The Stata Journal

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go "beyond the Stata manual" in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

For more information on the *Stata Journal*, including information for authors, see the webpage

http://www.stata-journal.com

The *Stata Journal* is indexed and abstracted in the following:

- CompuMath Citation Index®
- Current Contents/Social and Behavioral Sciences®
- RePEc: Research Papers in Economics
- Science Citation Index Expanded (also known as SciSearch®)
- Scopus™
- Social Sciences Citation Index®

The *Stata Journal*, electronic version (ISSN 1536-8734) is a publication of Stata Press. Stata, Mata, NetCourse, and Stata Press are registered trademarks of StataCorp LP.

The Stata Journal (2012)
12, Number 1, pp. 61–71

Estimating panel time-series models with heterogeneous slopes

Markus Eberhardt
School of Economics
University of Nottingham
Nottingham, UK
markus.eberhardt@nottingham.ac.uk

Abstract. This article introduces a new Stata command, `xtmg`, that implements three panel time-series estimators, allowing for heterogeneous slope coefficients across group members: the Pesaran and Smith (1995, *Journal of Econometrics* 68: 79–113) mean group estimator, the Pesaran (2006, *Econometrica* 74: 967–1012) common correlated effects mean group estimator, and the augmented mean group estimator introduced by Eberhardt and Teal (2010, Discussion Paper 515, Department of Economics, University of Oxford). The latter two estimators further allow for unobserved correlation across panel members (cross-section dependence).

Keywords: st0246, xtmg, nonstationary panels, parameter heterogeneity, cross-sectional dependence

1 Introduction

Over the past two decades, the study of panel data where both the cross-section (N) and the time-series (T) dimension are moderate to large has been a very active field within theoretical econometrics. This literature is dedicated to the analysis of macro panel datasets, where the cross-section dimension is typically represented by countries or states, provinces, or regions within countries. Examples for this type of data include the Penn World Table and macro data from organizations such as the World Bank, the Food and Agriculture Organization of the UN, the International Monetary Fund, and the Organization for Economic Co-operation and Development, all of which provide annual data for up to 60 years across many developing and developed economies.[1]

The theoretical literature on panel time-series econometrics began with a first generation of methods (unit-root tests, cointegration tests, and empirical estimators), which assumed that panel members were cross-sectionally independent (for example, Im, Pesaran, and Shin [2003]; Levin, Lin, and Chu [2002]; Maddala and Wu [1999]; and Pedroni [1999, 2004]). It then progressed to a second generation of methods that explicitly addressed the concerns of correlation across panel members (for example, Bai and Ng [2004]; Bai, Kao, and Ng [2009]; and Pesaran [2006, 2007]).

1. For links to these and other macro panel datasets, refer to the author's personal website at https://sites.google.com/site/medevecon.

On the applied side, however, there are still relatively few studies in mainstream economics journals that use panel time-series methods (examples include Cavalcanti, Mohaddes, and Raissi [2011]; Eberhardt, Helmers, and Strauss [forthcoming]; and Moscone and Tosetti [2010]). The analysis of macro panel data is still dominated by estimators developed for micro datasets (primarily the dynamic panel-data estimators by Arellano and Bond [1991] and Blundell and Bond [1998]).[2] The three empirical estimators introduced in this command relax the assumption of parameter homogeneity across panel members maintained by the aforementioned micro panel estimators.

2 Heterogeneous panel estimators

2.1 Empirical model

Assume the following simple model: for $i = 1, \ldots, N$ and $t = 1, \ldots, T$ let

$$y_{it} = \beta_i x_{it} + u_{it} \tag{1}$$

$$\text{where} \quad u_{it} = \alpha_{1i} + \lambda_i f_t + \varepsilon_{it} \tag{2}$$

