UNOBSERVED EFFECTS LINEAR PANEL DATA MODELS, I

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

1. Introduction
2. Assumptions
3. Estimation and Testing
4. Comparison of Estimators

1. INTRODUCTION

- We already covered panel data models where the error term had no particular structure. But we assumed either contemporaneous exogeneity (pooled OLS) or strict exogeneity (feasible GLS).
- Now we explicitly add a time constant, *unobserved effect* to the model. Often called *unobserved heterogeneity*.

- Start with the balanced panel case, and assume random sampling across i (the cross section dimension), with fixed time periods T. So $\{(\mathbf{x}_{it}, y_{it}) : t = 1,\ldots,T, c_i\}$ where c_i is the unobserved effect drawn along with the observed data.
- The unbalanced case is trickier because we must know why we are missing some time periods for some units. We consider this much later under missing data/sample selection issues.

• For a random draw i from the population, the basic model is

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

where $\{u_{it} : t = 1,\ldots,T\}$ are the *idiosyncratic errors*. The *composite error* at time t is

$$v_{it} = c_i + u_{it}$$

• Because of c_i, the sequence $\{v_{it} : t = 1,\ldots,T\}$ is almost certainly serially correlated, and definitely is if $\{u_{it}\}$ is serially uncorrelated.

• Useful to write a population version of the model in conditional expectation form:

$$E(y_t|\mathbf{x}_t, c) = \mathbf{x}_t\boldsymbol{\beta} + c,\ t = 1, \ldots, T.$$

Therefore,

$$\beta_j = \frac{\partial E(y_t|\mathbf{x}_t, c)}{\partial x_{tj}},$$

so that β_j is the partial effect of x_{tj} on $E(y_t|\mathbf{x}_t, c)$, so that we are "holding c fixed."

• Hope is that we can allow c to be correlated with $\mathbf{x}_t$.

- With a single cross section, there is nothing we can do unless we can find good observable proxies for c or IVs for the endogenous elements of $\mathbf{x}_t$. But with two or more periods we have more options.
- We can write the population model as

$$
\begin{aligned}
y_t &= \mathbf{x}_t\boldsymbol{\beta} + c + u_t \\
E(u_t|\mathbf{x}_t, c) &= 0
\end{aligned}
$$

Suppose we have $T = 2$ time periods:

$$
\begin{aligned}
y_1 &= \mathbf{x}_1\boldsymbol{\beta} + c + u_1 \\
y_2 &= \mathbf{x}_2\boldsymbol{\beta} + c + u_2
\end{aligned}
$$

- Subtract $t = 1$ from $t = 2$ and define $\Delta y = y_2 - y_1$, $\Delta \mathbf{x} = \mathbf{x}_2 - \mathbf{x}_1$, and $\Delta u = u_2 - u_1$:

$$\Delta y = \Delta \mathbf{x} \boldsymbol{\beta} + \Delta u,$$

which is now a cross section in the changes or differences.

- Sufficient for OLS on a random sample to consistently estimate $\boldsymbol{\beta}$:

$$E(\Delta \mathbf{x}' \Delta u) = \mathbf{0}$$
$$rank\, E(\Delta \mathbf{x}' \Delta \mathbf{x}) = K.$$

- The rank condition is violated if $\mathbf{x}_t$ has elements that do not change over time. Assume each element of $\mathbf{x}_t$ has some time variation (that is, for at least some members in the population).

- The orthogonality condition is

$$E[(\mathbf{x}_2 - \mathbf{x}_1)'(u_2 - u_1)] = \mathbf{0}.$$

But

$$\begin{aligned} E[(\mathbf{x}_2 - \mathbf{x}_1)'(u_2 - u_1)] &= E(\mathbf{x}_2'u_2) - E(\mathbf{x}_1'u_2) - E(\mathbf{x}_2'u_1) + E(\mathbf{x}_1'u_1) \\ &= -[E(\mathbf{x}_1'u_2) + E(\mathbf{x}_2'u_1)] \end{aligned}$$

because $E(\mathbf{x}_t'u_t) = \mathbf{0}$ under the conditional mean specification.

- OLS on the differences will only be consistent if we add

$$E(\mathbf{x}_s' u_t) = 0, s \neq t.$$

This is a kind of strict exogeneity assumption. However, we have removed c from the composite error. Assuming $\mathbf{x}_s$ is uncorrelated with u_t for all s and t is weaker than assuming $\mathbf{x}_s$ is uncorrelated with the composite error, $c + u_t$, for all s and t.

• Would we really omit an intercept from the differenced equation? Very unlikely. If we start with a model with different intercepts,

$$y_1 = \theta_1 + \mathbf{x}_1\boldsymbol{\beta} + c + u_1$$
$$y_2 = \theta_2 + \mathbf{x}_2\boldsymbol{\beta} + c + u_2$$

then

$$\Delta y = \alpha + \Delta\mathbf{x}\boldsymbol{\beta} + \Delta u,$$

where $\alpha = \theta_2 - \theta_1$ is the change in the aggregate time effects (intercepts). Now the rank condition also excludes variables that change by the same amount for each unit (such as age).

2. ASSUMPTIONS

- As mentioned earlier, we assume a balanced panel and all asymptotic analysis – implicit or explicit – is with fixed T and $N \to \infty$, where N is the size of the cross section.

- The basic *unobserved effects model* is

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

where $\mathbf{x}_{it}$ is $1 \times K$ and so $\boldsymbol{\beta}$ is $K \times 1$. In addition to unobserved effect and unobserved heterogeneity, c_i is sometimes called a *latent effect* or an individual effect, firm effect, school effect, and so on.

- An extension of the basic model is

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

where $\{\eta_t : t = 1, \ldots, T\}$ are unknown parameters (and we have to assume something like $\eta_1 = 1$). More on this later.

- As in the earlier treatment, the model is written with $\boldsymbol{\beta}$ not depending on time. But $\mathbf{x}_{it}$ can include time period dummies and interactions of variables with time periods dummies, so the model is quite flexible.

- A general specification is

$$y_{it} = \mathbf{g}_t\boldsymbol{\theta} + \mathbf{z}_i\boldsymbol{\delta} + \mathbf{w}_{it}\boldsymbol{\gamma} + c_i + u_{it}$$

where $\mathbf{g}_t$ is a vector of aggregate time effects (often time dummies), $\mathbf{z}_i$ is a set of time-constant observed variables, and $\mathbf{w}_{it}$ changes across i and t (for at least some units i and time periods t). $\mathbf{w}_{it}$ can include nteractions among time-constant and time varying variables.

- In microeconometric applications, best to avoid calling c_i a "random effect" or a "fixed effect." We are treating c_i always as a random variable.

Assumptions about the Unobserved Effect

• In modern applications, "random effect" essentially means

$$Cov(\mathbf{x}_{it}, c_i) = \mathbf{0}, t = 1, \ldots, T,$$

although we often will strengthen this.

• The term "fixed effect" means that no restrictions are placed on the relationship between c_i and $\{\mathbf{x}_{it}\}$.

• Recently, "correlated random effects" is used to denote situations where we model the relationship between c_i and $\{\mathbf{x}_{it}\}$, and it is especially useful for nonlinear models (but also for linear models, as we will see).

Exogeneity Assumptions on the Explanatory Variables

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

Contemporaneous Exogeneity Conditional on the Unobserved Effect:

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

or

$$E(y_{it}|\mathbf{x}_{it}, c_i) = \mathbf{x}_{it}\boldsymbol{\beta} + c_i.$$

- Ideally, we could proceed with just this assumption.

• *Strict Exogeneity Conditional on the Unobserved Effect:*

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

so that only $\mathbf{x}_{it}$ affects the expected value of y_{it} once c_i is controlled for.

• This is weaker than if we did not condition on c_i. Assuming the condition holds conditional on c_i,

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

So correlation between c_i and $(\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT})$ would invalidate the assumption without conditioning on c_i.

• But strict exogeneity conditional on c_i rules out lagged dependent variables and feedback. Written in terms of the idiosyncratic errors, strict exogeneity is

$$E(u_{it}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT},c_i) = 0,$$

and so $\mathbf{x}_{i,t+h}$ must be uncorrelated with u_{it} for all $h > 0$.

• In addition to ruling out feedback, strict exogeneity assumes we have any distributed lag dynamics correct, too. For example, if $\mathbf{x}_{it} = (\mathbf{z}_{it},\mathbf{z}_{i,t-1})$, then

$$E(y_{it}|\mathbf{z}_{i1},\ldots,\mathbf{z}_{it},\ldots,\mathbf{z}_{iT},c_i) = E(y_{it}|\mathbf{z}_{it},\mathbf{z}_{i,t-1},c_i).$$

• A more reasonable assumption that we will use later is

$$E(y_{it}|\mathbf{x}_{it}, \mathbf{x}_{i,t-1}, \ldots, \mathbf{x}_{i1}, c_i) = E(y_{it}|\mathbf{x}_{it}, c_i) = \mathbf{x}_{it}\boldsymbol{\beta} + c_i,$$

which is *sequential exogeneity conditional on the unobserved effect*.

• Sequential exogeneity assumes correct distributed lag dynamics but is silent on feedback.

3. ESTIMATION AND TESTING

- There are four common methods: pooled OLS, random effects, fixed effects, and first differencing.

3.1. Pooled OLS

- We already covered this. Now, we just recognize that the equation is

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

- Consistency (fixed T, $N \to \infty$) of the POLS estimator is ensured by

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

- Contemporaneous exogeneity is weaker than strict exogeneity, but it buys us little in practice because POLS also uses $E(\mathbf{x}_{it}'c_i) = \mathbf{0}$, which cannot hold for lagged dependent variables and is unlikely for other variables not strictly exogenous.
- Inference should be made robust to serial correlation and heteroskedasticity.

- Let $\hat{v}_{it} = y_{it} - \mathbf{x}_{it}\hat{\boldsymbol{\beta}}_{POLS}$ be the POLS residuals. Then

$$\widehat{Avar}(\hat{\boldsymbol{\beta}}_{POLS}) = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}'\mathbf{x}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=1}^{T}\hat{v}_{it}\hat{v}_{ir}\mathbf{x}_{it}'\mathbf{x}_{ir}\right)$$
$$\cdot\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{x}_{it}'\mathbf{x}_{it}\right)^{-1},$$

or sometimes with an adjustment, such as multiply by $N/(N-1)$.

- Can also write this estimator as

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

- In Stata:

```
reg y x1 x2 ... xK, cluster(id)
```

3.2. Random Effects Estimation

- State assumptions in conditional mean terms so that second moment derivations are are easier.

ASSUMPTION RE.1:

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

$$\text{(b)}\ \ E(c_i|\mathbf{x}_{i1},\mathbf{x}_{i2},\ldots,\mathbf{x}_{iT}) = E(c_i)$$

- Assume $\mathbf{x}_{it}$ includes (at least) unity, and probably time dummies in addition. Then $E(c_i) = 0$ is without loss of generality.

• A GLS approach also leaves c_i in the error term:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + v_{it},\ t = 1, 2, \ldots, T$$

and we know the properties of feasible GLS when

$$E(\mathbf{x}_{is}' v_{it}) = 0, \text{ all } s, t = 1, \ldots, T.$$

• This weaker version of strict exogeneity is implied by Assumption RE.1.

