Freie Universität Bozen
Libera Università di Bolzano
Università Liedia de Bulsan

Post Conference Workshop
Part 2: Large-T Dynamic Panel Data Analysis

Jan Ditzen

Free University of Bozen-Bolzano, Bozen, Italy
www.jan.ditzen.net, jan.ditzen@unibz.it

July 8, 2026

Panel-Time Series??? I

- Panel-Time Series models became very popular in the past years.
- Allow modelling of heterogeneity in time and space and can be applied to many datasets.
- Dependencies across space can be modelled while using time dimension at the same time.
- Growing body of literature on methods, for an overview see Smith and Fuertes (2012); Sarafidis and Wansbeek (2012); Baltagi (2015); Sul (2019); Elhorst et al. (2021); Ditzen and Karavias (2025).
- What are typical datasets?
  - Any dataset with a large number of observations over time and space.
  - Macro data: economic growth, house prices, stocks, GDP, inflation, etc.
  - Micro data: Firms, industries, sectors, consumer data,

Panel-Time Series??? II

- Panel-Time Series models are a mix of time series and panel data models with a large number of observations over time (T) and cross-section units (N):
- What is large?
- In *theory* it means that N and T grow jointly to infinity $(N, T) \xrightarrow{j} \infty$. Relative speed depends on estimator.
- In *practice* much harder.
  - Pesaran et al. (1999, p. 80): "When T is large enough that it is sensible to run separate regressions for each group..."
  - Smith and Fuertes (2012, p. 4): "Similar arguments hold for N being large if averaging across units is required for consistency or for central limit theorems to be valid."
  - In a nutshell: enough data to estimate the model in both dimensions.

Panel-Time Series??? III

- Topics when using panel time series models:
  - 'Classical' time series topics (Unit Roots, Stationarity, Cointegration etc.)
  - Slope Heterogeneity
  - Structural Breaks
  - Dependence over time
  - Dependence over space or Cross-Sectional Dependence
- In short: panel-time series models combine the 'best' from panel data and time series!

Table of Conents

1 Introduction

2 The Unobservables

3 The Slopes

4 Cross-Sectional Dependence

5 Estimation

6 ...in Stata

Econometric Model - General Model

- Most general model: dynamic panel model with heterogeneous slopes:

$$y_{i,t} = \lambda_{i,t} y_{i,t-1} + \beta_{i,t} x_{i,t} + u_{i,t} \tag{1}$$

- Main parameter of interest: $\boldsymbol{\pi}_{i,t} = (\lambda_{i,t}, \beta_{i,t})$.
- We allow the coefficients to vary over time and space.
- We can pool the model, assume a group structure or individual slope coefficients.

Additive Fixed Effects

- Very often we want to model the unobserved component including a unit specific μ_i and time fixed effect η_t:

$$u_{i,t} = \mu_i + \eta_t + \vartheta_{i,t} \tag{2}$$

- $u_{i,t}$ contain the additive fixed effects and $\vartheta_{i,t} \sim iid(0, \Sigma^2)$.
- The unit specific fixed effects can be correlated or uncorrelated with the observables in $x_{i,t}$.
- Examples for individual fixed effects are skills, technology, ability, geography, etc.
- The assumption is that the fixed effects capture all unobserved heterogeneity in the data.

$\Rightarrow$ Estimator: Fixed Effects estimator.

Group Fixed Effects

- It is possible to introduce further heterogeneity into the model by allowing group specific time varying unobservables:

$$u_{i,t} = \alpha_{g_i,t} + \vartheta_{i,t}, \;\; g_i \in \{1, ..., G\} \tag{3}$$

  where G is the number of groups and $\alpha_{g_i,t}$ the time varying group specific heterogeneity.
- The group specific effects capture group specific heterogeneity.
- For example effects of being in the same location, part of the same ethnic group, network etc.

⇒ Estimator: Group Fixed Effects estimator (Bonhomme and Manresa, 2015) .

