Praktikum z ekonometrie - Týden 7
Panel data models and tests

VŠE Praha

Tomáš Formánek

Content

Panel data models: quick repetition

LSDV regression

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

μ_i are usually regarded as unobservable variables.
This approach gives appropriate interpretation of $\boldsymbol{\beta}$.
Traditional (old) approaches to fixed effects estimation view the μ_i as parameters to be estimated along with $\boldsymbol{\beta}$.

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

- Define N dummy variables - one for each cross-section.
- Convenient LSDV model expansion: use interactions to control for individual slopes for chosen regressors.

FD estimator

We can eliminate unobserved individual heterogeneity from the regression: $y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + \mu_i + u_{it}$
by first differences (FD) transformation:
$\Delta y_{it} = y_{it} - y_{i,t-1} = \Delta\boldsymbol{x}_{it}\boldsymbol{\beta} + \Delta\mu_i + \Delta u_{it} = \Delta\boldsymbol{x}_{it}\boldsymbol{\beta} + \Delta u_{it}$

- ✓ 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 $t = 1; 2$ – 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: $y_{it} = \beta_1 x_{it1} + \cdots + \beta_k x_{itk} + \mu_i + u_{it}$,
$i = 1, \ldots, N, \;\; t = 1, \ldots, T$

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

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

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).

FD estimator

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).

FE estimator

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

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

$\overline{y}_i = \beta_1 \overline{x}_{i1} + \cdots + \beta_k \overline{x}_{ik} + \overline{\mu}_i + \overline{u}_i$
(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 μ_i element - as it does not vary over time

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

After FE estimation, μ_i elements may be estimated as follows:
$\hat{\mu}_i = \overline{y}_i - \hat{\beta}_1 \overline{x}_{i1} - \cdots - \hat{\beta}_k \overline{x}_{ik}, \ \ i = 1, \ldots, N$

However, in most practical applications, μ_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: $y_{it} = \beta_1 x_{it1} + \cdots + \beta_k x_{itk} + \mu_i + u_{it}$, $i = 1, \dots, N, \;\; t = 1, \dots, T$

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

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

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

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

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 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.

RE estimator

If μ_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 (appropriate if C-S units are from a large sample).

- RE models reduce the number of parameters estimated.
- RE estimator is potentially inconsistent, if assumption not met.
- $y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + \mu_i + u_{it}$
- If we can assume that μ_i is uncorrelated with each explanatory variable: $cov(\boldsymbol{x}_{it}, \mu_i) = 0; \;\; t = 1, 2, \dots, T$
  then we may drop μ_i from the equation and β_j estimates will remain unbiased.
- By dropping μ_i from the regression, we effectively create a new error term: $v_{it} = \mu_i + u_{it}$
- As μ_i is time-invariant, the random element v_{it} contains a lot of “inertia”, i.e. autocorrelation (unless $\mu_i = 0$).

RE estimator - FGLS

$y_{it} = \beta_0 + \beta_1 x_{it1} + \cdots + \beta_k x_{itk} + v_{it};$
The quasi-demeaning (quasi-differencing) parameter λ is used for the FGLS estimation:
$\theta = 1 - \left[\sigma_u^2 / (\sigma_u^2 + T\sigma_\mu^2)\right]^{1/2}, \quad 0 \leq \theta \leq 1$
where $var(\mu_i) = \sigma_\mu^2; \quad var(u_i) = \sigma_u^2$

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

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

$$\boxed{[y_{it} - \theta\overline{y}_i]} = \beta_1 \boxed{[x_{it1} - \theta\overline{x}_{i1}]} + \cdots + \beta_k \boxed{[x_{itk} - \theta\overline{x}_{ik}]} + \boxed{[\mu_i - \theta\overline{a}_i + u_{it} - \theta\overline{u}_i]}$$

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

RE estimator - FGLS

$$\boxed{[y_{it} - \theta \overline{y}_i]} = \beta_1 \boxed{[x_{it1} - \theta \overline{x}_{i1}]} + \cdots + \beta_k \boxed{[x_{itk} - \theta \overline{x}_{ik}]} + \boxed{[\mu_i - \theta \overline{a}_i + u_{it} - \theta \overline{u}_i]}$$

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

$$\begin{aligned} \hat{\theta} \to 1 \quad &\to \quad \text{RE} \to \text{ FE} \\ \hat{\theta} \to 0 \quad &\to \quad \text{RE} \to \text{ Pooled} \end{aligned}$$

RE estimator – Assumptions

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

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

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

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

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

FE.7 [normality of u_{it} has little actual importance for the 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 μ_i given all regressors is constant: $E(\mu_i \mid \boldsymbol{X}_i) = \beta_0$. [Rules out correlation between μ_i and $\boldsymbol{X}_i$]

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

CRE estimator

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

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

  Specifically, as μ_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 (compare to FE).
- CRE allows for convenient testing of FE vs. RE.

