UNOBSERVED EFFECTS LINEAR PANEL DATA MODELS, III

Econometric Analysis of Cross Section and Panel Data, 2e
MIT Press
Jeffrey M. Wooldridge

1. Random Trend Models
2. General Models with Unit-Specific Slopes
3. Robustness of Standard Fixed Effects Estimators
4. Testing for Correlated Random Slopes

1. RANDOM TREND MODELS

- Consider an extension of the usual unobserved effects model:

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

where g_i is an individual-specific linear trend. This model explicitly allows for two sources of heterogeneity: an intercept effect and a trend effect.

- Somewhat unfortunately, this is often called a **random trend model**. For us, the "random" is redundant because all heterogeneity is treated as random outcomes. "Heterogeneous trend" model would be better.

- Ideally, we can let the covariates, $\{\mathbf{x}_{it} : t = 1, \ldots, T\}$, be arbitrarily correlated with both sources of heterogeneity. This puts us in a "fixed effects" environment. "Random effects" approaches are also available, where we treat both sources of heterogeneity as uncorrelated with $\{\mathbf{x}_{it} : t = 1, \ldots, T\}$.

• For interpretative purposes, a useful assumption is

$$E(y_{it}|\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}, c_i, g_i) = E(y_{it}|\mathbf{x}_{it}, c_i, g_i) = c_i + g_i t + \mathbf{x}_{it}\boldsymbol{\beta}$$

which means

$$E(u_{it}|\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}, c_i, g_i) = 0,\ t = 1, \ldots, T.$$

• Although we are controlling for two sources of unit-specific heterogeneity, this is still a strict exogeneity assumption on $\{\mathbf{x}_{it}\}$ with respect to the $\{u_{it}\}$.

• In an RE framework, we would assume $E(c_i|\mathbf{x}_i) = E(c_i)$ and $E(g_i|\mathbf{x}_i) = E(g_i)$, and so

$$E(y_{it}|\mathbf{x}_i) = E(y_{it}|\mathbf{x}_{it}) = \mu_c + \mu_g t + \mathbf{x}_{it}\boldsymbol{\beta}.$$

- But the RE assumption rules out policy analysis where the intervention is related to unobserved heterogeneity. If we allow $D(c_i, g_i|\mathbf{x}_i)$ to be unrestricted, then the analysis is usually more convincing.
- The FE setup is useful for estimating economic relationships, too. In a production function environment, $g_i t$ captures firm-specific trends in productivity, u_{it} is shocks to output, and $\mathbf{x}_{it}$ contains inputs.
- When y_{it} is a natural log, say $\log(q_{it})$, the model is often called the "random growth" model. (Holding $\mathbf{x}_{it}$ and u_{it} fixed, g_i is the proportionate increase in q_{it} from one period to the next.)

- Under strict exogeneity and sufficient variation in $\{\mathbf{x}_{it}\}$, there are a variety of estimation methods. A simple one is to first difference the equation to remove c_i:

$$\Delta y_{it} = g_i + \Delta \mathbf{x}_{it} \boldsymbol{\beta} + \Delta u_{it}, t = 2, \ldots, T,$$

using $g_i t - g_i(t-1) = g_i$. This is now a standard unobserved effects model but where all variables are changes.

- To remove g_i, we can difference again or apply FE to the FD equation.
- Either approach requires $T \geq 3$ time periods.
- If $\{u_{it}\}$ has strong, positive serial correlation, FD followed by FE has appeal. If $\{u_{it}\}$ is a random walk, then $\{\Delta u_{it}\}$ is serially uncorrelated.

• If $\{u_{it}\}$ contains little serial correlation, differencing can be inefficient. An alternative is to use unit-specific linear detrending. Write the equations for all time periods as

$$\mathbf{y}_i = \mathbf{W}_T \mathbf{a}_i + \mathbf{X}_i \boldsymbol{\beta} + \mathbf{u}_i$$

where

$$\mathbf{W}_T = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & T \end{pmatrix}, \quad \mathbf{a}_i = \begin{pmatrix} c_i \\ g_i \end{pmatrix}.$$

- Let

$$\mathbf{Q}_T = \mathbf{I}_T - \mathbf{W}_T(\mathbf{W}_T'\mathbf{W}_T)^{-1}\mathbf{W}_T'$$

$$\mathbf{Q}_T\mathbf{y}_i = \mathbf{Q}_T\mathbf{W}_T\mathbf{a}_i + \mathbf{Q}_T\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Q}_T\mathbf{u}_i$$

or

$$\ddot{\mathbf{y}}_i = \ddot{\mathbf{X}}_i\boldsymbol{\beta} + \ddot{\mathbf{u}}_i$$

because $\mathbf{Q}_T\mathbf{W}_T = \mathbf{0}$. This is now an equation in unit-specific linearly detrended variables.