- Write the equation in system form (for all time periods) as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{v}_i = \mathbf{X}_i\boldsymbol{\beta} + c_i\mathbf{j}_T + \mathbf{u}_i$$

where $\mathbf{j}_T' = (1, 1, \ldots, 1)$.

- Define

$$\underset{T\times T}{\boldsymbol{\Omega}} = E(\mathbf{v}_i\mathbf{v}_i') = Var(\mathbf{v}_i).$$

ASSUMPTION RE.2: $\boldsymbol{\Omega}$ is nonsingular and *rank* $E(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i)$.

• RE imposes a special structure on $\mathbf{\Omega}$ (which could be wrong!). Under RE.1(a), c_i and u_{it} are uncorrelated. Assume further that

$$\begin{aligned} Var(u_{it}) &= \sigma_u^2, t = 1, \ldots, T \\ Cov(u_{it}, u_{is}) &= 0, t \neq s \end{aligned}$$

Then

$$\begin{aligned} Var(v_{it}) &= Var(c_i + u_{it}) = Var(c_i) + Var(u_{it}) \\ \sigma_v^2 &= \sigma_c^2 + \sigma_u^2 \end{aligned}$$

- Further, for $t \neq s$,

$$
\begin{aligned}
Cov(v_{it}, v_{is}) &= Cov(c_i + u_{it}, c_i + u_{is}) \\
&= Var(c_i) + Cov(c_i, u_{is}) + Cov(u_{it}, c_i) + Cov(u_{it}, u_{is}) \\
&= \sigma_c^2
\end{aligned}
$$

- This leads to the “random effects” or “exchangeable” structure for $\mathbf{\Omega}$:

$$
\begin{aligned}
\mathbf{\Omega} &= E[(c_i \mathbf{j}_T + \mathbf{u}_i)(c_i \mathbf{j}_T + \mathbf{u}_i)'] = E(c_i^2)\mathbf{j}_T\mathbf{j}_T' + E(\mathbf{u}_i\mathbf{u}_i') \\
&= \sigma_c^2 \mathbf{j}_T \mathbf{j}_T' + \sigma_u^2 \mathbf{I}_T
\end{aligned}
$$

or

$$\mathbf{\Omega} = \begin{pmatrix} \sigma_c^2 + \sigma_u^2 & \cdots & \sigma_c^2 & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2 + \sigma_u^2 & & \sigma_c^2 \\ \vdots & & \ddots & \vdots \\ \sigma_c^2 & \cdots & \sigma_c^2 & \sigma_c^2 + \sigma_u^2 \end{pmatrix},$$

so the $T \times T$ matrix depends on only two parameters, σ_c^2 and σ_u^2 or, more directly, σ_v^2 and σ_c^2.

- Feasible GLS requires estimating $\mathbf{\Omega}$, that is, the two parameters.
- Actually, it would be enough to know $\rho = \sigma_c^2/(\sigma_c^2 + \sigma_u^2)$, the fraction of the total variance accounted for by c_i. Notice that $\rho = Corr(v_{it}, v_{is})$ for all $t \neq s$.

- We can also write $\mathbf{\Omega}$ as

$$\mathbf{\Omega} = \sigma_v^2 \begin{pmatrix} 1 & \cdots & \rho & \rho \\ \rho & 1 & & \rho \\ \vdots & & \ddots & \vdots \\ \rho & \cdots & \rho & 1 \end{pmatrix}$$

which shows we only need to estimate ρ to proceed with FGLS.

- Typically, we estimate σ_v^2 and σ_c^2, but ρ is useful for summarizing the importance of c_i.

• We can use pooled OLS to get the residuals, $\check{v}_{it}$, across all i and t. Then a consistent estimator of σ_v^2 (not generally unbiased), as N gets large for fixed T, is

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

the usual variance estimator from OLS regression. This is based on, for each i, $\sigma_v^2 = T^{-1} \sum_{t=1}^{T} E(v_{it}^2)$ and then average across i, too. Then replace population with sample average, and $\boldsymbol{\beta}$ with pooled OLS estimates, and subtract K as a degrees-of-freedom adjustment.

- For σ_c^2, note that

$$\sigma_c^2 = [T(T-1)/2]^{-1} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} E(v_{it} v_{is}).$$

So a consistent "estimator" would be

$$\tilde{\sigma}_c^2 = [NT(T-1)/2]^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} v_{it} v_{is}.$$

- An actual estimator replaces v_{it} with the POLS residuals,

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

and subtracts K from $NT(T-1)/2$ as a df adjustment. By the usual argument,

$$\underset{N\to\infty}{\text{plim}}\ \hat{\sigma}_c^2 = \sigma_c^2$$

with T fixed.

- Now we can use

$$\hat{\mathbf{\Omega}} = \begin{pmatrix} \hat{\sigma}_v^2 & \cdots & \hat{\sigma}_c^2 & \hat{\sigma}_c^2 \\ \hat{\sigma}_c^2 & \hat{\sigma}_v^2 & & \hat{\sigma}_c^2 \\ \vdots & & \ddots & \vdots \\ \hat{\sigma}_c^2 & \cdots & \hat{\sigma}_c^2 & \hat{\sigma}_v^2 \end{pmatrix} \text{ or } \hat{\mathbf{\Lambda}} = \begin{pmatrix} 1 & \cdots & \hat{\rho} & \hat{\rho} \\ \hat{\rho} & 1 & & \hat{\rho} \\ \vdots & & \ddots & \vdots \\ \hat{\rho} & \cdots & \hat{\rho} & 1 \end{pmatrix}$$

where $\hat{\rho} = \hat{\sigma}_c^2/\hat{\sigma}_v^2$ in FGLS.

• It is possible for $\hat{\sigma}_c^2$ to be negative, which means the basic unobserved effects variance-covariance structure is faulty.

• Typically, $\hat{\sigma}_c^2 > 0$ unless the variables have been transformed in some way – such as being first differenced – before applying GLS.

• The FGLS estimator that uses this particular structure of $\hat{\mathbf{\Omega}}$ is the *random effects (RE) estimator*.

• Fully robust inference is available for RE, and there are good reasons for doing so.

(1) $\mathbf{\Omega}$ may not have the special (and restrictive, especially for large T) RE structure, that is, $E(\mathbf{v}_i\mathbf{v}_i)$ need not have the RE form. Serial correlation or changing variances in $\{u_{it} : t = 1,\ldots,T\}$ invalidate the RE structure.

(2) The system homoskedasticity requirement,

$$E(\mathbf{v}_i\mathbf{v}_i|\mathbf{X}_i) = E(\mathbf{v}_i\mathbf{v}_i)$$

might not hold.

• A fully robust estimator is

$$\widehat{Avar}(\hat{\boldsymbol{\beta}}_{RE}) = \left(\sum_{i=1}^{N} \mathbf{X}_i'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X}_i\right)^{-1}\left(\sum_{i=1}^{N} \mathbf{X}_i'\hat{\boldsymbol{\Omega}}^{-1}\hat{\mathbf{v}}_i\hat{\mathbf{v}}_i'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X}_i\right)^{-1} \cdot \left(\sum_{i=1}^{N} \mathbf{X}_i'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X}_i\right)^{-1},$$

where $\hat{\mathbf{v}}_i = \mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}_{RE}$ is the vector of RE (FGLS) residuals.

• Sometimes, an iterative procedure is used. These new residuals can be used to obtain a new estimate of $\boldsymbol{\Omega}$, and so on.

- For first order asymptotics, no efficiency gain from iterating. Might help with smaller N, though.
- What is the advantage of RE, which imposes specific assumptions on $\mathbf{\Omega}$, and the unrestricted FGLS we discussed earlier? Theoretically, nothing. We do not get more efficiency with large N and small T by imposing restrictions on $\mathbf{\Omega}$.
- If system homoskedasticity holds but $\mathbf{\Omega}$ is not of the RE form, an unrestricted FGLS analysis is more efficient than RE (again, fixed T, $N \rightarrow \infty$).
- As we will see later, RE does have some appeal because of its implicit transformation.

• A nonrobust variance matrix estimator can be used if we add an assumption:

ASSUMPTION RE.3:

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

• Under Assumptions RE.1 and RE.3, $\mathbf{\Omega}$ has the RE structure and system homoskedasticity holds. Part (a) is homoskedasticity and serial uncorrelatedness of $\{u_{it}\}$ conditional on $(\mathbf{x}_i, c_i)$, and (b) is homoskedasticity of c_i.

• Under RE.1, RE.2, and RE.3,

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

is a valid estimator.

• Inference is straightforward. Typically use Wald or robust Wald statistic for multiple restrictions.

• In Stata, fully robust inference uses the "cluster" option; for the "usual" variance matrix estimator, drop this option:

```
xtreg y x1 x2 ... xK, re cluster(id)
```

- Occasionally, one might want to test

$$H_0 : \sigma_c^2 = 0$$
$$H_1 : \sigma_c^2 > 0$$

It's rare that one cannot strongly reject this because of the strong positive serial correlation in the POLS residuals in most applications. The formal test, derived under joint normality for $(c_i, \mathbf{u}_i)$, is called the Breusch-Pagan test.

• A fully robust test does not add any additional assumptions, and allows for heteroskedasticity. The key is that if $\hat{v}_{it}$ now denotes the POLS residuals – which is what the B-P test uses – then

$$N^{-1/2} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \hat{v}_{it}\hat{v}_{is} = N^{-1/2} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} v_{it}v_{is} + o_p(1)$$

- Therefore, under

$$H_0 : E(v_{it}v_{is}) = 0, \text{ all } t \neq s,$$

it follows that

$$\frac{N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\hat{v}_{it}\hat{v}_{is}}{\left\{E\left[\left(\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}v_{it}v_{is}\right)^2\right]\right\}^{1/2}} \xrightarrow{d} Normal(0,1)$$

- Now estimate the denominator and cancel the sample sizes:

$$\frac{\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\hat{v}_{it}\hat{v}_{is}}{\left[\sum_{i=1}^{N}\left(\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\hat{v}_{it}\hat{v}_{is}\right)^{2}\right]^{1/2}} \xrightarrow{d} Normal(0,1).$$

- Later, show how to test $\{u_{it}\}$ for serial correlation allowing for c_i, which is more interesting.

3.3 Fixed Effects Estimation

- Unlike POLS and RE, fixed effects estimation removes c_i to form an estimating equation.
- Average the original equation,

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

across t to get a cross-sectional equation:

$$\bar{y}_i = \bar{\mathbf{x}}_i\boldsymbol{\beta} + c_i + \bar{u}_i,$$

where the overbar indicates *time averages*:

$$\bar{y}_i = T^{-1}\sum_{t=1}^{T} y_{it},\ \bar{\mathbf{x}}_i = T^{-1}\sum_{t=1}^{T}\mathbf{x}_{it},\ \bar{u}_i = T^{-1}\sum_{t=1}^{T}u_{it}$$

• The equation $\bar{y}_i = \bar{\mathbf{x}}_i\boldsymbol{\beta} + c_i + \bar{u}_i$ is often called the *between equation* because it relies on variation in the data between cross section observations. The *between estimator* is the OLS estimator from the cross section regression

$$\bar{y}_i \text{ on } \bar{\mathbf{x}}_i,\ i = 1,\ldots,N.$$

[In practice, an intercept is included to account for nonzero $E(c_i)$.]