CRE estimator

CRE: The individual-specific effect μ_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 $y_{it} = \beta_1 x_{it} + \mu_i + u_{it}$, we assume (a single-regressor illustration):

$$\mu_i = \alpha + \gamma \overline{x}_i + r_i, \text{ now: } cor(r_i, \overline{x}_i) = 0 \Rightarrow cor(r_i, x_{it}) = 0$$

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

By substituting for μ_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>
As $\gamma \overline{x}_i$ controls for the correlation between μ_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 $y_{it} = \beta_1 x_{it} + \mu_i + u_{it}$:

with uncorrelated random effect r_i but with the time averages as additional regressors.

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

- CRE allows for incorporation of time-invariant regressors: Besides $\hat{\beta}_{CRE} = \hat{\beta}_{FE}$, we can include arbitrary time invariant regressors and estimate γ_{CRE} values.
- 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 $\gamma = 0$: if we reject H_0, we also reject RE in favor of FE]

Arellano-Bond estimator (dynamic panels)

Dynamic panel

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

…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 estimator (dynamic panels)

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 for the dynamic coefficients.
- AB approach is based on IV (we need instruments for the lagged dependent variable – this is an endogenous regressor, correlated with the errors in the FD model).
- 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!

Poolability tests

LSDV-based test for individual intercepts

- Null hypothesis of common intercept is tested against the alternative of individual-specific intercepts.
- Common slopes are assumed (not tested)
- Unrestricted model: $y_{it} = \beta_0 + \boldsymbol{d}'\boldsymbol{\delta}_0 + \beta_1 x_{it1} + \beta_2 x_{it2} + u_{it}$ where $\boldsymbol{d}$ is a vector of CSID-based dummy variables and $\boldsymbol{\delta}_0$ is a vector of regression coefficients ($N-1$ dummies used to avoid dummy variable trap).
- Restricted model: $y_{it} = \beta_0 \quad + \beta_1 x_{it1} + \beta_2 x_{it2} + u_{it}.$
- Can be implemented as an F-test for linear (zero) restrictions: Pooled regression vs LSDV model

Chow test for identical slopes

- `pooltest()` from the `{plm}` package
- We allow for different intercepts & test for equal slopes in all CS-units
  - Estimate model separately for each CS unit.
  - Compare with "FE" model (individual intercept, common slopes on regressors) using an F-test – are the slopes identical among CS-units?
- Drawback: test cannot handle time-invariant regressors (FE; also, as the unrestricted model is estimated individually for each CS-unit, such regressors are perfectly correlated with the intercept and μ_i elements)
- Unrestricted model: $y_{it} = \beta_0 + \beta_{i1} x_{it} + \mu_i + u_{it}$
- Restricted model: $y_{it} = \beta_0 + \beta_1 x_{it} + \mu_i + u_{it}$

$$H_0: \ \beta_{11} = \beta_{21} = \cdots = \beta_{N1}$$
$$H_1: \ \neg H_0$$

Chow test for identical slopes

SSR_r: restricted model – allow for different μ_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 no structural break, $F \sim F[(N-1)k, (NT - N - Nk)]$

- Alternatively, the restricted model can be amended to feature a single intercept (no μ_i individual effects).

Honda (1985) test for individual and time effects

- `plmtest(..., type="honda")` from the `{plm}` package
- Using OLS-based ("pooling") residuals, we test the null hypothesis of redundant individual (μ_i) and/or time (λ_t) effects.
- Individual effects:

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \cdots + \beta_k x_{itk} + \mu_i + \nu_{it}$$

- Time effects:

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \cdots + \beta_k x_{itk} + \lambda_t + \nu_{it}$$

- Twoways effects:

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \cdots + \beta_k x_{itk} + \mu_i + \lambda_t + \nu_{it}$$

- Note: for this LM-based tests, we only use the residuals of the pooling model (if performed on RE of FE model, corresponding pooling model is calculated internally first).

