Freie Universität Bozen
Libera Università di Bolzano
Università Liedia de Bulsan

Panel-data models with large N and large T: An overview

2021 Stata Economics Virtual Symposium

Jan Ditzen

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

November 10, 2021

Motivation

- This talk will be an overview and not very technical.
- For most topics it is a touch-and-go rather than a deep discussion.
- Panel-Time Series regression became very popular in the past years.
- Growing body of literature on methods, for an overview see Smith and Fuertes (2012); Baltagi (2015); Sul (2019); Elhorst et al. (2021).
- Aim of this talk: Overview of the literature and methods and their applications in Stata.

Panel-Time Series??? I

- 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 with the same speed to infinity $(N, T) \xrightarrow{j} \infty$ with $T/N \rightarrow \kappa, \; 0 < \kappa < \infty$.
- 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??? II

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

Econometric Model - General Model

- Most general model: dynamic panel model with heterogeneous slopes and interactive fixed effects:

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

- We observe $y_{i,t}$ and $x_{i,t}$, the common factors $(f_{l,t})$ and the loadings $(\gamma_{k,l,i})$ are unobserved.
- $\xi_{i,t}$ and $\epsilon_{i,t}$ are both IID white noise.
- Heterogeneous slopes for $\lambda_i \sim IID(\lambda, \sigma^2_\lambda)$ and $\beta_i \sim IID(\beta, \sigma^2_\beta)$
- Unit specific and time fixed effects can be nested in the interactive fixed effects.
- Potential dependence across units via the common factors $(f_{m,t}, m = 1, 2, 3)$.

The classics

Unit Roots, Stationary, Cointegration

- The 'classics' receive a lot of attention and there are plenty of packages in Stata
- Unit Roots:
  - `xtbunitroot` (Chen et al., 2021)
  - `xtunitroot`
  - `multipurt`, `pescadf`,....
- Granger Causality:
  - `xtgranger` (Xiao et al., 2021)
  - `xtgcause` (Lopez and Weber, 2017)
- Cointegration:
  - `xtwest` (Persyn and Westerlund, 2008)
  - `xtpedroni` (Neal, 2014)

Cross-Section Dependence (CSD) I

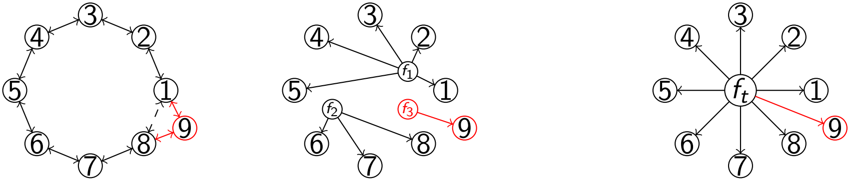
$$x_{i,t} = \gamma_{x,1,i} f_{1,t} + \gamma_{x,2,i} f_{2,t} + \xi_{i,t}$$
$$u_{i,t} = \gamma_{u,1,i} f_{1,t} + \gamma_{u,3,i} f_{3,t} + \epsilon_{i,t}$$

- Cross-section dependence occurs if the factor loadings are not equal to zero.
- It implies that all units are exposed to the same common factor (or shock).
- If it is not accounted for, then CSD potentially leads to:
  1. Omitted variable bias if $\gamma_{x,1,i} \neq 0$ and $\gamma_{u,1,i} \neq 0$
  2. Residuals can be correlated across units if $\gamma_{u,1,i} \neq 0$ and $\gamma_{u,3,i} \neq 0$
- If $\gamma_{x,1,i} = \gamma_{x,2,i} = 0$ then no first order problem for estimator.

Cross-Section Dependence (CSD) II

- Dependence is measured by constant α (Chudik et al., 2011)

$$\lim_{N\to\infty} N^{-\alpha} \sum_{i=1}^{N} |\gamma_{k,i,l}| = K < \infty$$



(Semi-) Weak CSD: $0 \leq \alpha < 0.5$

(Semi-) Strong CSD: $0.5 \leq \alpha \leq 1$

Weak and strong cross-sectional dependence with additional unit 9 as $N \to \infty$

- We can a) estimate the number of factors, b) estimate the exponent of cross-section dependence and c) test for weak cross-section dependence

Estimating number of factors and exponent of CSD

Estimating number of factors

- Either directly estimated (Onatski, 2010; Ahn and Horenstein, 2013; Gagliardini et al., 2019) or determined by information criteria (Bai and Ng, 2002).
- `xtnumfac`[1] (Reese and Ditzen, 2021) implements the methods above.