- In particular, the SOLS estimator is a pooled OLS estimator of

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

where now the “double-dot” variables denote unit-specific detrending. For example, for each i, run the regression

$$y_{it} \text{ on } 1, t,\quad t = 1, \ldots, T$$

and obtain the residuals, $\ddot{y}_{it}$. And similarly for each element of $\ddot{\mathbf{x}}_{it}$. (This is called the “fixed effects” estimator for the random trend model.)

• Fully robust inference looks the same as before. Let $\hat{\ddot{\mathbf{u}}}_i = \ddot{\mathbf{y}}_i - \ddot{\mathbf{X}}_i\hat{\boldsymbol{\beta}}_{FE}$ be the residuals. The fully robust variance matrix estimator is

$$\widehat{Avar}(\hat{\boldsymbol{\beta}}_{FE}) = \left(\sum_{i=1}^{N}\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i\right)^{-1}\left(\sum_{i=1}^{N}\ddot{\mathbf{X}}_i'\hat{\ddot{\mathbf{u}}}_i\hat{\ddot{\mathbf{u}}}_i'\ddot{\mathbf{X}}_i\right)\left(\sum_{i=1}^{N}\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i\right)^{-1}$$

with sometimes with a df adjustment.

- For estimating σ_u^2, we have effectively $N(T-2)-K$ degrees-of-freedom: we lose df two for each i.
- If

$$E(\mathbf{u}_i\mathbf{u}_i'|\mathbf{x}_i, c_i, g_i) = \sigma_u^2\mathbf{I}_T,$$

then

$$\hat{\sigma}_u^2 = [N(T-2)-K]^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{\ddot{u}}_{it}^{2} \xrightarrow{p} \sigma_u^2$$

and $N \to \infty$.

- Then,

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

- Similar results go through for IV approaches. In effect, the IVs, regressors, and response variable are linearly detrended at the unit level, followed by pooled 2SLS.
- One could apply GLS methods, too, or GMM.

EXAMPLE: Random Trend Model for Airfares

```
. do rantrend1

. use airfare, clear

.
. xtset id year
       panel variable:  id (strongly balanced)
        time variable:  year, 1997 to 2000
                delta:  1 unit
.
. gen lfare_dt = .
(4596 missing values generated)

. gen concen_dt = .
(4596 missing values generated)

. gen y99_dt = .
(4596 missing values generated)

. gen y00_dt = .
(4596 missing values generated)
```

```
local i = 1
while ‘i’ <= 1149 {
       qui reg lfare year if id == ‘i’
       predict lfare_t, xb
       qui replace lfare_dt = lfare - lfare_t if id == ‘i’

       qui reg concen year if id == ‘i’
       predict concen_t, xb
       qui replace concen_dt = concen - concen_t if id == ‘i’

       qui reg y99 year if id == ‘i’
       predict y99_t, xb
       qui replace y99_dt = y99 - y99_t if id == ‘i’

       qui reg y00 year if id == ‘i’
       predict y00_t, xb
       qui replace y00_dt = y00 - y00_t if id == ‘i’

       drop lfare_t concen_t y99_t y00_t

       local i = ‘i’ + 1

}
```

```
. reg lfare_dt concen_dt y99_dt y00_dt, nocons cluster(id)

Linear regression                                      Number of obs =    4596
                                                       F(  3,  1148) =   33.64
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0459
                                                       Root MSE      =  .05894

                                  (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
   lfare_dt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   concen_dt |   .1590414   .0463449     3.43   0.001     .0681113    .2499715
      y99_dt |  -.0095344   .0058903    -1.62   0.106    -.0210914    .0020226
      y00_dt |   .0289026   .0100883     2.86   0.004     .0091089    .0486962
------------------------------------------------------------------------------
.
end of do-file
```

2. GENERAL MODELS WITH UNIT-SPECIFIC SLOPES

- Consider now a model written as

$$y_{it} = \mathbf{w}_{it}\mathbf{a}_i + \mathbf{x}_{it}\boldsymbol{\beta} + u_{it}$$

where $\mathbf{w}_{it}$ is $1 \times J$, $\mathbf{x}_{it}$ is $1 \times K$. Here, $\mathbf{a}_i$ is a $J \times 1$ vector of unobserved heterogeneity. We can hope to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\alpha} = E(\mathbf{a}_i)$ (and maybe other features of the distribution of $\mathbf{a}_i$).

