--------------------------------------------------------
              concen |    .168456   .0490272     3.44   0.001     .0722631     .264649
                 y98 |   .0228364   .0041548     5.50   0.000     .0146846    .0309881
                 y99 |   .0363788   .0050698     7.18   0.000     .0264317    .0463259
                 y00 |   .0977672   .0053842    18.16   0.000     .0872033    .1083312
                     |
c.y98#c.concenbar_dm |   .0616642   .0232067     2.66   0.008     .0161318    .1071965
                     |
c.y99#c.concenbar_dm |   .1307868   .0285379     4.58   0.000     .0747945    .1867791
                     |
c.y00#c.concenbar_dm |   .1960431   .0318083     6.16   0.000     .1336342     .258452
                     |
               _cons |   4.953577   .0293317   168.88   0.000     4.896028    5.011127
---------------------+--------------------------------------------------------
```

```
. * Interact distance with time dummies:

. reg lfare concen concenbar ldist c.ldist_dm#c.ldist_dm y98 y99 y00 ///
>         c.y98#c.concenbar_dm c.y99#c.concenbar_dm c.y00#c.concenbar_dm ///
>         c.y98#c.ldist_dm c.y99#c.ldist_dm c.y00#c.ldist_dm, ///
>         vce(cluster id)

Linear regression                               Number of obs     =      4,596
                                                F(13, 1148)       =     106.35
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4082
                                                Root MSE          =     .33618

                                   (Std. Err. adjusted for 1,149 clusters in id)
---------------------------------------------------------------------------------------
                      |               Robust
                lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------------+----------------------------------------------------------------
               concen |   .1681968   .0492655     3.41   0.001     .0715364    .2648573
            concenbar |   .1497355   .0869212     1.72   0.085    -.0208067    .3202777
                ldist |   .4986949   .0197181    25.29   0.000     .4600074    .5373823
                      |
c.ldist_dm#c.ldist_dm |   .1038426   .0202043     5.14   0.000     .0642011    .1434841
                      |
                  y98 |   .0228387   .0041539     5.50   0.000     .0146886    .0309888
                  y99 |   .0363768   .0050668     7.18   0.000     .0264355     .046318
                  y00 |   .0977644   .0053295    18.34   0.000     .0873077     .108221
```

```
                  |
c.y98#c.concenbar_dm |   .0297475   .0267611     1.11   0.267    -.0227587    .0822536
                  |
c.y99#c.concenbar_dm |   .1160657   .0343829     3.38   0.001     .0486053    .1835261
                  |
c.y00#c.concenbar_dm |   .1124318    .036763     3.06   0.002     .0403016    .1845619
                  |
    c.y98#c.ldist_dm |  -.0165298   .0068546    -2.41   0.016    -.0299786   -.0030809
                  |
    c.y99#c.ldist_dm |  -.0076258    .009357    -0.81   0.415    -.0259845    .0107329
                  |
    c.y00#c.ldist_dm |  -.0433015   .0099287    -4.36   0.000    -.0627819   -.0238211
                  |
               _cons |   1.477749   .1633311     9.05   0.000     1.157288     1.79821
--------------------------------------------------------------------------------------

. test c.y98#c.concenbar_dm c.y99#c.concenbar_dm c.y00#c.concenbar_dm ///
>        c.y98#c.ldist_dm c.y99#c.ldist_dm c.y00#c.ldist_dm

 ( 1)  c.y98#c.concenbar_dm = 0
 ( 2)  c.y99#c.concenbar_dm = 0
 ( 3)  c.y00#c.concenbar_dm = 0
 ( 4)  c.y98#c.ldist_dm = 0
 ( 5)  c.y99#c.ldist_dm = 0
 ( 6)  c.y00#c.ldist_dm = 0

