CESifo *Working Paper Series*

# TESTING FOR OUTPUT CONVERGENCE: A RE-EXAMINATION

Yin-Wong Cheung
Antonio Garcia Pascual

Working Paper No. 319

July 2000

*CESifo*
*Poschingerstr. 5*
*81679 Munich*
*Germany*
*Phone: +49 (89) 9224-1410/1425*
*Fax: +49 (89) 9224-1409*
*http://www.CESifo.de*

CESifo *Working Paper No. 319*
*July 2000*

# TESTING FOR OUTPUT CONVERGENCE: A RE-EXAMINATION


## Abstract

This paper investigates output convergence for the G7 countries using multivariate 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 multivariate unit root and stationarity tests.


Keywords: Output convergence, multivariate test, unit root test, stationarity test

JEL Classification: O40, C32

*Yin-Wong Cheung*
*Department of Economics*
*University of California*
*Santa Cruz, CA 95064*
*USA*

*Antonio Garcia Pascual*
*Department of Economics*
*University of Munich*
*Ludwigstr. 28*
*80539 Munich*
*Germany*
*email: antonio.pascual@lrz.uni-muenchen.de*

# 1 Introduction

One of the major differences between neoclassical and endogenous growth models is their predictions of the behavior of national output. 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 the 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 the country-specific factors can evolve endogenously according to the environment unique to a nation, countries with dissimilar initial endowments and attributes can have per capita output not converging over time. The different views on national output have spurred considerable interest on testing if 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.

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 the issue of growth dynamics. Output convergence, under the time series framework, requires the cross-country per capita output differentials to be stationary; that is, the levels of national output per capita are not diverging over time.

The existing evidence derived from the time series approach is not very favorable to the notion of convergence. Bernard and Durlauf (1995), using the 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 U.S. data. Using a multivariate 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 and their per capita output relative to the group aggregate. After allowing for structural breaks in their relative per capita output series, the authors uncover a strong evidence of convergence.

Compared with the cross-country analysis, the time series approach yields some relatively unconvincing 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.

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 logarithmic per capita output relative to the U.S. 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 a statistical procedure will not reject either the null hypothesis of convergence or no convergence. To highlight the power issue, we consider two types of multivariate time series tests - one has unit root (no convergence) as the null hypothesis and the other has stationarity (convergence) as the null. Multivariate procedures are employed because their ability to reject a false null is higher than the corresponding univariate procedures. The results from procedures with different specifications of the null hypothesis helps 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 multivariate time series procedures. First, we account for the effect of cross-country output correlations on the multivariate tests. The second issue is related to the interpretation of the joint hypothesis considered by multivariate tests. For a typical multivariate procedure, the rejection of, say, the no-convergence null hypothesis means the rejection of the joint hypothesis that individual output pairs are divergent. However, it may only take convergence in a subset of countries, which is not identified by the test, to obtain the rejection result. To address the possible ambiguity, we will examine the convergence behavior of individual country pairs in a multivariate 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 the Monte Carlo method to generate critical values from artificial data generated according to parameters estimated from actual output series.

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 multivariate unit root tests (Im *et al.*, 1997; Taylor and Sarno, 1998), a multivariate framework to evaluate the convergence behavior of individual country pairs (Breuer *et al.*, 1999), and the multivariate 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, one from 1950 to 1992 and the other from 1885 to 1994, are considered. 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. 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 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 is taken as a requirement that, for two countries, their per capita output differentials are stationary. In this case, the test for convergence is translated to a test for the stationarity of output differentials.

For the empirical studies adopting the notion of time series convergence, they 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 per capita output at time t and $Y_{*,t}$ is the output variable of the benchmark country. The no-convergence hypothesis is stated as

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

where $x_{i,t}$ is the 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 nonstationary process. Standard procedures such as the 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 persistence alternatives. The use of (1) as the null 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. When the data offer a clear indication of nonconvergence, the stationary null hypothesis will be rejected. The null of convergence is given by

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

where $I(0)$ denotes a stationary stochastic process.

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 leveled field to evaluate output convergence dynamics. However, it does not directly address the issue of the test power. One way to extract more information from cross-country data is to use a multivariate procedure, which incorporates interactions between data series. In this exercise, we employ a few multivariate 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-WW II sample, we examine a longer history data set to see if it gives a more definite conclusion on convergence dynamics.

