Panel Econometrics

Paulo M. M. Rodrigues

February 2022

Today's outline

Estimation of linear panel data models II

1. The first difference estimator
2. The between estimator
3. The random effects estimator
4. Up next

Outline

FD estimator

Another way to get rid of c_i is to difference the model

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

yielding

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

The FD estimator applies pooled OLS to the differenced equation:

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{FD} &= \left(\sum_{i=1}^{N} \Delta \mathbf{X}_i' \Delta \mathbf{X}_i \right)^{-1} \sum_{i=1}^{N} \Delta \mathbf{X}_i' \Delta \mathbf{y}_i \\
&= \left(\sum_{i=1}^{N} \sum_{t=2}^{T} \Delta \mathbf{x}_{it}' \Delta \mathbf{x}_{it} \right)^{-1} \sum_{i=1}^{N} \sum_{t=2}^{T} \Delta \mathbf{x}_{it}' \Delta y_{it}.
\end{aligned}$$

Assumptions

Assumption FD.1 (strict exogeneity): $\mathrm{E}(u_{it}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT},c_i)=0$ for all $t=1,\ldots,T$.

Assumption FD.2 (invertibility): rank $\sum_{t=2}^{T}\mathrm{E}(\Delta\mathbf{x}_{it}'\Delta\mathbf{x}_{it})=K$.

Discussion:

- FD.1 implies $\mathrm{E}(\Delta u_{it}|\Delta\mathbf{x}_{i1},\ldots,\Delta\mathbf{x}_{iT},c_i)=0$ for all $t=1,\ldots,T$.
- Hence, $\mathrm{E}(\Delta\mathbf{x}_{it}'\Delta u_{it})=0$ for all t which guarantees consistency of the FD estimator.
- FD.2 guarantees invertibility (and requires to exclude time-invariant regressors).

Estimator of the variance matrix

Under assumptions FD.1 and FD.2, the asymptotic distribution is

$$\sqrt{N}\left(\hat{\boldsymbol{\beta}}_{FD} - \boldsymbol{\beta}\right) \xrightarrow{\mathrm{d}} \text{Normal}\left(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\right),$$

where

$$\mathbf{A} \equiv \mathrm{E}(\Delta\mathbf{X}_i'\Delta\mathbf{X}_i) \quad \text{and} \quad \mathbf{B} \equiv \mathrm{E}\left(\Delta\mathbf{X}_i'\mathbf{e}_i\mathbf{e}_i'\Delta\mathbf{X}_i\right).$$

Like always, the variance estimator replaces population moments by sample averages.

And an unplausible extra assumption

Assumption FD.3 (error variance): $\mathrm{E}(\mathbf{e}_i\mathbf{e}_i'|\mathbf{x}_{i1},\ldots,\mathbf{x}_{it},c_i) = \sigma_e^2\mathbf{I}_{T-1}$, where $\mathbf{e}_i$ is the $(T-1)\times 1$ vector containing $e_{it}\equiv \Delta u_{it}$, $t=2,\ldots,T$.

If one is willing to additionally make assumption FD.3, the asymptotic variance matrix simplifies, since

$$\mathbf{B} \equiv \sigma_e^2\,\mathrm{E}(\Delta\mathbf{X}_i'\Delta\mathbf{X}_i),$$

which is estimated in a straightforward way using the FD residuals.

But FD.3 says that $\Delta u_{it} = u_{it} - u_{it-1} = e_{it}$ serially uncorrelated, implying that $u_{it} = u_{it-1} + e_{it}$ is a random walk.

Policy analysis using the FD estimator

Consider a two-period program evaluation like before.

Assume that participation starts in period 2.

- Hence, $\texttt{prog}_{i1} = 0$ for all i.
- In period 2, $\texttt{prog}_{i2} = 1$ for the treatment group and $\texttt{prog}_{i2} = 0$ for the control group.

The structural equation is

$$y_{it} = \alpha + \theta_2 d_t^{(2)} + \mathbf{z}_{it}\gamma + \delta\texttt{prog}_{it} + c_i + u_{it},$$

- where $\mathbf{z}_{it}$ are observable time-varying characteristics,
- c_i summarizes time-invariant variables that might be correlated with $\mathbf{x}_{it}$ (e.g., due to self selection into the program), and
- $d_t^{(2)}$ is a time dummy for period 2 to control for macroeconomic developments. (More on time dummies when we meet the two-way error model.)

