

Testing for output convergence: a re-examination

By Yin-Wong Cheung* and Antonio Garcia Pascual†

*Department of Economics, SS1, University of California, Santa Cruz, CA 95064, USA; email: cheung@cats.ucsc.edu

†International Monetary Fund, 700 19th Street NW, Washington, DC 20431, USA; email: agarciapascual@imf.org.

This paper investigates output convergence for the G7 countries using panel time-series techniques. We consider both the null hypotheses of no convergence and convergence. It is shown that inferences on output convergence depend on which one of the two null hypotheses is considered. Further, the no convergence results reported in previous studies using the time-series definition may be attributed to the low power of the test procedures being used. Our results also highlight some potential problems on interpreting results from some typical panel unit root and stationarity tests.

1. Introduction

One of the major differences between neoclassical and endogenous growth models is their prediction of national output dynamics. A strong result from the standard neoclassical growth model (Solow, 1956; Swan, 1956) is the convergence of *per capita* output across countries with a similar productivity level, savings rate, depreciation rate, productivity growth and population growth. It means that differences in national output, in *per capita* terms, are going to disappear over time. The endogenous growth model (Romer, 1990; Aghion and Howitt, 1992; Grossman and Helpman, 1991), on the other hand, asserts that country-specific factors play a role in determining aggregate income. Since country-specific factors can evolve endogenously according to the environment unique to a country, countries with dissimilar initial endowments and attributes can have *per capita* output that do not converge over time. The different views on national output have spurred considerable interest on testing whether the observed *per capita* output data are converging or not. Various statistical techniques and samples on output data are used to evaluate growth theories. In their review article, Durlauf and Quah (1999) point out that the growing empirical literature on economic growth has generated ‘fresh stylized facts on growth with important implications for theory.’

The empirical literature on output convergence has undergone some changes in recent years. Many of the early empirical studies are based on the cross-country analysis, which regresses average *per capita* output growth rates on initial output

levels. Usually, conditioning variables such as education attainment, government spending, political instability, and the growth rate of the terms of trade are included in such an output regression equation to control for effects of other growth factors (Barro and Sala-i-Martin, 1995). Under the convergence hypothesis, countries starting with a low *per capita* income should have a higher growth rate. Thus, an inverse relationship between output growth and initial output is interpreted as evidence in favor of the convergence hypothesis. A sample of studies pursuing this methodology is Baumol (1986), DeLong (1988), Barro (1991), Barro and Sala-i-Martin (1992), and Mankiw *et al.* (1992). In general, these studies report results in favor of convergence.¹

However, the appropriateness of the cross-country regression approach is challenged by, for example, Quah (1993), Bernard and Durlauf (1996), and Evans (1997). Quah (1993) shows that a negative correlation between output growth and initial output is consistent with a stable variance in cross-country output. Bernard and Durlauf (1996) argue that the initial-output regression approach tends to reject the null hypothesis of no convergence too often in the presence of multiple output equilibria. Evans (1997) points out that the cross-sectional approach may generate inconsistent convergence rate estimates, which lead to incorrect inferences. Instead, these authors propose the use of time-series methods to study convergence. Under the time-series framework, output convergence requires real *per capita* cross-country output differentials to be stationary; that is, the levels of *per capita* national output are not diverging over time.

The existing evidence from the time-series approach is not very favorable to the notion of convergence. Bernard and Durlauf (1995), using unit root and cointegration techniques, detect the presence of multiple integrated processes driving the output data of the OECD countries. The result is interpreted as not supportive of the convergence hypothesis. Evidence against the convergence hypothesis is also reported in Quah (1992), which examines the unit root property of *per capita* output relative to the US data. Using a panel unit root test, Evans (1998) shows that convergence occurs within a group of developed countries and different growth patterns are observed among countries with different literacy rates. Li and Papell (1999) consider a group of OECD countries. Only after allowing for structural breaks, the authors uncover strong evidence of convergence using *per capita* national output relative to the group aggregate.

Compared with the cross-country analysis, the time-series approach yields less convincing findings for the convergence hypothesis. One possible reason for the no-convergence outcome, however, may be related to the empirical procedures used in these studies. The typical time-series test has no convergence (unit root) under the null hypothesis. Since it is commonly known that unit root tests tend to have a low power against persistent but stationary alternatives, the inability of these studies to reveal evidence of convergence is not too surprising.

¹ See Temple (1999) and Durlauf and Quah (1999) for surveys of the empirical growth literature.

In this study, we use recently developed statistical techniques to investigate the convergence property of national output. Data from the G7 countries are used. The stationarity property of the real *per capita* output relative to the US is used to infer convergence. Data quality has some bearings on the ability of a test to distinguish the null from the alternative. If the data are not informative enough to discriminate between the convergence and no-convergence specifications, then statistical procedures will not be able to reject either the null hypothesis of convergence or no convergence. To highlight the power issue, we consider two types of panel time-series tests—one has unit root (no convergence) as the null hypothesis, and the other has stationarity (convergence) as the null. Panel time-series procedures are employed because their ability to reject a false null hypothesis is higher than the corresponding univariate procedures. The results from procedures with different specifications of the null hypothesis help determine the usefulness of the data in terms of their ability to identify the convergence property.

In analyzing output convergence, we control for a few issues related to the use of panel time-series procedures. First, we account for the effect of cross-country output correlations on the panel tests. The second issue is related to the joint-hypothesis nature of panel time-series tests. For a typical panel procedure, the rejection of, say, the no-convergence null hypothesis is commonly interpreted as the rejection of the joint hypothesis that individual output differential series are divergent. However, it may only take convergence in a subset of countries, which are not identified by the test, to obtain the rejection result. To address the possible ambiguity, we will examine the convergence behavior of individual output differential series in a panel framework. The third issue is the use of sample-specific critical values. Even though the statistical procedures usually have well defined asymptotic behavior, their performance in finite samples is quite difficult to evaluate analytically. To ensure proper inferences, we use artificial data generated according to parameters estimated from actual output differential series to derive the critical values.