One important qualification of the time-series based approach to study convergence is that the data are generated by a time-invariant data-generating process. The time-invariant property requires the countries under investigation to be near their long-run equilibria such that transient dynamics do not affect their steady-state output behavior. Thus, the time series test may have poor performance when applied to a group of diverse countries at different stages of economic development. To alleviate the effects of transitional dynamics on our empirical results, we work with data from the G7 countries.

# 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 multivariate 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 on to the next one.

## 3.1 Univariate Unit Root Test

The augmented Dickey-Fuller test (ADF, hereafter) is arguably the most widely used unit root test in economics. The ADF test is also commonly used to investigate the stationarity property of individual output differential series. The test is based on the regression equation

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

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) labelled 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 $\alpha_i$ (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 multivariate procedures are adopted to enhance the power performance.

## 3.2 Multivariate Unit Root Tests

Over the past decade, several statistical procedures are developed to test for unit roots in a multivariate setting (Levin and Lin, 1992, 1993; Quah, 1994). One potential benefit of using a multivariate 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 multivariate 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 multivariate framework. The LM statistic is based on the average of individual LM statistics for testing $b_{i,0} = 0$, for all $i = 1, ..., N$. Specifically, rewriting (3) in matrix-notation:

$$\Delta X_i = b_{i,0} X_{i,-1} + Q_i \gamma_i + \varepsilon_i$$

where

$$\begin{aligned}
\Delta X_i &= \left(x_{i,1+p_i}, ..., x_{i,T}\right), \\
X_{i,-1} &= \left(x_{i,p_i}, x_{i,1+p_i}, ..., x_{i,T-1}\right), \\
\gamma_i &= \left(\alpha_i, b_{i,1}, ..., b_{i,p_i-1}\right)', \\
Q_i &= \left(1_{T_i}, \Delta X_{i,-1}, \Delta X_{i,-2}, ..., \Delta X_{i,-p_i+1}\right),
\end{aligned}$$

and $1_T$ is a $(T - p_i) \times 1$ vector of unity. The LM statistic is given by

$$LM = \frac{1}{N} \sum_{i=1}^{N} LM_i \tag{4}$$

where $LM_i$ is

$$LM_i = T \frac{(\Delta X_i' P_i^* \Delta X_i)}{(\Delta X_i' M_{Q_i} \Delta X_i)}$$

$$\begin{aligned} M_{Q_i} &= I_T - Q_i \left(Q_i' Q_i\right)^{-1} Q_i' \\ P_i^* &= M_{Q_i} X_{i,-1} \left(X_{i,-1}' M_{Q_i} X_{i,-1}\right)^{-1} X_{i,-1}' M_{Q_i}. \end{aligned}$$

The statistic (4), under conditions given in Im *et al.* (1997), has an asymptotic standard normal distribution under the null hypothesis that $b_{i,0} = 0$, for all $i = 1, ..., N$. The condition that $b_{i,0} = 0$ for all $i = 1, ..., N$ implies 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 Multivariate ADF Test

Sarno and Taylor (1998) and Taylor and Sarno (1998) propose a multivariate version of the ADF (MADF) test. To implement the test, (3) is estimated as a system of $N$ equations using the feasible generalized least squares (FGLS) technique and the standard Wald test statistic is used to evaluate the joint hypothesis

$$H_0 : b_{1,0} = b_{2,0} = ... = b_{N,0} = 0.$$

Again, the case of no convergence between any pair of countries is the null hypothesis. 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 Multivariate 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, the multivariate test results should be interpreted with caution. The joint hypothesis of all country pairs do not converge is invalidated if one or more 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. The possibility that countries may converge to different growth rates is consistent with the notion of convergence clubs reported in, for example, Ben-David (1994), Quah (1997), Evans (1998), and Durlauf and Quah (1999).

Breuer *et al.* (1999) devise a multivariate procedure to test for a unit root in an individual series. The procedure constructs the test statistics within a multivariate 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 stationarity/nonstationary. The FGLS method is used to estimate the system of $N$ equations and the null hypothesis

$$H_0 : b_{i,0} = 0 \ , \tag{5}$$

is examined individually for $i = 1, ..., N$. 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.* (1999) recommend the use of sample-specific critical values to conduct the statistical inference.

