INTRODUCTION AND LINEAR MODELS

Correlated Random Effects Panel Data Models

IZA Summer School in Labor Economics
May 13-19, 2013
Jeffrey M. Wooldridge
Michigan State University

1. Introduction
2. The Linear Model with Additive Heterogeneity
3. Assumptions
4. Estimation and Inference
5. The Robust Hausman Test
6. Correlated Random Slopes
1. Introduction

• Why correlated random effects (CRE) models?

1. In some cases, CRE approaches lead to widely used estimators, such as fixed effects (FE) in a linear model.

2. The CRE approach leads to simple, robust tests of correlation between heterogeneity and covariates. Hausman test comparing random effects (RE) and fixed effects in a linear model.
3. For nonlinear models, avoids the incidental parameters problem (at the cost of restricting conditional heterogeneity distributions).
4. Average partial effects – not just parameters – are generally identified using CRE approaches.
5. Can combine CRE and the related control function approach for nonlinear models with heterogeneity and endogeneity.
6. Recent work on extending the CRE approach to unbalanced panels.
7. Can use the CRE approach for dynamic models. Helps solve the initial conditions problem.
2. The Linear Model with Additive Heterogeneity

- Assume a large population of cross-sectional units (say, individuals or families or firms) that we can observe over time.
- We randomly sample from the cross section, so observations are necessarily independent in the cross section.
• With a large cross section \((N)\) and relatively few time periods \((T)\), we can allow arbitrary time series dependence when conducting inference. Asymptotics is with fixed \(T\) and \(N \to \infty\).

• For inference, we need not worry about “unit roots” in the time series dimension.

• Start with the balanced panel case, and denote the observed random draw for unit \(i\) as \(\{(x_{it}, y_{it}) : t = 1, \ldots, T\}\).
• The unobserved heterogeneity, denoted $c_i$, is drawn along with the observed data.

• View the $c_i$ as random draws. The “fixed” versus “random” debate is counterproductive. The key is what we assume about the relationship between the unobserved $c_i$ and the observed covariates, $x_{it}$.

• The labels “unobserved effect” or “heterogeneity” are neutral. The “fixed effects” and “random effects” labels are best attached to common estimation methods.
• The basic linear model with additive heterogeneity is

\[ y_{it} = \eta_t + x_{it} \beta + c_i + u_{it}, \quad t = 1, \ldots, T, \]

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

\[ v_{it} = c_i + u_{it} \]

• The sequence \( \{v_{it} : t = 1, \ldots, T\} \) is almost certainly serially correlated, and definitely is if \( \{u_{it}\} \) is serially uncorrelated.

• \( x_{it} \) is a \( 1 \times K \) row vector; at this point, it can contain variables that change across \( i \) only, or across \( i \) and \( t \).
• With a short panel, the time period intercepts, $\eta_t$, are treated as parameters that can be estimated by including dummy variables for different time periods.

• With a different setup, such as small $N$ and large $T$, it makes sense to view the $\eta_t$ as random variables that induce cross-sectional correlation.

• When convenient, we absorb time dummies into $x_{it}$. 
3. Assumptions

• Absorb aggregate time effects into $x_{it}$ and write, for a random draw $i$,

$$y_{it} = x_{it}\beta + c_i + u_{it}, \quad t = 1, \ldots, T.$$  

• $x_{it}$ can include interactions of variables with time periods dummies, and general nonlinear functions and interactions, so the model is quite flexible.
Especially for the CRE approach, useful to separate out different kinds of covariates:

\[ y_{it} = g_t \theta + z_i \delta + w_{it} \gamma + c_i + u_{it} \]

- \( g_t \) is a vector of aggregate time effects (often but not necessarily time dummies)
- \( z_i \) is a set of time-constant observed variables
- \( w_{it} \) changes across \( i \) and \( t \) (for at least some units \( i \) and time periods \( t \)).
- Depending on assumptions, we may not be able to consistently estimate \( \delta \).
• Airfare example:

\[
\log(fare_{it}) = \eta_t + \beta_1 \text{concen}_{it} + \beta_2 \log(dist_i) + \beta_3 [\log(dist_i)]^2 \\
+ c_i + u_{it}
\]

• \text{concen}_{it} is a measure of concentration on route \(i\) in year \(t\). The \(\eta_t\) are unrestricted year effects capturing secular changes in airfare.

