Block 3
Panel data – models, estimation and testing

Advanced Econometrics 4EK608

Vysoká škola ekonomická v Praze

Outline

1. Panel data – basics (repetition from BSc courses)
2. Short panels – estimation, inference & testing
3. Long panels – models and estimation
4. Large panel data sets – introduction
5. Panel data – additional topics and extensions

Panel data

Panel data – basics (repetition from BSc courses)

- Pooled cross sections
- Longitudinal data
- Panel data
- Balanced & unbalanced panel data sets
- Dimensions of panel data sets & analysis implications
- Basic features and motivation for panel data use

Pooled cross sections

- **<u>Pooled cross sections:</u>** Random sampling from a large population at different time periods (i.e. for each time period, we have a different - randomly chosen - set of CS units).
- Should not be confused with "actual" panel data.
- Pooled cross sections: sampling from a changing population at different points in time generates **independent, not identically distributed** (*inid*) observations.
- Pooled cross sections are easy to deal with, simply by allowing the intercept (and perhaps some selected slopes) in a LRM to vary across time.
- Can be used for policy analysis (difference-in-differences estimator).

Pooled cross sections - model example

$$\log(wage_{it}) = \theta_0 + \theta_1 d91_t + \theta_2 d92_t + \delta_1 female_{it} + \delta_2 educ_{it} + \\ + \gamma_1 exper_{it} + \gamma_2 (female \times d91)_{it} + \gamma_3 (female \times d92)_{it} + u_{it}$$

where $t = 1990, 1991, 1992; \quad i = 1, 2, \ldots, 500$

Each year, we draw 500 individuals at random. Individual respondents are not followed. Total observations: $N \times T = 1.500$

$d91_t$ and $d92_t$ are time dummies,

$female_{it}$, $educ_{it}$ and $exper_{it}$ describe the gender, education and work experience of the i-th individual at time t,

$(female \times d91)_{it}$ is an interaction element, may be used to describe whether changes in wages over time are statistically different for man and woman.

Pooled cross sections - model example contd.

$$\log(wage_{it}) = \beta_0 + \beta_1 d91_t + \beta_2 d92_t + \beta_3 female_{it} + \\ + \beta_4 educ_{it} + \beta_5 exper_{it} + u_{it}$$

Chow test for structural changes across time

Basically an F-test for linear restrictions, can be used to determine whether the estimated slope coefficients change across time.

In our $\log(wage)$ equation, we would test the H_0 of "time-invariant" β_3, β_4 and β_5 coefficients, while allowing for time dummies (time-specific intercepts).

Pooled cross sections: Chow test

SSR_r: restricted model – pooled regression, allowing for different time intercepts.

SSR_{ur}:run a regression for each of the time periods. $SSR_{ur} = SSR_1 + SSR_2 + \cdots + SSR_T$

$T + Tk$ parameters estimated in the unrestricted model

$$F = \frac{SSR_r - SSR_{ur}}{SSR_{ur}} \cdot \frac{(n-T-Tk)}{(T-1)k};$$

under H_0 of no structural break, $F \sim F((T-1)k, (n - T - Tk))$

Note: This test is not robust to heteroscedasticity (including changing variance across time). Robust variants of the test exist, based on interaction terms.

Longitudinal data

- N individual CS units are followed over time.

- The observation set $\{y_{it}, \boldsymbol{x}_{it}\}$ denotes some ith individual observed at a time period t. The number of observations in time may differ among CS units and observations may occur at different time points.

  Example: For a medical study, we measure child's weight (plus other data) at birth and repeatedly over a period of one year. For some y_{it} observation, index t denotes days from birth. Due to doctor visit scheduling, children are weighted at different t "values". Typically, the number of doctor visits (observations) differs across children. Children in the study are born on different dates (say, Jan 2015 - Oct 2019).

  Example extends easily to economic environment
  (we can follow newly founded companies, etc.).

- Longitudinal data are typically used in Linear mixed effects (LME) models (discussed separately).

- Note: Distinction between longitudinal and panel data may be subtle and different authors may use conflicting terminologies ...

Panel data

- Here, N individual CS units are followed over T time periods. Index t denotes a common time period (year, quarter, month) at which CS units are observed.

- Regression model of the form

$$y_{it} = \boldsymbol{x}'_{it}\boldsymbol{\beta} + a_i + \varepsilon_{it},$$

  where i denotes CS units and t identifies time periods, a_i is the individual unobserved element (person, company, group or other CS-unit).

- In this course (Block 3), we focus on panel data.

  Different data dimensions, model types, estimators and tests discussed next.

Balanced & unbalanced panel data sets

- **Balanced panels:** observations available for all time periods on all CS units. Often assumed for simplicity of interpretation.

- **Unbalanced panels:** mechanics of coefficient estimation do not differ. Model interpretation may require formal description of why the panel may be unbalanced. Does the random sampling assumption (CS units) hold?

- Problems in unbalanced panels may be caused by:
  - **Sample selection bias:** with e.g. self-selection, coefficients can be be biased and inconsistent.
  - **Attrition bias:** even if participants are randomly selected at the beginning of observation, they often leave (medical study, school, etc.) on a non-random basis.

Dimensions of panel data sets

- Short panels: $N \gg T$
  Working with short panels is similar to CS data analysis. If CS units are randomly drawn from a population and T is small and fixed, then asymptotic analysis – asymptotic properties – hold for arbitrary time dependence and distributional heterogeneity across time.

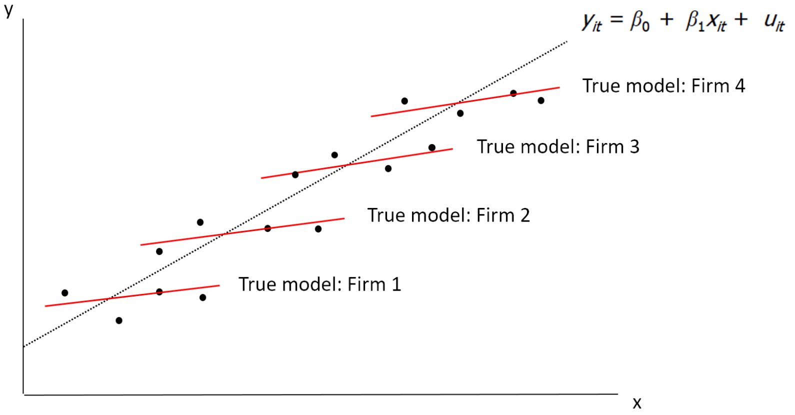
- Long panels: $T \gg N$
  Working with long panels is similar to time-series analysis. In TS analysis, stationarity & weak dependency conditions apply. SURE (Seemingly Unrelated Regression Equation) approach can be used: for the regression equations under scrutiny (typically with a common model specification), we estimate contemporaneous error covariances and use this information to improve efficiency of the estimate
  (see Greene, chapter 10.2)

- Large panel datasets: T and N large
  Both CS and TS analysis assumptions apply, specialized estimators exist for large (heterogeneous) panels.
  Cointegrated series in panels: estimation and tests by Pesaran.

Basic features and motivation for panel data use

Pooled regression with panel data:

- Heterogeneity bias
- Example similar in principle to the Simpson's paradox



Variation for the dependent variable and regressors:

- overall variation – variation over time and individuals
- between variation – variation between individuals
- within variation – variation within individuals (over time)

Id	Time	Variable	Individual mean	Overall mean	Overall deviation	Between deviation	Within deviation	Within deviation (modified)
i	t	x_{it}	$\overline{x}_i$	$\overline{x}$	$x_{it} - \overline{x}$	$\overline{x}_i - \overline{x}$	$x_{it} - \overline{x}_i$	$x_{it} - \overline{x}_i + \overline{x}$
1	1	9	10	20	-11	-10	-1	19
1	2	10	10	20	-10	-10	0	20
1	3	11	10	20	-9	-10	1	21
2	1	20	20	20	0	0	0	20
2	2	20	20	20	0	0	0	20
2	3	20	20	20	0	0	0	20
3	1	25	30	20	5	10	-5	15
3	2	30	30	20	10	10	0	20
3	3	35	30	20	15	10	5	25

Panel data model – a structured notation example

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

where $\quad i = 1, 2, \ldots, N;\ t = 1, 2, \ldots, T,$

$\boldsymbol{g}_t'$ is a row-vector of aggregate time effects (often time dummies),

$\boldsymbol{z}_i$ is a set of time-constant observed variables,

$\boldsymbol{w}_{it}$ changes across i and t (for at least some units i and time periods t), can include interactions among time-constant and time varying variables,

$\boldsymbol{\theta}, \boldsymbol{\delta}$ and $\boldsymbol{\gamma}$ – column vectors of regression coefficients

Panel data model - a structured notation example

$$\log(wage_{it}) = \theta_0 + \theta_1 d91_t + \theta_2 d92_t + \delta_1 female_i + \delta_2 educ_i + \\ + \gamma_1 exper_{it} + \gamma_2 (female \times exper)_{it} + a_i + u_{it}$$

Where $t = 1990, 1991, 1992; \quad i = 1, 2, \ldots, 100.$
For a balanced panel, $T \times N = 300$

We follow 100 individuals across three years.

$d91_t$ and $d92_t$ are time dummies,
$female_i$ and $educ_i$ do not change over time (individuals in our dataset are not active students ...),
$exper_{it}$ changes between individuals and across time periods,
$(female \times exper)_{it}$ is an interaction element, changes between individuals and across time.

Short panels – estimation, inference & testing

- Estimation methods – repetition from BSc courses
- Choosing adequate estimators: assumptions and tests
- Robust inference (autocorrelation and heteroscedasticity)

LSDV regression

In the model $y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + a_i + u_{it}$,

- Elements a_i are usually regarded as unobservable variables.
- Accounting for a_i can provide appropriate interpretation of $\boldsymbol{\beta}$.
- Traditional (old) approaches to fixed effects estimation view the a_i as parameters to be estimated along with $\boldsymbol{\beta}$.

How to estimate a_i values along with $\boldsymbol{\beta}$?

- Define N dummy variables - one for each cross-section. (Amendment for dummy-variable trap is necessary.)
- Convenient LSDV model expansion: use interactions to control for individual slopes for chosen regressors.

LSDV regression – example

$$y_{it} = \alpha_1 \boxed{ind1_i} + \alpha_2 \boxed{ind2_i} + \cdots + \alpha_N \boxed{indN_i} + \beta_1 x_{it1} + \cdots + \beta_K x_{itK} + u_{it}$$

Dummy equals 1 only if observations (time-invariant) relate to i-th C-S unit.

- $\hat{\boldsymbol{\beta}}_{\text{LSDV}}$ is identical to $\hat{\boldsymbol{\beta}}_{\text{FE}}$ (explained next).
- $\hat{\boldsymbol{\beta}}_{\text{LSDV}}$ is a consistent estimator of $\boldsymbol{\beta}$ if we hold T fixed and $N \rightarrow \infty$.
- For $\hat{\boldsymbol{\alpha}}$ (vector of individual $\hat{\alpha}_i$ values), LSDV-estimator consistency does not hold: as $N \rightarrow \infty$, information does not accumulate for a_i.