       F(  6,  1148) =   12.72
            Prob > F =    0.0000
```

5. Allowing for Heterogeneous Slopes

- A model where everything is heterogeneous:

$$y_{it} = \mathbf{x}_{it}\mathbf{b}_i + c_i + \mathbf{d}_t\mathbf{g}_i + u_{it}$$

$$\mathbf{b}_i = \boldsymbol{\beta} + \mathbf{a}_i$$
$$\mathbf{g}_i = \boldsymbol{\gamma} + \mathbf{h}_i$$

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + \mathbf{d}_t\boldsymbol{\gamma} + \mathbf{x}_{it}\mathbf{a}_i + \mathbf{d}_t\mathbf{h}_i + u_{it}$$

$$\mathbf{a}_i = \boldsymbol{\Lambda}(\bar{\mathbf{x}}_i - \boldsymbol{\mu}_{\bar{\mathbf{x}}})' + \mathbf{q}_i$$
$$\mathbf{h}_i = \boldsymbol{\Xi}(\bar{\mathbf{x}}_i - \boldsymbol{\mu}_{\bar{\mathbf{x}}})' + \mathbf{f}_i$$

$$
\begin{aligned}
y_{it} &= \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{d}_t\boldsymbol{\gamma} + \mathbf{x}_{it}[\boldsymbol{\Lambda}(\bar{\mathbf{x}}_i - \boldsymbol{\mu}_{\bar{\mathbf{x}}})'] + \mathbf{d}_t[\boldsymbol{\Lambda}(\bar{\mathbf{x}}_i - \boldsymbol{\mu}_{\bar{\mathbf{x}}})'] + c_i + v_{it} \\
&= \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{d}_t\boldsymbol{\gamma} + [(\bar{\mathbf{x}}_i - \boldsymbol{\mu}_{\bar{\mathbf{x}}}) \otimes \mathbf{x}_{it}]\boldsymbol{\lambda} + [(\bar{\mathbf{x}}_i - \boldsymbol{\mu}_{\bar{\mathbf{x}}}) \otimes \mathbf{d}_t]\boldsymbol{\xi} + c_i + v_{it}
\end{aligned}
$$

$$
E(v_{it}|\mathbf{x}_i) = 0,\ t = 1,\ldots,T
$$

- Remove c_i using fixed effects.
- Or apply Mundlak, adding

$$
\bar{\mathbf{x}}_i,\ (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}) \otimes \bar{\mathbf{x}}_i
$$

- Can also add

$$
(\mathbf{z}_i - \bar{\mathbf{z}}) \otimes \mathbf{x}_{it}
$$

```
. * Now heterogeneous slope on concen, too.
.
. reg lfare concen c.concen#c.concenbar_dm c.concen#c.ldist_dm ///
>         concenbar ldist c.ldist_dm#c.ldist_dm y98 y99 y00 ///
>         c.concenbar#c.concenbar_dm c.concenbar#c.ldist_dm ///
>         c.y98#c.concenbar_dm c.y99#c.concenbar_dm c.y00#c.concenbar_dm ///
>         c.y98#c.ldist_dm c.y99#c.ldist_dm c.y00#c.ldist_dm, ///
>         vce(cluster id)

