ARE FUTURE ENLARGEMENT CANDIDATE COUNTRIES CONVERGING WITH THE EU?

Christos Kollias[1], Petros Messis[2]

[1]Laboratory of Economic Policy and Strategic Planning, Department of Economics, University of Thessaly, Volos, Greece
[2]Department of Accounting and Finance, University of Macedonia, Thessaloniki, Greece

Abstract: The paper addresses the issue of convergence with the EU for nine countries: Albania, Bosnia and Herzegovina, Georgia, Moldova, Serbia, North Macedonia, Turkey, Montenegro and Ukraine. All that are at different stages of EU candidacy that could eventually lead to full membership. Some are officially recognized as candidate countries while others are at the stage of an association agreement. The presence of convergence is examined in terms of two macroeconomic indices: GDP per capita and GDP per person employed. Panel unit root tests as well as univariate unit root tests are estimated for the period 1997-2016. In broad terms, the empirical findings reported herein indicate a lack of convergence with the EU irrespective of the metric used. However, they indicate a process of in-group convergence mostly in terms of GDP per person employed.

Keywords: Convergence, panel unit root tests, EU candidate countries

JEL Classification: C23, O47, O52

1. Introduction

A long-standing pivotal policy objective of the EU has been to implement growth-inducing policies in lagging national economies or regions. Such policies aim to reduce interregional and/or national disparities and inequalities and thus promote cohesion and economic convergence among EU members and regions. Not surprisingly, the process of European integration has attracted a steadily growing body of literature that examines the presence (or not) of convergence between countries or regions in various spheres (*inter alia*: Belke *et al*. 2018; Firgo and Huber, 2014; Borsi and Metiu, 2015; Sondermann, 2014; Emvalomatis, 2017; Lyncker and Thoennessen, 2017; Belke and Schneider, 2013). If established, the degree and speed of convergence can be regarded as a measure of the effectiveness of the EU policies that aim to promote cohesion and decrease national or regional disparities by spurring growth in lagging regions (Apergis *et al*. 2010; Chapsa and Katrakilidis, 2014). In fact, as the EU enlarged with the accession of new member states that more often than not lagged in many spheres compared to the European core, EU cohesion policies emerged as the main convergence policy tool.

Gradually, such cohesion aiming policies steadily increased and specialized in scope. Moreover, they also started to include candidate states as they prepared for future accession. Catching-up during the pre-accession phase facilitates the process of integration in the EU when full membership is eventually achieved. As Bongardt and Torres (2013) note, one of the major attractions of EU membership to potential candidate countries has always been the opportunity it offers to catch up with EU income levels, living standards and social conditions. In the broader spirit of studies that empirically address the theme of EU convergence, this paper focuses on the countries that currently are at various stages in the process of EU candidacy. In particular, we examine whether Albania, Bosnia and Herzegovina, Georgia, Moldova, Serbia, North Macedonia, Turkey, Montenegro and Ukraine are catching-up with the EU in terms of two macroeconomic indices: GDP per capita and GDP per person employed.

All are candidates for accession in the next enlargements if and when eventually they take place given the current important challenges faced by the EU such as the planned withdrawal of the UK and the rise of Euroscepticism (*inter alia*: Danderstadt, 2014; Tsoukalis, 2014). Both datasets used in the empirical investigation that follows are drawn from the World Bank's World Development Indicators database and are expressed in PPPs. The former of the two is the typical income level index used to compare intercountry economic performance and developmental levels. It is invariably used whenever convergence is examined. The latter metric, i.e. GDP per person employed, is an index that captures labour productivity since it measures output per unit of labour input. Rising productivity is an important contributor to growth. It facilitates catching-up efforts by economies with income levels that lag behind the EU's averages as these countries progress in the road towards eventual membership.

In line with similar methodological strategies adopted by recent studies, we employ two averages as benchmarks against which the convergence hypothesis is investigated for the period 1997-2016 (Ceylan and Abiyev, 2016; Tsanana and Katrakilidis, 2014). One benchmark is the average of the member states that thus far make up the core of European integration. That is the EU15. The average is calculated for each of the two variables i.e. per capita GDP and GDP per person employed. In line with the aforementioned studies, the other benchmark is the corresponding average for each of the two indices for the group of the nine countries (AV9) examined here. The first benchmark is used to investigate the convergence hypothesis in relation to the EU. Moreover, given that all nine are at various stages in the process towards EU accession, it is also possible that this process is generating a comparatively stronger in-group convergence. Hence the choice to add this in-group benchmark in the empirical tests that follow. We calculate the difference $(y_i - \bar{y})$ between the natural logarithm of the time series of per capita GDP (and GDP per person employed) for each country *i* and the natural logarithm of the benchmark index. The section that follows includes a brief comparative overview of the nine countries for which the convergence hypothesis is examined here. Section three outlines the methodology employed in the empirical investigation. The findings are presented and discussed in section four. Section five concludes the paper.

2. Structural characteristics and comparative economic performance

As already noted, the nine countries examined here form a rather heterogeneous group in terms of their links to the EU. They are at different phases of their potential accession and EU membership. Turkey is by far the country with the longest association and candidacy history. The initial treaty – the Ankara Association Agreement - dates back to 1963 (Icoz, 2011; Phinnemore and İçener, 2016). In other words, it is more than half a century old. It was signed at a time when some of the current EU members did not exist as independent state entities. Turkey formally applied for membership in 1987. The candidate country status was awarded to Turkey in 1999 and accession negotiations started in 2005. However, their progress is rather sluggish and frequently is slowed down or halted by political developments. Albania, Serbia, Bosnia and Herzegovina and North Macedonia were identified as potential candidate countries in 2003. Subsequently, they were awarded the status of candidate countries except Bosnia and Herzegovina. In particular, Albania formally applied in 2009 and was officially awarded the status of candidate country in 2014. Serbia was granted the candidate country status in 2012 and accession negotiations started in 2014. North Macedonia, applied in 2004 and was declared a candidate country in 2005. These Western Balkans countries will probably achieve full membership over the next few years. Montenegro is the newest independent state of the sample. Its independence was declared in 2006. Two years later, in 2008, it applied for EU membership. It was officially declared a candidate country in 2010 and accession negotiations started in 2012. The remaining three countries, i.e. Georgia, Ukraine and Moldova are at an earlier stage of their association with the EU. They all have association agreements and are regarded as potential future candidates for EU membership. The association agreements with Ukraine and Moldova were signed in 2014, while that with Georgia in 2016.

The nine countries present significant differences in terms of their structural economic characteristics compared to the corresponding EU averages as well as between them. Table 1, offers a summary view of their respective economies. For comparison purposes, the corresponding EU indicators are also presented. As can be seen, noteworthy differences are identifiable in terms of the sectoral structure of the economy both within the sample, as well as compared to the EU. The share of the agricultural sector in GDP is probably the feature that stands out the most. In all nine countries, the value added in agriculture as a share of GDP is considerably higher when compared to the EU average. It is also appreciably higher when compared with Bulgaria and Romania, the two EU member states with the highest share[1]. Significant differences are also present within the group. Albania and Moldova are the two countries with the highest shares of agricultural value

[1] The corresponding 1997-2016 average is 8.9% for Bulgaria and 9.5% for Romania

added to GDP and with lowest shares of value added in the manufacturing sector. Ukraine, Turkey and Serbia are the ones with the highest manufacturing shares. Less heterogeneous is the picture within the group in the case of the services sector. Notable is the difference with the EU service sector average in all cases except Montenegro.