• The between estimator is inconsistent unless

$$Cov(\bar{\mathbf{x}}_i, c_i) = \mathbf{0},\ Cov(\bar{\mathbf{x}}_i, \bar{u}_i) = \mathbf{0}.$$

- Instead, subtract off the time-averaged equation from the original equation to eliminate c_i:

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\boldsymbol{\beta} + u_{it} - \bar{u}_i,\ t = 1,\ldots,T$$

or

$$\ddot{y}_{it} = \ddot{\mathbf{x}}_{it}\boldsymbol{\beta} + \ddot{u}_{it},\ t = 1,\ldots,T$$

where $\ddot{y}_{it} = y_{it} - \bar{y}_i$ and so on.

- We call this the *time demeaned equation*, and the transformation is *time demeaning*, *fixed effects*, or *within* (time variation within each i is used).

• Key is that c_i is gone from the time demeaned equation. So, we can use pooled OLS:

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

This is the *fixed effects (FE) estimator* or the *within estimator*.

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{FE} &= \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{\mathbf{x}}_{it}'\ddot{\mathbf{x}}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{\mathbf{x}}_{it}'\ddot{y}_{it}\right)\\ &= \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{\mathbf{x}}_{it}'\ddot{\mathbf{x}}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{\mathbf{x}}_{it}'y_{it}\right)\end{aligned}$$

because $\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'(y_{it} - \bar{y}_i) = \sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'y_{it}$.

• What is the weakest orthogonality assumption for consistency? We can just apply the results for POLS, but it is useful to see it directly.

• Write the estimator by substituting $\ddot{y}_{it} = \ddot{\mathbf{x}}_{it}\boldsymbol{\beta} + \ddot{u}_{it}$:

$$\hat{\boldsymbol{\beta}}_{FE} = \boldsymbol{\beta} + \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{\mathbf{x}}_{it}' \ddot{\mathbf{x}}_{it} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{\mathbf{x}}_{it}' \ddot{u}_{it} \right)$$

$$= \boldsymbol{\beta} + \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{\mathbf{x}}_{it}' \ddot{\mathbf{x}}_{it} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{\mathbf{x}}_{it}' u_{it} \right)$$

- By the WLLN as $N \to \infty$ with fixed T, the key moment condition for consistency is

$$\sum_{t=1}^{T} E(\ddot{\mathbf{x}}_{it}' u_{it}) = \sum_{t=1}^{T} E[(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' u_{it}] = \mathbf{0}.$$

- In addition to contemporaneous exogeneity, $E(\mathbf{x}_{it}' u_{it}) = \mathbf{0}$, we need a kind of strict exogeneity:

$$E(\bar{\mathbf{x}}_i' u_{it}) = T^{-1} \sum_{s=1}^{T} E(\mathbf{x}_{is}' u_{it}) = \mathbf{0},\ t = 1, 2, \ldots, T.$$

ASSUMPTION FE.1: Same as RE.1(a), that is,

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

- This implies $E(\mathbf{x}_{is}' u_{it}) = \mathbf{0}$, all $s, t = 1, \ldots, T$, and so $E(\ddot{\mathbf{x}}_{it}' u_{it}) = \mathbf{0}$, $t = 1, \ldots, T$.

- The rank condition is directly from POLS.2:

ASSUMPTION FE.2:

$$rank\left[\sum_{t=1}^{T} E(\ddot{\mathbf{x}}_{it}' \ddot{\mathbf{x}}_{it})\right] = K.$$

• The rank condition rules out elements in $\mathbf{x}_{it}$ that have no time variation for any unit in the population. Such variables get swept away by the within transformation.

• Under FE.1 and FE.2,

$$\hat{\boldsymbol{\beta}}_{FE} \xrightarrow{p} \boldsymbol{\beta} \text{ as } N \rightarrow \infty$$

• The FE estimator works well for large T, too, but showing that requires putting restrictions on the time series process $\{(\mathbf{x}_{it}, y_{it}) : t = 1, 2, \ldots\}$.

- What parameters can we identify with FE? Suppose we start with

$$y_{it} = \theta_1 + \theta_2 d2_t + \ldots + \theta_T dT_t + \mathbf{z}_i\boldsymbol{\gamma}_1 + d2_t\mathbf{z}_i\boldsymbol{\gamma}_2 + \ldots + dT_t\mathbf{z}_i\boldsymbol{\gamma}_T$$
$$+ \mathbf{w}_{it}\boldsymbol{\delta} + c_i + u_{it}$$

- Using FE, we cannot estimate θ_1 or $\boldsymbol{\gamma}_1$, but all other parameters are generally identified.
- FE allows c_i to be arbitrarily correlated with $(\mathbf{z}_i, \mathbf{w}_{it})$, and so we cannot distinguish $\theta_1 + \mathbf{z}_i\boldsymbol{\gamma}_1$ from c_i.

- We can estimate $\theta_2, \ldots, \theta_T$ and $\boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_T$. So we can estimate whether the effect of the time constant variables has changed over time. We cannot estimate the effect in any period t because it is $\boldsymbol{\gamma}_1$ for $t = 1$ and $\boldsymbol{\gamma}_1 + \boldsymbol{\gamma}_t$ for $t = 2, \ldots, T$.

• As another example, suppose w_{it} is a scalar policy variable and $\mathbf{z}_i$ are time-constant characteristics, and the model is

$$
\begin{aligned}
y_{it} &= \theta_1 + \theta_2 d2_t + \ldots + \theta_T dT_t + \mathbf{z}_i \boldsymbol{\gamma}_1 + d2_t \mathbf{z}_i \boldsymbol{\gamma}_2 + \ldots + dT_t \mathbf{z}_i \boldsymbol{\gamma}_T \\
&\quad + \delta w_{it} + w_{it}(\mathbf{z}_i - \boldsymbol{\mu}_{\mathbf{z}})\boldsymbol{\psi} + c_i + u_{it}
\end{aligned}
$$

where $\boldsymbol{\mu}_{\mathbf{z}} = E(\mathbf{z}_i)$.

• We can estimate δ (the average partial effect) as well as $\boldsymbol{\psi}$, which means we can see how the policy effects change with individual characters (and test $H_0 : \boldsymbol{\psi} = \mathbf{0}$). As a practical matter, we would replace the population mean $\boldsymbol{\mu}_{\mathbf{z}}$ with the sample average, $\bar{\mathbf{z}} = N^{-1} \sum_{i=1}^{N} \mathbf{z}_i$.

- We can obtain a variance matrix estimator valid under Assumptions FE.1 and FE.2.
- Define the FE residuals as

$$\widehat{\ddot{u}}_{it} = \ddot{y}_{it} - \ddot{\mathbf{x}}_{it}\hat{\boldsymbol{\beta}}_{FE}, \; t = 1, \ldots, T; i = 1, \ldots, N$$

- These are "estimates" of the $\ddot{u}_{it}$, not the u_{it}. This has implications for estimating the error variance, σ_u^2.

- Without additional assumptions, use the "cluster-robust" matrix

$$\widehat{Avar}(\hat{\boldsymbol{\beta}}_{FE}) = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{\mathbf{x}}_{it}'\ddot{\mathbf{x}}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{r=1}^{T}\hat{\ddot{u}}_{it}\hat{\ddot{u}}_{ir}\ddot{\mathbf{x}}_{it}'\ddot{\mathbf{x}}_{ir}\right)$$

$$\cdot\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{\mathbf{x}}_{it}'\ddot{\mathbf{x}}_{it}\right)^{-1}.$$

• In Stata, again use the "cluster" option:

```
xtreg y x1 x2 ... xK, fe cluster(id)
```

• Of course, a nonrobust form requires an extra assumption:

ASSUMPTION FE.3: Same as RE.3(a), that is,

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

• To find the asymptotic variance under this assumption, remember the general form for a pooled OLS estimator – in this case, on the time demeaned data – is the sandwich form

$$[E(\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i)]^{-1}E(\ddot{\mathbf{X}}_i'\ddot{\mathbf{u}}_i\ddot{\mathbf{u}}_i'\ddot{\mathbf{X}}_i)[E(\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i)]^{-1}.$$

Under FE.3, we can simplify the middle matrix. First, use

$$\ddot{\mathbf{X}}_i'\ddot{\mathbf{u}}_i = \sum_{t=1}^{T} \ddot{\mathbf{x}}_{it}'\ddot{u}_{it} = \sum_{t=1}^{T} \ddot{\mathbf{x}}_{it}'u_{it} = \ddot{\mathbf{X}}_i'\mathbf{u}_i.$$

Therefore,

$$\begin{aligned} E(\ddot{\mathbf{X}}_i'\ddot{\mathbf{u}}_i\ddot{\mathbf{u}}_i'\ddot{\mathbf{X}}_i) &= E(\ddot{\mathbf{X}}_i'\mathbf{u}_i\mathbf{u}_i'\ddot{\mathbf{X}}_i) = E[E(\ddot{\mathbf{X}}_i'\mathbf{u}_i\mathbf{u}_i'\ddot{\mathbf{X}}_i|\ddot{\mathbf{X}}_i)] \\ &= E[\ddot{\mathbf{X}}_i'E(\mathbf{u}_i\mathbf{u}_i'|\ddot{\mathbf{X}}_i)\ddot{\mathbf{X}}_i] = E[\ddot{\mathbf{X}}_i'(\sigma_u^2\mathbf{I}_T)\ddot{\mathbf{X}}_i] \\ &= \sigma_u^2 E(\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i) \end{aligned}$$

because $E(\mathbf{u}_i\mathbf{u}_i'|\ddot{\mathbf{X}}_i) = \sigma_u^2\mathbf{I}_T$ under FE.3.

• So $Avar[\sqrt{N}(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta})] = \sigma_u^2[E(\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i)]^{-1}$ under FE.1, FE.2, and FE.3.

• Estimating σ_u^2 requires some care because we effectively observe $\ddot{u}_{it}$, not u_{it}.

• Under the constant variance and no serial correlation assumptions on $\{u_{it}\}$,

$$\begin{aligned} Var(\ddot{u}_{it}) &= Var(u_{it} - \bar{u}_i) = \sigma_u^2 + \sigma_u^2/T - 2Cov(u_{it}, \bar{u}_i) \\ &= \sigma_u^2 + \sigma_u^2/T - 2\sigma_u^2/T = \sigma_u^2(1 - 1/T) \end{aligned}$$

- So

$$\sum_{t=1}^{T} E(\ddot{u}_{it}^2) = (T-1)\sigma_u^2.$$

- One degree of freedom is lost for each unit i because of the time demeaning: $\sum_{t=1}^{T} \ddot{u}_{it} = 0$.

- Therefore,

$$\sigma_u^2 = [N(T-1)]^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} E(\ddot{u}_{it}^2)$$

and now take away expectation, insert $\hat{\boldsymbol{\beta}}_{FE}$ for $\boldsymbol{\beta}$, and use a df adjustment to account for estimating the K-vector $\boldsymbol{\beta}$:

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

- $\hat{\sigma}_u^2$ is actually unbiased under FE.1, FE.2, and FE.3. It is consistent as $N \to \infty$.