• Distance between cities does not change over time.
• Main interest is in the coefficient on a variable that changes across $i$ and $t$, $\text{concen}_{it}$. Distance is a control.
• Are there time-constant differences in routes not captured by distance? Almost certainly. Are those factors, in $c_i$, correlated with $\text{concen}_{it}$? Probably.
• How can we get the most convincing estimate of $\beta_1$ and obtain a reliable confidence interval?
Exogeneity Assumptions on the Explanatory Variables

\[ y_{it} = x_{it}\beta + c_i + u_{it} \]

*Contemporaneous Exogeneity (Conditional on the Unobserved Effect):*

\[ E(u_{it}|x_{it}, c_i) = 0 \]

or

\[ E(y_{it}|x_{it}, c_i) = x_{it}\beta + c_i, \]

which gives the \( \beta_j \) partial effects interpretations *holding* \( c_i \) fixed.
• $\beta$ is not identified without more assumptions.
• And still, contemporaneous exogeneity already rules out standard kinds of endogeneity where some elements of $x_{it}$ are correlated with $u_{it}$: measurement error, simultaneity, and time-varying omitted variables.
• In terms of zero correlation, the CE assumption is

$$Cov(x_{it}, u_{it}) = 0, \ t = 1, \ldots, T$$
• **Strict Exogeneity (Conditional on the Unobserved Effect):**

\[
E(y_{it}|x_{i1}, \ldots, x_{iT}, c_i) = E(y_{it}|x_{it}, c_i) = x_{it}\beta + c_i,
\]
so that only \(x_{it}\) 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 strict exogeneity condition holds conditional on \(c_i\),

\[
E(y_{it}|x_{i1}, \ldots, x_{iT}) = x_{it}\beta + E(c_i|x_{i1}, \ldots, x_{iT}).
\]

So correlation between \(c_i\) and \((x_{i1}, \ldots, x_{iT})\) would invalidate the assumption without conditioning on \(c_i\).
• Strict exogeneity implies, for example, correct distributed lag dynamics (a challenge with small $T$).
• Strict exogeneity definitely rules out lagged dependent variables.
• Rules out other situations where shocks today affect future movements in covariates:

$$E(u_{it}|x_{i1}, \ldots, x_{iT}, c_i) = 0.$$
An important implication of strict exogeneity – sometimes used as the definition – is

$$Cov(x_{is}, u_{it}) = 0, s, t = 1, \ldots, T.$$  

In other words, the covariates at any time $s$ are uncorrelated with the idiosyncratic errors at any time $t$.

As an example, suppose we want to estimate a distributed lag model with a single lag, so $x_{it} = (z_{it}, z_{i,t-1})$, and

$$y_{it} = \eta_t + z_{it} \delta_0 + z_{i,t-1} \delta_1 + c_i + u_{it}$$
• Then strict exogeneity means

\[ E(y_{it}|z_{i1}, \ldots, z_{it}, \ldots, z_{iT}, c_i) = E(y_{it}|z_{it}, z_{i,t-1}, c_i) \]
\[ = \eta_t + z_{it}\delta_0 + z_{i,t-1}\delta_1 + c_i \]

• We must have the distributed lag dynamics correct and we cannot allow the shocks \( u_{it} \) to be correlated with, say, \( z_{i,t+1} \). In other words, there can be no feedback.
• RE, FE, and CRE all rely on strict exogeneity when $T$ is not large.
• In applications we need to ask: Why are the explanatory variables changing over time, and might those changes be related to past shocks to $y_{it}$?
• Example: If a worker changes his union status, is he reacting to past shocks to earnings?
• Example: Do shocks to air fares feed back into future changes in route concentration?
Sequential Exogeneity (Conditional on the Unobserved Effect):

- A more natural assumption is

\[
E(y_{it}|x_{it}, x_{i,t-1}, \ldots, x_{i1}, c_i) = E(y_{it}|x_{it}, c_i) = x_{it}\beta + c_i. 
\]

- Sequential exogeneity is a middle ground between contemporaneous and strict exogeneity. It allows lagged dependent variables and other variables that change in reaction to past shocks.

- Strict exogeneity effectively imposes restrictions on economic behavior while sequential exogeneity is less restrictive.
Assumptions about the Unobserved Effect (Heterogeneity)

- In modern applications, treating $c_i$ as a “random effect” essentially means

$$\text{Cov}(x_{it}, c_i) = 0, \ t = 1, \ldots, T,$$

although we sometimes strengthen this to

$$E(c_i|x_i) = E(c_i)$$

where $x_i = (x_{i1}, x_{i2}, \ldots, x_{iT})$.