… and a nice interpretation

First differencing leaves us with a single cross section:

$$\Delta y_{i2} = \theta_2 + \Delta \mathbf{z}_{i2}\boldsymbol{\gamma} + \delta \mathtt{prog}_{i2} + \Delta u_{i2},$$

where we use $\Delta d_t^{(2)} = 1$ and $\Delta \mathtt{prog}_{i2} = \mathtt{prog}_{i2}$.

For a moment (but not in real life), let us neglect observable individual characteristics:

$$\Delta y_{i2} = \theta_2 + \delta \mathtt{prog}_{i2} + \Delta u_{i2}.$$

Note that $\mathtt{prog}_{i2}$ is a simple 0-1 dummy. Applying OLS thus yields

$$\hat{\theta}_2 = \overline{\Delta y}_{control} \xrightarrow{\text{p}} \mathrm{E}(\Delta y_{control})$$

and

$$\hat{\delta} = \overline{\Delta y}_{treat} - \overline{\Delta y}_{control} \xrightarrow{\text{p}} \mathrm{E}(\Delta y_{treat}) - \mathrm{E}(\Delta y_{control}).$$

The latter is a basic difference-in-difference estimator: it measures how the changing outcomes of the treated and non-treated differs.

Outline

A sideline remark

The between estimator considers only the time-averaged information that is discarded by the within transformation:

$$\bar{y}_i = \bar{\mathbf{x}}_i \boldsymbol{\beta} + c_i + \bar{u}_i, \qquad i = 1, \ldots, N.$$

Since this estimator uses only N of all NT observations, it is inefficient.

In addition, it requires not only strict exogeneity but also uncorrelatedness of c_i and $\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}$ which was not necessary for the FE estimator. (Sufficient conditions for consistency are RE.1 discussed below plus an invertibility condition.)

Hence, it is rather an auxiliary estimator. We discuss it briefly because the RE estimator introduced below can be shown to be a linear combination of the FE estimator and the between estimator.

Details

OLS applied to the $i = 1, \ldots, N$ observations (one time average per individual) of this equation yields the between estimator

$$\hat{\boldsymbol{\beta}}_B = \left(\sum_{i=1}^{N} \bar{\mathbf{x}}_i' \bar{\mathbf{x}}_i \right)^{-1} \sum_{i=1}^{N} \bar{\mathbf{x}}_i' \bar{y}_i.$$

Note again that this estimator neglects any variation "within" the individuals.

If the regressors are uncorrelated with the two error components, the between estimator is consistent and asymptotically normal as can be shown using standard arguments.

A different representation ⋆

Let us stack observations $t = 1, \dots, T$ for individual i

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

and apply the between transformation (i.e. time averaging)

$$\underbrace{\mathbf{J}_T \mathbf{y}_i}_{\bar{\mathbf{y}}_i} = \underbrace{\mathbf{J}_T \mathbf{X}_i}_{\bar{\mathbf{X}}_i} \boldsymbol{\beta} + \underbrace{\mathbf{J}_T \boldsymbol{\iota}_T}_{\boldsymbol{\iota}_T} c_i + \underbrace{\mathbf{J}_T \mathbf{u}_i}_{\bar{\mathbf{u}}_i},$$

where

$$\bar{\mathbf{y}}_i = T^{-1} \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} y_{i,1} \\ \vdots \\ y_{i,T} \end{bmatrix} = \begin{bmatrix} \bar{y}_i \\ \vdots \\ \bar{y}_i \end{bmatrix} = \boldsymbol{\iota}_T \bar{y}_i$$

$$\bar{\mathbf{X}}_i = T^{-1} \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} x_{i,11} & \dots & x_{i,K1} \\ \vdots & \ddots & \vdots \\ x_{i,1T} & \dots & x_{i,KT} \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{x}}_i \\ \vdots \\ \bar{\mathbf{x}}_i \end{bmatrix} = \boldsymbol{\iota}_T \bar{\mathbf{x}}_i$$