FD estimator

We can eliminate unobserved individual heterogeneity from the regression: $y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + a_i + u_{it}$

by first differences (FD) transformation:

$\Delta y_{it} = y_{it} - y_{i,t-1} = \Delta\boldsymbol{x}_{it}\boldsymbol{\beta} + \Delta a_i + \Delta u_{it} = \Delta\boldsymbol{x}_{it}\boldsymbol{\beta} + \Delta u_{it}$

- ✓ Removes any unobserved heterogeneity.
- × We remove all time-invariant factors in $\boldsymbol{x}$.
  If the time-invariant regressors are of no interest, this is a robust estimator.

Estimation can be done with FGLS (autocorrelation of transformed residuals), or OLS with HAC robust errors.

FD is most suitable when we have $t = \{1; 2\}$, i.e. for a two period panel. FD may be used with more time periods, we have $N(T-1)$ observations after differencing.

FD estimator – assumptions

FD.1 Functional form: $y_{it} = \beta_1 x_{it1} + \cdots + \beta_K x_{itK} + a_i + u_{it}$, $i = 1, \ldots, N, \;\; t = 1, \ldots, T$

FD.2 We have random sample from cross-sectional units.

FD.3 Each regressor changes in time at least for some i and no perfect linear combination exists among regressors.

FD.4 For each i and t, $E(u_{it} \mid \boldsymbol{X}_i, a_i) = 0$. [Alt.: regressors are strictly exogenous conditional on unobserved effects: $\mathrm{corr}(x_{itj}, u_{is} \mid a_i) = 0, \quad \forall \; t, s$]

FD.5 Variance of differenced errors conditional on all regressors is constant: $\mathrm{var}(\Delta u_{it} \mid \boldsymbol{X}_i) = \sigma^2, \quad t = 2, 3, \ldots, T$. [homoscedasticity]

FD.6 No serial correlation exists among differenced errors. $cov(\Delta u_{it}, \Delta u_{is} \mid \boldsymbol{X}_i) = 0, \quad t \neq s$

FD.7 Differenced errors are normally distributed conditional on all regressors $\boldsymbol{X}_i$.

FD estimator – assumptions

Under **FD.1 - FD.4**
FD estimator is unbiased.
FD estimator is consistent for fixed T as $N \to \infty$.
For unbiasedness, $E(\Delta u_{it} \mid \boldsymbol{X}_i) = 0$ (for $t = 2, 3, \dots$) is sufficient (instead of FD.4)

Under **FD.1 - FD.6**
FD estimator is BLUE (conditional on explanatory variables).
Asymptotic inference for FD estimator holds (t and F statistics asymptotically follow corresponding distributions).

Under **FD.1 - FD.7**
FD estimator is BLUE (conditional on explanatory variables).
FD estimators - i.e. pooled OLS on first differences - are normally distributed (t and F statistics have exact t and F distributions).

Problems related to the FD estimator:

- First-differenced estimates will be imprecise if explanatory variables vary only to a small extent over time (no estimate possible if regressors are time-invariant).
- Potentially, there is insufficient (lower) variability in differenced variables.
- Without strict exogeneity of regressors (e.g. in the case of a lagged dependent variable /say, $y_{i,t-1}$/ among regressors or with measurement errors), adding further periods does not reduce inconsistency.
- FD estimator may be worse than pooled OLS if explanatory variables are subject to measurement errors (errors in variables - EIV).

FD estimator example

$$crmrte_{it} = \beta_0 + \delta_0 d87_{it} + \beta_1 unem_{it} + a_i + u_{it},$$

$$t = 1982, 1987$$

Dummy for the second time period

Model expanded: written separately for each time period

$$crmrte_{i1987} = \beta_0 + \delta_0 \cdot 1 + \beta_1 unem_{i1987} + a_i + u_{i1987}$$

$$crmrte_{i1982} = \beta_0 + \delta_0 \cdot 0 + \beta_1 unem_{i1982} + a_i + u_{i1982}$$

FD applied

$$\Rightarrow \Delta crmrte_i = \delta_0 + \beta_1 \Delta unem_i + \Delta u_i$$

δ_0 has a time effect interpretation

Individual unobserved effect disappears

$$\Delta \widehat{crmrte} = \underset{(4.70)}{15.40} + \underset{(.88)}{2.22}\ \Delta unem$$

With OLS estimation, HAC errors should be used

FE estimator

"Fixed" means correlation of a_i and $\boldsymbol{x}_{it}$, not that a_i is non-stochastic.

We can rewrite $y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + a_i + u_{it}$ as follows:
$y_{it} = \beta_1 x_{it1} + \cdots + \beta_K x_{itK} + a_i + u_{it}, \qquad i = 1, \ldots, N, \;\; t = 1, \ldots, T$
Now, <u>for each i</u>, we <u>average</u> the above equation <u>over time</u>:

$\overline{y}_i = \beta_1 \overline{x}_{i1} + \cdots + \beta_K \overline{x}_{iK} + \overline{a}_i + \overline{u}_i$
(N equations with individual averages)

By subtracting individual averages from the original observations (time-demeaning), we get:

$$\Rightarrow \boxed{[y_{it} - \overline{y}_i]} = \beta_1 \boxed{[x_{it1} - \overline{x}_{i1}]} + \cdots + \beta_K \boxed{[x_{itK} - \overline{x}_{iK}]} + \boxed{[u_{it} - \overline{u}_i]}$$

Alternative notation: $\ddot{y}_{it} = \ddot{\boldsymbol{x}}_{it}\boldsymbol{\beta} + \ddot{u}_{it}$; where $\ddot{y}_{it} = y_{it} - \overline{y}_i$, etc.

FE estimator, denoted $\hat{\boldsymbol{\beta}}_{FE}$, is the pooled OLS estimator applied to time-demeaned data.

FE estimator

FE estimator: by time demeaning, we get rid of the a_i element - as it does not vary over time

- $a_i = \overline{a}_i \;\rightarrow\; a_i - \overline{a}_i = 0$
- Intercept and all time-invariant regressors are also eliminated using the FE (within) transformation.

After FE estimation, a_i elements may be estimated as follows:

$$\hat{a}_i = \overline{y}_i - \hat{\beta}_1 \overline{x}_{i1} - \cdots - \hat{\beta}_K \overline{x}_{iK}, \;\; i = 1, \dots, N$$

However, in most practical applications, a_i values bear limited useful information.

For each C-S observation i, we loose one d.f. in estimation …for each i, the demeaned errors $\ddot{u}_{it}$ add up to zero when summed over time. Hence $df = N(T-1) - k$

FE estimator – assumptions

FE.1 Functional form: $y_{it} = \beta_1 x_{it1} + \cdots + \beta_K x_{itK} + a_i + u_{it}$, $i = 1, \ldots, N,\ \ t = 1, \ldots, T$

FE.2 We have random sample from cross-sectional units.

FE.3 Each regressor changes in time at least for some i and no perfect linear combination exists among regressors.

FE.4 For each i and t, $E(u_{it} \mid \boldsymbol{X}_i, a_i) = 0$. [Alt.: regressors are strictly exogenous conditional on unobserved effects: $\mathrm{corr}(x_{itj}, u_{is} \mid a_i) = 0, \quad \forall\ t, s$]

FE.5 Variance of errors conditional on all regressors is constant: $\mathrm{var}(u_{it} \mid \boldsymbol{X}_i, a_i) = \mathrm{var}(u_{it}) = \sigma_u^2, \quad t = 1, 2, \ldots, T$. [homoscedasticity]

FE.6 No serial correlation exists among idiosyncratic errors. $\mathrm{cov}(u_{it}, u_{is} \mid \boldsymbol{X}_i, a_i) = 0, \quad t \neq s$

FE.7 Errors are normally distributed conditional on all regressors $(\boldsymbol{X}_i, a_i)$.

FE estimator – assumptions

Under **FE.1 - FE.4** (identical to **FD.1 - FD.4**)
FE estimator is unbiased.
FE estimator is consistent for fixed T as $N \to \infty$.

Under **FE.1 - FE.6**
FE estimator is BLUE.
FD is unbiased
. . . **FE.6** makes FE better (less variance) than FD.
Asymptotically valid inference for FE estimator holds (t and F).

Under **FE.1 - FE.7**
FE estimator is BLUE and t and F statistics have exact t and F distributions.
FE estimators - i.e. pooled OLS on time demeaned data - are normally distributed.

FE estimator – example

Example: Effect of training grants on firm scrap rate

$$scrap_{it} = \beta_1 d88_{it} + \beta_2 d89_{it} + \beta_3 grant_{it} + \beta_4 grant_{it-1} + a_i + u_{it}$$

Time-invariant reasons why one firm is more productive than another are controlled for. The important point is that these may be correlated with other explanatory variables.

Stars denote time-demeaning

Fixed-effects estimation using the years 1987, 1988, 1989:

$$\underset{(.109)}{\widehat{scrap}_{it}^{*} = -.080}\ d88_{it}^{*} - \underset{(.133)}{.247}\ d89_{it}^{*} \boxed{-\underset{(.151)}{.252}}\ grant_{it}^{*} \boxed{-\underset{(.210)}{.422}}\ grant_{it-1}^{*}$$

$n = 162, \quad R^2 = .201$

Training grants significantly improve productivity (with a time lag)

Between estimator

- **Within estimator** $\iff$ **FE estimator**
  For equation $\ddot{y}_{it} = \ddot{\boldsymbol{x}}_{it}\boldsymbol{\beta} + \ddot{u}_{it}$, the FE estimator (pooled OLS on time-demeaned data) is often called "within" estimator, as it uses variation within each cross-section.

- **Between estimator**
  Is obtained as the OLS estimation of
  $\overline{y}_i = \beta_1 \overline{x}_{i1} + \cdots + \beta_K \overline{x}_{iK} + \overline{a}_i + \overline{u}_i$ (i-avgs. over time)
  where we add an intercept and "ignore" a_i (assume $\overline{a}_i = 0$):

  $$\overline{y}_i = \beta_0 + \beta_1 \overline{x}_{i1} + \cdots + \beta_K \overline{x}_{ik} + \overline{u}_i$$

  The between estimator uses only variation between the CS observations (ignores information on how the variables change over time). Consistent for a_i and $\boldsymbol{X}_i$ independent.

- $\hat{\boldsymbol{\beta}}_{Between}$ is not consistent if a_i is correlated with $\boldsymbol{X}_i$.
  If we can reasonably assume no correlation between $\boldsymbol{X}_i$ and a_i, the "between" estimator is consistent, yet not efficient - we would use the RE estimator (explained next).

RE estimator

If a_i are uncorrelated with $\boldsymbol{x}_{it}$, then it may be appropriate to model the individual constant terms as randomly distributed across cross-sectional units. RE models are appropriate if C-S units are from **a large sample** (good asymptotic properties).

