Estimating long-run coefficients and bootstrapping in large panels with cross-sectional dependence using xtdcce2

2019 London Stata User Group Meeting

Jan Ditzen

Heriot-Watt University, Edinburgh, UK
Center for Energy Economics Research and Policy (CEERP)

September 05, 2019

Introduction I

- Setting: Dynamic panel model with heterogeneous slopes and an unobserved common factor (f_t) and a heterogeneous factor loading (γ_i):

$$\begin{aligned} y_{i,t} &= \lambda_i y_{i,t-1} + \beta_i x_{i,t} + u_{i,t}, \qquad (1) \\ u_{i,t} &= \gamma_i' f_t + e_{i,t} \\ \beta_{MG} &= \frac{1}{N}\sum_{i=1}^{N}\beta_i, \quad \lambda_{MG} = \frac{1}{N}\sum_{i=1}^{N}\lambda_i \\ i = 1, ..., N &\; and \; t = 1, ..., T \end{aligned}$$

- Aim: consistent estimation of β_i and β_{MG} :
  - Large N, T = 1: Cross Section; $\hat{\beta} = \hat{\beta}_i, \ \forall \ i$
  - N=1 , Large T: Time Series; $\hat{\beta}_i$
  - Large N, Small T: Micro-Panel; $\hat{\beta} = \hat{\beta}_i, \ \forall \ i$
  - Large N, Large T: Panel Time Series; $\hat{\beta}_i$ and $\hat{\beta}_{MG}$

Introduction II

- If the common factors are left out, they become an omitted variable, leading to the omitted variable bias.
- Pesaran (2006); Chudik and Pesaran (2015) suggest to approximate the common factors with cross-sectional averages.
- Estimation method implemented by `xtdcce2`, on SSC since August 2016. Described in *The Stata Journal*, Vol 18, Number 3, (Ditzen, 2018) and in Ditzen (2019).
- `xtdcce2` includes test for cross-sectional dependence (Pesaran, 2015), `xtcd2`, and estimation of exponent of cross-sectional dependence (Bailey et al., 2016, 2019), `xtcse2`.

Introduction

- Estimation of most economic models requires heterogeneous coefficients. Examples: growth models (Lee et al., 1997), development economics (McNabb and LeMay-Boucher, 2014), productivity analysis (Eberhardt et al., 2012), consumption models (Shin et al., 1999) ,...
- Vast econometric literature on heterogeneous coefficients models (Zellner, 1962; Pesaran and Smith, 1995; Shin et al., 1999).
- Theoretical literature how to account for unobserved dependencies between cross-sectional units evolved (Pesaran, 2006; Chudik and Pesaran, 2015).

Dynamic Common Correlated Effects I

$$\begin{aligned} y_{i,t} &= \lambda_i y_{i,t-1} + \beta_i x_{i,t} + u_{i,t}, \\ u_{i,t} &= \gamma_i' f_t + e_{i,t} \end{aligned} \tag{2}$$

- Individual fixed effects (α_i) or deterministic time trends can be added, but are omitted in the remainder of the presentation.
- The heterogeneous coefficients are randomly distributed around a common mean, $\beta_i = \beta + v_i,\ v_i \sim IID(0, \Omega_v)$ and $\lambda_i = \lambda + \varsigma_i,\ \varsigma_i \sim IID(0, \Omega_\varsigma)$.
- f_t is an unobserved common factor and γ_i a heterogeneous factor loading.
- In a static model $\lambda_i = 0$, Pesaran (2006) shows that equation (2) can be consistently estimated by approximating the unobserved common factors with cross section averages $\bar{x}_t$ and $\bar{y}_t$ under strict exogeneity.

Dynamic Common Correlated Effects II

- In a dynamic model, the lagged dependent variable is not strictly exogenous and therefore the estimator becomes inconsistent. Chudik and Pesaran (2015) show that the estimator gains consistency if the floor of $p_T = \left[\sqrt[3]{T}\right]$ lags of the cross-sectional averages are added.
- Estimated Equation:

$$
\begin{aligned}
y_{i,t} &= \lambda_i y_{i,t-1} + \beta_i x_{i,t} + \sum_{l=0}^{p_T} \gamma_{i,l}' \bar{\mathbf{z}}_{t-l} + \epsilon_{i,t} \\
\bar{\mathbf{z}}_t &= (\bar{y}_t, \bar{x}_t)
\end{aligned}
$$

- The Mean Group Estimates are: $\hat{\pi}_{MG} = \frac{1}{N}\sum_{i=1}^{N} \hat{\pi}_i$ with $\hat{\pi}_i = (\hat{\lambda}_i, \hat{\beta}_i)$ and the asymptotic variance is

$$
\widehat{Var}(\hat{\pi}_{MG}) = \frac{1}{N(N-1)} \sum_{i=1}^{N} (\hat{\pi}_i - \hat{\pi}_{MG})(\hat{\pi}_i - \hat{\pi}_{MG})'
$$

Estimation of Long Run Coefficients

- A more general representation of eq (1) with further lags of the dependent and independent variable in the form of an ARDL(p_y, p_x) model is:

$$y_{i,t} = \sum_{l=1}^{p_y} \lambda_{l,i} y_{i,t-l} + \sum_{l=0}^{p_x} \beta_{l,i} x_{i,t-l} + u_{i,t}. \tag{3}$$

- where p_y and p_x is the lag length of y and x.
- The long run coefficient of β and the mean group coefficient are:

$$\theta_i = \frac{\sum_{l=0}^{p_x} \beta_{l,i}}{1 - \sum_{l=1}^{p_y} \lambda_{l,i}}, \quad \bar{\theta}_{MG} = \frac{1}{N} \sum_{i=1}^{N} \theta_i \tag{4}$$

- 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 (ECM).

CS-DL, CS-ARDL, CS-ECM

- CS-DL
  - Idea: directly estimate the long run coefficients, by adding differences of the explanatory variables and their lags.

$$y_{i,t} = \theta_i x_{i,t} + \sum_{l=0}^{p_x - 1} \delta_{i,l} \Delta x_{i,t-l} + \sum_{l=0}^{p_T} \gamma'_{i,l} \bar{\mathbf{z}}_{t-l} + e_{i,t}$$

- CS-ARDL and CS-ECM
  - Idea: first estimate short run coefficients, then calculate long run coefficients.