Estimating the exponent of cross-section dependence

- Bailey et al. (2016, 2019) propose an estimator for the exponent of CSD.
- Estimation only possible for $\alpha > 1/2$.
- In Stata implemented by `xtcse2` (Ditzen, 2021).

[1] available upon request/soon

Testing for weak cross-sectional dependence

...or a zoo of tests...

- Pesaran (2015, 2021) proposes a test for weak cross-section dependence, the CD-test:

  H_0 weak dependence vs. H_1 strong dependence
- Further developments: CDw (Juodis and Reese, 2021), CDw with power enhancement (Fan et al., 2015) and CD* (Pesaran and Xie, 2021)
- CDw, CDw+ and CD* should be applied to residuals, CD can be generally applied.
- In Stata there is a zoo of tests: `xtcsd` (De Hoyos and Sarafidis, 2006), `xtcd` (Eberhardt, 2011), `xtcdf` (Wursten, 2017) and `xtcd2` (Ditzen, 2018).
- The latest version of `xtcd2`[2] implements the CD, CDw, CDw+ and CD*.

[2] update available soon

Testing for slope homogeneity

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

- With a large number of cross-sectional units, we are able to estimate coefficients for each unit, i.e. λ_i and β_i.
- Is this necessary? Pesaran and Yamagata (2008) and Blomquist and Westerlund (2013) develop test for slope homogeneity $\beta_i = \beta \forall i$.

  H_0 : slope homogeneity vs. H_1 : slope heterogeneity
- In Stata: `xthst` (Bersvendsen and Ditzen, 2021).

Structural Breaks

- We have a large number of observations over time, possible that coefficients break.
- Assume we have s breaks, then:

$$y_{i,t} = \alpha_i + \beta_i x_{i,t} + \delta_{l,i} z_{i,t} + u_{i,t}, \quad l = 1, ..., s$$

- Bai and Perron (1998, 2003) propose methods to estimate number of breaks and location in time series. Three tests:
  1. No breaks vs. s breaks
  2. No breaks vs. up to s breaks
  3. s breaks vs. $s+1$ breaks
- Karavias et al. (2021); Ditzen et al. (2021a) generalize the methods for panel time series models with cross-section dependence.
- In Stata: `xtbreak` (Ditzen et al., 2021b).

Estimation

...finally...

- Pesaran (2006); Chudik and Pesaran (2015) proposes to approximate the common factors using cross-section averages.

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

where $p_T = \sqrt[3]{T}$ are the number of lags of the cross-section averages $\bar{x}_t = 1/N \sum_{i=1}^{N} x_{i,t}$ and $\bar{y}_t = 1/N \sum_{i=1}^{N} y_{i,t}$.

- We can use the pooled or mean group estimator.
- In a dynamic model the pooled estimator will be biased however!
- Long run relationships can be estimated using CS-ECM, CS-ARDL and CS-DL estimator.

Estimation

CCE-MG and CCE-P

Mean Group Estimator (Pesaran and Smith, 1995; Pesaran, 2006; Chudik and Pesaran, 2019)

- $\hat{\beta}_{MG} = \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_i$
- Variance estimator $V(\hat{\beta}_{MG}) = \frac{1}{N(N-1)}\sum_{i=1}^{N}\left(\hat{\beta}_i - \hat{\beta}_{MG}\right)^2$
- Note: Variance estimator is independent (!!) of residuals. It relies on the assumption that we consistently estimate the individual slope coefficients and that they are distributed around a common mean.

Pooled CCE Estimator (Pesaran, 2006)

- Estimate β_p directly with the condition $\beta_i = \beta_p$.
- Various variance estimators, such as $V(\hat{\beta}_p)_{np} = f(\hat{\beta}_i, \hat{\beta}_{MG}, \tilde{X}'\tilde{X})$ or $V(\hat{\beta}_p)_{hac} = f(\hat{\beta}_p, \tilde{X}'\tilde{X}, \hat{\epsilon}_{i,t})$.
- Depending on the estimator, we need to make sure that the residuals are cross-section dependence free!