We define the empirical specification of output convergence in the next section. The definition is similar to the time-series version of convergence introduced in Bernard and Durlauf (1996). The statistical tests are described in Section 3. In this section, we introduce two panel unit root tests (Im *et al.*, 1997; Taylor and Sarno, 1998), a panel framework to evaluate the convergence behavior of individual output differential series (Breuer *et al.*, 2002), and the panel stationarity tests (Choi and Ahn, 1999). In the same section, we also detail the generation of the sample-specific finite sample critical values. The empirical results are presented in Section 4. Two sample periods are considered—one from 1950 to 1992 and the other from 1885 to 1994. A larger sample size usually entails better information. Indeed, our exercise shows that the information in the shorter sample does not give an unequivocal inference on convergence. The longer output data series, however, yield unambiguous evidence on output convergence. Section 5 contains some concluding remarks.

2. Empirical specification of output convergence

A time-series based approach to investigate output convergence has been proposed by Bernard and Durlauf (1996) and Quah (1992). According to Bernard and Durlauf (1996), there is output convergence between two countries if the long-run forecasts of their real *per capita* outputs are the same. If pairwise convergence holds for all the pairs of countries under consideration, then there is convergence of all the countries simultaneously. To make the definition operational, time-series convergence between two countries requires that their real *per capita* output differentials are stationary. In this case, the test for convergence is translated to a test for the stationarity of output differentials.

Empirical studies adopting the notion of time-series convergence usually test the null hypothesis of no convergence against the alternative of convergence (Bernard and Durlauf, 1995; Evans, 1998; Li and Papell, 1999). Specifically, let $Y_{i,t}$ be the (logarithm of the) country i 's real *per capita* output at time t and let $Y_{*,t}$ be the output variable of the benchmark country. The no-convergence hypothesis is stated as

$$H_0 : x_{i,t} \equiv (Y_{i,t} - Y_{*,t}) = I(1), \quad \forall i = 1, \dots, N \quad (1)$$

where $x_{i,t}$ is the real *per capita* output of country i relative to the benchmark country, $N+1$ is the total number of countries in the sample, and $I(1)$ denotes a unit root non-stationary process. Standard procedures, such as augmented Dickey-Fuller tests, are commonly used to test the $I(1)$ null against the stationary alternative.

It is well known that the standard unit root tests have low power against stationary but persistent alternatives. The use of (1) as the null hypothesis may lead to the bias of accepting the no-convergence hypothesis. One way to address the issue is to consider a different null hypothesis. For instance, if the information content of the data is not rich enough to differentiate the different types of output dynamics, then the data will not reject the convergence or the no-convergence hypothesis. Taking stationarity as the null hypothesis offers an avenue to check if the acceptance of no convergence is due to the lack of power or if the data actually display diverging dynamics. The null hypothesis of convergence is given by

$$H_0 : x_{i,t} \equiv (Y_{i,t} - Y_{*,t}) = I(0) \quad \forall i = 1, \dots, N \quad (2)$$

where $I(0)$ denotes a stationary stochastic process. When the data offer a clear indication of non-convergence, the stationary null hypothesis will be rejected.

The results from testing both the non-stationary null hypothesis (1) and the stationary null hypothesis (2) offer some complementary information to evaluate the competing hypotheses. When we fail to reject both (1) and (2), the data just do not allow us to separate these two types of convergence dynamics. However, if we reject (1) as the null hypothesis and do not reject (2) as the null, then the evidence

is in favor of cross-country convergence. Alternatively, a strong evidence of no convergence is established if the null (2) is rejected but (1) is not rejected.

The examination of both the null hypotheses of stationarity and non-stationarity provides a level field to evaluate output convergence dynamics. However, it does not directly address the power issue. One way to extract more information from cross-country data is to use a panel procedure, which incorporates interactions between data series. In this exercise, we employ a few panel procedures to improve the power performance. Another way to improve the power performance is to work with a long data series. Thus, in addition to a post-WWII sample, we examine a longer history data set to see if it gives a more definite conclusion on convergence dynamics.

3. Statistical procedures

In this section, we describe the statistical procedures used to evaluate the output convergence dynamics. While the univariate unit root ADF test is a standard technique, the recently developed panel stationary/non-stationary procedures are not very commonly utilized in this literature. Thus, we provide a brief description of these procedures in the following subsection. However, readers who are mainly interested in the empirical results can skip this section and move directly to the next one.

3.1 Univariate unit root test

The augmented Dickey-Fuller test (ADF, hereafter) is quite commonly used to investigate the stationarity property of individual output differential series. The test is based on the regression equation

$$\Delta x_{it} = \alpha_i + \beta_i t + b_{i,0} x_{it-1} + \sum_{j=1}^{p_i-1} b_{i,j} \Delta x_{it-j} + \varepsilon_{it} \quad t = 1, \dots, T \quad (3)$$

The null hypothesis of unit root non-stationarity is rejected if the coefficient $b_{i,0}$ is significantly less than zero. The inference is based on the usual t-statistic of $b_{i,0}$, which has a non-standard distribution. The Akaike information criterion is used to determine the lag length parameter p_i .

The test procedure allows for both a constant and a time trend. The presence of the constant term is to accommodate the possibility of parallel *per capita* output paths, which Li and Papell (1999) labeled deterministic convergence. As pointed out by Evans (1998), many interesting exogenous growth models predict that countries have the same long-run output growth rate, which are determined by technical knowledge and have parallel output paths. The trend term is included to ensure the test result does not depend on the value of α_i (Evans and Savin, 1984). West (1987) also points out that the ADF test is inconsistent if the process is

stationary around a time trend and the trend term is not included. In fact, for the results reported in the next section, the trend term is always not significant. It is recognized that the inclusion of such a trend term will lower the power of the test. However, as one of the safeguards against misleading inferences, we choose to keep the trend term in the regression and accept a power loss. Instead, the panel procedures are adopted to enhance the power performance.