Interactive Fixed Effects

- Closely related to the fixed effects model is the interactive fixed effects model.
- The unit specific fixed effects are allowed to be time varying:

$$u_{i,t} = \gamma_{1,i} f_{1,t} + ... + \gamma_{M,i} f_{M,t} + \vartheta_{i,t} \tag{4}$$

- where we have M interactions between unit specific effects and time effects.
- For example $\gamma_{1,i} f_{1,t}$ can be a individual effect of family upbringing and changes over time.
- In a macro context, f_t is a common factor which affects all individuals differently at a specific point in time.
- The interactive fixed effects model nests the additive fixed effects model.
- The factor structure also introduces cross-section dependence. We will discuss this issue on more detail in the next lecture.

Why slope heterogeneity??

"We are all individuals", Life of Brian (1979)

- The assumption that in economic models causal effects are the same for all firms, countries, sectors, time periods is very questionable.
- We cannot estimate $\pi_{i,t}$, so we need to restrict it, possibilities:
  1. π_i varies in the cross-section dimension: slope homogeneity/heterogeneity

  $\Rightarrow$ Literature: Zellner (1962); Pesaran and Smith (1995); Pesaran et al. (1999); Chudik and Pesaran (2019)
  2. π_t varies in the time dimension: structural breaks

  $\Rightarrow$ Literature: Bai and Perron (1998, 2003); Karavias et al. (2021); Ditzen et al. (2025b)
- In the last decade models with heterogeneous coefficients across space received more attention.
- For structral breaks in Stata see `xtbreak` (Ditzen et al., 2025a).

Testing for slope homogeneity

- In a nutshell, we want to find out if $\boldsymbol{\pi}_i = \boldsymbol{\pi}, \ \forall i \ \in \ N$.
- Tests based on Pesaran and Yamagata (2008) and Blomquist and Westerlund (2013) with hypothesis: H_0 : slope homogeneity vs. H_1 : slope heterogeneity

$$\tilde{\Delta} = \frac{1}{\sqrt{N}} \left(\frac{\sum_{i=1}^{N} \tilde{d}_i - k_2}{\sqrt{2k_2}} \right), \tag{5}$$

- Under the H_0 asymptotically $\tilde{\Delta} \sim_a \mathcal{N}(0, 1)$.
- `xthst`

Stata Syntax

xtbreak and xthst

- *xtbreak*
  - Automatic estimation of number of breaks and break dates:

    `xtbreak` *depvar* [*indepvars*] [*if*] [, *options1 options2 options3 options5 options6*]
  - Known Breaks:

    `xtbreak test` *depvar* [*indepvars*] [*if*] , *breakpoints(numlist—datelist* [`,index|fmt(`*string*`)`] *)* [*options1 options5*]
  - Unknown Breaks

    `xtbreak test` *depvar* [*indepvars*] [*if*] [, `hypothesis(1|2|3) breaks(real)` *options1 options2 options3 options4 options5*]
- *xthst*

  `xthst` *depvar* [*indepvars*] [`,partial(varlist_p) noconstant ar hac bw(integer) whitening kernel(qs|bartlett|truncated) crosssectional(varlist_cr [,cr_lags(numlist)]) nooutput comparehac absorb(varlist_a)`]

Motivation I

- In panel models with $(N, T) \Rightarrow \infty$ the unobserved heterogeneity can be modelled using interactive fixed effects

$$y_{i,t} = \beta_i x_{i,t} + u_{i,t} \tag{6}$$
$$u_{i,t} = \gamma_{u,1,i} f_{u,1,t} + ... + \gamma_{u,M_u,i} f_{u,M_u,t} + \vartheta_{i,t} \tag{7}$$

- with M_u interactions between unit specific effects and time effects.
- $\gamma_{u,m,i}$ is the factor loading of factor $f_{u,m,t}$.
- The explanatory variables can also consist of a factor structure:

$$x_{i,t} = \gamma_{x,1,i} f_{x,1,t} + ... + \gamma_{x,M_x,i} f_{x,M_x,t} + \epsilon_{i,t}$$

- Assumption: $corr(\vartheta_{i,t}, \epsilon_{i,t}) = 0$!