- Under this key RE assumption, we can technically estimate $\beta$ with a single cross section.
• The label “fixed effect” means that no restrictions are placed on the relationship between $c_i$ and $\{x_{it}\}$. It can be (and traditionally was) taken to mean the $c_i$ are parameters to be estimated. In the standard linear model, these two views lead essentially to the same place – at least for estimating $\beta$. But one has to be careful in general.
• The term *correlated random effects* is used to denote situations where we model the relationship between \( c_i \) and \( \{x_{it}\} \).
• A CRE approach allows us to unify the fixed and random effects estimation approaches.
• Often,

\[
E(c_i|x_{i1}, \ldots, x_{iT}) = E(c_i|\bar{x}_i) = \psi + \bar{x}_i \xi
\]

where \( \bar{x}_i = T^{-1} \sum_{r=1}^{T} x_{ir} \) is the vector of time averages.
• Proposed by Mundlak (1978) and relaxed by Chamberlain (1980, 1982).
• Useful to decompose $c_i$ as

$$c_i = \psi + \bar{x}_i \xi + a_i$$

$$E(a_i|x_i) = 0$$

• Then

$$y_{it} = x_{it} \beta + \psi + \bar{x}_i \xi + a_i + u_{it}$$

• If we also assume strict exogeneity,

$$E(y_{it}|x_i) = E(y_{it}|x_{it}, \bar{x}_i) = x_{it} \beta + \psi + \bar{x}_i \xi, \ t = 1, \ldots, T.$$
4. Estimation and Inference

- Use the CRE approach as a unifying theme.
- The estimating equation is

\[ y_{it} = x_{it} \beta + \psi + \bar{x}_i \xi + a_i + u_{it} \]

\[ \equiv x_{it} \beta + \psi + \bar{x}_i \xi + v_{it} \]

- If

\[ E(a_i|x_i) = 0 \]
\[ E(u_{it}|x_i) = 0, \ t = 1, \ldots, T \]

then \( E(v_{it}|x_i) = 0. \)
• We can use pooled OLS to consistently estimate all parameters, including $\xi$.

• With the CRE approach we can include time-constant variables.

• If we start with

$$y_{it} = g_t \theta + z_i \delta + w_{it} \gamma + c_i + u_{it}$$

then we can use the CRE estimating equation

$$y_{it} = g_t \theta + z_i \delta + w_{it} \gamma + \psi + \bar{w}_i \xi + a_i + u_{it}$$

• However, we must use caution in interpreting $\hat{\delta}$. 
• Well-known algebraic equivalence: The pooled OLS estimator on

\[ y_{it} = g_t \theta + z_i \delta + w_i \gamma + \psi + \bar{w}_i \xi + \nu_{it} \]

gives the fixed effects estimates of \( \theta \) and \( \gamma \), the coefficients on the
time-varying covariates.

• Important consequence: For estimating \( \theta \) and \( \gamma \), the CRE approach is
robust to arbitrary violations of

\[ E(c_i|w_i) = \psi + \bar{w}_i \xi \]

• Strict exogeneity of \( \{x_{it} : t = 1, \ldots, T\} \) with respect to \( u_{it} \) is still
required.
Because of the presence of $a_i$ the error $v_{it} = a_i + u_{it}$ likely has lots of serial correlation.

Alternative to pooled OLS is feasible GLS based on a specific variance-covariate matrix:

$$Cov(a_i, u_{it}) = 0, \text{ all } t$$

$$Cov(u_{it}, u_{is}) = 0, \text{ all } t \neq s$$

$$Var(u_{it}) = \sigma_u^2, \text{ all } t$$

Technically, we should condition on $x_i$. 
• If we write

\[ y_{it} = \mathbf{x}_{it} \mathbf{\beta} + \psi + \bar{\mathbf{x}}_i \xi + v_{it} \]

and let \( \mathbf{v}_i \) be \( T \times 1 \), then \( \mathbf{\Omega} = E(\mathbf{v}_i \mathbf{v}_i^\top) \) has the RE structure:

\[
\mathbf{\Omega} = \begin{pmatrix}
\sigma_a^2 + \sigma_u^2 & \cdots & \sigma_a^2 & \sigma_a^2 \\
\sigma_a^2 & \sigma_a^2 + \sigma_u^2 & \sigma_a^2 \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_a^2 & \cdots & \sigma_a^2 & \sigma_a^2 + \sigma_u^2
\end{pmatrix}.
\]
• Feasible GLS is straightforward – *xtreg* in Stata.

• Another algebraic fact: If we apply RE to

\[ y_{it} = g_i \theta + z_i \delta + w_{it} \gamma + \psi + \bar{w}_i \xi + a_i + u_{it} \]

we still obtain the FE estimates of \( \theta \) and \( \gamma \).