3.2 Panel unit root tests

Over the past decade, several statistical procedures have been developed to test for unit roots in a panel setting (Levin and Lin, 1992, 1993; Quah, 1994). One potential benefit of using a panel procedure is the information gained from pooling data across different series. The information gain can improve the estimation efficiency and the power of the testing procedure. In the following subsections, we describe three recently advanced panel unit root tests and their special features.

3.2.1 Lagrange multiplier test Im *et al.* (1997) propose a Lagrange multiplier (LM) statistic to test for the presence of unit roots in the panel framework. The LM statistic is based on the average of individual Lagrange multiplier statistics (LM_i 's) for testing $b_{i,0} = 0$, for all $i = 1, \dots, N$. Specifically, the LM test statistic and the null hypothesis are given by

$$LM = \frac{1}{N} \sum_{i=1}^N LM_i \quad (4)$$

and

$$H_0 : b_{1,0} = b_{2,0} = \dots = b_{N,0} = 0 \quad (5)$$

Under conditions given in Im *et al.* (1997), the standardized LM statistic has an asymptotic standard normal distribution under the null hypothesis (5). See Im *et al.* (1997) for a detailed discussion of the test statistics. Note that under the null hypothesis given by (5), there is no convergence between any pair of countries. Compared with some previous procedures (Levin and Lin, 1992, 1993; Quah 1994), the LM test allows a higher degree of heterogeneity in the cross-sectional data. For instance, the length parameter p_i can vary across individual series. Also, the cross-sectional correlation induced by a common time effect can be handled using cross-sectionally demeaned data.

3.2.2 Panel ADF test Sarno and Taylor (1998) and Taylor and Sarno (1998) propose a panel version of the ADF (MADF hereafter) test. To implement the test, (3) is estimated as a system of N equations using the feasible generalized least squares technique and the standard Wald test statistic is used to evaluate the null hypothesis given by eq. (5).

Again, the null hypothesis of the MADF test is a joint hypothesis of no convergence between any pair of countries. The MADF test allows for different lag parameters in individual series. Compared with the Im *et al.* LM test, the MADF test has a more flexible structure to accommodate cross-sectional correlation. Taylor and Sarno (1998) show that, using the Monte Carlo method, the MADF test has better power properties than the univariate ADF procedure.

3.2.3 SURADF test for a unit root in individual series The LM and MADF tests discussed in the previous subsections achieve power improvement by exploiting the multivariate nature of the system. However, a rejection of the null hypothesis by the LM and MADF tests should be interpreted with caution. Because of the joint nature of the null hypothesis, the null hypothesis that all country pairs do not converge can be rejected if one or more real *per capita* output differential series display convergence. Thus, when the null hypothesis is rejected, we do not know if convergence exists in some or all the countries under consideration. These two panel unit root tests are not able to identify which members of the panel achieve convergence. Thus, the two procedures do not yield unambiguous evidence on convergence even the no-convergence null hypothesis is rejected.

Breuer *et al.* (2002) devise a panel procedure, labeled SURADF test, to test for the presence of a unit root in an individual series. The procedure constructs the test statistics within a panel framework but evaluates the unit root property series by series. In doing so, the procedure exploits information embedded in the system and yields evidence on which member of the system is (non-)stationary. That is the SURADF test offers a way to improve the chance of rejecting a false null hypothesis of non-stationarity and alleviates the ambiguity that arises with a joint null and alternative hypotheses of ‘all series are non-stationary/not all series are non-stationary’.

In essence, the feasible generalized least squares method is used to estimate the system of N equations and the null hypothesis

$$H_0 : b_{i,0} = 0 \quad (6)$$

is examined individually for $i = 1, \dots, N$. As the statistics are estimated within a system, it is not appropriate to use the standard ADF critical values to appraise their significance. Breuer *et al.* (2002) recommend the use of sample-specific critical values to conduct the statistical inference.

3.3 Panel stationarity tests

The tests presented in the previous subsections have non-stationarity, or no convergence, as the null hypothesis. If these tests have low power, then the no-convergence result will be erroneously established. Recently, Choi and Ahn (1998) developed a panel test for stationarity against non-stationary alternatives.

Their procedures are semi-nonparametric and do not require information on the parametric ARMA structure of the data. Compared with the panel unit root tests, the Choi and Ahn test offers a different perspective to evaluate the output convergence property. One statistic considered here is a multivariate version of the Sargan and Bhargava (1983) test

$$SBDH = tr \left[\left(\frac{1}{T^2} \sum_{t=1}^T \tilde{S}_t \tilde{S}_t' \right) \tilde{\Omega}_t \right] \quad (7)$$

tr denotes the trace of a matrix. \tilde{S}_t is the partial sum $\sum_{k=1}^t \tilde{X}_k$ where the transformed variable \tilde{X}_t is derived from

$$X_t = \beta_0 + \beta_1 t + \tilde{X}_t$$

where X_t is the $N \times 1$ vector containing $x_{i,t}$, $i = 1, \dots, N$. $\tilde{\Omega}_t$ is the heteroscedasticity and autocorrelation consistent covariance matrix estimator

$$\begin{aligned} \tilde{\Omega}_t &= \sum_{n=1}^l C(n) k(n/l) \\ C(n) &= \begin{cases} T^{-1} \sum_{t=2}^{T-n} \Delta \tilde{S}_{t-n} \Delta \tilde{S}_t', & \text{for } n \geq 0 \\ T^{-1} \sum_{t=2}^{T+n} \Delta \tilde{S}_t \Delta \tilde{S}_{t+n}', & \text{for } n \geq 0 \end{cases} \end{aligned}$$

where $k(n/l)$ is a quadratic spectral kernel and l is a bandwidth parameter capturing the serial correlation in the data.