                                  (Std. Err. adjusted for 1,149 clusters in
-----------------------------------------------------------------------------
                         |               Robust
                   lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval
-------------------------+---------------------------------------------------------------
                  concen |   .1690247   .0500444     3.38   0.001     .0708359    .2672134
                         |
   c.concen#c.concenbar_dm |   .3101472   .4075416     0.76   0.447    -.4894626    1.109757
                         |
       c.concen#c.ldist_dm |  -.2097112   .0907065    -2.31   0.021    -.3876804   -.0317421
                         |
               concenbar |   .1396369   .2389655     0.58   0.559    -.3292212    .6084949
                   ldist |   .3880281   .0856013     4.53   0.000     .2200756    .5559807
                         |
   c.ldist_dm#c.ldist_dm |   .1286053   .0225076     5.71   0.000     .0844446     .172766
                         |
                     y98 |   .0230028   .0041427     5.55   0.000     .0148747    .0311308
                     y99 |   .0355884   .0050872     7.00   0.000     .0256071    .0455698
                     y00 |   .0976618    .005291    18.46   0.000     .0872806     .108043
```

c.concenbar#c.concenbar_dm	-.2588744	.5716612	-0.45	0.651	-1.380492	.8627436
c.concenbar#c.ldist_dm	.4014202	.1687987	2.38	0.018	.0702316	.7326088
c.y98#c.concenbar_dm	.0262151	.0272121	0.96	0.336	-.027176	.0796061
c.y99#c.concenbar_dm	.1131609	.0340452	3.32	0.001	.0463631	.1799587
c.y00#c.concenbar_dm	.1065938	.0363228	2.93	0.003	.0353272	.1778604
c.y98#c.ldist_dm	-.0144794	.0069094	-2.10	0.036	-.0280359	-.000923
c.y99#c.ldist_dm	-.0080089	.0093604	-0.86	0.392	-.0263743	.0103564
c.y00#c.ldist_dm	-.0442223	.0096118	-4.60	0.000	-.0630809	-.0253637
_cons	2.225276	.6831325	3.26	0.001	.8849479	3.565604

6. Allowing for Unit-Specific Trends

- Now allow the base model to have heterogeneous linear trends:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{d}_t\boldsymbol{\gamma} + c_i + g_i t + \mathbf{d}_t\mathbf{h}_i + u_{it}$$

- In removing the linear trends, lose one element of $\mathbf{d}_t$.
- In running the regressions (for each i),

$$\mathbf{x}_{it} \text{ on } 1, t, \;\; t = 1, \ldots, T,$$

let

$$\hat{\mathbf{x}}_{it} = \hat{\mathbf{a}}_{i0} + \hat{\mathbf{a}}_{i1} t$$

- Detrended values:

$$\ddot{\mathbf{x}}_{it} = \mathbf{x}_{it} - \hat{\mathbf{x}}_{it}$$

$$\ddot{y}_{it} = \ddot{\mathbf{x}}_{it}\boldsymbol{\beta} + \ddot{\mathbf{d}}_t\boldsymbol{\gamma} + \ddot{\mathbf{d}}_t\mathbf{h}_i + \ddot{u}_{it}$$

- Model $\mathbf{h}_i$ as a (linear) function of

$$\hat{\mathbf{a}}_i = (\hat{\mathbf{a}}_{i0}, \hat{\mathbf{a}}_{i1})$$

- Use pooled OLS on

$$\ddot{y}_{it} = \ddot{\mathbf{x}}_{it}\boldsymbol{\beta} + \ddot{\mathbf{d}}_t\boldsymbol{\gamma} + (\ddot{\mathbf{d}}_t \otimes \hat{\mathbf{a}}_i)\boldsymbol{\xi} + \ddot{e}_{it}$$

- Using an extension of the Mundlak result, can show it is algebraically equivalenet to using POLS on

$$y_{it} = \alpha + \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{d}_t\boldsymbol{\gamma} + \bar{\mathbf{x}}_i\boldsymbol{\eta} + \hat{\mathbf{x}}_{it}\boldsymbol{\theta} + (\mathbf{d}_t \otimes \hat{\mathbf{a}}_i)\boldsymbol{\xi} + r_{it}$$

$$\hat{\mathbf{x}}_{it} = \hat{\mathbf{a}}_{i0} + \hat{\mathbf{a}}_{i1}t$$

- Apply to a passenger “demand” model.

```
. * First apply the previous model where only time averages are used.

. reg lpassen concen concenbar ldist ldistsq y98 y99 y00 ///
>         c.y98#c.concenbar_dm c.y99#c.concenbar_dm c.y00#c.concenbar_dm ///
>         lfare lfarebar c.y98#c.lfarebar_dm c.y99#c.lfarebar_dm ///
>         c.y00#c.lfarebar_dm, vce(cluster id)

Linear regression                               Number of obs     =      4,596
                                                F(15, 1148)       =      39.97
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0770
                                                Root MSE          =     .85065