Introduction

- We want to estimate a simple Solow-style growth model.
- Data: Penn World Tables with 93 countries over years 1960 - 2007
- Static model:

$$log\_rgdpo_{i,t} = \beta_{0,i} + \beta_{1,i} log\_hc_{i,t} \\ + \beta_{2,i} log\_ck_{i,t} + \beta_{3,i} log\_ngd_{i,t} + u_{i,t}$$

- Variables:
  - `log_rgdpo`: Real GDP per capita
  - `log_hc`: human capital
  - `log_ngd`: population growth rate
  - `log_ck`: capital stock

Testing for cross-section dependence

- First step we are using `xtcd2` to test for weak cross-section dependence.[3]

  H_0 : weak dependence, H_1: strong dependence

```
. xtcd2 log_rgdpo log_hc log_ck log_ngd, pesaran
Testing for weak cross-sectional dependence (CSD)
   H0: weak cross-section dependence
   H1: strong cross-section dependence
```

	CD
log_rgdpo	145.97 (0.000)
log_hc	427.70 (0.000)
log_ck	417.23 (0.000)
log_ngd	75.83 (0.000)

```
p-values in parenthesis.
References
  CD:        Pesaran (2015, 2021)
```

- We only use the CD test (Pesaran, 2015) as indicated by the option `pesaran`.
- We find strong cross-section dependence for all variables.

[3] Preliminary version, results might change!

Estimating the exponent and number of common factors I

- We use xtcse2

```
. xtcse2 log_rgdpo log_hc log_ck log_ngd, nocd
Cross-Sectional Dependence Exponent Estimation and Test
Panel Variable (i): id
Time Variable (t): year
Estimation of Cross-Sectional Exponent (alpha)
```

variable	alpha	Std. Err.	[95% Conf.	Interval]
log_rgdpo	.9554411	.0399948	.8770528	1.033829
log_hc	1.002321	.0544753	.8955511	1.10909
log_ck	1.002318	.0676462	.8697337	1.134902
log_ngd	.9790158	.133342	.7176703	1.240361

```
0.5 <= alpha < 1 implies strong cross-sectional dependence.
Variables are centered around zero.
```

- Again, test for cross-section dependence confirms earlier results, strong cross-section dependence found.

Estimating the exponent and number of common factors II

- Next we estimate the number of factors using `xtnumfac` (only for the dependent variable)

```
. xtnumfac log_rgdpo
Information criteria for number of common factors in log_rgdpo
                                   N  =        93
                                   T  =        48
```

IC	# factors	IC	# factors
PC_{p1}	8	IC_{p1}	8
PC_{p2}	8	IC_{p2}	8
PC_{p3}	8	IC_{p3}	8
ER	1	GR	1
GOL	1	ED	4

```
8 factors maximally considered.
PC_{p1},...,IC_{p3} from Bai and Ng (2002)
ER, GR from Ahn and Horenstein (2013)
ED from Onatski (2010)
GOL from Gagliardini, Ossola, Scaillet (2019)
```

- Evidence for common factors, but results differ.
- Ignore GOL as it is only for residuals.
- Advantage of the CCE estimator, no detailed knowledge of number of factors required.

Testing for slope homogeneity

- `xthst` implements tests for slope homogeneity:

  H_0 homogeneous slopes vs. H_1 heterogeneous slopes

- We control for cross-section dependence using the `cr()` option:

```
. xthst log_rgdpo log_hc log_ck log_ngd, ///
> cr(log_rgdpo log_hc log_ck log_ngd) hac
Testing for slope heterogeneity
(Blomquist, Westerlund. 2013. Economic Letters)
H0: slope coefficients are homogenous
-------------------------------------
         Delta          p-value
         45.915           0.000
 adj.    48.842           0.000
-------------------------------------
HAC Kernel: bartlett
with average bandwith 3
Variables partialled out: constant
Cross Sectional Averaged Variables: log_rgdpo log_hc log_ck log_ngd
```

- We find that the slopes are heterogenous.

Number of breaks

- xtbreak implements estimators for number of and location of breaks.
- As before, we control for strong cross-section dependence using the csd option:

```
. xtbreak log_rgdpo log_hc log_ck log_ngd, csd vce(hac)
Sequential test for multiple breaks at unknown breakpoints
(Ditzen, Karavias & Westerlund. 2021)
```

	Test Statistic	Bai & Perron Critical Values: 1% Critical Value	5% Critical Value	10% Critical Value
F(1|0)	0.08	18.26	13.98	12.08
F(2|1)	5.21	19.77	15.72	13.91
F(3|2)	1.66	20.75	16.83	14.96
F(4|3)	0.62	21.98	17.61	15.68
F(5|4)	0.63	22.46	18.14	16.35
Detected number of breaks:		.	.	.

```
The detected number of breaks indicates the highest number of
 breaks for which the null hypothesis is rejected.
No breaks found, cannot estimate breakpoints.
```

- We find no evidence for structural breaks.
- Note: assumption that slope coefficients are homogenous!