A similarly varied picture appears in terms of trade openness and the ratio between exports and imports. Bosnia and Herzegovina and Moldova are the two countries with the highest trade deficit. Turkey and the Ukraine are the two with the lowest trade deficit. As a percentage of GDP, their imports are marginally higher than exports. Given that urbanization is associated with more developed countries, the share of the rural population among the total population is a strong indicator of the developmental differences between most of the countries in the group compared to the EU average. In Bosnia and Herzegovina, Albania and Moldova, more than half of the population lives in rural areas (Table 1). During the period examined here, Albania followed by Turkey and Montenegro are the three countries that exhibit the largest decline in the share of rural population during 1997-2016. It declined from 60% to 41.6% in Albania, from 36.8% to 26.1% in Turkey and from 44.5% to 35.8% in Montenegro. In Serbia and the Ukraine, the reduction in the share of rural population is smaller: from 47.6% to 44.3% and 33% to 30.1% respectively. In the remaining four countries - Bosnia and Herzegovina, Georgia, Moldova and North Macedonia – the share of rural population has more or less remained unchanged during 1997-2016.

Table 1: Structural characteristics (Averages 1997-2016)

	Agriculture	Manufacturing	Services	Exports	Imports	Rural population
	Value added as % of GDP			*as % of GDP*		*as % of total*
Albania	23.4	7.2	53.6	23.0	44.9	51.4
Bosnia and Herzegovina	10.5	12.7	62.7	30.2	66.3	60.7
Georgia	15.4	11.7	60.7	31.6	50.6	47.1
Moldova	19.1	15.4	62.0	46.6	80.7	54.7
North Macedonia	11.8	13.0	62.0	39.1	56.2	42.3
Montenegro	10.3	7.6	69.1	40.1	64.9	38.5
Serbia	13.0	19.1	56.5	28.7	43.9	45.5
Turkey	10.0	19.5	60.0	22.3	25.2	31.4
Ukraine	11.7	19.8	56.0	49.5	51.9	31.8
EU average	1.9	16.8	71.9	37.0	35.6	27.0

For convergence in income levels and living standards to be achieved, a candidate country or a lagging member state, needs to grow systematically faster in order to catch up with the EU's average level. In terms of growth performance, the nine countries of the sample have achieved an annual average considerably higher than the corresponding EU annual average. This can be seen in Table 2 where two growth metrics are presented: the average rate of change of GDP and that of per capita GDP for 1997-2016. The only exception to this general observation is Ukraine. With an average annual GDP growth of 1.9%, its growth performance is the group's lowest. Nevertheless, it is marginally higher than the EU average of 1.7% for the same period. The domestic acute political tensions, frictions and instability that have marred Ukraine over the past years along with the volatile geopolitical environment that culminated in the conflict with Russia and the annexation of Crimea can be cited as explanatory factors for this relative underperformance compared to the rest of the group. Bosnia and Herzegovina emerges as the growth champion of the group with an average annual GDP growth of 5.9%. A strong contributor to this performance are the high rates recorded immediately after the end of the Bosnian War. The establishment and consolidation of peace with the concomitant reconstruction that followed yielded many tangible developmental benefits mirrored in the growth performance of the Bosnia and Herzegovina economy. Georgia and Turkey follow in terms of GDP growth performance (5.5% and 4.7% respectively). Albania has the fourth highest average annual growth of GDP (4.1%). Overall, it appears that the economies of all the nine candidate or potential candidate countries are growing at a faster rate than the EU average. This is the case for both growth performance metrics that are presented in Table 2.

Table 2: GDP and per capita GDP growth rates (Average 1997-2016)*

	GDP growth	*per capita GDP growth*
Albania	4.1	4.6
Bosnia and Herzegovina	5.9	6.3
Georgia	5.5	6.6
Moldova	3.5	3.7
North Macedonia	2.9	2.7
Montenegro	2.3	2.2
Serbia	2.7	3.1
Turkey	4.7	3.2
Ukraine	1.9	2.8
EU average	1.7	1.5

* For Montenegro 1998-2016
Source: World Development Indicators

In terms of income levels, all nine countries lag behind the EU average despite the higher growth rates. In terms of per capita GDP in 2016, out of the nine countries, Turkey and Montenegro are the two with the highest per capita GDP followed by Serbia. Their respective GDP per capita is $23,756; $15,725 and $13,723. Still, even these three significantly trail behind the EU average per capita GDP of $36,330 for the same year. Perhaps a reminder is in order here. The average per capita GDP in the EU hides the significant income level differences that do exist between the current member-states. Thus, if the per capita GDP of the nine countries in the sample is compared to that of countries such as Bulgaria and Romania, the income gap is less pronounced. Bulgaria and Romania are the two EU members with the lowest per capita GDP: $17,000 and $20,545 respectively[2]. In other words, their per capita income level is comparable to countries such as Turkey or Montenegro in our sample. Moldova is the country that has the lowest per capita GDP out of the nine, followed by Ukraine and Georgia. The same broad picture emerges if we compare the nine in terms of GDP per person employed. Once again, in 2016 Turkey and Montenegro are the two countries with significantly higher values compared to the rest of the group. At the other end of the spectrum, Ukraine, Georgia and Moldova have the lowest GDP per person employed in 2016 (Table 3).

Table 3: Per capita GDP and GDP per person employed (1997 & 2016)

	Albania	Bosnia and Herzegovina	Georgia	Moldova	North Macedonia	Montenegro	Serbia	Turkey	Ukraine	EU
GDP per capita										
1997	4214	4750	2946	2507	7779	10531	8147	13776	4497	27689
2016	11359	11327	9277	4944	13055	15725	13723	23756	7668	36330
GDP per person employed										
1997	11011	16967	6206	4952	29616	30615	16865	40487	10266	66604
2016	31329	39464	18273	12619	36626	47168	29963	59288	16192	81844

* Constant 2011 international $ in PPPs
** Constant 2011 $ in PPPs
Source: World Development Indicators

2 For 2016, in PPPs and constant 2011 international $ as the data in Table 3.

The two macroeconomic metrics presented in Table 3 are used in the empirical tests that follow in order to examine whether or not this group of nine countries is converging with the EU. As pointed out above, some of them are already officially recognized as candidate states while others are at earlier stages of the road to eventual EU accession. Of the two indices, per capita GDP is the one invariably used by studies that focus on income level convergence between countries (*inter alia*: Chapsa *et al*. 2015; Ayala *et al*. 2013; Vojinovic *et al*. 2010; Dobrinsky, 2003; Martin and Sanz, 2003). The other index, i.e. GDP per person employed, is widely used as an index of labour productivity measuring output per unit of labour input. As noted in the introduction, rising productivity is an important contributor to growth. It spurs the catching-up efforts by countries with lower income levels compared to the EU as it is the case for all the nine countries examined here (Table 3).