Motivation II

- This setting poses two challenges:
  1. Correlation across units, cross-sectional dependence
  2. If the factors in $x_{i,t}$ and $u_{i,t}$ overlap, the observables and unobservables are correlated
- Our estimator needs to take care of both!

Cross-Sectional Dependence I

- Cross-sectional dependence can origin from a series, unit or a factor, but it boils down that there is a factor driving the dependence.
- Dependence is measured by constant α (Chudik et al., 2011):

$$\lim_{N\rightarrow\infty} N^{-\alpha} \sum_{i=1}^{N} |\gamma_{k,l_k,i}| = K < \infty,$$
$$k = [u, x], \quad l_k = 1, ..., M_k$$

- 4 cases:
  1. $\alpha = 0$ No dependence (independence)
  2. $0 < \alpha < 0.5$ (semi-) weak dependence
  3. $0.5 \leq \alpha < 1$ (semi-) strong dependence
  4. $\alpha = 1$ strong dependence

Cross-Sectional Dependence

- Estimation Number of factors
  - `xtnumfac` (Ditzen and Reese, 2023)
- Estimation exponent of cross-section dependence
  - `xtcse2` (Ditzen, 2021)
- Test for weak cross-sectional dependence

  H_0 weak dependence vs. H_1 strong dependence
  - Test statistic:

$$CD = \sqrt{\frac{2T}{N(N-1))}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \rho_{i,j},$$

  - `xtcd2` (Ditzen, 2018)
  - `xtcsd` (De Hoyos and Sarafidis, 2006)
- Dominant units
  - Work in progress....
  - See `xtpervasive`

Stata Syntax

xtcd2, xtcd2 and xtnumfac

xtcd2

xtcse2 [*varlist*] [*if*] [pca(integer) standardize nocenter nocd residual reps(integer) size(real) tuning(real) lags(integer)]

xtcse2

xtcd2 [*varlist*] [*if*] [pesaran cdw pea cdstar rho pca(integer) reps(integer) kdensity name(string) heatplot[(...)] contour[(...)] noadjust]

xtnumfac

xtnumfac [varlist] [if] [in] [, kmax(#) detail standardize(#)]

Testing for weak cross-sectional dependence

CD Test

- Often it is sufficient to establish if there is weak or strong cross-section dependence.
- Pesaran (2015, 2021) proposes a test for weak cross-section dependence, the CD-test:
  H_0 weak dependence vs. H_1 strong dependence
- Test statistic:
  $$CD = \sqrt{\frac{2T}{N(N-1))}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \rho_{i,j},$$
  with $\rho_{i,j}$ is the correlation coefficient.
- Under the null hypothesis asymptotically: $CD \sim N(0,1)$.
- Problem of the CD test: tends to over reject.
- Further developments: CDw (Juodis and Reese, 2021), CDw with power enhancement (Fan et al., 2015) and CD* (Pesaran and Xie, 2021)

Mean Group Estimator I

- Idea: the time series data is rich enough to use it to estimate the model for each cross-section individually:

$$y_{i,t} = \lambda_i y_{i,t-1} + \beta_i x_{i,t} + u_{i,t} \tag{8}$$

- We can apply all sorts of time series variance estimators to account for time dependence.
- Problem: estimation leads to N estimators.
- Idea: Calculate the average, the mean group estimator (Pesaran et al., 1999; Chudik and Pesaran, 2019):

$$\boldsymbol{\pi}_{MG} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\pi}_i \tag{9}$$

- Implicit assumption: individual coefficients are distributed around a common mean $\boldsymbol{\pi}$ and a variance $\boldsymbol{\Sigma}^2$.
- MG estimator estimates this common mean and the variance.

Mean Group Estimator II

- Variance estimator is based on the individual variation:

$$\boldsymbol{\Sigma}^2_{MG} = \frac{1}{N(N-1)} \sum_{i=1}^{N} (\boldsymbol{\pi}_i - \boldsymbol{\pi}_{MG})^2 \tag{10}$$

- Disadvantages: Receptive to outliers; assumes some sort of "pooled" effect, why not pool in the first instance?
  Assumption: independence of covariates!
- Advantages: Easy implement/estimate, variance estimator does not depend on error distribution
- Possible to use weighted averages.
- Special R^2 required, Holly et al. (2010) propose a mean group R^2.
- Estimator implemented in Stata by the community contributed packages `xtmg` (Eberhardt, 2012) and `xtdcce2` (Ditzen, 2018).