… and its consequences ⋆

Hence, the transformed equation

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

includes T identical observations per individual. Applying OLS to it again yields the between estimator (now based on $N \cdot T$ observations)

$$\begin{aligned}\left(\sum_{i=1}^{N}\bar{\mathbf{X}}_i'\bar{\mathbf{X}}_i\right)^{-1}\sum_{i=1}^{N}\bar{\mathbf{X}}_i'\bar{\mathbf{y}}_i &= \left(\sum_{i=1}^{N}\bar{\mathbf{x}}_i'\underbrace{\boldsymbol{\iota}_T'\boldsymbol{\iota}_T}_{T}\bar{\mathbf{x}}_i\right)^{-1}\sum_{i=1}^{N}\bar{\mathbf{x}}_i'\underbrace{\boldsymbol{\iota}_T'\boldsymbol{\iota}_T}_{T}\bar{y}_i \\ &= \left(\sum_{i=1}^{N}\bar{\mathbf{x}}_i'\bar{\mathbf{x}}_i\right)^{-1}\sum_{i=1}^{N}\bar{\mathbf{x}}_i'\bar{y}_i = \hat{\boldsymbol{\beta}}_B\end{aligned}$$

Estimator of the error variance

The variance estimator of the between regression may be used by the RE estimator discussed below. This is why it pays off to consider it here.

Recall the time-averaged equation is

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

with error variance

$$\mathrm{Var}(\bar{v}_i) = \mathrm{Var}(c_i) + \mathrm{Var}(\bar{u}_i) = \sigma_c^2 + \sigma_u^2/T.$$

Now, OLS applied to the time-averaged equation (i.e., the between estimator) yields the residuals $\hat{\bar{v}}_i$ on which a (consistent) error variance estimator is based:

$$\widehat{\mathrm{Var}(\bar{v}_i)} = \hat{\sigma}_c^2 + \hat{\sigma}_u^2/T = \frac{1}{N-K} \sum_{i=1}^{N} \hat{\bar{v}}_i^2.$$

Outline

Assumptions

Assumption RE.1:

a. $\mathrm{E}(u_{it}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT},c_i)=0$ for all $t=1,\ldots,T$
b. $\mathrm{E}(c_i|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT})=\mathrm{E}(c_i)$
c. $\mathrm{E}(c_i)=0$

Discussion:

- Part (a) is strict exogeneity as discussed above.
- Part (b) rules out correlation between c_i and any $\mathbf{x}_{it}$, $t=1,\ldots,T$. Hence, the "omitted variable problem" used to motivate panel analysis cannot be handled here. Hence, the RE estimator is less robust than the FE estimator (but potentially efficient).
- Part (c) is without loss of generality if we include an intercept.

Generalized LS

Under assumption RE.1 we have

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

with

$$\mathrm{E}(v_{it}|\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}) = 0, \qquad t = 1, \ldots, T.$$

Stacking observations $t = 1, \ldots, T$ for individual i yields

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

We have already seen that the unconditional variance matrix of $\mathbf{v}_i$,

$$\boldsymbol{\Omega} \equiv \mathrm{E}(\mathbf{v}_i\mathbf{v}_i'),$$

is nondiagonal. It thus suggests GLS estimation.

... and an assumption

Assumption RE.2:

(a) rank $\mathrm{E}(\mathbf{X}_i'\boldsymbol{\Omega}\mathbf{X}_i) = K$

Discussion:

- This is just the usual rank condition for GLS estimation.
- Under RE.1 and RE.2, GLS and FGLS that uses an unrestricted variance estimator of $\boldsymbol{\Omega}$ is consistent and asymptotically normal as $N \to \infty$.
- But it would be inefficient because an unrestricted estimator of $\boldsymbol{\Omega}$ would need to estimate $T(T+1)/2$ parameters while the classical random effects assumptions are much stricter as we present next.