• Estimates of \( \delta \) and \( \xi \) differ from pooled OLS.
• Summary: The CRE approach allows us to include the time-constant variables $z_i$ and at the same time delivers the FE estimates on the time-varying covariates.

• Setting $\xi = 0$ gives the usual RE estimates.
• Two ways that RE can fail to be true GLS.

(i) $\text{Var}(v_i)$ does not have the RE form.

(ii) $\text{Var}(v_i|x_i) \neq \text{Var}(v_i)$ (“system heteroskedasticity”).

• Important: Either way, RE is generally consistent provided a mild rank condition holds.
If we apply RE but it is not truly GLS then we are performing a “quasi-” GLS procedure: the variance-covariance matrix we are using is wrong. But the RE estimator is still consistent under the exogeneity conditions.

The main implication of serial correlation in \( \{u_{it}\} \), or heteroskedasticity in \( a_i \) or \( u_{it} \), is that we should make our inference fully robust.

Might enhance efficiency by using an unrestricted \( T \times T \) variance-covariance matrix.
• In Stata, fully robust inference requires the “cluster” option; nonrobust inference drops it.

    xtset id year
    egen w1bar = mean(w1), by(id)
    :  
    egen wMbar = mean(wM), by(id)
    xtreg y d2 ... dT z1 ... zJ w1 w2 ... wM w1bar ... wMbar, re cluster(id)
EXAMPLE: 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 \([\log(dist_i)]^2\).

(AIRFARE.DTA)

```
. sum fare concen dist

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>fare</td>
<td>4596</td>
<td>178.7968</td>
<td>74.88151</td>
<td>37</td>
<td>522</td>
</tr>
<tr>
<td>concen</td>
<td>4596</td>
<td>.6101149</td>
<td>.196435</td>
<td>.1605</td>
<td>1</td>
</tr>
<tr>
<td>dist</td>
<td>4596</td>
<td>989.745</td>
<td>611.8315</td>
<td>95</td>
<td>2724</td>
</tr>
</tbody>
</table>
```

. xtset id year
panel variable: id (strongly balanced)
time variable: year, 1997 to 2000
delta: 1 unit
Pooled OLS:

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

(Std. Err. adjusted for 1149 clusters in id)

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

* Indirect evidence of plenty of serial correlation in the composite error.
. xtreg lfare concen ldist ldistsq y98 y99 y00, re

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

R-sq: within = 0.1348
between = 0.4176
overall = 0.4030

Obs per group: min = 4
avg = 4.0
max = 4

Random effects u_i ~ Gaussian
corr(u_i, X) = 0 (assumed)

Wald chi2(6) = 1360.42
Prob > chi2 = 0.0000

| Coef. Std. Err. z P>|z| [95% Conf. Interval] |
|---------------------|---------------------|---------------------|---------------------|
| lfare               | 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) |
. xtreg lfare concen ldist ldistsq y98 y99 y00, re cluster(id)

(Std. Err. adjusted for 1149 clusters in id)

|              | Coef.  | Std. Err. | z   | P>|z| | [95% Conf. Interval] |
|--------------|--------|-----------|-----|------|----------------------|
| lfare        | .2089935 | .0422459 | 4.95 | 0.000 | .126193 .2917939     |
| 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) |   |      |                     |

* Even though we have done "GLS," the robust standard error is still much larger than the nonrobust one. The robust GLS standard error is substantially below the robust POLS standard error.
Now fixed effects, first with nonrobust inference:

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

Fixed-effects (within) regression

- Number of obs = 4596
- Number of groups: id = 1149

R-sq:  
- within = 0.1352
- between = 0.0576
- overall = 0.0083

- Obs per group: min = 4
- avg = 4.0
- max = 4

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

|     | Coef. | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|-----|-------|-----------|-------|------|----------------------|
| lfare | .168859 | .0294101 | 5.74  | 0.000 | .1111959 .226522   |
| concen |       |          |       |      |                      |
| 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