$$x_{it} = \alpha_{2i} + \lambda_i f_t + \gamma_i g_t + e_{it} \tag{3}$$

where x_{it} and y_{it} are observables, β_i is the country-specific slope on the observable regressor, and u_{it} contains the unobservables and the error terms ε_{it}. The unobservables in (2) are made up of group fixed effects α_{1i}, which capture time-invariant heterogeneity across groups, as well as an unobserved common factor f_t with heterogeneous factor loadings λ_i, which can capture time-variant heterogeneity and cross-section dependence. The factors f_t and g_t are not limited to linear evolution over time; they can be nonlinear and nonstationary, with obvious implications for cointegration.[3] Additional problems arise because the regressors are driven by some of the same common factors as the observables: the presence of f_t in (2) and (3) induces endogeneity in the estimation equation (see discussions by Coakley, Fuertes, and Smith [2006] and Eberhardt and Teal [2011]). ε_{it} and e_{it} are assumed white noise. For simplicity of exposition, the model developed here includes only one covariate and one unobserved common factor in the estimation equation of interest; the principle extends to multiple covariates and factors.

All mean group (MG) type estimators follow the same principle methodology:

1. Estimate N group-specific ordinary least-squares (OLS) regressions.

2. Average the estimated coefficients across groups.

2. The discussion by Roodman (2009) is particularly illuminating in this context because all empirical examples provided in the article use macro panel data. The prevalence of the "dynamic panel-data estimators" in empirical application is at least in part because of the `xtabond2` command written by David Roodman, which made these methods available to Stata users.
3. g_t is included to highlight that the observables x will also be driven by factors other than f_t.

The first of these steps is made up of standard OLS regressions where for the common correlated effects mean group (CCEMG) and the augmented mean group (AMG) estimators, each empirical equation is simply augmented with additional covariates (to be detailed below).

The (weighted or unweighted) average of country-specific estimates for β_i provides a first benchmark of comparison for these heterogeneous parameter model results with pooled model results (including pooled OLS, two-way fixed effects, and Arellano–Bond-type estimators), and we will view this average as the parameter of interest. The `xtmg` command results thus indicate the *average* relationship across panel members. In principle, however, allowing the slope coefficients to differ across panel members opens up a further dimension of inquiry, namely, the analysis of the patterns and the ultimate source of this parameter heterogeneity.[4]

The following sections describe the three estimators implemented in this routine in more detail.

2.2 Pesaran and Smith (1995)

The Pesaran and Smith (1995) MG estimator does not concern itself with cross-section dependence and assumes away $\lambda_i f_t$ or models these unobservables with a linear trend. Thus (1) above is estimated for each panel member i, including an intercept to capture fixed effects and (optionally) a linear trend to capture time-variant unobservables. The estimated coefficients $\widehat{\beta}_i$ are subsequently averaged across panel members—here weights can be applied, but in the standard implementation this is just the unweighted average.[5]

2.3 Pesaran (2006)

The Pesaran (2006) CCEMG estimator allows for the empirical setup as laid out in (1), (2), and (3). The empirical setup induces cross-section dependence, time-variant unobservables with heterogeneous impact across panel members, and problems of identification (β_i is unidentified if the regressor contains f_t).[6] The CCEMG estimator solves this problem with a simple but powerful augmentation of the group-specific regression equation: apart from the regressors x_{it} and an intercept, this equation now includes the cross-section averages of the dependent and independent variables, $\overline{y}_t$ and $\overline{x}_t$, as additional regressors. The combination of $\overline{y}_t$ and $\overline{x}_t$ can account for the unobserved common factor f_t. Because the relationship is estimated for each panel member separately, the heterogeneous impact (λ_i) is also given by construction (for an accessible discussion, see Eberhardt, Helmers, and Strauss [forthcoming]). Thus, in practical terms, cross-section

4. Using an alternative approach, Durlauf, Kourtellos, and Minkin (2001) were among the first to emphasize this issue. See Eberhardt and Teal (2010, 2011) for a detailed discussion.
5. Note that the `xtpmg` command by Blackburne and Frank (2007) and the `xtwest` command by Persyn and Westerlund (2008) optionally provide MG estimates for dynamic specifications.
6. The latter issue is comparable with the "transmission bias" problem in micro production function models, whereby inputs x_{it} are correlated with (from the econometrician's perspective) unobserved productivity shocks f_t.

averages $\overline{y}_t$ and $\overline{x}_t$ for all observable variables in the model are computed (using the data *for the entire panel*) and then added as explanatory variables in each of the N regression equations. Subsequently, the estimated coefficients $\widehat{\beta}_i$ are averaged across panel members, where different weights may be applied.

