

CONVERGENCE OF OUTPUT IN THE G-7 COUNTRIES

Neil R. Ericsson and Jonathan R. Halket*

February 12, 2002

Abstract: Tests of convergence in a time series framework have historically been formulated as univariate tests of stationarity of the output differential. However, convergence is a multivariate hypothesis, so a multivariate framework is more natural for testing the hypothesis of convergence. One such framework is the cointegrated vector autoregression, with the Johansen procedure providing the statistical basis for tests of convergence. This approach yields novel results relative to those obtained from the univariate approach. For instance, the multivariate approach detects output convergence among France, Germany, and Italy over the last century, whereas the univariate approach fails to detect any such convergence.

Keywords: cointegration, convergence, economic growth, long run, output.

JEL classifications: F43, C12.

All correspondence should be addressed to the first author (NRE).

Neil R. Ericsson

Stop 20, Division of International Finance, Federal Reserve Board

2000 C Street, N.W., Washington, D.C. 20551 U.S.A.

(1) (202) 452-3709 (office) (1) (202) 263-4850 (fax)

ericsson@frb.gov (email)

Jonathan R. Halket

Stop 20, Division of International Finance, Federal Reserve Board

2000 C Street, N.W., Washington, D.C. 20551 U.S.A.

(1) (202) 452-2902 (office) (1) (202) 263-4850 (fax)

jonathan.rl.halket@frb.gov (email)

*The authors are a staff economist and a research assistant in the Division of International Finance, Board of Governors of the Federal Reserve System, Washington, D.C. 20551 U.S.A. They may be reached on the Internet at ericsson@frb.gov and jonathan.rl.halket@frb.gov respectively. The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System. We are grateful to Antonio Garcia Pascual for invaluable assistance in replicating results in Cheung and Pascual (2000). All numerical results were obtained using PcGive Professional Versions 9.2 and 10.0b: see Doornik and Hendry (1996, 2001).

1 Introduction

Many papers have investigated the convergence of output across countries. The standard neoclassical growth model implies that per capita real incomes in different countries should tend to converge, or at least should differ by only a stationary amount, whereas endogenous growth models may permit divergence in long-term growth paths. Whether output converges or not has substantive economic consequences for the countries involved. Convergence, or the lack thereof, also has possible policy implications, in so far as policy might affect economic growth.

Historically, tests of convergence in a time series framework have been formulated as univariate tests of stationarity of the output differential. In this framework, empirical evidence for convergence turns on whether (e.g.) the corresponding Dickey–Fuller statistic rejects the null hypothesis of a nonstationary output differential. Recent studies in this vein include Lee, Pesaran, and Smith (1997), Evans (1998), Li and Papell (1999), and Linden (2000).

This univariate approach has economic and econometric shortcomings. Multivariate tests of cointegration using the Johansen procedure help address both concerns. Multivariate tests also can yield novel results relative to those obtained from the univariate approach, as this paper shows. The immediate application herein is to convergence of output across the G-7 countries. The multivariate approach, however, has many other potential applications, including the analysis of long-run purchasing power parity, and analyses involving real variables or relative prices.

This paper is organized as follows. Section 2 considers how univariate, bivariate, and multivariate tests of stationarity may be used to test the convergence hypothesis. Section 3 briefly describes the data, which are on real per capita GDP, as taken from Maddison (1995) and from Summers and Heston’s (1991) Penn World Table. Section 4 empirically implements these tests using those data and shows how the results can depend upon the form of the test. At a substantive economic level, Section 4 demonstrates output convergence among France, Germany, and Italy over the last century. Section 5 concludes.

2 Empirical Methodology

This section considers different empirical methodologies for testing convergence. Section 2.1 considers univariate tests of convergence—the most common approach used in the literature. Section 2.2 develops bivariate tests of convergence, which generally have greater power than the univariate tests. Section 2.3 examines the choice of numeraire. Section 2.4 generalizes the approach in Section 2.2 to multivariate tests of convergence.

Before turning to specific tests of convergence, this section delineates three mean-

ings of convergence for time series data. For ease of discussion, consider two output series y_t^a and y_t^b , where y denotes the log of real per capita output, a and b are country indexes, and t is the time index. For the definitions of convergence, y_t^a and y_t^b are assumed to be $I(1)$.

