Fakultät für Wirtschaftswissenschaften
Facoltà di Economia
Faculty of Economics and Management

BEMPS –
Bozen Economics & Management Paper Series

NO 81/ 2021

Estimating long run effects and the exponent of cross-sectional dependence: an update to *xtdcce2*

Jan Ditzen

Estimating long run effects and the exponent of cross-sectional dependence: an update to `xtdcce2`

Jan Ditzen
Free University of Bozen-Bolzano
Bozen, Italy
jan.ditzen@unibz.it

February 25, 2021

Abstract

In this paper I describe several updates to `xtdcce2` (Ditzen, 2018). First I explain how to estimate long run effects in models with cross-sectional dependence. Three methods to estimate the long run effects are reviewed and their implementation into Stata using `xtdcce2` discussed. Two of the estimation methods build on Chudik et al. (2016); the CS-DL and the CS-ARDL estimator. As a third alternative I review an error correction model in the presence of cross-sectional dependence. Second, I explain how to estimate the exponent of cross-sectional dependence using `xtcse2` following Bailey et al. (2016, 2019).
Keywords: xtdcce2, xtcse2, xtcd2, parameter heterogeneity, dynamic panels, cross section dependence, common correlated effects, pooled mean-group estimator, mean-group estimator, error correction model, ardl, long run coefficients

1 Introduction

Estimation of long run relationships is important in empirical applications of economic and in particular macroeconomic models. Long run relationships describe how one or more variables react to changes in the steady state. An example would be the relationships between macroeconomic variables such as GDP and inflation. Another would be the effects of investments, exchange rates, education or technological progress on economic growth.

With pure time series data the autoregressive distributed lag (ARDL) model is widely used to estimate long run relationships. ARDL models estimate the short run coefficients and then back out the long run coefficients. They were implemented by the community-contributed `ardl` command in Stata (Kripfganz and Schneider, 2018). A related model is the error correction model. The model consists of two terms, one term captures the short run deviations from equilibrium and the other the long run movements (Engle and Granger, 1987). Both models can be applied to panel data (Pesaran and Smith, 1995; Pesaran et al., 1999). Panel data models add an extra layer of dimension compared to time series models. Time series models cover one panel unit and slope heterogeneity across units is not an issue. Panel models include many panel units and long and/or short run coefficients can vary across those. A popular method is the pooled mean group (PMG) estimator, which assume heterogeneous short run and homogeneous long run effects in a panel error correction model (Pesaran et al., 1999). Blackburne and Frank (2007) implemented this method into Stata with the community-contributed command `xtpmg`.

The estimation of unit specific coefficients require datasets with a large number of observations across time periods and cross-sectional units. Such datasets often inhibit cross-sectional dependence. It implies that cross-sectional units depend on each other, for instance by sharing a common factor. If this dependence is ignored, estimation results can be biased and inconsistent. Therefore the extent of cross-sectional dependence needs to be understood and the estimation method chosen accordingly. The literature proposes two methods to identify cross-sectional dependence. The first is to estimate the strength of the dependence (Bailey et al., 2016), the other is to test for cross-sectional dependence (Pesaran, 2015). The Stata community-contributed command `xtcd2` (Ditzen, 2018) tests for cross-sectional dependence. This paper introduces the first method, the estimation of the exponent of cross-sectional dependence using `xtcse2`.

After establishing the existence of strong cross-sectional dependence, it can be approximated or controlled for by either principle components (Bai and Ng, 2002; Bai, 2009) or by adding cross-sectional averages (Pesaran, 2006), for a comparison see Westerlund and Urbain (2015). Due to its simplicity the approach using cross-sectional averages is very popular and started its own literature, Everaert and De Groote (2016); Chudik et al. (2011); Chudik and Pesaran (2015a) provide overviews. The estimation method, called the common correlated effects (CCE) estimator, applies to static- (Pesaran, 2006) and dynamic panel models (Chudik and Pesaran, 2015b; Karabiyik et al., 2017), pooled- (Karabiyik et al., 2020) and mean group estimators (Chudik and Pesaran, 2019). The idea of the estimator is to add cross-sectional averages of the independent and dependent variables which approximate the cross-sectional dependence. This estimator was implemented into Stata in the static version by the community contributed command `xtmg` (Eberhardt, 2012) and in the dynamic version by `xtdcce2` (Ditzen, 2018).

Neither of the commands were able to estimate long run relationships directly. In this paper I introduce an extended version of `xtdcce2` which allows the estimation of the long run coefficients.[1] The estimation method is based on Chudik et al. (2016) and an augmented error correction model.

[1]The estimation of long run coefficients is possible with `xtdcce2` version 1.33 and later. This paper refers to version 2.0 or later. See the author's webpage www.jan.ditzen.net for updates.

The remainder of the paper is structured as follows. The next section introduces the panel model, cross-sectional dependence and the CCE estimator. Then three different methods to estimate the long run coefficients are discussed, first from a theoretical perspective and then from an applied perspective. Examples on how to estimate the models using `xtdcce2` are given. The paper closes with a conclusion.