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

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

- For all estimators the mean group estimates are $\hat{\bar{\theta}}_{MG} = 1/N \sum_{i=1}^{N} \hat{\theta}_i$.
- The variance/covariance matrix for the mean group coefficients is the same as for the "normal" (D)CCE estimator.
- For the calculation of the variance/covariance matrix of the individual long run coefficients θ_i, the delta method is used. Delta Method

Next Steps...

1. Monte Carlo simulation
2. Bootstrapping in large panels
3. Description of `xtdcce2`
4. Examples

Monte Carlo Simulation

- Aims: Assess the bias of the point estimate and standard error of the long run coefficient.
- Simulation follows Chudik et al. (2016).
- The DGP is an ARDL(2,1) model:

$$
\begin{aligned}
y_{i,t} &= \alpha_i + \lambda_{1,i} y_{i,t-1} + \lambda_{2,i} y_{t-2} + \beta_{0,i} x_{i.t} + \beta_{1,i} x_{i,t-1} + u_{i,t} \\
u_{i,t} &= \gamma_u f_t + \epsilon_{i,t}
\end{aligned}
$$

- The coefficients are generated as:

$$
\begin{aligned}
\theta_i &\sim IIDN(1, \sigma_\theta^2) & \lambda_{1,i} &= (1 + \xi_{\lambda i}) \eta_{\lambda i} & \lambda_{2,i} &= -\xi_{\lambda i} \eta_{\lambda i} \\
\beta_{0,i} &= \xi_{\beta i} \eta_{\beta i}, & \beta_{1,i} &= (1 - \xi_{\beta i}) \, \eta_{\beta i} & \eta_{\lambda i} &= IIDU(0, \lambda_{max}) \\
\eta_{\beta i} &= \theta_i / (1 - \lambda_{i,1} - \lambda_{2,i}), & \xi_{\lambda i} &\sim IIDU(0.2, 0.3), & \xi_{\beta i} &\sim IIDU(0, 1)
\end{aligned}
$$

- $(\sigma_\theta^2, \lambda_{max})$ are varied between $(0.2, 0.6)$ and $(0.8, 0.8)$. Details

Monte Carlo Results

Bias and RMSE of $\hat{\theta}_{MG}$.

(N,T)	Bias of $\hat{\theta}_{MG}$ (x100)					RMSE of $\hat{\theta}_{MG}$ (x100)				
	40	50	100	150	200	40	50	100	150	200
CS-DL										
40	-21.57	-21.04	-19.52	-18.73	-18.26	23.50	22.48	20.10	19.04	18.46
50	-19.41	-19.15	-17.09	-16.64	-16.42	21.12	20.19	17.51	16.84	16.52
100	-20.04	-18.76	-17.40	-17.08	-16.93	20.39	19.02	17.25	16.81	16.61
150	-16.99	-16.41	-15.06	-14.72	-14.56	17.35	16.64	15.05	14.62	14.46
200	-20.73	-19.62	-18.20	-17.72	-17.37	21.04	19.80	18.24	17.70	17.31
CS-ARDL										
40	-2.63	-1.64	-1.94	-0.64	-0.48	192.31	13.65	8.01	5.58	4.80
50	-2.13	-186.07	-1.45	-0.75	-0.58	40.85	4049.97	6.53	5.47	4.36
100	-3.53	-0.43	-1.21	-0.94	-0.65	182.04	24.21	4.64	3.46	2.96
150	-4.93	-2.29	-1.31	-0.95	-0.59	34.46	7.20	3.69	2.69	2.48
200	-2.63	-2.29	-1.63	-1.11	-0.61	23.47	8.54	3.76	2.73	2.22

Monte Carlo results for $\hat{\theta}_{MG} = 1/N \sum_{i=1}^{N} \hat{\theta}_i$ with $p_T = [T^{1/3}]$, $\rho_f = 0$ and $(\sigma_\theta^2, \lambda_{max}) = (0.2, 0.6)$.

- CS-ARDL performs better in terms of bias, bias of both estimators decline with an increase in T.

Monte Carlo Results

Bias and RMSE of $SE(\hat{\theta}_{MG})$.

(N,T)	Bias of $SE(\hat{\theta}_{MG})$ (x100)					RMSE of $SE(\hat{\theta}_{MG})$ (x100)				
	40	50	100	150	200	40	50	100	150	200
CS-DL										
40	-53.83	-60.79	-71.47	-75.26	-77.61	12.06	13.54	15.85	16.68	17.19
50	-54.64	-60.85	-71.95	-75.80	-78.13	11.40	12.63	14.87	15.66	16.13
100	-67.21	-71.64	-79.56	-82.30	-83.81	12.91	13.75	15.26	15.79	16.07
150	-73.50	-76.87	-83.12	-85.09	-86.19	14.17	14.81	16.01	16.39	16.60
200	-76.23	-79.50	-85.22	-87.17	-88.23	14.77	15.40	16.51	16.88	17.09
CS-ARDL										
40	-46.24	-43.80	-65.46	-71.38	-74.85	187.57	10.94	14.57	15.84	16.59
50	-10.73	836.47	-66.20	-72.09	-75.85	36.00	4048.46	13.72	14.91	15.67
100	-42.71	-53.72	-75.66	-80.31	-82.62	180.31	24.47	14.53	15.41	15.85
150	-35.95	-67.29	-80.78	-84.14	-85.84	32.86	13.31	15.56	16.21	16.53
200	-39.30	-68.12	-82.47	-85.69	-87.39	21.64	14.47	15.98	16.60	16.93

Monte Carlo results for $SE(\hat{\theta}_{MG}) = \sqrt{1/N \sum_{i=1}^{N}(\hat{\theta}_i - \hat{\theta}_{MG})^2}$ with $p_T = [T^{1/3}]$, $\rho_f = 0$ and $(\sigma^2_\theta, \lambda_{max}) = (0.2, 0.6)$.

- Standard errors are downward biased, increase with number of time periods.