It is known that pre-whitening the data with a low-order AR regression yields a more accurate covariance estimate. Thus, in our exercise, the covariance matrix is estimated by the Andrews and Monahan (1992) prewhitened kernel estimator, which applies the Andrews (1991) data-dependent method to determine the bandwidth parameter. Following Choi and Ahn (1998), we set l equal to integer $[\delta(T/100)^{0.25}]$ if the data-dependent method gives a bandwidth parameter that is greater than $T^{0.65}$. In this exercise, we have $\delta = 8, 10$, and 12 . The prewhitening of $\Delta \tilde{S}_t$ is accomplished using a VAR(1) operator, and the filtered data are used to construct $C(n)$.

Choi and Ahn (1998) suggest the SBDH statistic can be calculated using another transformed variable S_t^* , defined by

$$\sum_{k=1}^t X_k = \beta_0 \sum_{i=1}^t i^0 + \beta_1 \sum_{i=1}^t i^1 + S_t^*$$

When \tilde{S}_t is replaced by S_t^* in (7), we label the resulting statistic SBDHT. Both the SBDH and SBDHT statistics test the null hypothesis that all individual series are

stationary. The alternative is that not all of the series in the panel are stationary. That is, under the null hypothesis, output data series under consideration converge simultaneously. Under the alternative hypothesis, convergence does not hold for at least one country. Thus, results from the SBDH and SBDHT tests are complementary to those from the panel unit root tests. However, similar to the LM and MADF panel unit root tests, a rejection of the SBDH and SBDHT tests has to be interpreted with caution. The rejection of stationarity by the SBDH and SBDHT tests imply some or all the series are non-stationary, but the tests do not identify which ones are non-stationary.

3.4 Sample-specific critical values

The finite-sample behavior of the panel test statistics can be very different from the one implied by their asymptotic distributions. To minimize the finite-sample effects on our empirical results, we rely on sample-specific critical values to draw statistical inferences. Since both the null hypotheses of stationarity and non-stationarity are considered, we fit an ARIMA and an ARMA model to each output differential series. The model specification is determined by the information criterion. The estimated residuals are used to produce the covariance matrix of the error term. These parameter estimates, then, define the data generating process.

For each panel unit root test, a pseudo-random sample $N \times (T + 100)$ is generated according to the estimated ARIMA and covariance matrix specifications under the null hypothesis. The first 100 observations of each series are dropped to minimize the effect of initial conditions. The sample test statistic is then calculated. The sample-specific critical values for the panel stationary tests are generated in a similar fashion using the stationary specifications. All the sample-specific critical values reported below are based on 10,000 replications in each experiment.

4. Empirical results

The procedures described in the previous section are used, together with the corresponding sample-specific critical values, to evaluate the output convergence hypothesis. Data from the G7 countries are used.

4.1 The Summers and Heston data set: 1950–1992

The output data from Penn World Table are widely used in empirical studies of growth and convergence. The current version, Mark 5.6, is a revised and updated version of the preceding Mark 5 (Summers and Heston, 1991). The annual data on GDP *per capita* in real terms from the G7 countries are considered. Purchasing power parity conditions are incorporated in the compilation process to make these output data comparable across countries. The real series are computed using a chain index expressed in international prices with 1985 as the base year. The sample period is from 1950 to 1992. In this exercise, the US is the benchmark

country. For the term ‘output differential’ is used to denote the real *per capita* output in a country relative to the real *per capita* output in the US.

4.1.1 Univariate analysis First, we apply the ADF test to the output differentials. The null hypothesis of non-stationarity is interpreted as no convergence with respect to the US data. The results, presented in Table 1, show limited evidence of convergence. The estimate that indicates output differential persistence displays a wide range. The half-lives implied by these estimates range from less than a year (United Kingdom) to over 10 years (Japan). Despite the estimated wide spectrum of persistence, the data on output differentials seem quite noisy. With the exception of the German output differential series, the null hypothesis of non-stationarity is not rejected. The no-convergence result is similar to those reported in earlier studies adopting the time-series approach (Bernard and Durlauf, 1995; Quah, 1992; Li and Papell, 1999).

While the univariate ADF test does not offer positive evidence on convergence, the univariate analysis provides two useful pieces of information about output dynamics. First, individual output differential series tend to have dissimilar lag structures. Second, there is considerable comovement among the output differential series as indicated by the correlation of residuals from the individual least squares estimation of (3). The Canadian output differential series has the lowest level of comovement with others—the correlation coefficients between the Canadian and other national series vary from -0.02 (Canada-UK) to 0.27 (Canada-France). For the other output differential series, the correlation coefficients of the estimated residuals are usually positive and large. In fact, the residual correlation coefficients range from 0.41 (Italy-UK) to 0.76 (France-Italy).

Both the heterogeneous lag structure and the correlation between series have implications for panel tests. For instance, if heterogeneous lag structure and cross-equation correlation are not properly accounted for, they have non-trivial effects on the size and power of panel tests (O’Connell, 1998; Papell and Theodoridis, 2000; Taylor and Sarno, 1998). Thus, the panel results reported in the following are all based on the tests that allow for both heterogeneous lag structure and cross-equation residual correlation. The effects of these two factors on finite-sample performance are accounted for using sample-specific critical values.

Table 1 Univariate ADF test (Summers-Heston data)

Country	$\hat{b}_{i,0} + 1$	Statistic	1%	5%	10%
Canada	0.8548	-1.6402	-4.1719	-3.5048	-3.1792
France	0.9175	-1.2987	-4.1948	-3.5253	-3.1988
Germany	0.8266	-4.0251^{**}	-4.1719	-3.5048	-3.1792
Italy	0.8869	-1.8415	-4.1948	-3.5253	-3.1988
Japan	0.9350	-1.4318	-4.1381	-3.4667	-3.1404
United Kingdom	0.6573	-2.5844	-4.1381	-3.4667	-3.1404

Note: Critical values from Cheung and Lai (1995) are used. ** indicates significance at the 5% level.

4.1.2 Panel analysis The results of the Im *et al.* LM test and the Taylor and Sarno MADF test are reported in Table 2. The joint null of non-stationarity is not rejected by both tests. The Im *et al.* statistic has a p-value of 26.75%, which is well above the usual 5% or 10% levels. The MADF statistic gives an even higher p-value of 74.63% and clearly indicates no convergence among the national output data. Even though these two panel procedures have better power than the univariate ADF test, they do not provide more favorable evidence of output convergence.

