Economics Letters 117 (2012) 837–840

Contents lists available at SciVerse ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

GDP clustering: A reappraisal

Michele Battisti [a,b,c], Christopher F. Parmeter [d,*]

[a] Dipartimento di Studi su Politica, Diritto e Societá, Palermo, Italy
[b] CeLEG LUISS Rome, Italy
[c] RCEA Rimini, Italy
[d] Department of Economics, University of Miami, Coral Gables, FL 33124-6520, United States

ARTICLE INFO

Article history:
Received 13 August 2012
Accepted 24 August 2012
Available online 11 September 2012

JEL classification:
C33
O15
O16
O47

Keywords:
Mixture densities
Output
Autoregressive
Isotropic

ABSTRACT

This note explores clustering in cross country GDP per capita using recently developed model based clustering methods for panel data. Previous research characterizing the components of the overall distribution of output either use ad hoc methods, or methods which ignore/subvert the panel nature of the data. These new methods allow the characterization of the possible autoregressive relationship of output between time points. We show that traditional static clustering decade by decade gives mixed results regarding clustering over time, while the application of longitudinal mixtures presents three distinct clusters at all periods of time.

1. Introduction

There has been a growing use of mixture models in the growth empirics literature. These methods acknowledge heterogeneity across a variety of dimensions in the growth spectrum, allowing the characterization of club convergence and the measurement of polarization. The primary use of these methods has been to study either direct features of the distribution of growth, such as the number of distinct components, or to determine the variables which influence membership into a specific club. Additionally, within group features have also been investigated, such as club σ-convergence. Overall, mixture methods have provided an array of insights into the cross country growth process.

One shortcoming of the existing methods is that they have been administered in a primarily static setting, even with the use of panel data. For example, Pittau et al. (2010) and Battisti and Parmeter (2011) both investigate the distribution of cross country output with access to panel data, but their main analyses hinge on treating each data period as distinct from one another. However, given the repeated measurement of countries over time, a more rigorous analysis can be conducted by allowing for cross period covariance within the mixture setup.

Using recently developed mixture methods for repeatedly measured data, this paper takes a fresh look at clustering in the distribution of cross country output. While we find the number of components to be consistent with previous studies,[1] the mean-variance dynamics over time suggest different behavior across the identified components, notably a disappearance of the middle class as measured by the component averages. Further, for our dataset, we see that using static decade by decade clustering does not produce a consistent number of components over time. On the contrary, exploiting the panel nature we show that there exist three components, with the poorest component (mainly sub-Saharan nations) not growing over time, while there is a tendency of within and across club convergence for the other two components.

* Corresponding author. Tel.: +1 305 284 4397; fax: +1 305 284 2985.
E-mail addresses: mbattisti@unipa.it (M. Battisti), cparmeter@bus.miami.edu (C.F. Parmeter).

[1] Pittau et al. (2010) find three groups using weighted average workforce GDP per worker PWT data over the period 1960–2000, while we have a longer time frame and different output data.

0165-1765/$ – see front matter
doi:10.1016/j.econlet.2012.08.038

2. Empirical framework

The basic framework for our empirical analysis is the Gaussian multivariate mixture model:

$$f(y) = \sum_{g=1}^{G} \psi_g \phi(y|\mu_g, \Sigma_g), \tag{1}$$

where ψ_s is the probability of membership in group g and $\phi(y|\mu, \Sigma)$ is the multivariate Gaussian distribution with mean vector μ and covariance matrix Σ. The covariance structure may be decomposed to test alternative shape, volume and orientation restrictions across the components (Fraley and Raftery, 2002). Note that the setup ignoring the panel nature of the data would estimate a univariate Gaussian mixture model. The multivariate version emerges since each period of measurement acts like a single variable in the multivariate framework.

McNicholas and Murphy (2010) allow for an extended Cholesky decomposition of the covariance matrix Σ in order to take into account the relationship between measurements at different points in time. This decomposition separates Σ into generalized autoregressive parameters T and innovation variances D (Pourahmadi, 1999), so that is possible to distinguish between the different sources of covariance:

$$\Sigma = T'D^{-1}T. \tag{2}$$

This setup for Σ has a very natural interpretation from least squares prediction theory. The least squares predictor of y_t given $y_{t-1}, \ldots, y_1$ is

$$\hat{y}_t = \mu_t + \sum_{s=1}^{t-1}(-\vartheta_{ts})(y_s - \mu_s) + \sqrt{d_t}\varepsilon_t. \tag{3}$$

The ϑ_{ts} are the sub-diagonal elements of T, while the d_ts are the diagonal elements of D and μ_s is the sth period mean. Taken together this format for constructing Σ allows for previous observations to influence current observations, something that is missing from static mixture analysis methods. As in static mixture modeling, the optimal solution can be obtained via the EM algorithm (Dempster et al., 1977).

