Panel Data Models

Ercan Karadas

New York University
Department of Economics

Spring, 2018

Content

- **Data structures: Time Series, Cross Sectional, Panel Data, Pooled Data**
- **Static linear panel data models: Fixed Effects, Random Effects, Estimation, Testing**
- **Dynamic panel data models: Estimation**

Resources:

- Verbeek, Chapter 11
- Greene, Chapter 11.1-5

Data structures

Data structures

We distinguish the following data structures

- **Time series data**:
  - $\{x_t, t = 1, \ldots, T\}$, univariate series, e.g. a price series:
    Its path over time is modeled. The path may also depend on third variables.
  - Multivariate, e.g. several price series:
    Their individual as well as their common dynamics is modeled. Third variables may be included.
- **Cross sectional data** are observed at a single point of time for several individuals, countries, assets, etc.,
  x_i, $i = 1, \ldots, N$.
  The interest lies in modeling the distinction of single individuals, the *heterogeneity across individuals*.

Data structures: Pooling

Pooling data refers to two or more *independent* data sets of the same type.

- **Pooled time series**:
  We observe e.g. return series of several sectors, which are assumed to be independent of each other, together with explanatory variables. The number of sectors, N, is usually small.
  Observations are viewed as repeated measures at each point of time. So parameters can be estimated with higher precision due to an increased sample size.

Data structures: Pooling

- **Pooled cross sections**:
  Mostly these type of data arise in surveys, where people are asked about e.g. their attitudes to political parties. This survey is repeated, T times, before elections every week. T is usually small.
  So we have several cross sections, but the persons asked are chosen randomly. Hardly any person of one cross section is member of another one. The cross sections are independent.
  Only overall questions can be answered, like the attitudes within males or women, but no individual (even anonymous) paths can be identified.

Data structures: Panel data

A **panel data** set (also **longitudinal data**) has both a cross-sectional and a time series dimension, where all cross section units are observed during the whole time period.
x_{it}, $i = 1, \ldots, N$, $t = 1, \ldots, T$. T is usually small.

We can distinguish between *balanced* and *unbalanced* panels.

Example for a balanced panel:
The Mikrozensus in Austria is a household, hh, survey, with the same size of 22.500 each quarter. Each hh has to record its consumption expenditures for 5 quarters. So each quarter 4500 members enter/leave the Mikrozensus. This is a *balanced panel*.

Data structures: Panel data

A special case of a balanced panel is a *fixed panel*. Here we require that all individuals are present in all periods.

An *unbalanced panel* is one where individuals are observed a different number of times, e.g. because of missing values.

We are concerned only with balanced/fixed panels.

In general panel data models are more 'efficient' than pooling cross-sections, since the observation of one individual for several periods reduces the variance compared to repeated random selections of individuals.

Pooling time series: estimation

We consider T relatively large, N small.

$$y_{it} = \alpha + \beta\, x_{it} + u_{it}$$

In case of heteroscedastic errors, $\sigma_i^2 \neq \sigma^2 (= \sigma_u^2)$, individuals with large errors will dominate the fit. A correction is necessary. It is similar to a GLS and can be performed in 2 steps.

First estimate under assumption of const variance for each indiv i and calculate the individual residual variances, s_i^2.

$$s_i^2 = \frac{1}{T-2} \sum_t (y_{it} - a - b\, x_{it})^2$$

Pooling time series: estimation

Secondly, normalize the data with s_i and estimate

$$(y_{it}/s_i) = \alpha\,(1/s_i) + \beta\,(x_{it}/s_i) + \tilde{u}_{it}$$

$\tilde{u}_{it} = u_{it}/s_i$ has (possibly) the required constant variance, is homoscedastic.

Remark: $\mathrm{V}(\tilde{u}_{it}) = \mathrm{V}(u_{it}/s_i) \approx 1$

Dummies may be used for different cross sectional intercepts.

Panel data modeling

Example

Say, we observe the weekly returns of 1000 stocks in two consecutive weeks.

The pooling model is appropriate, if the stocks are chosen randomly in each period. The panel model applies, if the same stocks are observed in both periods.

We could ask the question, what are the characteristics of stocks with high/low returns in general.

For panel models we could further analyze, whether a stock with high/low return in the first period also has a high/low return in the second.

