Panel Econometrics

Paulo M. M. Rodrigues

February 2022

Concluding the basics

1. Relation between RE, FE, and pooled OLS
2. Which estimator should one choose?
3. More on the Hausman test
4. Up next

Outline

Transformation associated with the FE estimator $\star$

An illuminating comparison of the different panel estimators is based on how they transform the data.

Recall that the FE estimator is based on the within transformation

$$\ddot{y}_{it} = y_{it} - \bar{y}_i \quad \text{and} \quad \ddot{x}_{it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_i.$$

The FE estimator simply applies OLS to the transformed equation

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

Thereby, any variation "between" the individuals, i.e., the variation between individuals' time averages, is neglected.

More within ⋆

It will be helpful to stack observations $t = 1, \ldots, T$ for individual i

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

and recall that the within transformation can be expressed as

$$\underbrace{\mathbf{Q}_T \mathbf{y}_i}_{\ddot{\mathbf{y}}_i} = \underbrace{\mathbf{Q}_T \mathbf{X}_i}_{\ddot{\mathbf{X}}_i} \boldsymbol{\beta} + \underbrace{\mathbf{Q}_T \mathbf{u}_i}_{\ddot{\mathbf{u}}_i} \qquad \text{because } \mathbf{Q}_T \boldsymbol{\iota}_T = 0.$$

Hence, the FE estimator is

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{FE} &= \left(\sum_{i=1}^{N} (\mathbf{Q}_T \mathbf{X}_i)'(\mathbf{Q}_T \mathbf{X}_i) \right)^{-1} \sum_{i=1}^{N} (\mathbf{Q}_T \mathbf{X}_i)'(\mathbf{Q}_T \mathbf{y}_i) \\
&= \left(\sum_{i=1}^{N} \ddot{\mathbf{X}}_i' \ddot{\mathbf{X}}_i \right)^{-1} \sum_{i=1}^{N} \ddot{\mathbf{X}}_i' \ddot{\mathbf{y}}_i .
\end{aligned}$$

Transformation associated with the between estimator ⋆

The between estimator is based on the time-averaged information that is discarded by the within transformation:

$$\bar{y}_i = \bar{\mathbf{x}}_i \boldsymbol{\beta} + \bar{v}_i.$$

In matrix notation,

$$\mathbf{J}_T \mathbf{y}_i = \mathbf{J}_T \mathbf{X}_i \boldsymbol{\beta} + \mathbf{J}_T \mathbf{v}_i \qquad \Leftrightarrow \qquad \bar{\mathbf{y}}_i = \bar{\mathbf{X}}_i \boldsymbol{\beta} + \bar{\mathbf{v}}_i$$

Applying OLS to this equation yields the between estimator

$$\begin{aligned} \hat{\boldsymbol{\beta}}_B &= \left(\sum_{i=1}^{N} (\mathbf{J}_T \mathbf{X}_i)' (\mathbf{J}_T \mathbf{X}_i) \right)^{-1} \sum_{i=1}^{N} (\mathbf{J}_T \mathbf{X}_i)' (\mathbf{J}_T \mathbf{y}_i) \\ &= \left(\sum_{i=1}^{N} \bar{\mathbf{X}}_i' \bar{\mathbf{X}}_i \right)^{-1} \sum_{i=1}^{N} \bar{\mathbf{X}}_i' \bar{\mathbf{y}}_i. \end{aligned}$$

The between estimator uses $\mathbf{J}_T$ rather than FE's $\mathbf{Q}_T$.

Transformation associated with the RE estimator $\star$

Recall that the RE estimator is (multiplying both nominator and denominator with $\hat{\sigma}_u^2$)

$$\hat{\boldsymbol{\beta}}_{RE} = \left(\sum_{i=1}^{N} \mathbf{X}_i' \hat{\sigma}_u^2 \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^{N} \mathbf{X}_i' \hat{\sigma}_u^2 \hat{\boldsymbol{\Omega}}^{-1} \mathbf{y}_i \right).$$

Employing GLS algebra, we have with $\hat{\boldsymbol{\Omega}}^{-1} = \hat{\boldsymbol{\Omega}}^{-\frac{1}{2}} \hat{\boldsymbol{\Omega}}^{-\frac{1}{2}}$ that