- RE estimator potentially inconsistent if assumptions not met.
- $y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + a_i + u_{it}$

  If we can assume that a_i is uncorrelated with each explanatory variable: $\text{corr}(\boldsymbol{x}_{it}, a_i) = 0; \;\; t = 1, 2, \ldots, T$
  then we may simply drop a_i from the equation and OLS-based β_j estimates will remain unbiased & consistent – yet inefficient.
- By dropping a_i from the regression, we effectively create a new error term: $v_{it} = a_i + u_{it}$.
- As a_i is time-invariant, the random element v_{it} contains a lot of "inertia", i.e. autocorrelation (unless $a_i = 0$).

RE estimator - FGLS

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \dots + \beta_K x_{itK} + v_{it};$$

The quasi-demeaning (quasi-differencing) parameter θ is used for the FGLS estimation:

$$\theta = 1 - \left[\sigma_u^2/(\sigma_u^2 + T\sigma_a^2)\right]^{1/2}, \quad 0 \leq \theta \leq 1$$

where $\mathrm{var}(a_i) = \sigma_a^2; \quad \mathrm{var}(u_i) = \sigma_u^2$

- For each dataset, consistent estimators of σ_a^2 and σ_u^2 are available.
- Their estimation is based on pooled OLS or FE.
  Also, we use the fact that $\sigma_v^2 = \sigma_a^2 + \sigma_u^2$

RE estimator is a pooled OLS used on the quasi-demeaned data:

$$\boxed{[y_{it} - \theta\overline{y}_i]} = \beta_1 \boxed{[x_{it1} - \theta\overline{x}_{i1}]} + \dots + \beta_K \boxed{[x_{itK} - \theta\overline{x}_{iK}]} + \boxed{[a_i - \theta\overline{a}_i + u_{it} - \theta\overline{u}_i]}$$

(transformed errors follow G-M assumptions – not autocorrelated)

RE estimator - FGLS

$$\boxed{[y_{it} - \theta \overline{y}_i]} = \beta_1 \boxed{[x_{it1} - \theta \overline{x}_{i1}]} + \cdots + \beta_K \boxed{[x_{itK} - \theta \overline{x}_{iK}]} + \boxed{[a_i - \theta \overline{a}_i + u_{it} - \theta \overline{u}_i]}$$

Interestingly, the FGLS equation is a general form that encompasses both FE and pooled OLS:

$$\hat{\theta} \to 1 \quad \Rightarrow \quad \text{RE} \to \text{FE}$$

$$\hat{\theta} \to 0 \quad \Rightarrow \quad \text{RE} \to \text{Pooled}$$

RE estimator – Assumptions

FE.1 Functional form: $y_{it} = \beta_1 x_{it1} + \cdots + \beta_K x_{itK} + a_i + u_{it}$, $i = 1, \ldots, N$, $t = 1, \ldots, T$

FE.2 We have random sample from cross-sectional units.

FE.4 $\forall\ i, t$: $E(u_{it} \mid \boldsymbol{X}_i, a_i) = 0$. [Alt.: $\operatorname{corr}(x_{itj}, u_{is} \mid a_i) = 0, \quad \forall\ t, s$]

FE.5 Variance of idiosyncratic errors conditional on all regressors is constant: $\operatorname{var}(u_{it} \mid \boldsymbol{X}_i, a_i) = \operatorname{var}(u_{it}) = \sigma_u^2, \quad t = 1, 2, \ldots, T$. [homoscedasticity]

FE.6 No serial correlation exists among idiosyncratic errors. $\operatorname{cov}(u_{it}, u_{is} \mid \boldsymbol{X}_i, a_i) = 0, \quad t \neq s$

FE.7 [small sample normality of u_{it} has little importance for RE estimator]

RE.1 There are no perfect linear relationships among explanatory variables. [replaces **FE.3**]

RE.2 In addition to **FE.4**, the expected value of a_i given all regressors is constant: $E(a_i \mid \boldsymbol{X}_i) = \beta_0$. [Rules out correlation between a_i and $\boldsymbol{X}_i$]

RE.3 In addition to **FE.5**, variance of a_i given all regressors is constant: $\operatorname{var}(a_i \mid \boldsymbol{X}_i) = \sigma_a^2$ [homoscedasticity imposed on a_i]

RE estimator – Assumptions

Under **FE.1+FE.2+RE.1+(FE.4+RE.2)**
RE estimator is consistent and asymptotically normal
(for fixed T as $N \to \infty$).
RE standard errors and statistics are not valid unless
(FE.5+RE.3) and **FE.6** conditions are met.

Under
FE.1-FE.2+RE.1+(FE.4+RE.2)+(FE.5+RE.3)+FE.6
RE estimator is consistent and asymptotically normal
(for fixed T as $N \to \infty$).
RE standard errors and statistics are valid.
RE is asymptotically efficient

- lower st.errs. than pooled OLS
- for time-varying variables, RE estimator is more efficient than FE (FE cannot be used on time-invariant variables).

RE estimator – Example

RE approach is used because many of the variables are time-invariant. But is the random effects assumption realistic?

Example:
Estimated wage equation:

$$
\begin{aligned}
\widehat{\log}(wage_{it}) = {} & \underset{(.011)}{.092}\ \boxed{educ_{it}} - \underset{(.048)}{0.213}\ black_{it} + \underset{(.043)}{0.054}\ hisp_{it} \\
& + \underset{(.015)}{.106}\ exper_{it} - \underset{(.0007)}{.0047}\ exper_{it}^2 + \underset{(.017)}{.064}\ married_{it} \\
& + \underset{(.018)}{.106}\ union_{it} + time\ dummies
\end{aligned}
$$

Random effects or fixed effects? In economics, unobserved individual effects are rarely uncorrelated with explanatory variables (say, individual ability and education would be correlated). CRE model/estimation may be more convincing.

CRE estimator

Correlated Random Effects (CRE) estimator - a synthesis of the RE and FE approaches:

- a_i viewed as random, yet they can be correlated with $\boldsymbol{x}_{it}$.

  Specifically, as a_i do not vary over time, it makes sense to allow for their correlation with the time average of $x_{it} : \overline{x}_i = T^{-1} \sum_{t=1}^{T} x_{it}$

- CRE allows for incorporation of time-invariant regressors into a FE-like estimator (combines RE and FE features).

- CRE allows for convenient testing of FE vs. RE.

CRE estimator

CRE: The individual-specific effect a_i is split up into a part that is related to the time-averages of the explanatory variables and a part r_i (a time-constant unobservable) that is unrelated to the explanatory variables:

For $y_{it} = \beta_1 x_{it} + a_i + u_{it}$, (a single-regressor illustration)

we assume:

$$a_i = \alpha + \gamma \overline{x}_i + r_i,$$

now: $\mathrm{corr}(r_i, \overline{x}_i) = 0 \Rightarrow \mathrm{corr}(r_i, x_{it}) = 0$

because $\overline{x}_i$ is a linear function of x_{it})

By substituting for a_i into the first equation, we obtain:
$y_{it} = \alpha + \beta_1 x_{it} + \gamma \overline{x}_i + r_i + u_{it}$

<u>This equation can be estimated using RE</u>
Element $\gamma \overline{x}_i$ controls for the correlation between a_i and x_{it},
r_i is uncorrelated with regressors.

CRE estimator

CRE: $y_{it} = \alpha + \beta_1 x_{it} + \gamma \overline{x}_i + r_i + u_{it}$

CRE is a modified RE of the original equation $y_{it} = \beta_1 x_{it} + a_i + u_{it}$:

with random effect r_i uncorrelated to other regressors & with time averages as additional regressors.

The resulting CRE estimate for β is identical to the FE estimator.

CRE allows for convenient testing of FE vs. RE:

H_0: $\gamma = 0$ can be evaluated using $\hat{\gamma}_{CRE}$ and appropriate (HCE) standard errors against

H_1: $\gamma \neq 0$ [RE assumes $\gamma = 0$: reject $H_0 \rightarrow$ reject RE in favor of FE]

- CRE is a versatile estimator. In terms of model specification, it allows for incorporation of time-invariant regressors into panel data models where a_i is correlated with regressors.

Arellano-Bond estimator (dynamic panels)

Dynamic panel model:

$$y_{it} = \delta_1 y_{i,t-1} + \boldsymbol{x}'_{it}\boldsymbol{\beta} + a_i + u_{it}$$

…may be expanded using additional lags of the dependent variable or using lagged exogenous regressors.

Nickel Bias

- Related (mostly) to the lagged exogenous regressors $\boldsymbol{x}$
- FEs take up some part of the dynamic effect and therefore dynamic panel data models lead to overestimated FEs and underestimated dynamic interactions.
- Whether the Nickel bias is significant in a particular model/dataset situation is an empirical question. Nevertheless, in theory this bias persists unless the number of time observations goes to infinity.
- The inclusion of additional cross-sections to the dataset would worsen the bias in most cases.

Arellano-Bond (AB) estimator

- The model is transformed into first differences to eliminate the individual effects:
  $\Delta y_{it} = \delta_1 \Delta y_{i,t-1} + \Delta \boldsymbol{x}'_{it} \boldsymbol{\beta} + \Delta u_{it}$,
- then a generalized method of moments (GMM) approach is used to produce asymptotically efficient estimates of the coefficients.
- AB is based on IVR (we need instruments for lagged dependent variable as this is an endogenous regressor in the FD-transformed model ($\Delta y_{i,t-1}$ correlated to Δu_{it}).
- Warning: AR(2) / not AR(1) / autocorrelation in residuals of the AB-estimated model renders the AB estimator inconsistent. After using the AB estimator, always test for AR(2) autocorrelation in the residuals!

Arellano-Bond estimator example

- Gross fixed capital formation model:

$$I_{it} = \beta_1 I_{i,t-1} + \boldsymbol{k}_{it}'\boldsymbol{\beta}_2 + \boldsymbol{x}_{it}'\boldsymbol{\beta}_3 + a_i + \varepsilon_{it}$$

  where I_{it} is the GFCF, $\boldsymbol{k}_{it}$ is a vector of foreign sources (FDI, loans, etc.) and $\boldsymbol{x}_{it}$ contains control variables (e.g. M2 deviations from 3-year trend, GDP growth, etc.).

- FE creates regressors ($\ddot{\boldsymbol{x}}_{it}$) which cannot be distributed independently of errors. Inconsistency of $\hat{\beta}_1$ is of order $1/T$ as $N \to \infty$ (T fixed). If $\beta_1 > 0$, the bias is invariably negative, β_1 will be underestimated (Nickel bias). More precisely, $(\hat{\beta}_1 - \beta_1) \approx -(1+\beta_1)/(T-1)$ for $N \to \infty$ and for T reasonably large (say, 10). Remaining β_j estimates are inconsistent as well.

- AB estimator: FD removes all constant terms (including a_i):

$$\Delta I_{it} = \beta_1 \Delta I_{i,t-1} + \Delta\boldsymbol{k}_{it}'\boldsymbol{\beta}_2 + \Delta\boldsymbol{x}_{it}'\boldsymbol{\beta}_3 + \Delta\varepsilon_{it}$$

  with $\Delta I_{i,t-1}$ still correlated to $\Delta\varepsilon_{it}$. However, this transformed model can be consistently estimated by GMM (usually, we use lags of regressors as IVs).