Panel data model

The standard static model with $i = 1, \ldots, N$, $t = 1, \ldots, T$ is

$$y_{it} = \beta_0 + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \epsilon_{it}$$

$\boldsymbol{x}_{it}$ is a K-dimensional vector of explanatory variables, without a const term.
β_0, the intercept, is independent of i and t.
$\boldsymbol{\beta}$, a $(K \times 1)$ vector, the slopes, is independent of i and t.
ϵ_{it}, the error, varies over i and t.

Individual characteristics (which do not vary over time), $\boldsymbol{z}_i$, may be included

$$y_{it} = \beta_0 + \boldsymbol{x}_{it}'\boldsymbol{\beta}_1 + \boldsymbol{z}_i'\boldsymbol{\beta}_2 + \epsilon_{it}$$

Two problems: endogeneity and autocorr in the errors

- *Consistency/exogeneity:*
  Assuming iid errors and applying OLS we get consistent estimates, if $\mathrm{E}(\epsilon_{it}) = 0$ and $\mathrm{E}(\boldsymbol{x}_{it}\epsilon_{it}) = \mathbf{0}$, if the $\boldsymbol{x}_{it}$ are weakly exogenous.
- *Autocorrelation in the errors:*
  Since individual i is repeatedly observed (contrary to pooled data)

  $$\mathrm{Corr}(\epsilon_{i,s}, \epsilon_{i,t}) \neq 0$$

  with $s \neq t$ is very likely. Then,
  - standard errors are misleading (similar to autocorr residuals),
  - OLS is inefficient (cp. GLS).

Common solution for individual unobserved heterogeneity

Unobserved (const) individual factors, i.e. if not all z_i variables are available, may be captured by α_i. E.g. we decompose ϵ_{it} in

$$\epsilon_{it} = \alpha_i + u_{it} \quad \text{with} \quad u_{it} \;\; iid(0, \sigma_u^2)$$

u_{it} has mean 0, is homoscedastic and not serially correlated.

In this decomposition *all individual characteristics* - including all observed, $\boldsymbol{z}_i'\beta_2$, as well as all unobserved ones, which do not vary over time - are *summarized in the* α_i's.

We distinguish *fixed effects* (FE), and *random effects* (RE) models.

Fixed effects model, FE

- **Fixed effects model**, FE:
  α_i are *individual intercepts* (fixed for given N).

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

  No overall intercept is (usually) included in the model.

  Under FE, consistency does not require, that the individual intercepts (whose coefficients are the α_i's) and u_{it} are uncorrelated. Only $\mathrm{E}(\boldsymbol{x}_{it} u_{it}) = \mathbf{0}$ must hold.

  There are $N - 1$ additional parameters for capturing the individual heteroscedasticity.

Random effects model, RE

- **Random effects model**, RE:

$$\alpha_i \sim iid(0, \sigma_\alpha^2)$$
$$y_{it} = \beta_0 + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i + u_{it}, \quad u_{it} \sim iid(0, \sigma_u^2)$$

The α_i's are *rvs* with the same variance. The value α_i is specific for individual i. The α's of different indivs are independent, have a mean of zero, and their distribution is assumed to be not too far away from normality. The overall mean is captured in β_0.
α_i is time invariant and homoscedastic across individuals.

There is only one additional parameter σ_α^2.

Only α_i contributes to $\mathrm{Corr}(\epsilon_{i,s}, \epsilon_{i,t})$. α_i determines both $\epsilon_{i,s}$ and $\epsilon_{i,t}$.

RE some discussion

- *Consistency*:
  As long as $\mathrm{E}[\boldsymbol{x}_{it}\epsilon_{it}] = \mathrm{E}[\boldsymbol{x}_{it}(\alpha_i + u_{it})] = 0$, i.e. $\boldsymbol{x}_{it}$ are uncorrelated with α_i *and* u_{it}, the explanatory vars are exogenous, the estimates are consistent.

  There are relevant cases where this exogeneity assumption is likely to be *violated*:
  E.g. when modeling investment decisions the firm specific heteroscedasticity α_i might correlate with (the explanatory variable of) the cost of capital of firm *i*.
  The resulting *inconsistency* can be avoided by considering a FE model instead.
- *Estimation*:
  The model can be estimated by (feasible) GLS which is in general more 'efficient' than OLS.