2 Panel Model and Common Correlated Effects Estimators

For this section, assume a dynamic ARDL(1,1) panel model with heterogeneous coefficients in the form of:[2]

$$y_{i,t} = \mu_i + \lambda_i y_{i,t-1} + \beta_{0,i} x_{i,t} + \beta_{1,i} x_{i,t-1} + u_{i,t} \tag{1}$$

$$u_{i,t} = \sum_{l=1}^{m} \varrho_{y,i,l} f_{t,l} + e_{i,t} \tag{2}$$

$$x_{i,t} = \sum_{l=1}^{m} \varrho_{x,i,l} f_{t,l} + \xi_{i,t} \tag{3}$$

with $i = 1, ..., N$ and $t = 1, ..., T_i$,

where $y_{i,t}$ is the dependent variable and $x_{i,t}$ an observed independent variable, which includes m unobserved common factors $f_{t,l}$. The estimation of the long run effect of x on y is the main point of interest. $e_{i,t}$ is a cross-section unit-specific IID error term. The factor loadings $\varrho_{x,i,l}$ and $\varrho_{y,i,l}$ are heterogeneous across units and μ_i is a unit-specific fixed effect. The heterogeneous coefficients are randomly distributed around a common mean, such that $\beta_i = \beta + v_i$, and $\lambda_i = \lambda + a_i$, where v_i and a_i are random deviations with mean zero, independent of the error term and the common factors. λ_i lies strictly inside the unit circle to ensure a non explosive series.

2.1 Estimating and testing for cross-sectional dependence

The strength of the factors can be measured by a constant $0 \leq \alpha \leq 1$, the so-called exponent of cross-sectional dependence. Depending on its limiting behaviour Chudik et al. (2011) propose four types of cross sectional dependence, weak ($\alpha = 0$), semi-weak ($0 < \alpha < 0.5$), semi-strong ($0.5 \leq \alpha < 1$) and strong ($\alpha = 1$) cross-sectional dependence. (Semi-)Weak cross-sectional dependence can be thought of as the following: even if the number of cross-sectional units increases to infinity, the sum of the effect of the common factors remains constant. In the case of strong cross-sectional dependence, the sum of the effect

[2]A more in-depth discussion of the model and the assumptions is provided in Chudik et al. (2011); Chudik and Pesaran (2015a); Ditzen (2018).

of the common factors becomes stronger with an increase in the number of cross-sectional units.

Bailey et al. (2016) propose a method for the estimation of the exponent of a variable under semi-strong and strong cross-sectional dependence. They derive a bias-adjusted estimator for α and its standard error based on auxiliary regressions using principle components and cross-sectional averages. In the case of estimating the exponent of cross-sectional dependence in residuals Bailey et al. (2019) propose to use significant pair-wise correlations of the residuals after multiple testing. A closed form solution for standard errors is not available and confidence intervals are constructed using a simple bootstrap. The community-contributed Stata command `xtcse2` estimates the exponent of a variable and residual.

Another possibility to determine the strength of cross-sectional dependence is to test for (semi-)weak cross-sectional dependence (Pesaran, 2015). Thus the so-called CD test indirectly tests for $\alpha < 0.5$. The test statistic is the sum across all pair-wise correlations and under the null asymptotically standard normal distributed. For a further theoretical discussion of the CD test see Pesaran (2015). The CD test is implemented in Stata by the community-contributed command `xtcd2` (Ditzen, 2018).

2.2 Common Correlated Effects Estimator

Given the model in equation (1), leaving the factor structure unaccounted for leads to an omitted variable bias and OLS becomes inconsistent (Everaert and De Groote, 2016). Pesaran (2006) and Chudik and Pesaran (2015b) propose an estimator to estimate equation (1) consistently by approximating the common factors with cross sectional averages. In a dynamic model the floor of $\sqrt[3]{T}$ lags of the cross-sectional averages are added. The estimated equation becomes:

$$y_{i,t} = \mu_i + \lambda_i y_{i,t-1} + \beta_{0,i} x_{i,t} + \beta_{1,i} x_{i,t-1} + \sum_{l=0}^{p_T} \boldsymbol{\gamma}'_{i,l} \bar{\mathbf{z}}_{t-l} + e_{i,t} \tag{4}$$

where $\bar{\mathbf{z}}_t = (\bar{y}_t, \bar{x}_t)' = (1/N \sum_{i=1}^N y_{i,t}, 1/N \sum_{i=1}^N x_{i,t})'$ are the cross sectional averages of the dependent and independent variables. $\boldsymbol{\gamma}_{i,l} = (\gamma_{y,i,l}, \gamma_{x,i,l})'$ are the estimated coefficients of the cross-sectional averages and generally treated as nuisance parameters. The model can be estimated by either a mean group estimator (Pesaran and Smith, 1995; Pesaran, 2006; Chudik and Pesaran, 2019) or by a pooled estimator (Pesaran, 2006; Karabiyik et al., 2020).[3] This estimator is known as the common correlated effects mean group estimator (CCE-MG) or common correlated effects pooled estimator (CCE-P). The CCE-MG estimator is implemented in Stata by `xtmg` (Eberhardt, 2012) and both estimators by `xtdcce2` (Ditzen, 2018).

[3] The assumption of heterogeneous slopes can be tested, see Pesaran and Yamagata (2008); Blomquist and Westerlund (2013) and in Stata Bersvendsen and Ditzen (2020).

3 Estimating Long Run Relationships

Dynamic models allow the estimation of long run relationships. They measure the effect of an explanatory variable on the steady state value of the dependent variable. Following the notation from equation (1) and assuming that model is in its steady state with $y_t^* = y_{t-1}^* = y^*$ and $x_t^* = x_{t-1}^* = x^*$, the long run effect of variable x is defined as:

$$\theta_i = \frac{\beta_{0,i} + \beta_{1,i}}{1 - \lambda_i}. \tag{5}$$

The long run effect in equation (5) can be estimated by an ARDL, DL (distributed lag) and ECM approach. All three can be augmented by cross-sectional averages to approximate cross-sectional dependence.