Estimation in the presence cross-section dependence I

- When do we need estimator that take care of cross-section dependence?
- Assume the model with three common factors:

$$y_{i,t} = \lambda_i y_{i,t-1} + \beta_i x_{i,t} + u_{i,t} \qquad x_{i,t} = \gamma_{x,1,i} f_{1,t} + \gamma_{x,2,i} f_{2,t} + \theta_{i,t}$$
$$u_{i,t} = \gamma_{u,1,i} f_{1,t} + \gamma_{u,3,i} f_{3,t} + \vartheta_{i,t}$$

- Cross-section dependence occurs if the factor loadings are not equal to zero.
- Weak dependence vanishes with $(N, T) \rightarrow \infty$, strong dependence remains.
- It implies that all units are exposed to the same common factor (or shock).
- If not accounted for, strong CSD leads to i) OMVB $\gamma_{x,1,i} \neq 0$ and $\gamma_{u,1,i} \neq 0$; ii) Residual correlation $\gamma_{u,1,i} \neq 0$ and $\gamma_{u,3,i} \neq 0$.

Estimation in the presence cross-section dependence II

$$
\begin{aligned}
y_{i,t} &= \lambda_i y_{i,t-1} + \beta_i x_{i,t} + u_{i,t} \\
x_{i,t} &= \gamma_{x,1,i} f_{1,t} + \gamma_{x,2,i} f_{2,t} + \theta_{i,t} \\
u_{i,t} &= \gamma_{u,1,i} f_{1,t} + \gamma_{u,3,i} f_{3,t} + \vartheta_{i,t}
\end{aligned}
$$

- How do we “control” for the (strong) cross-sectional dependence?
  1. Principal Components (Bai, 2009)
  2. Common Correlated Effects (Pesaran, 2006; Chudik and Pesaran, 2015)
  3. Two Stage IV Estimator (Norkute et al., 2021)
- Differentiation between static ($\lambda_i = 0$) and dynamic models ($\lambda_i \neq 0$).

IFE Estimator

- Idea is to approximate the common factors with principal components.
- Advantage: robust to correlations and heteroskedasticty in time and space, can be applied to static and dynamic panels.
- Disadvantage: number of factors needs to be known, iterative estimator.

$$\widehat{\beta}_{IFE} = \left(\sum_{i=1}^{N} \mathbf{x}_i' \mathbf{M}_{\widehat{F}} \mathbf{x}_i\right)^{-1} \left(\sum_{i=1}^{N} \mathbf{x}_i' \mathbf{M}_{\widehat{F}} \mathbf{y}_i\right), \tag{11}$$

where $\mathbf{M}_{\widehat{F}} \equiv \mathbf{I}_T - \widehat{\mathbf{F}}\left(\widehat{\mathbf{F}}'\widehat{\mathbf{F}}\right)^{-1}\widehat{\mathbf{F}}'$ is the projection matrix that projects out $\widehat{\mathbf{F}}$ from the data

regife *depvar* [*indepvars*] [ife(idvar timevar, ndmis) absorb(absvar) vce(vcetype) tolerance(real) maxiterations(real) residuals(varname) bstart(matrix)]

Static CCE

- Pesaran (2006) proposes to span the space of the common factors using cross-section averages in a static model:

$$y_{i,t} = \beta_i x_{i,t} + \psi_{x,i}\bar{x}_t + \psi_{y,i}\bar{y}_t + \epsilon_{i,t}$$

  where $\bar{x}_t = 1/N \sum_{i=1}^{N} x_{i,t}$ and $\bar{y}_t = 1/N \sum_{i=1}^{N} y_{i,t}$ are the cross-section averages.
- Requires at least as many cross-section averages as factors (rank condition).
- Proved to be versatile in many conditions (Kapetanios et al., 2011; Chudik et al., 2011; Westerlund, 2018)
- Implemented in Stata by `xtdcce2` (Ditzen, 2018, 2021)
- Good recent discussion in Reese and Juodis (2026).
- Can be combined with mean group or pooled estimator.