The static linear model

– The fixed effects model

3 Estimators:
– Least square dummy variable estimator, LSDV
– Within estimator, FE
– First difference estimator, FD

[LSDV] Fixed effects model: LSDV estimator

We can write the FE model using N dummy vars indicating the individuals.

$$y_{it} = \sum_{j=1}^{N} \alpha_j d_{it}^j + \boldsymbol{x}_{it}'\boldsymbol{\beta} + u_{it} \qquad u_{it} \sim iid(0, \sigma_u^2)$$

with dummies d^j, where $d_{it}^j = 1$ if $i = j$, and 0 else.

The parameters can be estimated by OLS. The implied estimator for $\boldsymbol{\beta}$ is called the **LS dummy variable estimator**, LSDV.
Instead of exploding computer storage by increasing the number of dummy variables for large N the *within estimator* is used.

[LSDV] Testing the significance of the group effects

Apart from *t*-tests for single α_i (which are hardly used) we can test, whether the indivs have 'the same intercepts' wrt 'some have different intercepts' by an *F*-test.

The pooled model (all intercepts are *restricted* to be the same), H_0, is

$$y_{it} = \beta_0 + \boldsymbol{x}'_{it}\boldsymbol{\beta} + u_{it}$$

the fixed effects model (intercepts may be different, are *unrestricted*), H_A,

$$y_{it} = \alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + u_{it} \quad i = 1, \ldots, N$$

The *F* ratio for comparing the pooled with the FE model is

$$F_{N-1,NT-N-K} = \frac{(R^2_{LSDV} - R^2_{Pooled})/(N-1)}{(1 - R^2_{LSDV})/(NT - N - K)}$$

[FE] Within transformation, within estimator

The FE estimator for β is obtained, if we use the deviations from the individual means as variables. The model in individual means is with $\bar{y}_i = \sum_t y_{it}/T$ and $\bar{\alpha}_i = \alpha_i$, $\bar{u}_i = 0$

$$\bar{y}_i = \alpha_i + \bar{\boldsymbol{x}}_i'\boldsymbol{\beta} + \bar{u}_i$$

Subtraction from $y_{it} = \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + u_{it}$ gives

$$y_{it} - \bar{y}_i = (\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)'\boldsymbol{\beta} + (u_{it} - \bar{u}_i)$$

where the intercepts vanish. Here the deviation of y_{it} from $\bar{y}_i$ is explained (not the difference between different individuals, $\bar{y}_i$ and $\bar{y}_j$).

The estimator for β is called the **within** or **FE estimator**.

Within refers to the variability (over time) among observations of individual i.

[FE] Within/FE estimator $\hat{\beta}_{FE}$

The within/FE estimator is

$$\hat{\beta}_{FE} = \left(\sum_i \sum_t (\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)' \right)^{-1} \sum_i \sum_t (\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)(y_{it} - \bar{y}_i)$$

This expression is identical to the well known formula $\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ for N (demeaned wrt individual i) data with T *repeated* observations.

[FE] Finite sample properties of $\hat{\beta}_{FE}$

Finite samples:

- *Unbiasedness:* if all $\boldsymbol{x}_{it}$ are independent of all u_{js}, strictly exogenous.
- *Normality:* if in addition u_{it} is normal.

[FE] Asymptotic samples properties of $\hat{\beta}_{FE}$

Asymptotically:

- Consistency wrt $N \to \infty$, T fix: if $\mathrm{E}[(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)u_{it}] = \mathbf{0}$.
  I.e. *both* $\boldsymbol{x}_{it}$ *and* $\bar{\boldsymbol{x}}_i$ are uncorrelated with the error, u_{it}.
  This (as $\bar{\boldsymbol{x}}_i = \sum_t x_{i,t}/T$) implies that $\boldsymbol{x}_{it}$ is *strictly exogenous*:

  $$\mathrm{E}[\boldsymbol{x}_{is}u_{it}] = \mathbf{0} \quad \text{for all } s, t$$

  $\boldsymbol{x}_{it}$ has not to depend on current, past or future values of the error term of individual i. This excludes
  - lagged dependent vars
  - any x_{it}, which depends on the history of y

  Even large N do not mitigate possible violations.
- Asymptotic normality: under consistency and weak conditions on u.