                                 (Std. Err. adjusted for 1,149 clusters in id)
------------------------------------------------------------------------------
                     |               Robust
             lpassen |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------------+--------------------------------------------------------
              concen |   .1390506   .0899681     1.55   0.122    -.0374697    .3155709
           concenbar |  -.6315775   .2026034    -3.12   0.002    -1.029092   -.2340631
               ldist |  -1.303931   .7013855    -1.86   0.063    -2.680072    .0722106
             ldistsq |   .0949819     .05304     1.79   0.074    -.0090842     .199048
                 y98 |    .045365    .004841     9.37   0.000     .0358667    .0548632
                 y99 |   .1035984    .006301    16.44   0.000     .0912357    .1159612
                 y00 |   .1966456   .0101262    19.42   0.000     .1767777    .2165135
```

c.y98#c.concenbar_dm	-.0011952	.026645	-0.04	0.964	-.0534735	.0510832
c.y99#c.concenbar_dm	-.035792	.0343854	-1.04	0.298	-.1032573	.0316732
c.y00#c.concenbar_dm	-.1049551	.0379231	-2.77	0.006	-.1793614	-.0305489
lfare	-1.15989	.1091893	-10.62	0.000	-1.374123	-.9456567
lfarebar	.7215856	.1377878	5.24	0.000	.4512414	.9919297
c.y98#c.lfarebar_dm	-.053968	.0112635	-4.79	0.000	-.0760674	-.0318685
c.y99#c.lfarebar_dm	-.054394	.0133306	-4.08	0.000	-.0805491	-.028239
c.y00#c.lfarebar_dm	-.0275061	.0166339	-1.65	0.098	-.0601424	.0051302
_cons	12.89575	2.323339	5.55	0.000	8.337285	17.45422

```
. * Now use the unit-specific linear trend heterogeneity:

. gen double lpassen_h = .
(4,596 missing values generated)

. gen double lpassen_dt = .
(4,596 missing values generated)

. gen double lfare_h = .
(4,596 missing values generated)

. gen double lfare_dt = .
(4,596 missing values generated)

. gen double lfare_a0 = .
(4,596 missing values generated)

. gen double lfare_a1 = .
(4,596 missing values generated)

. gen double concen_h = .
(4,596 missing values generated)

. gen double concen_dt = .
(4,596 missing values generated)

. gen double concen_a0 = .
(4,596 missing values generated)
```

```
. gen double concen_a1 = .
(4,596 missing values generated)

. gen double y98_h = .
(4,596 missing values generated)

. gen double y99_h = .
(4,596 missing values generated)

. gen double y00_h = .
(4,596 missing values generated)

. gen double y98_dt = .
(4,596 missing values generated)

. gen double y99_dt = .
(4,596 missing values generated)

. gen double y00_dt = .
(4,596 missing values generated)

. gen double y98_a1_t = .
(4,596 missing values generated)

. gen double y99_a1_t = .
(4,596 missing values generated)

. gen double y00_a1_t = .
(4,596 missing values generated)
```

```
. gen t = year - 1997

. local i = 1

. while `i' <= 1149 {
  2.
.         qui reg lpassen t if id == `i'
  3.         predict lpassen_t, xb
  4.         qui replace lpassen_h = lpassen_t if id == `i'
  5.         qui replace lpassen_dt = lpassen - lpassen_t if id == `i'
  6.
.         qui reg lfare t if id == `i'
  7.         qui replace lfare_a0 = _b[_cons] if id == `i'
  8.         qui replace lfare_a1 = _b[t] if id == `i'
  9.         predict lfare_t, xb
 10.         qui replace lfare_h = lfare_t if id == `i'
 11.         qui replace lfare_dt = lfare - lfare_t if id == `i'
 12.
.         qui reg concen t if id == `i'
 13.         qui replace concen_a0 = _b[_cons] if id == `i'
 14.         qui replace concen_a1 = _b[t] if id == `i'
 15.         predict concen_t, xb
 16.         qui replace concen_h = concen_t if id == `i'
 17.     qui replace concen_dt = concen - concen_t if id == `i'
 18.
```

```
.         qui reg y98 t if id == `i'
 19.     predict y98_t, xb
 20.         qui replace y98_h = y98_t if id == `i'
 21.     qui replace y98_dt = y98 - y98_t if id == `i'
 22.         qui replace y98_a1_t = _b[t]*t if id == `i'
 23.
.         qui reg y99 t if id == `i'
 24.     predict y99_t, xb
 25.         qui replace y99_h = y99_t if id == `i'
 26.     qui replace y99_dt = y99 - y99_t if id == `i'
 27.         qui replace y99_a1_t = _b[t]*t if id == `i'
 28.
.     qui reg y00 t if id == `i'
 29.     predict y00_t, xb
 30.         qui replace y00_h = y00_t if id == `i'
 31.     qui replace y00_dt = y00 - y00_t if id == `i'
 32.         qui replace y00_a1_t = _b[t]*t if id == `i'
 33.
.     drop lfare_t concen_t lpassen_t y98_t y99_t y00_t
 34.
.     local i = `i' + 1
 35. }
```