40
```
.xtreg lfare concen ldist ldistsq y98 y99 y00, fe cluster(id)

(Std. Err. adjusted for 1149 clusters in id)

|            | Coef.  | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|------------|--------|-----------|-------|------|----------------------|
| lfare      | 0.1687 | 0.0494    | 3.41  | 0.001| 0.0718 to 0.2659     |
| concen     |        |           |       |      |                      |
| ldist      | (dropped) |        |       |      |                      |
| ldistsq    | (dropped) |        |       |      |                      |
| y98        | 0.0228 | 0.0042    | 5.48  | 0.000| 0.0147 to 0.0310     |
| y99        | 0.0363 | 0.0052    | 7.10  | 0.000| 0.0263 to 0.0464     |
| y00        | 0.0977 | 0.0055    | 17.76 | 0.000| 0.0869 to 0.1085     |
| _cons      | 4.95   | 0.0297    | 166.91| 0.000| 4.895 to 5.012      |

|            |        |           |       |      |                      |
| sigma_u    | 0.43   |           |       |      |                      |
| sigma_e    | 0.11   |           |       |      |                      |
| rho        | 0.94   |           |       |      |                      |

* Again there is indirect evidence of serial correlation in the
* idiosyncratic errors.
```
* Now the CRE approach. concen is the only variable that varies
* across both i and t.

. 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)

|            | Coef.   | Robust Std. Err. | z     | P>|z|    | [95% Conf. Interval] |
|------------|---------|------------------|-------|--------|---------------------|
| lfare      |         |                  |       |        |                     |
| concen     | 0.168859| 0.0494749        | 3.41  | 0.001  | 0.07189 - 0.2658279 |
| concenbar  | 0.213635| 0.0816403        | 2.62  | 0.009  | 0.0536227 - 0.3736466|
| ldist      | -0.9089297| 0.2721637 | -3.34 | 0.001  | -1.442361 - -0.3754987|
| ldistsq    | 0.1038426| 0.0201911        | 5.14  | 0.000  | 0.0642688 - 0.1434164|
| y98        | 0.0228328| 0.0041643        | 5.48  | 0.000  | 0.0146708 - 0.0309947|
| y99        | 0.0363819| 0.0051292        | 7.09  | 0.000  | 0.0263289 - 0.0464349|
| y00        | 0.0977717| 0.0055072        | 17.75 | 0.000  | 0.0869777 - 0.1085656|
| _cons      | 6.207889| 0.9118109        | 6.81  | 0.000  | 4.420773 - 7.995006  |
| sigma_u    | 0.3193384|                 |       |        |                     |
| sigma_e    | 0.10651186|                |       |        |                     |
| rho        | 0.89988885| (fraction of variance due to u_i) |       |        |                     |

. The coefficient on concen is the FE estimate.
5. The Robust Hausman Test

- FE removes within-group means. RE removes a fraction of the within-group means, with the fraction given by

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

estimated by \( \hat{\theta} \).

- In other words, the RE estimates can be obtained from the pooled OLS regression

\[
y_{it} - \hat{\theta} \bar{y}_i \text{ on } x_{it} - \hat{\theta} \bar{x}_i, t = 1, \ldots, T; i = 1, \ldots, N.
\]
\[ \hat{\theta} \approx 0 \Rightarrow \hat{\beta}_{RE} \approx \hat{\beta}_{POLS} \]
\[ \hat{\theta} \approx 1 \Rightarrow \hat{\beta}_{RE} \approx \hat{\beta}_{FE} \]

- If \( x_{it} \) includes time-constant variables \( z_i \), then \((1 - \hat{\theta})z_i \) appears as a regressor.

- Using the CRE setup it is easy to see why RE is more efficient than FE under the full set of RE assumptions: RE is true GLS and correctly imposes \( \xi = 0 \) in

\[ y_{it} = g_t \theta + z_i \delta + w_{it} \gamma + \psi + \tilde{w}_i \xi + v_{it} \]
Testing the Key RE Assumption

- Both RE and FE require $\text{Cov}(x_{is}, u_{it}) = 0$, all $s$ and $t$.
- RE adds the assumption $\text{Cov}(x_{it}, c_{i}) = 0$ for all $t$. With lots of good time-constant controls ("observed heterogeneity," such as industry dummies) might be able to make this condition roughly true. But we should test it.
1. The Traditional Hausman Test: Compare the coefficients on the time-varying explanatory variables, and compute a chi-square statistic. Cautions:
(i) Usual Hausman test maintains the second moment assumptions yet has no systematic power for detecting violations of these assumptions. (Analogy is using a standard $t$ statistic and thinking that, by itself, it has some potential to detect heteroskedasticity.)
(ii) With aggregate time effects, must use generalized inverse, and even then it is easy to get the degrees of freedom wrong. In Stata, must use the same estimated covariance matrix for both estimators to get proper df.
2. Variable Addition Test (VAT): Write the model as

\[ y_{it} = g_t \theta + z_i \delta + w_{it} \gamma + c_i + u_{it}. \]