[FE] Properties of the N intercepts, $\hat{\alpha}_i$, and $\mathrm{V}(\hat{\beta}_{FE})$

- Consistency of $\hat{\alpha}_i$ wrt $T \to \infty$: if $\mathrm{E}[(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i) u_{it}] = \mathbf{0}$.
  There is no convergence wrt to N, even if N gets large. Cp. $\bar{y}_i = (1/T) \sum_t y_{it}$.

The estimates for the N intercepts, $\hat{\alpha}_i$, are simply

$$\hat{\alpha}_i = \bar{y}_i - \bar{\boldsymbol{x}}_i' \hat{\beta}_{FE}$$

Reliable estimates for $\mathrm{V}(\hat{\beta}_{FE})$ are obtained from the LSDV model. A consistent estimate for σ_u^2 is

$$\hat{\sigma}_u^2 = \frac{1}{NT - N - K} \sum_i \sum_t \hat{u}_{it}^2$$

with $\hat{u}_{it} = y_{it} - \hat{\alpha}_i - \boldsymbol{x}_{it}' \hat{\beta}_{FE}$

[FD] The first difference, FD, estimator for the FE model

An alternative way to eliminate the individual effects α_i is to take first differences (wrt time) of the FE model.

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

or

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

Here also, all variables, which are only indiv specific - they do not change with t - drop out.

Estimation with OLS gives the **first-difference estimator**

$$\hat{\boldsymbol{\beta}}_{FD} = \left(\sum_i \sum_{t=2}^{T} \Delta \boldsymbol{x}_{it} \Delta \boldsymbol{x}_{it}' \right)^{-1} \sum_i \sum_{t=2}^{T} \Delta \boldsymbol{x}_{it} \Delta y_{it}$$

[FD] Properties of the FD estimator

- Consistency for $N \to \infty$: if $\mathrm{E}[\Delta \boldsymbol{x}_{it} \Delta u_{it}] = 0$.
  This is slightly less demanding than for the within estimator, which is based on strict exogeneity.
  FD allows e.g. correlation between $\boldsymbol{x}_{it}$ and $u_{i,t-2}$.
- The FD estimator is slightly less efficient than the FE, as
  Δu_{it} exhibits serial correlation, even if u_{it}'s are uncorrelated.
- FD looses one time dimension for each i.
  FE looses one degree of freedom for each i by using $\bar{y}_i, \bar{\boldsymbol{x}}_i$.

The static linear model
– estimation of the random effects model

Estimation of the random effects model

$$y_{it} = \beta_0 + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \alpha_i + u_{it}, \quad u_{it} \sim iid(0, \sigma_u^2), \quad \alpha_i \sim iid(0, \sigma_\alpha^2)$$

where $(\alpha_i + u_{it})$ is an error consisting of 2 components:

- an individual specific component, which does not vary over time, α_i.
- a remainder, which is uncorrelated wrt i and t, u_{it} .
- α_i and u_{it} are mutually independent, and indep of all $\boldsymbol{x}_{js}$.

As simple OLS does not take this specific error structure into account, so GLS is used.

[RE] GLS estimation

In the following we stack the errors of individual i, α_i and u_{it}. Ie, we put each of them into a column vector. $(\alpha_i + u_{it})$ reads then as

$$\alpha_i \, \boldsymbol{\iota}_T + \boldsymbol{u}_i$$

α_i is constant for individual i
$\boldsymbol{\iota}_T = (1, \ldots, 1)'$ is a $(T \times 1)$ vector of only ones.
$\boldsymbol{u}_i = (u_{i1}, \ldots, u_{iT})'$ is a vector collecting all u_{it}'s for indiv i.
$\boldsymbol{I}_T$ is the T-dimensional identity matrix.

Since the errors of different indiv's are independent, the covar-matrix consists of N *identical blocks in the diagonal*. Block i is

$$\mathrm{V}(\alpha_i \, \boldsymbol{\iota}_T + \boldsymbol{u}_i) = \boldsymbol{\Omega} = \sigma_\alpha^2 \, \boldsymbol{\iota}_T \boldsymbol{\iota}_T' + \sigma_u^2 \, \boldsymbol{I}_T$$

Remark: The inverse of a block-diag matrix is a block-diag matrix with the inverse blocks in the diag. So it is enough to consider indiv blocks.