```
. sum concen_h

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
    concen_h |      4,596    .6101149    .1929988     .15482     1.0747

. gen double concen_h_dm = concen_h - r(mean)

. sum lfare_h

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
     lfare_h |      4,596    5.095601    .4322098   3.836819    6.26246

. gen double lfare_h_dm = lfare_h - r(mean)

. sum concen_a0

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
   concen_a0 |      4,596    .6175428    .1959924     .15482    1.01686

. gen double concen_a0_dm = concen_a0 - r(mean)
```

```
. sum concen_a1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
   concen_a1 |      4,596   -.0049519    .0351504    -.16346      .1463

. gen double concen_a1_dm = concen_a1 - r(mean)

. sum lfare_a0

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
    lfare_a0 |      4,596    5.050825    .4526443   3.836819   6.069202

. gen double lfare_a0_dm = lfare_a0 - r(mean)

. sum lfare_a1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
    lfare_a1 |      4,596    .0298502    .0637264  -.3344203   .5016729

. gen double lfare_a1_dm = lfare_a1 - r(mean)
```

```
. reg lpassen_dt concen_dt lfare_dt y98_dt y99_dt y00_dt ///
>         c.y98_dt#c.concen_a0_dm c.y99_dt#c.concen_a0_dm c.y00_dt#c.concen_a0_dm ///
>         c.y98_dt#c.lfare_a0_dm c.y99_dt#c.lfare_a0_dm c.y00_dt#c.lfare_a0_dm ///
>         c.y98_dt#c.concen_a1_dm c.y99_dt#c.concen_a1_dm c.y00_dt#c.concen_a1_dm ///
>         c.y98_dt#c.lfare_a1_dm c.y99_dt#c.lfare_a1_dm c.y00_dt#c.lfare_a1_dm,
>         nocons vce(cluster  id)

note: y00_dt omitted because of collinearity
note: c.y00_dt#c.concen_a0_dm omitted because of collinearity
note: c.y00_dt#c.lfare_a0_dm omitted because of collinearity
note: c.y00_dt#c.concen_a1_dm omitted because of collinearity
note: c.y00_dt#c.lfare_a1_dm omitted because of collinearity

Linear regression                               Number of obs     =      4,596
                                                F(12, 1148)       =      34.15
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4038
                                                Root MSE          =     .07314

                                    (Std. Err. adjusted for 1,149 clusters in id)
------------------------------------------------------------------------------------------
                         |               Robust
             lpassen_dt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------------+-----------------------------------------------------------------
              concen_dt |   .0049276   .0951786     0.05   0.959     -.181816    .1916712
               lfare_dt |  -1.006965   .0660814   -15.24   0.000    -1.136619   -.8773108
                 y98_dt |   -.017328    .004024    -4.31   0.000    -.0252232   -.0094328
                 y99_dt |  -.0231335   .0044971    -5.14   0.000    -.0319568   -.0143101
                 y00_dt |          0  (omitted)


                        |
```

c.y98_dt#c.concen_a0_dm	.0247107	.0192804	1.28	0.200	-.013118	.0625394
c.y99_dt#c.concen_a0_dm	.0308164	.0193918	1.59	0.112	-.007231	.0688639
c.y00_dt#c.concen_a0_dm	0	(omitted)				
c.y98_dt#c.lfare_a0_dm	-.0329983	.0083758	-3.94	0.000	-.0494319	-.0165648
c.y99_dt#c.lfare_a0_dm	-.0375524	.0079381	-4.73	0.000	-.0531273	-.0219776
c.y00_dt#c.lfare_a0_dm	0	(omitted)				
c.y98_dt#c.concen_a1_dm	.0450727	.1310509	0.34	0.731	-.2120534	.3021988
c.y99_dt#c.concen_a1_dm	.0196799	.1975331	0.10	0.921	-.3678863	.4072462
c.y00_dt#c.concen_a1_dm	0	(omitted)				
c.y98_dt#c.lfare_a1_dm	.198575	.1346018	1.48	0.140	-.0655181	.4626681
c.y99_dt#c.lfare_a1_dm	.0075498	.2000518	0.04	0.970	-.3849584	.400058
c.y00_dt#c.lfare_a1_dm	0	(omitted)				