  Notation follows Baltagi (2008)

Honda (1985) test for individual and time effects

Panel model

- $y_{it} = \alpha + \boldsymbol{x}'_{it}\boldsymbol{\beta} + u_{it}$ where $u_{it} = \mu_i + \lambda_t + \nu_{it}$
- Assumptions for Honda (1985) test:
  i.i.d. individual effects: $\mu_i \sim N(0, \sigma^2_\mu)$;
  i.i.d. time effects: $\lambda_t \sim N(0, \sigma^2_\lambda)$;
  i.i.d. idiosyncratic errors: $\nu_{it} \sim N(0, \sigma^2_\nu)$.
- Null hypotheses to be tested:
  - $H_0^\mu : \sigma^2_\mu = 0$ (no individual effects)
  - $H_0^\lambda : \sigma^2_\lambda = 0$ (no time effects)
  - $H_0^{\mu\lambda} : \sigma^2_\mu = \sigma^2_\lambda = 0$ (no individual nor time effects)

Honda (1985) test for individual and time effects

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

Balanced panel assumed.

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

- In matrix form, $\boldsymbol{u}$ can be cast as:
  $\boldsymbol{u} = \boldsymbol{D}_\mu \boldsymbol{\mu} + \boldsymbol{D}_\lambda \boldsymbol{\lambda} + \boldsymbol{\nu}$
  where
  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_N)'$,
  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_T)'$,
  $\boldsymbol{\nu}$ follows the structure of $\boldsymbol{u}$,
  $\boldsymbol{D}_\mu = (\boldsymbol{I}_N \otimes \boldsymbol{\iota}_T)$ i.e. $\boldsymbol{I}_N$ with each row repated 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

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

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

- $\boldsymbol{D}_\mu \boldsymbol{D}'_\mu = (\boldsymbol{I}_N \otimes \boldsymbol{J}_T)$ 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 $r = \mu$ or $r = \lambda$.

Honda (1985) test for individual and time effects

$$y_{it} = \alpha + \boldsymbol{x}'_{it}\boldsymbol{\beta} + u_{it} \qquad \text{where } u_{it} = \mu_i + \lambda_t + \nu_{it} \qquad \text{(balanced panel)}$$

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

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

- 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

$$y_{it} = \alpha + \boldsymbol{x}'_{it}\boldsymbol{\beta} + u_{it} \qquad \text{where } u_{it} = \mu_i + \lambda_t + \nu_{it} \qquad \text{(balanced panel)}$$

- Honda (1985) provides a test statistic for $H_0^{\mu\lambda} : \sigma_\mu^2 = \sigma_\lambda^2 = 0$ against a one-sided alternative
(not derived as a uniformly most powerful LM statistics):

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

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

F-test for unobserved effects (FE-based) vs pooling model

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

- `pFtest()` from the `{plm}` package
- F-test of effects based on the comparison of "pooling" and "within" models (either "individual", "time" or "twoways" effects can be tested).
- Hence, two main arguments to the test function are `plm`-estimated "pooling" and "within" models.
- d.f. of the F-test depend on the number of observations and parameters restricted:
  `df1` is the number of parameters restricted,
  `df2` $= N(T-1) -$ (# parameters est. in the unrestricted model)
  ...remember that for each C-S observation i, we loose one d.f. as the demeaned errors $\ddot{\nu}_{it}$ add up to zero when summed over time.

Estimator selection & serial correlation tests

Hausman test: RE vs FE estimator

- `phtest()` from the `{plm}` package
- Hausman test is based on the comparison of two sets of estimates
- A classical application of the Hausman test for panel data is to compare the fixed and the random effects models:

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

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

H_0: $\text{cov}(\boldsymbol{x}_{it}, \mu_i) = 0$ ...i.e. the crucial RE assumption holds

H_1: RE assumptions violated.

Hausman test: RE vs FE estimator

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

- 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, and RE is efficient.
- For asymptotic variance estimators ($\widehat{Avar}$), 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].

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

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

- `pwfdtest()` from the `{plm}` package
- 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:
  - FE is more efficient than FD.
  - Residuals in the FD model: $e_{it} = \nu_{it} - \nu_{i,t-1}$ are correlated, with $\operatorname{cor}(e_{it}, e_{i,t-1}) = -0.5$.
- Test (for models with individual effects) 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}.
- If this H_0 is not rejected, we would prefer FE.
- Test performs well for T asymptotics. For short panels, other serial correlation tests are available.