- Under FE.1, FE.2, and FE.3,

$$\widehat{Avar(\hat{\boldsymbol{\beta}}_{FE})} = \hat{\sigma}_u^2\left(\sum_{i=1}^{N}\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i\right)^{-1} = \hat{\sigma}_u^2(\ddot{\mathbf{X}}'\ddot{\mathbf{X}})^{-1}$$

and this is the “usual” asymptotic variance estimator.

• If you do the time-demeaning and run pooled OLS, the usual statistics do not reflect the lost degrees of freedom (N of them). The estimate of σ_u^2 will be $SSR/(NT - K)$, which is too small. Canned FE packages properly compute the statistics.

• The FE estimator $\hat{\boldsymbol{\beta}}_{FE}$ can also be obtained by running a long regression on the original data, and including dummy variables for each cross section unit:

$$y_{it} \text{ on } d1_i, d2_i, \ldots, dN_i, \mathbf{x}_{it},\ t = 1, \ldots, T; i = 1, \ldots, N,$$

often called the *dummy variable regression*. The statistics are properly computed because the inclusion of the N dummy variables.

• Only danger: treating the c_i as parameters to estimate, while sensible with "large" T, can lead to trouble later with nonlinear models. Here, we get a consistent estimator of $\boldsymbol{\beta}$ for fixed T.

• Sometimes we want to estimate the c_i using the T time periods. Do not have to run the dummy variable regression:

$$\hat{c}_i = \bar{y}_i - \bar{\mathbf{x}}_i \hat{\boldsymbol{\beta}}_{FE}, \ i = 1, \ldots, N.$$

• With small T, this is not a good "estimate" of c_i, but it is unbiased. We can estimate features of the distibution of c_i well:

$$
\begin{aligned}
\hat{\mu}_c &= N^{-1}\sum_{i=1}^{N}\hat{c}_i = N^{-1}\sum_{i=1}^{N}(\bar{y}_i - \bar{\mathbf{x}}_i\hat{\boldsymbol{\beta}}_{FE}) = N^{-1}\sum_{i=1}^{N}[c_i + \bar{u}_i + \bar{\mathbf{x}}_i(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_{FE})] \\
&= N^{-1}\sum_{i=1}^{N}c_i + N^{-1}\sum_{i=1}^{N}\bar{u}_i + \left(N^{-1}\sum_{i=1}^{N}\bar{\mathbf{x}}_i\right)(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_{FE}) \\
&= N^{-1}\sum_{i=1}^{N}c_i + o_p(1) + O_p(1)o_p(1) \overset{p}{\rightarrow} \mu_c.
\end{aligned}
$$

- Stata, for example, reports $\hat{\mu}_c$ as the "intercept" or "constant" in FE regressions.
- This consistency argument uses only FE.1 and FE.2.

• Can estimate other features of the distribution, too, although some "obvious" estimators are inconsistent. For example, we might try to estimate σ_c^2 using the sample variance of $\{\hat{c}_i : i = 1,\ldots,N\}$:

$$\tilde{\sigma}_c^2 = (N-1)^{-1}\sum_{i=1}^{N}(\hat{c}_i - \hat{\mu}_c)^2.$$

But under FE.1 to FE.3 it can be shown that

$$\text{plim}(\tilde{\sigma}_c^2) = \sigma_c^2 + Var(\bar{u}_i) = \sigma_c^2 + \sigma_u^2/T.$$

• We can adjust for the “bias” using the estimate $\hat{\sigma}_u^2$:

$$\hat{\sigma}_c^2 = \tilde{\sigma}_c^2 - \hat{\sigma}_u^2/T = (N-1)^{-1}\sum_{i=1}^{N}(\hat{c}_i - \hat{\mu}_c)^2 - \hat{\sigma}_u^2/T$$

is consistent for σ_c^2 for any T as $N \to \infty$.

• If we treat the c_i as parameters, can test the null that they are the same. This is easy to see if we add to FE.1 to FE.3 the assumption $u_{it}|\mathbf{x}_i, c_i \sim Normal(0, \sigma_u^2)$. Then the classical linear model assumptions hold, and so $H_0 : c_1 = c_2 = \ldots = c_N$ can be tested using an F statistic with $N-1$ and $N(T-1)-K$ degrees of freedom.

- We can also obtain an estimate of σ_c^2 using

$$\sigma_c^2 = \sigma_v^2 - \sigma_u^2$$

We already have $\hat{\sigma}_u^2 = SSR/[N(T-1) - K$, which is consistent for σ_u^2 under FE.1 to FE.3. Also, $v_{it} = y_{it} - \mathbf{x}_{it}\boldsymbol{\beta}$ and so a consistent estimator of σ_v^2 is

$$\hat{\sigma}_v^2 = (NT-K)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \mathbf{x}_{it}\hat{\boldsymbol{\beta}}_{FE} - \hat{\mu}_v)^2,$$

where $\hat{\mu}_v$ is the sample average of the $\{y_{it} - \mathbf{x}_{it}\hat{\boldsymbol{\beta}}_{FE}\}$.

- Recent work by Orme and Yamagata (2006, Econometric Reviews) has shown that the F statistic is approximately valid if we drop the normality assumption on u_{it}, but it is still unknown how to test constancy of the c_i with serial correlation or heteroskedasticity in $\{u_{it}\}$.

Testing for Serial Correlation

- Because we can obtain fully robust inference, why should we test for serial correlation in the $\{u_{it}\}$? The answer is that we might be able to improve efficiency using a GLS-type method.
- We can test for serial correlation in $\{u_{it}\}$, but it is tricky because we effectively only have $\{\ddot{u}_{it}\}$.
- When $\{u_{it}\}$ is serially uncorrelated with constant variance, for $t \neq r$ we have

$$\begin{aligned} E(\ddot{u}_{it}\ddot{u}_{ir}) &= E[(u_{it} - \bar{u}_i)(u_{ir} - \bar{u}_i)] \\ &= -2\sigma_u^2/T + \sigma_u^2/T = -\sigma_u^2/T. \end{aligned}$$

Therefore,

$$Corr(\ddot{u}_{it}, \ddot{u}_{ir}) = \frac{-\sigma_u^2/T}{\sigma_u^2[(T-1)/T]} = -\frac{1}{T-1}.$$

• If the original errors are serially uncorrelated, the time-demeaned errors have a negative correlation, which is smaller as T increases.

• Cannot (and need not) test for serial correlation when $T = 2$ because $\ddot{u}_{i1} = -\ddot{u}_{i2}$.

• But for $T > 2$, can examine whether the fixed effects residuals are consistent with correlation of roughly $-(T-1)^{-1}$.

• A simple test is based on a pooled AR(1) regression. First obtain the FE residuals, $\hat{\ddot{u}}_{it}$. (In Stata, use the "areg" command.) Then run the pooled OLS regression

$$\hat{\ddot{u}}_{it} \text{ on } \hat{\ddot{u}}_{i,t-1},\ t = 3,\ldots,T; i = 1,\ldots,N$$

and let the coefficient on $\hat{\ddot{u}}_{i,t-1}$ be $\hat{\delta}$. The tricky thing is that, under the null, the $\ddot{u}_{it}$ are serially correlated.

• We obtain a simple statistic using a fully robust standard error for $\hat{\delta}$, $se(\hat{\delta})$ (available from the "cluster" option in POLS). The t statistic is

$$\frac{[\hat{\delta} + (T-1)^{-1}]}{se(\hat{\delta})}.$$

- Typically observe $\hat{\delta} > 0$ if $\{u_{it}\}$ is positively serially correlated. A positive, significant estimate of $\hat{\delta}$ reveals some positive serial correlation. If $\hat{\delta} \approx -(T-1)^{-1}$, no serial correlation in $\{u_{it}\}$ might be reasonable.
- If we find strong evidence of serial correlation in $\{u_{it}\}$, we might want to exploit it in estimation rather than just making FE inference robust.

Fixed Effects GLS

- Write the T time periods for a random draw i as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + c_i\mathbf{j}_T + \mathbf{u}_i$$

and let the variance matrix of $\mathbf{u}_i$ to be a $T \times T$ unrestricted matrix $\boldsymbol{\Lambda}$.

- When we eliminate c_i by demeaning we get

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

where, for example, $\ddot{\mathbf{u}}_i = \mathbf{Q}_T\mathbf{u}_i$ and $\mathbf{Q}_T = \mathbf{I}_T - \mathbf{j}_T(\mathbf{j}_T'\mathbf{j}_T)^{-1}\mathbf{j}_T'$ is symmetric, idempotent with rank $T - 1$.

- Because $\mathbf{j}_T'\ddot{\mathbf{u}}_i = 0$, we know the $T \times T$ matrix $E(\ddot{\mathbf{u}}_i\ddot{\mathbf{u}}_i')$ has rank less than T. In fact, (unconditional) variance covariance matrix of $\ddot{\mathbf{u}}_i$ is

$$\mathbf{\Omega} \equiv E(\ddot{\mathbf{u}}_i\ddot{\mathbf{u}}_i') = E(\mathbf{Q}_T\mathbf{u}_i\mathbf{u}_i'\mathbf{Q}_T) = \mathbf{Q}_T\mathbf{\Lambda}\mathbf{Q}_T,$$

which has rank $T - 1$.

- Applying FGLS to

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

is tricky (generalized inverse required).

• There is a simple solution. After demeaning to obtain $\ddot{\mathbf{y}}_i$ and $\ddot{\mathbf{X}}_i$ using all T time periods and obtaining

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

drop one of the time periods. It does not matter which one is dropped (but the first or last are easiest).

• Apply FGLS to the $T - 1$ remaining equations using $\hat{\mathbf{\Omega}}$.

• Remember, can still make a case for robust inference because system heteroskedasticity is always a possibility.

Some Practical Hints in Applying Fixed Effects

• Possible confusion concerning the term "fixed effects." Suppose i is a firm. Then the phrase "firm fixed effect" corresponds to allowing c_i in the model to be correlated with the covariates. If c_i is called a firm "random effect" then it is being assumed to be uncorrelated with $\mathbf{x}_{it}$.

• Suppose that we cannot, or do not want to, use FE estimation. This might occur because the key variable at the firm level is constant across time for all firms – and so the FE transformation sweeps it away – or there is little time variation within firm in the key variable, leading to large standard errors.

- Instead, we might use a random effects analysis at the firm level but include industry dummy variables to account for systematic differences across industries. So, we include in $\mathbf{x}_{it}$ a set of industry dummy variables while also allowing a firm effect c_i in a “random effects” analysis.
- If there are many firms per industry, the industry “fixed effects” – the coefficients on the industry dummies – can be precisely estimated. So the industry “fixed effects” are really parameters to estimate whereas the c_i are not.

- Generally, including dummies for more aggregated levels and then applying RE is common when the covariates of interest vary in the cross section but not (much) over time.
- Keep in mind that an RE analysis at the firm level with industry dummies need not be entirely convincing: the key elements of $\mathbf{x}_{it}$ might be correlated with unoberved firm features that are not adequately captured by industry differences.