Dynamic CCE

- Static models assume that there is no dependence over time, often Economists are interested in long run effects.
- Dynamic Model:

$$\begin{aligned} y_{i,t} &= \lambda_i y_{i,t-1} + \beta_i x_{i,t} + u_{i,t} \\ x_{i,t} &= \gamma_{x,1,i} f_{1,t} + \gamma_{x,2,i} f_{2,t} + \theta_{i,t} \\ u_{i,t} &= \gamma_{u,1,i} f_{1,t} + \gamma_{u,3,i} f_{3,t} + \vartheta_{i,t} \end{aligned}$$

- Chudik and Pesaran (2015) propose to add $p_T = \sqrt[3]{T}$ lags of the dependent and independent variables:

$$y_{i,t} = \lambda_i y_{i,t-1} + \beta_i x_{i,t} + \sum_{l=0}^{p_T} \left(\psi_{l,x,i} \bar{x}_{t-l} + \psi_{l,y,i} \bar{y}_{t-l} \right) + \epsilon_{i,t}$$

- Factors are allowed to be autocorrelated and covariance stationary.
- Slopes can be pooled and mean group.
- Long run effects via CS-DL or CS-ARDL (Chudik et al., 2016).

Dynamic CCE I

Long Run Effects

- Long run coefficients estimate the long run effect of a change of x on y. This implies the change is measured over several periods.
- A more general representation with further lags of the dependent and independent variable in the form of an ARDL(p_y, p_x) model is:

$$y_{i,t} = \sum_{l=1}^{p_y} \lambda_{l,i} y_{i,t-l} + \sum_{l=0}^{p_x} \beta_{l,i} x_{i,t-l} + u_{i,t}. \tag{12}$$

- where p_y and p_x is the lag length of y and x.

Dynamic CCE II

Long Run Effects

- The long run coefficient of θ and the mean group coefficient are:

$$\theta_i = \frac{\sum_{l=0}^{p_x} \beta_{l,i}}{1 - \sum_{l=1}^{p_y} \lambda_{l,i}}, \quad \bar{\theta}_{MG} = \frac{1}{N} \sum_{i=1}^{N} \theta_i \tag{13}$$

- How to estimate θ_i and $\bar{\theta}_{MG}$?
- Chudik et al. (2016) propose two methods, the cross-sectionally augmented ARDL (CS-ARDL) and the cross-sectionally augmented distributed lag (CS-DL) estimator.
- Using an error correction model (CS-ECM).
- Advantage: consistent estimates irrespective x and y are I(1) or I(0) (Chudik et al., 2016), unit roots in regressors and/or factors; serial correlation in errors and factors.

Dynamic CCE

A word of caution

- General Advice: Be careful in dynamic models with common factors/strong dependence.
  1. Lag Selection difficult
  2. Different order of integration or autocorrelation of factors + many cross-sectional units.
  3. Postestimation (CD test; Rank Condition; ICs) only applicable with care.
  4. Requires large T!
  5. Nickell Bias (Nickell, 1981) is of order $O(T^{-1})$.
- Partial solution: half panel jackknife bias correction and recursive mean adjustment (Chudik and Pesaran, 2015; Chudik et al., 2018) and bias corrected CCEP (De Vos and Everaert, 2021).

Two Stage IV Estimator I

- Alternative to CCE and PCA: 2SIV (Norkute et al., 2021).
- Mix of IV estimator using lags and principal components.
- Idea: use instruments to approximate common factors.
- Advantage: versatile, allows for weak and strong cross-section dependence, no bias correction needed, no rank condition necessary, slope heterogeneity and long and short run coefficients.
- $\sqrt{NT}$ consistent and symbiotically normal distributed.
- Pooled and mean group estimates.
- Can be combined with external instruments in case of endogeneity.