Bootstrapping in large panels

- Monte Carlo results show that standard errors are downward biased.
- Bootstrap often useful in small samples.
- No closed form solution for standard errors of individual long run coefficients. Delta method can fail.
- Bootstrap has to maintain the following properties of the DGP:
  - Dynamic nature of the model
  - Common factor structure
  - Error structure across time and cross-sectional units
  - N and T jointly to infinity
- Kapetanios (2008) and Westerlund et al. (2019) propose to re-sample cross-sectional units, but common factor structure changes.
- Gonçalves and Perron (2018) show that resampling over time is invalid in the presence of cross-sectional dependence.
- Praskova (2018) shows that if the common factors are known a wild bootstrap can be used.
- Idea: Wild Bootstrap

Wild Bootstrap

- Steps:
  1. Estimate Model, eg: $y_{i,t} = \lambda_i y_{i,t-1} + \beta_i x_{i,t} + \sum_{l=0}^{p_T} \gamma_{i,l}' \bar{\mathbf{z}}_{t-l} + \epsilon_{i,t}$
  2. Remove residual: $\tilde{y}_{i,t} = y_{i,t} - \hat{\epsilon}_{i,t}$
  3. Following Roodman et al. (2018) generate weights

$$k_{i,t}^{(b)} = \begin{cases} 1 & \text{with } p = 0.5 \\ -1 & \text{with } p = 0.5 \end{cases}$$

     and calculate $y_{i,t}^{(b)} = \tilde{y}_{i,t} + k_{i,t}^{(b)} \hat{\epsilon}_{i,t}$
  4. Estimate model and save coefficients.
  5. Repeat 3 - 4 B times and calculate standard errors or percentile confidence interval.

xtdcce2

General Syntax

Syntax:

xtdcce2 *depvar* [*indepvars*] [*varlist2* = *varlist_iv*] [*if*] crosssectional(*varlist_cr*) [, nocrosssectional pooled(*varlist_p*) cr_lags(#) ivreg2options(*string*) e_ivreg2 ivslow lr(*varlist_lr*) lr_options(*string*) pooledconstant noconstant reportconstant trend pooledtrend jackknife recursive exponent xtcse2options(string) nocd fullsample showindividual fast blockdiaguse nodimcheck useinvsym useqr noomitted showomitted]

More Details, Stored in e(), Bias Correction

For Bootstrap:

bootstrap_xtdcce2 [, reps(intger) seed(string) cfresiduals percentile showindividual]

xtdcce2

General Syntax

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

- `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)`

- Chudik et al. (2013) estimate the long run effect of public debt on output growth with the following equation:

$$\Delta y_{i,t} = c_i + \boldsymbol{\theta}_i' \mathbf{x}_{i,t} + \sum_{l=0}^{p_x - 1} \boldsymbol{\beta}_{i,l} \Delta \mathbf{x}_{i,t-l} + \gamma_{y,i} \Delta \bar{y}_t + \sum_{l=0}^{3} \boldsymbol{\gamma}_{x,i,l} \bar{\mathbf{x}}_{i,t-l} e_{i,t}$$

- where $y_{i,t}$ is the log of real GDP, $\mathbf{x}_{i,t} = (\Delta d_{i,t}, \pi_{i,t})'$, $d_{i,t}$ is log of debt to GDP ratio and π is the inflation rate.
- The results from Chudik et al. (2013, Table 18) with 1 lag of the explanatory variables ($p_x = 1$) in the form of an ARDL(1,1,1) and three lags of the cross sectional averages are estimated with:

```
xtdcce2 d.y dp d.gd d.(dp d.gd) , cr(d.y dp d.gd)
cr_lags(0 3 3) fullsample
```

xtdcce2

CS-DL Example

```
. xtdcce221 d.y dp d.gd d.(dp d.gd) ///
> , cr(d.y dp d.gd) cr_lags(0 3 3) fullsample
(Dynamic) Common Correlated Effects Estimator - Mean Group

Panel Variable (i): ccode                        Number of obs      =      1601
Time Variable (t): year                          Number of groups   =        40

Degrees of freedom per group:                    Obs per group (T)  =        40
 without cross-sectional averages    = 35.025
 with cross-sectional averages       = 26.025
Number of                                        F(560, 1041)       =      0.90
 cross-sectional lags                0 to 3      Prob > F           =      0.93
 variables in mean group regression = 160        R-squared          =      0.67
 variables partialled out            = 400       R-squared (MG)     =      0.40
                                                 Root MSE           =      0.03
                                                 CD Statistic       =      1.11
                                                    p-value         =    0.2667
------------------------------------------------------------------------------
         D.y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 Mean Group: |
          dp | -.0889339   .0256445    -3.47   0.001     -.1391961   -.0386717
        D.gd | -.0865123      .0143    -6.05   0.000     -.1145398   -.0584849
        D.dp |  .0053284   .0413629     0.13   0.897     -.0757413    .0863981
       D2.gd |  .0068065   .0148306     0.46   0.646      -.022261     .035874
------------------------------------------------------------------------------
Mean Group Variables: dp D.gd D.dp D2.gd
Cross Sectional Averaged Variables: D.y(0) dp(3) D.gd(3)
Heterogenous constant partialled out.
```

- The long run coefficients are $\hat{\theta}_{\pi,MG} = -0.0889$ and $\hat{\theta}_{\Delta d,MG} = -0.0865$.

xtdcce2

CS-DL Example

```
. bootstrap_xtdcce2 , reps(500)
(running on xtdcce2 sample)
Wild-Bootstrap replications ( 500 ) using residuals
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100
..................................................   150
..................................................   200
..................................................   250
..................................................   300
..................................................   350
..................................................   400
..................................................   450
..................................................   500
```

	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf.	Interval]
Mean Group:						
dp	-.0889339	.1532996	-0.58	0.562	-.3893956	.2115279
D.gd	-.0865123	.0600872	-1.44	0.150	-.204281	.0312563
D.dp	.0053284	.2093117	0.03	0.980	-.404915	.4155719
D2.gd	.0068065	.094243	0.07	0.942	-.1779065	.1915195

- The long run coefficients are not significant any longer.

xtdcce2

CS-ARDL

- Assume an ARDL(1,2) and $p_T = (p_{\bar{y}}, p_{\bar{x}}) = (2, 2)$ such as:

$$y_{i,t} = \lambda_i y_{i,t-1} + \beta_{0,i} x_{i,t} + \beta_{1,i} x_{i,t-1} + \beta_{2,i} x_{i,t-2} + \sum_{l=0}^{2} \gamma_{y,i,l} \bar{y}_t + \sum_{l=0}^{2} \gamma_{x,i,l} \bar{x}_{t-l} + e_{i,t}$$

- The short run coefficients are estimated and then the long run coefficients are calculated as:

$$\hat{\theta}_{CS-ARDL,i} = \frac{\hat{\beta}_{0,i} + \hat{\beta}_{1,i} + \hat{\beta}_{2,i}}{1 - \hat{\lambda}_i}$$