Application

For $N = 1,149$ U.S. air routes and the years 1997 through 2000, y_{it} is $\log(fare_{it})$ and the key explanatory variable is $concen_{it}$, the concentration ratio for route i. Other covariates are year dummies and the time-constant variables $\log(dist_i)$ and $[\log(dist_i)]^2$. Note that what I call c_i Stata refers to as u_i.

```
. use airfare

. tab year

1997, 1998, |
 1999, 2000 |      Freq.     Percent        Cum.
------------+-----------------------------------
       1997 |      1,149       25.00       25.00
       1998 |      1,149       25.00       50.00
       1999 |      1,149       25.00       75.00
       2000 |      1,149       25.00      100.00
------------+-----------------------------------
      Total |      4,596      100.00

. sum fare concen dist

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        fare |      4596    178.7968    74.88151         37        522
      concen |      4596    .6101149     .196435      .1605          1
        dist |      4596     989.745    611.8315         95       2724
```

```
. reg lfare concen ldist ldistsq y98 y99 y00

      Source |       SS       df       MS              Number of obs =    4596
-------------+------------------------------           F(  6,  4589) =  523.18
       Model |  355.453858     6  59.2423096           Prob > F      =  0.0000
    Residual |  519.640516  4589  .113236112           R-squared     =  0.4062
-------------+------------------------------           Adj R-squared =  0.4054
       Total |  875.094374  4595  .190444913           Root MSE      =  .33651

------------------------------------------------------------------------------
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .3601203   .0300691    11.98   0.000     .3011705    .4190702
       ldist |  -.9016004    .128273    -7.03   0.000    -1.153077   -.6501235
     ldistsq |   .1030196   .0097255    10.59   0.000     .0839529    .1220863
         y98 |   .0211244   .0140419     1.50   0.133    -.0064046    .0486533
         y99 |   .0378496   .0140413     2.70   0.007      .010322    .0653772
         y00 |     .09987   .0140432     7.11   0.000     .0723385    .1274015
       _cons |   6.209258   .4206247    14.76   0.000     5.384631    7.033884
------------------------------------------------------------------------------
```

```
. reg lfare concen ldist ldistsq y98 y99 y00, cluster(id)

                                  (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .3601203    .058556     6.15   0.000     .2452315    .4750092
       ldist |  -.9016004   .2719464    -3.32   0.001    -1.435168   -.3680328
     ldistsq |   .1030196   .0201602     5.11   0.000     .0634647    .1425745
         y98 |   .0211244   .0041474     5.09   0.000     .0129871    .0292617
         y99 |   .0378496   .0051795     7.31   0.000     .0276872     .048012
         y00 |     .09987   .0056469    17.69   0.000     .0887906    .1109493
       _cons |   6.209258   .9117551     6.81   0.000     4.420364    7.998151
------------------------------------------------------------------------------
```

```
. xtreg lfare concen ldist ldistsq y98 y99 y00, re

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

R-sq:  within  = 0.1348                         Obs per group: min =         4
       between = 0.4176                                        avg =       4.0
       overall = 0.4030                                        max =         4

Random effects u_i ~Gaussian                   Wald chi2(6)       =   1360.42
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
       lfare |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .2089935   .0265297     7.88   0.000     .1569962    .2609907
       ldist |  -.8520921   .2464836    -3.46   0.001    -1.335191   -.3689931
     ldistsq |   .0974604   .0186358     5.23   0.000     .0609348     .133986
         y98 |   .0224743   .0044544     5.05   0.000     .0137438    .0312047
         y99 |   .0366898   .0044528     8.24   0.000     .0279626    .0454171
         y00 |    .098212   .0044576    22.03   0.000     .0894752    .1069487
       _cons |   6.222005   .8099666     7.68   0.000       4.6345     7.80951
-------------+----------------------------------------------------------------
     sigma_u |  .31933841
     sigma_e |  .10651186
         rho |  .89988885   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

```
. * The coefficient on the time-varying variable concen drops quite a bit.
. * Notice that the RE and POLS coefficients on the time-constant
. * distance variables are pretty similar, something that often occurs.

. xtreg lfare concen 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 |   .2089935   .0422459     4.95   0.000      .126193    .2917939
       ldist |  -.8520921   .2720902    -3.13   0.002    -1.385379   -.3188051
     ldistsq |   .0974604   .0201417     4.84   0.000     .0579833    .1369375
         y98 |   .0224743   .0041461     5.42   0.000      .014348    .0306005
         y99 |   .0366898   .0051318     7.15   0.000     .0266317     .046748
         y00 |    .098212   .0055241    17.78   0.000     .0873849     .109039
       _cons |   6.222005   .9144067     6.80   0.000     4.429801    8.014209
-------------+----------------------------------------------------------------
     sigma_u |  .31933841
     sigma_e |  .10651186
         rho |  .89988885   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. * Robust standard error on concen is quite a bit larger.
```

```
. * What if we do not control for distance in RE?

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

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


                                   (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .0468181   .0427562     1.09   0.274    -.0369826    .1306188
         y98 |   .0239229   .0041907     5.71   0.000     .0157093    .0321364
         y99 |   .0354453   .0051678     6.86   0.000     .0253167     .045574
         y00 |   .0964328   .0055197    17.47   0.000     .0856144    .1072511
       _cons |   5.028086   .0285248   176.27   0.000     4.972178    5.083993
-------------+----------------------------------------------------------------
     sigma_u |  .40942871
     sigma_e |  .10651186
         rho |  .93661309   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. * The RE estimate is now much smaller than when ldist and ldistsq are
. * controlled for, and much smaller than the FE estimate. Thus, it can be
. * very harmful to omit time-constant variables in RE estimation.
```

```
. * Allow an unrestricted unconditional variance-covariance matrix, but
. * make robust to system heteroskedasticity:

. xtgee lfare concen ldist ldistsq y98 y99 y00, corr(uns) robust

GEE population-averaged model                   Number of obs      =      4596
Group and time vars:                 id year    Number of groups   =      1149
Link:                               identity    Obs per group: min =         4
Family:                             Gaussian                   avg =       4.0
Correlation:                    unstructured                   max =         4
                                                Wald chi2(6)       =   1246.97
Scale parameter:                    .1135142    Prob > chi2        =    0.0000

                                     (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |             Semi-robust
       lfare |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .2364893   .0406545     5.82   0.000     .1568079    .3161706
       ldist |  -.8806104     .26696    -3.30   0.001    -1.403842   -.3573785
     ldistsq |   .0992803   .0197484     5.03   0.000     .0605741    .1379866
         y98 |   .0222287   .0041432     5.37   0.000     .0141082    .0303492
         y99 |   .0369008   .0051386     7.18   0.000     .0268293    .0469724
         y00 |   .0985136   .0055411    17.78   0.000     .0876533     .109374
       _cons |   6.313734   .8977898     7.03   0.000     4.554098     8.07337
------------------------------------------------------------------------------
```

```
. xtreg lfare concen ldist ldistsq y98 y99 y00, fe

Fixed-effects (within) regression               Number of obs      =      4596
Group variable: id                              Number of groups   =      1149

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

                                                F(4,3443)          =    134.61
corr(u_i, Xb)  = -0.2033                        Prob > F           =    0.0000

------------------------------------------------------------------------------
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |    .168859   .0294101     5.74   0.000     .1111959     .226522
       ldist |  (dropped)
     ldistsq |  (dropped)
         y98 |   .0228328   .0044515     5.13   0.000     .0141048    .0315607
         y99 |   .0363819   .0044495     8.18   0.000     .0276579    .0451058
         y00 |   .0977717   .0044555    21.94   0.000      .089036    .1065073
       _cons |   4.953331   .0182869   270.87   0.000     4.917476    4.989185
-------------+----------------------------------------------------------------
     sigma_u |  .43389176
     sigma_e |  .10651186
         rho |  .94316439   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(1148, 3443) =    36.90          Prob > F = 0.0000
```

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

                                 (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |    .168859   .0494587     3.41   0.001     .0718194    .2658985
       ldist |  (dropped)
     ldistsq |  (dropped)
         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)
------------------------------------------------------------------------------
```

```
. * Let the effect of concen depend on route distance.

. sum ldist if y00

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       ldist |      1149    6.696482    .6595331   4.553877   7.909857

. gen ldistconcen = (ldist - 6.7)*concen
```

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

Fixed-effects (within) regression               Number of obs      =      4596
Group variable: id                              Number of groups   =      1149

                                  (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .1652538   .0482782     3.42   0.001     .0705304    .2599771
 ldistconcen |  -.2498619   .0828545    -3.02   0.003    -.4124251   -.0872987
         y98 |   .0230874   .0041459     5.57   0.000      .014953    .0312218
         y99 |   .0355923   .0051452     6.92   0.000     .0254972    .0456874
         y00 |   .0975745   .0054655    17.85   0.000     .0868511    .1082979
       _cons |    4.93797   .0317998   155.28   0.000     4.875578    5.000362
-------------+----------------------------------------------------------------
     sigma_u |  .50598296
     sigma_e |  .10605257
         rho |  .95791776   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

```
. * Effect at the average of ldist is similar to before. But at one standard
. * deviation of ldist above its mean, the effect of concen is zero:

. lincom concen + .66*ldistconcen

 ( 1)  concen + .66 ldistconcen = 0

------------------------------------------------------------------------------
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .0003449   .0554442     0.01   0.995    -.1084383    .1091281
------------------------------------------------------------------------------

. count if ldist > 6.7 + .66 & y00
  209

. di 209/1149
.1818973

. * So about 18.2* of the routes of ldist greater than one standard deviation
. * above the mean.
```

3.4. First-Differencing Estimation

• Like FE, FD removes c_i. But it does it by differencing adjacent observations. FE and FD are the same when $T = 2$, but differ otherwise. Again, start with the original equation:

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

For FD, we explicitly lose the first time period:

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

The FD estimator is pooled OLS on the first differences.

• In practice, might not difference period dummies, unless interested in the year intercepts in the original levels.

• FD also requires a kind of strict exogeneity. The weakest assumption is

$$E(\Delta\mathbf{x}_{it}'\Delta u_{it}) = 0,\ t = 2,\ldots,T.$$

• Failure of strict exogeneity will cause different inconsistencies in FE and FD when $T > 2$.

• (For later: In unbalanced cases, FD requires that data exists in adjacent time periods. FE does not.)

• A sufficient condition is

ASSUMPTION FD.1: Same as FE.1, $E(u_{it}|\mathbf{x}_i, c_i) = 0$, $t = 1, \ldots, T$.

ASSUMPTION FD.2: Let $\Delta\mathbf{X}_i$ be the $(T-1) \times K$ matrix with rows $\Delta\mathbf{x}_{it}$. Then,

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

• Should make inference robust to serial correlation and heteroskedasticity in the differenced errors, $e_{it} \equiv u_{it} - u_{i,t-1}$. For example, if $\{u_{it}\}$ is uncorrelated, $Corr(e_{it}, e_{i,t+1}) = -.5$.