- Write as

$$\mathbf{y}_i = \mathbf{W}_i\mathbf{a}_i + \mathbf{X}_i\boldsymbol{\beta} + \mathbf{u}_i$$

$$E(\mathbf{u}_i|\mathbf{w}_i, \mathbf{x}_i, \mathbf{a}_i) = \mathbf{0}.$$

• Define $\mathbf{M}_i = \mathbf{I}_T - \mathbf{W}_i(\mathbf{W}_i'\mathbf{W}_i)^{-1}\mathbf{W}_i'$ and so

$$\mathbf{M}_i\mathbf{y}_i = \mathbf{M}_i\mathbf{W}_i\mathbf{a}_i + \mathbf{M}_i\mathbf{X}_i\boldsymbol{\beta} + \mathbf{M}_i\mathbf{u}_i$$

so

$$\ddot{\mathbf{y}}_i = \ddot{\mathbf{X}}_i\boldsymbol{\beta} + \ddot{\mathbf{u}}_i$$

because $\mathbf{M}_i\mathbf{W}_i = \mathbf{0}$.

• Again, we can now use system OLS, which is a pooled OLS estimator on unit-specific residuals. The $\ddot{y}_{it}$ are from the unit-specific regression

$$y_{it} \text{ on } \mathbf{w}_{it},\ t = 1,\ldots,T$$

and similarly for $\ddot{\mathbf{x}}_{it}$.

• This gives us the fixed effects estimator of $\boldsymbol{\beta}$, with fully robust inference straightforward. We need the standard rank condition

$$rank\, E(\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i) = K$$

Because $rank(\mathbf{M}_i) = T - J$, a necessary condition is $T > J$.

• Under

$$E(\mathbf{u}_i\mathbf{u}_i'|\mathbf{w}_i,\mathbf{x}_i,\mathbf{a}_i) = \sigma_u^2\mathbf{I}_T,$$

use

$$\hat{\sigma}_u^2 = [N(T-J)-K]^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{\ddot{u}}_{it}^2,$$

and so on.

• What about estimating $\boldsymbol{\alpha}$? For example, perhaps a policy indicator has a unit-specific effect, and we want to estimate the average effect in the population (average partial effect).

• Pre-multiply the equation by $(\mathbf{W}_i'\mathbf{W}_i)^{-1}\mathbf{W}_i'$ assuming that $\mathbf{W}_i'\mathbf{W}_i$ is nonsingular with probability one:

$$(\mathbf{W}_i'\mathbf{W}_i)^{-1}\mathbf{W}_i'\mathbf{y}_i = \mathbf{a}_i + (\mathbf{W}_i'\mathbf{W}_i)^{-1}\mathbf{W}_i'\mathbf{X}_i\boldsymbol{\beta} - (\mathbf{W}_i'\mathbf{W}_i)^{-1}\mathbf{W}_i'\mathbf{u}_i$$

Solving for $\mathbf{a}_i$ gives

$$\mathbf{a}_i = (\mathbf{W}_i'\mathbf{W}_i)^{-1}\mathbf{W}_i'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) - (\mathbf{W}_i'\mathbf{W}_i)^{-1}\mathbf{W}_i'\mathbf{u}_i.$$

- Now take the expected value

$$\boldsymbol{\alpha} \equiv E(\mathbf{a}_i) = E[(\mathbf{W}_i'\mathbf{W}_i)^{-1}\mathbf{W}_i'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})].$$

- Replace population means with sample averages and replace $\boldsymbol{\beta}$ with $\hat{\boldsymbol{\beta}}_{FE}$:

$$\hat{\boldsymbol{\alpha}} = N^{-1}\sum_{i=1}^{N}(\mathbf{W}_i'\mathbf{W}_i)^{-1}\mathbf{W}_i'(\mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}_{FE}) \equiv N^{-1}\sum_{i=1}^{N}\hat{\mathbf{a}}_i.$$

• Can get an expression for the asymptotic variance matrix estimator of $\hat{\boldsymbol{\alpha}}$.

$$\hat{\mathbf{C}} = N^{-1} \sum_{i=1}^{N} (\mathbf{W}_i'\mathbf{W}_i)^{-1}\mathbf{W}_i'\mathbf{X}_i$$

$$\hat{\mathbf{A}} = N^{-1} \sum_{i=1}^{N} \ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i$$

$$\widehat{Avar}(\hat{\boldsymbol{\alpha}}) = \sum_{i=1}^{N} [(\hat{\mathbf{a}}_i - \hat{\boldsymbol{\alpha}}) - \hat{\mathbf{C}}\hat{\mathbf{A}}^{-1}\ddot{\mathbf{X}}_i'\widehat{\ddot{\mathbf{u}}}_i][(\hat{\mathbf{a}}_i - \hat{\boldsymbol{\alpha}}) - \hat{\mathbf{C}}\hat{\mathbf{A}}^{-1}\ddot{\mathbf{X}}_i'\widehat{\ddot{\mathbf{u}}}_i]'.$$

• Generally, $\hat{\boldsymbol{\alpha}}$ is not the asymptotically efficient estimator.