- Using `xtdcce2` the command line is:
```
xtdcce2 y , lr(L.y x L.x L2.x) lr_options(ardl) cr(y x) cr_lags(2)
```
- `lr()` defines the long run variables.
- `xtdcce2` automatically detects the variables and their lags if time series operators are used. Alternatively variables can be enclosed in parenthesis, for example `lr(L.y (x lx l2x))`, with `lx = L.x` and `l2x = L2.x`.

xtdcce2

CS-ARDL Example - ARDL(1,1,1) from Chudik et al. (2013, Table 17).

```
. xtdcce221 d.y , lr(L.d.y L.dp dp L.d.gd d.gd) ///
> lr_options(ardl) cr(d.y dp d.gd) cr_lags(3) ///
> fullsample
(Dynamic) Common Correlated Effects Estimator -  (CS-ARDL)

Panel Variable (i): ccode                       Number of obs      =       1599
Time Variable (t): year                         Number of groups   =         40

Degrees of freedom per group:                   Obs per group (T) =         40
 without cross-sectional averages      = 33.975
 with cross-sectional averages         = 21.975
Number of                                       F(720, 879)        =       0.79
 cross-sectional lags                  = 3      Prob > F           =       1.00
 variables in mean group regression = 200       R-squared          =       0.61
 variables partialled out              = 520    R-squared (MG)     =       0.44
                                                Root MSE           =       0.03
                                                CD Statistic       =       0.57
                                                   p-value         =     0.5690
```

D.y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Short Run Est.						
Mean Group:						
LD.y	.0475614	.0393514	1.21	0.227	-.0295659	.1246888
dp	-.1036029	.0402888	-2.57	0.010	-.1825675	-.0246383
D.gd	-.0745686	.0122305	-6.10	0.000	-.0985398	-.0505973
L.dp	-.0199465	.0462873	-0.43	0.667	-.1106679	.0707749
LD.gd	-.0132482	.0156115	-0.85	0.396	-.0438463	.0173498
Long Run Est.						
Mean Group:						
lr_dp	-.1639748	.0378594	-4.33	0.000	-.2381778	-.0897718
lr_gd	-.0873991	.0164432	-5.32	0.000	-.1196271	-.0551711
lr_y	-.9524386	.0393514	-24.20	0.000	-1.029566	-.8753112

```
Mean Group Variables:
Cross Sectional Averaged Variables: D.y dp D.gd
Long Run Variables:  lr_dp lr_gd lr_y
Cointegration variable(s): lr_y
Heterogenous constant partialled out.
```

xtdcce2

CS-ARDL Example - ARDL(1,1,1) from Chudik et al. (2013, Table 17), bootstrapped.

```
. bootstrap_xtdcce2 , reps(500) percentile
(running on xtdcce2 sample)

Wild-Bootstrap replications ( 500 ) using residuals
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100
..................................................   150
..................................................   200
..................................................   250
..................................................   300
..................................................   350
..................................................   400
..................................................   450
..................................................   500
```

	Observed Coef.	Observed Std. Err.	z	P>|z|	percentile t [95% Conf.	Interval]
Short Run Est.						
Mean Group:						
LD.y	.0475614	.0393514	1.21	0.227	.7006659	1.241012
dp	-.1036029	.0402888	-2.57	0.010	-.2056815	-.1088788
D.gd	-.0745686	.0122305	-6.10	0.000	-.0594676	-.0451482
L.dp	-.0199465	.0462873	-0.43	0.667	-.0205935	.0986097
LD.gd	-.0132482	.0156115	-0.85	0.396	-.0282485	.0010115
Long Run Est.						
Mean Group:						
lr_dp	-.1639748	.0378594	-4.33	0.000	-.2314161	-.142728
lr_gd	-.0873991	.0164432	-5.32	0.000	-.0905856	-.0492529
lr_y	-.9524386	.0393514	-24.20	0.000	-.2993341	.241012

xtdcce2

CS-ARDL Example - ARDL(3,3,3) from Chudik et al. (2013, Table 17).

```
. xtdcce221 d.y , cr_lags(3) fullsample ///
> lr(L(1/3).(d.y) (L(0/3).dp) (L(0/3).d.gd) ) ///
> lr_options(ardl) cr(d.y dp d.gd)
(Dynamic) Common Correlated Effects Estimator -  (CS-ARDL)

Panel Variable (i): ccode                        Number of obs      =       1562
Time Variable (t): year                          Number of groups   =         40

Degrees of freedom per group:                    Obs per group (T) =         39
 without cross-sectional averages    = 27.05
 with cross-sectional averages       = 15.05
Number of                                        F(960, 602)        =       0.96
 cross-sectional lags                = 3         Prob > F           =       0.71
 variables in mean group regression = 440        R-squared          =       0.39
 variables partialled out            = 520       R-squared (MG)     =       0.51
                                                 Root MSE           =       0.02
                                                 CD Statistic       =      -0.51
                                                    p-value         =     0.6108
------------------------------------------------------------------------------
         D.y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Short Run Est. |
-------------+----------------------------------------------------------------
  Mean Group:|
        LD.y |   .0123776   .0349374    0.35   0.723    -.0560984    .0808536
       L2D.y |  -.1395721   .0948493   -1.47   0.141    -.3254733     .046329
       L3D.y |  -.0829106   .1072972   -0.77   0.440    -.2932092    .1273881
          dp |  -.0707066   .0503045   -1.41   0.160    -.1693015    .0278883
        D.gd |  -.0853072   .0137595   -6.20   0.000    -.1122754   -.0583391
        L.dp |  -.0312738   .0513445   -0.61   0.542    -.1319071    .0693595
       L2.dp |    .098219    .101743    0.97   0.334    -.1011937    .2976317
       L3.dp |  -.0424672   .0581718   -0.73   0.465    -.1564818    .0715474
       LD.gd |  -.0270313   .0204755   -1.32   0.187    -.0671624    .0130999
      L2D.gd |  -.0114101    .012726   -0.90   0.370    -.0363525    .0135324
      L3D.gd |   .0283559   .0177672    1.60   0.110    -.0064671    .0631789
-------------+----------------------------------------------------------------
Long Run Est. |
-------------+----------------------------------------------------------------
  Mean Group:|
       lr_dp |  -.0795232   .0587003   -1.35   0.176    -.1945738    .0355274
       lr_gd |  -.1198351   .0402246   -2.98   0.003    -.1986738   -.0409964
        lr_y |  -1.210105   .2006012   -6.03   0.000    -1.603276   -.8169339
------------------------------------------------------------------------------
Mean Group Variables:
Cross Sectional Averaged Variables: D.y dp D.gd
Long Run Variables:  lr_dp lr_gd lr_y
Cointegration variable(s): lr_y
Heterogenous constant partialled out.
```

xtdcce2

CS-ARDL Example - ARDL(3,3,3) from Chudik et al. (2013, Table 17), bootstrapped.

```
. bootstrap_xtdcce2 , reps(500)
(running on xtdcce2 sample)
Wild-Bootstrap replications ( 500 ) using residuals
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................     50
..................................................    100
..................................................    150
..................................................    200
..................................................    250
..................................................    300
..................................................    350
..................................................    400
..................................................    450
..................................................    500
```