Choosing adequate estimators – assumptions and tests

- Poolability tests (pooled regression vs other estimators)
- Cross sectional dependency
- Estimator selection (FD vs FE; FE vs RE)
- Autocorrelation, heteroscedasticity, and robust inference

Choosing adequate estimators – assumptions and tests

We start by generalization of the (short) panel data model:

a) Model with individual effects:

$$y_{it} = \alpha + \beta_1 x_{it1} + \cdots + \beta_K x_{itK} + a_i + \nu_{it}$$

b) Model with time effects:

$$y_{it} = \alpha + \beta_1 x_{it1} + \cdots + \beta_K x_{itK} + \lambda_t + \nu_{it}$$

c) Model with twoways effects:

$$y_{it} = \alpha + \beta_1 x_{it1} + \cdots + \beta_K x_{itK} + a_i + \lambda_t + \nu_{it}$$

- For short panels, we often apply models with individual effects only – and use time dummies if necessary.
- **Some** of the following test are designed for different variants of unobserved effects.
- In principle, unobserved time effects are dealt with the same way as unobserved individual effects.

LSDV-based test for individual intercepts

- General principle of the test: Null hypothesis of common intercept ($H_0: \ a_1 = a_2 = \cdots = a_N$) is tested against the alternative of individual-specific intercepts. Common slopes (the same β-coefficients across CS units) are assumed (not tested).

- Test designed to evaluate significance of unobserved individual effects (time and twoways effect test – by analogy).

- Unrestricted model: $y_{it} = \alpha + \boldsymbol{d}'\boldsymbol{\delta} + \boldsymbol{x}_t'\boldsymbol{\beta} + u_{it}$

  where $\boldsymbol{d}$ is a vector of CS-ID dummy variables and $\boldsymbol{\delta}$ is a vector of regression coefficients ($N-1$ dummies used to avoid dummy variable trap).

- Restricted model: $y_{it} = \alpha + \boldsymbol{x}_t'\boldsymbol{\beta} + u_{it}$.

- Can be implemented as an F-test for linear (zero) restrictions: Pooled regression is compared to LSDV model.

`pooltest()` – F-test of stability (Chow test)

- Test for data poolability. Test of stability (or Chow test) for the coefficients of a panel model.
- We allow for different intercepts & tests for equal slopes in all CS-units. `R` implementation compares pooling and FE estimators. Algorithm outline:
  1. Estimate model separately for each CS unit (ignore a_i).
  2. Compare with FE estimator (allow individual effects, impose common slopes on regressors) using an F-test
  - Are the slopes (β-coefficients) identical among CS-units?

$$H_0: \ \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_N$$
$$H_1: \ \neg H_0$$

- **Drawback:** test cannot handle time-invariant regressors – as the unrestricted model is estimated individually for each CS-unit, such regressors are perfectly correlated with the intercept. With FE estimator, all time-invariant regressors are eliminated.
- Single-regressor example:

Unrestricted model: $y_{it} = \alpha_i + \beta_{i1} x_{it} + u_{it}, \quad i = 1, 2, \dots, N$

Restricted model: $y_{it} = \alpha + \beta_1 x_{it} + a_i + u_{it}, \quad$ use FE

pooltest() – F-test of stability (Chow test)

SSR_r: restricted model – allow for different a_i, impute common slopes.

SSR_{ur}:run a regression for each of the CS units. $SSR_{ur} = SSR_1 + SSR_2 + \cdots + SSR_N$

$N + NK$ parameters estimated in the unrestricted model, K is # regressors

$$F = \frac{SSR_r - SSR_{ur}}{SSR_{ur}} \cdot \frac{(NT - N - NK)}{(N-1)K};$$

under H_0 of common slopes (no structural break),

$$F \sim F[(N-1)K, (NT - N - NK)]$$

- R implementation: `pooltest()` from the `{plm}` package.

`pFtest()` for unobserved effects

$$y_{it} = \alpha + \boldsymbol{x}'_{it}\boldsymbol{\beta} + a_i + \lambda_t + \nu_{it}$$

- Alternative test for panel model validity.

  F-test for significance of unobserved effects. Significances of either "individual", "time" or "twoways" effects can be tested.

- Based on comparing FE-estimator against the pooling model.

- d.f. of the F-test statistic depend on the number of observations and parameters restricted:
  `df1` is the number of restrictions (parameters restricted),
  `df2` $= N(T-1) - (\#$ parameters in the unrestricted model$)$

- Hence, two main arguments to the test function are `plm`-estimated "pooling" and "within" models.

- Implementation: `pFtest()` from the `{plm}` package

Honda (1985) test for individual and time effects

- Using OLS-based ("pooling") residuals, we test the null hypothesis of redundant individual (a_i) and/or time (λ_t) effects.
- This LM-based tests uses residuals of the pooling model.

  In `R`, if this test is performed on RE of FE model, corresponding pooling model is calculated internally first.
- Implementation:

  `plmtest(..., type="honda")` from the `{plm}` package

Honda (1985) test for individual and time effects

To describe Honda test, we start by casting the panel model:

- $y_{it} = \alpha + \boldsymbol{x}'_{it}\boldsymbol{\beta} + u_{it}$ where $u_{it} = a_i + \lambda_t + \nu_{it}$
- Assumptions for Honda (1985) test:

  i.i.d. individual effects: $a_i \sim N(0, \sigma_a^2)$;
  i.i.d. time effects: $\lambda_t \sim N(0, \sigma_\lambda^2)$;
  i.i.d. idiosyncratic errors: $\nu_{it} \sim N(0, \sigma_\nu^2)$.

- Null hypotheses to be tested:
  - $H_0^a : \sigma_a^2 = 0$ (no individual effects)
  - $H_0^\lambda : \sigma_\lambda^2 = 0$ (no time effects)
  - $H_0^{a\lambda} : \sigma_a^2 = \sigma_\lambda^2 = 0$ (no individual nor time effects)

Honda (1985) test for individual and time effects

$$y_{it} = \alpha + \boldsymbol{x}'_{it}\boldsymbol{\beta} + u_{it} \qquad \text{where } u_{it} = a_i + \lambda_t + \nu_{it}$$

Balanced panel assumed.

- Error component in vector form:
  $\boldsymbol{u}_i = (u_{i1}, u_{i2}, \ldots, u_{iT})'$ and $\boldsymbol{u} = (\boldsymbol{u}'_1, \boldsymbol{u}'_2, \ldots, \boldsymbol{u}'_N)'$
  $\boldsymbol{u}_i$ is $T \times 1$ and $\boldsymbol{u}$ is $NT \times 1$.

- In matrix form, $\boldsymbol{u}$ can be cast as:
  $\boldsymbol{u} = \boldsymbol{D}_a\boldsymbol{a} + \boldsymbol{D}_\lambda\boldsymbol{\lambda} + \boldsymbol{\nu}$
  where
  $\boldsymbol{a} = (a_1, \ldots, a_N)'$,
  $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_T)'$,
  $\boldsymbol{\nu}$ follows the structure of $\boldsymbol{u}$,
  $\boldsymbol{D}_a = (\boldsymbol{I}_N \otimes \boldsymbol{\iota}_T)$ i.e. $\boldsymbol{I}_N$ with each row repeated T-times; $(NT \times N)$,
  $\boldsymbol{D}_\lambda = (\boldsymbol{\iota}_N \otimes \boldsymbol{I}_T)$ i.e. $\boldsymbol{I}_T$ stacked vertically N-times; $(NT \times T)$, note that time is the "fast index" here.

Honda (1985) test for individual and time effects

$$y_{it} = \alpha + \boldsymbol{x}'_{it}\boldsymbol{\beta} + u_{it} \qquad \text{where } u_{it} = a_i + \lambda_t + \nu_{it}$$

$$\boldsymbol{u} = \boldsymbol{D}_a\boldsymbol{a} + \boldsymbol{D}_\lambda\boldsymbol{\lambda} + \boldsymbol{\nu}$$

- $\boldsymbol{D}_a\boldsymbol{D}'_a = (\boldsymbol{I}_N \otimes \boldsymbol{J}_T)$ i.e. block-diagonal matrix of $\boldsymbol{J}_T$-matrices
  where $\boldsymbol{J}_T = \iota_T\iota'_T$ ($\boldsymbol{J}_T$ is a $T \times T$ matrix of ones).
- $\boldsymbol{D}_\lambda\boldsymbol{D}'_\lambda = (\boldsymbol{J}_N \otimes \boldsymbol{I}_T)$ i.e. $N \times N$ array of $\boldsymbol{I}_T$-matrices.
- Now, we define
  $A_r = \left[\left(\frac{\boldsymbol{u}'\boldsymbol{D}_r\boldsymbol{D}'_r\boldsymbol{u}}{\boldsymbol{u}'\boldsymbol{u}}\right) - 1\right]$ for $r = a$ or $r = \lambda$.

Honda (1985) test for individual and time effects

$y_{it} = \alpha + \boldsymbol{x}'_{it}\boldsymbol{\beta} + u_{it} \qquad$ where $u_{it} = a_i + \lambda_t + \nu_{it}$
Balanced panel assumed.

- Honda (1985) provides (uniformly most powerful) *LM* statistics for $H_0^a : \sigma_a^2 = 0$ against a one-sided $H_1^a : \sigma_a^2 > 0$:

$$\mathrm{HO}_a = \sqrt{\frac{NT}{2(T-1)}}\, A_a \underset{H_0}{\rightarrow} N(0,1)$$

- Similarly, for $H_0^\lambda : \sigma_\lambda^2 = 0$ against a one-sided $H_1^\lambda : \sigma_\lambda^2 > 0$:

$$\mathrm{HO}_\lambda = \sqrt{\frac{NT}{2(T-1)}}\, A_\lambda \underset{H_0}{\rightarrow} N(0,1)$$

Honda (1985) test for individual and time effects

$y_{it} = \alpha + \boldsymbol{x}'_{it}\boldsymbol{\beta} + u_{it} \qquad$ where $u_{it} = a_i + \lambda_t + \nu_{it}$
Balanced panel assumed.

- Honda (1985) provides a test statistic for

  $H_0^{a\lambda} : \sigma_a^2 = \sigma_\lambda^2 = 0$ against a one-sided alternative.

$$\mathrm{HO}_{a\lambda} = \frac{\mathrm{HO}_a + \mathrm{HO}_\lambda}{\sqrt{2}} \rightarrow N(0,1)$$

- Honda (1985) statistics can be generalized to the unbalanced case – see e.g.: http://www.eviews.com/help/

Cross-sectional dependency (XSD)

- In principle, XSD is similar to serial correlation in TS data.
- Can arise if individuals respond to common shocks or if spatial autocorrelation processes are present (i.e. processes relating individuals based on their distances).
- If XSD is present, the consequence is, at a minimum, inefficiency of the usual estimators and invalid inference when using the standard covariance matrix.
- In `{plm}`, only misspecification tests to detect XSD are available – no robust method to perform valid inference in its presence.
- In case of spatially determined XSD, spatial (spatial panel) econometric models should be used (discussed separately).

