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`, which implements a number of panel time series estimators allowing for heterogeneous slope coefficients across group members: the Pesaran and Smith (1995) Mean Group (MG) estimator, the Pesaran (2006) Common Correlated Effects Mean Group (CCEMG) estimator and the Augmented Mean Group (AMG) estimator introduced in Eberhardt and Teal (2010). The latter two estimators further allow for unobserved correlation across panel members (cross-section dependence).

Keywords: st0001, 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 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 and regions within countries. Examples for this type of data include the Penn World Table and macro data from organisations such as the World Bank, Food and Agriculture Organisation of the UN (FAO), the International Monetary Fund (IMF), or the Organisation for Economic Co-operation and Development (OECD), all of which provide annual data for at times up to 60 years across a significant number of developing and developed economies.¹

The theoretical literature on panel time series econometrics has progressed from a first generation of panel unit root tests, cointegration tests and empirical estimators, which assumed that panel members were cross-sectionally independent (e.g. Im et al. 2003; Levin et al. 2002; Maddala and Wu 1999; Pedroni 1999, 2004), to a second generation of methods which explicitly addressed the concerns of correlation across panel members (e.g. Bai and Ng 2004; Bai et al. 2009; Pesaran 2006, 2007). On the applied side, however, there are still relatively few studies in mainstream economics journals which employ panel time series methods (examples include Cavalcanti et al. 2011; Eberhardt et al. forthcoming; Moscone and Tosetti 2010) and the analysis of macro panel data is still dominated by estimators developed for micro datasets (primarily the

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

© yyyy StataCorp LP
dynamic panel data estimators by Arellano and Bond 1991; 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

\[
\begin{align*}
y_{it} &= \beta_i x_{it} + u_{it} \\
\text{where } u_{it} &= \alpha_{1i} + \lambda_i f_t + \varepsilon_{it} \\
x_{it} &= \alpha_{2i} + \lambda_i g_t + \gamma_i x_{it} + \varepsilon_{it}
\end{align*}
\]

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

All Mean Group type estimators follow the same principle methodology:

1. estimate \(N\) group-specific OLS regressions,
2. average the estimated coefficients across groups.

The first of these steps is made up of standard OLS regressions where in case of the CCEMG and 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 \(\beta_i\) provides a first benchmark of comparison for these heterogeneous parameter model results with

\(^2\) The discussion in Roodman (2009) is particularly illuminating in this context, since all empirical examples provided in the article employ macro panel data. It should be noted here that the prevalence of the ‘dynamic panel data estimators’ in empirical application is at least in part due to the \texttt{xtabond2} command written by David Roodman which made these methods available to \texttt{Stata} users.

\(^3\) \(g_t\) is included to highlight that the observables \(x\) will also be driven by factors other than \(f_t\).
pooled model results (pooled OLS, two-way fixed effects, Arellano-Bond-type estimators, among others) and in the present article we shall view this average as the parameter of interest. The \texttt{xtmg} results thus indicate the average relationship across panel members. In principle, however, it is important to note that allowing the slope coefficients to differ across panel members opens up a further dimension of enquiry, namely the analysis of the patterns as well as the ultimate source of this parameter heterogeneity.\footnote{Using an alternative approach Durlauf et al. (2001) were among the first to emphasise this issue. See Eberhardt and Teal (2010, 2011) for a detailed discussion.}

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

\subsection{Pesaran and Smith (1995)}

The Pesaran and Smith (1995) Mean Group (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, equation (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 $\hat{\beta}_i$ are subsequently averaged across panel members — here weights can be applied but in the standard implementation this is just the unweighted average.\footnote{Note that the \texttt{xtmg} command by Blackburne III and Frank (2007) as well as the \texttt{xtwest} command by Persyn and Westerlund (2008) optionally provide MG estimates for dynamic specifications.}

\subsection{Pesaran (2006)}

The Pesaran (2006) Common Correlated Effects Mean Group (CCEMG) estimator allows for the empirical setup as laid out in equations (1) to (3), which induces cross-section dependence, time-variant unobservables with heterogeneous impact across panel members and problems of identification ($\beta_i$ is unidentified if the regressor contains $f_t$).\footnote{The latter issue is comparable to the `transmission bias' problem in micro production function models, whereby inputs $x_{it}$ are correlated with (from the econometrician’s perspective) unobserved productivity shocks $f_t$.}

The CCEMG 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 $f_t$ and as the relationship is estimated for each panel member separately the heterogeneous impact ($\lambda_i$) is also given by construction (for an accessible discussion see Eberhardt et al. forthcoming). Thus, in practical terms, cross-section 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 $\hat{\beta}_i$ are averaged across panel members, where different weights may be applied.