Two Stage IV Estimator II

- Two Stages:
  1. Eliminate Common Factors in X using principal components and obtain consistent estimator of β using defactored regressors as instruments.
  2. Defactor entire model based on factors obtained from residuals from first stage and use same instruments as before.
- Over-identification restrictions (OIR) test statistic can be used to asses
  1. Exogeneity of defactored covariates with respect to idiosyncratic error
  2. If coefficients are heterogeneous, OIR should reject null hypothesis asymptotically
- Null hypothesis: instruments are exogenous, "non-defactored "instruments are valid.
- Especially 1) is informative if covariates are correlated to factors and loadings.

Stata Syntax
IFE

regife *depvar* [*indepvars*] [ife(idvar timevar, ndmis)
absorb(absvar) vce(vcetype) tolerance(real) maxiterations(real)
residuals(varname) bstart(matrix)]

Stata Syntax

xtdcce2

Syntax:

xtdcce2 *depvar* [*indepvars*] [*varlist2* = *varlist_iv*] [*if*]
$\boxed{\textit{crosssectional(varlist\_cr [,lags(\#)])}}$ [, nocrosssectional pooled(*varlist_p*)
$\boxed{\texttt{cr\_lags(\#)}}$ ivreg2options(*string*) e_ivreg2 ivslow $\boxed{\texttt{lr(}\textit{varlist\_lr}\texttt{)}}$
$\boxed{\texttt{lr\_options(}\textit{string}\texttt{)}}$ pooledconstant noconstant reportconstant trend
pooledtrend jackknife recursive exponent xtcse2options(string)
nocd fullsample showindividual fast blockdiaguse nodimcheck
useinvsym useqr noomitted showomitted]

Stata Syntax

`xtdcce2`

$$
\begin{aligned}
y_{i,t} =& \alpha_i + \sum_{l=1}^{p_y} \lambda_{l,i} y_{i,t-l} + \sum_{l=0}^{p_x} \beta_{l,i} x_{i,t-l} \\
& + \sum_{l=0}^{p_{\bar{y}}} \gamma_{y,i,l} \bar{y}_{t-l} + \sum_{l=0}^{p_{\bar{x}}} \gamma_{x,i,l} \bar{x}_{t-l} + e_{i,t}
\end{aligned}
$$

- `crosssectional`(*varlist*) specifies cross sectional means, i.e. variables in $\bar{z}_t$. These variables are partialled out.
- `cr_lags`(#) defines number of lags (p_T) of the cross sectional averages. The number of lags can be variable specific. The same order as in `cr`() applies, hence if `cr(y x)`, then `cr_lags(`$p_{\bar{y}}$ $p_{\bar{x}}$`)`.
- `lr(`*varlist_lr*`)` and `lr_options(`*string*`)` define the long run coefficients and options. For an ARDL (2,2) model it would be: `lr(L(1/2).y L(0/2).x) lr_options(ardl)`

Dynamic CCE

CS-DL

- Lags of the cross-sectional averages are added to account for cross-sectional dependence.
- Together with the lags, the equation can be written as:

$$y_{i,t} = \theta_i x_{i,t} + \delta_i(L)\Delta x_{i,t} + \tilde{u}_{i,t}$$

$$\Rightarrow y_{i,t} = \theta_i x_{i,t} + \sum_{l=0}^{p_x-1} \delta_{i,l}\Delta x_{i,t-l} + \sum_{l=0}^{p_{\bar{y}}} \gamma_{y,i,l}\bar{y}_{i,t-l} + \sum_{l=0}^{p_{\bar{x}}} \gamma_{x,i,l}\bar{x}_{i,t-l} + e_{i,t}$$

- where $p_{\bar{y}}$ and $p_{\bar{x}}$ is the number of lags of the cross-sectional averages.
- The mean group estimates are then: $\hat{\hat{\theta}}_{MG} = \sum_{i=1}^{N} \hat{\theta}_i$
- The variance/covariance matrix for the mean group coefficients is the same as for the "normal" (D)CCE estimator.