Cross-sectional dependency (XSD)

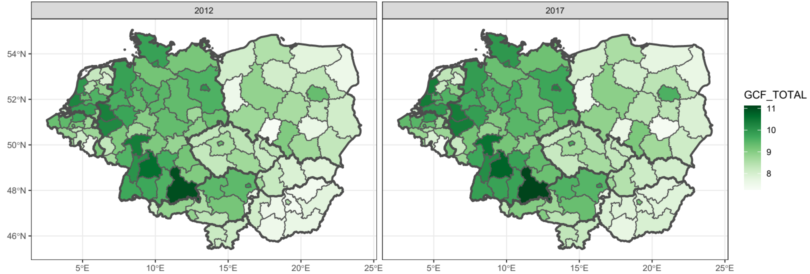


Figure 1: Total gross fixed capital formation; 2015 fixed prices, log-transformed EUR values, years 2012 and 2017 shown

- Spatial dependency is a common form of XSD. For details, see:
  https://cran.r-project.org/web/packages/spatialreg/index.html
- Other forms of XSD may be linked to non-spatially defined groups (e.g. on social networks), etc.

Cross-sectional dependency (XSD) test

- `pcdtest()` from the `{plm}` package:
- Test based on transformations of the correlation coefficient of model residuals, defined as

$$\hat{\rho}_{ij} = \frac{\sum_{t=1}^{T} \hat{u}_{it}\hat{u}_{jt}}{\left(\sum_{t=1}^{T} \hat{u}_{it}^2\right)^{1/2} \left(\sum_{t=1}^{T} \hat{u}_{jt}^2\right)^{1/2}}$$

  i.e. – we use averages over the time dimension of pairwise correlation coefficients for each pair of CS-units.

- Pesaran's CD test (Pesaran, 2004):

$$\mathrm{CD} = \sqrt{\frac{2}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \sqrt{T_{ij}}\, \hat{\rho}_{ij} \right) \underset{H_0}{\rightarrow} N(0,1)$$

  CD test is appropriate in both N and T-asymptotic settings. Also, CD test has good performance in samples of any practically relevant size and is robust to a variety of settings.

Estimator selection (FD vs FE; FE vs RE)

- FD vs FE estimators
- FE vs RE estimators

FE vs FD estimator

- For $T = 2$, FE and FD estimators produce identical estimates and inference. (FE must include a time dummy for the second period to be actually identical to the FD estimation output)
- For $T > 2$, FE and FD are both unbiased under FE.1 - FE.4. Both FE and FD are consistent for fixed T as $N \to \infty$
- If u_{it} is not serially correlated, FE is more efficient than FD
- If u_{it} follows a random walk (hence Δu_{it} is serially uncorrelated) FD is better than FE.
- If u_{it} shows some level of positive serial correlation (not a random walk), FD and FE may not be easily compared. For negative correlation of u_{it}, we prefer FE.

FE vs FD estimator

- As the time dimension increases, especially if non-stationary series are involved, FE may lead to spurious regression problems, while the FD-approach helps us with transforming integrated series into weakly dependent series.

- If strict exogeneity is violated, both FE and FD are biased. However, FE is likely to have less bias than FD (unless $T = 2$). The bias of FD does not depend on T, while the bias in FE tends to zero at rate $1/T$.

- ...it may be a good idea to use both FD and FE. If the results are not method-sensitive, so much the better. If the results from FE and FD differ significantly, we sometimes report both.

FD vs FE estimator: Wooldridge's FD-based test

$y_{it} = \alpha + \boldsymbol{x}'_{it}\boldsymbol{\beta} + a_i + \nu_{it}$

- Serial correlation test that can be used as a specification test to choose the most efficient estimator – FD vs FE.

★ If ν_{it} are not serially correlated, then:
  - Residuals in the FD model: $e_{it} = \nu_{it} - \nu_{i,t-1}$ are correlated, with $\operatorname{cor}\left(e_{it}, e_{i,t-1}\right) = -0.5$.
  - FE is more efficient than FD.

- For models with individual effects, the test can be based on estimating the model $\hat{e}_{it} = \delta \hat{e}_{i,t-1} + \eta_{it}$ based on residuals of the FD model, where we test $H_0 : \delta = -0.5$, corresponding to the null of no serial correlation in the original (undifferenced) residuals ν_{it}.

- Implementation: `pwfdtest(..., h0="fe")`
  H_0 : no serial correlation in FE-errors ν_{it},
  if not rejected, use FE.

- Test does not rely on large–T asymptotics and has good properties in short panels.

FD vs FE estimator: Wooldridge's FD-based test

$$y_{it} = \alpha + \boldsymbol{x}'_{it}\boldsymbol{\beta} + a_i + \nu_{it}$$

★ If ν_{it} follow a random walk (RW):

- Residuals in the FE model: $\nu_{it} = \nu_{i,t-1} + e_{it}, \quad$ (RW).
- Residuals in the FD model: $e_{it} = \nu_{it} - \nu_{i,t-1}$ are not serially correlated.
- FD is more efficient than FE.

- `pwfdtest(..., h0="fd")`
  H_0 : no serial correlation in FD-errors e_{it},
  if not rejected, use FD.

- If both null hypotheses `pwfdtest(..., h0="fe")` and `pwfdtest(..., h0="fd")` are rejected, whichever estimator is chosen will have serially correlated errors: use the autocorrelation-robust covariance estimators.

RE vs FE estimator: Hausman test

- **Hausman test** is based on the comparison of two sets of estimates – RE and FE.

- A classical application of the Hausman test for panel data is to compare the coefficient vectors and corresponding covariance matrices of FE and RE estimators:

$$H = (\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE})^T [\widehat{\text{Avar}}(\hat{\boldsymbol{\beta}}_{FE}) - \widehat{\text{Avar}}(\hat{\boldsymbol{\beta}}_{RE})]^{-1} (\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE}) \underset{H_0}{\sim} \chi^2(m)$$

  where m is the number of regressors varying across i and t.

  H_0 $\text{cov}(\boldsymbol{x}_{it}, a_i) = 0$ ...i.e. the crucial RE assumption holds, both FE and RE are consistent (RE is efficient).

  H_1 RE assumptions violated.

- Implementation: `phtest()` from the `{plm}` package

RE vs FE estimator: Hausman test

$$H = (\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE})^T [\widehat{\text{Avar}}(\hat{\boldsymbol{\beta}}_{FE}) - \widehat{\text{Avar}}(\hat{\boldsymbol{\beta}}_{RE})]^{-1} (\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE}) \underset{H_0}{\sim} \chi^2(m)$$

- If $\hat{\boldsymbol{\beta}}_{FE}$ and $\hat{\boldsymbol{\beta}}_{RE}$ do not differ too much [or when the asymptotic variances are relatively large] we do not reject H_0.
- If we may assume RE assumptions hold, both RE and FE are consistent, RE is efficient.
- For asymptotic variance estimators ($\widehat{\text{Avar}}$), see Wooldridge (2010).
- If we reject H_0, we need to assume that RE assumptions are violated $\rightarrow$ RE is not consistent [we use FE].

RE vs FE estimator: CRE-based test

$$\text{CRE:} \;\; y_{it} = \alpha + \beta_1 x_{it} + \gamma \bar{x}_i + r_i + u_{it}$$

CRE allows for FE vs. RE testing:

H_0: $\gamma = 0$, i.e. RE assumptions hold – can be evaluated using $\hat{\gamma}_{CRE}$ and appropriate (HCE) standard errors.

H_1: $\gamma \neq 0$ [reject $H_0 \rightarrow$ reject RE in favor of FE]

Autocorrelation, heteroscedasticity, and robust inference

- Autocorrelation & heteroscedasticity in short panels
- Autocorrelation & heteroscedasticity tests
- Robust inference

Autocorrelation & heteroscedasticity in short panels

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

- Serial correlation (between-period correlation)

  $$u_{it} = \begin{cases} \boxed{\rho} u_{i,t-1} + \varepsilon_{it} \\ \boxed{\rho_i} u_{i,t-1} + \varepsilon_{it} \end{cases}$$

- Correlation between cross-sectional units (XSD)

  H_0 of no C-S dependence may be written as follows:

  $\rho_{ij} = corr(u_{it}, u_{jt}) = 0 \quad for \quad i \neq j$

  (XSD discussed separately, worth mentioning here as it is a type of autocorrelation).

- Heteroscedasticity (RE-model example):

  $$var(v_{it} \mid \boldsymbol{X}_i) = \sigma^2_{a_i} + var(u_{it} \mid \boldsymbol{X}_i) = \begin{cases} \sigma^2_{a_i} + \boxed{\sigma^2_{u_i}} \\ \sigma^2_{a_i} + \boxed{\sigma^2_{u_t}} \end{cases}$$

Serial correlation tests (RE model)

- `pwtest()` Unobserved effects: "Wooldridge"-type test

- $$W = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \hat{u}_{it}\hat{u}_{is}}{\left[\sum_{i=1}^{N} \left(\sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \hat{u}_{it}\hat{u}_{is}\right)^2\right]^{1/2}} \underset{H_0}{\sim} N(0,1); \quad \text{(asympt.)},$$

  test does not rely on homoscedasticity assumptions.

- $H_0 : \sigma_a^2 = 0$, i.e., no unobserved effects in the residuals of RE model. [Note: technically, H_0 only states $\text{var}(a_i) = 0$].
- Test has power both against the RE specification ($\sigma_a^2 = 0$), as well as against any kind of serial correlation in error terms. Test "nests" both RE and serial correlation tests, trading some power (against more specific alternatives) in exchange for robustness.
- Not rejecting the null favours the use of pooled OLS. Rejection may follow from two sources (including serial correlation) & doesn't truly support RE specification.

Serial correlation tests (RE model)

- `pbsytest()`   Bera, Sosa-Escudero, Yoon (2001)
- Locally robust LM-tests for serial correlation or random effects. Solution to the previous problem: can distinguish between random effect and serial correlation.
- Three tests (of the RE-type model):
  - `test = "ar"` for H_0 : no serial correlation while controlling for random effects
  - `test = "re"` for H_0 : no random effects (while controlling for possible ser. corr.)
  - `test = "j"` for H_0 : no random effects & no serial correlation.
- `R` implementation can handle both balanced and unbalanced panels. For detailed description of both tests, see: Wooldridge, 2002 & `https://www.jstatsoft.org/article/view/v027i02`

Serial correlation tests (general)

- `pbgtest()` Direct generalization of the Breusch-Godfrey test for panels, mainly for RE (and pooling) models.
- Under RE assumptions of homoskedasticity and no serial correlation in the idiosyncratic error, residuals of the quasi-demeaned regression must be spherical as well. Hence, serial correlation test (BG test) is applied to residuals in the quasi-demeaned model (may be applied to pooled OLS residuals as well).
- Technically, `pbgtest()` is a wrapper to `bgtest()` from the `lmtest()` package.
- With BG-test, we can test for different orders of serial correlation.
- NOT suited for FE-estimated models, for $N \gg T$, test is severely biased towards rejecting H_0 of no ser. corr.
- `pdwtest()` Durbin-Watson test for panels (...analogous).