```
. reg lpassen concen lfare c.ldist c.ldist_dm#c.ldist_dm y98 y99 y00 ///
>         concenbar lfarebar concen_h lfare_h ///
>         c.y98#c.concen_a0_dm c.y99#c.concen_a0_dm c.y00#c.concen_a0_dm ///
>         c.y98#c.lfare_a0_dm c.y99#c.lfare_a0_dm c.y00#c.lfare_a0_dm ///
>         c.y98#c.concen_a1_dm c.y99#c.concen_a1_dm c.y00#c.concen_a1_dm ///
>         c.y98#c.lfare_a1_dm c.y99#c.lfare_a1_dm c.y00#c.lfare_a1_dm, vce(cluster id)

Linear regression                               Number of obs     =      4,596
                                                F(23, 1148)       =      33.03
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1057
                                                Root MSE          =     .83806

                                    (Std. Err. adjusted for 1,149 clusters in id)
-------------------------------------------------------------------------------------
                      |               Robust
              lpassen |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------------+--------------------------------------------------------------
               concen |   .0049276   .0953035     0.05   0.959    -.1820609    .1919161
                lfare |  -1.006965   .0661681   -15.22   0.000    -1.136789    -.877141
                ldist |    -.01635   .0592676    -0.28   0.783     -.132635     .099935
                      |
c.ldist_dm#c.ldist_dm |    .121346   .0530347     2.29   0.022     .0172902    .2254018
                      |
                  y98 |   .0242774   .0112826     2.15   0.032     .0021406    .0464141
                  y99 |   .0600772   .0212456     2.83   0.005     .0183926    .1017618
                  y00 |    .124816   .0300917     4.15   0.000     .0657752    .1838568
```

concenbar	-2.458434	.4773988	-5.15	0.000	-3.395106	-1.521762
lfarebar	-.475291	.3082788	-1.54	0.123	-1.080144	.1295621
concen_h	2.055851	.4549934	4.52	0.000	1.163139	2.948563
lfare_h	.962927	.3047852	3.16	0.002	.3649286	1.560925
c.y98#c.concen_a0_dm	-.0040696	.024358	-0.17	0.867	-.0518608	.0437216
c.y99#c.concen_a0_dm	-.026744	.0318316	-0.84	0.401	-.0891988	.0357107
c.y00#c.concen_a0_dm	-.0863407	.0366607	-2.36	0.019	-.1582702	-.0144112
c.y98#c.lfare_a0_dm	-.0423505	.0109446	-3.87	0.000	-.0638241	-.0208769
c.y99#c.lfare_a0_dm	-.0562567	.0141338	-3.98	0.000	-.0839876	-.0285258
c.y00#c.lfare_a0_dm	-.0280564	.0171804	-1.63	0.103	-.061765	.0056521
c.y98#c.concen_a1_dm	-1.814706	.474012	-3.83	0.000	-2.744733	-.8846795
c.y99#c.concen_a1_dm	-3.699879	.9978239	-3.71	0.000	-5.657642	-1.742116
c.y00#c.concen_a1_dm	-5.579338	1.415692	-3.94	0.000	-8.356972	-2.801704
c.y98#c.lfare_a1_dm	-1.028896	.3812075	-2.70	0.007	-1.776838	-.2809546
c.y99#c.lfare_a1_dm	-2.447393	.7684671	-3.18	0.001	-3.955151	-.9396358
c.y00#c.lfare_a1_dm	-3.682415	1.09094	-3.38	0.001	-5.822874	-1.541956
_cons	8.910393	.4332756	20.57	0.000	8.060292	9.760494

7. Extension to Unbalanced Panels

- Unbalanced panels, using the complete cases.
- Use functions of

$$\{(s_{it}, s_{it}\mathbf{x}_{it}) : t = 1, \ldots, T\}$$

- Now more care is needed with the aggregate time dummies [Wooldridge (2019)].