Taylor and Sarno (1998), for example, allude to the possibility that the presence of one stationary series can lead to the rejection of the joint hypothesis that all the individual series are non-stationary. That is, rejecting the joint null does not tell us whether all or some of the elements of the multivariate system are stationary. However, the current exercise illustrates the other possibility. The non-rejection of the null does not necessarily mean all the member series are non-stationary. Due to the lack of power, a panel unit root test may disguise the stationarity of a member series. As indicated in Table 1, the German output differential series is stationary. The panel unit root tests, nonetheless, do not reject the null that all the series are non-stationary. Thus, one should interpret these panel test results with caution.

We used the Breuer *et al.* SURADF test to investigate individual output differential convergence in the panel setting. The results are given in Table 3. Both the estimates $(\hat{b}_{i,0} + 1)$'s and the test results are similar to those based on the univariate ADF tests.² There is only evidence of convergence between the German and US output data as the German output differential series is found to be stationary. For the other countries, there is no sign of convergence with the US output. Thus, the potential power gain from using a panel framework does not yield stronger evidence for output convergence. Compared with the panel tests, the univariate ADF seems to still have a role in studying output convergence.

The test results derived from both univariate and panel unit root tests offer limited support for the output convergence hypothesis. However, it is possible that the finding of no convergence is due to the inability of the procedures to reject the null hypothesis—either because of the low power of the procedures or

Table 2 Panel unit root tests for the joint null hypothesis of non-stationarity (Summers-Heston data)

	Statistic	p-value	1%	5%	10%
LM	5.6097	0.2675	7.9988	7.0324	6.4983
MADF	23.6492	0.7463	57.4831	48.2556	43.6976

Note: The Im *et al.* LM and Taylor and Sarno MADF panel unit root test statistics are given in rows labeled LM and MADF. The p-values and the 1%, 5%, and 10% sample-specific critical values are computed from 10,000 Monte Carlo simulations.

² As observed by an anonymous referee, the estimates of $(\hat{b}_{i,0} + 1)$ from the univariate and SURADF procedures are not the same as the latter takes into account the contemporaneous correlation of the error terms.

Table 3 Test for unit root in individual series using panel estimation (Summers-Heston data)

Country	$\hat{b}_{i,0} + 1$	Statistic	p-value	1%	5%	10%
Canada	0.8833	-1.3889	0.8624	-4.6653	-3.9788	-3.6132
France	0.9209	-1.7821	0.7859	-4.9229	-4.0977	-3.7222
Germany	0.8231	-5.4472	0.0019	-4.6977	-3.9768	-3.5889
Italy	0.8650	-2.8308	0.3567	-4.8495	-4.0354	-3.6833
Japan	0.9503	-1.4775	0.8012	-4.8029	-3.9876	-3.6008
United Kingdom	0.6760	-2.9617	0.2632	-4.7801	-4.0186	-3.6187

Note: Results of the Breuer *et al.* SURADF test are reported. The p-values and the 1%, 5%, and 10% sample-specific critical values are computed from 10,000 Monte Carlo simulations.

Table 4 Panel tests for the joint null hypothesis of stationarity (Summers-Heston data)

	Statistic	δ	p-value	1%	5%	10%
SBDHT	1.0914	8	0.4010	9.0292	2.7976	1.9529
SBDH	1.3075	8	0.3948	12.1740	3.3111	2.3090
SBDHT	1.6601	10	0.3768	39.5396	8.0292	4.4393
SBDH	1.6155	10	0.4535	41.8087	8.4362	4.8938
SBDHT	1.6576	12	0.3446	40.9670	9.1615	4.8888
SBDH	1.7212	12	0.3872	44.1652	10.1339	5.5214

Note: The Choi and Ahn panel stationarity SBDH and SBDHT test statistics are given. The p-values and the 1%, 5%, and 10% sample-specific critical values are computed from 10,000 Monte Carlo simulations.

the uninformative data. To shed light on such a possibility, we apply the Choi and Ahn panel stationarity test to the output differential data. Following Choi and Ahn (1999), we set the lag truncation parameter δ to 8, 10, and 12. Table 4 presents the test results.

The stationarity test results are in stark contrast with the unit root test results. Both the SBDHT and SBDH statistics do not reject the null hypothesis that all the series are stationary. Specifically, for $\delta = 10$, the SBDHT statistic has a p-value of 37.68% and the SBDH statistic has a p-value of 45.35%—both fail to reject the joint null of stationarity even at the 10% level. That is, if the null hypothesis is ‘convergence exists simultaneously in all country pairs’, then the data do not contradict the notion of output convergence.

Given the unit root and stationarity test results, what can we say about the output convergence hypothesis? The inference is disconcertingly dependent on the way the null hypothesis is set up. While the unit root tests do not reject the no-convergence hypothesis, a result that is consistent with previous studies adopting the time-series approach, the stationarity tests reveal the data can be supportive of the convergence hypothesis. There is one probable interpretation of the non-confirmatory results. The information content of the output differential data is not sharp enough for the testing procedures to discriminate between convergence and no convergence.

4.2 The Maddison data set: 1885–1994

In the previous subsection, the results based on the post-WWII sample period indicate that output data may not be informative enough to yield an unequivocal conclusion on output convergence. One possible remedy is to employ a more informative data set. Maddison (1995) provides an alternative data set to examine output convergence. The data set is also commonly used in output dynamics analysis (Bernard and Durlauf, 1995; Evans, 1998; Li and Papell, 1999). One advantage of the Maddison data set is that, compared with the Penn World Table, it covers a longer sample period. The data are comparable across countries and have been adjusted for changes in national boundaries. The sample period is from 1885 to 1994.