- Obvious we cannot compare FE and RE estimates of \( \delta \) because the former is not defined.
- Less obvious (but true) that we cannot compare FE and RE estimates of \( \theta \).
- We can only compare \( \hat{\gamma}_{FE} \) and \( \hat{\gamma}_{RE} \). Let \( w_{it} \) be \( 1 \times J \).
• Use the CRE formulation

\[ y_{it} = g_i \theta + z_i \delta + w_{it} \gamma + \psi + \bar{w}_i \xi + a_i + u_{it}. \]

• Estimate this equation using RE and test \( H_0 : \xi = 0. \)

• Should make test fully robust to serial correlation in \( \{u_{it}\} \) and heteroskedasticity in \( a_i, \{u_{it}\} \).
• Conundrum: When FE and RE are similar, it does not really matter which we choose (although statistical significance can be an issue). When they differ by a lot and in a statistically significant way, RE is likely to be inappropriate.
• Classic pre-testing problem: Decision to include \( \bar{w}_i \) is made with error.
• Apply to airfare model:

  . * First use the Hausman test that maintains all of the RE assumptions under
  . * the null:

  . 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 ----
<table>
<thead>
<tr>
<th></th>
<th>(b)</th>
<th>(B)</th>
<th>(b-B)</th>
<th>sqrt(diag(V_b-V_B))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b_fe</td>
<td>b_re</td>
<td>Difference</td>
<td>S.E.</td>
</tr>
<tr>
<td>-----</td>
<td>-------</td>
<td>-------</td>
<td>------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>concen</td>
<td>.168859</td>
<td>.2089935</td>
<td>-.0401345</td>
<td>.0126937</td>
</tr>
<tr>
<td>y98</td>
<td>.0228328</td>
<td>.0224743</td>
<td>.0003585</td>
<td>.</td>
</tr>
<tr>
<td>y99</td>
<td>.0363819</td>
<td>.0366898</td>
<td>-.000308</td>
<td>.</td>
</tr>
<tr>
<td>y00</td>
<td>.0977717</td>
<td>.098212</td>
<td>-.0004403</td>
<td>.</td>
</tr>
</tbody>
</table>

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

\[ \text{chi2}(4) = (b-B)'[(V_b-V_B)^{-1}](b-B) \]
\[ = 10.00 \]
\[ \text{Prob}>\text{chi2} = 0.0405 \]
\[ (V_b-V_B \text{ is not positive definite}) \]

. di -.0401/.0127
-3.1574803

* This is the nonrobust H 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.
* 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.

<table>
<thead>
<tr>
<th></th>
<th>(b)</th>
<th>(B)</th>
<th>(b-B)</th>
<th>sqrt(diag(V_b-V_B))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b_fe</td>
<td>b_re</td>
<td>Difference</td>
<td>S.E.</td>
</tr>
<tr>
<td>concen</td>
<td>.168859</td>
<td>.2089935</td>
<td>-.0401345</td>
<td>.0127597</td>
</tr>
<tr>
<td>y98</td>
<td>.0228328</td>
<td>.0224743</td>
<td>.0003585</td>
<td>.000114</td>
</tr>
<tr>
<td>y99</td>
<td>.0363819</td>
<td>.0366898</td>
<td>-.000308</td>
<td>.0000979</td>
</tr>
<tr>
<td>y00</td>
<td>.0977717</td>
<td>.098212</td>
<td>-.0004403</td>
<td>.00014</td>
</tr>
</tbody>
</table>

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
```

53
. * 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)

| 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 | .0016419 | 5.14  | 0.000 | .0642688 .1434164    |
| y98   | .0228328 | .0016419 | 5.48  | 0.000 | .0146708 .0309947    |
| y99   | .0363819 | .0016419 | 5.48  | 0.000 | .0263289 .0464349    |
| y00   | .0977717 | .0016419 | 5.48  | 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.
* What if we do not control for distance in RE?

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

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