The focus of the estimator is to obtain consistent estimates of the parameters related to the observable variables. In empirical application, the estimated coefficients on the cross-section–averaged variables as well as their average estimates are not interpretable in a meaningful way; they are merely present to blend out the biasing impact of the unobservable common factor. The CCEMG approach is robust to the presence of a limited number of "strong" factors and an infinite number of "weak" factors—the latter can be associated with local spillover effects, whereas the former can represent global shocks, such as the recent global financial crisis (Chudik, Pesaran, and Tosetti 2011; Pesaran and Tosetti 2011). Furthermore, the estimator is robust to nonstationary common factors (Kapetanios, Pesaran, and Yamagata 2011).

2.4 Eberhardt and Teal (2010)

The AMG estimator was developed by Eberhardt and Teal (2010) as an alternative to the Pesaran (2006) CCEMG estimator with macro production function estimation in mind. In the CCEMG estimator, the unobservable common factor f_t is treated as a nuisance, something to be accounted for that is not of particular interest for the empirical analysis. In cross-country production functions, however, unobservables represent total factor productivity (TFP). Note that standard panel approaches to cross-country empirics are commonly based on a production function of Cobb–Douglas form; see Eberhardt and Teal (2011) for a detailed discussion of the growth empirics literature.

The AMG procedure is implemented in three steps:

1. A pooled regression model augmented with year dummies is estimated by first difference OLS, and the coefficients on the (differenced) year dummies are collected. They represent an estimated cross-group average of the evolution of unobservable TFP over time. This is referred to as the "common dynamic process".

2. The group-specific regression model is then augmented with this estimated TFP process: either a) as an explicit variable or b) imposed on each group member with a unit coefficient by subtracting the estimated process from the dependent variable. Like in the MG case, each regression model includes an intercept that captures time-invariant fixed effects (TFP levels).

3. Like in the MG and CCEMG estimators, the group-specific model parameters are averaged across the panel (weights may be applied).

In Monte Carlo simulations (Eberhardt and Bond 2009), the AMG and CCEMG performed similarly well in terms of bias or root mean squared error (RMSE) in panels with nonstationary variables (cointegrated or not) and multifactor error terms (cross-section dependence).

The standard errors reported in the averaged regression results of all three estimators are constructed following Pesaran and Smith (1995), thus testing the significant difference of the average coefficient from zero. In practice, the group-specific coefficients are regressed on an intercept, either without any weighting or with less weight attached to "outliers" (see `rreg` by Hamilton [1991] for more details on the latter).

3 The xtmg command

3.1 Syntax

`xtmg` *depvar* [*indepvars*] [*if*] [*in*] [, `cce` `augment` `impose` `trend` `robust` `full` `noconstant` `level(`#`)` `res(`*varname*`)` `pred(`*varname*`)`]

3.2 Options

`cce` implements the Pesaran (2006) CCEMG estimator. The Pesaran and Smith (1995) MG estimator is set as the default. The regression output includes the averaged coefficients on the cross-section averages of the dependent and independent variables. These are identified in the results table as *varname*`_avg`.

`augment` implements the AMG estimator. This option cannot be used with `cce`.

`impose` specifies that the AMG estimator be implemented by imposing the "common dynamic process" with unit coefficient—by subtracting it from the dependent variable. This option works only if used with `augment`. The default is for the "common dynamic process" to enter as an additional covariate.

`trend` specifies each group-specific regression to be augmented with a linear trend term.

`robust` specifies that the `rreg` command be used to construct the coefficient averages across N panel members reported (see Hamilton [1991] for details). This puts less emphasis on outliers while computing the average coefficient. The default is unweighted averages.

`full` specifies that all N regression results be listed. Individual results will be numbered from 1 to N in the order given in the cross-section identifier of `xtset`. Only the averaged coefficients are listed by default.

`noconstant` suppresses the constant term. This option is generally not recommended.