The results of the univariate and panel unit root tests are reported in Tables 5, 6, and 7. In general, the $(\hat{b}_{i,0} + 1)$'s suggest that the point estimates of output persistence in the post-WWII sample and the Maddison data set are quite comparable.

Table 5 Univariate ADF test (Maddison data)

Country	$\hat{b}_{i,0} + 1$	Statistic	1%	5%	10%
Canada	0.8217	−2.9644	−4.0239	−3.4377	−3.1414
France	0.8761	−3.3140*	−4.0239	−3.4377	−3.1414
Germany	0.8968	−2.7634	−4.0030	−3.4201	−3.1252
Italy	0.9019	−2.7765	−4.0239	−3.4377	−3.1414
Japan	0.9575	−1.4747	−4.0351	−3.4466	−3.1494
United Kingdom	0.8827	−2.4949	−4.0351	−3.4466	−3.1494

Note: see the note to Table 1. “*” indicates significance at the 10% level.

Table 6 Panel unit root tests for the joint null hypothesis of non-stationarity (Maddison data)

	Statistic	p-value	1%	5%	10%
LM	9.3506	0.0039	8.5550	7.3725	6.7644
MADF	48.0690	0.0457	55.6803	47.6319	43.2772

Note: see the note to Table 2.

Table 7 Test for unit root in individual series using panel estimation (Maddison data)

Country	$\hat{b}_{i,0} + 1$	Statistic	p-value	1%	5%	10%
Canada	0.8011	−4.3299	0.0101	−4.3401	−3.6721	−3.3629
France	0.8254	−5.6418	0.0005	−4.3775	−3.7625	−3.4232
Germany	0.8869	−3.3587	0.0929	−4.3401	−3.6401	−3.3174
Italy	0.8816	−4.9693	0.0024	−4.3982	−3.7873	−3.4514
Japan	0.9376	−2.6350	0.3261	−4.3964	−3.6778	−3.3348
United Kingdom	0.8845	−3.6285	0.0552	−4.3812	−3.6761	−3.3501

Note: see the note to Table 3.

The United Kingdom output differential series is the notable exception; ($\hat{b}_{i,0} + 1$) in Table 5 is larger than the corresponding one in Table 1. Thus, if the Maddison data set rejects the unit root hypothesis, the result is likely driven by the improved information in the longer sample.

Despite the increase in the sample size, the univariate ADF test still gives limited support for convergence. Only the French output differential series rejects the unit root hypothesis at the 10% level; indicating the US and French output data converge over time. Other output differential series show no significant evidence of convergence.

The Im *et al.* and Taylor and Sarno panel unit root tests, on the other hand, deliver a completely different picture (Table 6). Both tests convincingly reject the joint null of non-stationarity. Thus, with the information in the long historical data, the LM and MADF panel unit-root tests are able to provide strong evidence in favor of convergence, which is in sharp contrast with the results for the post-WWII data. Thus the combination of a long data series and efficient panel tests contributes to a positive result of convergence. However, there is still a question of whether all or just some countries display convergence.

The Breuer *et al.* SURADF test results highlight the possibility that the rejection of the joint non-stationary null hypothesis does not necessarily mean all the output differential series are stationary. Indeed, the SURADF test reveals a diverse pattern of convergence behavior among individual output differential series and suggests the rejection of the joint non-stationarity null hypothesis is driven by the stationarity of some but not all the output differential series. Table 7 indicates that there is no convergence between the US and Japanese output series. Again, the results point to the potential ambiguity in interpreting findings from panel unit root tests. While the results with the Maddison data set are more favorable to the convergence hypothesis, the results from the SURADF test suggests that a subgroup in the panel may be responsible for the findings from the other tests.

Broadly speaking, the results of the Choi and Ahn tests reported in Table 8 corroborate the unit root test results. For the recommended δ values, both the SBDHT and SBDH statistics fail to reject the null of simultaneous convergence between the G7 countries. These results reinforce those in Tables 6 and 7 and constitute strong evidence of convergence. Nevertheless, given the SURADF test

Table 8 Panel tests for the joint null hypothesis of stationarity (Maddison data)

	Statistic	δ	p-value	1%	5%	10%
SBDHT	0.61362	10	0.0652	2.0712	0.6797	0.5517
SBDH	0.73289	10	0.1094	3.1892	1.0205	0.7505
SBDHT	0.61750	12	0.1479	3.0010	0.9023	0.6951
SBDH	0.73526	12	0.2028	3.8024	1.2270	0.9036
SBDHT	0.64328	14	0.3003	5.2921	1.5270	1.0166
SBDH	0.75565	14	0.3520	6.2656	1.8393	1.2345

Note: see the note to Table 4.

results, the non-rejection of the joint stationarity null also disguises the possibility that some series in the system are non-stationary.

4.3 Discussion

When the same set of procedures are applied to the output differential series from Summers and Heston (1991) and Maddison (1995), they generate different inferences on output convergence. It can be argued that the discrepancy is attributed to sample differences. As the Maddison sample is longer than the Summers and Heston sample, the former contains more information about output dynamics than the latter. This observation is consistent with the findings that the Maddison sample offers much more precise statistical evidence of convergence than the Summers and Heston data set. Thus, data information, which has significant implications for discriminating among different types of output convergence behavior, is an important factor in studying the output dynamics. Empirical studies on output convergence should benefit from both the use of a longer sample period and more efficient panel procedures.

It is known that the data from Summers and Heston (1991) and Maddison (1995) are constructed differently. The different ways to compile the data can impute dissimilar dynamics in the output data and lead to different convergence results. To explore this possibility, we apply the same tests to the Maddison data set for the sample period 1950–1992; that is the same period covered by the Summers and Heston data set. The results, reported in the Appendix, are qualitatively the same as those in Tables 1 to 4. Again, the inference on output convergence in the shorter sample period depends crucially on whether convergence or no convergence is taken as the null hypothesis. Thus, the information content of the data, rather than data construction method, is the probable reason behind the discrepancy in inferences on output convergence reported in the previous subsections.