3. Empirical methodology: an outline

A number of different methodologies ranging from *σ* and *β*-convergence to unit root tests can be used in order to test for convergence. The latter have emerged as a prevalent tool in empirical studies that examine convergence hypotheses (*inter alia*: Lau *et al* 2016; Beyaert and Camacho, 2008; Dawson and Strazicich, 2010). When income convergence across countries is examined, panel unit roots present a useful econometric tool on which reliable inferences can be drawn. The EU and OECD members as well as Latin American countries have attracted a fair share of the income convergence literature with many studies opting to use unit root tests (*inter alia*: Ceylan and Abiyev, 2016; Ayala *et al*. 2013; Galvão and Gomes, 2007; Strazicich *et al*. 2004). In line with recent studies, we apply a battery of panel unit root tests for the period 1997-2016 to address the convergence issue

(Beyaert and Camacho, 2008; Chapsa *et al*. 2015; Tsanana and Katrakilidis, 2014). We employ the deviation of per capita GDP and GDP per person employed from the EU15 average and the average of the group of nine countries examined here (AV9). Hence, $(y_i - \bar{y})$ demonstrates the difference of the natural logarithm of each series with the benchmark index. Testing for convergence among countries indicates the identification of time series whose means and variances remain constant over time (Evans, 1998). In what follows, we briefly describe the tests that will be conducted in the next section. We start with the conventional unit root tests, before we move to the presentation of the tests that allow structural breaks in the series.

Unit root tests: a primer

One of the most widely used tests for stationarity is the Augmented Dickey-Fuller (ADF) (1979) test but has been criticized by Perron (1989) for failing to allow for structural breaks as will be discussed in the following subsection. The limited power in small samples of the Dickey Fuller and Augmented Dickey Fuller (1979, 1981) tests gave also rise to unit root tests with panel data. These tests present higher testing power by letting $N \to \infty$ for fixed *T*. Hence, the analysis of panel data increases the size of data sample since they use both cross sectional and time series observations. Different univariate and panel data unit root tests are used in order to increase the power of the empirical results. Such tests are estimated under the assumption that all the examined series have a common Autoregressive (AR) structure. For a panel dataset, an AR(1) process has the following form:

$$y_{it} = \rho_i y_{i,t-1} + X'_{it}\,\delta_i + u_{it}\,,$$

with *i=1*,…,N being the cross-section series, in our case the countries, that are observed over the t=1,…,T periods. X_{it} includes any fixed effects or individual trends, ρ_i are the autoregressive coefficients and u_{it} are the i.i.d. error terms. All tests we examine below, apart from that of Im, Pesaran and Shin (2003), assume that ρ_i is identical across cross-sections.

Levin, Lin and Chu (2002) propose a more powerful panel unit root test (henceforth LLC) than performing individual unit root tests for each cross section (Baltagi, 2005). The null hypothesis is that each individual time series contains a unit root against the alternative that each time series is stationary. Hence, the null hypothesis is $H_0 : \rho_i = 1$ for i=1,…,N against $H_1 : -1 < \rho_i = \rho < 1$ for i=1,…,N. Thus, the first-order serial correlation coefficient ρ is required to be identical in all units since the test statistic is computed in a pooled model (Hlouskova and Wagner, 2006). According to LLC, the test procedure follows three steps. In the first step, for each individual series an ADF regression of the form

$$\Delta y_{it} = \rho_i y_{it-1} + \sum_{L=1}^{p_i} \gamma_{iL} \Delta y_{it-L} + \delta_{mi} d_{mt} + u_{it}, \qquad \text{m=1,2,3} \tag{1}$$

is performed where d_{mt} indicates the vector of deterministic variables and δ_m demonstrates the corresponding vector of coefficients for each particular model m (*i.e.* m=1,2,3). Hence, $d_{1t} = \{\varnothing\}$ constitutes the first case with no deterministic component. $d_{2t} = \{1\}$ is the second case and allows for intercept in the data generating process, whereas $d_{3t} = \{1, t\}$ is the third case which contains intercept and time trend. Finally, *L* is the lag operator. In the second step, the ratio of long-run to short-run standard deviations is estimated while the third step requires the estimation of the following adjusted *t*-statistic:

$$t_\rho^* = \frac{t_\rho - \mathrm{NT} \hat{S}_N \sigma_u^{-2} \sigma_\rho \mu_{mT}^*}{\sigma_{mT}^*} \tag{2}$$

where σ^*_{mT} and μ^*_{mT} are the standard deviation and the mean adjustments respectively provided for various panel dimensions in Table 4 for LLC. NT is the total number of observations, and $\hat{S}_N = \frac{1}{N}\sum_{i=1}^{N} s_i$ with $s_i = \sigma_{yi} / \sigma_{ei}$. The null hypothesis is rejected if the adjusted t-statistic $t^*_\rho < criticalvalue$ of the standard normal distribution for the cases m=2,3.

The test of Im, Pesaran and Shin (henceforth IPS) (2003) is not as restrictive as that of LLC since it does not require ρ to be homogeneous across *i.* The null hypothesis is that all series in the panel follow a unit root process, i.e., $H_0 : \rho_i = 0$ for all *i.* The alternative hypothesis allows some (but not all) of the individual series to have unit roots, i.e., $H_1 : \rho_i < 0$ at least for an *i.* The IPS *t*-bar statistic, which is defined as the average of the individual ADF statistics, has the following shape (Baltagi, 2005):

$$\bar{t} = \frac{1}{N}\sum_{i=1}^{N} t_{\rho i} \tag{3}$$

where $t_{\rho i}$ is the individual t-statistic for all i cross-section units. The null hypothesis is rejected if $\bar{t} < Z_{criticalvalue}$.

The Breitung (2000) testing procedure constitutes three steps and develops a *t*-statistic, which follows a standard normal distribution. The first step is similar to that of LLC, except that deterministic terms are excluded. A regression of the first differences of y_{it} (i.e. Δy_{it}), on $\Delta y_{i,t-L}$ is performed and the residuals e_{it} are obtained. A further regression is estimated: that of $y_{i,t-1}$ on $\Delta y_{i,t-L}$ in order to obtain the residuals v_{it}. Next, the residuals e_{it} are transformed using the forward orthogonalization transformation (i.e. to each of the first *T-1* observations, the mean of the remaining future observations available in the sample is subtracted) taking e_{it}^{*}. The final step is to run the pooled regression $e_{it}^{*} = \rho v_{i,t-1}^{*} + u_{it}^{*}$ and obtain the t-statistic for $H_0 : \rho = 0$ which is asymptotically N(0,1) distributed. Once again, the null hypothesis is rejected if $\bar{t} < Z_{criticalvalue}$.

The Hadri (2000) test is a residual-based Langrange multiplier (LM) test. The Ordinary Least Squares (OLS) residuals are obtained from regressing y_{it} on a constant (referred to as case two) or a constant and time trend (referred to as case three). The null hypothesis is that there is no unit root in any series, implying stationarity and the alternative hypothesis is that the panel has a unit root. Assuming that the individual time series y_{it} are generated by $y_{it} = r_{it} + \varepsilon_{it}$ with i=1,…,N and t=1,…,T (a time trend β_{it} can also be added) where r_{it} is a random walk, i.e., $r_{it} = r_{i,t-1} + u_{it}$, the stationarity hypothesis is $H_0 : \sigma_u^2 = 0$. Hence, if variance u_{it} is zero, then r_{it} becomes a constant and thus y_{it} is stationary. The LM statistic is given by

$$LM_H = \frac{1}{N}\left(\sum_{i=1}^{N} \frac{1}{T^2} S_{it}^2\right) / \sigma_{\varepsilon}^{2} , \qquad (4)$$

where S_{it} denote the partial sum of OLS residuals and $\sigma_{\varepsilon}^2 = \frac{1}{\mathrm{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{it}^2$. The rejection of the null is confirmed when $LM_H < Z_{criticalvalue}$. We also apply tests with individual unit root processes starting with Fischer type tests (ADF and Phillips-Perron). These specific tests proposed by Maddala and Wu (1999) and Choi (2001) combine the *p*-values from unit root tests for each cross-section *i* and constitute an alternative approach of Fischer's (1932) results. This test has a χ^2 distribution and if *pi* are the probability values then

$$P = -2\sum_{i=1}^{N} \ln(pi) \ . \qquad (5)$$

Choi (2001) proposes an inverse normal test that has the following form:

$$Z = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Phi^{-1}(\text{pi}) \ , \tag{6}$$

with Φ being the standard normal cumulative function. The asymptotic χ^2 and the standard normal statistics are reported using ADF and PP unit root tests. The null hypothesis is $H_0 : \rho_i = 0$ against the alternative $H_1 : \rho_i < 0$ that indicates stationarity.