	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf.	Interval]
Short Run Est.						
Mean Group:						
LD.y	.0123776	.0523665	0.24	0.813	-.0902588	.115014
L2D.y	-.1395721	.0619368	-2.25	0.024	-.2609661	-.0181782
L3D.y	-.0829106	.0383575	-2.16	0.031	-.1580899	-.0077312
dp	-.0707066	.1551891	-0.46	0.649	-.3748717	.2334585
D.gd	-.0853072	.0408892	-2.09	0.037	-.1654486	-.0051658
L.dp	-.0312738	.2973936	-0.11	0.916	-.6141546	.551607
L2.dp	.098219	.3611995	0.27	0.786	-.6097191	.8061571
L3.dp	-.0424672	.2644735	-0.16	0.872	-.5608257	.4758914
LD.gd	-.0270313	.07608	-0.36	0.722	-.1761453	.1220827
L2D.gd	-.0114101	.1131884	-0.10	0.920	-.2332553	.2104352
L3D.gd	.0283559	.0859342	0.33	0.741	-.1400721	.1967839
Long Run Est.						
Mean Group:						
lr_dp	-.0795232	.0790329	-1.01	0.314	-.2344249	.0753785
lr_gd	-.1198351	.0324949	-3.69	0.000	-.183524	-.0561463
lr_y	-1.210105	.1392513	-8.69	0.000	-1.483033	-.9371776

Conclusion

`xtdcce2`...

- introduced a routine to estimate a panel model with heterogeneous slopes and dependence across cross-sectional untis by using the dynamic common correlated effects estimator.
- supports estimation of long run coefficients using three different models, using the
  - CS-DL estimator - direct estimation of the long run coefficients
  - CS-ARDL estimator - calculation of long run coefficients out of short run coefficients
  - an ECM approach
- is available on SSC (current version 2.01).
- standard errors and confidence intervals can be bootstrapped.
- includes estimation of cross-sectional exponent.
- Further developments:
  - Two-step ECM.
  - Speed improvements and fitting it for "big" data.
  - Compare bootstrapped standard errors and delta method standard errors.

The Delta Method

back

- Allows calculation of an approximate probability distribution for a matrix function $a(\beta)$ based on a random vector with a known variance.
- Assume $\beta_i \to_p \beta$ and $\sqrt{n}(\beta_i - \beta) \to_d N(0, \sigma)$ and first derivate of $a(\beta)$:

$$A(\beta) \equiv \frac{\partial a(\beta)}{\partial \beta'}$$

- then the distribution of the function $a()$ is

$$\sqrt{n}\left[a(\beta_i) - a(\beta)\right] \to_d N\left(0, A(\beta)\Sigma A(\beta)'\right).$$

The Delta Method I

back

- Assume an ARDL(2,1) model with the following long run coefficients:

$$
\begin{aligned}
y_{i,t} &= \alpha_i + \lambda_{1,i} y_{i,t-1} + \lambda_{2,i} y_{i,t-2} + \beta_{0,i} x_{i,t} + \beta_{1,i} x_{i,t-1} + e_{i,t} \\
\phi_i &= -\left(1 - \lambda_{1,i} - \lambda_{2,i}\right) \\
\theta_{1,i} &= \frac{\beta_{0,i} + \beta_{1,i}}{1 - \lambda_{1,i} - \lambda_{2,i}}
\end{aligned}
$$

- Stack the short run coefficients into $\pi_i = (\lambda_{1,i}, \lambda_{2,i}, \beta_{0,i}, \beta_{1,i})$
- The vector function $a(\pi_i)$ maps the short run coefficients into a vector of the short run and long run coefficients:

  $a(\pi_i) = (\lambda_{1,i}, \lambda_{2,i}, \beta_{0,i}, \beta_{1,i}, \phi_i, \theta_{1,i})$, where $\phi_i = -1 + \lambda_{1,i} + \lambda_{2,i}$ and $\theta_{1,i} = \frac{\beta_{0,i}+\beta_{1,i}}{1-\lambda_{1,i}-\lambda_{2,i}}$.

The Delta Method II

back

- The covariance matrix is:

$$\Sigma_i = \begin{pmatrix} Var(\lambda_{1,i}) & Cov(\lambda_{1,i}, \lambda_{2,i}) & Cov(\lambda_{1,i}, \beta_{0,i}) & Cov(\lambda_{1,i}, \beta_{1,i}) \\ & \ddots & & \\ & & \ddots & \\ & & & Var(\beta_{1,i}) \end{pmatrix}$$

- The first derivative of $a(\pi_i)$ is:

The Delta Method III

back

$$A(\pi_i) = \begin{pmatrix}
\frac{\partial \lambda_{1,i}}{\partial \lambda_{1,i}} & \frac{\partial \lambda_{1,i}}{\partial \lambda_{2,1}} & \frac{\partial \lambda_{1,i}}{\partial \beta_{0,i}} & \frac{\partial \lambda_{1,i}}{\partial \beta_{1,i}} \\
\frac{\partial \lambda_{2,i}}{\partial \lambda_{1,i}} & \frac{\partial \lambda_{2,i}}{\partial \lambda_{2,i}} & \frac{\partial \lambda_{2,i}}{\partial \beta_{0,i}} & \frac{\partial \lambda_{2,i}}{\partial \beta_{1,i}} \\
\frac{\partial \beta_{0,i}}{\partial \lambda_{1,i}} & \frac{\partial \beta_{0,i}}{\partial \lambda_{2,i}} & \frac{\partial \beta_{0,i}}{\partial \beta_{0,i}} & \frac{\partial \beta_{0,i}}{\partial \beta_{1,i}} \\
\frac{\partial \beta_{1,i}}{\partial \lambda_{1,i}} & \frac{\partial \beta_{1,i}}{\partial \lambda_{2,i}} & \frac{\partial \beta_{1,i}}{\partial \beta_{0,i}} & \frac{\partial \beta_{1,i}}{\partial \beta_{1,i}} \\
\frac{\partial \phi_{i}}{\partial \lambda_{1,i}} & \frac{\partial \phi_{i}}{\partial \lambda_{2,i}} & \frac{\partial \phi_{i}}{\partial \beta_{0,i}} & \frac{\partial \phi_{i}}{\partial \beta_{1,i}} \\
\frac{\partial \theta_{1,i}}{\partial \lambda_{1,i}} & \frac{\partial \theta_{1,i}}{\partial \lambda_{2,i}} & \frac{\partial \theta_{1,i}}{\partial \beta_{0,i}} & \frac{\partial \theta_{1,i}}{\partial \beta_{1,i}}
\end{pmatrix}$$