Estimation

- We established we have a) cross-section dependence, b) heterogeneous slopes, and c) no breaks.
- Hence, we need an estimator which accounts for a) and b): the CCE-MG estimator and estimate the following equation:

$$\begin{aligned} log\_rgdpo_{i,t} &= \beta_{0,i} + \beta_{1,i} log\_hc_{i,t} + \beta_{2,i} log\_ck_{i,t} \\ &+ \beta_{3,i} log\_ngd_{i,t} + \gamma_i \bar{z}_t + \epsilon_{i,t} \end{aligned}$$

- where $\bar{z}_t$ is a vector with the cross-section averages of the dependent and independent variables.
- The CCE-MG estimator will be used and remaining cross-section dependence in the residuals is tested after.
- The analysis will be done using `xtdcce2`.[4]

[4] update available soon

Static MG

```
. xtdcce2 log_rgdpo log_hc log_ck log_ngd, fast2 nocross
(Dynamic) Common Correlated Effects Estimator - Mean Group (xtdcce2fast)

Panel Variable (i): id                              Number of obs      =       4371
Time Variable (t): year                             Number of groups   =         93

Degrees of freedom per group:                       Obs per group (T) =         47
 without cross-sectional averages    = 43
 with cross-sectional averages       = 43
Number of                                           R-squared (mg)     =       0.92
 cross-sectional lags                none           CD Statistic       =      31.22
 variables in mean group regression = 372              p-value         =     0.0000
 variables partialled out            = 0
--------------------------------------------------------------------------------
   log_rgdpo |      Coef.   Std. Err.       z    P>|z|      [95% Conf. Interval]
-------------+------------------------------------------------------------------
 Mean Group: |
      log_hc | -.1585699    .2677876    -0.59    0.554     -.683424    .3662842
      log_ck |  .3685834    .0416629     8.85    0.000     .2869257    .4502412
     log_ngd |   .318421    .1599968     1.99    0.047     .0048329     .632009
       _cons |  4.900441    .5768085     8.50    0.000     3.769917    6.030965
--------------------------------------------------------------------------------
Mean Group Variables: log_hc log_ck log_ngd _cons
```

- We use the option `fast2` which uses a speed optimized version of `xtdcce`.
- No cross-sectional averages added, a lot of cross-section dependence left in the model.
- Likely that the remaining CSD is correlated with the CSD in the variables.
- Next step: add cross-section averages.

Static CCE-MG

```
. xtdcce2 log_rgdpo log_hc log_ck log_ngd, fast2 ///
> cr(log_rgdpo log_hc log_ck log_ngd)
(Dynamic) Common Correlated Effects Estimator - Mean Group (xtdcce2fast)
Panel Variable (i): id                            Number of obs      =      4371
Time Variable (t): year                           Number of groups   =        93
Degrees of freedom per group:                     Obs per group (T)  =        47
 without cross-sectional averages    = 43
 with cross-sectional averages       = 39
Number of                                         R-squared (mg)     =      0.96
 cross-sectional lags                = 0          CD Statistic       =      0.29
 variables in mean group regression = 372            p-value         =    0.7754
 variables partialled out            = 372
-------------------------------------------------------------------------------
   log_rgdpo|     Coef.   Std. Err.      z     P>|z|     [95% Conf. Interval]
-------------------------------------------------------------------------------
 Mean Group:|
      log_hc| -.6393411   .3986611   -1.60    0.109    -1.420702   .1420203
      log_ck|  .2714685   .0535893    5.07    0.000     .1664354   .3765017
     log_ngd| -.0349364   .1418044   -0.25    0.805    -.312868    .2429953
       _cons| -2.315308   1.330156   -1.74    0.082    -4.922366   .2917494
-------------------------------------------------------------------------------
Mean Group Variables: log_hc log_ck log_ngd _cons
Cross Sectional Averaged Variables: log_rgdpo log_hc log_ck log_ngd
```

- Cross-section dependence is well taken out of the model.
- Only physical capital is significant.
- Comparing the model to a pooled model possible but does not improve results.