Tests for structural breaks

Unit root tests with structural break(s) are also used since it is important to take into account structural breaks in the series examined. Perron (1989) suggests that structural change in time series can influence the results of tests for unit roots. Hence, he proposes a test for unit root hypothesis against trend stationarity with an exogenous structural break. However, the method of assuming the break date as known ex ante has been considered inappropriate by several authors in the subsequent theoretical and empirical literature. For example, Zivot and Andrews (1992) (ZA test in the results section), among others, propose a unit root test that treats the breakpoint as endogenous and hence the null of a unit root is less favorable. The ZA test is a sequential test that utilizes a different dummy variable for each possible break date.

Lumsdaine and Pappell (1997) further developed the ZA test to allow for two endogenous breaks under the alternative hypothesis (LP test in the results section). However, Lee and Strazicich (2003) (LS test in the results section) provide a minimum Langrange Multiplier test with breaks in the level and trend, arguing that spurious regression problems may arise similar to that of ZA test with a break under the null hypothesis. The ZA test is a sequential Dickey-Fuller unit root test and accommodates three alternative models for possible breaks. Model A stands for a break only in the intercept, model B accounts for break only in trend while model C incorporates the possibility of a change in the intercept as well as in the trend. Model A that is going to be used[3] in this paper is given as follows:

$$y_t = \mu + \theta DU_t(T_b) + \beta t + ay_{t-1} + \sum_{j=1}^{k} c_j \Delta y_{t-j} + e_t \qquad (7)$$

Where DU_t is a dummy variable capturing a break in the intercept. The dummy variable $DU_t = 1$ if $t>Tb$ and 0 otherwise. T_b denotes the time of break and it is chosen to minimize the one-sided t-statistic for testing $\alpha=1$ in (7). The null hypothesis is rejected if the coefficient is statistically significant.

The Lumsdaine and Papell (1997) test tries to capture two structural breaks. They advocate that taking into account two structural breaks for testing unit root hypothesis is more powerful compared to the ones that allow for one break. The model that allows two structural breaks in the intercept has the following form:

$$\Delta y_t = \mu + \theta_1 DU1_t(TB_1) + \theta_2 DU2_t(TB_2) + \beta_t t + \alpha y_{t-1} + \sum_{j-1}^{k} c_j \, \Delta y_{t-j} + e_t \qquad (8)$$

The two indicator dummy variables $DU1_t$ and $DU2_t$ capture structural changes in the intercept occurring at times TB_1 and TB_2. Hence, $DU1_t = 1$ if $t > TB_1$ and 0 otherwise and $DU2_t = 1$ if $t > TB_2$ and 0 otherwise. The null hypothesis of a unit root without structural change is $\alpha=0$ and the alternative hypothesis is that α is significantly different from zero in equation (8). The break points are chosen so that the maximum evidence against the unit root null (minimum *t*-statistic on α) is obtained.

[3] It should be noted here that there is no formal criterion that can be used in order to choose among the three models (Altinay, 2005). However, Ben-David and Papell (1997) report that if a trend is present in a series then using in the estimations a model without trend is possible to fail to capture important data characteristics. Most of the series used in analysis here do not exhibit an upward or downward trend. Nevertheless, we also estimate Model C (not reported for reasons of brevity but available from the authors upon request) but the general findings and concomitant conclusion dot not change in any substantive manner.

Finally the Lee and Strazicich (2003) minimum LM test procedures (LS test in the results section) comprise alternative unit root tests that are unaffected by structural breaks under the null. The authors report in their study that one important issue common to the ZA and LP tests is that both of them assume no break(s) under the unit root null hypothesis. It means that rejecting the null does not necessarily imply rejection of a unit root but rather rejection of a unit root without breaks. The procedure allowing for two structural breaks in level starts by assuming that the DGP of the series y_t is: $y_t = \delta' Z_t + \varepsilon_t$ with $\varepsilon_t = b\varepsilon_{t-1} + u_t$, where Z_t constitutes a vector with all exogenous variables while the residuals follow the classical assumptions, i.e., $\varepsilon_t \sim iidN(0,\sigma^2)$. The model (A in our case) that allows for two shifts in the level has the following regressors in $Z_t = [1, t, D_{1t}, D_{2t}]'$. Hence, $D_{it} = 1$ if $t > TB_i + 1$ for i=1,2, and 0 otherwise. Once again, TB_i denotes the time period of breaks. The two-break minimum LM unit root test statistic can be obtained from the following regression[4]:

$$\Delta y_t = \delta' \Delta Z_t + \varphi S_{t-1} + \varepsilon_t, \qquad (9)$$

Where *Δ* is the first difference operator, $S_t = y_t - \psi_x - Z_t \delta$ is the de-trended series with *t=2,…,T* being the time parameter. δ constitutes a vector of coefficients after regressing Δy_t on ΔZ_t. ψ_x is the difference between y_1 and $Z_1 \delta$ where y_1 and Z_1 are the first observations of y_t and Z_t respectively. The LM t-statistic for $\varphi = 0$ which is the unit root hypothesis is given by:

[4] It is possible to correct for possible serial correlation by including an additional term in the model such as ΔS_{t-1} (Amsler and Lee, 1995).

$\rho = \mathrm{T} \cdot \varphi$ and $\tilde{\tau} = t - statistic$. Through a grid search utilization, the min LM unit root test determines the break points $\lambda_i = TB_i / T$ as follows: $LM_\rho = \inf_\lambda \rho(\lambda)$, $LM_\tau = \inf_\tau \tilde{\tau}(\lambda)$. The break point(s) are determined to be where the minimization of the test statistic occurs and the method is analogous to that in the ZA and LP tests. Finally, in line with the methodological approach adopted by Tsanana and Katrakilidis (2014), the panel LM unit root test proposed by Im *et al*. (2005) is calculated by averaging the univariate LM unit root t-test statistics for each country as follows:

$$\overline{LM}_{NT} = \frac{1}{N}\sum_{i=1}^{N} LM_i^T \quad . \tag{10}$$

After normalization, we take the following result:

$$\Gamma_{\mathrm{NT}} = \frac{\sqrt{\mathrm{N}}(\overline{LM}_{NT} - E(LM_i^T))}{\sqrt{\mathrm{V}(LM_i^T)}} \Rightarrow N(0,1)\,, \tag{11}$$

where $E(LM_i^T)$ and $\mathrm{V}(LM_i^T)$ denote the expected value and variance of LM_i^T statistic respectively under the null hypothesis. Simulated values of means and variances for different time periods, starting from *T=10*, can be found in Im *et al*. (2005). The asymptotic distribution of the test is not affected by the presence of structural breaks and is standard normal.