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{RE} &= \left(\sum_{i=1}^{N} (\hat{\sigma}_u \hat{\boldsymbol{\Omega}}^{-\frac{1}{2}} \mathbf{X}_i)' (\hat{\sigma}_u \hat{\boldsymbol{\Omega}}^{-\frac{1}{2}} \mathbf{X}_i) \right)^{-1} \left(\sum_{i=1}^{N} (\hat{\sigma}_u \hat{\boldsymbol{\Omega}}^{-\frac{1}{2}} \mathbf{X}_i)' (\hat{\sigma}_u \hat{\boldsymbol{\Omega}}^{-\frac{1}{2}} \mathbf{y}_i) \right) \\
&= \left(\sum_{i=1}^{N} \tilde{\mathbf{X}}_i' \tilde{\mathbf{X}}_i \right)^{-1} \left(\sum_{i=1}^{N} \tilde{\mathbf{X}}_i' \tilde{\mathbf{y}}_i \right)
\end{aligned}$$

of the transformed observations

$$\tilde{\mathbf{y}}_i = \hat{\sigma}_u \hat{\boldsymbol{\Omega}}^{-\frac{1}{2}} \mathbf{y}_i \quad \text{and} \quad \tilde{\mathbf{X}}_i = \hat{\sigma}_u \hat{\boldsymbol{\Omega}}^{-\frac{1}{2}} \mathbf{X}_i.$$

RE: quasi-demeaning ⋆

Let us define

$$\hat{\phi}^2 = \frac{\hat{\sigma}_u^2}{T\hat{\sigma}_c^2 + \hat{\sigma}_u^2}. \qquad \text{(Note: the textbook uses } \hat{\lambda} = 1 - \hat{\phi}.\text{)}$$

It can be show that $\hat{\sigma}_u \hat{\mathbf{\Omega}}^{-\frac{1}{2}} = [\mathbf{I}_T - (1-\hat{\phi})\mathbf{J}_T] = \mathbf{Q}_T + \hat{\phi}\mathbf{J}_T$ and thus

$$\begin{aligned}\tilde{\mathbf{y}}_i &= [\mathbf{I}_T - (1-\hat{\phi})\mathbf{J}_T]\mathbf{y}_i = \mathbf{y}_i - (1-\hat{\phi})\bar{\mathbf{y}}_i \qquad \leftarrow \text{“quasi demeaning”} \\ &= [\mathbf{Q}_T + \hat{\phi}\mathbf{J}_T]\mathbf{y}_i = \ddot{\mathbf{y}}_i + \hat{\phi}\bar{\mathbf{y}}_i\end{aligned}$$

and

$$\begin{aligned}\tilde{\mathbf{X}}_i &= [\mathbf{I}_T - (1-\hat{\phi})\mathbf{J}_T]\mathbf{X}_i = \mathbf{X}_i - (1-\hat{\phi})\bar{\mathbf{X}}_i \qquad \leftarrow \text{“quasi demeaning”} \\ &= [\mathbf{Q}_T + \hat{\phi}\mathbf{J}_T]\mathbf{X}_i = \ddot{\mathbf{X}}_i + \hat{\phi}\bar{\mathbf{X}}_i\end{aligned}$$

Hence, the RE estimator is equivalent to POLS applied to the quasi-demeaned obbservations!

Comparison of estimators and associated transformations

OLS:	regression of $\mathbf{y}_i$ on $\mathbf{X}_i$ $\rightarrow$ weights within and between variance equally
FE (within):	regression of $\ddot{\mathbf{y}}_i$ on $\ddot{\mathbf{X}}_i$ $\rightarrow$ uses only within variance
Between:	regression of $\bar{\mathbf{y}}_i$ on $\bar{\mathbf{X}}_i$ $\rightarrow$ uses only between variance
RE:	regression of $\mathbf{y}_i - (1-\hat{\phi})\bar{\mathbf{y}}_i$ on $\mathbf{X}_i - (1-\hat{\phi})\bar{\mathbf{X}}_i$
	$\rightarrow$ can be interpreted as combination of OLS and between estimators
	regression of $\ddot{\mathbf{y}}_i + \hat{\phi}\bar{\mathbf{y}}_i$ on $\ddot{\mathbf{X}}_i + \hat{\phi}\bar{\mathbf{X}}_i$
	$\rightarrow$ can be interpreted as combination of within and between estimators
	Important: combination is function of $\hat{\phi}$ and thus data-dependent
	$\rightarrow$ weights within and between variance in data-dependent way