[RE] GLS estimation

GLS corresponds to premultiplying the vectors $(y_{i1}, \ldots, y_{iT})'$, etc. by $\boldsymbol{\Omega}^{-1/2}$. Where

$$\boldsymbol{\Omega}^{-1} = \sigma_u^{-2} \left[\boldsymbol{I}_T - \tilde{\psi}\, \boldsymbol{\iota}_T \boldsymbol{\iota}_T' \right] \quad \text{with} \quad \tilde{\psi} = \frac{\sigma_\alpha^2}{\sigma_u^2 + T\sigma_\alpha^2}$$

and

$$\psi = \frac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}$$

Remark: The GLS for the classical regression model is

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

where $\boldsymbol{u} \sim .(\boldsymbol{0}, \boldsymbol{\Omega})$ is heteroscedastic.

[RE] GLS estimator

The GLS estimator can be written as

$$\hat{\beta}_{GLS} = \left(\sum_i \sum_t (\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)' + \psi \sum_i T(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}})' \right)^{-1}$$

$$\times \left(\sum_i \sum_t (\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)(\boldsymbol{y}_{it} - \bar{\boldsymbol{y}}_i)' + \psi \sum_i T(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}})(\bar{\boldsymbol{y}}_i - \bar{\boldsymbol{y}})' \right)$$

- If $\psi = 0$, ($\sigma_u^2 = 0$), the GLS estimator is equal to the FE estimator. As if N individual intercepts were included.
- If $\psi = 1$, ($\sigma_\alpha^2 = 0$). The GLS becomes the OLS estimator. Only 1 overall intercept is included, as in the pooled model.
- If $T \to \infty$ then $\psi \to 0$. I.e.: The FE and RE estimators for β are equivalent for large T. (But *not* for T fix and $N \to \infty$.)
  Remark: $(\sum_i \sum_t + \sum_i T)/(N\,T)$ is finite as N or T or both $\to \infty$.

Interpretation of within and between, $\psi = 1$

Within refers to the variation within one individual (over time), *between* measures the variation between the individuals.

The within and between components

$$(y_{it} - \bar{y}) = (y_{it} - \bar{y}_i) + (\bar{y}_i - \bar{y})$$

are orthogonal. So

$$\sum_i \sum_t (y_{it} - \bar{y})^2 = \sum_i \sum_t (y_{it} - \bar{y}_i)^2 + T \sum_i (\bar{y}_i - \bar{y})^2$$

The variance of y may be decomposed into the sum of variance within and the variance between.

$$\bar{y}_i = (1/T) \sum_t y_{it} \quad \text{and} \quad \bar{y} = (1/(N\,T) \sum_i \sum_t y_{it}$$

[RE] Properties of the GLS

- GLS is unbiased, if the x's are independent of all u_{it} and α_i.
- The GLS will be more efficient than OLS in general under the RE assumptions.
- Consistency for $N \to \infty$ (T fix, or $T \to \infty$):
  if $\mathrm{E}[(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)u_{it}] = 0 \quad \text{and} \quad \mathrm{E}[\bar{\boldsymbol{x}}_i \alpha_i] = 0$
  holds.
- Under weak conditions (errors need not be normal) the feasible GLS is asymptotically normal.

Comparison and testing of FE and RE

Interpretation: FE and RE

- Fixed effects: The distribution of y_{it} is seen conditional on $\boldsymbol{x}_{it}$ and individual dummies d_i.

  $$\mathrm{E}[y_{it}|\boldsymbol{x}_{it}, d_i] = \boldsymbol{x}_{it}\boldsymbol{\beta} + \alpha_i$$

  This is plausible if the individuals are *not a random draw*, like in samples of large companies, countries, sectors.
- Random effects: The distribution of y_{it} is not conditional on single individual characteristics. Arbitrary indiv effects have a fixed variance. Conclusions *wrt a population* are drawn.

  $$\mathrm{E}[y_{it}|\boldsymbol{x}_{it}, 1] = \boldsymbol{x}_{it}\boldsymbol{\beta} + \beta_0$$

If the d_i are expected to be correlated with the x's, FE is *preferred*. RE increases only efficiency, if exogeneity is guaranteed.