The results of the various tests are presented and discussed in the section that follows. As invariably is the case, the estimations of the various unit root tests yield results that are not uniform across all tests and differ given the different assumptions and powers of each individual test. For instance, as pointed out by Tsanana and Katrakilidis (2014), conventional unit root tests are affected by the limited number of observations, a weakness that is not present in the univariate and panel LM unit root tests.

4. Empirical findings

In this section, we apply the unit root tests briefly presented in the previous section to the series under investigation, i.e. GDP per capita and GDP per person employed for the nine countries. The data used are drawn for the *Word Development Indicators* database and cover the period 1997-2016. The period was strictly determined by the availability of data for all the nine countries for both indicators used here. As noted in the introduction, we use two benchmarks. The first one is the EU15 average in both metrics employed to explore the convergence issue. The other one is the average of the group of the nine countries examined here (AV9). We calculate the differences in income levels in terms of GDP per capita and productivity as reflected in the latter of the two aforementioned series (GDP per person employed) with the benchmark by taking the difference of the natural logarithm of each series, so that $(y_i - \bar{y})$ being the difference of country *i* with the benchmark index.

Table 4: Panel unit root tests 1997-2016

Panel A: GDP per capita							
Tests		LLC	IPS	Breitung	ADF-Fischer	PP-Fischer	Hadri
Benchmark	EU15	-1.114	-0.948	1.373	27.83	23.75	5.403
		(0.132)	(0.171)	(0.915)	(0.163)	(0.163)	(0.000)
	AV9	-2.076	-2.024	-0.619	36.28	47.16	6.407
		(0.019)	(0.021)	(0.268)	(0.006)	(0.000)	(0.000)
Panel B: GDP per person employed							
Tests		LLC	IPS	Breitung	ADF-Fischer	PP-Fischer	Hadri
Benchmark	EU15	1.482	1.150	0.164	13.24	9.396	5.252
		(0.931)	(0.874)	(0.565)	(0.777)	(0.949)	(0.000)
	AV9	-1.749	-1.762	-0.676	31.16	35.21	5.181
		(0.040)	(0.039)	(0.249)	(0.028)	(0.009)	(0.000)

Notes: *p*-values in parentheses. The tested hypothesis is the unit root but Hadri's test adopts stationarity as the null hypothesis. Probabilities for Fischer tests are computed using an asymptotic Chi square distribution. All other tests assume asymptotic normality. LLC stands for the Levin, Lin and Chu (2002) test and IPS for the Im, Pesaran and Shin (2003) test as presented in the preceding section.

The results of the five different panel unit root tests (LLC, IPS, Breitung, ADF-Fischer, PP-Fischer) and the panel stationarity test proposed by Hadri (2000) are presented in Table 4. In the panel unit root tests individual country-specific fixed effects are used. As pointed out by Baltagi (2005), this specification is more appropriate when a specific set of *N* countries is examined. As a broad observation, when the benchmark is the average EU15 per capita GDP, the findings indicate lack of convergence in five out of the six testing procedures. Specifically, lack of convergence is suggested by LLC, IPS, Breitung, PP-Fischer and Hadri tests at the 10% level of significance. On the contrary, a process of in-group convergence in terms of per capita GDP is observed in four cases when the average of the nine examined countries (AV9) is used as the benchmark. A similar finding is the case for the GDP per person-employed index compared to the EU15 average. All the tests point to the absence of any convergence process (Table 4). However, this is not the case when the benchmark is the AV9. In this case, a process of in-group convergence seems to be the dominant finding. Only two tests - Breitung and Hadri – suggest absence of convergence. Apart from these two exceptions, all testing procedures indicate a process of in-group convergence in terms of GDP per person employed but not with the EU15 average. However, as already noted in the previous section, failing to take into account the possible presence of structural breaks in the series can yield misleading results and lead to erroneous inferences. Consequently, following Tsanana and Katrakilidis (2014), in the next step of the empirical strategy we use the tests outlined in the previous section that allow for endogenously determined breaks. In particular, we apply the univariate tests (ZA, LP, LS)[5] that allow for possible shifts in the level. As Altinay (2005) notes, the choice of lag length (*k*)

[5] ZA for the Zivot and Andrews (1992) test, LP for the Lumsdaine and Pappell (1997) test and LS for Lee and Strazicich (2003) test as presented in the preceding section.

is important for all unit root tests with structural breaks. The general to specific procedure in lag length selection is what a number of authors suggest and we opt to follow here (Perron, 1989; Zivot and Andrews, 1992; Lee and Strazicich, 2003).

Table 5: Min LM unit root test with no structural breaks

	GDP per capita				GDP per person employed			
Benchmark	EU15		AV9		EU15		AV9	
Countries	LM-stat	Lag	LM-stat	Lag	LM-stat	Lag	LM-stat	Lag
Albania	0.646	(0)	-0.909	(2)	-1.607	(0)	-2.423	(2)
Bosnia and Herzegovina	0.403	(0)	-0.039	(0)	-1.001	(0)	-1.245	(0)
Georgia	-1.634	(0)	-2.224	(0)	-2.362	(1)	-1.775	(0)
Moldova	-2.565	(3)	-2.372	(0)	-2.616	(0)	-2.663	(0)
North Macedonia	-1.572	(1)	-3.795***	(1)	-1.461	(0)	-6.062***	(1)
Serbia	-2.805*	(0)	-2.854*	(1)	-2.278	(1)	-2.526	(1)
Turkey	-2.016	(2)	-0.680	(0)	-3.598**	(1)	-4.402***	(3)
Ukraine	-1.817	(0)	-1.262	(0)	-1.355	(0)	-1.293	(0)
Montenegro	-0.833	(0)	-2.253	(1)	-1.210	(0)	-0.697	(0)

***, **, * indicate the 1%, 5% and 10% levels of significance. The corresponding critical values for the test are -3.63, -3.06 and -2.77. The chosen lag length is in parentheses.

We start with Table 5, where we present the results of the min LM unit root test of Lee and Strazicich (2003) test for the individual time series without structural breaks. From the reported values, we observe that using as benchmark the EU15 for the GDP per capita series, the null hypothesis of a unit root (i.e. absence of convergence) is rejected only for Serbia. For the same series and when the AV9 is the benchmark, once again the null of no convergence is rejected for Serbia at the 10% level of significance as well as for North Macedonia at the 1% level of significance. For the rest of the countries, this lack of stationarity as Tsanana and Katrakilidis (2014) observe, may be due to the limited explanatory power of the test under the presence of significant structural breaks. The results do not change significantly in the case of the GDP per person-employed series. Convergence in terms of this index is established for Turkey both with the EU15 and with the AV9 benchmark. The same applies for North Macedonia but only in the case of the AV9 benchmark. In order to check the robustness of the results obtained thus far, we proceed with the estimation of the Zivot and Andrews (1992) ZA and Lee and Strazicich (2003) LS tests with one structural break on the individual time series.