3.1 CS-ECM

The cross-sectionally augmented error correction approach (CS-ECM) follows on the lines of Lee et al. (1997) and Pesaran et al. (1999). Equation (4) is transformed into an error correction model (ECM):[4]

$$\Delta y_{i,t} = \mu_i - \phi_i \left[y_{i,t-1} - \theta_{1,i} x_{i,t} \right] - \beta_{1,i} \Delta x_{i,t} + \sum_{l=0}^{p_T} \boldsymbol{\gamma}_{i,l}' \bar{\mathbf{z}}_{t-l} + e_{i,t} \tag{6}$$

with Δ the first difference operator, θ_i defined as in (5) and

$$\phi_i = (1 - \lambda_i)$$

is the error-correction speed of adjustment parameter and $[y_{i,t-1} - \theta_{1,i} x_{i,t}]$ is the error correction term. A long run relationship exists if $\phi_i \neq 0$ (Pesaran et al., 1999). $\beta_{0,i}$ captures the immediate or short run effect of $x_{i,t}$ on $y_{i,t}$. The long run or equilibrium effect is captured by θ_i. The long run effect measures how the equilibrium changes and ϕ_i represents how fast the adjustment occurs.

In the case without cross-sectional dependence and homogeneous long run coefficients ($\theta_i = \theta \ \forall \ i$), the model can be estimated by the pooled mean group (PMG) estimator (Pesaran et al., 1999).

3.2 CS-ARDL

An alternative to the CS-ECM is the cross-sectionally augmented ARDL (CS-ARDL) approach (Chudik et al., 2016). First the short run coefficients are estimated and then the long run coefficients are calculated. The advantage of this approach is that a full set of estimates for the long and the short run

[4]The ECM can be expressed in terms of regressors in time $t-1$ instead of time t. In this case Equation (6) would be: $\Delta y_{i,t} = \mu_i - \phi_i \left[y_{i,t-1} - \theta_{1,i} x_{i,t-1} \right] + \beta_{0,i} \Delta x_{i,t} + \sum_{l=0}^{p_T} \boldsymbol{\gamma}_{i,l}' \bar{\mathbf{z}}_{t-l} + e_{i,t}$. This is only a different parametrisation and long run estimates will remain the same. For a more detailed discussion see the helpfile of the community contributed `ardl` command (Kripfganz and Schneider, 2018).

coefficients is obtained. An ARDL model can be rewritten as an ECM and therefore the long run estimates from the CS-ECM and CS-ARDL approaches are numerically equivalent.

Equation (1) can be generalised to an $ARDL(p_y, p_x)$ model:

$$y_{i,t} = \mu_i + \sum_{l=1}^{p_y} \lambda_{l,i} y_{i,t-l} + \sum_{l=0}^{p_x} \beta_{l,i} x_{i,t-l} + \sum_{l=0}^{p} \boldsymbol{\gamma}'_{i,l} \bar{\mathbf{z}}_{t-l} + e_{i,t}. \tag{7}$$

The individual long run coefficients are calculated as:

$$\hat{\theta}_{CS-ARDL,i} = \frac{\sum_{l=0}^{p_x} \hat{\beta}_{l,i}}{1 - \sum_{l=1}^{p_y} \hat{\lambda}_{l,i}}. \tag{8}$$

The coefficients can be directly estimated by the mean group or pooled estimator. The mean group variance estimator can be applied (Chudik et al., 2016), if the mean group estimator is used.

3.3 CS-DL

Under the assumption that λ_i lies in the unit circle, the general representation of an ARDL(p_y, p_x) model can be written in distributed lag form:[5]

$$y_{i,t} = \mu_i + \theta_{1,i} x_{i,t} + \delta_i(L) \Delta x_{i.t} + \tilde{u}_{i,t} \tag{9}$$

Chudik et al. (2016) show that Equation (9) can be directly estimated by the common correlated effects estimator, named the cross-sectionally augmented DL (CS-DL) approach. The regression is augmented with the differences of the explanatory variables (x), their lags and the cross-sectional averages. Following Pesaran (2006) the estimation is consistent even if the errors are serially correlated.

For a general ARDL(p_y, p_x) model with added cross-sectional averages to take out strong cross-sectional dependence, the CS-DL estimator is based on the following equation:

$$\begin{aligned} y_{i,t} =& \mu_i + \theta_{1,i} x_{i,t} + \sum_{l=0}^{p_x - 1} \delta_{i,l} \Delta x_{i,t-l} \\ &+ \sum_{l=0}^{p_{\bar{y}}} \gamma_{y,i,l} \bar{y}_{t-l} + \sum_{l=0}^{p_{\bar{x}}} \gamma_{x,i,l} \bar{x}_{t-l} + e_{i,t}, \end{aligned} \tag{10}$$

where $\bar{y}_{t-l}$ and $\bar{x}_{t-l}$ are the cross sectional averages and $p_{\bar{x}} = \lfloor T^{1/3} \rfloor$ and $p_{\bar{y}} = 0$.

[5]The other parameters are defined as: $\delta_i(L) = -\sum_{l=0}^{\infty} \left[\lambda_i^{l+1} (1 - \lambda_i)^{-1} \beta_{1,i} \right] L^l, \theta_{0,i} = (1 - \lambda_i L)^{-1} \mu_i$, $\tilde{u}_{i,t} = (1 - \lambda_i L)^{-1} u_{i,t}$ and L is the lag operator.

4 Updates to `xtdcce2` command

4.1 Syntax

The updated syntax is described below. The options in the second pair of brackets are new additions or updated to the version explained in Ditzen (2018):

`xtdcce2` *depvar* [*indepvars*] [*varlist2* = *varlist_iv*] [*if*]
crosssectional(varlist_cr) [, `nocrosssectional` `pooled(`*varlist_p*`)`
`cr_lags(`#`)` `ivreg2options(`*options1*`)` `e_ivreg2` `ivslow`
`pooledconstant` `noconstant` `reportconstant` `trend` `pooledtrend`
`jackknife` `recursive` `nocd` `fullsample` `showindividual`] [`fast`
`lr(`*varlist_lr*`)` `lr_options(`*options2*`)` `exponent`
`xtcse2options(`*options3*`)` `blockdiaguse` `nodimcheck` `useinvsym`
`useqr` `noomitted` `showomitted`]

4.2 New and updated options

In the following the updated or new options are explained. For a full explanation see Ditzen (2018, 2019) and the helpfile for `xtdcce2`.