And then some

Note now that

$$\hat{\boldsymbol{\beta}}_{RE} = \mathbf{W}_1 \hat{\boldsymbol{\beta}}_B + (\mathbf{I} - \mathbf{W}_1)\hat{\boldsymbol{\beta}}_{FE}$$

where

$$\mathbf{W}_1 = \left(\hat{\phi}^2 \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{J}_T \mathbf{X}_i + \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{Q}_T \mathbf{X}_i \right)^{-1} \hat{\phi}^2 \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{J}_T \mathbf{X}_i.$$

Alternatively, we can write

$$\hat{\boldsymbol{\beta}}_{RE} = \mathbf{W}_2 \hat{\boldsymbol{\beta}}_B + (\mathbf{I} - \mathbf{W}_2)\hat{\boldsymbol{\beta}}_{OLS}$$

where

$$\mathbf{W}_2 = \left([\hat{\phi}^2 - 1] \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{J}_T \mathbf{X}_i + \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{X}_i \right)^{-1} [\hat{\phi}^2 - 1] \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{J}_T \mathbf{X}_i.$$

RE as a function of within to between regressor dispersion

Discussion (for given $\hat{\phi}^2$):

- Between regressor dispersion dominates within regressor dispersion, $\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{J}_T \mathbf{X}_i \gg \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{Q}_T \mathbf{X}_i$:

  $\mathbf{W}_1 \approx \mathbf{I} \Rightarrow \hat{\boldsymbol{\beta}}_{RE} \approx \hat{\boldsymbol{\beta}}_B$

- Within regressor dispersion dominates between regressor dispersion, $\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{J}_T \mathbf{X}_i \ll \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{Q}_T \mathbf{X}_i$:

  $\mathbf{W}_1 \approx 0 \Rightarrow \hat{\boldsymbol{\beta}}_{RE} \approx \hat{\boldsymbol{\beta}}_{FE}$

RE as a function of within to between error variance

Write

$$\hat{\phi}^2 = \frac{\hat{\sigma}_u^2}{T\hat{\sigma}_c^2 + \hat{\sigma}_u^2} = \frac{1}{T\hat{\sigma}_c^2/\hat{\sigma}_u^2 + 1} = \frac{\hat{\sigma}_u^2/\hat{\sigma}_c^2}{T + \hat{\sigma}_u^2/\hat{\sigma}_c^2}$$

Discussion (for given regressor dispersion):

- Between to within error variance small, $\hat{\sigma}_c^2/\hat{\sigma}_u^2 \approx 0$:

  $\hat{\phi}^2 \approx 1 \Rightarrow \mathbf{W}_2 \approx 0 \Rightarrow \hat{\boldsymbol{\beta}}_{RE} \approx \hat{\boldsymbol{\beta}}_{OLS}$

- Within to between error variance small, $\hat{\sigma}_u^2/\hat{\sigma}_c^2 \approx 0$:

  $\hat{\phi}^2 \approx 0 \Rightarrow \mathbf{W}_1 \approx 0 \Rightarrow \hat{\boldsymbol{\beta}}_{RE} \approx \hat{\boldsymbol{\beta}}_{FE}$

Efficiency of the RE estimator

Based on the previous results we can understand why the RE estimator (if it is consistent) is more efficient than the FE estimator (and also the between estimator).

- It optimally adjusts to the within versus between regressor dispersion.
- It optimally adjusts to the within versus between error variance.

In contrast, the FE estimator neglects any between variance and thus is only as good as the RE estimator if the between dimension is negligible.

(But recall the exogeneity assumptions...)

Outline

FE estimator or FD estimator?

The assumptions of the FE and FD estimators differ in one respect:

- FE: u_{it} is white noise over t
- FD: u_{it} is a random walk (at least under classical error assumptions)

Choose according to what is more likely.

Often, reality is in between: there is some serial correlation but not as much as predicted by the random walk assumption. In these cases, it might be helpful to apply both estimators with robust variance matrix $\rightarrow$ both point estimators and SE's are consistent and the differences should be relatively small.

If the differences are large, the strict exogeneity assumption may be invalid (for a test see next page).