Table 6: Unit root tests with one structural break for GDP per capita

Panel A: EU15						
Countries	ZA t-stat	Lags	Break	LS-stat	Lags	Break
Albania	-2.187	(0)	2012	-0.806	(2)	2010
Bosnia and Herzegovina	-5.755***	(0)	2004	0.2471	(0)	2013
Georgia	-4.010	(2)	2014	-2.809	(1)	2003
Moldova	-4.266	(3)	2014	-3.883**	(3)	2012
North Macedonia	-3.035	(1)	2006	-1.752	(1)	2004
Serbia	-3.248	(0)	2003	-3.293*	(2)	2007
Turkey	-4.054	(2)	2010	-2.286	(2)	2005
Ukraine	-1.852	(0)	2002	-3.034	(3)	2011
Montenegro	-8.465***	(1)	2005	-1.721	(1)	2012
Panel B: AV9						
Countries	ZA t-stat	Lags	Break	LS-stat	Lags	Break
Albania	-4.581*	(2)	2008	-1.354	(2)	2008
Bosnia and Herzegovina	-7.917***	(1)	2010	-1.163	(3)	2010
Georgia	-4.982**	(2)	2005	-2.644	(0)	2001
Moldova	-6.576***	(0)	2006	-3.386*	(0)	2012
North Macedonia	-3.674	(1)	2002	-5.254***	(1)	2001
Serbia	-3.043	(0)	2001	-3.781**	(0)	2013
Turkey	-2.834	(0)	2010	-2.485	(2)	2009
Ukraine	-2.135	(0)	2000	-1.977	(1)	2008
Montenegro	-4.253	(1)	2006	-2.589	(1)	2012

Notes: The critical values for the Zivot and Andrews (1992) ZA test are -5.34, -4.80 and -4.58 at the 1%, 5% and 10% level respectively. The LS-stat are the corresponding critical values for the Lee and Strazicich (2003) test with one break on the intercept are -4.239, -3.566 and -3.211.

The results for the GDP per capita series are shown in Table 6 and in Table 7 the corresponding ones of the GDP per person-employed. In Table 6 in Panel A, based on the ZA test convergence can be inferred only in two out of the nine countries: Montenegro and Bosnia and Herzegovina. The LS test results also suggest convergence only for two countries: Moldova and Serbia. It would appear that on balance, the tests do not yield strong evidence in favor of a convergence process in terms of per capita GDP. Slightly stronger is the evidence in favor of an in-group convergence hypothesis in terms of per capita GDP (Panel

B, Table 6). In Table 7 where the unit root tests for the GDP per person employed are presented, the null is rejected in the case of Moldova (by both the ZA and LS tests) for Bosnia and Herzegovina, Albania, and Georgia (ZA test) and for Serbia and North Macedonia (LS test).

Table 7: Unit root tests with one structural break for GDP per person employed

Panel A: EU15						
Countries	ZA t-stat	Lags	Break	LS-stat	Lags	Break
Albania	-2.526	(0)	2014	-2.181	(1)	2001
Bosnia and Herzegovina	-3.942	(0)	2003	-1.331	(3)	2005
Georgia	-3.962	(1)	2004	-3.152	(1)	2002
Moldova	-5.258**	(0)	2002	-3.741**	(2)	2004
North Macedonia	-3.797	(1)	2003	-1.868	(0)	2001
Serbia	-3.435	(1)	2013	-2.797	(2)	2004
Turkey	-8.131***	(3)	2003	-4.084	(2)	2001
Ukraine	-1.638	(0)	2002	-2.112	(1)	2001
Montenegro	-3.419	(0)	2006	-2.339	(3)	2006
Panel B: AV9						
Countries	ZA t-stat	Lags	Break	LS-stat	Lags	Break
Albania	-5.167**	(3)	2008	-2.974	(2)	2003
Bosnia and Herzegovina	-6.370***	(3)	2010	-2.354	(3)	2005
Georgia	-4.329	(1)	2006	-2.637	(1)	2002
Moldova	-7.034***	(0)	2002	-2.953	(0)	2009
North Macedonia	-6.296***	(1)	2013	-7.382***	(1)	2004
Serbia	-4.650*	(1)	2013	-3.035	(1)	2002
Turkey	-4.154	(3)	2008	-6.978***	(3)	2003
Ukraine	-2.325	(0)	2000	-2.308	(1)	2008
Montenegro	-2.098	(0)	2006	-0.837	(0)	2004

Notes: The critical values for the ZA test are -5.34, -4.80 and -4.58 at the 1%, 5% and 10% level respectively. The corresponding critical values for the LS test with one break on the intercept are -4.239, -3.566 and -3.211.

Clearly, the results and the concomitant inferences are sensitive to the test adopted in order to investigate the presence of convergence. The break years identified by the tests vary between countries. In a number of cases, they appear to broadly coincide with the recent international economic crisis while in other cases with country specific reasons. For example, in the case of Montenegro, the break year – 2005 – identified by the ZA tests almost coincides with the country’s independence in 2006. The 2005 break identified in the case of Georgia (Panel B) could be tentatively associated with the troubles in South Ossetia and Abkhazia in 2004 and 2005 and/or the trade related problems with Russia in 2006. A similar explanation can be cautiously proposed for Moldova and the break years identified in Panel B (Table 6). Domestic political strife may explain the break year identified by the LS test (2012 in Panel B) or the bilateral frictions over the price of gas supplied by Russia in 2006 (Panel B, ZA test).

Table 8: Unit root tests with two structural breaks for GDP per capita

Countries	LP t-stat	Lags	Break points	LS-stat	Lags	Break points
Panel A: EU15						
Albania	-2.856	(0)	2010 2013	-3.453	(3)	2009 2011
Bosnia and Herzegovina	-7.327***	(0)	2004 2009	0.089	(0)	2008 2013
Georgia	-8.368***	(2)	2005 2009	-3.386	(1)	2001 2003
Moldova	-4.605	(3)	2008 2014	-5.087***	(3)	2009 2012
North Macedonia	-3.285	(1)	2006 2012	-1.895	(1)	2004 2012
Serbia	-3.751	(0)	2003 2013	-3.995**	(2)	2002 2007
Turkey	-4.009	(2)	2007 2010	-2.496	(2)	2005 2009
Ukraine	-3.654	(0)	2008 2014	-4.538**	(3)	2007 2011
Montenegro	-9.484***	(1)	2004 2006	-2.519	(3)	2006 2013
Panel B: AV9						
Countries	LP t-stat	Lags	Break points	LS-stat	Lags	Break points
Albania	-7.591***	(2)	2008 2010	-1.828	(2)	2005 2008
Bosnia and Herzegovina	-8.460***	(1)	2003 2010	-1.829	(3)	2001 2003
Georgia	-6.165**	(2)	2005 2011	-3.029	(2)	2003 2006
Moldova	-7.061**	(1)	2004 2010	-2.512	(3)	2007 2009
North Macedonia	-4.631	(1)	2008 2013	-6.499***	(1)	2001 2008
Serbia	-3.568	(0)	2001 2013	-4.563***	(0)	2009 2013
Turkey	-4.727	(0)	2000 2006	-3.179	(2)	2009 2010
Ukraine	-3.070	(0)	2000 2014	-2.550	(1)	2001 2008
Montenegro	-6.533**	(1)	2006 2014	-3.404	(3)	2007 2012

Notes: The critical values for the LP test are -6.74, -6.16 and -5.89 at the 1%, 5% and 10% level respectively. The corresponding critical values for the LS test with one break on the intercept are -4.545, -3.842 and -3.504. ***, **, * represent the 1%, 5% and 10% significant levels.

Similar tentative inferences on the break years identified by the two tests can be drawn in the case of the results for the GDP per person-employed series presented in Table 7. For instance the break year for Turkey - 2003 – can be associated with the culmination of the 2000-01 economic crisis (Panel A ZA test & Panel B LS test). Similarly, domestic strife in Moldova could be cited as a possible explanation for the 2002 break (Panel A & B, ZA test) or the 2004 break identified by the LS test in Panel A. Again, as a broad observation, the convergence test results in Table 7, paint a rather hazy picture. There is no systematic and strong evidence in favor of a convergence process in terms of the GDP per person-employed index.