Interpretation: FE and RE (T fix)

The FE estimator, $\hat{\beta}_{FE}$, is

- consistent ($N \to \infty$) and efficient under the FE model assumptions.
- consistent, but not efficient, under the RE model assumptions, as the correlation structure of the errors is not taken into account (replaced only by different intercepts). As the error variance is not estimated consistently, the t-values are not correct.

The RE estimator is

- consistent ($N \to \infty$) and efficient under the assumptions of the RE model.
- not consistent under the FE assumptions, as the true explanatory vars $(d_i, \boldsymbol{x}_{it})$ are correlated with $(\alpha_i + u_{it})$ [with $\hat{\psi} \neq 0$ and $\bar{\boldsymbol{x}}_i \neq \bar{\boldsymbol{x}}$ biased for fix T].

Hausman test

Hausman tests the H_0 that $\boldsymbol{x}_{it}$ and α_i are uncorrelated. We compare therefore two estimators

- one, that is consistent under both hypotheses, and
- one, that is consistent (and efficient) only under the null.

A significant difference between both indicates that H_0 is unlikely to hold.

H_0 is the RE model

$$y_{it} = \beta_0 + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \alpha_i + u_{it}$$

H_A is the FE model

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

- $\hat{\beta}_{RE}$ is consistent (and efficient), under H_0, not under H_A.
- $\hat{\beta}_{FE}$ is consistent, under H_0 and H_A.

Test FE against RE: Hausman test

We consider the difference of both estimators. If it is large, we reject the null.

$$\hat{\beta}_{FE} - \hat{\beta}_{RE}$$

Since $\hat{\beta}_{RE}$ is efficient under H_0, it holds

$$\mathrm{V}(\hat{\beta}_{FE} - \hat{\beta}_{RE}) = \mathrm{V}(\hat{\beta}_{FE}) - \mathrm{V}(\hat{\beta}_{RE})$$

The Hausman statistic under H_0 is

$$(\hat{\beta}_{FE} - \hat{\beta}_{RE})'[\hat{\mathrm{V}}(\hat{\beta}_{FE}) - \hat{\mathrm{V}}(\hat{\beta}_{RE})]^{-1}(\hat{\beta}_{FE} - \hat{\beta}_{RE}) \overset{asy}{\sim} \chi^2(K)$$

If in finite samples $[\hat{\mathrm{V}}(\hat{\beta}_{FE}) - \hat{\mathrm{V}}(\hat{\beta}_{RE})]$ is not positive definite, testing is performed only for a subset of $\boldsymbol{\beta}$.

Comparison via goodness-of-fit, R^2 wrt $\hat{\beta}$

We are often interested in a distinction between the *within* R^2 and the *between* R^2. The *within* R^2 for any arbitrary estimator is given by

$$R^2_{within}(\hat{\beta}) = \mathrm{Corr}^2[\hat{y}_{it} - \hat{y}_i,\ y_{it} - \bar{y}_i]$$

and *between* R^2 by

$$R^2_{between}(\hat{\beta}) = \mathrm{Corr}^2[\hat{y}_i,\ \bar{y}_i]$$

where $\quad \hat{y}_{it} = \boldsymbol{x}_{it}'\hat{\beta}, \quad \hat{y}_i = \bar{\boldsymbol{x}}_i'\hat{\beta} \quad$ and $\quad \hat{y}_{it} - \hat{y}_i = (\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)'\hat{\beta}$.

The *overall* R^2 is

$$R^2_{overall}(\hat{\beta}) = \mathrm{Corr}^2[\hat{y}_{it},\ y_{it}]$$

Goodness-of-fit, R^2

- The usual R^2 is valid for comparison of the pooled model estimated by OLS and the FE model.
- The comparison of the FE and RE via R^2 is not valid as in the FE the α_i's are considered as explanatory variables, while in the RE model they belong to the unexplained error.
- Comparison via R^2 should be done only within the same class of models and estimators.

Testing for autocorrelation

A generalization of the **Durbin-Watson statistic** for autocorr of order 1 can be formulated

$$DW_p = \frac{\sum_i \sum_{t=2}^{T}(\hat{u}_{it} - \hat{u}_{i,t-1})^2}{\sum_i \sum_t \hat{u}_{it}^2}$$

The critical values depend on T, N and K. For $T \in [6, 10]$ and $K \in [3, 9]$ the following approximate lower and upper bounds can be used.