```
(Std. Err. adjusted for 1149 clusters in id)

|         | Coef. | Std. Err. | z   | P>|z| | [95% Conf. Interval] |
|---------|-------|-----------|-----|------|----------------------|
| lfare   |       |           |     |      |                      |
| concen  | 0.0468181 | 0.0427562 | 1.09 | 0.274 | -0.0369826 -0.1306188 |
| y98     | 0.0239229 | 0.0041907 | 5.71 | 0.000 | 0.0157093 0.0321364  |
| y99     | 0.0354453 | 0.0051678 | 6.86 | 0.000 | 0.0253167 0.045574  |
| y00     | 0.0964328 | 0.0055197 | 17.47 | 0.000 | 0.0856144 0.1072511 |
| _cons   | 5.028086  | 0.0285248 | 176.27 | 0.000 | 4.972178 5.083993  |

|         |       |           |     |      |                      |
| sigma_u | 0.40942871 |            |      |      |                      |
| sigma_e | 0.10651186 |            |      |      |                      |
| rho     | 0.93661309 | (fraction of variance due to u_i) |      |      |                      |
```

* The RE estimate is now way below the FE estimate, .169. Thus, it can be
* very harmful to omit time-constant variables in RE estimation.
6. Correlated Random Slopes

• Linear model but where all slopes may vary by unit:

\[ y_{it} = c_i + x_{it}b_i + u_{it} \]

\[ E(u_{it}|x_i, c_i, b_i) = 0, t = 1, \ldots, T, \]

where \( b_i \) is \( K \times 1 \).

• Question: What if we ignore the heterogeneity in the slopes and act as if \( b_i \) is constant all \( i \)? We apply the usual FE estimator that only eliminates \( c_i \).

• Define the average partial effect (population average effect) as \( \beta = E(b_i) \).
Wooldridge (2005, *REStat*): A sufficient condition for FE to consistently estimate $\beta$ is

$$E(b_i|\overline{x}_{it}) = E(b_i) = \beta, \quad t = 1, \ldots, T.$$ 

This condition allows the slopes, $b_i$, to be correlated with the regressors $x_{it}$ through permanent components. What it rules out is correlation between idiosyncratic movements in $x_{it}$. 
• For example, suppose \( x_{it} = f_i + r_{it}, t = 1, \ldots, T \) where \( f_i \) is the unit-specific “level” of the process and \( \{r_{it}\} \) are the deviations from this level. Because \( \bar{x}_{it} = \bar{r}_{it} \) it suffices that

\[
E(b_i | r_{i1}, r_{i2}, \ldots, r_{iT}) = E(b_i)
\]

• Note that any kind of serial correlation and changing variances/covariances are allowed in \( \{r_{it}\} \).
• Extension to random trend settings. Write

\[ y_{it} = w_t a_i + x_{it} b_i + u_{it}, \quad t = 1, \ldots, T \]

where \( w_t \) is a set of deterministic functions of time. Now the fixed effects estimator sweeps away \( a_i \) by netting out \( w_t \) from \( x_{it} \).

• The \( \tilde{x}_{it} \) are the residuals from the regression \( x_{it} \) on \( w_t \), \( t = 1, \ldots, T \).

• In the random trend model, \( w_t = (1, t) \), and so the elements of \( x_{it} \) have unit-specific linear trends removed in addition to a level effect.
• Removing more of the heterogeneity from \( \{ x_{it} \} \) makes it even more likely that the FE estimator is consistent. For example, if \( w_t = (1, t) \) and \( x_{it} = f_i + h_i t + r_{it} \), then \( b_i \) can be arbitrarily correlated with \( (f_i, h_i) \).

• Adding to \( w_t \) – such as polynomials in \( t \) – requires more time periods, and it decreases the variation in \( \ddot{x}_{it} \) compared to the usual FE estimator. Requires \( \text{dim}(w_t) < T \).
Modelling the Correlated Random Slopes

- Again consider the model

\[ y_{it} = c_i + x_{it}b_i + u_{it}, \quad t = 1, \ldots, T \]

where we assume

\[ b_i = \beta + \Gamma(\bar{x}_i - \mu_{\bar{x}})' + d_i \]

\[ E(d_i|x_i) = 0 \]
• After simple algebra:

\[ y_{it} = \theta_t + c_i + x_{it}\beta + \left[ (\bar{x}_i - \mu_{\bar{x}}) \otimes x_{it} \right] \gamma + x_{it}d_i + u_{it}, \]

• A test of \( H_0 : \gamma = 0 \) is simple to carry out.

1. Create interactions

\[ (\bar{x}_i - \bar{x}) \otimes x_{it} \]

where \( \bar{x} \) is the vector of overall averages, \( \bar{x} = N^{-1} \sum_{i=1}^{N} \bar{x}_i \). (Or use a subset of \( \bar{x}_i \) to interact with a subset of \( x_{it} \).)

2. Estimate the equation by FE (to account for \( c_i \)) and obtain a fully robust test of the interaction terms.
