Economics Letters 138 (2016) 92–95

Contents lists available at ScienceDirect

Economics Letters

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

A simple proposal to improve the power of income convergence tests

Artur Silva Lopes *

ISEG, Universidade de Lisboa, Portugal

HIGHLIGHTS

- I propose to replace trend with level stationarity in income convergence tests.
- This replacement improves the power of the tests.
- It is supported by a clear and well known definition of convergence.
- It is illustrated with simulation and empirical results.

ARTICLE INFO

Article history:
Received 28 July 2015
Received in revised form
19 November 2015
Accepted 27 November 2015
Available online 15 December 2015

JEL classification:
C22
F43
O47

Keywords:
Economic growth
Income convergence
Unit root tests

ABSTRACT

I make a simple proposal to replace trend stationarity with level stationarity as the alternative hypothesis in unit root tests for income convergence. This replacement improves the power of the tests and is supported by a clear and well known definition of convergence. A small Monte Carlo study and an empirical example illustrate this improvement.

1. Introduction

The neoclassical growth model predicts that, provided two countries possess similar technologies, preferences and population growth rates, the difference in their per capita incomes must be transitory. Regardless of their initial conditions, in the long-run this difference must exhibit a stationary behavior. Therefore, in a time series framework, support to this income convergence hypothesis requires the rejection of the presence of a unit root in the autoregressive representation of such differences or gaps.

The quest for power for these income convergence tests has led researchers to resort to innovative frameworks, e.g., allowing for non-linearities in deviations from the linear trend (Chong et al., 2008), or even introducing non-linear trends approximated by Fourier expansions (King and Dobson, 2014, 2015). In this note, I propose instead a (very) simple modification of the usual procedure to test the necessary conditions for convergence. This simply amounts to dropping the linear trend term from the unit root test regression and concentrating instead on level stationarity of output gaps. Although this small change is known to generally improve power against level stationary processes, it has not been adopted in the literature on convergence testing. It appears that the reason for this lies not only on a purpose to employ a different notion of convergence but also on a misguided efficiency justification, trying to extend the usefulness of the testing equation to further classification of the type of convergence.

I consider pairwise comparisons between individual countries, not between these and some group average, and I am mostly interested on insights about the particular countries, not on the general support to the hypothesis. Therefore, as is usually the case, a benchmark or reference country is nominated and this is the technology leader, the US.

Moreover, the proposal is restricted to the case where only mature economies are considered. Besides aiming at a comparison with the results of Chong et al. (2008, CHLL), this restriction is imposed to comply with Bernard and Durlauf's (1996, BD)

* Correspondence to: ISEG, Rua do Quelhas 6, Gab. 107, 1200–781 Lisboa, Portugal.

E-mail address: asl@iseg.utl.pt.

http://dx.doi.org/10.1016/j.econlet.2015.11.041
0165-1765/

requirement to consider only economies close to their steady state, to avoid spurious rejections of the hypothesis.

2. Definition and DF test regressions

A proper assessment of the income convergence hypothesis requires that it is clearly defined. The most clear definition appears to be the one from BD, requiring that the long-run optimal forecasts for the logs of both outputs do not diverge:

$$\lim_{k\to\infty} \mathrm{E}(y_{i,t+k} - y_{j,t+k} \mid \mathcal{F}_t) = 0, \tag{1}$$

at any fixed time t, $y_{i,t}$ and $y_{j,t}$ denoting the logarithms of per capita output for countries i and j, respectively, and $\mathcal{F}_t$ representing the set of all available information at time t, containing at least all historical information on both incomes.

Assuming, as usual, that both logged outputs contain a linear deterministic trend and a stochastic one, BD establish the testing conditions in their proposition 5: if the gap or discrepancy $y_{i,t} - y_{j,t}$ contains a nonzero mean or a unit root, Eq. (1) is violated. Hence, there must be cotrending, both stochastic and deterministic.

However, the zero mean condition is usually considered as overly stringent. In Pesaran's (2007) simple decomposition model, for instance, it requires that several structural parameters must be identical for both economies. Different savings rates or population growth rates may imply a (constant) non-zero mean for the output gap. This is the condition most commonly adopted, and corresponds to "long-run convergence", in Oxley and Greasley (1995), or "deterministic convergence", in Li and Papell (1999), or "asymptotically relative convergence", in Hobijn and Franses (2000).

A weaker notion of convergence is that of *catching-up*. Again following BD, countries i and j are said to converge in this sense between "*dates t and $t+H$ if the (log) per capita output disparity at t is expected to decrease in value*". In particular, if i refers to the reference economy, then $y_{i,t} > y_{j,t}$ and the conditional forecast for time $t+H$ must satisfy

$$\mathrm{E}(y_{i,t+H} - y_{j,t+H} \,|\mathcal{F}_t) < y_{i,t} - y_{j,t}, \tag{2}$$

that is, the difference is forecasted to diminish over the specified time interval. Although the presence of a unit root in the output gap is still ruled out, that of a deterministic time trend now is not.