To shed some light on the power of the joint hypothesis unit root and stationarity tests, we evaluate the sample-specific empirical power of the LM, MADF, SBDH, and SBDHT. In evaluating the power of the LM and MADF tests, we generate the data according to the VAR models that best fit the data and use the sample-specific finite sample critical values to determine the rejection frequency. The number of replication is set to 1,000.³ For the post-WWII sample period, the rejection rates of the LM and MADF tests are, respectively, 89.4% and 95.6% at the 5% significance level. For the Maddison sample period, the rejection rates are 97.5% (LM) and 99.3% (MADF). The simulation exercise indicates both the LM and MADF tests have very good empirical power.

The best-fitted integrated VAR models are used to generate artificial data to assess the sample-specific empirical power of the panel stationarity tests. Using the sample-specific finite sample critical values, the SBDH and SBDHT tests appear to have limited power to reject the integrated VAR specifications. For

³ The detailed simulation results are available from the authors.

instance, consider the 5% significance level and $\delta = 10$. In the post-WWII sample period, both SBDH and SBDHT tests have a rejection rate lower than the size of the test. For the Maddison sample period, the empirical power improves slightly to 10.0% (SBDH) and 16.6% (SBDHT). Apparently, the simulation results reflect the usual power concern about statistics based on nonparametric kernel methods to control for heteroskedasticity and serial correlation.

5. Concluding remarks

The time-series framework is used to investigate the presence (absence) of output convergence among the G7 countries. Arguments for and against the output convergence hypothesis are made based on whether an output differential series is stationary or has a unit root. In this exercise, we approach the empirical issue of output convergence from two perspectives. First, we consider no convergence as the null hypothesis. In this case, the hypothesis of no convergence enjoys the benefit of doubt in the sense that its rejection requires some strong evidence against it. Second, we assess the null hypothesis of convergence. The study of cross-country output dynamics from both viewpoints gives an equal opportunity for both convergence and no convergence to be validated by the data as the null hypothesis.

Our empirical results suggest that the inference about output convergence can be dictated by the choice of a null hypothesis. A conclusion of no output convergence can be reached just because no convergence is considered as the null hypothesis. Further, the no-convergence result reported in previous studies pursuing the time-series definition may be attributed to the low power of the test procedures being used. While short output data series or the use of univariate unit root procedures yields very limited support for the convergence hypothesis, the combination of long sample and efficient panel procedures delivers a more favorable result for the same hypothesis.

In addition to the issue of output convergence, the empirical exercise raises a few interesting observations. For example, the results from a typical panel unit root or stationarity test have to be interpreted with caution. A non-rejection of a joint non-stationarity (stationarity) null hypothesis is not a *sine qua non* for all the series to be non-stationary (stationary). Similarly, the stationarity (non-stationarity) of a subset of series can lead to the rejection of a joint non-stationarity (stationarity) null hypothesis. Specific panel procedures have to be implemented to determine the stationarity property of individual series. In our exercise, the use of the Breuer *et al.* SURADF test helps identify the countries in the sample that converge.

Another issue is related to the presence of convergence clubs (Quah, 1997; Evans, 1998). In the two data samples examined, there are signs that countries have diverse convergence patterns. Even among the G7 nations that are quite homogenous the convergence in output does not occur simultaneously across all the countries. The result lends considerable support to the notion of convergence clubs in which member countries converge to a club-specific steady state. Thus, empirical studies of output convergence should allow for the presence of convergence clubs.

Acknowledgements

The views contained herein do not necessarily represent those of the International Monetary Fund or any other organizations the authors are associated with.

References

- Aghion, P. and Howitt, P. (1992). 'A model of growth through creative destruction', *Econometrica*, **60**, 323–51.
- Andrews, D. (1991). 'Heteroskedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica*, **59**, 817–58.
- Andrews, D. and Monahan, J. (1992). 'An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator', *Econometrica*, **60**, 953–66.
- Barro, R.J. (1991). 'Economic growth in a cross-section of countries', *Quarterly Journal of Economics*, **106**, 407–43.
- Barro, R.J. and Sala-i-Martin, X. (1992). 'Convergence', *Journal of Political Economy*, **100**, 223–51.
- Barro, R.J. and Sala-i-Martin, X. (1995). *Economic Growth*, McGraw Hill Inc., New York, 414–61.
- Baumol, W.J. (1986). 'Productivity growth, convergence and welfare: what the long-run data show', *American Economic Review*, **76**, 1072–85.
- Ben-David, D. (1994). 'Convergence clubs and diverging economies', Discussion Paper No. 922, CEPR, London.
- Bernard, A. and Durlauf, S. (1995). 'Convergence in international output', *Journal of Applied Econometrics*, **10**, 97–108.
- Bernard, A. and Durlauf, S. (1996). 'Interpreting tests of the convergence hypothesis', *Journal of Econometrics*, **71**, 161–74.
- Breuer, J.B., McNown, R., and Wallace, M. (2002). 'Series-specific unit root tests with panel data', *Oxford Bulletin of Economics and Statistics*, **64**, 527–46.
- Cheung, Y.-W. and Lai, K.S. (1995). 'Lag order and critical values of the augmented Dickey-Fuller test', *Journal of Business and Economic Statistics*, **13**, 227–80.
- Choi, I. and Ahn, B.C. (1998). 'Testing the null of stationarity for multiple time series', *Journal of Econometrics*, **88**, 41–77.
- DeLong, J.B. (1988). 'Productivity growth, convergence, and welfare: comment', *American Economic Review*, **78**, 1138–54.
- 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**, 427–31.
- Durlauf, S. and Quah, D. (1999). 'The new empirics of economic growth', in J.B. Taylor and M. Woodford (eds), *Handbook of Macroeconomics*, North-Holland Elsevier Science, Amsterdam, 231–304.
- Evans, G.B.A. and Savin, N.E. (1984). 'Testing for unit roots: 2', *Econometrica*, **52**, 1241–69.
- Evans, P. (1997). 'How fast do economies converge?', *Review of Economics and Statistics*, **79**, 219–25.
- Evans, P. (1998). 'Using panel data to evaluate growth theories', *International Economic Review*, **39**, 295–306.