`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`.

`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 the equivalent. 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 Pesaran et al. (1999) Pooled Mean Group estimator.

`lr(`*varlist_lr*`)`: 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 ECM approach. In this case each estimated coefficient is divided by the negative of the long-run cointegration coefficient (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.

`lr_options(`*options2*`)` passes options for the long run estimation. *options2* 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)` and `e(V)` match the name convention from `xtpmg`.

`exponent` uses `xtcse2` to estimate the exponent of the cross-sectional dependence of the residuals. A value above 0.5 indicates strong cross-sectional dependence.

`xtcse2options(`*options3*`)` passes options to `xtcse2`.

`fast` omit calculation of unit specific standard errors.

`useqr` calculates the generalized inverse via QR decomposition. The default is `mata cholinv`. QR decomposition was the default for rank-deficient matrices for xtdcce2 pre version 1.35.

`useinvsym` calculates the generalized inverse via `mata invsym`.

`showomitted` displays a cross-sectional unit - variable breakdown of omitted coefficients.

`nomitted` suppress checks for collinearity.

4.2.1 New stored values

The new version stores the following two additional results:

Matrices

`e(alpha)`	estimated of exponent of cross-section dependence	`e(alphaSE)`	estimated standard error of exponent of cross-section dependence

5 The `xtcse2` command

5.1 Syntax

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

5.2 Options

`pca(`*integer*`)` sets the number of principle components for the calculation of *cn*. Default is to use the first 4 components.

`standardize` standardizes variables.

`nocenter` do not center variables (i.e. cross-sectional mean is zero).

`nocd` suppresses test for weak cross-sectional dependence using xtcd2.

`residual` estimates the exponent of cross-sectional dependence in residuals, following Bailey et al. (2019).

reps(*integer*) number of repetitions for bootstrap for calculation of standard error and confidence interval for exponent in residuals. Default is 100.

size(*real*) size of the test. Default is 10% (0.1).

tuning(*real*) tuning parameter for estimation of the exponent in residuals. Default is 0.5.

lags(*integer*) number of lags (or training period) for calculation of recursive residuals when estimating the exponent after a regression with weakly exogenous regressors.

5.3 Stored Values

Matrices

r(alpha)	matrix of estimated α	r(alphaSE)	matrix with standard errors of α
r(N_g)	matrix with number of cross-sectional units	r(T)	matrix with number of time periods
r(CD)	matrix with values of CD test statistic (if requested)	r(CDp)	matrix of p values of CD test statistic (if requested)
r(alphas)	matrix with estimated $\tilde{\alpha}$, $\hat{\alpha}$ and α		

6 Empirical Examples

6.1 Estimating and testing for cross-sectional dependence

Blackburne and Frank (2007) explain the use of xtpmg by estimating the long-run consumption function from Lee et al. (1997) and Pesaran et al. (1999):[6]

$$c_{i,t} = \theta_{0t} + \theta_{1t} y_{i,t} + \theta_{2t} \pi_{i,t} + \mu_i + \epsilon_{i,t}, \tag{11}$$

where $c_{i,t}$ log of consumption per capita, $y_{i,t}$ is log of real per capita income and $\pi_{i,t}$ is the inflation rate.

Before estimating the model it is necessary to evaluate if the variables inhibit cross-sectional dependence. xtcse2 is used to estimate the exponent of and test for cross-sectional dependence for the variables $c_{i,t}$ (c), $y_{i,t}$ (y) and $\pi_{i,t}$ (pi):

[6] The following example uses the *jasa2* dataset, which is available with the xtpmg command.

```
. xtcse2 c pi y
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]	
c	1.004833	.0544669	.8980796	1.111586
pi	1.004841	1.763292	-2.451148	4.460831
y	1.004833	.0466978	.913307	1.096359

```
0.5 <= alpha < 1 implies strong cross-sectional dependence.
Pesaran (2015) test for weak cross-sectional dependence.
H0: errors are weakly cross-sectional dependent.
```

variable	CD	p-value	N_g	T
c	89.656	0.000	24	33
pi	96.751	0.000	24	33
y	89.659	0.000	24	33

The CD test rejects the null of weak cross-sectional dependence for all variables and the estimated exponent of cross-sectional dependence is well above 0.5. This is evidence that an estimation method taking cross-sectional dependence into account is necessary. All remaining examples are dynamic models. Following Chudik and Pesaran (2015b) the contemporaneous levels of the dependent and independent variables and the floor of $T^{1/3}$ lags of the cross-sectional averages will be added to approximate strong cross-sectional dependence. After each regression the residuals are tested for strong cross-sectional dependence using the CD test and the exponent of cross-sectional dependence estimated.

6.2 CS-ECM

The ECM representation of the equation (11) reads:

$$\Delta c_{i,t} = \mu_i - \phi_i(c_{i,t-1} - \theta_{1,i}y_{i,t} - \theta_{2,i}\pi_{i,t}) - \beta_{1,i}\Delta y_{i,t} - \beta_{2,i}\Delta\pi_{i,t} + \epsilon_{i,t}. \quad (12)$$

Blackburne and Frank (2007) and Ditzen (2018) estimate a pooled mean group model without and with contemporaneous cross-sectional averages using `xtpmg` and `xtdcce2`. This exercises focuses on the CS-ECM model and all coefficients are assumed to be heterogeneous. Following Chudik and Pesaran (2015b) $p = \lfloor T^{1/3} \rfloor = \lfloor 29^{1/3} \rfloor = 3$ lags of the cross-sectional averages are added to estimated equation (12):[7]

[7] $\lfloor . \rfloor$ denotes the floor of a number.