3. ROBUSTNESS OF STANDARD FIXED EFFECTS ESTIMATORS

- To apply the previous results, need $T > J$, and so if we have many covariates relative to time periods, we have to (arbitrarily) set some coefficients to constants, allow others to be random.

- Now we study what happens if we mistakenly treat some random slopes as if they are fixed and apply standard fixed effects methods. We might ignore some heterogeneity because we are ignorant of scope of heterogeneity in the model or because we simply do not have enough time periods to proceed with a general analysis.

• Start with the basic model but where all slopes allowed to vary by unit:

$$
\begin{aligned}
y_{it} &= c_i + \mathbf{x}_{it}\mathbf{b}_i + u_{it} \\
E(u_{it}|\mathbf{x}_i, c_i, \mathbf{b}_i) &= 0, t = 1, \ldots, T,
\end{aligned}
$$

where $\mathbf{b}_i$ is $K \times 1$. We now ignore the heterogeneity in the slopes and act as if $\mathbf{b}_i$ is constant all i. We think c_i might be correlated with at least some elements of $\mathbf{x}_{it}$.

• Question: If we apply the usual fixed effects estimator (that only eliminates c_i), when does it consistently estimate the average partial effect (population average effect), $\boldsymbol{\beta} = E(\mathbf{b}_i)$?

• We need the usual rank condition, FE.2, but we also need more than just strict exogeneity of $\{\mathbf{x}_{it}\}$ conditional on $(c_i, \mathbf{b}_i)$. Write $\mathbf{b}_i = \boldsymbol{\beta} + \mathbf{d}_i$ where the unit-specific deviation from the average, $\mathbf{d}_i$, necessarily has a zero mean. Then

$$y_{it} = c_i + \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{x}_{it}\mathbf{d}_i + u_{it} \equiv c_i + \mathbf{x}_{it}\boldsymbol{\beta} + v_{it}$$

where $v_{it} \equiv \mathbf{x}_{it}\mathbf{d}_i + u_{it}$. A sufficient condition for consistency of the FE estimator (along with Assumption FE.2) is

$$E(\ddot{\mathbf{x}}_{it}'\ddot{v}_{it}) = \mathbf{0}, t = 1, \ldots, T.$$

But $\ddot{v}_{it} = \ddot{\mathbf{x}}_{it}\mathbf{d}_i + \ddot{u}_{it}$ and $E(\ddot{\mathbf{x}}_{it}'\ddot{u}_{it}) = \mathbf{0}$. The extra assumption is

$$E(\ddot{\mathbf{x}}_{it}'\ddot{\mathbf{x}}_{it}\mathbf{d}_i) = \mathbf{0}, \text{ all } t.$$

A sufficient condition, and one that is easier to interpret, is

$$E(\mathbf{b}_i|\ddot{\mathbf{x}}_{it}) = E(\mathbf{b}_i) = \boldsymbol{\beta}, \quad t = 1, \ldots, T.$$

• This condition allows the slopes, $\mathbf{b}_i$, to be correlated with the regressors $\mathbf{x}_{it}$ through permanent components. What it rules out is correlation between idiosyncratic movements in $\mathbf{x}_{it}$.

- For example, suppose $\mathbf{x}_{it} = \mathbf{f}_i + \mathbf{r}_{it}, t = 1,\ldots,T$ where $\mathbf{f}_i$ is the unit-specific "level" of the process and $\{\mathbf{r}_{it}\}$ are the deviations from this level. Because $\ddot{\mathbf{x}}_{it} = \ddot{\mathbf{r}}_{it}$ it suffices that $E(\mathbf{b}_i|\mathbf{r}_{i1},\mathbf{r}_{i2},\ldots,\mathbf{r}_{iT}) = E(\mathbf{b}_i)$. Note that any kind of serial correlation and changing variances/covariances are allowed in $\{\mathbf{r}_{it}\}$.

• Extension to random trend settings. Write

$$y_{it} = \mathbf{w}_t\mathbf{a}_i + \mathbf{x}_{it}\mathbf{b}_i + u_{it}, \quad t = 1, \ldots, T$$

where $\mathbf{w}_t$ is a set of deterministic functions of time. Now the "fixed effects" estimator sweeps away $\mathbf{a}_i$ by netting out $\mathbf{w}_t$ from $\mathbf{x}_{it}$. In particular, now let $\ddot{\mathbf{x}}_{it}$ denote the residuals from the regression $\mathbf{x}_{it}$ on $\mathbf{w}_t, t = 1, \ldots, T$.