3. Analysis

Our measurement of cross country output is GDP per capita taken from the Groningen Growth and Development Center and measured in constant 1990 prices for 101 countries over the period 1950–2010. Analyzing the data across each decade provides us with seven time points for each country. Prior to estimating the clustering of the distribution of world-wide output, we focus on the results from a traditional, static implementation.

Table 1 shows the stylized fact of emerging multimodality (see Quah, 1993). The LR bootstrap test suggests a single group of countries from 1950 until 1970, with the number of components ranging from two and four after 1970. The static clustering makes interpreting and discussing polarization and club σ-convergence difficult given the different numbers of groups emerging in each decade. In Pittau et al.'s (2010) analysis they always find three groups of countries which makes these types of statements more natural. The use of longitudinal mixture methods will help to alleviate these issues.

Turning our attention to the longitudinal mixture results, we test for the presence of a maximum of six components for each of the eight possible covariance decomposition structures listed in McNicholas and Murphy (2010).[2] The eight possible covariance structures for Σ arise over choices concerning the parameters in (3). The within component variance coefficients can be restricted to be equal across time, known as an isotropic constraint ($D_g = \delta_g I$), the across component variance coefficients in each time period can be equal ($d_t^g = d_t^m$ for $g \neq m$) and the across component autoregressive parameters can be equal ($\vartheta_{ts}^g = \vartheta_{tk}^m$ for $g \neq m$). Given that there exists a trade off between a greater description of the clustering process (with more groups) and more coefficients to estimate, we use two criteria to select the optimal solution, the traditional BIC and the ICL (integrated completed log-likelihood) that is a correction of BIC that penalizes components that are more spread out.

Table 1
LR bootstrap test computed over 1000 replications.

	$G = 1$	$G = 2$	$G = 3$	$G = 4$
1950	**0.139**			
1960	**0.764**			
1970	**0.894**			
1980	0.004	0.002	**0.407**	
1990	0.006	**0.781**		
2000	0.004	0.049	**0.181**	
2010	0.004	0.030	0.043	**0.481**

Notes: the table reports the probability of rejection of the null hypothesis of components equal to G segments, with an alternative hypothesis that the number of segments is greater than G.

Table 2
Log GDP per capita: longitudinal cluster characteristics.

	Cluster 1	Cluster 2	Cluster 3
Mean 1950	8.017	6.544	7.430
Mean 1960	8.256	6.768	7.810
Mean 1970	8.553	6.935	8.336
Mean 1980	8.851	6.942	8.754
Mean 1990	8.950	6.944	8.824
Mean 2000	9.149	7.035	8.947
Mean 2010	9.329	7.355	9.259
△ Variance 1960 (%)	10.8	20.3	3.2
△ Variance 1970 (%)	5.2	23.4	−15.7
△ Variance 1980 (%)	−15.2	−12.5	−8.0
△ Variance 1990 (%)	4.5	26.0	98.9
△ Variance 2000 (%)	3.2	48.3	79.9
△ Variance 2010 (%)	−15.6	27.0	−8.3
Overall change (%)	−5.6	213.1	172.8
Weight (%)	48.0	29.9	22.1

The best model determined using either the BIC or the ICL is a mixture with three components that restricts the generalized autoregressive parameters across components to be equal while allowing different component innovation variances, without isotropic constraints. The restriction on the generalized autoregressive parameters for Σ_g suggests that the components behave in a similar fashion over time, relative to previous realizations of output, but given different component means, certain clubs will inevitably have higher levels of output on average. The fact that the component innovation variances differ across time and across components also implies that the members of each component react differently to shocks than members in other components.

To study the estimated clusters in more detail Table 2 presents the cluster means over the seven decades as well as changes in the within group variance across the decades (Pittau et al., 2010, initially propose this idea). Component 1 represents mainly rich countries (containing oil producers, OECD and several Latin American countries), component 2 is primarily made up of the poorest nations (sub-Saharan African and some Asian countries) while component 3 is an intermediate group having some Asian, Eastern European and Middle East/North African countries. These groups are consistent with the univariate results of Pittau et al.

[2] We use the `longclust` R module, see McNicholas and Subedi (2012).