## 3.3 Multivariate Stationarity Tests

The tests presented in the previous subsections have nonstationarity; that is, 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) develop a multivariate test for stationarity against nonstationary alternatives. Their procedures are semi-nonparametric and do not require information on the parametric ARMA structure of the data. Compared with the multivariate 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:

$$SBDH = tr\left[\left(\frac{1}{T^2}\sum_{t=1}^{T}\tilde{S}_t\tilde{S}_t'\right)\tilde{\Omega}_l\right]. \tag{6}$$

$tr$ 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

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

$\tilde{\Omega}_l$ is the heteroscedasticity and autocorrelation consistent covariance matrix estimator,

$$\begin{aligned} \tilde{\Omega}_l &= \sum_{n=-l}^{l} C(n)k(n/l), \qquad (7) \\ C(n) &= \begin{cases} T^{-1}\sum_{t=2}^{T-n}\Delta\tilde{S}_t\Delta\tilde{S}_{t+n}' \text{ , for } n \geq 0 \\ \\ T^{-1}\sum_{t=2}^{T+n}\Delta\tilde{S}_{t-n}\Delta\tilde{S}_t' \text{ , for } n < 0 \end{cases} \end{aligned}$$

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 prewhitening 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 $[\delta(T/100)^{.25}]$ when the data-dependent bandwidth parameter is greater than $T^{.65}$. 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, instead of $\Delta\tilde{S}_t$, 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 (6), we label the resulting statistic SBDHT. Both the SBDH and SBDHT statistics test the null hypothesis of all individual series are stationary against the alternative of at least one series is nonstationary. That is, under the null hypothesis output data from all the countries 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 multivariate unit root tests.

## 3.4 Sample-Specific Critical Values

The finite-sample behavior of the multivariate 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 multivariate unit root test, a pseudo-random sample $N \times (T + 100)$ is generated according to the estimated ARIMA specifications and covariance matrix 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 multivariate stationary test 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 as the dynamics of these data are likely to be characterized as time-invariant, which is a maintained assumption of the time-series based approach to study output convergence.

## 4.1 The Summers and Heston Data Set: 1950-1992

The output data from Penn World Table are perhaps the most 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. Before we conduct any statistical tests, we subtract the U.S. real per capita GDP data from the other G7 countries and call the resulting series output differentials for convenience.

### 4.1.1 Univariate Analysis

First, we apply the ADF test to the output differentials. The null hypothesis of nonstationarity is interpreted as no convergence with respect to the U.S. data. The results, presented in Table 1, show limited evidence of convergence. 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-U.K.) 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-U.K.) to 0.76 (France-Italy).

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

#### 4.1.2 Multivariate Analysis

The results of the Im *et al.* multivariate unit root 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 multivariate procedures have better power than the univariate ADF test, they do not provide a 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 nonstationary. 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 nonstationary. Due to the lack of power, a multivariate unit root test may disguise the stationarity of a member series. As indicated in Table 1, the German output differential series is stationary. The multivariate unit root tests, nonetheless, do not reject the null that all the series are nonstationary. Thus, one should interpret these multivariate test results with caution.

To investigate pairwise output convergence in the multivariate setting we used the Breuer *et al.* test. The results are given in Table 3 and are similar to those based on the univariate ADF tests. There is only evidence of convergence between the German and U.S. output data as the German output differential series is found to be stationary. For the other countries, there is no sign of pairwise convergence with the U.S. output. Thus, the potential power gain from using a multivariate framework does not yield stronger evidence for pairwise convergence. Compared with the multivariate tests, the univariate ADF seems to still have a role in studying pairwise convergence.

The test results derived from both univariate and multivariate 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 rejection the null hypothesis - either because of the low power of the procedures or the uninformative data. To shed light on such a possibility, we apply the Choi and Ahn multivariate stationarity test to the output differential data. Following Choi and Ahn (1999), we set the lag truncation parameter $\delta$ 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

Results in the previous subsection indicate that the post-WWII 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. Again, the G7 data on real annual per capita output differential series are examined. The sample period is from 1885 to 1994.

The result of the univariate and multivariate unit root tests are reported in Tables 5, 6, and 7. Again the univariate ADF test gives limited support for convergence. Only the French output differential series rejects the unit root hypothesis at the 10% level; indicating the U.S and French output data converge over time. Other country pairs show no significant evidence of convergence.

The Im *et al.* and Taylor and Sarno multivariate 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, the multivariate unit-root tests based on a long historical data set 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 multivariate tests contributes to a positive result of convergence. However, there is still a question of whether all or just some country pairs show convergence.