Dynamic CCE

CS-ARDL

- Idea: Estimate the short run coefficients first and then calculate the long run coefficients:

$$y_{i,t} = \sum_{l=1}^{p_y} \lambda_{l,i} y_{i,t-l} + \sum_{l=0}^{p_x} \beta_{l,i} x_{i,t-l} + \sum_{l=0}^{p_T} \gamma'_{i,l} \bar{\mathbf{z}}_{t-l} + e_{i,t}.$$

- with $\bar{\mathbf{z}}_{t-l} = (\bar{y}_{i,t-l}, \bar{x}_{i,t-l})$
- and the long run coefficients and the mean group estimates are

$$\hat{\theta}_{CS-ARDL,i} = \frac{\sum_{l=0}^{p_x} \hat{\beta}_{l,i}}{1 - \sum_{l=1}^{p_y} \hat{\lambda}_{l,i}}, \quad \hat{\bar{\theta}}_{MG} = \sum_{i=1}^{N} \hat{\theta}_i$$

- The variance/covariance matrix for the mean group coefficients is the same as for the "normal" (D)CCE estimator.

Stata Syntax

`xtivdfreg`

`xtivdfreg` *depvar* [*indepvars*] [`absorb(varlist) iv(iv_spec)`
`factmax(#) mg`]

- `absorb` categorical variables that identify the fixed effects to be absorbed
- `iv(iv_spec)` instruments; can be specified more than once
- `factmax(#)` specify the maximum number of factors
- `mg` compute the mean-group estimator

References I

Bai, B. Y. J., and P. Perron. 1998. Estimating and Testing Linear Models with Multiple Structural Changes. Econometrica, 66(1): 47–78.

Bai, J. 2009. Panel Data Models With Interactive Fixed Effects. Econometrica 77(4): 1229–1279.

Bai, J., and P. Perron. 2003. Computation and analysis of multiple structural change models. Journal of Applied Econometrics 18(1): 1–22.

Baltagi, B. H. 2015. The Oxford Handbook of Panel Data, vol. 53. Oxford University Press.

Bersvendsen, T., and J. Ditzen. 2021. Testing for slope heterogeneity in Stata. The Stata Journal 21(1): 51–80.

Blomquist, J., and J. Westerlund. 2013. Testing slope homogeneity in large panels with serial correlation. Economics Letters 121(3): 374–378.

Bonhomme, S., and E. Manresa. 2015. Grouped Patterns of Heterogeneity in Panel Data. Econometrica 83(3): 1147–1184.

References II

Chudik, A., K. Mohaddes, M. H. Pesaran, and M. Raissi. 2016. Long-Run Effects in Large Heterogeneous Panel Data Models with Cross-Sectionally Correlated Errors. In Essays in Honor of Aman Ullah, 85–135.

Chudik, A., and M. H. Pesaran. 2015. Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. Journal of Econometrics 188(2): 393–420.

———. 2019. Mean group estimation in presence of weakly cross-correlated estimators. Economics Letters 175: 101–105.

Chudik, A., M. H. Pesaran, and E. Tosetti. 2011. Weak and strong cross-section dependence and estimation of large panels. The Econometrics Journal 14(1): C45–C90.

References III

Chudik, A., M. H. Pesaran, and J. C. Yang. 2018. Half-panel jackknife fixed-effects estimation of linear panels with weakly exogenous regressors. Journal of Applied Econometrics 33(6): 816–836.

De Hoyos, R. E., and V. Sarafidis. 2006. Testing for cross-sectional dependence in panel-data models. The Stata Journal 6(4): 482–496.

De Vos, I., and G. Everaert. 2021. Bias-Corrected Common Correlated Effects Pooled Estimation in Dynamic Panels. Journal of Business and Economic Statistics 39(1): 294–306. URL `https://doi.org/10.1080/07350015.2019.1654879`.

Ditzen, J. 2018. Estimating dynamic common-correlated effects in Stata. The Stata Journal 18(3): 585 – 617.

———. 2021. Estimating long run effects and the exponent of cross-sectional dependence: an update to xtdcce2. The Stata Journal 21(3): 687–707. URL `https://ideas.repec.org/p/bzn/wpaper/bemps81.html`.

References IV

Ditzen, J., and Y. Karavias. 2025. Interactive, Grouped and Non-separable Fixed Effects: A Practitioner's Guide to the New Panel Data Econometrics . URL https://arxiv.org/abs/2507.19099.