The focus of the estimator is on obtaining 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 as well as 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 et al. 2011; Pesaran and Tosetti 2011). Furthermore, as shown in Kapetanios et al. (2011), the estimator is robust to nonstationary common factors.

2.4 Eberhardt and Teal (2010)

The Augmented Mean Group estimator (AMG) was developed in Eberhardt and Teal (2010) as an alternative to the Pesaran (2006) CCEMG with macro production function estimation in mind. In the CCEMG the unobservable common factor \( f_t \) is treated as a nuisance, something to be accounted for which 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 unit coefficient by subtracting the estimated process from the dependent variable. Like in the MG case each regression model includes an intercept, which captures time-invariant fixed effects (TFP level).

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 (Bond and Eberhardt 2009) the AMG performed similarly well as the CCEMG in terms of bias or 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 attaching less weight to ‘outliers’ (see \texttt{rreg} by Hamilton (1992) for more details on the latter).
3 The xtmg command

3.1 Syntax

\[
\text{xtmg} \ depvar \ [ \ indepvars ] \ [ \ if \ ] \ [ \ in \ ] \ [, \ \text{cce aug imp trend full noconstant level(\#) res(varname) pred(varname)} \]
\]

3.2 Options

The Pesaran and Smith (1995) Mean Group 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 \textit{varname avg}.

cce aug implements the Augmented Mean Group estimator.

imp specifies that the Augmented Mean Group estimator is implemented by imposing the ‘common dynamic process’ with unit coefficient — by subtracting it from the dependent variable. This option only works in combination with aug.

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

robust specifies the use of the \textit{rreg} command to construct the coefficient averages across \(N\) panel members reported (see Hamilton 1992, for details). This puts less emphasis on outliers in the computation of 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 \textit{xtset}. Only the averaged coefficients are listed by default.

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

level(\#) specifies the confidence level, as a percentage, for confidence intervals. The default is level(95) or as set by \textit{set level}; see \textit{[U] 23.5 Specifying the width of confidence intervals}.

res(varname) provides residuals which are stored in \textit{varname}. These can then be subjected to diagnostic tests, including testing for cross-section dependence (see \textit{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 post-estimation command \textit{predict} with the option \textit{group(varname)} in the Random Coefficient Model estimator \textit{xtrc}, although in the latter this only allows predicted values but not residuals to be computed.

pred(varname) provides predicted values which are stored in \textit{varname}. These series are again based on the linear prediction of the group-specific regressions.
3.3 Saved results

The \texttt{xtmg} routine environment saves the following information to \texttt{e()}: 

**Scalars**

- \texttt{e(N)}: number of observations
- \texttt{e(g_min)}: lowest number of observations in an included group
- \texttt{e(g_max)}: highest number of observations in an included group
- \texttt{e(N_g)}: number of groups
- \texttt{e(g_avg)}: average number of observations in an included group
- \texttt{e(rmse)}: root mean squared error
- \texttt{e(df_m)}: model degrees of freedom
- \texttt{e(chi2)}: Wald chi-squared statistic
- \texttt{e(trend_sig)}: share of statistically significant linear trends

**Macros**

- \texttt{e(cmd)}: \texttt{xtmg}
- \texttt{e(ivar)}: group (panel) variable
- \texttt{e(depvar)}: dependent variable
- \texttt{e(tvar)}: time variable
- \texttt{e(title2)}: estimator selected: MG, CCEMG or AMG

**Matrices**

- \texttt{e(b)}: coefficient vector
- \texttt{e(V)}: variance–covariance matrix of the estimators
- \texttt{e(betas)}: group-specific regression coefficients (vector)
- \texttt{e(varbetas)}: variances for group-specific regression coefficients (vector)
- \texttt{e(stebetas)}: st. errors for group-specific regression coefficients (vector)
- \texttt{e(tbetas)}: t-statistics for group-specific regression coefficients (vector)