The Breuer *et al.* test results show that the rejection of the joint non-stationary null hypothesis does not necessarily mean all the output differential series are stationary. In fact, there is a diverse pattern of convergence behavior among individual output differential series behind the rejection result from the Im *et al.* and Taylor and Sarno tests. Table 7 indicates that there is no convergence between the U.S. and Japanese output series. Again, the results point to the potential ambiguity in interpreting findings from multivariate unit root tests. While the Maddison data set is more favorable to the convergence hypothesis, there is still a sign of different subgroup output behavior among the G7 countries.

Broadly speaking, the results of the Choi and Ahn tests reported in Table 8 collaborate the unit root test results. For the recommended $\delta$ 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 Breuer *et al.* test 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. The discrepancy is likely due to sample differences. As the Maddison sample is longer than the Summers and Heston one, 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 multivariate 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 contraction method, is probably the reason behind the discrepancy in inferences on output convergence reported in the previous subsections.

# 5 Concluding Remarks

The time-series framework is used to examine if the data on real per capita output of the G7 countries converge over time. Under the time-series framework, a stationary output differential series is interpreted as evidence of convergence and the presence of a unit root in the series is viewed as proof against convergence. 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. With short output data series or univariate unit root procedures, there is only very limited support for the convergence hypothesis. On the other hand, a combination of long sample and efficient multivariate 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 multivariate unit root or stationarity test have to be interpreted with caution. A non-rejection of a joint non-stationarity (stationarity) null 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 multivariate procedures have to be implemented to identify the stationarity property of individual series.

Another issue is related to the presence of convergence clubs (Quah, 1997; Evans, 1998). The choice of the G7 countries should make our sample close to the requirement that the underlying data generating process is time-invariant. In the two data samples examined, there are signs that different country pairs have diverse convergence patterns. Even among the G7 group, 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.

# References


Aghion P, Howitt P. 1992. A model of growth through creative destruction. *Econometrica 60*: 323-351.

Andrews D. 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*: 817-858.

Andrews D, Monahan J. 1992. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica 60*: 953-966.

Barro RJ. 1991. Economic growth in a cross-section of countries. *Quarterly Journal of Economics 106*: 407-443.

Barro RJ, Sala-i-Martin X. 1992. Convergence. *Journal of Political Economy 100*: 223-251.

Barro RJ, Sala-i-Martin X. 1995. *Economic Growth*, McGraw Hill Inc. Chapter 12: 414-461.

Baumol WJ. 1986. Productivity growth, convergence and welfare: what the long-run data show. *American Economic Review 76*: 1072-1085.

Ben-David D. 1994. Convergence clubs and diverging economies. CEPR Discussion Paper # 922.

Bernard A, Durlauf S. 1995. Convergence in international output. *Journal of Applied Econometrics 10*: 97-108.

Bernard A, Durlauf S. 1996. Interpreting tests of the convergence hypothesis. *Journal of Econometrics 71*: 161-174.

Breuer JB, McNown R, Wallace M. 1999. Series-specific tests for a unit root in a panel setting with an application to real exchange rates. Mimeo.

Cheung Y-W, Lai KS. 1995. Lag order and critical values of the augmented Dickey-Fuller test. *Journal of Business and Economic Statistics 13*: 227-280.

Choi I, Ahn BC. 1998. Testing the null of stationarity for multiple time series. *Journal of Econometrics 88*: 41-77.

DeLong JB. 1988. Productivity growth, convergence, and welfare: Comment. *American Economic Review 78*: 1138-54.

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

Durlauf S, and Quah D. 1999. The new empirics of economic growth. *Handbook of Macroeconomics*, Taylor JB, Woodford M (eds), North-Holland Elsevier Science. Chapter 4: 231–304.

Evans GBA, Savin NE. 1984. Testing for unit roots: 2. *Econometrica 52*: 1241-1269.

Evans P. 1997. How fast do economies converge. *Review of Economics and Statistics 79*: 219-225.

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.

Grossman GM, Helpman E. 1991. *Innovation and Growth in the Global Economy.* The MIT Press, Cambridge, MA. Chapters 3-4: 43-111.

Im KS, Pesaran MH, Shin Y. 1997. Testing for unit roots in heterogeneous panels. Mimeo.

Levin A, Lin C-F. 1992. Unit root in panel data: asymptotic and finite-sample properties. University of California, San Diego Working Paper 92-63.