… and another one

Assumption RE.3:

(a) $\mathrm{E}(\mathbf{u}_i\mathbf{u}_i'|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT},c_i)=\sigma_u^2\mathbf{I}_T$

(b) $\mathrm{E}(c_i^2|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT})=\sigma_c^2$

Discussion:

- Part (a) implies $\mathrm{E}(u_{it}^2|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT},c_i)=\sigma_u^2$ and $\mathrm{E}(u_{it}u_{is}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT},c_i)=0$ if $t\neq s$.
- By the law of iterated expectations, it also implies

$$\mathrm{E}(u_{it}^2)=\sigma_u^2 \text{ and } \mathrm{E}(u_{it}u_{is})=0.$$

... and the implications

- Part (b) together with RE.1 is a homoskedasticity assumption: $\mathrm{Var}(c_i|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}) = \mathrm{Var}(c_i) = \sigma_c^2$.
- It implies the variance structure for the stacked equation

$$\mathrm{Var}(\boldsymbol{\iota}_T c_i) = \mathrm{E}(\boldsymbol{\iota}_T\boldsymbol{\iota}_T' c_i^2) = \boldsymbol{\iota}_T\boldsymbol{\iota}_T'\sigma_c^2$$

  where $\boldsymbol{\iota}_T\boldsymbol{\iota}_T'$ is a $T \times T$ matrix of ones.

- Taken together the $T \times 1$ regression disturbance $\mathbf{v}_i$ of the stacked equation has variance

$$\boldsymbol{\Omega} = \mathrm{E}(\mathbf{v}_i'\mathbf{v}_i) = \sigma_c^2\boldsymbol{\iota}_T\boldsymbol{\iota}_T' + \sigma_u^2\mathbf{I}_T = \begin{bmatrix} \sigma_c^2+\sigma_u^2 & & \sigma_c^2 \\ & \ddots & \\ \sigma_c^2 & & \sigma_c^2+\sigma_u^2 \end{bmatrix}$$

  which is the classical random effects structure.

- $\boldsymbol{\Omega}$ depends on only two parameters to be estimated for feasible GLS.

Random effects estimator

Suppose you have initial estimates of σ_c^2 and σ_u^2 and construct

$$\hat{\mathbf{\Omega}} = \hat{\sigma}_c^2 \boldsymbol{\iota}_T \boldsymbol{\iota}_T' + \hat{\sigma}_u^2 \mathbf{I}_T.$$

Then the FGLS or random effects (RE) estimator is

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

Discussion:

- Consistency hinges on assumptions RE.1 and RE.2. Hence, the RE estimator is consistent (though not efficient) even if RE.3 is violated and $\mathbf{\Omega}$ is misspecified. (More on this later.)
- There are many ways to compute initial estimates of σ_c^2 and σ_u^2. In the following, we present the easiest ones.

Wooldridge's estimators of the variance components

Recall that the main diagonal of $\mathbf{\Omega}$ has elements

$$\sigma_v^2 \equiv \mathrm{Var}(v_{it}) = \sigma_c^2 + \sigma_u^2.$$

To estimate it, use the pooled OLS estimator and construct

$$\hat{\hat{v}}_{it} = y_{it} - \mathbf{x}_{it}\hat{\boldsymbol{\beta}}_{POLS}.$$

Then compute

$$\hat{\sigma}_v^2 = \frac{1}{NT - K}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\hat{v}}_{it}^2.$$

Next recall that the off-diagonal elements of $\mathbf{\Omega}$ are $\sigma_c^2 = \mathrm{E}(v_{it}v_{is})$, $t \neq s$.

The off-diagonal elements

Due to symmetry, it is sufficient to estimate the lower triangular part of $\boldsymbol{\Omega}$ which contains the following $T(T-1)/2$ elements:

$$\mathrm{E}(v_{it}v_{is}), \qquad t = 1, \ldots, T-1, s = t+1, \ldots, T.$$

To estimate σ_c^2, use again the pooled OLS residuals and average across all i:

$$\hat{\sigma}_c^2 = \frac{1}{NT(T-1)/2 - K} \sum_{i=1}^{N} \underbrace{\sum_{t=1}^{T-1} \sum_{s=t+1}^{T}}_{T(T-1)/2 \text{ elements}} \hat{\hat{v}}_{it} \hat{\hat{v}}_{is} \, .$$

Note that $\hat{\sigma}_c^2$ may turn out to be negative. This might be either due to sampling error (then you may set it to zero or use a different estimator) or because assumption RE.3 is invalid (then you may use an unrestricted FGLS estimator as presented below).