• We can also allow $b_i$ to depend on time constant observable variables:

$$b_i = \beta + \Gamma(\bar{x}_i - \mu_{\bar{x}})' + \Psi(h_i - \mu_h)' + d_i$$

where $h_i$ is a row vector of time-constant variables that we think might influence $b_i$.

• The new equation is

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

and we can test $H_0 : \gamma = 0, \psi = 0$ or a subset, using the fixed effects estimator (to remove $c_i$).
• Can also model $c_i$ using a CRE approach, which means add $\bar{x}_i$ and $h_i$ as separate regressors and estimate the model by random effects. This is common in the hierarchical linear models literature.

• Then estimate the equation

$$y_{it} = \theta_t + x_{it}\beta + \bar{x}_i\xi + h_i\zeta + [(\bar{x}_i - \bar{x}) \otimes x_{it}]\gamma + [(h_i - \bar{h}) \otimes x_{it}]\psi + \text{error}_{it}$$

using RE with fully robust inference.

**EXAMPLE**: Effects of Concentration on Airfares

• Interact $\text{concen}_{it}$ with the time average and the distance variables. Center about averages to obtain the average partial effects.
. egen concenb = mean(concen), by(id)

. sum concenb ldist ldistsq

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>concenb</td>
<td>4596</td>
<td>.6101149</td>
<td>.1888741</td>
<td>.1862</td>
<td>.9997</td>
</tr>
<tr>
<td>ldist</td>
<td>4596</td>
<td>6.696482</td>
<td>.6593177</td>
<td>4.553877</td>
<td>7.909857</td>
</tr>
<tr>
<td>ldistsq</td>
<td>4596</td>
<td>45.27747</td>
<td>8.726898</td>
<td>20.73779</td>
<td>62.56583</td>
</tr>
</tbody>
</table>

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

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

| lfare | Coef. | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|-------|-------|-----------|------|------|----------------------|
| concen| .1682492 | .0496695 | 3.39 | 0.001 | .0708988 - .2655996 |
| concenb| .157291 | .2085049 | 0.75 | 0.451 | -.2513711 .565953 |
| cbconcen| .0635453 | .3033809 | 0.21 | 0.834 | -.531073 .6581609 |
| ldconcen| -.2994869 | .9930725 | -0.30 | 0.763 | -2.245873 1.646899 |
| ldsqconcen| .0112477 | .0746874 | 0.15 | 0.880 | -.135137 .1576324 |
| ldist| -.4394368 | .6713288 | -0.65 | 0.513 | -1.755217 .8763435 |
| ldistsq| .0752147 | .0494201 | 1.52 | 0.128 | -.0216469 .1720764 |
| y98| .0229684 | .0041542 | 5.53 | 0.000 | .0148262 .031105 |
| y99| .0358549 | .0051298 | 6.99 | 0.000 | .0258007 .0459091 |
| y00| .0976256 | .005461 | 17.88 | 0.000 | .0869221 .108329 |
| _cons| 4.382552 | 2.272566 | 1.93 | 0.054 | -.0715953 8.836699 |
. test concenb cbconcen ldconcen ldsqconcen

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

    chi2( 4) =  14.02
    Prob > chi2 =  0.0072

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

. test cbconcen ldconcen ldsqconcen

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

    chi2( 3) =  5.47
    Prob > chi2 =  0.1407
. * Estimated coefficient on concen very close to omitting interactions:

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

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

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

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

(Std. Err. adjusted for 1149 clusters in id)

| Robust       | Coef.     | Std. Err. | z    | P>|z|   | [95% Conf. Interval] |
|--------------|-----------|-----------|------|-------|----------------------|
| lfare        |           |           |      |       |                      |
| concen       | 0.168859  | 0.0494749 | 3.41 | 0.001 | 0.07189  0.2658279   |
| concenb      | 0.2136346 | 0.0816403 | 2.62 | 0.009 | 0.0536227 0.3736466 |
| ldist        | -0.9089297| 0.2721637 | -3.34| 0.001 | -1.442361 -0.3754987|
| ldistsq      | 0.1038426 | 0.0201911 | 5.14 | 0.000 | 0.0642688 0.1434164 |
| y98          | 0.0228328 | 0.0041643 | 5.48 | 0.000 | 0.0146708 0.0309947 |
| y99          | 0.0363819 | 0.0051292 | 7.09 | 0.000 | 0.0263289 0.0464349 |
| y00          | 0.0977717 | 0.0055072 | 17.75| 0.000 | 0.0869777 0.1085656 |
| _cons        | 6.207889  | 0.9118109 | 6.81 | 0.000 | 4.420773  7.995006  |

sigma_u | .31933841
sigma_e | .10651186
rho     | .89988885 (fraction of variance due to u_i)