```
. xtdcce2 d.c d.y d.pi  if year >= 1962, ///
> lr(L.c y pi)  cr(_all) cr_lags(3) exponent
(Dynamic) Common Correlated Effects Estimator - Mean Group (CS-ECM)

Panel Variable (i): id                                Number of obs     =        695
Time Variable (t): year                               Number of groups  =         24

Degrees of freedom per group:                         Obs per group:
 without cross-sectional avg. min   = 22                           min =         28
                              max   = 23                           avg =         29
 with cross-sectional avg.    min   = 10                           max =         29
                              max   = 11
Number of                                             F(432, 263)       =       2.90
 cross-sectional lags               = 3               Prob > F          =       0.00
 variables in mean group regression = 120             R-squared         =       0.17
 variables partialled out           = 312             R-squared (MG)    =       0.83
                                                      Root MSE          =       0.01
                                                      CD Statistic      =       0.27
                                                         p-value        =     0.7899
```

D.c	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Short Run Est.						
Mean Group:						
D.y	.0088767	.0511634	0.17	0.862	-.0914017	.109155
D.pi	.0146379	.0412939	0.35	0.723	-.0662966	.0955725
Adjust. Term						
Mean Group:						
L.c	-.6112082	.056361	-10.84	0.000	-.7216738	-.5007426
Long Run Est.						
Mean Group:						
pi	-.5976237	.275682	-2.17	0.030	-1.13795	-.057297
y	.7872628	.0995928	7.90	0.000	.5920646	.982461

```
Mean Group Variables: D.y D.pi pi y
Cross Sectional Averaged Variables: pi y c
Long Run Variables:  pi y
Cointegration variable(s): L.c
Heterogenous constant partialled out.
Estimation of Cross-Sectional Exponent (alpha)
```

variable	alpha	Std. Err.	[95% Conf.	Interval]
residuals	.5844011	.0243676	.5366414	.6321607

```
0.5 <= alpha < 1 implies strong cross sectional dependence.
SE and CI bootstrapped with 100 repetitions.
```

The mean group estimate of the partial adjustment coefficients is $\hat{\phi} = -0.611$ (`L.c`), the long run effect of income on consumption is $\hat{\theta}_1 = 0.787$ (`y`) and of inflation on consumption is $\hat{\theta}_2 = -0.598$ (`pi`). The results imply that 61.1% of the disequilibrium is adjusted every period. An increase in income increases consumption in the long run, while an increase in prices hampers consumption in the long run.

There are some notable differences between `xtpmg` and `xtdcce2`. `xtpmg` calculates the long run coefficients using maximum likelihood. `xtdcce2` internally estimates (leaving out any cross-sectional averages):

$$\Delta c_{i,t} = \mu_i - \phi_i c_{i,t-1} + \kappa_{1,i} y_{i,t} + \kappa_{2,i} \pi_{i,t} - \beta_{1,i} \Delta y_{i,t} - \beta_{2,i} \Delta \pi_{i,t} + \epsilon_{i,t}, \quad (13)$$

using OLS with $\kappa_{1,i} = -\theta_{1,i}\phi_i$ and $\kappa_{2,i} = -\theta_{2,i}\phi_i$. The long run coefficients and the mean group coefficients are estimated in three steps and the variances are calculated using the Delta method. First, the cross-section specific coefficients $\mu_i, \phi_i, \kappa_{1,i}, \kappa_{2,i}, \beta_{1,i}$ and $\beta_{2,i}$ are estimated. Then the cross-section specific long run coefficients are calculated. Lastly, the mean group coefficients are calculated as the unweighed average over the unit specific long run coefficients. As an example, the average long run unit specific coefficient for $\hat{\theta}_{1,i}$ is derived as $\hat{\theta}_{1,i} = -\hat{\kappa}_{1,i}/\hat{\phi}_i$. Then the mean group estimator is: $\hat{\hat{\theta}}_1 = 1/N \sum_{i=1}^N \hat{\theta}_{1,i} = 1/N \sum_{i=1}^N (-\hat{\kappa}_{1,i}/\hat{\phi}_i)$.

The PMG estimator assumes homogeneous long run and heterogeneous short run coefficients. `xtdcce2` is build to handle both coefficients to be heterogeneous and/or homogeneous. If the long run coefficients are homogeneous but the short run coefficients heterogeneous, then the mean group estimate of the error speed of correction term is used to calculate the long run coefficient. They then become $\theta_1^p = -\kappa_1^p/\phi_{MG}$.

The option `exponent` is used to calculate the exponent of the cross-sectional dependence using `xtcse2`. Standard errors and confidence intervals can be obtained by a simple bootstrap in which the cross-sectional units are drawn with replacement. `xtdcce2` automatically runs a bootstrap with 100 repetitions. Further options to `xtcse2` can be passed by the option `xtcse2option(options)`. In the example above, the p-value of the CD test is 0.79 and the test cannot reject the null hypothesis of (semi-)weak cross-sectional dependence. Bailey et al. (2019, p. S92) state that the estimated exponent of cross-sectional dependence should be close to 0.5 if the residuals are weakly cross-sectional dependent. The estimated exponent of cross-sectional dependence is 0.588 and close to the threshold of 0.5.

6.3 CS-ARDL

The ECM in equation (12) can be transferred into an ARDL(1,1,1) model:

$$c_{i,t} = \mu_i + \lambda_i c_{i,t-1} + \beta_{10,i} y_{i,t} + \beta_{11,i} y_{i,t-1} + \beta_{20,i} \pi_{i,t} + \beta_{20,i} \pi_{i,t-1} + \epsilon_{i,t}. \quad (14)$$

Using `xtdcce2`, all short run variables are added to the `lr` option and the ARDL routine is invoked by using `lr_options(ardl)`:[8]

[8]There is no need to specify the long run variables separately as `xtdcce2` automatically detects the common base of variables if time series operators are used. If lags are created as variables via `gen lx = L.x`, then the variables with the same base which form a long run coefficient need to be enclosed in parenthesis, for example `lr((y ly) (x lx))`.