As a final step in the empirical investigation, we control for the possibility of two structural breaks. The results using the per capita GDP and the GDP per person-employed series are shown in Table 8 and 9 respectively. As pointed out by Lumsdaine and Papel (1997), allowing for the presence of only one endogenous break may not be sufficient and could lead to loss of information contained in the series. A careful inspection of the results presented in Table 8, suggests that there is no significant difference in terms of the broad, overall conclusions drawn previously. The dominant picture remains that there is a lack of convergence in terms of per capita GDP. Once again, the break years identified can be associated either with the recent financial crisis or with country specific reasons. For example, this is the case with Georgia (break years 2005 & 2009 in Panel A and 2005 & 2011 in Panel B, Table 8). The 2004 and 2009 break points for Bosnia and Herzegovina (Panel A) and 2003 and 2010 (Panel B) can tentatively be associated with significant domestic developments. The Constitutional stalemate in 2009, the handing over of the peacekeeping operation from NATO to an EU led force in 2004 and/or the potential candidate country status offered by the EU summit meeting in 2003 are potential explanations for the two break points.

Turning to the GDP per person-employed findings (Table 9), the results once again do not present a uniform picture and do not allow for any strong and inequivalent inferences to be drawn concerning the convergence hypothesis. Allowing for two structural breaks, the null is rejected by both tests in the cases Serbia and Turkey when the EU15 is used as the benchmark and for Albania, North Macedonia and Bosnia and Herzegovina in the case of the AV9 benchmark (Panel B, Table 9). Based on the LS-stat, the null with two structural breaks is also rejected for Georgia (Panel A, Table 9). Apart from Albania and Bosnia and Herzegovina where the null is rejected by both test statistics with the AV9 as the benchmark (Panel B, Table 9), the null is also rejected for Moldova and Serbia by the LP *t*-statistic and for Georgia and Turkey by the LS statistic.

Table 9: Unit root tests with two structural breaks for GDP per person employed

Panel A: EU15						
Countries	LP t-stat	Lags	Break points	LS-stat	Lags	Break points
Albania	-3.301	(0)	2011 2014	-3.264	(3)	2001 2003
Bosnia and Herzegovina	-5.423	(0)	2005 2007	-2.538	(3)	2004 2005
Georgia	-5.861	(1)	2004 2006	-3.517*	(1)	2002 2004
Moldova	-6.622**	(0)	2000 2012	-3.530*	(0)	2012 2015
North Macedonia	-4.505	(1)	2003 2006	-2.295	(0)	2001 2003
Serbia	-6.050*	(1)	2003 2013	-3.708*	(2)	2002 2004
Turkey	-8.936***	(3)	2003 2008	-6.803***	(2)	2001 2008
Ukraine	-3.552	(0)	2008 2013	-3.301	(3)	2007 2011
Montenegro	-3.443	(0)	2004 2006	-2.696	(3)	2006 2010
Panel B: AV9						
Countries	LP t-stat	Lags	Break points	LS-stat	Lags	Break points
Albania	-7.166***	(3)	2008 2011	-3.720*	(3)	2004 2010
Bosnia and Herzegovina	-8.542***	(3)	2005 2014	-3.632*	(3)	2004 2005
Georgia	-4.712	(1)	2006 2011	-4.038**	(3)	2006 2007
Moldova	-8.873***	(0)	2002 2009	-3.385	(0)	2001 2009
North Macedonia	-6.488**	(1)	2009 2013	-9.295***	(1)	2004 2015
Serbia	-6.058*	(1)	2003 2013	-3.281	(1)	2002 2008
Turkey	-4.752	(3)	2003 2014	-10.14***	(3)	2002 2003
Ukraine	-3.788	(0)	2008 2013	-2.929	(1)	2001 2008
Montenegro	-2.957	(0)	2006 2012	-1.080	(0)	2004 2005

The critical values for the LP test are -6.74, -6.16 and -5.89 at the 1%, 5% and 10% level respectively. The corresponding critical values for the LS test with one break on the intercept are -4.545, -3.842 and -3.504.
***, **, * represent the 1%, 5% and 10% significant levels.

Table 10: Panel LM unit root test statistic without and with breaks

		GDP per capita			**GDP per person-employed**		
Number of Breaks		No Break	One Break	Two Breaks	No Break	One Break	Two Breaks
Benchmark	EU15	2.960	-0.727	-4.827***	0.227	-2.930***	-6.672***
	AV9	0.795	-3.461***	-5.992***	-2.663***	-6.984***	-8.907***

***, **, * represent the 1%, 5% and 10% significant levels. The corresponding critical values for the panel LM test are -2.326, -1.645 and -1.282.

Finally, in Table 10, we present the panel LM unit root test for the two series with and without structural breaks. As can be seen, without allowing for a structural break, the test does not reject the joint null hypothesis of a unit root for both the GDP per person and GDP per person-employed series. This is the case irrespective of the benchmark used. The findings accord with those of the conventional testing approaches presented earlier (Table 4). Allowing for one or two breaks significantly alters the results (Table 10). The panel LM test statistics are negative and significant rejecting the joint unit root null hypothesis in seven cases. As Huang *et al*. (2011) stress, failing to allow for structural breaks in the data used may lead to misleading conclusions. In broad terms, the LM test results in Table 10 seem to point to a stronger in-group convergence process in both metrics used here whereas in comparative terms the results are less robust when it comes to convergence with the EU15. Allowing for one or two structural breaks, seems to point to a convergence process in terms of the productivity index used. Regarding the findings for the individual countries (Tables 5-9) no strong pattern seems to emerge that could be used to support reliable inferences. As in most cases where such tests are used, the findings are seldom uniform. Given the different assumptions and powers of each individual test, this should not come as a surprise. Nevertheless, some weak patterns in the findings can be identified. Focusing on convergence with the EU15, none of the tests conducted points to a convergence process for Albania and North Macedonia. The same applies for Ukraine with the exception of the LS test in Panel A of Table 8. For Serbia, Turkey and Moldova a number of tests indicate convergence with the EU15. However, this does not apply across all the estimated unit root tests. Finally, in terms of in-group convergence, Ukraine is the country that stands out since none of the tests conducted points to a process of convergence. With the exception of the LP test result in Table 8 (panel B), the same applies for Montenegro.

Conclusions

The paper used two different data series in order to explore the issue of convergence with the EU15 for nine countries: Albania, Bosnia and Herzegovina, Georgia, Moldova, Serbia, North Macedonia, Turkey, Montenegro and Ukraine over the period 1997-2016. All are at different stages of EU candidacy that could eventually lead to full membership. Some are officially recognized as candidate countries while others are at the stage of an association agreement. The convergence issue was examined in terms of two macroeconomic indices: GDP per capita and GDP per person-employed. To this effect, panel unit root tests as well as univariate unit root tests were used. To allow for endogenously determined years when structural breaks occur, we used the ZA, the LP and the LS tests. The latter, as noted by Tsanana and Katrakilidis, (2014) is more flexible in detecting structural breaks of unknown dates. The results reported herein varied across the various tests. No robust and clear evidence emerged in favor of a convergence process. As a broad and general observation, the empirical findings indicate lack of convergence with the EU irrespective of the metric used. On the other hand, a noteworthy process of in-group convergence seems to be the case mostly in terms of GDP per person-employed. However, this is a finding that does not universally apply for all the countries. Nevertheless, despite the lack of an empirically traceable convergence process and irrespective of the significant differences in terms of income and productivity, accession to the EU will probably be decided mostly on the basis of geopolitical and other non-economic considerations. This partially explains the accelerating process towards full membership in the case of the Western Balkans countries such as North Macedonia, Montenegro and Serbia. In a similar vein, Turkey's protracted candidacy is probably better explained by political and other factors rather than purely economic ones.