Levin A, Lin C-F. 1993. Unit root tests in panel data : new results. University of California, San Diego Working Paper 93-56.

Li Q, Papell D. 1999. Convergence of international output: Time series evidence for 16 OECD countries. *International Review of Economics and Finance 8*: 267-280.

Maddison A. 1995. *Monitoring the World Economy, 1820-1992*, OECD, Washington, DC.

Mankiw G, Romer D, Weil D. 1992. A contribution to the empirics of economic growth. *Quarterly Journal of Economics 107*: 407-437.

O'Connell P. 1998. The overvaluation of purchasing power parity. *Journal of International Economics 44*: 1-19.

Papell D, Theodoridis H. 2000. The choice of numeraire currency in panel tests of purchasing power parity. Forthcoming in *Journal of Money, Credit, and Banking.*

Quah D. 1992. International patterns of growth: I. Persistence in cross-country disparities. LSE Working Paper, October.

Quah D. 1993. Galton's fallacy and tests of the convergence hypothesis. *The Scandinavian Journal of Economics 95:* 427-443.

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 JD, Bhargava A. 1983. Testing for residuals from least squares regression for being generated by the Gaussian random walk. *Econometrica 51*: 153-174.

Sarno L, Taylor M. 1998. Real exchange rates under the recent float: unequivocal evidence of mean reversion. *Economics Letters 60*: 131-137.

Solow RM. 1956. A contribution to the theory of economic growth. *Quarterly Journal of Economics 70*: 65-94.

St Aubyn M. 1999. Convergence across industrialised countries (1980-1989): New results using time series methods. *Empirical Economics 24*: 23-44.

Summers R, Heston A. 1991. Penn World Table (Mark 5): An expanded set of international comparisons, 1955-1988. *Quarterly Journal of Economics 106*: 327-368.

Swan T. 1956. Economic growth and capital accumulation. *Economic Record 32*: 334-361.

Taylor M, Sarno L. 1998. The behavior of real exchange rates during the post-Bretton Woods period. *Journal of International Economics 46*: 281-312.

West, KD. 1987. A note on the power of least squares tests for a unit root. *Economics Letters 24*: 249-252.

**Table 1**: Univariate ADF test (Penn World Table)

| Country | statistic | 1% | 5% | 10% |
|---|---|---|---|---|
| Canada | -1.6402 | -4.1719 | -3.5048 | -3.1792 |
| France | -1.2987 | -4.1948 | -3.5253 | -3.1988 |
| Germany | (-4.0251)** | -4.1719 | -3.5048 | -3.1792 |
| Italy | -1.8415 | -4.1948 | -3.5253 | -3.1988 |
| Japan | -1.4318 | -4.1381 | -3.4667 | -3.1404 |
| United Kingdom | -2.5844 | -4.1381 | -3.4667 | -3.1404 |

Note: Critical values from Cheung and Lai (1995) are used.

( )** indicates significance at the 5% level.

**Table 2**: Multivariate unit root tests for the joint null of no stationarity $H_0 : b_{1,0} = b_{2,0} = ... = b_{N,0} = 0$ (Penn World Table)

| test | statistic | p-value | 1% | 5% | 10% |
|---|---|---|---|---|---|
| IPS | 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.* and Taylor and Sarno multivariate unit root tests are given in rows labelled IPS and MADF. The p-values and the 1%, 5%, and 10% sample-specific critical values are computed from 10,000 Monte Carlo simulations.

**Table 3**: Individual unit root test using panel estimation, $H_0 : b_{i,0} = 0$, (Penn World Table)

| Country | statistic | p-value | 1% | 5% | 10% |
|---|---|---|---|---|---|
| Canada | -1.3889 | 0.8624 | -4.6653 | -3.9788 | -3.6132 |
| France | -1.7821 | 0.7859 | -4.9229 | -4.0977 | -3.7222 |
| Germany | -5.4472 | 0.0019 | -4.6977 | -3.9768 | -3.5889 |
| Italy | -2.8308 | 0.3567 | -4.8495 | -4.0354 | -3.6833 |
| Japan | -1.4775 | 0.8012 | -4.8029 | -3.9876 | -3.6008 |
| United Kingdom | -2.9617 | 0.2632 | -4.7801 | -4.0186 | -3.6187 |

Note: Results of the Breuer *et al.* test, which investigates pairwise output convergence in the multivariate setting, 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**: Multivariate tests for the joint null of stationarity (Penn World Table)

| test | statistic | $\delta$ | 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 multivariate stationarity tests are given. The p-values and the 1%, 5%, and 10% sample-specific critical values are computed from 10,000 Monte Carlo simulations.