The Delta Method IV

▸ back

- with

$$\frac{\partial \phi_i}{\partial \lambda_{1,i}} = \frac{\partial \phi_i}{\partial \lambda_{2,i}} = 1$$
$$\frac{\partial \theta_{1,i}}{\partial \beta_{0,i}} = \frac{\partial \theta_{1,i}}{\partial \beta_{1,i}} = \frac{1}{1 - \lambda_{1,i} - \lambda_{2,i}}$$
$$\frac{\partial \theta_{1,i}}{\partial \lambda_{1,i}} = \frac{\partial \theta_{1,i}}{\partial \lambda_{2,i}} = \frac{\beta_{0,i} + \beta_{1,i}}{(1 - \lambda_{1,i} - \lambda_{2,i})^2}$$

- Then $A(\pi_i)$ becomes:

The Delta Method V

back

$$A(\pi_i) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ \frac{\beta_{0,i}+\beta_{1,i}}{\left(1-\lambda_{1,i}-\lambda_{2,i}\right)^2} & \frac{\beta_{0,i}+\beta_{1,i}}{\left(1-\lambda_{1,i}-\lambda_{2,i}\right)^2} & \frac{1}{1-\lambda_{1,i}-\lambda_{2,i}} & \frac{1}{1-\lambda_{1,i}-\lambda_{2,i}} \end{pmatrix}$$

- Then the covariance matrix including the long run coefficients is

$$\Sigma_i^{lr} = A(\pi_i)\Sigma_i A(\pi_i)'$$

Monte Carlo Setup ▸ back

As in Chudik et al. (2016) the data generating processes are the following:[1]

$$y_{i,t} = \alpha_i + \lambda_{1,i} y_{i,t-1} + \lambda_{2,i} y_{t-2} + \beta_{0,i} x_{i.t} + \beta_{1,i} x_{i,t-1} + u_{i,t}$$
$$u_{i,t} = \gamma_i' f_t + \epsilon_{i,t}$$
$$x_{i,t} = c_{xi} + \alpha_{xi} y_{i,t-1} + \gamma_{xi} f_t + v_{xi,t}$$

$y_{i,t}$ is the dependent variable and $x_{i,t}$ the only independent variable. For a matter of ease, it is assumed that only one explanatory variable exists. The common factors are calculated as below:

$$f_t = \rho_f f_{t-1} + \varsigma_{ft}, \varsigma_{ft} \sim IIDN(0, 1 - \rho_f^2)$$
$$v_{xi,t} = \rho_{xi} v_{xi,t-1} + \varsigma_{xi,t}, \varsigma_{xi,t} \sim IIDN(0, \sigma^2_{vxi})$$
$$\rho_{xi} \sim IIDU(0, 0.95)$$
$$\rho_f = 0 \text{ if serially uncorrelated factors, or if correlated } \rho_f = 0.6$$
$$\sigma^2_{vxi} = \sigma^2_{vi} = \left(\beta_{0i} \sqrt{1 - [E(\rho_{xi})]^2} \right)^2$$

[1] This paper focuses on the baseline cases with heterogenous slopes and stationary factors.

Monte Carlo Setup back I

Fixed Effects The cross-section specific fixed effects are generated as:

$$c_{yi} \sim IIDN(1,1)$$
$$c_{xi} = c_{yi} + \varsigma_{c_x i}, \varsigma_{c_x i} \sim IIDN(0,1).$$

Dependence between $x_{i,t}, g_{i,t}$ and c_{yi} is introduced by adding c_{yi} to the equations for c_{xi} and c_{gi}.

Coefficients First the long run coefficient θ is drawn and then the short run coefficients are backed out.

$$\theta_i \sim IIDN(1, \sigma_\theta^2)$$
$$\lambda_{1,i} = (1 + \xi_{\lambda i})\eta_{\lambda i}, \qquad \lambda_{2,i} = -\xi_{\lambda i}\eta_{\lambda i}$$
$$\beta_{0,i} = \xi_{\beta i}\eta_{\beta i}, \qquad \beta_{1,i} = (1 - \xi_{\beta i})\,\eta_{\beta i}$$
$$\eta_{\lambda i} = IIDU(0, \lambda_{max}), \qquad \eta_{\beta i} = \theta_i / \left(1 - \lambda_{i,1} - \lambda_{2,i}\right)$$
$$\xi_{\lambda i} \sim IIDU(0.2, 0.3), \qquad \xi_{\beta i} \sim IIDU(0,1)$$

Monte Carlo Setup back II

Factor Loadings

$$\gamma_i = \gamma + \eta_{i\gamma}, \qquad \eta_{i\gamma} \sim IIDN(0, \sigma^2_\gamma)$$

$$\gamma_{xi} = \gamma_x + \eta_{i\gamma x}, \qquad \eta_{i\gamma x} \sim IIDN(0, \sigma^2_{\gamma x})$$

$$\sigma^2_\gamma = \sigma^2_{\gamma x} = 0.2^2$$

$$\gamma = \sqrt{b_\gamma}, \qquad b_\gamma = \frac{1}{m} - \sigma^2_\gamma$$

$$\gamma_x = \sqrt{b_x}, \qquad b_x = \frac{2}{m(m+1)} - \frac{2}{m+1}\sigma^2_{\gamma x}$$

where m is the number of unobserved factors. In comparison to Chudik and Pesaran (2015) it is restricted to 1.