```
. xtdcce2 c if year >= 1962, ///
> lr(L.c L(0/1).y pi L.pi) lr_options(ardl) ///
> cr(_all) cr_lags(3)
(Dynamic) Common Correlated Effects Estimator -  (CS-ARDL)

Panel Variable (i): id                          Number of obs     =        695
Time Variable (t): year                         Number of groups  =         24

Degrees of freedom per group:                   Obs per group:
 without cross-sectional avg. min    = 22                     min =         28
                              max    = 23                     avg =         29
 with cross-sectional avg.    min    = 10                     max =         29
                              max    = 11
Number of                                       F(432, 263)       =       3.27
 cross-sectional lags                = 3        Prob > F          =       0.00
 variables in mean group regression = 120       R-squared         =       0.16
 variables partialled out           = 312       R-squared (MG)    =       1.00
                                                Root MSE          =       0.01
                                                CD Statistic      =       0.27
                                                   p-value        =     0.7899
--------------------------------------------------------------------------------
               c|     Coef.   Std. Err.       z    P>|z|     [95% Conf. Interval]
----------------+---------------------------------------------------------------
 Short Run Est. |
----------------+---------------------------------------------------------------
   Mean Group:  |
             L.c|   .3887918    .056361     6.90   0.000      .2783262   .4992574
              pi| -.1113299    .0760736    -1.46   0.143     -.2604314   .0377716
               y|   .486285    .0598417     8.13   0.000      .3689975   .6035726
             L.y| -.0088767    .0511634    -0.17   0.862      -.109155   .0914017
            L.pi| -.0146379    .0412939    -0.35   0.723     -.0955725   .0662966
----------------+---------------------------------------------------------------
 Adjust. Term   |
----------------+---------------------------------------------------------------
   Mean Group:  |
            lr_c| -.6112082     .056361   -10.84   0.000     -.7216738  -.5007426
----------------+---------------------------------------------------------------
 Long Run Est.  |
----------------+---------------------------------------------------------------
   Mean Group:  |
           lr_pi| -.5976237     .275682    -2.17   0.030      -1.13795   -.057297
            lr_y|  .7872628    .0995928     7.90   0.000      .5920646    .982461
--------------------------------------------------------------------------------
Mean Group Variables: L.c pi y L.y L.pi lr_pi lr_y
Cross Sectional Averaged Variables: pi y c
Long Run Variables:  lr_pi lr_y
Adjustment variable(s): lr_c (L.c)
Heterogenous constant partialled out.
```

As expected, the regression results are the same as above for the CS-ECM model. In the output the long run coefficient estimates have the prefix `lr_` and the adjustment parameter (ϕ) is displayed in a separate section. If the long run coefficients are pooled, `xtdcce2` uses the delta method to calculate the variance/covariance matrix of the long run coefficients.

For the remaining examples, the results in Chudik et al. (2013) will be replicated. The authors estimate the long run effect of public debt on output growth with the following equation:

$$\Delta y_{i,t} = \mu_i + \sum_{l=1}^{p} \lambda_{i,l} \Delta y_{i,t-l} + \sum_{l=0}^{p} \boldsymbol{\beta}_{i,l}' \mathbf{x}_{i,t-l} + \sum_{l=0}^{3} \boldsymbol{\gamma}_{i,l}' \bar{\mathbf{z}}_{t-l} + e_{i,t} \qquad (15)$$

where $y_{i,t}$ is the logarithm of real GDP and $\Delta y_{i,t}$ its growth rate. $\mathbf{x}_{i,t} = (\Delta d_{i,t}, \pi_{i,t})'$, $d_{i,t}$ is log of debt to GDP ratio and π is the log of the inflation rate and p the number of lags. The cross-sectional averages are $\bar{\mathbf{z}}_t = (\bar{\mathbf{x}}_t, \bar{\Delta} y_t)'$. The variables in the example dataset are `dy` for $\Delta y_{i,t}$, `dgd` for $\Delta d_{i,t}$ and `dp` for the inflation rate $\pi_{i,t}$.

The degree of cross-sectional dependence is checked with:

```
. xtcse2 y dp gd, standardize
output omitted
```

All variables are strongly cross-sectional dependent with $\hat{\alpha}_y = 1$, $\hat{\alpha}_{dp} = 0.94$ and $\hat{\alpha}_{dgd} = 0.92$. The CD-test statistic yields the same conclusion, all variables contain strong cross-sectional dependence.

Next, we can turn to estimate the ARDL model. As before 3 lags of the cross sectional averages are added to take out any strong cross-sectional dependence. To replicate the results of the ARDL(1,1,1) model from Chudik et al. (2013, Table 17), the first lag of the dependent and the base and the first lag of the dependent variables are added:

```
. xtdcce2 dy , lr(L.dy L.dp dp L.dgd dgd) ///
> lr_options(ardl) cr(dy dp dgd) 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
```

dy	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Short Run Est.						
Mean Group:						
L.dy	.0475615	.0393516	1.21	0.227	-.0295662	.1246891
dp	-.1036032	.0402887	-2.57	0.010	-.1825676	-.0246389
dgd	-.0745686	.0122305	-6.10	0.000	-.0985399	-.0505974
L.dp	-.019946	.0462871	-0.43	0.667	-.1106671	.070775
L.dgd	-.0132481	.0156115	-0.85	0.396	-.0438461	.0173498
Adjust. Term						
Mean Group:						
lr_dy	-.9524385	.0393516	-24.20	0.000	-1.029566	-.8753109
Long Run Est.						
Mean Group:						
lr_dgd	-.0873993	.0164431	-5.32	0.000	-.1196272	-.0551713
lr_dp	-.1639757	.0378599	-4.33	0.000	-.2381797	-.0897717