- After POLS on the first differences, let

$$\hat{e}_{it} = \Delta y_{it} - \Delta \mathbf{x}_{it}\hat{\boldsymbol{\beta}}_{FD},\ t = 2,\ldots,T; i = 1,\ldots,N$$

and let $\hat{\mathbf{e}}_i = (\hat{e}_{i2},\ldots,\hat{e}_{iT})'$ be the $(T-1) \times 1$ residuals. Then

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

is the fully robust variance matrix estimator.

- Use pooled OLS, on the first differences and then use a “cluster” option.

ASSUMPTION FD.3:

$$E(\mathbf{e}_i\mathbf{e}_i'|\Delta\mathbf{X}_i) = \sigma_e^2\mathbf{I}_T$$

where $\sigma_e^2 = E(e_{it}^2)$ for all t.

- Under Assumption FE.3, the usual POLS statistics in the FD regression are asymptotically valid.
- If we believe FD.3, then $u_{it} = u_{i,t-1} + e_{it}$ is a random walk. In a pure time series setting, this means the regression would be "spurious."

- For a given i, the time series "model" would be

$$
\begin{aligned}
y_{it} &= c_i + \mathbf{x}_{it}\boldsymbol{\beta} + u_{it} \\
u_{it} &= u_{i,t-1} + e_{it},
\end{aligned}
$$

where c_i is the intercept for unit i. This does not define a sensible time series regression because $\{u_{it}\}$ is not "mean reverting." One way to see this is $Var(u_{it}) = \sigma_e^2 t$, and so the idiosyncratic error variance grows as a linear function of t.

- Here we can allow random walk behavior in $\{u_{it}\}$ with a short T because we have cross section variation driving the large-sample analysis.

- Testing for serial correlation in $\{e_{it} = \Delta u_{it}\}$ is easy. If we start with $T \geq 3$, then use a t test or heteroskedasticity-robust version for $\hat{\delta}$, where $\hat{\delta}$ is the coefficient on $\hat{e}_{i,t-1}$ in the pooled dynamic OLS regression

$$\hat{e}_{it} \text{ on } \hat{e}_{i,t-1},\ t = 3,\ldots,T; i = 1,\ldots,N.$$

- We can also use this regression to test whether $Corr(e_{it}, e_{i,t-1}) = -.5$, as implied by FE.3. But then the standard error of $\hat{\delta}$ should be made robust to serial correlation. The t statistic in this case is

$$\frac{(\hat{\delta} +.5)}{se(\hat{\delta})}.$$

• Can use the FD residuals to recover an estimate of ρ if we think $\{u_{it} : t = 1, 2, \ldots T\}$ follows a stationary AR(1) process. Then $Cov(u_{it}, u_{i,t-h}) = \rho^h \sigma_u^2,\ h = 0,\ 1\ , \ldots.$ Therefore

$$\begin{aligned} Cov(e_{it}, e_{i,t-1}) &= Cov(u_{it} - u_{i,t-1}, u_{i,t-1} - u_{i,t-2}) \\ &= \rho\sigma_u^2 - \rho^2\sigma_u^2 - \sigma_u^2 + \rho\sigma_u^2 \\ &= -\sigma_u^2(1 - 2\rho + \rho^2) \\ &= -\sigma_u^2(1 - \rho)^2 \end{aligned}$$

- Further,

$$\begin{aligned} Var(e_{it}) &= \sigma_u^2 - 2Cov(u_{it}, u_{i,t-1}) + \sigma_u^2 \\ &= 2\sigma_u^2(1-\rho) \end{aligned}$$

- It follows that

$$Corr(e_{it}, e_{i,t-1}) = \frac{-\sigma_u^2(1-\rho)^2}{2\sigma_u^2(1-\rho)} = \frac{(\rho-1)}{2}.$$

Letting $\delta \equiv Corr(e_{it}, e_{i,t-1})$, we can write

$$\rho = 1 + 2\delta$$

- Notice we get the right answer when $\delta = 0$: namely, $\rho = 1$ (so that $\{u_{it}\}$ follows a random walk). So we can use

$$\hat{\rho} = 1 + 2\hat{\delta}$$

as a consistent estimator of ρ for $\delta \leq 0$.

- If $[\hat{\delta}_L, \hat{\delta}_U]$ is a 95% CI for δ, then we get a 95% CI for ρ by finding $\hat{\rho}_L = 1 + 2\hat{\delta}_L$ and $\hat{\rho}_U = 1 + 2\hat{\delta}_U$.

• Applying feasible GLS after differencing is especially easy because the lost degree of freedom for each i is automatically incorporated by losing the first time period.

• Resulting estimator is the FDGLS estimator. It uses an unrestricted $(T-1) \times (T-1)$ variance matrix in the FD equation

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

where $\Delta \mathbf{u}_i$ is $(T-1) \times 1$.

• Easy to use the `xtgee` command in Stata.

```
. sort id year

. gen clfare = lfare - lfare[_n-1] if year > 1997
(1149 missing values generated)

. gen cconcen = concen - concen[_n-1] if year > 1997
(1149 missing values generated)

. reg clfare cconcen y99 y00

      Source |       SS       df       MS              Number of obs =    3447
-------------+------------------------------           F(  3,  3443) =   45.61
       Model |  2.14076964     3   .71358988           Prob > F      =  0.0000
    Residual |  53.8669392  3443   .01564535           R-squared     =  0.0382
-------------+------------------------------           Adj R-squared =  0.0374
       Total |  56.0077088  3446  .016252963           Root MSE      =  .12508

------------------------------------------------------------------------------
      clfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     cconcen |   .1759764   .0284387     6.19   0.000     .1202181    .2317348
         y99 |  -.0091019   .0052688    -1.73   0.084    -.0194322    .0012284
         y00 |   .0386441   .0052301     7.39   0.000     .0283897    .0488985
       _cons |   .0227692   .0036988     6.16   0.000     .0155171    .0300212
------------------------------------------------------------------------------

. predict eh, resid
```

```
. * Fairly close to FE estimate of .169, but standard errors are probably
. * not correct. The R-squared gives us a measure of how well changes
. * in concentration explain changes in lfare.

. reg clfare cconcen y99 y00, cluster(id)

Linear regression                                      Number of obs =    3447
                                                       F(  3,  1148) =   34.36
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0382
                                                       Root MSE      =  .12508

                                  (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
      clfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     cconcen |   .1759764   .0430367     4.09   0.000     .0915371    .2604158
         y99 |  -.0091019   .0058305    -1.56   0.119    -.0205416    .0023378
         y00 |   .0386441   .0055658     6.94   0.000     .0277239    .0495643
       _cons |   .0227692   .0041573     5.48   0.000     .0146124     .030926
------------------------------------------------------------------------------

. * We can estimate the intercepts in the original model, too, by
. * differencing the year dummies.

. gen cy98 = y98 - y98[_n-1] if year > 1997
(1149 missing values generated)
```

```
. gen cy99 = y99 - y99[_n-1] if year > 1997
(1149 missing values generated)

. gen cy00 = y00 - y00[_n-1] if year > 1997
(1149 missing values generated)

. reg clfare cconcen cy98 cy99 cy00, nocons cluster(id)

Linear regression                                      Number of obs =    3447
                                                       F(  4,  1148) =  118.18
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0952
                                                       Root MSE      =  .12508

                                  (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
      clfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     cconcen |   .1759764   .0430367     4.09   0.000     .0915371    .2604158
        cy98 |   .0227692   .0041573     5.48   0.000     .0146124     .030926
        cy99 |   .0364365    .005153     7.07   0.000      .026326    .0465469
        cy00 |   .0978497   .0055468    17.64   0.000     .0869666    .1087328
------------------------------------------------------------------------------

. * All estimates are now similar to FE. This R-squared is less useful
. * than when a constant is included because it does not remove the average.
. * It is the "uncentered" R-squared.
```

```
. * Test for serial correlation using FD.

. predict eh, resid
(1149 missing values generated)

. gen eh_1 = eh[_n-1] if year > 1998
(2298 missing values generated)

. reg eh eh_1, robust

Linear regression                                      Number of obs =    2298
                                                       F(  1,  2296) =   21.60
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0197
                                                       Root MSE      =   .1169

------------------------------------------------------------------------------
             |               Robust
          eh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        eh_1 |  -.1275163   .0274343    -4.65   0.000    -.1813148   -.0737177
       _cons |  -3.30e-11   .0024386    -0.00   1.000    -.0047821    .0047821
------------------------------------------------------------------------------

. * We can reject zero correlation in FD errors. (Robust to heteroskedasticity.)
```

```
. * Can use xtgee to obtain the FGLS estimator on the FD equation:

. xtgee clfare cconcen y99 y00, corr(uns)

GEE population-averaged model                   Number of obs      =      3447
Group and time vars:                id year     Number of groups   =      1149
Link:                              identity     Obs per group: min =         3
Family:                            Gaussian                    avg =       3.0
Correlation:                   unstructured                    max =         3
                                                Wald chi2(3)       =    119.43
Scale parameter:                   .0156274     Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      clfare |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     cconcen |    .169649   .0285421     5.94   0.000     .1137076    .2255904
         y99 |  -.0092635   .0054855    -1.69   0.091    -.0200149     .001488
         y00 |   .0385667   .0054062     7.13   0.000     .0279707    .0491627
       _cons |   .0228257   .0036967     6.17   0.000     .0155802    .0300712
------------------------------------------------------------------------------
```

```
. xtgee clfare cconcen y99 y00, corr(uns) robust

GEE population-averaged model                   Number of obs      =      3447
Group and time vars:                id year     Number of groups   =      1149
Link:                              identity     Obs per group: min =         3
Family:                            Gaussian                    avg =       3.0
Correlation:                   unstructured                    max =         3
                                                Wald chi2(3)       =    101.68
Scale parameter:                   .0156274     Prob > chi2        =    0.0000

                                     (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |             Semirobust
      clfare |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     cconcen |    .169649    .042983     3.95   0.000     .0854038    .2538942
         y99 |  -.0092635   .0058158    -1.59   0.111    -.0206622    .0021352
         y00 |   .0385667   .0055622     6.93   0.000     .0276651    .0494683
       _cons |   .0228257   .0041575     5.49   0.000     .0146771    .0309743
------------------------------------------------------------------------------

. * The robust standard error for FGLS is about 50% larger than the nonrobust
. * one.
```

```
. reg eh eh_1, cluster(id)

                                   (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
          eh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        eh_1 |  -.1275163   .0272003    -4.69   0.000    -.1808841   -.0741485
       _cons |  -3.30e-11   .0023264    -0.00   1.000    -.0045644    .0045644
------------------------------------------------------------------------------

. lincom eh_1 + .5

 ( 1)  eh_1 = -.5

------------------------------------------------------------------------------
          eh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .3724837   .0272003    13.69   0.000     .3191159    .4258515
------------------------------------------------------------------------------

. * And we can easily reject -.5, too, which is what would happen under FE.3.

. * If we believe u(i,t) follows an AR(1), then we can use
. * rho = 1 + 2*Corr(eh,eh_1)

. di 1 + 2*(-.128)
.744

. * So the estimated rho is pretty high at .744.
```

```
. * Test for serial correlation using FE. Use "areg" to get the FE
. * residuals.

. areg lfare concen y98 y99 y00, absorb(id)

Linear regression, absorbing indicators                Number of obs =    4596
                                                       F(  4,  3443) =  134.61
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.9554
                                                       Adj R-squared =  0.9404
                                                       Root MSE      =  .10651

------------------------------------------------------------------------------
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |    .168859   .0294101     5.74   0.000     .1111959     .226522
         y98 |   .0228328   .0044515     5.13   0.000     .0141048    .0315607
         y99 |   .0363819   .0044495     8.18   0.000     .0276579    .0451058
         y00 |   .0977717   .0044555    21.94   0.000      .089036    .1065073
       _cons |   4.953331   .0182869   270.87   0.000     4.917476    4.989185
-------------+----------------------------------------------------------------
          id |     F(1148, 3443) =     60.521   0.000       (1149 categories)

. predict udh, resid

. sort id year

. gen udh_1 = udh[_n-1] if year > 1998
(2298 missing values generated)
```

```
. reg udh udh_1, cluster(id)

Linear regression                                      Number of obs =    2298
                                                       F(  1,  1148) =    0.87
                                                       Prob > F      =  0.3498
                                                       R-squared     =  0.0006
                                                       Root MSE      =  .08806

                                  (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
         udh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       udh_1 |  -.0285168   .0304886    -0.94   0.350    -.0883364    .0313028
       _cons |   1.45e-11   .0019846     0.00   1.000    -.0038938    .0038938
------------------------------------------------------------------------------

. lincom udh_1 + .333

 ( 1)  udh_1 = -.333

------------------------------------------------------------------------------
         udh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .3044832   .0304886     9.99   0.000     .2446636    .3643028
------------------------------------------------------------------------------

.* -1/(T-1) = -.333 when T = 4. Strongly reject FE.3; appears to be positive
. * serial correlation, as we already concluded using FD.
```