Serial correlation tests (general & FE)

$y_{it} = \alpha + \boldsymbol{x}_{it}'\boldsymbol{\beta} + a_i + u_{it}$

- `pwartest()` Wooldridge test for FE model (short panels).
- Under the null hypothesis of no serial correlation in the idiosyncratic errors u_{it}, residuals in the FE-estimated model (time demeaned data) are correlated:

$$\text{cor}\left(e_{it}, e_{i,t-1}\right) = -1/(T-1).$$

- H_0 of no serial correlation in u_{it} can be tested using residuals from the FE-estimated model and auxiliary regression:

$$\hat{e}_{it} = \alpha + \delta\, \hat{e}_{i,t-1} + \eta_{it}$$

  By rejecting $H_0 : \delta = -1/(T-1)$, we reject the original null hypothesis of no serial correlation in u_{it}.
- Applicable to any "FE model", particularly with $N \gg T$.
- As T grows, $-1/(T-1) \rightarrow 0$ & `pbgtest()` can be used.

Robust inference in short panel data models

- Robust inference
- Covariance matrix – White 1
- Covariance matrix – White 2
- Covariance matrix – Arellano

Robust statistical inference

- Implementation: `vcovHC()` from the `{plm}` package, used together with functions from `{lmtest}`
- Three types of HC/HAC covariance matrix estimators are based on the general White's "sandwich estimator". The CS-data version can be cast as:

$$\text{var}\left(\hat{\boldsymbol{\beta}}|\boldsymbol{X}\right) = \left[\boldsymbol{X}'\boldsymbol{X}\right]^{-1}\left[\boldsymbol{X}'\boldsymbol{\Sigma}\boldsymbol{X}\right]\left[\boldsymbol{X}'\boldsymbol{X}\right]^{-1}$$

- For the panel extension of White's HC/HAC estimator, we assume XSD-independence: no correlation between errors of different CS-units (groups), while allowing for heteroscedasticity across CS-units (and for serial correlation).

Robust statistical inference

- `vcovHC(...  , method="white1")`
- `"white1"`: heteroscedasticity-consistent approach to covariance matrix $\mathbf{\Sigma}$ estimation. Allows for general heteroscedasticity but no XSD nor serial correlation, i.e., we assume:

$$\mathbf{\Sigma}_i = \begin{bmatrix} \sigma_{i1}^2 & \dots & \dots & 0 \\ 0 & \sigma_{i2}^2 & & 0 \\ \vdots & & \ddots & 0 \\ 0 & \dots & \dots & \sigma_{iT}^2 \end{bmatrix}$$

and $\mathbf{\Sigma}$ is a block-diagonal matrix of $\mathbf{\Sigma}_i$ matrices.

✓ `white1` can be used for RE models, does not rely on large N asymptotics.

✗ Even if errors are uncorrelated, FE induces autocorrelation in residuals of transformed model $[\mathrm{cor}\,(e_{it}, e_{i,t-1}) = -1/(T-1)]$. Hence, `white1` is inconsistent (fixed T as $N \to \infty$). In this case it is advisable to use the `arellano` version.

Robust statistical inference

- `vcovHC(...   , method="white2")`
- `"white2"` is a special case of `"white1"`, with constant variance "inside" every CS unit: $\boldsymbol{\Sigma}_i = \sigma_i^2 \boldsymbol{I}_T$. Again, $\boldsymbol{\Sigma}$ is a block-diagonal matrix of $\boldsymbol{\Sigma}_i$ matrices.
- FE/RE features analogous to `"white1"`.

- Note (relevant for all three robust estimators):

  The counterpart to CS-related sandwich estimator element $[\boldsymbol{X}'\boldsymbol{\Sigma}\boldsymbol{X}]$ would be:

$$\ddot{\boldsymbol{X}}'\boldsymbol{\Sigma}\ddot{\boldsymbol{X}} = \sum_{i=1}^{N} \left(\ddot{\boldsymbol{X}}_i'\boldsymbol{\Sigma}_i\ddot{\boldsymbol{X}}_i \right)$$

  where $\ddot{\boldsymbol{X}}$ are the transformed regressors.

Robust statistical inference

- `vcovHC(...  , method="arellano")`
- `"arellano"` allows a fully general structure w.r.t. heteroscedasticity and serial correlation (no XSD):

$$\mathbf{\Sigma}_i = \begin{bmatrix} \sigma_{i1}^2 & \sigma_{i1,i2} & \dots & \dots & \sigma_{i1,iT} \\ \sigma_{i2,i1} & \sigma_{i2}^2 & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \sigma_{iT-1}^2 & \sigma_{iT-1,iT} \\ \sigma_{iT,i1} & \dots & \dots & \sigma_{iT,iT-1} & \sigma_{iT}^2 \end{bmatrix}$$

  and $\mathbf{\Sigma}$ is a block-diagonal matrix of $\mathbf{\Sigma}_i$ matrices
- `"arellano"`: consistent w.r.t. timewise correlation of the errors, but (unlike `"white1"`,`"white2"`), it relies on large N asymptotics with small T (short panels).

  Typical `"arellano"` use: FE & large N.

Long panels – models and estimation

- Quick repetition of relevant topics from BSc courses
- Long panels and the SUR model / SURE
- Long panels and the general SUR model
- SURE & equations with identical regressors
- SURE & “pooled” model
- SURE – FGLS

Long panels – models and estimation (BSc repetition)

General LRM (TS-based): $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$

The following cases of $\boldsymbol{\Omega} = \text{var}(\boldsymbol{\varepsilon}|\boldsymbol{X})$ can occur:

(a) ε_t *i.i.d.* – corresponds to a CLRM:

$$\text{var}(\boldsymbol{\varepsilon}|\boldsymbol{X}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\boldsymbol{X}) = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \boldsymbol{I}_T$$

(b) ε_t under heteroscedasticity (no `ar(p)` process present)

$$\text{var}(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & \sigma_N^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} h_1 & 0 & \cdots & 0 \\ 0 & h_2 & \cdots & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \cdots & h_N \end{bmatrix} = \sigma^2 \boldsymbol{H}$$

i.e. $\sigma_t^2 = \sigma^2 h_{tt}$ and $[h_{ts}] \geq 0$.

Long panels – models and estimation (BSc repetition)

General LRM (TS-based): $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

(c) ε_t with `ar(1)` (no heteroscedasticity):

$$\mathrm{var}(\boldsymbol{\varepsilon}|\boldsymbol{X}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\boldsymbol{X}) = \frac{\sigma_e^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{n-2} & \dots & \rho & 1 & \rho \\ \rho^{n-1} & \dots & \rho^2 & \rho & 1 \end{bmatrix} = \frac{\sigma_e^2}{1-\rho^2}\boldsymbol{H}$$

- From $y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + \varepsilon_t$ and $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$, we get:
  $\varepsilon_t = u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \cdots$, by repeated substitution.
  Now, $\mathrm{var}(u_t) = \sigma_u^2 + \rho^2\sigma_u^2 + \rho^4\sigma_u^2 + \cdots$, since u are *i.i.d.*
  and the variance-covariance matrix follows from
  $\mathrm{cov}(\varepsilon_t, \varepsilon_{t-s}) = \frac{\rho^s\sigma_u^2}{1-\rho^2}$, provided $|\rho| < 1$.
  (see Greene, Econometric analysis 7th ed., ch. 20.3.20)

(d) ε_t: general case (both heteroscedasticity and `ar(p)` may be present

$\mathrm{var}(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \boldsymbol{\Omega}$, where $\boldsymbol{\Omega}$ is a $(T \times T)$ PSD matrix.

Long panels – models and estimation (BSc repetition)

General LRM: $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ & OLS vs GLS:
(for $t = 1, 2, \ldots, T$ observations)

- $\text{var}(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \sigma^2 \boldsymbol{I}_N \quad \rightarrow \quad$ use OLS (BLUE, assumptions apply):

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

- $\text{var}(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \sigma^2 \boldsymbol{H} \quad \rightarrow \quad$ use GLS (efficent w.r.t. OLS):

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{H}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{H}^{-1}\boldsymbol{y}$$

- FGLS: For empirical applications, we usually have to find $\hat{\boldsymbol{H}}$, i.e. some "good" estimate of the unobserved $\boldsymbol{H}$.

Kronecker product (for two general matrices $\boldsymbol{A}$ and $\boldsymbol{B}$):

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \cdots & a_{1K}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \cdots & a_{2K}\boldsymbol{B} \\ & & \cdots & \\ a_{N1}\boldsymbol{B} & a_{N2}\boldsymbol{B} & \cdots & a_{NK}\boldsymbol{B} \end{bmatrix}$$

For the Kronecker product:

- $(\boldsymbol{A} \otimes \boldsymbol{B})' = \boldsymbol{A}' \otimes \boldsymbol{B}'$
- $(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}$
- $(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = \boldsymbol{AC} \otimes \boldsymbol{BD}$

  ... given conforming dimensions of the matrices.

Kronecker product example:

- $\boldsymbol{A} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}, \qquad \boldsymbol{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
- $\boldsymbol{A} \otimes \boldsymbol{I}_2 = \begin{bmatrix} a & 0 & b & 0 \\ 0 & a & 0 & b \\ b & 0 & c & 0 \\ 0 & b & 0 & c \end{bmatrix}$
- $\boldsymbol{I}_2 \otimes \boldsymbol{A} = \begin{bmatrix} a & b & 0 & 0 \\ b & c & 0 & 0 \\ 0 & 0 & a & b \\ 0 & 0 & b & c \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A} \end{bmatrix}$

  …i.e. the result is a block-diagonal matrix

Long panels and SUR models

Seemingly unrelated regression equations (SUR/SURE):

- Consider $i = 1, \ldots, M$ individuals (CS units) and $t = 1, \ldots, T$ observations for each individual (while t suggest time, SURE may extend to hierarchical CS data as well).