Testing for strict exogeneity

Recall that strict exogeneity means $\mathrm{E}(y_{it}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT},c_i) = \mathrm{E}(y_{it}|\mathbf{x}_{it},c_i)$.

A simple test for $T \geq 3$ is to apply FE or FD to the augmented regression

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

where $\mathbf{w}_{i,t+1}$ is a subset of $\mathbf{x}_{i,t+1}$.

Under the null of strict exogeneity, $\boldsymbol{\gamma} = \mathbf{0}$. This can be tested with a simple F test.

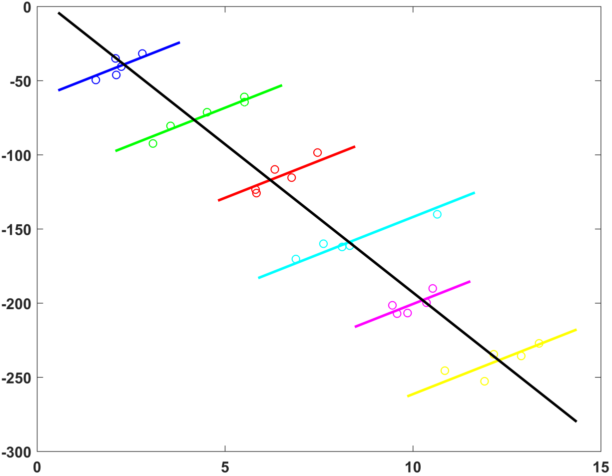
FE estimator or RE estimator?

The main difference between FE and RE assumptions is whether the c_i are allowed to correlate with $\mathbf{x}_i$ or not.

- Consistency of the RE estimator requires this correlation be zero. We have discussed many examples where this assumption is likely to fail. In such cases, you should not use the RE estimator.
- Instead, when c_i is expected to correlate with $\mathbf{x}_i$ you should use the FE (or FD) estimator which is still consistent.
- If you are unsure, you can use a Hausman test to compare FE and RE estimators (see below). Beware of effects of such pretests, though.
- Given the robustness of the FE estimator with respect to the question whether the c_i are allowed to correlate with $\mathbf{x}_i$, it is natural to ask why one should use the RE estimator at all. The answer is efficiency (see above).

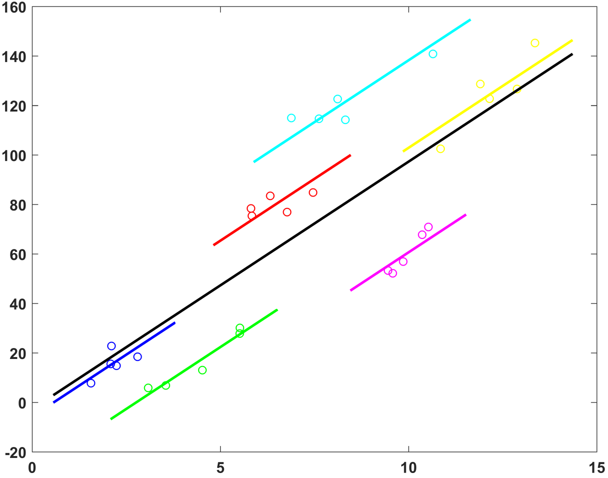
Data example: pooled OLS / RE estimator inconsistent

Positive partial effect of x_i on y_i but c_i negatively correlated with x_i



Data example: pooled OLS / RE estimator consistent

Positive partial effect of x_i on y_i and c_i uncorrelated with x_i



Hausman testing principle

The Hausman test is a general testing principle that compares two estimators $\hat{\boldsymbol{\beta}}_A$ and $\hat{\boldsymbol{\beta}}_B$.

Under the null, both estimators are consistent but only $\hat{\boldsymbol{\beta}}_B$ is efficient, i.e., $\text{Avar}(\hat{\boldsymbol{\beta}}_A) - \text{Avar}(\hat{\boldsymbol{\beta}}_B) > 0$.

Under the alternative, $\hat{\boldsymbol{\beta}}_B$ is inconsistent while $\hat{\boldsymbol{\beta}}_A$ remains consistent.