Static CCE-MG

Testing for CSD

- Re-run the tests for weak cross-section dependence[5]

```
. xtcd2
Residuals calculated using  predict, residuals from xtdcce2.
Testing for weak cross-sectional dependence (CSD)
   H0: weak cross-section dependence
   H1: strong cross-section dependence
```

	CD	CDw	CDw+	CD*
residuals	0.29	-0.82	5550.32	-1.32
	(0.775)	(0.415)	(0.000)	(0.188)

```
p-values in parenthesis.
References
  CD:        Pesaran (2015, 2021)
  CDw:       Juodis, Reese (2021)
  CDw+:      CDw with power enhancement from Fan et. al. (2015)
  CD*:       Pesaran, Xie (2021) with 4 PC(s)
```

[5] Preliminary version, results might change!

Dynamic Models

- It is more standard to estimate a dynamic model:

$$log\_rgdpo_{i,t} = \beta_{0,i} + \lambda_i log\_rgdpo_{i,t-1} + \beta_{1,i} log\_hc_{i,t} + \beta_{2,i} log\_ck_{i,t} + \beta_{3,i} log\_ngd_{i,t} + \sum_{l=1}^{p_T} \gamma_{i,l} \bar{z}_{t-l} + \epsilon_{i,t}$$

- To account for the dynamics, we add $p_T = \sqrt[3]{T}$ lags of the cross-section averages (Chudik and Pesaran, 2015).
- Some notes:
  - Interpretation of the CD test requires care.
  - CCE Pooled would be biased.

Dynamic CCE-MG

```
. xtdcce2 log_rgdpo L.log_rgdpo log_hc log_ck log_ngd, fast2 ///
> cr(log_rgdpo log_hc log_ck log_ngd) cr_lags(3)
(Dynamic) Common Correlated Effects Estimator - Mean Group (xtdcce2fast)

Panel Variable (i): id                          Number of obs      =       4092
Time Variable (t): year                         Number of groups   =         93

Degrees of freedom per group:                   Obs per group (T)  =         44
 without cross-sectional averages    = 39
 with cross-sectional averages       = 23
Number of                                       R-squared (mg)     =       0.98
 cross-sectional lags                = 3        CD Statistic       =       1.40
 variables in mean group regression = 465          p-value         =     0.1601
 variables partialled out            = 1488
------------------------------------------------------------------------------
   log_rgdpo |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 Mean Group:
 L.log_rgdpo |   .3864255   .0311066    12.42   0.000     .3254577    .4473934
      log_hc |  -1.286829   .3900789    -3.30   0.001     -2.05137   -.5222887
      log_ck |   .2098063     .04939     4.25   0.000     .1130036    .3066089
     log_ngd |   .0052778   .1006462     0.05   0.958    -.1919851    .2025408
       _cons |  -2.174129   1.786946    -1.22   0.224    -5.676478    1.328221
------------------------------------------------------------------------------
Mean Group Variables: L.log_rgdpo log_hc log_ck log_ngd _cons
Cross Sectional Averaged Variables: log_rgdpo log_hc log_ck log_ngd
```

- Human and physical capital are now significant.
- Note the number of variables in the mean group regression and being partialled out.
- Value of CD-test statistic still low.
- We can estimate the long run relationships using the CS-ARDL estimator.

Dynamic CCE - CS-ARDL

```
. xtdcce2 log_rgdpo , fast2 cr(log_rgdpo log_hc log_ck log_ngd) cr_lags(3) ///
> lr(L.log_rgdpo log_hc log_ck log_ngd) lr_options(ardl)
(Dynamic) Common Correlated Effects Estimator - Mean Group (xtdcce2fast)
Panel Variable (i): id                          Number of obs     =       4092
Time Variable (t): year                         Number of groups  =         93
Degrees of freedom per group:                   Obs per group (T) =         44
 without cross-sectional averages    = 39
 with cross-sectional averages       = 23
Number of                                       R-squared (mg)    =       0.98
 cross-sectional lags                = 3        CD Statistic      =       1.40
 variables in mean group regression = 465          p-value        =     0.1601
 variables partialled out            = 1488
------------------------------------------------------------------------------
     log_rgdpo |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------+--------------------------------------------------------------
 Short Run Est.|
---------------+--------------------------------------------------------------
 ...
---------------+--------------------------------------------------------------
 Adjust. Term  |
---------------+--------------------------------------------------------------
   Mean Group: |
   lr_log_rgdpo| -.6135745   .0311066  -19.72   0.000    -.6745423   -.5526066
---------------+--------------------------------------------------------------
 Long Run Est. |
---------------+--------------------------------------------------------------
   Mean Group: |
      lr_log_hc| -1.854644   .7475237   -2.48   0.013    -3.319763   -.3895246
      lr_log_ck|  .0895815   .2191939    0.41   0.683    -.3400306    .5191935
     lr_log_ngd| -.3731947   .3800709   -0.98   0.326     -1.11812    .3717306
------------------------------------------------------------------------------
Mean Group Variables: L.log_rgdpo log_hc log_ck log_ngd _cons
Cross Sectional Averaged Variables: log_rgdpo log_hc log_ck log_ngd
Long Run Variables: L.log_rgdpo log_hc log_ck log_ngd
Cointegration variable(s): lr_log_rgdpo
```