`level(`#`)` specifies the confidence level, as a percentage, for confidence intervals. The default is `level(95)` or as set by `set level`; see [U] **20.7 Specifying the width of confidence intervals**.

`res(`*varname*`)` provides residuals, which are stored in *varname*. These can then be subjected to diagnostic tests, including testing for cross-section dependence (see `xtcd` if installed). Note that these residual series are not based on the linear prediction of the averaged MG estimates but are derived from the group-specific regressions. This is similar to the postestimation command `predict` with the option `group(`*varname*`)` in the random coefficient model estimator `xtrc`, although the latter allows only predicted values (not residuals) to be computed.

`pred(`*varname*`)` provides predicted values, which are stored in *varname*. These series are based on the linear prediction of the group-specific regressions.

3.3 Saved results

`xtmg` saves the following in `e()`:

Scalars	
`e(N)`	number of observations
`e(N_g)`	number of groups
`e(g_min)`	fewest number of observations in an included group
`e(g_max)`	greatest number of observations in an included group
`e(g_avg)`	average number of observations in an included group
`e(df_m)`	model degrees of freedom
`e(chi2)`	χ^2
`e(trend_sig)`	share of statistically significant linear trends
Macros	
`e(cmd)`	`xtmg`
`e(depvar)`	dependent variable
`e(ivar)`	group (panel) variable
`e(tvar)`	time variable
`e(title2)`	estimator selected: MG, CCEMG, or AMG
Matrices	
`e(b)`	coefficient vector
`e(V)`	variance–covariance matrix of the estimators
`e(betas)`	group-specific regression coefficients (vector)
`e(varbetas)`	variances for group-specific regression coefficients (vector)
`e(stebetas)`	standard errors for group-specific regression coefficients (vector)
`e(tbetas)`	t statistics for group-specific regression coefficients (vector)
Functions	
`e(sample)`	marks estimation sample

4 Empirical example: Cross-country productivity analysis

In this section, I illustrate the use of `xtmg` by investigating a cross-country production function for the manufacturing sector, taken from Eberhardt and Teal (2010). The dataset consists of aggregate sectoral data for manufacturing in a panel of 48 developing and developed countries from 1970 to 2002 (unbalanced panel), taken from the United Nations Industrial Development Organization's Industrial Statistics database (UNIDO 2004). Preliminary investigation of the annual data suggests that the variables used are integrated of order one. The dataset must be `xtset` before use.

```
. use manu_prod
(Manufacturing productivity analysis (1970-2002))
. xtset nwbcode year
       panel variable:  nwbcode (strongly balanced)
        time variable:  year, 1970 to 2002
                delta:  1 unit
```

The data have been deflated to constant US$ 1990 values and are investigated in a standard Cobb–Douglas production function of the form

$$Y = AK^{\alpha_i}L^{1-\alpha_i}$$

where Y is value-added, K is capital stock (constructed using the permanent inventory method), and L is the labor force. A captures TFP. This model is taken to the data in a log-linearized form with a technology parameter α that is heterogeneous across countries and constant returns to scale imposed (value-added and capital stock are now in per-worker terms, indicated by lowercase letters):

$$\ln\, y_{it} = A_{it} + \alpha_i \ln\, k_{it} + \varepsilon_{it}$$

We implement the MG, AMG, and CCEMG estimators, reporting unweighted coefficient averages; results are contained in table 1. These are the results reported by Eberhardt and Teal (2010), which are qualitatively identical to weighted (outlier-robust) averages, indicating that outliers do not influence the results.

Table 1. Country regression averages (imposed)

	[1]	[2]	[3]	[4]	[5]
	MG	AMG	AMG	CCEMG	CCEMG
dep. variable	ly	ly-$\widehat{\mu}_t^{va\bullet}$	ly	ly	ly
log capital per worker	0.179	0.290	0.298	0.466	0.312
	[2.22]*	[3.91]**	[3.66]**	[6.69]**	[3.68]**
common dynamic process			0.879		
			[4.35]**		
country trend	0.017	0.000	0.002		0.011
	[5.89]**	[0.04]	[0.55]		[3.06]**
intercept	7.653	6.382	6.243	0.896	4.786
	[8.95]**	[8.33]**	[7.32]**	[0.88]	[3.62]**
# of sign. trends	33	24	15	n/a	18
RMSE	0.100	0.097	0.091	0.099	0.088

Notes: t statistics are reported in square brackets. Statistical significance at the 5% and 1% level is indicated with * and **, respectively. $\widehat{\mu}_t^{va\bullet}$ signifies the "common dynamic process".