**Table 5**: Univariate ADF test (Maddison data)

| Country | lag | statistic | 1% | 5% | 10% |
|---|---|---|---|---|---|
| Canada | 2 | -2.9644 | -4.0239 | -3.4377 | -3.1414 |
| France | 2 | (-3.3140)* | -4.0239 | -3.4377 | -3.1414 |
| Germany | 4 | -2.7634 | -4.0030 | -3.4201 | -3.1252 |
| Italy | 2 | -2.7765 | -4.0239 | -3.4377 | -3.1414 |
| Japan | 1 | -1.4747 | -4.0351 | -3.4466 | -3.1494 |
| United Kingdom | 1 | -2.4949 | -4.0351 | -3.4466 | -3.1494 |

Note: see the note to Table 1.

( )* indicates significance at the 10% level.

**Table 6**: Multivariate unit root tests for the joint null of no stationarity, $H_0 : b_{1,0} = b_{2,0} = ... = b_{N,0} = 0$ (Maddison data)

| test | statistic | p-value | 1% | 5% | 10% |
|---|---|---|---|---|---|
| IPS | 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**: Individual unit root test using panel estimation, $H_0 : b_{i,0} = 0$ (Maddison data).

| Country | statistic | p-value | 1% | 5% | 10% |
|---|---|---|---|---|---|
| Canada | -4.3299 | 0.0101 | -4.3401 | -3.6721 | -3.3629 |
| France | -5.6418 | 0.0005 | -4.3775 | -3.7625 | -3.4232 |
| Germany | -3.3587 | 0.0929 | -4.3401 | -3.6401 | -3.3174 |
| Italy | -4.9693 | 0.0024 | -4.3982 | -3.7873 | -3.4514 |
| Japan | -2.6350 | 0.3261 | -4.3964 | -3.6778 | -3.3348 |
| United Kingdom | -3.6285 | 0.0552 | -4.3812 | -3.6761 | -3.3501 |

Note: see the note to Table 3.

**Table 8**: Multivariate tests for the joint null of stationarity (Maddison data)

| test | statistic | $\delta$ | 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.

# Appendix: **Results for the 1950-1992 Maddison data set**

**Table A1**: Univariate ADF test

| Country | statistic | 1% | 5% | 10% |
|---|---|---|---|---|
| Canada | -2.6841 | -4.1948 | -3.5253 | -3.1988 |
| France | -1.1421 | -4.1948 | -3.5253 | -3.1988 |
| Germany | (-4.1593)** | -4.1719 | -3.5048 | -3.1792 |
| Italy | -1.6579 | -4.1948 | -3.5253 | -3.1988 |
| Japan | -1.3279 | -4.1719 | -3.5048 | -3.1792 |
| United Kingdom | (-3.3271)* | -4.1381 | -3.4667 | -3.1404 |

Note: see the note to Table 1.

**Table A2**: Multivariate unit root tests for the joint null of non-stationarity, $H_0 : b_{1,0} = b_{2,0} = ... = b_{N,0} = 0$

| test | statistic | p-value | 1% | 5% | 10% |
|---|---|---|---|---|---|
| IPS | 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**: Individual unit root test using panel estimation, $H_0 : b_{i,0} = 0$

| Country | statistic | p-value | 1% | 5% | 10% |
|---|---|---|---|---|---|
| Canada | -1.7019 | 0.7780 | -4.8399 | -4.0188 | -3.6294 |
| France | -0.9348 | 0.9614 | -4.9131 | -4.1562 | -3.7735 |
| Germany | -5.6886 | 0.0015 | -4.8686 | -4.0653 | -3.6765 |
| Italy | -2.5893 | 0.4808 | -4.8910 | -4.1588 | -3.7871 |
| Japan | -1.2061 | 0.9060 | -4.8778 | -4.0738 | -3.6994 |
| United Kingdom | -2.9249 | 0.2687 | -4.7443 | -3.9378 | -3.5703 |

Note: see the note to Table 3.

**Table A4**: Multivariate tests for the joint null of stationarity

| test | statistic | $\delta$ | 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.