Under general conditions, the Hausman statistic

$$(\hat{\boldsymbol{\beta}}_A - \hat{\boldsymbol{\beta}}_B)'[\widehat{\text{Avar}}(\hat{\boldsymbol{\beta}}_A) - \widehat{\text{Avar}}(\hat{\boldsymbol{\beta}}_B)]^{-1}(\hat{\boldsymbol{\beta}}_A - \hat{\boldsymbol{\beta}}_B) \xrightarrow{\text{d}} \chi^2_r,$$

where r is the number of parameters.

Hausman test to compare FE and RE estimators

We want to test $H_0 : c_i$ uncorrelated with $\mathbf{x}_i$ versus $H_1 : \neg H_0$.

Under the null both RE and FE estimators are consistent but RE is more efficient, while under the alternative only FE is consistent.

Suppose the strict exogeneity, invertibility and homoskedasticity/white noise assumptions (RE.1a, RE.2, RE.3) hold throughout. Further assume the regressors do not include variables that vary solely across t (such as time dummies). Will return to this.

If the regressors include time-invariant variables, their parameters are not identified by the FE estimator. Hence, only the parameters of the time-varying regressors can be compared (Stata automatically excludes the other). In the following, for simplicity we assume $\mathbf{x}_{it}$ includes only variables that vary both with i and t.

Concretely

In terms of assumption RE.1 (b)-(c) the hypotheses are:

H_0: $\mathrm{E}(c_i|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}) = \mathrm{E}(c_i) = 0$ vs. H_1: $\mathrm{E}(c_i|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}) \neq \mathrm{E}(c_i)$

The classical Hausman statistic:

$$H = (\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE})'[\widehat{\mathrm{Avar}}(\hat{\boldsymbol{\beta}}_{FE}) - \widehat{\mathrm{Avar}}(\hat{\boldsymbol{\beta}}_{RE})]^{-1}(\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE}) \xrightarrow{\mathrm{d}} \chi^2_K,$$

where K is the number of parameters (= the length of the vectors $\hat{\boldsymbol{\beta}}_{FE}$ and $\hat{\boldsymbol{\beta}}_{RE}$).

The null is rejected if H exceeds the critical value derived from the χ^2_K distribution.

Hausman test – implementation

The tricky thing is estimating the difference between the FE and RE (homoskedastic) variance matrices,

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

In finite samples, this difference may not be positive definite.

To mitigate this problem, Wooldridge (p. 331) suggests to use the same estimator of σ_u^2 to estimate the FE variance

$$\hat{\sigma}_u^2 \left\{ \sum_{i=1}^{N} \ddot{\mathbf{X}}_i' \ddot{\mathbf{X}}_i \right\}^{-1}$$

and the RE variance

$$\left\{ \sum_{i=1}^{N} \mathbf{X}_i' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X}_i \right\}^{-1}.$$

Remarks

The Hausman test has some important details:

- It is (both under the null and the alternative) based on the strict exogeneity assumption. If this assumption fails, the plims of the RE and FE estimators will generally differ and the test will tend to reject.
- It is—at least in the conventional form presented here—based on the assumptions RE.3. If this assumption fails, the asymptotic χ^2 distribution will not hold and test decisions will be biased. (But a robust form is available, see below.)

... there is more

In addition, the Hausman test can only compare estimators of regressors that vary both with i and t:

- The parameters of time-invariant regressors are not identified by the FE estimator und thus cannot be compared to the RE estimator.
- The parameters of regressors that vary solely across t (such as time dummies) have the same asymptotic variance when estimated by FE or RE. Hence, the test cannot distinguish the two estimation approaches.
- Fortunately, Stata will automatically apply the Hausman test only to those parameters that are eligible.

Note: Be sure that K is only the number of regressors that vary across both i and t. Regressors that are time-invariant or vary solely across t are excluded! (Again, Stata...)