Acknowledgments: The authors are grateful to two anonymous referees for spotting flaws and inconsistencies in earlier versions and for their insightful detailed comments and constructive suggestions that helped improve the paper. They also wish to sincerely thank Junsoo Lee and Constantinos Katrakilidis for their assistance with the panel LM unit root test code.

References

Altinay, G. 2005. Structural breaks in long-term Turkish macroeconomic data, 1923-2003. *Applied Econometrics and International Development*, 5, 117-130.

Amsler, C. and Lee, J. 1995. An LM test for a unit root in the presence of a structural change. *Econometric Theory*, 11, 359-368.

Apergis, N., Panopoulou, E. and Tsoumas, C. 2010. Old wine in a new bottle: growth convergence dynamics in the EU, *Atlantic Economic Journal*, 38, 169–81.

Ayala, A., J. Cunado and L. Gil-Alana 2013. Real convergence: empirical evidence for Latin America, *Applied Economics*, 45(22), 3220-3229

Baltagi, B. 2005. *Econometric Analysis of Panel Data*. John Wiley and Sons, 3rd Edition.

Belke, A., U. Haskamp and G. Schnabl. 2018. Beyond Balassa and Samuelson: real convergence, capital flows, and competitiveness in Greece. *Empirica*, 45, 409-424

Belke, A. and J. Schneider 2013. Portfolio choice of financial investors and European business cycle convergence: a panel analysis for EU countries. *Empirica*, 40, 175–196

Ben-David, D. and Papell, D. 1997. Slowdowns and meltdowns: post war growth evidence from 74 countries, *Review of Economics and Statistics*, 28, 561-571.

Beyaert, A. and Camacho, M. 2008. Panel unit root tests and real convergence: an application to the EU enlargement process, *Review of Development Economics*, 12, 668–81.

Bongardt, A. and F. Torres 2013. Forging Sustainable Growth: The Issue of Convergence of Preferences and Institutions in EMU. *Intereconomics* 2, 72-92

Borsi, M. and N. Metiu 2015. The evolution of economic convergence in the European Union. *Empirical Economics*, 48(2), 657–681

Breitung, J. 2000. The local power of some unit root tests for panel data. *Advances in Econometrics*, 15, 161-177.

Ceylan, R. and V. Abiyev, 2016. An examination of convergence hypothesis for EU-15 countries. *International Review of Economics and Finance* 45, 96–105

Chapsa, X. and C. Katrakilidis 2014. Assessing economic convergence in the EU: is there a perspective for the 'cohesion countries'?, *Applied Economics*, 46(33), 4025-4040

Chapsa, X., C. Katrakilidis and N. Tabakis 2015. Investigating the convergence hypothesis in the EU: more evidence accounting for structural breaks, in A. Karasavvoglou *et al*. (eds.), *EU Crisis and the Role of the Periphery*, Springer International Publishing Switzerland pp. 21-39

Choi, I. 2001. Unit root tests for panel data. *Journal of International Money and Finance*, 20, 249-272.

Dauderstadt, M. 2014. Convergence in crisis: European integration in jeopardy, Friedrich Ebert Stiftung http://library.fes.de/pdf-files/id/ipa/11001.pdf

Dawson, J. W., & Strazicich, M. C. 2010. Time-series tests of income convergence with two structural breaks: evidence from 29 countries. *Applied Economics Letters*, 17, 909–912.

Dickey, D., and Fuller, W. 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74, 427-431.

Dickey, D., and Fuller, W. 1981. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49, 1057-1072.

Dobrinsky, R. 2003. Convergence in per capita income levels, productivity dynamics and real exchange rates in the EU acceding countries. *Empirica* 30, 305–334

Emvalomatis, G. 2017. Is productivity diverging in the EU? Evidence from 11 Member States. *Empirical Economics* 53: 1171–1192

Evans, P. 1998. Using panel data to evaluate growth theories. *International Economic Review*, 39, 295-306.

Firgo, M. and P. Huber 2014. Convergence as a heterogeneous process: what can be learnt about convergence in EMU from regional experiences? *Empirica*, 41, 129–151

Galvão, A. and F. A. Reis Gomes 2007. Convergence or divergence in Latin America? A time series analysis, *Applied Economics*, 39(11), 1353-1360

Hadri, K. 2000. Testing for stationarity in heterogeneous panel data. *Econometrics Journal*, 3, 148-161.

Hlouskova, J. and Wagner, M. 2006. The performance of panel unit root and stationarity tests: Results from a large scale simulation study. *Econometric Reviews*, 25, 85-116.

Huang, H., P. Lin and C. Yeh. 2011. Price level convergence across cities? Evidence from panel unit root tests. *Applied Economics Letters*, 18, 87-93.

Icoz, G. 2011. Turkey's path to EU membership: an historical institutionalist perspective. *Journal of Contemporary European Studies*, 19(4), 511-521

Im, K., Lee, J. and Tieslau, M. 2005. Panel LM unit-root tests with level shifts. *Oxford Bulletin of Economics and Statistics*, 67, 393-419.

Im, K., Pesaran, M. and Shin, Y. 2003. Testing for unit roots in heterogenous panels. *Journal of Econometrics*, 115, 53-74.

Lau, C., K. Marco, E. Demir and M. Bilgin. 2016. A nonlinear model of military expenditure convergence: evidence from Estar nonlinear unit root test, *Defence and Peace Economics*, 27(3), 392-403

Lee, J. and Strazicich, M. 2003. Minimum Lagrange multiplier unit root test with two structural breaks, *Review of Economics and Statistics*, 85, 1082-1089.

Levin, A., Lin, C. and Chu, C. 2002. Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of Econometrics*, 108, 1-24.

Lumsdaine, R. and Papel, D. 1997. Multiple trend breaks and the unit-root hypothesis. *Review of Economics and Statistics*, 79, 212-218.

Lyncker, K. and R. Thoennessen 2017. Regional club convergence in the EU: evidence from a panel data analysis *Empirical Economics*, 52, 525–553

Maddala, G. and Wu, S. 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.

Martin, C. and I. Sanz. 2003. Real convergence and European integration: the experience of the less developed EU Members. *Empirica* 30, 205–236

Perron, P. 1989. The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, 57, 1361-1401.

Sondermann, D. 2014. Productivity in the euro area: any evidence of convergence? *Empirical Economics*, 47, 999–1027

Strazicich, M., J. Lee and E. Day 2004. Are incomes converging among OECD countries? Time series evidence with two structural breaks. *Journal of Macroeconomics* 26, 131–145

Tsanana, E. and Katrakilidis, C. 2014. Do Balkan economies catch up with EU? New evidence from panel unit root analysis. *Empirica*, 41, 641-662.

Tsoukalis, L. 2014. *The Unhappy State of the Union. Europe Needs a New Grand Bargain*. Policy Network, www.policy-network.net

Vojinovic, B., Z. Oplotnik and M. Prochniak. 2010. EU enlargement and real economic convergence. *Post-Communist Economies*, 22(3), 303–322

Zivot, E. and Andrews, D. 1992. Further evidence on the great crash, the oil-price shock and the unit root hypothesis. *Journal of Business and Economic Statistics*, 10, 251-270.