```
Mean Group Variables: L.dy dp dgd L.dp L.dgd lr_dgd lr_dp
Cross Sectional Averaged Variables: dy dp dgd
Long Run Variables:  lr_dgd lr_dp
Adjustment variable(s): lr_dy (L.dy)
Heterogenous constant partialled out.
```

The long run coefficients for the logarithm of debt to GDP ratio and inflation are both significant and negative. A decrease in the debt burden and inflation will increase GDP growth. A one percent decrease of the debt to GDP growth is associated with an increase of the GDP growth rate of 0.16%. A one percent decrease in the inflation rate leads to an increase of the GDP growth rate of 0.087%. The partial adjustment to the long run equilibrium appears to be very quick, 95% of the gap is closed within one year.

For the ARDL(3,3,3) the 3 lags of the explanatory variables and the dependent variable are added. To improve readability, the different bases are enclosed into parenthesis:

```
. xtdcce2 dy , cr_lags(3) fullsample ///
> lr(L(1/3).(dy) (L(0/3).dp) (L(0/3).dgd) ) ///
> lr_options(ardl) cr(dy dp dgd)
(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
-------------------------------------------------------------------------------
             dy|     Coef.   Std. Err.       z    P>|z|     [95% Conf. Interval]
---------------+---------------------------------------------------------------
 Short Run Est.|
---------------+---------------------------------------------------------------
   Mean Group: |
           L.dy|  .0123738   .0349377     0.35   0.723     -.0561029   .0808506
          L2.dy| -.1395645   .0948427    -1.47   0.141     -.3254529   .0463238
          L3.dy|  -.082903   .1072901    -0.77   0.440     -.2931877   .1273817
             dp|  -.070708   .0503039    -1.41   0.160     -.1693018   .0278858
            dgd|  -.085307   .0137595    -6.20   0.000     -.1122752  -.0583388
           L.dp| -.0312712   .0513435    -0.61   0.542     -.1319025   .0693601
          L2.dp|  .0982105   .1017365     0.97   0.334     -.1011893   .2976103
          L3.dp| -.0424631   .0581692    -0.73   0.465     -.1564726   .0715464
          L.dgd| -.0270311   .0204753    -1.32   0.187     -.0671619   .0130997
         L2.dgd| -.0114103    .012726    -0.90   0.370     -.0363528   .0135322
         L3.dgd|  .0283551   .0177666     1.60   0.110     -.0064667   .0631769
---------------+---------------------------------------------------------------
 Adjust. Term  |
---------------+---------------------------------------------------------------
   Mean Group: |
          lr_dy| -1.210094   .2005902    -6.03   0.000     -1.603243  -.8169442
---------------+---------------------------------------------------------------
 Long Run Est. |
---------------+---------------------------------------------------------------
   Mean Group: |
         lr_dgd| -.1198362   .0402251    -2.98   0.003      -.198676  -.0409965
          lr_dp| -.0795245   .0586992    -1.35   0.175     -.1945727   .0355238
-------------------------------------------------------------------------------
Mean Group Variables: L.dy L2.dy L3.dy dp dgd L.dp L2.dp L3.dp L.dgd L2.dgd L3.dgd lr_dgd lr_dp
Cross Sectional Averaged Variables: dy dp dgd
Long Run Variables:  lr_dgd lr_dp
Adjustment variable(s): lr_dy (L.dy L2.dy L3.dy)
Heterogenous constant partialled out.
```

6.4 CS-DL

Besides the ARDL model, Chudik et al. (2013) estimate an CS-DL model. Equation (15) in CS-DL form is:

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

The results from Chudik et al. (2013, Table 18) with 1 lag ($p = 1$) in the form of an ARDL(1,1,1) model can be replicated as follows

```
. xtdcce2 dy dp dgd d.(dp dgd) ///
> , cr(dy dp dgd) 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
------------------------------------------------------------------------------
          dy |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 Mean Group: |
          dp | -.0889337   .0256445    -3.47    0.001     -.1391959  -.0386715
         dgd | -.0865123      .0143    -6.05    0.000     -.1145398  -.0584849
        D.dp |  .0053277   .0413627     0.13    0.898     -.0757417   .0863971
       D.dgd |  .0068065   .0148306     0.46    0.646      -.022261   .0358739
------------------------------------------------------------------------------
Mean Group Variables: dp dgd D.dp D.dgd
Cross Sectional Averaged Variables: dy(0) dp(3) dgd(3)
Heterogenous constant partialled out.
```

The first differences as part of the vector $\Delta\mathbf{x}_{i,t}$ are added as `d.(dp dgd)`. The `fullsample` option is used to make use of the entire sample. The long run coefficients are -0.0889 (`dp`) and -0.0865 (`dgd`). While the coefficient on the inflation rate is almost identical to the CS-ARDL model, the coefficient on the debt to GDP is about half the absolute size. A (dis-)advantage of the CS-DL model is that no partial adjustment coefficient is estimated because the long run coefficients are directly estimated.

An ARDL(3,3,3) model is estimated using three rather than one lag for the differences and `L(0/2).d.(dp dgd)` replaces `d.(dp dgd)`:

```
. xtdcce2 dy dp dgd L(0/2).d.(dp dgd) ///
> , cr(dy dp dgd) cr_lags(0 3 3) fullsample
(Dynamic) Common Correlated Effects Estimator - Mean Group