• In the random trend model, $\mathbf{w}_t = (1, t)$, and so the elements of $\mathbf{x}_{it}$ have unit-specific linear trends removed in addition to a level effect.

• Removing even more of the heterogeneity from $\{\mathbf{x}_{it}\}$ makes it even more likely that $E(\mathbf{b}_i|\ddot{\mathbf{x}}_{it}) = E(\mathbf{b}_i)$. For example, if $\mathbf{x}_{it} = \mathbf{f}_i + \mathbf{h}_i t + \mathbf{r}_{it}$, then $\mathbf{b}_i$ can be arbitrarily correlated with $(\mathbf{f}_i, \mathbf{h}_i)$. Of course, adding to $\mathbf{w}_t$ – such as polynomials in t – requires more time periods, and it decreases the variation in $\ddot{\mathbf{x}}_{it}$ compared to the usual FE estimator. Cannot do it at all unless $\dim(\mathbf{w}_t) < T$.

• If we first difference followed by the within transformation, a condition sufficient for consistency of the resulting estimator for $\boldsymbol{\beta}$ is

$$E(\mathbf{b}_i|\Delta\ddot{\mathbf{x}}_{it}) = E(\mathbf{b}_i), \ \ t = 2, \ldots, T,$$

where $\Delta\ddot{\mathbf{x}}_{it} = \Delta\mathbf{x}_{it} - \overline{\Delta\mathbf{x}_i}$ are the demeaned first differences.

• Can apply to models with time-varying "factor loads":

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

Because $\ddot{\eta}_t$ is just a function of the parameters $(\eta_1,\ldots,\eta_T)$, a sufficient condition for the usual FE estimator (that ignores the η_t) is

$$E[(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' c_i] = \mathbf{0}.$$

• Often want to estimate the η_t, with $\eta_1 = 1$ as a normalization. Requires nonlinear GMM.

• Similar results hold for FEIV methods. In $y_{it} = \mathbf{w}_t \mathbf{a}_i + \mathbf{x}_{it} \mathbf{b}_i + u_{it}$, let $\ddot{\mathbf{z}}_{it}$ be the unit-specific detrended IVs ($\mathbf{z}_{it}$ on $\mathbf{w}_t$, $t = 1, \ldots, T$), and assume

$$E(\mathbf{b}_i | \ddot{\mathbf{z}}_{it}) = E(\mathbf{b}_i) = \boldsymbol{\beta}, \quad t = 1, \ldots, T.$$

This turns out not to be enough for the FEIV estimator to identify the APE. An additional sufficient condition is

$$Cov(\ddot{\mathbf{x}}_{it}, \mathbf{b}_i | \ddot{\mathbf{z}}_{it}) = Cov(\ddot{\mathbf{x}}_{it}, \mathbf{b}_i), t = 1, \ldots, T.$$

• Allows $Cov(\ddot{\mathbf{x}}_{it}, \mathbf{b}_i)$, a $K \times K$ matrix, to be any function of t, but not $\ddot{\mathbf{z}}_{it}$.

• Why is this condition sufficient? From $E(\mathbf{b}_i|\ddot{\mathbf{z}}_{it}) = E(\mathbf{b}_i)$, $Cov(\ddot{\mathbf{x}}_{it}, \mathbf{d}_i|\ddot{\mathbf{z}}_{it}) = E(\ddot{\mathbf{x}}_{it}\mathbf{d}_i'|\ddot{\mathbf{z}}_{it})$, and so $E(\ddot{\mathbf{x}}_{it}\mathbf{d}_i|\ddot{\mathbf{z}}_{it}) = E(\ddot{\mathbf{x}}_{it}\mathbf{d}_i) \equiv \gamma_t$ under the previous assumptions. Write $\ddot{\mathbf{x}}_{it}\mathbf{d}_i = \gamma_t + r_{it}$ where $E(r_{iti}|\ddot{\mathbf{z}}_{it}) = 0, t = 1, \ldots, T$. Write the transformed equation as

$$\ddot{y}_{it} = \ddot{\mathbf{x}}_{it}\boldsymbol{\beta} + \ddot{\mathbf{x}}_{it}\mathbf{d}_i + \ddot{u}_{it} = \ddot{y}_{it} = \ddot{\mathbf{x}}_{it}\boldsymbol{\beta} + \gamma_t + r_{it} + \ddot{u}_{it}.$$

• If $\mathbf{x}_{it}$ contains a full set of time period dummies, then we can absorb γ_t into $\ddot{\mathbf{x}}_{it}$, and we assume that here. Then the sufficient condition for consistency of IV estimators applied to the transformed equations is $E[\ddot{\mathbf{z}}_{it}'(r_{it} + \ddot{u}_{it})] = \mathbf{0}$,.and this condition is met under the maintained assumptions.