- Individual regression equations have a common structure:

$$\begin{aligned}
\boldsymbol{y}_1 &= \boldsymbol{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1, \\
\boldsymbol{y}_2 &= \boldsymbol{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2, \\
&\cdots \\
\boldsymbol{y}_M &= \boldsymbol{X}_M \boldsymbol{\beta}_M + \boldsymbol{\varepsilon}_M;
\end{aligned}$$

general form notation: $\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i, \qquad i = 1, \ldots, M.$

- Example: Unemployment dynamics in Germany (NUTS1, $M = 16$):

$$Unemp_{it} = \beta_{1i} + \beta_{2i} \log(GDP_{it}) + \quad \cdots \quad + \varepsilon_{it}$$

Long panels and SUR models

- $\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i, \qquad i = 1, \dots M, \qquad \boldsymbol{y}_i = (y_{i1}, y_{i2}, \dots, y_{iT})',$

  can be written in stacked matrix form as:

- $$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_M \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{X}_2 & \cdots & \mathbf{0} \\ & & \vdots & \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{X}_M \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_M \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_M \end{bmatrix} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

- For the $MT \times 1$ vector of disturbances $\boldsymbol{\varepsilon}$, we assume:
  - Strict exogeneity: $E[\boldsymbol{\varepsilon} | \boldsymbol{X}_1, \dots, \boldsymbol{X}_M] = \mathbf{0}$,
  - Homoscedasticity in CS units: $E[\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i' | \boldsymbol{X}_1, \dots, \boldsymbol{X}_M] = \sigma_{ii} \boldsymbol{I}_T$ (σ_{ii} – error variance for ith unit, notation follows Greene).
  - Disturbances uncorrelated across T but contemporaneously correlated between CS units (equations):
    $E[\varepsilon_{it}\varepsilon_{js} | \boldsymbol{X}_1, \dots, \boldsymbol{X}_M] = \sigma_{ij} \quad$ if $t = s$; 0 otherwise.

- Equation by equation OLS estimation: consistent.

- GLS is efficent w.r.t. OLS: uses information on contemporaneous correlation among errors as in the matrix $\boldsymbol{\Sigma} = [\sigma_{ij}]$.

Long panels and SUR models

Three types of SUR/SURE models:

1. General SURE: Model with distinct (general) $\boldsymbol{X}_i$ matrices and distinct $\boldsymbol{\beta}_i$ vectors. Equation-by-equation OLS estimator is consistent, FGLS is efficient (details & conditions discussed next).

2. SURE model with identical $\boldsymbol{X}_i$ blocks and (generally) distinct $\boldsymbol{\beta}_i$ vectors. FGLS not relevant as GLS is identical to equation-by-equation OLS.

3. SURE model with distinct (general) $\boldsymbol{X}_i$ matrices and identical $\boldsymbol{\beta}_i$ coefficients. "Pooled" OLS estimation is consistent, FGLS is efficient (details & conditions discussed next).

Long panels – SUR "general model"

The general case of SUR model, with distinct $\boldsymbol{X}_i$ regressors and $\boldsymbol{\beta}_i$ coefficients:

- $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ model can be written as:

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_M \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{X}_2 & \cdots & \mathbf{0} \\ & & \vdots & \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{X}_M \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_M \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_M \end{bmatrix} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

- $\hat{\boldsymbol{\beta}}_{\text{GLS}} = [\boldsymbol{X}'(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_T)^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X}'(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_T)^{-1}\boldsymbol{y}$
  GLS estimator has the same form as in "pooled" case, yet $\boldsymbol{X}$ and $\boldsymbol{\beta}$ dimensions are different.

- GLS computation assumes $\boldsymbol{\Sigma}$ is known, which is unlikely (with FGLS, $\boldsymbol{\Sigma}$ is estimated).

Long panels – general SUR models

- Stacked matrix form of the SUR model:
  $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{y}$ is $(MT \times 1)$, $\boldsymbol{X}$ is block-diagonal, etc.
- $\boldsymbol{\Sigma}$ can be constructed from the vector of errors $\boldsymbol{\varepsilon}' = (\boldsymbol{\varepsilon}_1', \dots, \boldsymbol{\varepsilon}_M')'$ as follows:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2M} \\ & & \vdots & \\ \sigma_{M1} & \sigma_{M2} & \cdots & \sigma_{MM} \end{bmatrix} = [\sigma_{ij}],$$

- the variance-covariance matrix for $\boldsymbol{\varepsilon}$ is given as $\boldsymbol{\Omega}$ $(MT \times MT)$:

$$\boldsymbol{\Omega} = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_T = \begin{bmatrix} \sigma_{11}\boldsymbol{I}_T & \sigma_{12}\boldsymbol{I}_T & \cdots & \sigma_{1M}\boldsymbol{I}_T \\ \sigma_{21}\boldsymbol{I}_T & \sigma_{22}\boldsymbol{I}_T & \cdots & \sigma_{2M}\boldsymbol{I}_T \\ & & \vdots & \\ \sigma_{M1}\boldsymbol{I}_T & \sigma_{M2}\boldsymbol{I}_T & \cdots & \sigma_{MM}\boldsymbol{I}_T \end{bmatrix}.$$

  This implies both heteroscedasticity (non-constant elements on the main diagonal) and autocorrelation (non-zero off-diagonal elements).

Long panels – general SUR models

- Stacked matrix form of the SUR model:

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{X}\beta + \boldsymbol{\varepsilon}, \\ \mathrm{var}(\boldsymbol{\varepsilon}) &= \boldsymbol{\Omega} = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_T. \end{aligned}$$

- The GLS estimator for SUR model (SURE):

$$\begin{aligned} \hat{\beta}_{\mathrm{GLS}} &= [\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{y} \\ &= [\boldsymbol{X}'(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_T)^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X}'(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_T)^{-1}\boldsymbol{y} \end{aligned}$$

- Asymptotic covariance matrix of the GLS estimator:

$$\begin{aligned} \mathrm{Asy.Cov}(\hat{\beta}_{\mathrm{GLS}}) &= [\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}]^{-1} \\ &= [\boldsymbol{X}'(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_T)^{-1}\boldsymbol{X}]^{-1} \end{aligned}$$

Long panels – general SUR models

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad \text{var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Omega} = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_T.$$

$$\hat{\beta}_{\text{GLS}} = [\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{y}$$

- SURE: how much efficiency over OLS is gained by GLS (SURE)?
  - Higher correlation of disturbances $\rightarrow$ higher efficiency gain.
  - SUR equations actually unrelated ($\sigma_{ij} = 0$, for $i \neq j$): no payoff in GLS.
  - The less correlation between the $\boldsymbol{X}$ matrices, the greater is the gain in efficency in using GLS (w.r.t. OLS).
  - SUR model with identical regressors ($\boldsymbol{X}_1 = \boldsymbol{X}_2 = \cdots = \boldsymbol{X}_M$): OLS and GLS are identical (discussed on next page).
- Homogeneity restrictions – equal coefficients in all equations of the SUR model (analogous to 'pooling OLS' model: $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_M$, i.e. $(M-1)K$ restrictions on the $(KM \times 1)$ vector $\boldsymbol{\beta}$ ...can be tested using Wald, LR and/or LM tests.

Long panels – SUR models (identical regressors)

SUR models with identical $\boldsymbol{X}_i$ regressors $\boldsymbol{X}_1 = \boldsymbol{X}_2 = \cdots = \boldsymbol{X}_M$:

Topic is partially out of scope in terms of long-panel data. However, SUR models with identical regressors have important empirical applications:

- VAR models (discussed separately in this course).
- Capital asset pricing model (for a given financial instrument):

$$r_{it} - r_{ft} = \alpha_i + \beta_i(r_{mt} - r_{ft}) + \varepsilon_{it}$$

  where r_{it} is the return of instrument i over time period t, r_{ft} and r_{mt} describe risk-free and market returns respectively; α_i and β_i are parameters, estimated separately for each ith financial isntrument – same regressor $(r_{mt} - r_{ft})$ used in each regression equation.

SURE & equations with identical regressors

SUR models with identical $\boldsymbol{X}_i$ regressors $\boldsymbol{X}_1 = \boldsymbol{X}_2 = \cdots = \boldsymbol{X}_M$:

- $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where: $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_i & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{X}_i & \cdots & \mathbf{0} \\ & & \vdots & \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{X}_i \end{bmatrix} = \boldsymbol{I}_M \otimes \boldsymbol{X}_i$

- $\hat{\boldsymbol{\beta}}_{\text{GLS}} = [\boldsymbol{X}'(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_T)^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X}'(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_T)^{-1}\boldsymbol{y}$

  Note that:

  - $\boldsymbol{X}' = (\boldsymbol{I}_M \otimes \boldsymbol{X}_i)' = \boldsymbol{I}_M \otimes \boldsymbol{X}_i'$,
  - $(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_T)^{-1} = \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_T$,
  - $\boldsymbol{X}'(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_T)^{-1} = (\boldsymbol{I}_M \otimes \boldsymbol{X}_i')(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_T) = \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{X}_i'$,
  - $[\boldsymbol{X}'(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_T)^{-1}\boldsymbol{X}] = (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{X}_i')(\boldsymbol{I}_M \otimes \boldsymbol{X}_i) = \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{X}_i'\boldsymbol{X}_i$,
  - $[\boldsymbol{X}'(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_T)^{-1}\boldsymbol{X}]^{-1} = \boldsymbol{\Sigma} \otimes (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}$

SURE & equations with identical regressors

SUR models with identical $\boldsymbol{X}_i$ regressors $\boldsymbol{X}_1 = \boldsymbol{X}_2 = \cdots = \boldsymbol{X}_M$:

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{\text{GLS}} &= [\boldsymbol{X}'(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_T)^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X}'(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_T)^{-1}\boldsymbol{y} \\
&= (\boldsymbol{\Sigma} \otimes (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1})(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{X}_i')\boldsymbol{y} \\
&= \left[\boldsymbol{I}_M \otimes (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\right]\boldsymbol{y}
\end{aligned}$$

$$= \begin{bmatrix} (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i' & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i' & \cdots & \mathbf{0} \\ & & \vdots & \\ \mathbf{0} & \mathbf{0} & \cdots & (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i' \end{bmatrix} \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_M \end{bmatrix}$$

$$= \begin{bmatrix} (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{y}_1 \\ (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{y}_2 \\ \vdots \\ (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{y}_M \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_{1,\text{OLS}} \\ \hat{\boldsymbol{\beta}}_{2,\text{OLS}} \\ \vdots \\ \hat{\boldsymbol{\beta}}_{M,\text{OLS}} \end{bmatrix}$$

…equation-by-equation OLS (VAR-model implications).

Long panels – SUR "pooled model"

SUR models with the same regressors (identical dimensions & variable structure across $\boldsymbol{X}_i$, yet different 'it' observations) and with all coefficient vectors assumed the same ($\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_M$):

- $\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i, \qquad i = 1, \ldots M, \qquad \boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iT})'$, can be written in stacked matrix form as:
- $$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_M \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_M \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_M \end{bmatrix} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$
- For the $MT \times 1$ vector of disturbances $\boldsymbol{\varepsilon}$, we assume:
  - Strict exogeneity: $E[\boldsymbol{\varepsilon_i} | \boldsymbol{X}] = \boldsymbol{0}$,
  - Homoscedasticity: $E[\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i' | \boldsymbol{X}] = \sigma_{ii} \boldsymbol{I}_T$.
  - Disturbances uncorrelated across T but contemporaneously correlated between CS units (equations):
    $E[\varepsilon_{it}\varepsilon_{js} | \boldsymbol{X}] = \sigma_{ij} \quad$ if $t = s$; 0 otherwise.
    Hence:
    $E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' | \boldsymbol{X}] = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_T, \quad$ where $\boldsymbol{\Sigma} = [\sigma_{ij}]$.