$N = 100$		$N = 500$		$N = 1000$	
d_L	d_U	d_L	d_U	d_L	d_U
1.86	1.89	1.94	1.95	1.96	1.97

Testing for heteroscedasticity

A generalization of the **Breusch-Pagan test** is applicable.
σ_u^2 is tested whether it depends on a set of J third variables $\boldsymbol{z}$.

$$\mathrm{V}(u_{it}) = \sigma^2 h(\boldsymbol{z}_{it}'\gamma)$$

where for the function $h(.)$, $h(0) = 1$ and $h(.) > 0$ holds.

The null hypothesis is $\gamma = \mathbf{0}$.

The $N(T-1)$ multiple of R_u^2 of the auxiliary regression

$$\hat{u}_{it}^2 = \sigma^2 h(\boldsymbol{z}_{it}'\gamma) + v_{it}$$

is distributed under the H_0 asymptotically $\chi^2(J)$ with J degrees of freedom.

$$N(T-1)R_u^2 \overset{a}{\sim} \chi^2(J)$$

Dynamic panel data models

Dynamic panel data models

The dynamic model with one lagged dependent without exogenous variables, $|\gamma| < 1$, is

$$y_{it} = \gamma \, y_{i,t-1} + \alpha_i + u_{it}, \quad u_{it} \sim iid(0, \sigma_u^2)$$

Endogeneity problem

Here, $y_{i,t-1}$ *depends positively on* α_i:
This is simple to see, when inspecting the model for period $(t-1)$

$$y_{i,t-1} = \gamma\, y_{i,t-2} + \alpha_i + u_{i,t-1}$$

There is an *endogeneity problem*. OLS or GLS will be inconsistent for $N \to \infty$ and T fixed, both for FE and RE. (Easy to see for RE.)

The finite sample bias can be substantial for small T. E.g. if $\gamma = 0.5$, $T = 10$, and $N \to \infty$

$$\text{plim}_{N\to\infty}\, \hat{\gamma}_{FE} = 0.33$$

However, as in the univariate model, if in addition $T \to \infty$, we obtain a consistent estimator. But T is i.g. small for panel data.

The first difference estimator (FD)

Using the 1st difference estimator (FD), which eliminates the α_i's

$$\Delta y_{it} = \gamma \Delta y_{i,t-1} + \Delta u_{it}$$

$$y_{it} - y_{i,t-1} = \gamma(y_{i,t-1} - y_{i,t-2}) + (u_{it} - u_{i,t-1})$$

is no help, since $y_{i,t-1}$ and $u_{i,t-1}$ are correlated even when $T \to \infty$.

We stay with the FD model, as the exogeneity requirements are less restrictive, and use an IV estimator.

For that purpose we look also at $\Delta y_{i,t-1}$

$$y_{i,t-1} - y_{i,t-2} = \gamma(y_{i,t-2} - y_{i,t-3}) + (u_{i,t-1} - u_{i,t-2})$$

[FD] IV estimation, Anderson-Hsiao

Instrumental variable estimators, IV, have been proposed by Anderson-Hsiao, as they are consistent with $N \to \infty$ and *finite* T.

Choice of the instruments for $(y_{i,t-1} - y_{i,t-2})$:

- Instrument $y_{i,t-2}$ as proxy is correlated with $(y_{i,t-1} - y_{i,t-2})$, but not with $u_{i,t-1}$ or $u_{i,t}$, and so $\Delta u_{i,t}$.
- Instrument $(y_{i,t-2} - y_{i,t-3})$ as proxy for $(y_{i,t-1} - y_{i,t-2})$ sacrifies one more sample period.