- Provides further justification for including a full set of time period dummies, even if we have flexible unit-specific trends in $\mathbf{w}_t$.
- Can also use GMM to obtain a more efficient estimator. If $\mathbf{b}_i$ truly depends on i, then the composite error $r_{it} + \ddot{u}_{it}$ is likely serially correlated and heteroskedastic.
- $Cov(\ddot{\mathbf{x}}_{it}, \mathbf{b}_i | \ddot{\mathbf{z}}_{it}) = Cov(\ddot{\mathbf{x}}_{it}, \mathbf{b}_i)$ cannot really be expected to hold for discrete endogenous $\mathbf{x}_{it}$.

4. TESTING FOR CORRELATED RANDOM SLOPES

- Recall in the basic model

$$y_{it} = c_i + \mathbf{x}_{it}\boldsymbol{\beta} + u_{it},$$

the standard test for the presence of c_i is based on an estimate of σ_c^2. In the standard setup, all serial correlation in $\{v_{it} = c_i + u_{it}\}$ is assumed to come from c_i. Can and should allow for serial correlation in $\{u_{it}\}$, but it is rarely done.

- Now we want to test for the presence of $\mathbf{b}_i$ in

$$y_{it} = c_i + \mathbf{x}_{it}\mathbf{b}_i + u_{it}.$$

First, consider the null hypothesis

$$H_0 : Var(\mathbf{b}_i) = \mathbf{0}.$$

Under this null, must have $E(\mathbf{b}_i|\mathbf{x}_i) = E(\mathbf{b}_i)$. For now, maintain

$$Var(\mathbf{b}_i|\mathbf{x}_i) = Var(\mathbf{b}_i) \equiv \mathbf{\Lambda},$$

• We still do not have enough to proceed. We need to restrict the conditional variance matrix of the idiosyncratic errors, and the simplest (and most common) assumption is

$$Var(\mathbf{u}_i|\mathbf{x}_i, c_i, \mathbf{b}_i) = \sigma_u^2 \mathbf{I}_T.$$

• Write the time-demeaned equation as $\ddot{\mathbf{y}}_i = \ddot{\mathbf{X}}_i \boldsymbol{\beta} + \ddot{\mathbf{v}}_i$, where

$$E(\ddot{\mathbf{v}}_i|\mathbf{x}_i) = \ddot{\mathbf{X}}_i E(\mathbf{d}_i|\mathbf{x}_i) + E(\ddot{\mathbf{u}}_i|\mathbf{x}_i) = \mathbf{0},$$

$$Var(\ddot{\mathbf{v}}_i|\mathbf{x}_i) = \ddot{\mathbf{X}}_i \boldsymbol{\Lambda} \ddot{\mathbf{X}}_i' + \sigma_u^2 \mathbf{M}_T,$$

and $\mathbf{M}_T = \mathbf{I}_T - \mathbf{j}_T(\mathbf{j}_T'\mathbf{j}_T)^{-1}\mathbf{j}_T$.

• So $Var(\ddot{\mathbf{v}}_i|\mathbf{x}_i)$ does not depend on $\ddot{\mathbf{X}}_i$ if $\mathbf{\Lambda} = \mathbf{0}$. If $\mathbf{\Lambda} \neq \mathbf{0}$, then the composite error in the time-demeaned equation generally exhibits heteroskedasticity and serial correlation that are quadratic functions of the time-demeaned regressors. So, the method would be to estimate $\boldsymbol{\beta}$ by standard fixed effects, obtain the FE residuals, and then test whether the variance matrix is a quadratic function of the $\ddot{\mathbf{x}}_{it}$.

• Problem with this test is that it associates system heteroskedasticity – that is, variances and covariances depending on the regressors – with the presence of “random” coefficients. But if $\mathbf{b}_i = \boldsymbol{\beta}$ and $Var(\mathbf{u}_i|\mathbf{x}_i, c_i)$ depends on $\mathbf{x}_i$, $Var(\ddot{\mathbf{v}}_i|\mathbf{x}_i)$ generally depends on $\mathbf{x}_i$.