Monte Carlo Setup back III

Error Term The errors are generated such that heteroskedasticity, autocorrelation and weakly cross-sectional dependence is allowed.

$$
\begin{aligned}
\epsilon_{i,t} &= \rho_{\epsilon i}\epsilon_{i,t-1} + \zeta_{i,t} \\
\zeta_t &= (\zeta_{1,t}, \zeta_{2,t}, ..., \zeta_{N,t}) = \alpha_{CSD} S_{\epsilon_t} + e_{\epsilon t} \\
\Rightarrow \zeta_t &= (1 - \alpha_{CSD} S_\epsilon)^{-1} e_{\epsilon t} \\
e_{\epsilon t} &\sim IIDN(0, \frac{1}{2}\sigma_i^2 \left(1 - \rho_{\epsilon i}^2\right)), \text{ with } \sigma_i^2 \sim \chi^2(2) \\
\rho_{\epsilon i} &\sim IIDU(0, 0.8) \\
S_\epsilon &= \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & & 0 \\ 0 & \frac{1}{2} & 0 & \ddots & & \vdots \\ 0 & 0 & \ddots & \ddots & \frac{1}{2} & 0 \\ \vdots & & & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & \dots & 0 & 1 & 0 \end{pmatrix}
\end{aligned}
$$

xtdcce2

pmg-Options

- `lr`(*varlist*) defines the variables in the long run relationship.
- `xtdcce2` estimates internally

$$\Delta y_{i,t} = \phi_i y_{i,t-1} + \gamma_i x_{i,t-1} - \beta_i \Delta x_{i,t} + \sum_{l=0}^{p_T} \gamma_{i,l} \bar{\mathbf{z}}_{i,t} + u_{i,t} \qquad (5)$$

while `xtpmg` (with common factors) is based on:

$$\Delta y_{i,t} = \phi_i \left[y_{i,t-1} - \theta_i x_{i,t-1} \right] - \beta_i \Delta x_{i,t} + \sum_{l=0}^{p_T} \gamma_{i,l} \bar{\mathbf{z}}_{i,t} + u_{i,t}$$

- where $\theta_i = -\frac{\gamma_i}{\phi_i}$. θ_i is calculated and the variances calculated using the Delta method.
- `lr_option`(*string*)
  - `nodivide`, coefficients are not divided by the error correction speed of adjustment vector (i.e. estimate (5)).
  - `xtpmgnames`, coefficients names in `e(b_p_mg)` and `e(V_p_mg)` match the name convention from `xtpmg`.

xtdcce2

Test for cross sectional dependence

- `xtdcce2` package includes the `xtcd2` command, which tests for cross sectional dependence (Pesaran, 2015).
- Under the null hypothesis, the error terms are weakly cross sectional dependent.

$$H_0 : E(u_{i,t} u_{j,t}) = 0, \forall\ t \text{ and } i \neq j.$$

$$CD = \sqrt{\frac{2T}{N(N-1)}} \left(\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij} \right)$$

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

- Under the null the CD test statistic is asymptotically $CD \sim N(0,1)$.

Saved values ‣ back

Scalars

e(N)	number of observations	e(N_g)	number of groups
e(T)	number of time periods	e(K_mg)	number of regressors
e(N_partial)	number of variables partialled out	e(N_omitted)	number of omitted variables
e(N_pooled)	number of pooled variables	e(mss)	model sum of square
e(rss)	residual sum of squares	e(F)	F statistic
e(ll)	log-likelihood (only IV)	e(rmse)	root mean squared error
e(df_m)	model degrees of freedom	e(df_r)	residual degree of freedom
e(r2)	R-squared	e(r2_a)	R-squared adjusted
e(cd)	CD test statistic	e(cdp)	p-value of CD test statistic
e(cr_lags)	number of lags of cross sectional averages		

Scalars (unbalanced panel)

e(Tmin)	minimum time	e(Tmax)	maximum time
e(Tbar)	average time		

Macros

e(tvar)	name of time variable	e(idvar)	name of unit variable
e(depvar)	name of dependent variable	e(indepvar)	name of independent variables
e(omitted)	name of omitted variables	e(lr)	long run variables
e(pooled)	name of pooled variables	e(cmd)	command line
e(cmdline)	command line including options	e(version)	xtdcce2 version, if xtdcce2, version used
e(insts)	instruments (exogenous) variables	e(instd)	instrumented (endogenous) variables
e(alpha)	estimated of exponent of cross-section dependence	e(alphaSE)	estimated standard error of exponent of cross-section dependence

Matrices

e(b)	coefficient vector (mean group or individual)	e(V)	variance–covariance matrix (mean group or individual)
e(bi)	coefficient vector (individual and pooled)	e(Vi)	variance–covariance matrix (individual and pooled)

Functions

e(sample)	marks estimation sample

Options

back

- pooled(*varlist*) specifies homogeneous coefficients. For these variables the estimated coefficients are constrained to be equal across all units ($\beta_i = \beta \ \forall \ i$). Variable may occur in *indepvars*. Variables in exogenous_vars(), endogenous_vars() and lr() may be pooled as well.
- crosssectional(*varlist*) defines the variables which are included in z_t and added as cross sectional averages ($\bar{z}_{t-l}$) to the equation. Variables in crosssectional() may be included in pooled(), exogenous_vars(), endogenous_vars() and lr(). Variables in crosssectional() are partialled out, the coefficients not estimated and reported. crosssectional(_all) adds adds all variables as cross sectional averages. No cross sectional averages are added if crosssectional(_none) is used, which is equivalent to nocrosssectional. crosssectional() is a required option but can be substituted by nocrosssectional.

Options I

back

- cr_lags(#) specifies the number of lags of the cross sectional averages. If not defined but crosssectional() contains *varlist*, then only contemporaneous cross sectional averages are added, but no lags. cr_lags(0) is equivalent to. The number of lags can be different for different variables, following the order defined in cr().
- nocrosssectional prevents adding cross sectional averages. Results will be equivalent to the Pesaran and Smith (1995) Mean Group estimator, or if lr(*varlist*) specified to the Shin et al. (1999) Pooled Mean Group estimator.
- xtdcce2 supports instrumental variable regression using ivreg2. The IV specific options are:
  - ivreg2options passes further options on to ivreg2. See ivreg2, options for more information.
  - fulliv posts all available results from ivreg2 in e() with prefix *ivreg2_*.