**Functions**

- \texttt{e(sample)}: marks estimation sample

4 Empirical Example: cross-country productivity analysis

In this section we illustrate the use of \texttt{xtmg} by investigating a cross-country production function for the manufacturing sector, taken from Eberhardt and Teal (2010). The data 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 Nation Industrial Development Organisation’s Industrial Statistics database (UNIDO 2004, IndStat). Preliminary investigation of the annual data suggests the variables employed are integrated of order one. The dataset must be \texttt{tsset} before use.

```stata
. 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 constant returns to scale Cobb-Douglas production function of the form

\[ Y = AK^\alpha L^{1-\alpha}, \]  

where \( Y \) is value-added (VA), \( K \) is capital stock (constructed using the permanent inventory method) and \( L \) the labour force. \( \alpha \) captures Total Factor Productivity. This model is taken to the data in a log-linearised form with technology parameter \( \alpha \) hetero-
geneous across countries and constant returns to scale imposed (VA and capital stock are now in per worker terms, indicated by lower case letters)

\[ \ln y_{it} = A_{it} + \alpha_i \ln k_{it} + \varepsilon_{it} \] (5)

We implement the MG, AMG and CCEMG estimators reporting unweighted coefficient averages — results are contained in Table 1. These are the results reported in 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 (CRS imposed)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>log capital per worker</td>
<td>0.179</td>
<td>0.290</td>
<td>0.298</td>
<td>0.466</td>
<td>0.312</td>
</tr>
<tr>
<td>common dynamic process</td>
<td>0.879</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[4.35]**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>country trend</td>
<td>0.017</td>
<td>0.000</td>
<td>0.002</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[5.89]**</td>
<td>[0.04]</td>
<td>[0.55]</td>
<td>[3.06]**</td>
<td></td>
</tr>
<tr>
<td>intercept</td>
<td>7.653</td>
<td>6.382</td>
<td>6.243</td>
<td>0.896</td>
<td>4.786</td>
</tr>
<tr>
<td></td>
<td>[8.95]**</td>
<td>[8.33]**</td>
<td>[7.32]**</td>
<td>[0.88]</td>
<td>[3.62]**</td>
</tr>
<tr>
<td># of sign. trends</td>
<td>33</td>
<td>24</td>
<td>15</td>
<td>n/a</td>
<td>18</td>
</tr>
<tr>
<td>RMSE</td>
<td>.100</td>
<td>.097</td>
<td>.091</td>
<td>.099</td>
<td>.088</td>
</tr>
</tbody>
</table>

Notes: \(t\)-statistics reported in square brackets. Statistical significance at the 5% and 1% level is indicated with * and ** respectively. \(\hat{\mu}_t^{va}\) signifies the ‘common dynamic process’.

The MG estimator in column [1] does not explicitly account for cross-section dependence and yields a capital coefficient of around .18, considerably below the capital share in output (taken from aggregate macro data), which is typically around 1/3 (Mankiw et al. 1992). In contrast, the AMG and CCEMG estimators all yield capital coefficients around .3, in case of the CCEMG once each country regression is augmented with a linear country trend.

For illustration we 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 Root Mean Squared Error. If the option trend is selected the number of trends which are statistically significant at the specified significance level is also reported (here the default 5% level is taken). Residuals have been computed and stored in variables \(eMG\) and \(eCMGt\).
Panel time series models with heterogeneous slopes

.xtmg ly lk, trend res(eMG)

Pesaran & Smith (1995) Mean Group estimator

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

Mean Group type estimation
Group variable: list

| Coef. | Std. Err. | z   | P>|z| | [95% Conf. Interval] |
|-------|-----------|-----|-----|---------------------|
| lk    | 0.1792207 | 0.0805226 | 2.22 | 0.026 | 0.0210994 - 0.3367421 |
| trend | 0.0174254 | 0.0029601 | 5.89 | 0.000 | 0.0116238 - 0.023227 |
| _cons | 7.652843  | 0.8546496 | 8.95 | 0.000 | 5.977761 - 9.327926 |

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

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)


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

Mean Group type estimation
Group variable: list

| Coef. | Std. Err. | z   | P>|z| | [95% Conf. Interval] |
|-------|-----------|-----|-----|---------------------|
| ly_avg| 0.3124664 | 0.0849226 | 3.68 | 0.000 | 0.1460202 - 0.479127 |
| trend | 0.0108121 | 0.0035327 | 3.06 | 0.002 | 0.0038881 - 0.017736 |
| ly_avg| 0.6570663 | 0.1563127 | 4.20 | 0.000 | 0.350699 - 0.963435 |
| lk_avg| -0.4840624 | 0.1260282 | -3.68 | 0.000 | -0.7110731 -0.2170518 |
| _cons | 4.786033  | 1.322707 | 3.62 | 0.000 | 2.193575 - 7.378492 |

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

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 Acknowledgements

This routine builds on the existing code for the Swamy RCM estimator (\texttt{xtrc}), the Pesaran et al. (1999) Pooled Mean Group estimator written by Edward F. Blackburne III and Mark W. Frank (\texttt{xtpmg}) and the Westerlund (2007) error correction cointegration test (\texttt{xtwest}) written by Damiaan Persyn. Thanks to Kit Baum and a \textit{Stata Journal} reviewer for useful comments, help and support. Any remaining errors are my own.

6 References


Panel time series models with heterogeneous slopes


**About the author**

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