• Rather than try to test whether $Var(\mathbf{b}_i) \neq \mathbf{0}$, we can instead test whether $\mathbf{b}_i$ varies with observable variables. That is, we can test

$$H_0 : E(\mathbf{b}_i|\mathbf{x}_i) = E(\mathbf{b}_i).$$

A sensible alternative is that $E(\mathbf{b}_i|\mathbf{x}_i)$ depends with the time averages, something we can capture with

$$\mathbf{b}_i = \boldsymbol{\beta} + \boldsymbol{\Gamma}(\bar{\mathbf{x}}_i - \boldsymbol{\mu}_{\bar{\mathbf{x}}})' + \mathbf{d}_i.$$

The null hypothesis is $H_0 : \boldsymbol{\Gamma} = \mathbf{0}$.

- Explicitly allowing for aggregate time effects gives, after algebra,

$$\begin{aligned} y_{it} &= \theta_t + c_i + \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{x}_{it}\boldsymbol{\Gamma}(\bar{\mathbf{x}}_i - \boldsymbol{\mu}_{\bar{\mathbf{x}}})' + \mathbf{x}_{it}\mathbf{d}_i + u_{it} \\ &= \theta_t + c_i + \mathbf{x}_{it}\boldsymbol{\beta} + [(\bar{\mathbf{x}}_i - \boldsymbol{\mu}_{\bar{\mathbf{x}}}) \otimes \mathbf{x}_{it}]vec(\boldsymbol{\Gamma}) + \mathbf{x}_{it}\mathbf{d}_i + u_{it} \\ &\equiv \theta_t + c_i + \mathbf{x}_{it}\boldsymbol{\beta} + [(\bar{\mathbf{x}}_i - \boldsymbol{\mu}_{\bar{\mathbf{x}}}) \otimes \mathbf{x}_{it}]\boldsymbol{\gamma} + v_{it}, \end{aligned}$$

where $v_{it} = \mathbf{x}_{it}\mathbf{d}_i + u_{it}$ and $\boldsymbol{\gamma} = vec(\boldsymbol{\Gamma})$.

- The test of $H_0 : \boldsymbol{\gamma} = \mathbf{0}$ is simple to carry out. Create interactions

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

where $\bar{\mathbf{x}}$ is the vector of overall averages, $\bar{\mathbf{x}} = N^{-1} \sum_{i=1}^{N} \bar{\mathbf{x}}_i$. (Or, one can choose a subset of $\bar{\mathbf{x}}_i$ to interact with a subset of $\mathbf{x}_{it}$. Obtain a fully robust test of joint significance in the context of fixed effects estimation.

- A failure to reject means that, if the $\mathbf{b}_i$ vary by i, they apparently do not do so in a way that depends on the time averages of the covariates.
- Test cannot detect heterogeneity in $\mathbf{b}_i$ that is uncorrelated with $\bar{\mathbf{x}}_i$. (Like the previous test, this test is not intended to determine whether FE is consistent for $\boldsymbol{\beta} = E(\mathbf{b}_i)$.)
- If we reject the null, can use the expanded equation as a way to model random slopes.

• We can also allow $\mathbf{b}_i$ to depend on time constant observable variables:

$$\mathbf{b}_i = \boldsymbol{\beta} + \boldsymbol{\Gamma}(\bar{\mathbf{x}}_i - \boldsymbol{\mu}_{\bar{\mathbf{x}}})' + \boldsymbol{\Psi}(\mathbf{h}_i - \boldsymbol{\mu}_{\mathbf{h}})' + \mathbf{d}_i$$

where $\mathbf{h}_i$ is a row vector of time-constant variables that we think might influence $\mathbf{b}_i$.

• The new equation is

$$y_{it} = \theta_t + c_i + \mathbf{x}_{it}\boldsymbol{\beta} + [(\bar{\mathbf{x}}_i - \bar{\mathbf{x}}) \otimes \mathbf{x}_{it}]\boldsymbol{\gamma} + [(\mathbf{h}_i - \bar{\mathbf{h}}) \otimes \mathbf{x}_{it}]\boldsymbol{\psi} + error_{it}$$

and we can test $H_0 : \boldsymbol{\gamma} = \mathbf{0}, \boldsymbol{\psi} = \mathbf{0}$ or a subset, using the fixed effects estimator (to remove c_i).

• Alternatively, add $\bar{\mathbf{x}}_i$ and $\mathbf{h}_i$ as separate regressors (to allow correlation with c_i) and estimate the model by random effects. This is common in the hierarchical linear models literature.