Options II

▸ back

- noisily shows the output of wrapped ivreg2 regression command.
- ivslow For the calculation of standard errors for pooled coefficients an auxiliary regressions is performed. In case of an IV regression, xtdcce2 runs a simple IV regression for the auxiliary regressions. this is faster. If option is used ivslow, then xtdcce2 calls ivreg2 for the auxiliary regression. This is advisable as soon as ivreg2 specific options are used.

- xtdcce2 is able to estimate long run coefficients. Three models are supported, an error correction model, the CS-DL and CS-ARDL method. No options for the CS-DL method are necessary.
  - lr(*varlist*): Variables to be included in the long-run cointegration vector. The first variable(s) is/are the error-correction speed of adjustment term. The default is to use the pmg model. In this case each estimated coefficient is divided by the negative of the long-run cointegration vector (the first variable). If the option ardl is used, then the long run coefficients are estimated as the sum over the coefficients relating to a variable, divided by the sum of the coefficients of the dependent variable.

Options III

back

- `lr_options(`*string*`)` Options for the long run coefficients. Options may be:
  - `ardl` estimates the CS-ARDL estimator.
  - `nodivide`, coefficients are not divided by the error correction speed of adjustment vector.
  - `xtpmgnames`, coefficients names in `e(b_p_mg)` and `e(V_p_mg)` match the name convention from `xtpmg`.

- `noconstant` suppress constant term.
- `pooledconstant` restricts the constant to be the same across all groups ($\beta_{0,i} = \beta_0, \ \forall i$).
- `reportconstant` reports the constant. If not specified the constant is treated as a part of the cross sectional averages.
- `trend` adds a linear unit specific trend. May not be combined with `pooledtrend`.
- `pooledtrend` a linear common trend is added. May not be combined with `trend`.

Options IV

▸ back

- `jackknife` applies the jackknife bias correction for small sample time series bias. May not be combined with `recursive`.
- `recursive` applies recursive mean adjustment method to correct for small sample time series bias. May not be combined with `jackknife`. `exponent` uses `xtcse2` to estimate the exponent of the cross-sectional dependence of the residuals. A value above 0.5 indicates cross-sectional dependence.
- `nocd` suppresses calculation of CD test statistic. `blockdiaguse` uses mata blockdiag rather than an alternative algorithm. `mata blockdiag` is slower, but might produce more stable results.
- `showindividual` reports unit individual estimates in output.
- `fast` omit calculation of unit specific standard errors.

Options V

back

- `fullsample` uses entire sample available for calculation of cross sectional averages. Any observations which are lost due to lags will be included calculating the cross sectional averages (but are not included in the estimation itself).
- **xtdcce2** checks for collinearity in three different ways. It checks if matrix of the cross-sectional averages is of full rank. After partialling out the cross-sectional averages, it checks if the entire model across all cross-sectional units exhibits multicollinearity. The final check is on a cross-sectional level. The outcome of the checks influence which method is used to invert matrices. If a check fails **xtdcce2** posts a warning message. The default is **cholinv** and **invsym** if a matrix is of rank-deficient. The following options are available to alter the behaviour of xtdcce2 with respect to matrices of not full rank:
    - `useqr` calculates the generalized inverse via QR decomposition. This was the default for rank-deficient matrices for **xtdcce2** pre version 1.35.
    - `useinvsym` calculates the generalized inverse via **mata invsym**.

Options VI

back

- `showomitted` displays a cross-sectional unit - variable breakdown of omitted coefficients.
- `nomitted` suppress checks for collinearity.

xtdcce2

Small Sample Time Series Bias Corrections

"half panel" jackknife

$$\hat{\pi}^J_{MG} = 2\hat{\pi}_{MG} - \frac{1}{2}\left(\hat{\pi}^a_{MG} + \hat{\pi}^b_{MG}\right)$$

- where $\hat{\pi}^a_{MG}$ is the mean group estimate of the first half ($t = 1, ..., \frac{T}{2}$) of the panel and $\hat{\pi}^b_{MG}$ of the second half ($t = \frac{T}{2} + 1, ..., T$) of the panel.

Recursive mean adjustment

$$\tilde{w}_{i,t} = w_{i,t} - \frac{1}{t-1}\sum_{s=1}^{t-1} w_{i,s} \text{ with } \quad w_{i,t} = (y_{i,t}, X_{i,t}).$$

- Partial mean from all variables, except the constant, removed.
- Partial mean is lagged by one period to prevent it from being influenced by contemporaneous observations.

back

References I

Bailey, N., S. Holly, and M. H. Pesaran (2016): "A Two-Stage Approach to Spatio-Temporal Analysis with Strong and Weak Cross-Sectional Dependence," Journal of Applied Econometrics, 31, 249–280.

Bailey, N., G. Kapetanios, and M. H. Pesaran (2019): "Exponent of Cross-Sectional Dependence for Residuals," Sankhya B. The Indian Journal of Statistics, forthcomin.

Chudik, A., K. Mohaddes, M. H. Pesaran, and M. Raissi (2013): "Debt, Inflation and Growth: Robust Estimation of Long-Run Effects in Dynamic Panel Data Models," .

——— (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, 393–420.

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

References II

——— (2019): "Estimating long run effects in models with cross-sectional dependence using xtdcce2," CEERP Working Paper Series, No. 7.

Eberhardt, M., C. Helmers, and H. Strauss (2012): "Do Spillovers Matter When Estimating Private Returns to R&D?" Review of Economics and Statistics, 95, 436 – 448.

Gonçalves, S. and B. Perron (2018): "Bootstrapping factor models with cross sectional dependence," .

Kapetanios, G. (2008): "A bootstrap procedure for panel data sets with many cross-sectional units," Econometrics Journal, 11, 377–395.

Lee, K., M. H. Pesaran, and R. Smith (1997): "Growth and Convergence in a Multi-Country Empirical Stochastic Solow Model," Journal of Applied Economics, 12, 357–392.

McNabb, K. and P. LeMay-Boucher (2014): "Tax Structures, Economic Growth and Development," .

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

References III

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

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

Praskova, Z. (2018): "Bootstrap Change Point Testing for," in Statistics and Simulation, ed. by J. P. C. Kleijnen, Springer International Publishing, chap. 4, 543–67.

Roodman, D., M. Ø. Nielsen, M. D. Webb, and J. G. Mackinnon (2018): "Fast and Wild: Bootstrap Inference in Stata using boottest," .

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

Westerlund, J., Y. Petrova, and M. Norkute (2019): "CCE in fixed-T panels," Journal of Applied Econometrics, 1–16.

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