The MG estimator in column [1] does not explicitly account for cross-section dependence; it yields a capital coefficient of about 0.18, considerably below the capital share in output (taken from aggregate macro data), which is typically around 1/3

(Mankiw, Romer, and Weil 1992). In contrast, the AMG and CCEMG estimators all yield capital coefficients around 0.3 (in the case of the CCEMG, once each country regression is augmented with a linear country trend).

To illustrate, I report the Stata output for the MG and CCEMG models (in both cases including country-specific linear trend terms) below. This corresponds to the results in columns [1] and [5] of table 1. In addition to the standard Stata panel regression information, the routine reports the RMSE. If the option `trend` is selected, the number of trends that are statistically significant at the specified significance level is also reported (the default 5% level is used here). Residuals have been computed and stored in variables `eMG` and `eCMGt`.

```
. xtmg ly lk, trend res(eMG)

Pesaran & Smith (1995) Mean Group estimator

All coefficients represent averages across groups (group variable: nwbcode)
Coefficient averages computed as unweighted means

Mean Group type estimation                        Number of obs      =      1194
Group variable: nwbcode                           Number of groups   =        48

                                                  Obs per group: min =        11
                                                                 avg =      24.9
                                                                 max =        33

                                                  Wald chi2(1)       =      4.94
                                                  Prob > chi2        =    0.0263

------------------------------------------------------------------------------
          ly |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          lk |   .1789207   .0805226     2.22   0.026     .0210994    .3367421
       trend |   .0174254   .0029601     5.89   0.000     .0116238     .023227
       _cons |   7.652843   .8546496     8.95   0.000     5.977761    9.327926
------------------------------------------------------------------------------
Root Mean Squared Error (sigma): 0.0996
Residual series based on country regressions stored in variable: eMG
Variable trend refers to the group-specific linear trend terms.
Share of group-specific trends significant at 5% level: 0.688 (= 33 trends)
```

```
. xtmg ly lk, cce trend res(eCMGt)

Pesaran (2006) Common Correlated Effects Mean Group estimator

All coefficients represent averages across groups (group variable: nwbcode)
Coefficient averages computed as unweighted means

Mean Group type estimation                        Number of obs      =      1194
Group variable: nwbcode                           Number of groups   =        48

                                                  Obs per group: min =        11
                                                                 avg =      24.9
                                                                 max =        33

                                                  Wald chi2(1)       =     13.54
                                                  Prob > chi2        =    0.0002

------------------------------------------------------------------------------
          ly |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          lk |   .3124664   .0849231     3.68   0.000     .1460202    .4789127
       trend |   .0108121   .0035327     3.06   0.002     .0038881     .017736
      ly_avg |   .6570663   .1563127     4.20   0.000      .350699    .9634335
      lk_avg |  -.4640624   .1260282    -3.68   0.000    -.7110731   -.2170518
       _cons |   4.786033   1.322707     3.62   0.000     2.193575    7.378492
------------------------------------------------------------------------------

Root Mean Squared Error (sigma): 0.0877
Cross-section averaged regressors are marked by the suffix avg.
Residual series based on country regressions stored in variable: eCMGt
Variable trend refers to the group-specific linear trend terms.
Share of group-specific trends significant at 5% level: 0.375 (= 18 trends)
```

5 Acknowledgments

This routine builds on the existing Stata code for the Swamy random coefficients model estimator (`xtrc`), the Pesaran, Shin, and Smith (1999) pooled mean group estimator written by Edward F. Blackburne III and Mark W. Frank (`xtpmg`), and the Westerlund (2007) error-correction cointegration test (`xtwest`) written by Damiaan Persyn. Thanks to Kit Baum and a *Stata Journal* reviewer for useful comments, help, and support. Any remaining errors are my own.