Hausman test – Stata

You first have to tell Stata that you have panel data:

```
xtset id year
```

FE estimator with classical variance matrix is computed and stored as “fixed”:

```
xtreg y x1 x2 x3, fe
estimates store fixed
```

RE estimator with classical variance matrix is computed:

```
xtreg y x1 x2 x3, re
```

The Hausman test is computed based on the more efficient RE estimate of σ_u^2:

```
hausman fixed ., sigmamore
```

The Hausman test is computed based on the less efficient FE estimate of σ_u^2:

```
hausman fixed ., sigmaless
```

Example: Effects of job training grants on scrap rates

Example 10.4 taken from Wooldridge's textbook

Note: regression includes two time dummies and one time-invariant variable (union)!

```
*** load data and set panel ***
use "jtrain1.dta", clear
xtset fcode year

*** FE regression and store result (Stata skips union) ***
xtreg lscrap d88 d89 union grant grant_1, fe
estimates store fixed

*** run RE regression ***
xtreg lscrap d88 d89 union grant grant_1, re

*** compute test based on efficient estimate of Var(u) ***
hausman fixed ., sigmamore
```

Stata output

```
. hausman fixed ., sigmamore

Note: the rank of the differenced variance matrix (2) 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))
             |     fixed          .          Difference          S.E.
-------------+----------------------------------------------------------------
         d88 |   -.0802157    -.0934519        .0132363        .0107669
         d89 |   -.2472028    -.2698336        .0226308        .0219962
       grant |   -.2523149     -.214696       -.0376188        .0306006
     grant_1 |   -.4215895    -.3770698       -.0445197        .0467426
------------------------------------------------------------------------------
                           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(2) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        2.59
                Prob>chi2 =      0.2738
                (V_b-V_B is not positive definite)
```

Outline

Hausman variable addition test

Under maintained assumption RE.3, it can be shown that the Hausman statistic can also be obtained from estimating the augmented equation

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \bar{\mathbf{x}}_i\boldsymbol{\delta} + \varepsilon_{it}$$

by means of the RE estimator and computing the Wald statistic for exclusion of $\bar{\mathbf{x}}_i$:

$$W = \hat{\boldsymbol{\delta}}' \left[\widehat{\mathrm{Avar}}(\hat{\boldsymbol{\delta}})\right]^{-1} \hat{\boldsymbol{\delta}}.$$

The Hausman statistic is identical to this Wald statistic.

You will be asked in the exercises to verify this claim.

Hausman variable addition test with general regressors

A nice feature of the variable addition test is that we can use it even if we include regressors that do not vary across both i and t.

Let us split the regressor set into

- $\mathbf{x}_{it}$ which vary across both i and t,
- $\mathbf{z}_t$ which vary only across t (e.g., time dummies), and
- $\mathbf{h}_i$ which vary only across i (e.g., gender or race dummies).

Then the structural equation is written as

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{z}_t\boldsymbol{\gamma} + \mathbf{h}_i\boldsymbol{\theta} + v_{it}.$$

Since we can only compare the RE and FE estimators of $\boldsymbol{\beta}$, the augmented regression is

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{z}_t\boldsymbol{\gamma} + \mathbf{h}_i\boldsymbol{\theta} + \bar{\mathbf{x}}_i\boldsymbol{\delta} + v_{it}.$$

Estimating this equation by RE and computing the Wald statistic for $H_0 : \boldsymbol{\delta} = \mathbf{0}$ yields again the Hausman statistic.

Hausman variable addition test with general covariance

The variable addition test can even be used when assumption RE.3 does not hold.

In this case, we estimate the augmented equation

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{z}_t\boldsymbol{\gamma} + \mathbf{h}_i\boldsymbol{\theta} + \bar{\mathbf{x}}_i\boldsymbol{\delta} + v_{it}$$

again by RE but now compute a robust estimator of $\mathrm{Var}(\hat{\boldsymbol{\beta}}_{RE})$.

Based on this robust variance estimator, we compute the Wald statistic for $H_0 : \boldsymbol{\delta} = \mathbf{0}$.

This yields a robust version of the Hausman statistic.

Further insights from the variable addition test

Let us consider the structural equation

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

Now split $\mathbf{x}_{it} = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) + \bar{\mathbf{x}}_i$ and rewrite the equation accordingly:

$$y_{it} = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\boldsymbol{\beta} + \mathbf{z}_t\boldsymbol{\gamma} + \mathbf{h}_i\boldsymbol{\theta} + \bar{\mathbf{x}}_i\boldsymbol{\beta} + c_i + u_{it}.$$

In this equation, only $\mathbf{h}_i$ and $\bar{\mathbf{x}}_i$ can correlate with c_i. This implies:

- Without controls $\mathbf{h}_i$, the Hausman null is $H_0 : \mathrm{Corr}(\bar{\mathbf{x}}_i, c_i) = 0$.
- With controls $\mathbf{h}_i$, the Hausman null is $H_0 : \mathrm{Corr}(\bar{\mathbf{x}}_i - \mathsf{L}(\bar{\mathbf{x}}_i|\mathbf{h}_i), c_i) = 0$, where $\bar{\mathbf{x}}_i - \mathsf{L}(\bar{\mathbf{x}}_i|\mathbf{h}_i)$ is the linear projection error (the part of $\mathbf{x}_i$ that is left after controlling for $\mathbf{h}_i$).
- Hence, with a rich set of individual-specific controls $\mathbf{h}_i$, it is possible for $\bar{\mathbf{x}}_i - \mathsf{L}(\bar{\mathbf{x}}_i|\mathbf{h}_i)$ to be uncorrelated with c_i even though $\bar{\mathbf{x}}_i$ is correlated with c_i.
- Practical consequence: include many controls $\mathbf{h}_i$!

And another view

Compare the structural equation

$$y_{it} = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\boldsymbol{\beta} + \mathbf{z}_t\boldsymbol{\gamma} + \mathbf{h}_i\boldsymbol{\theta} + \bar{\mathbf{x}}_i\boldsymbol{\beta} + c_i + u_{it}.$$

with the augmented equation

$$y_{it} = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\boldsymbol{\beta} + \mathbf{z}_t\boldsymbol{\gamma} + \mathbf{h}_i\boldsymbol{\theta} + \bar{\mathbf{x}}_i\underbrace{(\boldsymbol{\delta} + \boldsymbol{\beta})}_{\boldsymbol{\kappa}} + c_i + u_{it}.$$

First note that the coefficient of $\mathbf{x}_{it} - \bar{\mathbf{x}}_i$ will be estimated consistently by RE, $\hat{\boldsymbol{\beta}}_{RE} \xrightarrow{\mathrm{p}} \boldsymbol{\beta}$, because $\mathbf{x}_{it} - \bar{\mathbf{x}}_i$ does not correlate with c_i.

- If the Hausman null $H_0 : \mathrm{Corr}(\bar{\mathbf{x}}_i - \mathsf{L}(\bar{\mathbf{x}}_i|\mathbf{h}_i), c_i) = 0$ holds, then the RE estimator of $\bar{\mathbf{x}}_i$ will also converge to $\boldsymbol{\beta}$: $\hat{\boldsymbol{\kappa}}_{RE} \xrightarrow{\mathrm{p}} \boldsymbol{\beta}$ and thus $\hat{\boldsymbol{\delta}}_{RE} \xrightarrow{\mathrm{p}} \mathbf{0}$. The null is thus equivalent to $H_0 : \boldsymbol{\delta} = \mathbf{0}$.
- If the null does not hold, then the RE estimator of $\bar{\mathbf{x}}_i$ will not converge to $\boldsymbol{\beta}$, and thus $\hat{\boldsymbol{\delta}}_{RE}$ will not converge to $\mathbf{0}$. (The correlation with the disturbance – here: c_i – leads to asymptotic bias in $\hat{\boldsymbol{\kappa}}_{RE}$.)

Hausman variable addition test – Stata

Classical Hausman test:

```
xtreg y x1 x2 z1 z2 h1 h2, fe
estimates store fixed
xtreg y x1 x2 z1 z2 h1 h2, re
hausman fixed ., sigmaless
```

Compute one time average per individual (assume x1 and x2 vary with i and t):

```
by id, sort:  egen x1bar = mean(x1)
by id, sort:  egen x2bar = mean(x2)
```

Classical Hausman variable addition test:

```
xtreg y x1 x2 z1 z2 h1 h2 x1bar x2bar, re
test x1bar x2bar
```

Robust Hausman variable addition test:

```
xtreg y x1 x2 z1 z2 h1 h2 x1bar x2bar, re vce(robust)
test x1bar x2bar
```

Example: Effects of job training grants on scrap rates

Example 10.4 taken from Wooldridge's textbook

Question: shall we use RE or FE?