• In other words, estimate he following equation in an RE framework:

$$y_{it} = \theta_t + \mathbf{x}_{it}\boldsymbol{\beta} + \bar{\mathbf{x}}_i\boldsymbol{\xi} + \mathbf{h}_i\boldsymbol{\zeta} + [(\bar{\mathbf{x}}_i - \bar{\mathbf{x}}) \otimes \mathbf{x}_{it}]\boldsymbol{\gamma} + [(\mathbf{h}_i - \bar{\mathbf{h}}) \otimes \mathbf{x}_{it}]\boldsymbol{\psi} + error_{it}$$

EXAMPLE: Effects of Concentration on Airfares

```
. egen concenb = mean(concen), by(id)

. sum concenb ldist ldistsq

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     concenb |      4596    .6101149    .1888741      .1862      .9997
       ldist |      4596    6.696482    .6593177   4.553877   7.909857
     ldistsq |      4596    45.27747    8.726898   20.73779   62.56583

. gen cbconcen = (concenb - .61)*concen
. gen ldconcen = (ldist - 6.696)*concen
. gen ldsqconcen = (ldistsq - 45.277)*concen
```

```
. xtreg lfare concen concenb cbconcen ldconcen ldsqconcen ldist ldistsq
       y98 y99 y00, re cluster(id)

                                  (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .1682492   .0496695     3.39   0.001     .0708988    .2655996
     concenb |    .157291   .2085049     0.75   0.451    -.2513711     .565953
    cbconcen |   .0635453   .3033809     0.21   0.834    -.5310704    .6581609
    ldconcen |  -.2994869   .9930725    -0.30   0.763    -2.245873    1.646899
  ldsqconcen |   .0112477   .0746874     0.15   0.880     -.135137    .1576324
       ldist |  -.4394368   .6713288    -0.65   0.513    -1.755217    .8763435
     ldistsq |   .0752147   .0494201     1.52   0.128    -.0216469    .1720764
         y98 |   .0229684   .0041542     5.53   0.000     .0148262    .0311105
         y99 |   .0358549   .0051298     6.99   0.000     .0258007    .0459091
         y00 |   .0976256    .005461    17.88   0.000     .0869221     .108329
       _cons |   4.382552   2.272566     1.93   0.054    -.0715953    8.836699
-------------+----------------------------------------------------------------
```

```
. test concenb cbconcen ldconcen ldsqconcen

 ( 1)  concenb = 0
 ( 2)  cbconcen = 0
 ( 3)  ldconcen = 0
 ( 4)  ldsqconcen = 0

           chi2(  4) =   14.02
         Prob > chi2 =    0.0072

. * If we test only the interactions, they are jointly insignificant:

. test cbconcen ldconcen ldsqconcen

 ( 1)  cbconcen = 0
 ( 2)  ldconcen = 0
 ( 3)  ldsqconcen = 0

           chi2(  3) =    5.47
         Prob > chi2 =    0.1407
```

```
. * Estimated coefficient on concen very close to omitting interations:

. xtreg lfare concen concenb ldist ldistsq y98 y99 y00, re cluster(id)

Random-effects GLS regression                   Number of obs      =      4596
Group variable: id                              Number of groups   =      1149

R-sq:  within  = 0.1352                         Obs per group: min =         4
       between = 0.4216                                        avg =       4.0
       overall = 0.4068                                        max =         4

Random effects u_i ~Gaussian                   Wald chi2(7)       =   1273.17
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

                                  (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |    .168859   .0494749     3.41   0.001       .07189    .2658279
     concenb |   .2136346   .0816403     2.62   0.009     .0536227    .3736466
       ldist |  -.9089297   .2721637    -3.34   0.001    -1.442361   -.3754987
     ldistsq |   .1038426   .0201911     5.14   0.000     .0642688    .1434164
         y98 |   .0228328   .0041643     5.48   0.000     .0146708    .0309947
         y99 |   .0363819   .0051292     7.09   0.000     .0263289    .0464349
         y00 |   .0977717   .0055072    17.75   0.000     .0869777    .1085656
       _cons |   6.207889   .9118109     6.81   0.000     4.420773    7.995006
-------------+----------------------------------------------------------------
```

• We can take the representation for $\mathbf{b}_i$ seriously and base a test for *unobserved* heterogeneity on

$$H_0 : Var(\mathbf{d}_i) = \mathbf{0}$$

using FE estimation with the interactions.

- Remember that $Var(\mathbf{b}_i)$ and $Var(\mathbf{d}_i)$ are not the same.
- Like the earlier test based on $Var(\mathbf{b}_i)$ under the assumption $E(\mathbf{b}_i|\mathbf{x}_i) = E(\mathbf{b}_i)$, a test based on $Var(\mathbf{d}_i) = \mathbf{0}$ relies on a particular structure for $Var(\mathbf{u}_i|\mathbf{x}_i, c_i, \mathbf{b}_i)$.
- As a testing strategy, it makes sense to first see if $\mathbf{b}_i$ depends on $(\bar{\mathbf{x}}_i, \mathbf{h}_i)$, and then to test $Var(\mathbf{d}_i) = \mathbf{0}$ (even though the latter imposes strong assumptions).