Since the most clear and precise definition requires level stationarity, the corresponding Dickey–Fuller (DF) test regression[1] should be

$$\Delta y_t = \alpha + \phi\, y_{t-1} + \sum_{i=1}^{k} \gamma_i\, \Delta y_{t-i} + \mathrm{error}_t, \quad t = 1, \ldots, T, \tag{3}$$

where y_t denotes the gap or discrepancy, $y_t = y_{i,t} - y_{j,t}$, and k is large enough to make the errors serially uncorrelated. That is, the only deterministic regressor should be the constant. However, the DF test regression used almost unanimously in convergence testing is

$$\Delta y_t = \alpha + \beta\, t + \phi\, y_{t-1} + \sum_{i=1}^{k} \psi_i\, \Delta y_{t-i} + \mathrm{error}_t, \tag{4}$$

where t denotes the usual (linear) trend term. A notable exception is Bernard and Durlauf (1995), who consider a framework analogous to that of (3) for a multivariate analysis.

However, it is clear that Eq. (3) should be preferred. This is a case where economic theory is clear in specifying the alternative hypothesis and the null follows straightforwardly. The presence of a linear trend is inadmissible both under the null and the alternative. This presence is scrutinized in *catching-up* testing but it is not admissible in a unit root testing framework, where the specification of deterministic regressors is acknowledged to be so important. It is at least inefficient to neglect the guidance that is provided by economic theory precisely in a situation where such guidance is so valuable due to the potentially dramatic implications on the test properties. Since the power of unit root tests is known to decrease as deterministic regressors are added, this very simple modification alone may provide a significant power improvement; for a recent discussion on this topic, addressing asymptotic power, see Harvey et al. (2009).

Furthermore, this is also a case where a sensitive issue of DF unit root testing may reverse its role and act positively, becoming useful for convergence testing. Actually, recall from the celebrated paper by Campbell and Perron (1991), here summarizing a finding in Perron (1988), that DF tests with only an intercept in the set of deterministic regressors have power that goes to zero as T grows in case the DGP is a trend stationary process (TSP), i.e.,

$$\lim_{T\to\infty} \Pr\left[\text{rejecting the unit root} \mid \text{process is TSP}\right] = 0.$$

Since the unit root hypothesis corresponds to non-convergence, rejecting it implies deciding for convergence. Hence,

$$\lim_{T\to\infty} \Pr\left[\text{deciding for convergence} \mid y_t \sim \mathit{TSP}\right] = 0,$$

that is,

$$\lim_{T\to\infty} \Pr\left[\text{deciding for non-convergence} \mid y_t \sim \mathit{TSP}\right] = 1,$$

i.e., as a convergence test the DF regression that omits the trend term is consistent, its power goes to 1 as T grows, in this case not because there is a unit root but because there is a trend in the discrepancy process, and hence there is no convergence. Albeit not strictly correct as a unit root test regression due to the presence of the trend, the test is consistent as a convergence test because such presence is not allowed by the hypothesis. And of course, in case of level stationarity, the standard properties of DF tests ensure its consistency.

Two further remarks follow:

(a) in case the most stringent definition is adopted, not even the intercept should be included in the test regression;
(b) obviously, the possibility of a segmented trend stationary process must be also excluded *a priori*.

3. Simulation results

In this section the results of a small simulation study are presented. Only two sets of simulation experiments are presented and the purpose is merely illustrative, far from any concerns of completeness or exhaustiveness. Two particular stationary processes are considered: an AR(1) and an ARMA(2, 1). In both cases the asymptotic level is 5%, 20.000 replications are used and the observations are generated for $t = -49, -48, \ldots, T$, to discard the first 50 observations. DF and ADF test statistics are used, the subscripts denoting the deterministic regressors included: c for the case with only an intercept, and ct for the case where a linear trend term is added. The ADF statistics are calculated using the general-to-specific (GTS) t-sig method with 10% asymptotic level simplifying tests for lag selection, which begins with $k_{\max} = 6, 8$ and 10 for the cases of $T = 50$, 100 and 200, respectively.

[1] Although popular wisdom indicates that (A)DF-GLS tests are more powerful, I follow the recommendation by Müller and Elliot (2003) and focus on the simpler (A)DF(-OLS) test statistics because for the great majority of countries the first observations of the gap series are far off their sample "equilibrium values". In spite of the restriction to OECD countries, this is due to the location of the beginning of the sample in 1950.