Ditzen, J., Y. Karavias, and J. Westerlund. 2025a. Testing and Estimating Structural Breaks in Time Series and Panel Data in Stata. The Stata Journal 25(3). URL https://arxiv.org/abs/2110.14550.

———. 2025b. Multiple Structural Breaks in Interactive Effects Panel Data Models. Journal of Applied Econometrics 40(1): 74–88.

Ditzen, J., and S. Reese. 2023. A battery of estimators for the number of common factors in time series and panel data models. The Stata Journal 23(2): 438–454. URL https://journals.sagepub.com/doi/abs/10.1177/1536867X231175305.

Eberhardt, M. 2012. Estimating panel time series models with heterogeneous slopes. The Stata Journal 12(1): 61–71.

References V

Elhorst, J. P., M. Gross, and E. Tereanu. 2021. Cross-Sectional Dependence and Spillovers in Space and Time: Where Spatial Econometrics and Global Var Models Meet. Journal of Economic Surveys 35(1): 192–226.

Fan, J., Y. Liao, and J. Yao. 2015. Power Enhancement in High-Dimensional Cross-Section Tests. Econometrica 83(4): 1497–1541.

Holly, S., M. H. Pesaran, and T. Yamagata. 2010. A spatio-temporal model of house prices in the USA. Journal of Econometrics 158(1): 160–173.

Juodis, A., and S. Reese. 2021. The Incidental Parameters Problem in Testing for Remaining Cross-Section Correlation. Journal of Business & Economic Statistics .

References VI

Kapetanios, G., M. H. Pesaran, and T. Yamagata. 2011. Panels with non-stationary multifactor error structures. Journal of Econometrics 160(2): 326–348.

Karavias, Y., J. Westerlund, and P. Narayan. 2021. Structural Breaks in Interactive Effects Panels and the Stock Market Reaction to COVID-19 .

Nickell, S. 1981. Biases in Dynamic Models with Fixed Effects. Econometrica 49(6): 1417–1426.

Norkute, M., V. Sarafidis, T. Yamagata, and G. Cui. 2021. Instrumental variable estimation of dynamic linear panel data models with defactored regressors and a multifactor error structure. Journal of Econometrics 220(2): 416–446.

Pesaran, M. H. 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74(4): 967–1012.

References VII

———. 2015. Testing Weak Cross-Sectional Dependence in Large Panels. Econometric Reviews 34(6-10): 1089–1117.

———. 2021. General diagnostic tests for cross-sectional dependence in panels. Empirical Economics 60(1): 13–50. URL https://doi.org/10.1007/s00181-020-01875-7.

Pesaran, M. H., Y. Shin, and R. P. Smith. 1999. Pooled Mean Group Estimation of Dynamic Heterogeneous Panels. Journal of the American Statistical Association 94(446): 621 –634.

Pesaran, M. H., and R. Smith. 1995. Estimating long-run relationships from dynamic heterogeneous panels. Journal of Econometrics 68(1): 79–113.

Pesaran, M. H., and Y. Xie. 2021. A Bias-Corrected CD Test for Error Cross- Sectional Dependence in Panel Data Models with Latent Factors. Cambridge Working Papers in Economics 2158.

References VIII

Pesaran, M. H., and T. Yamagata. 2008. Testing slope homogeneity in large panels. Journal of Econometrics 142(1): 50–93.

Reese, S., and A. Juodis. 2026. Five lessons for applied researchers from twenty years of common correlated effects estimation. Journal of Econometrics .

Sarafidis, V., and T. Wansbeek. 2012. Cross-Sectional Dependence in Panel Data Analysis. Econometric Reviews 31(5): 483–531.

Smith, R. P., and A.-M. Fuertes. 2012. Panel Time-Series. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.168.4383.

Sul, D. 2019. Panel data econometrics : common factor analysis for empirical researchers. Routledge.

References IX

Westerlund, J. 2018. CCE in panels with general unknown factors. Econometrics Journal 21(3): 264–276.

Zellner, A. 1962. An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias. Journal of the American Statistical Association 57(298): 348–368.