3.5. Prediction

• For prediction with unobserved effects models, we might include only lags of explanatory variables in $\mathbf{x}_{it}$ – so we do not have to forecast future values of the covariates – and then try to forecast $y_{i,T+1}$ based on data observed up through time T. Can show under the full RE assumptions

$$
\begin{aligned}
E(y_{i,T+1}|\mathbf{X}_i, \mathbf{x}_{i,T+1}, y_{i1}, \ldots, y_{iT}) &= \mathbf{x}_{i,T+1}\boldsymbol{\beta} + [\sigma_c^2/(\sigma_c^2 + \sigma_u^2/T)]\bar{v}_i \\
\bar{v}_i &= \bar{y}_i - \bar{\mathbf{x}}_i\boldsymbol{\beta},
\end{aligned}
$$

so the prediction for RE is

$$\mathbf{x}_{i,T+1}\hat{\boldsymbol{\beta}}_{RE} + [\hat{\sigma}_c^2/(\hat{\sigma}_c^2 + \hat{\sigma}_u^2/T)](\bar{y}_i - \bar{\mathbf{x}}_i\hat{\boldsymbol{\beta}}_{RE}).$$

- For fixed effects, the prediction would be

$$\mathbf{x}_{i,T+1}\hat{\boldsymbol{\beta}}_{FE} + (\bar{y}_i - \bar{\mathbf{x}}_i\hat{\boldsymbol{\beta}}_{FE}),$$

which does not shrink the influence of the second term. As $\hat{\sigma}_c^2$ increases relative to $\hat{\sigma}_u^2$, or for large T, the two predictions are similar.

- Seems unlikely that either of these can match dynamic models estimated by pooled OLS. The RE and FE methods each give the same weight to the most recent and earliest outcomes on y.

4. COMPARISON OF ESTIMATORS

FE versus FD.

- Estimates and inference are identical when $T = 2$. Generally, can see differences as T increases.

- Usually think a significant difference signals violation of $Cov(\mathbf{x}_{is}, u_{it}) = 0$, all s, t. FE has some robustness if $Cov(\mathbf{x}_{it}, u_{it}) = \mathbf{0}$ but $Cov(\mathbf{x}_{it}, u_{is}) = 0$, some $s \neq t$: The "bias" is of order $1/T$. FD does not average out the bias over T.

- To see this, maintain contemporaneous exogeneity:

$$E(\mathbf{x}_{it}' u_{it}) = \mathbf{0}.$$

- Generally, under Assumption FE.2, we can write

$$\underset{N\to\infty}{\text{plim}}\ (\hat{\boldsymbol{\beta}}_{FE}) = \boldsymbol{\beta} + \left[T^{-1} \sum_{t=1}^{T} E(\ddot{\mathbf{x}}_{it}' \ddot{\mathbf{x}}_{it}) \right]^{-1} \left[T^{-1} \sum_{t=1}^{T} E(\ddot{\mathbf{x}}_{it}' u_{it}) \right].$$

- Under contemporaneous exogeneity,

$$E(\ddot{\mathbf{x}}_{it}' u_{it}) = -E(\bar{\mathbf{x}}_i' u_{it})$$

and so

$$T^{-1} \sum_{t=1}^{T} E(\ddot{\mathbf{x}}_{it}' u_{it}) = T^{-1} \sum_{t=1}^{T} E(\bar{\mathbf{x}}_i' u_{it}) = -E(\bar{\mathbf{x}}_i' \bar{u}_i).$$

• Under stationarity and weak dependence, $E(\bar{\mathbf{x}}_i'\bar{u}_i) = O(T^{-1})$ because, by the Cauchy-Schwartz inequality, for each j,

$$|Cov(\bar{x}_{ij}, \bar{u}_i)| \leq sd(\bar{x}_{ij})sd(\bar{u}_i)$$

and $sd(\bar{x}_{ij})$, $sd(\bar{u}_i)$ are $O(T^{-1/2})$ where each series is weakly dependent. (If uncorrelated with constant variance, $sd(\bar{u}_i) = \sigma_u/\sqrt{T}$.)

• Further, $T^{-1}\sum_{t=1}^{T} E(\ddot{\mathbf{x}}_{it}'\ddot{\mathbf{x}}_{it})$ is bounded as a function of T. It follows that

$$\operatorname*{plim}_{N\to\infty} (\hat{\boldsymbol{\beta}}_{FE}) = \boldsymbol{\beta} + O(1) \cdot O(T^{-1}) = \boldsymbol{\beta} + O(T^{-1}).$$

• For the first difference estimator, the general probability limit is

$$\operatorname*{plim}_{N\to\infty} (\hat{\boldsymbol{\beta}}_{FD}) = \boldsymbol{\beta} + \left[(T-1)^{-1} \sum_{t=2}^{T} E(\Delta\mathbf{x}_{it}'\Delta\mathbf{x}_{it}) \right]^{-1} \cdot \left[(T-1)^{-1} \sum_{t=2}^{T} E(\Delta\mathbf{x}_{it}'\Delta u_{it}) \right]$$

- If $\{\mathbf{x}_{it} : t = 1, 2, \ldots\}$ is weakly dependent, so is $\Delta\mathbf{x}_{it}$, and so the first average is generally bounded. (In fact, under stationarity this average does not depend on T.
- As for the second average,

$$E(\Delta\mathbf{x}_{it}'\Delta u_{it}) = -[E(\mathbf{x}_{it}'u_{i,t-1}) + E(\mathbf{x}_{i,t-1}'u_{it})]$$

which is constant under stationarity (and generally nonzero). So

$$\underset{N\to\infty}{\text{plim}}\ (\hat{\boldsymbol{\beta}}_{FD}) = \boldsymbol{\beta} + O(1)$$

even if $E(\mathbf{x}_{i,t-1}'u_{it}) = \mathbf{0}$ (so the dynamics given the elements of $\mathbf{x}_{it}$ are correct).

- Can show the previous results hold even if $\{\mathbf{x}_{it}\}$ is $I(1)$ as a time series process (has a "unit root"), but it is crucial that $\{u_{it}\}$ is $I(0)$ (weakly dependent). If the regression is "spurious" in levels, it is better to first difference!

- In simple cases, such as the AR(1) model with $x_{it} = y_{i,t-1}$, can find what the $O(T^{-1})$ term is for FE. If write the model as

$$y_{it} = \beta y_{i,t-1} + (1-\beta)a_i + u_{it}$$

for $-1 < \beta \leq 1$, then $\operatorname{plim}_{N\to\infty}(\hat{\beta}_{FE}) = \beta + O(T^{-1})$. When $\beta = 1$, the second term is $-3/(T+1)$.

- Simple test for feedback when the model does not contain lagged dependent variables, that is, $Cov(\mathbf{x}_{i,t+1}, u_{it}) \neq 0$. Estimate

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

by FE and test $H_0 : \boldsymbol{\delta} = \mathbf{0}$ (fully robust, as usual).

- Only useful for $T \geq 3$ because lose last time period.

```
. * We found that the FE and FD estimates of concen coefficient were
. * pretty close.

. sort id year
. gen concenp1 = concen[_n+1] if year < 2000
. xtreg lfare concen concenp1 y98 y99 y00, fe cluster(id)

Fixed-effects (within) regression               Number of obs      =      3447
Group variable: id                              Number of groups   =      1149

R-sq:  within  = 0.0558                         Obs per group: min =         3
       between = 0.0535                                        avg =       3.0
       overall = 0.0347                                        max =         3

                                                F(4,1148)          =     25.63
corr(u_i, Xb)  = -0.2949                        Prob > F           =    0.0000

                                  (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .2983988    .054797     5.45   0.000     .1908854    .4059122
    concenp1 |  -.0659259   .0467578    -1.41   0.159    -.1576663    .0258145
         y98 |   .0205809   .0042341     4.86   0.000     .0122735    .0288883
         y99 |   .0360638   .0050754     7.11   0.000     .0261058    .0460218
         y00 |  (dropped)
       _cons |   4.914953   .0478488   102.72   0.000     4.821072    5.008834
-------------+----------------------------------------------------------------
```

• If do not reject strict exogeneity, can use serial correlation properties of $\{u_{it}\}$ to choose between FE and FD. Generally a good idea to do FE and FD and report robust standard errors.

• If we maintain system homoskedasticity (sufficient is $Var(\mathbf{u}_i|\mathbf{x}_i, c_i) = Var(\mathbf{u}_i)$), then unrestricted FDGLS and FEGLS (with a time period dropped) are asymptotically equivalent.

FE versus RE.

• Time-constant variables drop out of FE estimation. On the time-varying covariates, are FE and RE so different after all? Define the parameter

$$\lambda = 1 - \left[\frac{1}{1 + T(\sigma_c^2/\sigma_u^2)} \right]^{1/2},$$

which is consistently estimated (for fixed T) by $\hat{\lambda}$. (Some authors use θ as the symbol.) The, the RE estimate can be obtained from the pooled OLS regression

$$y_{it} - \hat{\lambda}\bar{y}_i \text{ on } \mathbf{x}_{it} - \hat{\lambda}\bar{\mathbf{x}}_i, t = 1, \ldots, T; i = 1, \ldots, N.$$

• Call $y_{it} - \hat{\lambda}\bar{y}_i$ a “quasi-time-demeaned” variable: only a fraction of the mean is removed.

$$\hat{\lambda} \approx 0 \Rightarrow \hat{\boldsymbol{\beta}}_{RE} \approx \hat{\boldsymbol{\beta}}_{POLS}$$
$$\hat{\lambda} \approx 1 \Rightarrow \hat{\boldsymbol{\beta}}_{RE} \approx \hat{\boldsymbol{\beta}}_{FE}$$

λ increases to unity as (i) σ_c^2/σ_u^2 increases or (ii) T increases. With large T, FE and RE are often similar.