Note: regression includes two time dummies (d88 and d89) and one time-invariant variable (union).

```
*** load data and set panel ***
use "jtrain1.dta", clear
xtset fcode year

*** compute Hausman test based efficient estimate of Var(u)
***
xtreg lscrap d88 d89 union grant grant_1, fe
estimates store fixed
xtreg lscrap d88 d89 union grant grant_1, re
hausman fixed ., sigmaless
```

Example continued

```
*** compute time averages ***
by fcode, sort:  egen gm = mean(grant)
by fcode, sort:  egen gm_1 = mean(grant_1)

*** classical Hausman variable addition test ***
xtreg lscrap d88 d89 union grant grant_1 gm gm_1, re
test gm gm_1

*** robust Hausman variable addition test ***
xtreg lscrap d88 d89 union grant grant_1 gm gm_1, re
vce(robust)
test gm gm_1
```

```
. hausman fixed ., sigmaless

Note: the rank of the differenced variance matrix (2) 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))
             |     fixed          .          Difference          S.E.
-------------+----------------------------------------------------------------
         d88 |   -.0802157    -.0934519        .0132363        .0107462
         d89 |   -.2472028    -.2698336        .0226308        .0219539
       grant |   -.2523149     -.214696       -.0376188        .0305419
     grant_1 |   -.4215895    -.3770698       -.0445197        .0466528
------------------------------------------------------------------------------
                           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(2) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        2.60
                Prob>chi2 =      0.2724
                (V_b-V_B is not positive definite)
```

```
. xtreg lscrap d88 d89 union grant grant_1 gm gm_1, re

Random-effects GLS regression                   Number of obs      =       162
Group variable: fcode                           Number of groups   =        54

R-sq:  within  = 0.2010                         Obs per group: min =         3
       between = 0.0658                                        avg =       3.0
       overall = 0.0780                                        max =         3

                                                Wald chi2(7)       =     29.69
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0001

------------------------------------------------------------------------------
      lscrap |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         d88 |  -.0802157   .1094751    -0.73   0.464     -.294783    .1343516
         d89 |  -.2472028   .1332183    -1.86   0.064    -.5083058    .0139003
       union |   .6391308   .4224104     1.51   0.130    -.1887784     1.46704
       grant |  -.2523149    .150629    -1.68   0.094    -.5475423    .0429125
     grant_1 |  -.4215895      .2102    -2.01   0.045    -.8335739   -.0096051
          gm |   2.429889   1.633967     1.49   0.137    -.7726267    5.632405
        gm_1 |  -.9712716   1.725502    -0.56   0.574    -4.353193     2.41065
       _cons |   .0633248   .3299351     0.19   0.848    -.5833361    .7099858
-------------+----------------------------------------------------------------
     sigma_u |  1.3900287
     sigma_e |  .49774421
         rho |  .88634984   (fraction of variance due to u_i)
------------------------------------------------------------------------------

. test gm gm_1

 ( 1)  gm = 0
 ( 2)  gm_1 = 0

           chi2(  2) =    2.60
         Prob > chi2 =    0.2724
```

```
. xtreg lscrap d88 d89 union grant grant_1 gm gm_1, re vce(robust)

Random-effects GLS regression                   Number of obs      =       162
Group variable: fcode                           Number of groups   =        54

R-sq:  within  = 0.2010                         Obs per group: min =         3
       between = 0.0658                                        avg =       3.0
       overall = 0.0780                                        max =         3

                                                Wald chi2(7)       =     30.20
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0001

                                 (Std. Err. adjusted for 54 clusters in fcode)
```

lscrap	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
d88	-.0802157	.0987892	-0.81	0.417	-.2738389	.1134076
d89	-.2472028	.1986894	-1.24	0.213	-.6366268	.1422212
union	.6391308	.4044357	1.58	0.114	-.1535487	1.43181
grant	-.2523149	.1448303	-1.74	0.081	-.536177	.0315472
grant_1	-.4215895	.2851984	-1.48	0.139	-.980568	.137389
gm	2.429889	1.534751	1.58	0.113	-.5781676	5.437946
gm_1	-.9712716	1.286396	-0.76	0.450	-3.492561	1.550017
_cons	.0633248	.4352352	0.15	0.884	-.7897204	.9163701
sigma_u	1.3900287					
sigma_e	.49774421					
rho	.88634984	(fraction of variance due to u_i)				

```
. test gm gm_1

 ( 1)  gm = 0
 ( 2)  gm_1 = 0

           chi2(  2) =    2.51
         Prob > chi2 =    0.2844
```

Outline

Coming up

Specification tests for panel data models