Long panels – SUR "pooled model"

SUR models with the same regressors (identical dimensions & variables across $\boldsymbol{X}_i$, yet different observations) and all coefficient vectors are assumed the same ($\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_M$):

- GLS estimator of the SUR "pooled" model:

  $$\hat{\boldsymbol{\beta}}_{\text{GLS}} = [\boldsymbol{X}'(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_T)^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X}'(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_T)^{-1}\boldsymbol{y}$$

  where $\boldsymbol{X}$ is a $(MT \times K)$ matrix – compare to the block diagonal $(MT \times MK)$ in the general SUR model

  and $\boldsymbol{\beta}$ is $(K \times 1)$ instead of the $(MK \times 1)$ for the general SUR model.

- General note: GLS computation assumes $\boldsymbol{\Sigma}$ is known, which is unlikely (with FGLS, $\boldsymbol{\Sigma}$ is estimated).

Long panels – SURE: FGLS estimator

- $\hat{\beta}_{\text{GLS}} \;\;= [\boldsymbol{X}'(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_T)^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X}'(\boldsymbol{\Sigma} \otimes \boldsymbol{I}_T)^{-1}\boldsymbol{y}$
- $\hat{\beta}_{\text{FGLS}} = [\boldsymbol{X}'(\hat{\boldsymbol{\Sigma}} \otimes \boldsymbol{I}_T)^{-1}\boldsymbol{X}]^{-1}\boldsymbol{X}'(\hat{\boldsymbol{\Sigma}} \otimes \boldsymbol{I}_T)^{-1}\boldsymbol{y}$

- FGLS estimator is based on OLS-estimated residuals $\boldsymbol{e}$:

$\hat{\sigma}_{ij} = \frac{1}{T}\boldsymbol{e}_i'\boldsymbol{e}_j$ and $\hat{\boldsymbol{\Sigma}} = [\hat{\sigma}_{ij}]$ is estimated as follows:

1. SURE, model with identical $\boldsymbol{X}_i$ blocks: FGLS not relevant as GLS = equation-by-equation OLS.
2. "pooled" case with identical $\boldsymbol{\beta}_i$ coefficients, where $\boldsymbol{X}$ is $(MT \times K)$: $\boldsymbol{e}_i$ is a subvector of OLS residuals from $\hat{\boldsymbol{\beta}}_{\text{OLS}} = [\boldsymbol{X}'\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{y}$.
3. "general case" SURE: $\boldsymbol{e}_i$ vectors come from equation-by-equation OLS: $\hat{\boldsymbol{\beta}}_{i,\text{OLS}} = [\boldsymbol{X}_i'\boldsymbol{X}_i]^{-1}\boldsymbol{X}_i'\boldsymbol{y}_i$, or as subvector of $\boldsymbol{e}$ from OLS estimation of the stacked model: $\hat{\boldsymbol{\beta}}_{\text{OLS}} = [\boldsymbol{X}'\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{y}$, where $\boldsymbol{X}$ is $(MT \times MK)$ – same residuals from both approaches.

Large panels – introduction

- Heterogeneous panels with strictly exogenous regressors
- Cross-sectional dependence in panels
  Spatial panel models
- Unit root and cointegration in panels

Models and notation in this section mostly follow from:
Pesaran, M.H.: Time series and panel data econometrics.

Heterogeneous panels with strictly exogenous regressors

- For stationary variables, the Swamy (1970) estimator is based on a panel model with K strictly exogenous regressors:

- $y_{it} = \sum_{k=1}^{K} \beta_{ki} x_{kit} + u_{it}, \qquad i = 1, \ldots, N; \quad t = 1, \ldots, T,$

  where coefficients $\boldsymbol{\beta}_i$ are random, with constant mean and variance-covariances:

  $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\eta}_i \ ,$

  with:

$$E(\boldsymbol{\eta}_i) = \mathbf{0},$$
$$E(\boldsymbol{\eta}_i \boldsymbol{x}'_{it}) = \mathbf{0},$$
$$E(\boldsymbol{\eta}_i, \boldsymbol{\eta}_j) = \begin{cases} \boldsymbol{\Omega}_\eta, & \text{if } \ i = j, \\ \mathbf{0}, & \text{if } \ i \neq j, \end{cases}$$

  and u_{it} is *iid* across i and t and $\text{var}(u_{it}) = \sigma_i^2$.

Swamy estimator:

- Using the substitution $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\eta}_i$, we can write the model in a stacked form:
- $\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{v}_i, \qquad$ with $\qquad \boldsymbol{v}_i = \boldsymbol{X}_i\boldsymbol{\eta}_i + \boldsymbol{u}_i,$

  which can be re-cast as:

  $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{v}$, where:

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_N \end{bmatrix}, \ \boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_N \end{bmatrix}, \ \boldsymbol{v} = \begin{bmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \\ \vdots \\ \boldsymbol{v}_N \end{bmatrix},$$

$$\boldsymbol{\Sigma} = E(\boldsymbol{v}\boldsymbol{v}') = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 & \cdots & \mathbf{0} \\ & & \ddots & \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}_N \end{bmatrix}, \quad \text{and where}$$

$\boldsymbol{\Sigma_i} = \sigma_i^2 \boldsymbol{I}_T + \boldsymbol{X}_i \boldsymbol{\Omega}_\eta \boldsymbol{X}_i'.$

Heterogeneous panels with strictly exogenous regressors

Swamy estimator:

- $$\hat{\beta}_{\text{SW}} = \left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$$
$$= \left(\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i\right)^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{y}_i,$$

- $$\operatorname{var}\left(\hat{\beta}_{\text{SW}}\right) = \left(\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i\right)^{-1},$$

  and the $\hat{\boldsymbol{\Sigma}}_i$ elements (i.e. $\hat{\sigma}_i^2$ and $\hat{\boldsymbol{\Omega}}_\eta$) can be obtained through separate OLS estimations across individual i-units.

- If errors u_{it} and $\boldsymbol{\eta}_i$ are normally distributed, parameters of the model $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\Omega}_\eta, \sigma_i^2)$ can be estimated by ML (may be computationally expensive).

Heterogeneous panels with strictly exogenous regressors

The mean group estimator (MGE):

- Alternative to Swamy's estimator (stationary variables). Defined as a simple average of OLS estimators for $\hat{\boldsymbol{\beta}}_i$:

- $\hat{\boldsymbol{\beta}}_{\mathrm{MG}} = \frac{1}{N} \sum_{i=1}^{N} \hat{\boldsymbol{\beta}}_i \, ,$

  where

  $\hat{\boldsymbol{\beta}}_i = (\boldsymbol{X}_i' \boldsymbol{X}_i)^{-1} \boldsymbol{X}_i' \boldsymbol{y}_i \, ,$

- $\mathrm{var}\left(\hat{\boldsymbol{\beta}}_{\mathrm{MG}}\right) = \frac{1}{N(N-1)} \sum_{i=1}^{N} \left(\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}}_{\mathrm{MG}}\right) \left(\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}}_{\mathrm{MG}}\right)',$

- MGE only possible if N and T are sufficiently large. It is applicable irrespective of random (Swamy-like) or "other" $\boldsymbol{\beta}$-parameter type of distribution.

Cross-sectional dependence in large panels

- Ignoring XSD may have serious consequences on estimator properties.
- **Residual multi-factor approach:** XSD can be characterized by a small number of unobserved common factors.
- **Spatial dependency approach:** discussed separately in the course 4EK417.

  https://github.com/formanektomas/4EK417/raw/master/Block3/Block_3.pdf

- Compare to other panel dimensions:

  With $N \gg T$, we may use spatial panels (data permitting).
  With $T \gg N$, we use SURE.

Residual multi-factor approach

- Outline of the approach only, detailed discussion is complex and requires definition and discussion of weak/strong XSD.
- $y_{it} = \boldsymbol{\alpha}_i' \boldsymbol{d}_t + \boldsymbol{\beta}_i' \boldsymbol{x}_{it} + u_{it}$

  is a heterogeneous panel data model where $\boldsymbol{d}_t$ is a $(N \times 1)$ vector of common effects (intercepts, seas. dummies, etc.), $\boldsymbol{x}_{it}$ is a $(K \times 1)$ vector of observed individual-specific regressors and u_{it} have the following common factor structure:

  $$u_{it} = \gamma_{i1} f_{1t} + \gamma_{i2} f_{2t} + \cdots + \gamma_{im} f_{mt} + e_{it}$$

  where $\boldsymbol{f}_t = (f_{1t}, f_{2t}, \ldots, f_{mt})'$ is an m-dimensional vector of unobserved common factors and $\boldsymbol{\gamma}$ is a corresponding vector of factor loadings (parameters).

Principal component estimator

- Model with strictly exogenous regressors and homogeneous slopes ($\boldsymbol{\beta}_i = \boldsymbol{\beta}$). Two stage approach:

1. In the first stage, principal components (PCs) are extracted from OLS residuals (which serve as proxies for unobserved variables).

2. In the second stage, an augmented regression model is estimated:

   $$y_{it} = \boldsymbol{\alpha}_i' \boldsymbol{d}_t + \boldsymbol{\beta} \boldsymbol{x}_{it} + \boldsymbol{\gamma}_i' \hat{\boldsymbol{f}}_t + e_{it},$$

   where $\hat{\boldsymbol{f}}_t$ is a vector of m "strong" components of the residuals computed in first stage. PCA & PCR discussed separately.

   https://github.com/formanektomas/4EK417/raw/master/Block2/Block_2.pdf

- In principle, this method aims at controlling for XSD through PCs, which leads to consistent $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}$ estimates. However, if the PCs and regressors are correlated (a common case), this estimator becomes inconsistent. To solve this problem (and related issues), various modifications of this estimator exist.

Unit root and cointegration in panels

Testing unit root and cointegration hypotheses on panels (as opposed to individual TS) involves multiple complications:

- large amount of unobserved heterogeneity (CS-specific parameters) obfuscates results,
- assumption of cross-sectional is often violated,
- complicated interpretation of tests, if null hypothesis (UR) is rejected – "at least some fraction of CS units is stationary",
- with $I(1)$ variables, the possibility of both "within group" and "across groups" cointegration exists,
- complicated asymptotic theory.

Unit root and cointegration in panels is a PhD-level topic, not covered by this course. For details, see e.g. Pesaran, M.H.: Time series and panel data econometrics (ch. 31).

Panel data – additional topics and extensions

- Advanced (PhD-level) course on panel data

  http://people.stern.nyu.edu/wgreene/Econometrics/PanelDataNotes.htm

- Linear/generalized linear mixed effects model
  Extension to the RE model
  (intercept and -some- coefficients have a random term):

  $$y_{it} = \boldsymbol{x}'_{it}\boldsymbol{\beta} + \boldsymbol{z}'_{it}(\boldsymbol{\gamma} + \boldsymbol{h}_i) + (\alpha + u_i) + \varepsilon_{it}$$

  where $\boldsymbol{h}_i$ describes random variation of the parameter(s) across individuals.

  http://www.bodowinter.com/tutorial/bw_LME_tutorial1.pdf