- The long run estimates are $\beta_{LR,i} = (\sum_{l=1}^{L_x} \beta_{l,i})/(1 - \sum_{l=1}^{L_y} \alpha_{l,i})$ and we can apply the MG estimator to $\beta_{LR,i}$.
- Adjustment is -0.614.
- CD test statistic still very low.

Conclusion

or take aways

- Panel-Time series models offer a lot of flexibility and insights into data.
- They require large N and large T.
- Account for cross-section dependence appropriately, there is some caution needed when using dynamic models.
- Cross-section dependence is a first order problem for estimators, using CSD robust standard errors does not help!
- Do you assume heterogeneous slopes or not?
- The road ahead:
  - Spatial Temporal Error correction models (Bhattacharjee et al., 2021)
  - Endogeneity (see next talk)
  - Lag selection
  - Correlation between common factors in exogenous variables (x) and error (u)

References I

Ahn, S. C., and A. R. Horenstein. 2013. Eigenvalue Ratio Test for the Number of Factors. Econometrica 81(3): 1203–1227.

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

Bai, J., and S. Ng. 2002. Determining the number of factors in approximate factor models. Econometrica 70(1): 191–221.

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

Bailey, N., G. Kapetanios, and M. H. Pesaran. 2016. Exponent of Cross-Sectional Dependence: Estimation and Inference. Journal of Applied Econometrics 31: 929–960.

———. 2019. Exponent of Cross-sectional Dependence for Residuals. Sankhya B 81: 46–102.

References II

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.

Bhattacharjee, A., J. Ditzen, and S. Holly. 2021. Spatial and Spatio-temporal Error Correction,Networks and Common Correlated Effects. BEMPS - Bozen Economics & Management Paper Series 76. URL https://ideas.repec.org/p/bzn/wpaper/bemps76.html.

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

Chen, P., Y. Karavias, and E. Tzavalis. 2021. Panel Unit Root Tests with Structural Breaks .

References III

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.

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

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

References IV

———. 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.

Ditzen, J., Y. Karavias, and J. Westerlund. 2021a. Testing for Multiple Structural Breaks in Panel Data .

———. 2021b. Testing and Estimating Structural Breaks in Time Series and Panel Data in Stata. arxiv [econ.EM] 2110.14550. URL https://arxiv.org/abs/2110.14550.

Eberhardt, M. 2011. XTCD: Stata module to investigate Variable/Residual Cross-Section Dependence. URL https://ideas.repec.org/c/boc/bocode/s457237.html.

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.

Gagliardini, P., E. Ossola, and O. Scaillet. 2019. A diagnostic criterion for approximate factor structure. Journal of Econometrics 212(2): 503–521. URL https://doi.org/10.1016/j.jeconom.2019.06.001.

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

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

References VI

Lopez, L., and S. Weber. 2017. Testing for Granger causality in panel data. The Stata Journal 17(4): 972–984.

Neal, T. 2014. Panel cointegration analysis with xtpedroni. The Stata Journal 14(3): 684–692.

Onatski, A. 2010. Determining the Number of Factors from Empirical Distribution of Eigenvalues. The Review ofEconomics and Statistics 92(4): 1004–1016.

Persyn, D., and J. Westerlund. 2008. Error-correction – based cointegration tests for. The Stata Journal 8(2): 232–241.

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

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

References VII

———. 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.

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

References VIII

Reese, S., and J. Ditzen. 2021. Determining number of common factors .

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.

Wursten, J. 2017. XTCDF: Stata module to perform Pesaran's CD-test for cross-sectional dependence in panel context. URL https://ideas.repec.org/c/boc/bocode/s458385.html.

Xiao, J., V. Sarafidis, Y. Karavias, and A. Juodis. 2021. Improved Tests for Granger Non-Causality in Panel Data (107180).