1. *Stochastic convergence* (Carlino and Mills (1993)). The series y_t^a and y_t^b exhibit stochastic convergence if $y_t^a - y_t^b$ is trend stationary. This form of convergence is relatively weak, in that the two countries' outputs could be diverging deterministically.
2. *Deterministic convergence* (Li and Papell (1999)). The series y_t^a and y_t^b exhibit stochastic convergence if $y_t^a - y_t^b$ is (level) stationary. That is, $y_t^a - y_t^b$ is $I(0)$. Deterministic convergence implies stochastic convergence, but not conversely. Deterministic convergence implies that the two output paths tend to move in parallel over the long run. However, the average of the output discrepancy $y_t^a - y_t^b$ may be nonzero, and nonzero in the long run.
3. *“Zero-mean” convergence* (Bernard and Durlauf (1996)). The series y_t^a and y_t^b exhibit zero-mean convergence if $y_t^a - y_t^b$ is (level) stationary with zero mean. Zero-mean convergence implies both deterministic and stochastic convergence, but not conversely. Under zero-mean convergence, the log-levels of output themselves tend to converge over time.

For the most part, the tests of convergence considered below are of stochastic and deterministic convergence, although implementing tests of zero-mean convergence is straightforward for the procedures discussed. See Carlino and Mills (1993), Bernard and Durlauf (1996), and Li and Papell (1999) for further discussion of the different senses of convergence. Rassekh (1998) summarizes the literature on the convergence hypothesis, and Temple (1999) discusses convergence in terms of evidence on economic growth.

The current paper does not consider convergence using cross-section (cross-country) data. As Bernard and Durlauf (1996) show, tests of convergence in cross-section data are weaker and less informative than tests of convergence using time series data. Furthermore, typical cross-country growth analyses impose a single relationship across countries with different actual relationships, confound the underlying dynamics, and misconstrue short-run for long-run properties; see Ericsson, Irons, and Tryon (2001). Such regressions are a doubtful and misleading basis for inferences about economic policy.

2.1 Univariate Tests for Convergence

Historically, tests of convergence have been formulated as univariate tests of stationarity of the output differential $y_t^a - y_t^b$. The regression for the Dickey–Fuller statistic

provides a frequently used univariate framework for testing convergence. Distinguishing between the three senses of convergence then turns on whether the intercept and the coefficient on a linear trend are nonzero in the corresponding regression. For an output differential $(y^a - y^b)_t$, that regression is:

$$\Delta(y^a - y^b)_t = \kappa(y^a - y^b)_{t-1} + \sum_{i=1}^k \gamma_i \Delta(y^a - y^b)_{t-i} + \theta_0 + \theta_1 t + u_t, \quad (1)$$

where Δ is the difference operator; κ , γ_i , θ_0 , and θ_1 are coefficients; and u_t is the error term, assumed normally and independently distributed with mean zero and constant variance σ^2 . The t -ratio on the least squares estimate of κ in (1) is the k th-order augmented Dickey–Fuller statistic, denoted $ADF(k)$. Stochastic convergence implies that $\kappa < 0$; deterministic convergence implies that $\kappa < 0$ and $\theta_1 = 0$; and zero-mean convergence implies that $\kappa < 0$ and $\theta_0 = \theta_1 = 0$. When testing for deterministic convergence, the linear trend is still sometimes included in (1) in order to obtain similarity with respect to θ_0 . See Dickey and Fuller (1979, 1981) for initial development of the statistic.

Some studies test for output convergence on a pairwise basis, as in Li and Papell (1999) and Linden (2000), the latter using non-parametric tests. Other studies implement panel unit-root tests, as in Lee, Pesaran, and Smith (1997), Evans (1998), and Cheung and Pascual (2000) for output across countries, Carlino and Mills (1993) for U.S. regional incomes, and Evans and Karras (1996) for U.S. state economies. Notably, the panel tests implicitly assume that either all regions converge or that none converge. If only some converge, inferences from the panel tests can be hazardous.

2.2 Bivariate Tests for Convergence

Each sense of convergence implies cointegration between y_t^a and y_t^b , with the cointegrating vector β' being equal to $(+1, -1)$. While several authors have noted this implication, only Bernard and Durlauf (1995) appear to have implemented tests of convergence in a *multivariate* cointegration framework. This subsection, and Section 2.4 below, develop multivariate tests of convergence, which generally have greater power to detect convergence than do univariate tests. The multivariate tests also have an increased ability to detect the *lack* of convergence. This subsection examines the analytical relationship between the univariate and bivariate approaches and then considers the relative merits of the two approaches. Section 2.4 generalizes the approach for more than two variables. For expositional convenience, the analytics in this subsection focus on the augmented Dickey–Fuller statistic and the Johansen procedure; see Johansen (1991, 1995) on the latter.