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

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

- If ν_{it} follow a random walk:
  - FD is more efficient than FE.
  - Residuals in the FE model: $\nu_{it} = \nu_{i,t-1} + e_{it}$.
  - Residuals in the FD model: $e_{it} = \nu_{it} - \nu_{i,t-1}$ are not serially correlated,
    (definition of a random walk for ν_{it}).
- `pwfdtest(..., h0="fd")`
  H_0 : no serial correlation in FD-errors e_{it},
  if not rejected, use FD.
- `pwfdtest(..., h0="fe")`
  H_0 : no serial correlation in FE-errors ν_{it},
  if not rejected, use FE.
- If both rejected, whichever estimator is chosen will have serially correlated errors: use the autocorrelation-robust covariance estimators.

Serial correlation tests

- `pwtest()` Unobserved effects: "Wooldridge"-type test
  $H_0 : \sigma_\mu^2 = 0$ for RE model. Under H_0, average value of the scaled elements of error covariance matrix (upper/lower triangle, excluding diagonal) asymptotically follow $N(0, 1)$.

  Non-rejection of H_0 favours pooled OLS, yet rejecting may be due to serial correlation (in random effects). H_0 rejection does not imply existence of individual effects (may be due to serial corr.).

- `pbsytest()` Bera, Sosa-Escudero, Yoon (2001)
  Solution to the previous problem: three tests:
  H_0 : no serial correlation while controlling for random effects
  H_0 : no random effects (while controlling for possible ser. corr.)
  H_0 : no random effects & no serial correlation.

- For detailed description of both tests, see:
  Wooldridge, 2002 (CS and panel data analysis)
  `https://www.jstatsoft.org/article/view/v027i02`

General serial correlation tests

- `pbgtest()` Breusch-Godfrey test for panels,
  Mainly for RE (and pooling) models.
- Under RE assumptions of homoskedasticity and no serial correlation in the idiosyncratic error, the 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. (see next page).
- `pdwtest()` Durbin-Watson test for panels (…analogous).

General serial correlation tests

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

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

$$\mathrm{cor}\,(e_{it}, e_{is}) = -1/(T-1).$$

- H_0 of no serial correlation in ν_{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 ν_{it}.
- Test applicable to any “FE model”, particularly with $N \gg T$.
- As T grows, $-1/(T-1) \to 0$ and `pbgtest()` can be used as well.

Robust statistical inference

Robust statistical inference

- `vcovHC()` from the `{plm}` package,
  used together with functions from `{lmtest}`
- three types of HC/HAC covariance matrix estimators (sandwich estimator)
- Based on White's general form (for CS data):

  $$\mathrm{var}(\hat{\boldsymbol{\beta}|\boldsymbol{X}}) = \left[\boldsymbol{X}'\boldsymbol{X}\right]^{-1}\left[\boldsymbol{X}'\sigma^2\boldsymbol{\Omega}\boldsymbol{X}\right]\left[\boldsymbol{X}'\boldsymbol{X}\right]^{-1}$$

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

Robust statistical inference

- `vcovHC(... , method="white1")`
- `"white1"` allows for general heteroskedasticity but no serial correlation, i.e.,

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

  and $\boldsymbol{\Sigma}$ is a block-diagonal matrix of $\boldsymbol{\Sigma}_i$ matrices
- `"white2"` is `"white1"` with common CS-variance: $\boldsymbol{\Sigma}_i = \sigma_i^2 \boldsymbol{I}_T$.
- The counterpart to CS-related $[\boldsymbol{X}'\boldsymbol{\Sigma}\boldsymbol{X}]$ 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 (time-demeaned) regressors.

Robust statistical inference

- `vcovHC(...  , method="arellano")`
- `"arellano"` allows a fully general structure w.r.t. heteroskedasticity and serial correlation:

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

  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).
- `"white1"` is inconsistent for fixed T as N grows
  $\rightarrow$ use `"arellano"` in such case

Cross-sectional dependence (XSD)

Cross-sectional dependence (XSD)

- `pcdtest()` from the `{plm}` package,
- Analogous (yet distinct) to the more familiar issue of serial correlation.
- Can arise, e.g., if individuals respond to common shocks or if spatial diffusion processes are present, relating individuals in a way depending on a measure of distance (spatial models)
- 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 misspeciffication tests to detect XSD are available – no robust method to perform valid inference in its presence.

Cross-sectional dependence (XSD)

- Test(s) based on on (transformations of) the product-moment correlation coefficient of a model's 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., as 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 both in N and T-asymptotic settings. Good performance in samples of any practically relevant size and is robust to a variety of settings.