Panel Variable (i): ccode                          Number of obs      =      1571
Time Variable (t): year                            Number of groups   =        40

Degrees of freedom per group:                      Obs per group (T)  =        39
 without cross-sectional averages   = 30.275
 with cross-sectional averages      = 21.275
Number of                                          F(720, 851)        =      1.12
 cross-sectional lags                 0 to 3       Prob > F           =      0.06
 variables in mean group regression = 320          R-squared          =      0.51
 variables partialled out           = 400          R-squared (MG)     =      0.47
                                                   Root MSE           =      0.03
                                                   CD Statistic       =      0.73
                                                      p-value         =    0.4680
-------------------------------------------------------------------------------
             dy|     Coef.   Std. Err.      z     P>|z|     [95% Conf. Interval]
-------------------------------------------------------------------------------
 Mean Group:   |
             dp| -.0855842   .0400845   -2.14    0.033     -.1641483    -.00702
            dgd| -.0816583   .0196252   -4.16    0.000     -.1201231  -.0431936
           D.dp|  .0183584   .0478696    0.38    0.701     -.0754643    .112181
          LD.dp|  .0015586   .0373619    0.04    0.967     -.0716695   .0747866
         L2D.dp|  .0034012   .0294771    0.12    0.908     -.0543729   .0611752
          D.dgd|  .0045224   .0144741    0.31    0.755     -.0238463   .0328912
         LD.dgd| -.0129675   .0134553   -0.96    0.335     -.0393395   .0134045
        L2D.dgd| -.0095151   .0090813   -1.05    0.295     -.0273142    .008284
-------------------------------------------------------------------------------
Mean Group Variables: dp dgd D.dp LD.dp L2D.dp D.dgd LD.dgd L2D.dgd
Cross Sectional Averaged Variables: dy(0) dp(3) dgd(3)
Heterogenous constant partialled out.
```

The first two variables (`dp` and `dgd`) represent the long run coefficients.

7 Conclusion

This paper explained how to test for cross-sectional dependence and estimate the exponent of cross-sectional dependence using the community contributed command `xtcse2`. It then reviewed three different methods to estimate long run coefficients in dynamic panels with a large number of observations over time and cross-sectional units and with cross-sectional dependence. It uses an extended version of `xtdcce2` (Ditzen, 2018) which allows for the estimation of long run coefficients using the CS-DL, CS-ARDL and CS-ECM estimator. Examples on how to apply `xtdcce2` are given and options are explained.

8 Acknowledgments

I am grateful to all participants of the Stata User Group Meeting in Zürich in 2018 and in London in 2019, in particular Achim Ahrens and David Drukker for valuable comments and feedback. I am grateful for help and comments from an anonymous referee and from Tore Bersvendsen, Sebastian Kripfganz, Kamiar Mohaddes, Mark Schaffer, Gregorio Tullio and plenty of users of `xtdcce2`, who gave valuable feedback. `xtcse2` benefited of help from Natalia Bailey and Sean Holly. All remaining errors are my own.

9 About the author

Jan Ditzen is a Forschungsassistent (Postdoc) at the Faculty of Economics and Management of the Free University of Bozen-Bolzano, Italy. His research interests are in the field of applied econometrics, with a focus on growth empirics and spatial econometrics, particularly cross-sectional dependence in large panels.

References

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

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

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.

Bersvendsen, T., and J. Ditzen. 2020. xthst : Testing for slope homogeneity in Stata. *CEERP Working Paper Series* (11).

Blackburne, E. F., and M. W. Frank. 2007. Estimation of nonstationary heterogeneous panels. *The Stata Journal* 7(2): 197–208.

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

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 (Advances in Econometrics, Volumn 36)*, ed. R. C. Hill, G. GonzÁlez-Rivera, and T.-H. Lee, 85–135. Emerald Group Publishing Limited.

Chudik, A., and M. H. Pesaran. 2015a. Large Panel Data Models with Cross-Sectional Dependence: A Survey. In *The Oxford Handbook Of Panel Data*, ed. B. H. Baltagi, chap. 1, 2–45. Oxford: Oxford University Press.

———. 2015b. 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.

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

———. 2019. Estimating long run effects in models with cross-sectional dependence using xtdcce2. *CEERP Working Paper Series* No. 7. URL `https://ceerp.hw.ac.uk/RePEc/hwc/wpaper/007.pdf.`

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

Engle, R. F., and C. W. J. Granger. 1987. Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica* 55(2): 251.

Everaert, G., and T. De Groote. 2016. Common Correlated Effects Estimation of Dynamic Panels with Cross-Sectional Dependence. *Econometric Reviews* 35(3): 1–31.

Karabiyik, H., S. Reese, and J. Westerlund. 2017. On the role of the rank condition in CCE estimation of factor-augmented panel regressions. *Journal of Econometrics* 197(1): 60–64. URL `http://dx.doi.org/10.1016/j.jeconom.2016.10.006.`

Karabiyik, H., J. Westerlund, and A. Juodis. 2020. On the Robustness of the Pooled CCE Estimator. *Journal of Econometrics* Forthcomin(xxxx).

Kripfganz, S., and D. Schneider. 2018. ardl: Estimating autoregressive distributed lag and equilibrium correction models. URL `https://ideas.repec.org/p/boc/usug18/09.html.`

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(4): 357–392.

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.

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 T. Yamagata. 2008. Testing slope homogeneity in large panels. *Journal of Econometrics* 142(1): 50–93.

Westerlund, J., and J. P. Urbain. 2015. Cross-sectional averages versus principal components. *Journal of Econometrics* 185(2): 372–377. URL `http://dx.doi.org/10.1016/j.jeconom.2014.09.014.`