The Swamy-Arora estimators of the variance components

- Estimate σ_u^2 from the within regression

$$\hat{\sigma}_u^2 = \frac{1}{N(T-1)-K} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{u}_{it}^2.$$

- Estimate $\sigma_c^2 + \sigma_u^2/T$ from the between regression

$$\hat{\sigma}_c^2 + \hat{\sigma}_u^2/T = \frac{1}{N-K} \sum_{i=1}^{N} \hat{\bar{v}}_i^2$$

- Estimate σ_c^2 from an appropriate linear combination

$$\hat{\sigma}_c^2 = \frac{\sum_{i=1}^{N} \hat{\bar{v}}_i^2}{N-K} - \hat{\sigma}_u^2/T.$$

When $\hat{\sigma}_c^2 < 0$, Stata sets it to zero.

Asymptotic distribution of the RE estimator

Recall from Econometrics I (look up the GLS estimator for the SUR model) that the FGLS estimator has limiting distribution

$$\sqrt{N}\left(\hat{\boldsymbol{\beta}}_{RE} - \boldsymbol{\beta}\right) \xrightarrow{\mathrm{d}} \text{Normal}\left(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\right),$$

where

$$\mathbf{A} \equiv \mathrm{E}(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i)$$

and

$$\mathbf{B} \equiv \mathrm{E}\left(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{v}_i\mathbf{v}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i\right).$$

Note that this expression for the variance does not require $\boldsymbol{\Omega}$ to be correctly specified and is thus robust to misspecification of $\boldsymbol{\Omega}$.

Estimating the variance matrix robust to misspecification $\star$

To estimate $\mathbf{A}$ and $\mathbf{B}$, use sample equivalents:

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

and

$$\hat{\mathbf{B}} \equiv N^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \hat{\mathbf{\Omega}}^{-1} \hat{\mathbf{v}}_i \hat{\mathbf{v}}_i' \hat{\mathbf{\Omega}}^{-1} \mathbf{X}_i,$$

where $\hat{\mathbf{v}}_i$ are the RE residuals.

Under correct specification ⋆

When we are sure that $\mathbf{\Omega}$ is correctly specified (=the random effects structure is correct), things simplify a lot:

$$\begin{aligned}\mathbf{B} &\equiv \mathrm{E}\left(\mathbf{X}_i'\mathbf{\Omega}^{-1}\mathbf{v}_i\mathbf{v}_i'\mathbf{\Omega}^{-1}\mathbf{X}_i\right) = \mathrm{E}\left(\mathbf{X}_i'\mathbf{\Omega}^{-1}\,\mathrm{E}[\mathbf{v}_i\mathbf{v}_i'|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}]\mathbf{\Omega}^{-1}\mathbf{X}_i\right)\\ &= \mathrm{E}\left(\mathbf{X}_i'\mathbf{\Omega}^{-1}\mathbf{\Omega}\mathbf{\Omega}^{-1}\mathbf{X}_i\right) = \mathrm{E}\left(\mathbf{X}_i'\mathbf{\Omega}^{-1}\mathbf{X}_i\right) = \mathbf{A}\end{aligned}$$

and thus

$$\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}.$$

Thus, we solely have to estimate

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

Note that this estimator is more efficient than the robust one if the random effects structure is correct. However, otherwise it is inconsistent.

Implementation in Stata

Example: Data set has identifier for each individual denoted `id` and for each time period denoted `year`.

You of course have to tell Stata that you have panel data:

```
xtset id year
```

The RE estimator with nonrobust variance matrix is computed using

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

The RE estimator with robust variance matrix is computed using

```
xtreg y x1 x2 x3, re vce(robust)
```

Example revisited: Effects of training grants on scrap rates

Sample: 54 firms reported scrap rates for 1987, 1988, and 1989. Some received a grant in one of the years 1988 or 1989 to initiate a training program.

Analysis: Regression of log scrap rates on yearly dummies, union membership dummy, grant dummy ("grant") and lagged grant dummy ("grant_1").

Note that we can afford to include the union membership dummy as we don't apply the within transformation.