• If $\mathbf{x}_{it}$ includes time-constant variables $\mathbf{z}_i$, then $(1 - \hat{\lambda})\mathbf{z}_i$ appears as a regressor.

```
. * Can get the quasi-time-demeaning parameter, which Stata calls “theta.”

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

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

Random effects u_i ~Gaussian                   Wald chi2(7)       = 386792.52
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000
theta              = .83550226

                                  (Std. Err. adjusted for 1149 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lfare |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      concen |   .2089935   .0422459     4.95   0.000      .126193    .2917939
       ldist |  -.8520921   .2720902    -3.13   0.002    -1.385379   -.3188051
     ldistsq |   .0974604   .0201417     4.84   0.000     .0579833    .1369375
         y98 |   .0224743   .0041461     5.42   0.000      .014348    .0306005
         y99 |   .0366898   .0051318     7.15   0.000     .0266317     .046748
         y00 |    .098212   .0055241    17.78   0.000     .0873849     .109039
       _cons |   6.222005   .9144067     6.80   0.000     4.429801    8.014209
-------------+----------------------------------------------------------------
     sigma_u |  .31933841
     sigma_e |  .10651186
         rho |  .89988885   (fraction of variance due to u_i)
------------------------------------------------------------------------------
 .* The value .836 makes it clear why FE and RE are pretty close.
```

Testing for Serial Correlation after RE

- Can show that under the RE variance matrix assumptions, $r_{it} \equiv v_{it} - \lambda\bar{v}_i = (1-\lambda)c_i + u_{it} - \lambda\bar{u}_i$ has constant (unconditional) variance and is serially uncorrelated.

- Suggests a way to test $\{u_{it}\}$ for serial correlation. After RE estimation, obtain $\hat{r}_{it}$ from the regression on the quasi-time-demeaned data, and use a standard test for, say, AR(1) serial correlation. (Can ignore estimation of parameters.)

Efficiency of RE

- Can show that RE is asymptotically more efficient than FE under RE.1, RE.2, FE.2, and RE.3. Assume, for simplicity, $\mathbf{x}_{it}$ has all time-varying elements. (See text Section 10.7.2 for more general case.)

- Then

$$Avar(\hat{\boldsymbol{\beta}}_{FE}) = \sigma_u^2[E(\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i)]^{-1}/N$$

- Let $\breve{\mathbf{x}}_{it} = \mathbf{x}_{it} - \lambda\bar{\mathbf{x}}_i$ be the quasi-time demeaned time-varying covariates. Then

$$Avar(\hat{\boldsymbol{\beta}}_{RE}) = \sigma_u^2[E(\breve{\mathbf{X}}_i'\breve{\mathbf{X}}_i)]^{-1}/N$$

- Using $\sum_{t=1}^{T} \ddot{\mathbf{x}}_{it} = \mathbf{0}$ we have

$$\begin{aligned}
\breve{\mathbf{X}}_i'\breve{\mathbf{X}}_i &= \sum_{t=1}^{T} \breve{\mathbf{x}}_{it}'\breve{\mathbf{x}}_{it} = \sum_{t=1}^{T}[\ddot{\mathbf{x}}_{it} + (1-\lambda)\bar{\mathbf{x}}_i]'[\ddot{\mathbf{x}}_{it} + (1-\lambda)\bar{\mathbf{x}}_i] \\
&= \sum_{t=1}^{T}[\ddot{\mathbf{x}}_{it}'\ddot{\mathbf{x}}_{it} + (1-\lambda)^2\bar{\mathbf{x}}_i'\bar{\mathbf{x}}_i] \\
&= \ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i + (1-\lambda)^2 T\bar{\mathbf{x}}_i'\bar{\mathbf{x}}_i
\end{aligned}$$

$$E(\breve{\mathbf{X}}_i'\breve{\mathbf{X}}_i) - E(\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i) = (1-\lambda)^2 TE(\bar{\mathbf{x}}_i'\bar{\mathbf{x}}_i)$$

which is positive semidefinite.

Testing the Key RE Assumption

- Recall the key RE assumption is $Cov(\mathbf{x}_{it}, c_i) = 0$. With lots of good time-constant controls ("observed heterogeneity") might be able to make this condition roughly true.
- a. The traditional Hausman Test: Compare the coefficients on the time-varying explanatory variables, and compute a chi-square statistic.
- Caution: Usual Hausman test maintains RE.3 – second moment assumptions – yet has no systematic power for detecting violations from this assumption.
- With time effects, must use generalized inverse. Easy to get the degrees of freedom wrong.

• b. Variable addition test. Write the model as

$$y_{it} = \mathbf{g}_t\boldsymbol{\theta} + \mathbf{z}_i\boldsymbol{\delta} + \mathbf{w}_{it}\boldsymbol{\gamma} + c_i + u_{it}.$$

Obvious we cannot compare FE and RE estimates of $\boldsymbol{\delta}$ because the former is not defined. Less obvious we cannot compare FE and RE estimates of $\boldsymbol{\theta}$ (because FE and RE both allow estimation). But it turns out we can only compare $\hat{\boldsymbol{\gamma}}_{FE}$ and $\hat{\boldsymbol{\gamma}}_{RE}$.

• Let $\mathbf{w}_{it}$ be $1 \times J$. Use a *correlated random effects (CRE)* formulation due to Mundlak (1978):

$$\begin{aligned} c_i &= \psi + \bar{\mathbf{w}}_i \boldsymbol{\delta} + a_i \\ E(a_i | \mathbf{z}_i, \mathbf{w}_i) &= 0. \end{aligned}$$

This allows c_i to be correlated with the time-varying explanatory variables through its average level over time. (We might think of this as a long-run component of $\{\mathbf{w}_{it} : t = 1, \ldots, T\}$.

- If we substitute $c_i = \psi + \bar{\mathbf{w}}_i \boldsymbol{\delta} + a_i$ into the original equation we get

$$y_{it} = \mathbf{g}_t \boldsymbol{\theta} + \mathbf{z}_i \boldsymbol{\delta} + \mathbf{w}_{it} \boldsymbol{\gamma} + \psi + \bar{\mathbf{w}}_i \boldsymbol{\delta} + a_i + u_{it}.$$

Estimate this model using RE and test $H_0 : \boldsymbol{\delta} = \mathbf{0}$ using RE estimation. Should make test fully robust if have any doubt about RE.3 (which we almost always should).

- The RE estimate of $\boldsymbol{\gamma}$ when $\bar{\mathbf{w}}_i$ is included is actually the FE estimate. For that matter, so is the POLS estimate. Including $\bar{\mathbf{w}}_i$ effectively proxies for c_i. (The remaining heterogeneity, a_i, is uncorrelated with all explanatory variables.)

• When we use the CRE formulation to obtain a test of

$$E(c_i|\mathbf{z}_i, \mathbf{w}_i) = E(c_i)$$

there is no mean relationship between c_i and $(\mathbf{w}_{i1}, \mathbf{w}_{i2}, \ldots, \mathbf{w}_{iT})$. The alternative $E(c_i|\mathbf{z}_i, \mathbf{w}_i) = E(c_i|\bar{\mathbf{w}}_i) = \psi + \bar{\mathbf{w}}_i\boldsymbol{\delta}$ is a convenient way to obtain a test.

• Nevertheless, if we believe $E(c_i|\mathbf{z}_i, \mathbf{w}_i) = \psi + \bar{\mathbf{w}}_i\boldsymbol{\delta}$ (or use linear projections) then the CRE formulation has the benefit of allowing us to estimate the coefficients on $\mathbf{z}_i$, the time-consant variables.

• Guggenberger (2010, *Journal of Econometrics*) has recently pointed out the pre-testing problem in using the Hausman test to decide between RE and FE. The regression-based version of the test shows it is related to the classic problem of pre-testing on a set of regressors – $\bar{\mathbf{w}}_i$ in this case – in order to decide whether or not to include them.

• If $\boldsymbol{\xi} \neq \mathbf{0}$ but the test has low power, we will omit $\bar{\mathbf{w}}_i$ when we should include it. That is, we will incorrectly opt for RE.

• As always, need to distinguish between a statistical and practical rejection.

Airfare Example

```
. * First use the Hausman test that maintains all of the RE assumptions under
. * the null and directly compares the RE and FE estimates:

. qui xtreg lfare concen ldist ldistsq y98 y99 y00, fe

. estimates store b_fe

. qui xtreg lfare concen ldist ldistsq y98 y99 y00, re

. estimates store b_re
```

```
. hausman b_fe b_re

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |      b_fe         b_re        Difference          S.E.
-------------+----------------------------------------------------------------
      concen |     .168859     .2089935       -.0401345        .0126937
         y98 |    .0228328     .0224743        .0003585               .
         y99 |    .0363819     .0366898        -.000308               .
         y00 |    .0977717      .098212       -.0004403               .
------------------------------------------------------------------------------
                           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(4) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =       10.00
                Prob>chi2 =      0.0405
                (V_b-V_B is not positive definite)
.
. di -.0401/.0127
-3.1574803

. * This is the nonrobust H t test based just on the concen variable. There is
. * only one restriction to test, not four. The p-value reported for the
. * chi-square statistic is incorrect. Notice that the rejection using the
. * correct df is much stronger than if we act as if there are four restrictions.
```

```
. * Using the same variance matrix estimator solves the problem of wrong df.
. * The next command uses the matrix of the relatively efficient estimator.

. hausman b_fe b_re, sigmamore

Note: the rank of the differenced variance matrix (1) does not equal the
      number of coefficients being tested (4); be sure this is what you expect,
      or there may be problems computing the test.  Examine the output of your
      estimators for anything unexpected and possibly consider scaling your
      variables so that the coefficients are on a similar scale.

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |      b_fe         b_re        Difference          S.E.
-------------+----------------------------------------------------------------
      concen |     .168859     .2089935       -.0401345        .0127597
         y98 |    .0228328     .0224743        .0003585         .000114
         y99 |    .0363819     .0366898        -.000308        .0000979
         y00 |    .0977717      .098212       -.0004403          .00014
------------------------------------------------------------------------------
                           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)
                          =        9.89
                Prob>chi2 =      0.0017
```

```
. * The regression-based test is better: it gets the df right AND is fully
. * robust to violations of the RE variance-covariance matrix:

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

. xtreg lfare concen concenbar 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 |    .168859   .0494749     3.41   0.001       .07189    .2658279
   concenbar |   .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
-------------+----------------------------------------------------------------
     sigma_u |  .31933841
     sigma_e |  .10651186
         rho |  .89988885   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. * So the robust t statistic is 2.62 --- still a rejection, but not as strong.
```

```
. * Using the CRE formulation, we get the FE estimate on the time-varying
. * covariate concen. In this case, the coefficients on the time-constant
. * variables are close to the usual RE estimates, and even closer to the
. * POLS estimates.
```