6 References

Arellano, M., and S. Bond. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58: 277–297.

Bai, J., C. Kao, and S. Ng. 2009. Panel cointegration with global stochastic trends. *Journal of Econometrics* 149: 82–99.

Bai, J., and S. Ng. 2004. A PANIC attack on unit roots and cointegration. *Econometrica* 72: 1127–1177.

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

Blundell, R., and S. Bond. 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87: 115–143.

Cavalcanti, T., K. Mohaddes, and M. Raissi. 2011. Growth, development and natural resources: New evidence using a heterogeneous panel analysis. *Quarterly Review of Economics and Finance* 51: 305–318.

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

Coakley, J., A.-M. Fuertes, and R. Smith. 2006. Unobserved heterogeneity in panel time series models. *Computational Statistics and Data Analysis* 50: 2361–2380.

Durlauf, S. N., A. Kourtellos, and A. Minkin. 2001. The local Solow growth model. *European Economic Review* 45: 928–940.

Eberhardt, M., and S. Bond. 2009. Cross-section dependence in nonstationary panel models: A novel estimator. MPRA Paper 17692, University Library of Munich. http://mpra.ub.uni-muenchen.de/17692/1/MPRA_paper_17692.pdf.

Eberhardt, M., C. Helmers, and H. Strauss. Forthcoming. Do spillovers matter when estimating private returns to R&D? *Review of Economics and Statistics*.

Eberhardt, M., and F. Teal. 2010. Productivity analysis in global manufacturing production. Discussion Paper 515, Department of Economics, University of Oxford. http://www.economics.ox.ac.uk/research/WP/pdf/paper515.pdf.

———. 2011. Econometrics for grumblers: A new look at the literature on cross-country growth empirics. *Journal of Economic Surveys* 25: 109–155.

Hamilton, L. C. 1991. How robust is robust regression? *Stata Technical Bulletin* 2: 21–26. Reprinted in *Stata Technical Bulletin Reprints*, vol. 1, pp. 169–175. College Station, TX: Stata Press.

Im, K. S., M. H. Pesaran, and Y. Shin. 2003. Testing for unit roots in heterogeneous panels. *Journal of Econometrics* 115: 53–74.

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

Levin, A. T., C.-F. Lin, and C.-S. J. Chu. 2002. Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics* 108: 1–24.

Maddala, G. S., and S. Wu. 1999. A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and Statistics* 61: 631–652.

Mankiw, N. G., D. Romer, and D. N. Weil. 1992. A contribution to the empirics of economic growth. *Quarterly Journal of Economics* 107: 407–437.

Moscone, F., and E. Tosetti. 2010. Health expenditure and income in the United States. *Health Economics* 19: 1385–1403.

Pedroni, P. 1999. Critical values for cointegration tests in heterogeneous panels with multiple regressors. *Oxford Bulletin of Economics and Statistics* 61: 653–670.

———. 2004. Panel cointegration: Asymptotic and finite sample properties of pooled time series tests with an applicaton to the PPP hypothesis. *Econometric Theory* 20: 597–625.

Persyn, D., and J. Westerlund. 2008. Error-correction–based cointegration tests for panel data. *Stata Journal* 8: 232–241.

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

———. 2007. A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics* 22: 265–312.

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: 621–634.

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

Pesaran, M. H., and E. Tosetti. 2011. Large panels with common factors and spatial correlation. *Journal of Econometrics* 161: 182–202.

Roodman, D. M. 2009. A note on the theme of too many instruments. *Oxford Bulletin of Economics and Statistics* 71: 135–158.

UNIDO. 2004. UNIDO Industrial Statistics 2004. Online database. Vienna: UNIDO (United Nations Industrial Development Organization).

Westerlund, J. 2007. Testing for error correction in panel data. *Oxford Bulletin of Economics and Statistics* 69: 709–748.

About the author

Markus Eberhardt is a lecturer (assistant professor) in the School of Economics at the University of Nottingham and a research associate in the Centre for the Study of African Economies in the Department of Economics at the University of Oxford.