- Hsiao, C. (1986). *Analysis of Panel Data*, Econometric Society monograph 11, Cambridge University Press, Cambridge.
- Grossman, G.M. and Helpman, E. (1991). *Innovation and Growth in the Global Economy*, The MIT Press, Cambridge, MA, 43–111.
- Im, K.S., Pesaran, M.H., and Shin, Y. (1997). 'Testing for unit roots in heterogeneous panels', mimeo, Department of Applied Economics, University of Cambridge.
- Levin, A. and Lin, C.F. (1992). 'Unit root in panel data: asymptotic and finite sample properties', Working Paper 92–63, Department of Economics, University of California, San Diego.
- Levin, A. and Lin, C.F. (1993). 'Unit root tests in panel data: new results', Working Paper 93–56, Department of Economics, University of California, San Diego.
- Li, Q. and Papell, D. (1999). 'Convergence of international output: time series evidence for 16 OECD countries', *International Review of Economics and Finance*, 8, 267–80.
- Maddison, A. (1995). *Monitoring the World Economy, 1820–1992*, OECD, Washington, DC.
- Mankiw, G., Romer, D., and Weil, D. (1992). 'A contribution to the empirics of economic growth', *Quarterly Journal of Economics*, 107, 407–37.
- O'Connell, P. (1998). 'The overvaluation of purchasing power parity', *Journal of International Economics*, 44, 1–19.
- Papell, D. and Theodoridis, H. (2001). 'The choice of numeraire currency in panel tests of purchasing power parity', *Journal of Money, Credit, and Banking*, 33, 790–803.
- Quah, D. (1992). 'International patterns of growth: I. Persistence in cross-country disparities', Working Paper, London School of Economics, October.
- Quah, D. (1993). 'Galton's fallacy and tests of the convergence hypothesis', *The Scandinavian Journal of Economics*, 95, 427–43.
- Quah, D. (1994). 'Exploiting cross-section variations for unit root inference in dynamic data', *Economic Letters*, 44, 9–19.
- Quah, D. (1997). 'Empirics for growth and distribution: stratification, polarization, and convergence clubs', *Journal of Economic Growth*, 2, 27–59.
- Romer, P. (1990). 'Endogenous technical change', *Journal of Political Economy*, 98, 71–102.
- Sargan, J.D. and Bhargava, A. (1983). 'Testing for residuals from least squares regression for being generated by the Gaussian random walk', *Econometrica*, 51, 153–74.
- Sarno, L. and Taylor, M. (1998). 'Real exchange rates under the recent float: unequivocal evidence of mean reversion', *Economics Letters*, 60, 131–7.
- Solow, R.M. (1956). 'A contribution to the theory of economic growth', *Quarterly Journal of Economics*, 70, 65–94.
- Summers, R. and Heston, A. (1991). 'PennWorld Table (Mark 5): an expanded set of international comparisons, 1955–1988', *Quarterly Journal of Economics*, 106, 327–68.
- Swan, T. (1956). 'Economic growth and capital accumulation', *Economic Record*, 32, 334–61.
- Taylor, M. and Sarno, L. (1998). 'The behavior of real exchange rates during the post-Bretton Woods period', *Journal of International Economics*, 46, 281–312.
- Temple, J. (1999). 'The new growth evidence', *Journal of Economic Literature*, 37, 112–56.
- West, K.D. (1987). 'A note on the power of least squares tests for a unit root', *Economics Letters*, 24, 249–52.

Appendix

Results for the 1950–1992 Maddison data set

Table A1 Univariate ADF test

Country	$\hat{b}_{i,0}+1$	statistic	1%	5%	10%
Canada	0.5966	−2.6841	−4.1948	−3.5253	−3.1988
France	0.9311	−1.1421	−4.1948	−3.5253	−3.1988
Germany	0.8612	−4.1593**	−4.1719	−3.5048	−3.1792
Italy	0.9169	−1.6579	−4.1948	−3.5253	−3.1988
Japan	0.9559	−1.3279	−4.1719	−3.5048	−3.1792
United Kingdom	0.6023	−3.3271*	−4.1381	−3.4667	−3.1404

Note: see the note to Table 1.

Table A2 Panel unit root tests for the joint null hypothesis of non-stationarity

	Statistic	p-value	1%	5%	10%
LM	5.3713	0.3294	8.1102	7.0742	6.5339
MADF	14.761	0.9746	58.9337	48.8404	44.6416

Note: see the note to Table 2.

Table A3 Test for unit root in individual series using panel estimation

Country	$\hat{b}_{i,0}+1$	Statistic	p-value	1%	5%	10%
Canada	0.6233	−1.7019	0.7780	−4.8399	−4.0188	−3.6294
France	0.9441	−0.9348	0.9614	−4.9131	−4.1562	−3.7735
Germany	0.8563	−5.6886	0.0015	−4.8686	−4.0653	−3.6765
Italy	0.8819	−2.5893	0.4808	−4.8910	−4.1588	−3.7871
Japan	0.9651	−1.2061	0.9060	−4.8778	−4.0738	−3.6994
United Kingdom	0.6637	−2.9249	0.2687	−4.7443	−3.9378	−3.5703

Note: see the note to Table 3.

Table A4 Panel tests for the joint null hypothesis of stationarity

	Statistic	δ	p-value	1%	5%	10%
SBDHT	1.1176	8	0.4007	10.8801	3.0511	2.1361
SBDH	1.6457	8	0.2357	11.1896	3.3939	2.4016
SBDHT	1.0645	10	0.5209	41.1462	8.9955	4.7205
SBDH	2.2710	10	0.2819	47.0899	9.2506	4.9326
SBDHT	1.0802	12	0.4324	40.3842	9.7757	4.9781
SBDH	2.0499	12	0.3154	43.1592	10.0445	5.4887

Note: see the note to Table 4.