Stata RE output

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

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

R-sq:  within  = 0.2006                         Obs per group: min =         3
       between = 0.0206                                        avg =       3.0
       overall = 0.0361                                        max =         3

                                                Wald chi2(5)       =     27.65
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
theta          = .79754262

                                 (Std. Err. adjusted for 54 clusters in fcode)
------------------------------------------------------------------------------
             |               Robust
      lscrap |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         d88 |  -.0934519   .0938166    -1.00   0.319    -.2773291    .0904253
         d89 |  -.2698336   .1885186    -1.43   0.152    -.6393232     .099656
       union |   .5478021   .4023672     1.36   0.173     -.240823    1.336427
       grant |   -.214696   .1311183    -1.64   0.102    -.4716832    .0422911
     grant_1 |  -.3770698   .2674417    -1.41   0.159    -.9012458    .1471062
       _cons |   .4148333   .2673996     1.55   0.121    -.1092603     .938927
-------------+----------------------------------------------------------------
     sigma_u |  1.3900287
     sigma_e |  .49774421
         rho |  .88634984   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Notes:

- R-sq within: squared correlation between $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\hat{\boldsymbol{\beta}}_{RE}$ and $y_{it} - \bar{y}_i$.
- R-sq between: squared correlation between $\bar{\mathbf{x}}_i\hat{\boldsymbol{\beta}}_{RE}$ and $\bar{y}_i$.
- R-sq overall: squared correlation between $\mathbf{x}_{it}\hat{\boldsymbol{\beta}}_{RE}$ and y_{it}.
- sigma_u: square root of $\mathrm{Var}(c_i) = \sigma_c^2$
- sigma_e: square root of $\mathrm{Var}(u_{it}) = \sigma_u^2$
- rho: variance share $\mathrm{Var}(c_i)/\,\mathrm{Var}(u_{it}) = \sigma_c^2/\sigma_u^2$
- theta: using the option theta, we obtain a measure of how near the RE estimator is to the pooled OLS or FE estimators (0=OLS, 1=FE). This is what we call $1 - \hat{\phi}$ below.

Generalized GLS estimator

If you suspect the RE variance structure is not valid **and** $N >> T$, you can leave $\mathbf{\Omega}$ fully unrestricted and estimate

$$\hat{\mathbf{\Omega}} = \frac{1}{N}\sum_{i=1}^{N} \hat{\hat{\mathbf{v}}}_i \hat{\hat{\mathbf{v}}}_i',$$

where $\hat{\hat{\mathbf{v}}}_i$ is again the pooled OLS residual vector. Then apply the GLS estimator as above.

A command that accomplishes this in Stata is:

```
xtgee y x1 x2 x3, family(gau) link(identity)
corr(unstructured) vce(robust)
```

Example: Effects of job training grants on scrap rates

```
. xtgee lscrap d88 d89 union grant grant_1, family(gau) link(identity) corr(unstructured) vce(robust)

Iteration 1: tolerance = .38513976
Iteration 2: tolerance = .00739269
Iteration 3: tolerance = .00013876
Iteration 4: tolerance = 2.490e-06
Iteration 5: tolerance = 7.070e-08

GEE population-averaged model                 Number of obs       =        162
Group and time vars:           fcode year     Number of groups    =         54
Link:                            identity     Obs per group: min =          3
Family:                          Gaussian                    avg =        3.0
Correlation:                 unstructured                    max =          3
                                              Wald chi2(5)        =      22.91
Scale parameter:                 2.124799     Prob > chi2         =     0.0004

                               (Std. Err. adjusted for clustering on fcode)
```

lscrap	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
d88	-.0769043	.0901083	-0.85	0.393	-.2535134	.0997048
d89	-.2499586	.1780286	-1.40	0.160	-.5988882	.098971
union	.5551608	.3934136	1.41	0.158	-.2159156	1.326237
grant	-.2617261	.1418729	-1.84	0.065	-.5397918	.0163396
grant_1	-.408804	.2657973	-1.54	0.124	-.9297573	.1121492
_cons	.4123804	.2622817	1.57	0.116	-.1016823	.9264432

Outline

Coming up

The relation between FE and RE