[FD] GMM estimation, Arellano-Bond

The **Arellano-Bond** (also Arellano-Bover) method of moments **estimator** is consistent.
The moment conditions use the properties of the instruments

$$y_{i,t-j},\ j \geq 2$$

to be uncorrelated with the future errors u_{it} and $u_{i,t-1}$. We obtain an increasing number of moment conditions for $t = 3, 4, \ldots, T$.

$$t = 3: \quad \mathrm{E}[(u_{i,3} - u_{i,2})y_{i,1}] = 0$$

$$t = 4: \quad \mathrm{E}[(u_{i,4} - u_{i,3})y_{i,2}] = 0, \quad \mathrm{E}[(u_{i,4} - u_{i,3})y_{i,1}] = 0$$

$$t = 5: \quad \mathrm{E}[(u_{i,5} - u_{i,4})y_{i,3}] = 0, \quad \ldots, \quad \mathrm{E}[(u_{i,5} - u_{i,4})y_{i,1}] = 0$$

$$\ldots \qquad \ldots \qquad \ldots$$

[FD] GMM estimation, Arellano-Bond

We define the $(T-2)\times 1$ vector

$$\Delta \boldsymbol{u}_i = [(u_{i,3} - u_{i,2}), \ldots, (u_{i,T} - u_{i,T-1})]'$$

and a $(T-2)\times(T-2)$ matrix of instruments

$$\boldsymbol{Z}_i' = \begin{bmatrix} y_{i,1} & y_{i,1} & \ldots & y_{i,1} \\ 0 & y_{i,2} & \ldots & y_{i,2} \\ 0 & 0 & \ddots & \vdots \\ 0 & \ldots & 0 & y_{i,T-2} \end{bmatrix}$$

[FD] GMM estimation, Arellano-Bond, without x's

Ignoring exogenous variables, for $\Delta y_{it} = \gamma \Delta y_{i,t-1} + \Delta u_{it}$,

$$\mathrm{E}[\boldsymbol{Z}_i' \, \Delta \boldsymbol{u}_i] = \mathrm{E}[\boldsymbol{Z}_i' \, (\Delta \boldsymbol{y}_i - \gamma \Delta \boldsymbol{y}_{i,-1})] = \mathbf{0}$$

The number of moment conditions are $1 + 2 + \ldots + (T - 2)$.
This number exceeds i.g. the number of unknown coefficients, so γ is estimated by minimizing the quadratic expression

$$\min_{\gamma} \left[\frac{1}{N} \sum_i \boldsymbol{Z}_i' (\Delta \boldsymbol{y}_i - \gamma \Delta \boldsymbol{y}_{i,-1}) \right]' \boldsymbol{W}_N \left[\frac{1}{N} \sum_i \boldsymbol{Z}_i' (\Delta \boldsymbol{y}_i - \gamma \Delta \boldsymbol{y}_{i,-1}) \right]$$

with a weighting matrix $\boldsymbol{W}_N$.
The optimal matrix $\boldsymbol{W}_N$ yielding an asy efficient estimator is the *inverse of the covariance matrix of the sample moments*.

[FD] GMM estimation, Arellano-Bond, without x's

- The $\boldsymbol{W}_N$ matrix can be estimated directly from the data after a first consistent estimation step.
- Under weak regularity conditions the GMM estimator is asy normal for $N \to \infty$ for fixed T, $T > 2$ using our instruments.
- It is also consistent for $N \to \infty$ and $T \to \infty$, though the number of moment conditions $\to \infty$ as $T \to \infty$.
- It is advisable to limit the number of moment conditions.

[FD] GMM estimation, Arellano-Bond, with x's

The dynamic panel data model with exogenous variables is

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + \gamma\, y_{i,t-1} + \alpha_i + u_{it}, \quad u_{it} \ \sim \ iid(0, \sigma_u^2)$$

As also exogenous x's are included in the model *additional moment conditions* can be formulated:

- For strictly exogenous vars, $\mathrm{E}[x_{is}\, u_{it}] = 0$ for all s, t,

$$\mathrm{E}[x_{is}\, \Delta u_{it}] = 0$$

- For predetermined (not strictly exogenous) vars, $\mathrm{E}[x_{is} u_{it}] = 0$ for $s > t$

$$\mathrm{E}[x_{i,t-j}\, \Delta u_{it}] = 0 \quad j = 2, \ldots, t-1$$

GMM estimation, Arellano-Bond, with x's

So there are a lot of possible moment restrictions both for differences *as well as for levels*, and so a variety of GMM estimators.

GMM estimation may be combined with both FE and RE.
Here also, the RE estimator is identical to the FE estimator with $T \to \infty$.