Convergence between y_t^a and y_t^b can be tested with the Johansen procedure. Specifically, letting x'_t be the vector of outputs (y_t^a, y_t^b) , the Johansen procedure analyzes

the bivariate vector autoregression (VAR) of x :

$$\Delta x_t = \pi x_{t-1} + \sum_{i=1}^k \Gamma_i \Delta x_{t-i} + \phi_0 + \phi_1 t + \varepsilon_t, \quad (2)$$

where π and Γ_i are 2×2 matrices of coefficients, ϕ_0 and ϕ_1 are 2×1 vectors of coefficients, and ε_t is the 2×1 vector error term, assumed normally and independently distributed with mean zero and constant variance matrix Σ . If y_t^a and y_t^b are cointegrated, then π has a (reduced) rank of one and can be written as $\alpha\beta'$, where α and β are 2×1 vectors of full rank. Under convergence, β' equals $(+1, -1)$, which is a testable hypothesis. Typically, the linear trend t is restricted to enter the cointegrating vector, such that $\phi_1 = \alpha\psi$ for some unrestricted scalar ψ .

The explicit representation of (2) clarifies the relationship between it and the ADF regression (1). For ease of exposition, consider $k = 1$. Under convergence, (2) is then:

$$\Delta y_t^a = \alpha_1(y^a - y^b)_{t-1} + \Gamma_{11}\Delta y_{t-1}^a + \Gamma_{12}\Delta y_{t-1}^b + \phi_{01} + \phi_{11}t + \varepsilon_{1t}, \quad (3)$$

$$\Delta y_t^b = \alpha_2(y^a - y^b)_{t-1} + \Gamma_{21}\Delta y_{t-1}^a + \Gamma_{22}\Delta y_{t-1}^b + \phi_{02} + \phi_{12}t + \varepsilon_{2t}, \quad (4)$$

where $\alpha' = (\alpha_1, \alpha_2)$, $\Gamma_1 = \{\Gamma_{ij}\}$ (for simplicity of notation), $\phi'_0 = (\phi_{01}, \phi_{02})$, $\phi'_1 = (\phi_{11}, \phi_{12})$, and $\varepsilon'_t = (\varepsilon_{1t}, \varepsilon_{2t})$. Subtraction of (4) from (3) and rearrangement of terms yields:

$$\begin{aligned} \Delta(y_t^a - y_t^b) &= (\alpha_1 - \alpha_2)(y^a - y^b)_{t-1} + (\Gamma_{11} - \Gamma_{21})\Delta(y^a - y^b)_{t-1} \\ &\quad + (\Gamma_{11} + \Gamma_{12} - \Gamma_{21} - \Gamma_{22})\Delta y_{t-1}^b \\ &\quad + (\phi_{01} - \phi_{02}) + (\phi_{11} - \phi_{12})t + (\varepsilon_{1t} - \varepsilon_{2t}). \end{aligned} \quad (5)$$

Equation (5) is thus (1) as an ADF(1), with $\kappa = \alpha_1 - \alpha_2$, $\gamma_1 = \Gamma_{11} - \Gamma_{21}$, $\theta_0 = \phi_{01} - \phi_{02}$, $\theta_1 = \phi_{11} - \phi_{12}$, and:

$$u_t = (\Gamma_{11} + \Gamma_{12} - \Gamma_{21} - \Gamma_{22})\Delta y_{t-1}^b + (\varepsilon_{1t} - \varepsilon_{2t}). \quad (6)$$

Testing for convergence via the Johansen procedure has several notable features that differentiate it from univariate procedures.

First, with the Johansen procedure, the hypothesis of convergence is the null hypothesis. That contrasts with convergence being the alternative hypothesis when using univariate test statistics such as the Dickey–Fuller statistic. This reversal of null and alternative hypotheses occurs because, in the Johansen procedure, the test of convergence (which is a restriction on β) is conditional on the number of cointegrating vectors. The question is whether or not the cointegrating vector β can be simplified such that $\beta' = (+1, -1)$.

This reversal of null and alternative hypotheses is particularly appealing when testing deterministic convergence or zero-mean convergence. For the Dickey–Fuller

statistic, both forms imply restrictions on θ_0 and/or θ_1 in (1), yet deterministic convergence (the weakest of the three forms considered) requires *rejecting* the hypothesis that $\kappa = 0$. Thus, in the univariate framework, deterministic convergence and zero-mean convergence each mix a null hypothesis about θ_0 and/or θ_1 with an alternative hypothesis about κ . For the Johansen-based approach, deterministic convergence corresponds to the null hypothesis that $\beta' = (+1, -1)$, with deterministic convergence and zero-mean convergence implying additional restrictions on ϕ_0 and/or ϕ_1 .

Second, because the Johansen-based test of convergence is conditional on the number of cointegrating vectors, the asymptotic distribution of the test statistic under the null hypothesis of convergence is χ^2 , rather than being a function of Wiener processes.

Third, the Johansen-based test of convergence can have increased power relative to univariate tests. The derivation of (5) from (3) and (4) shows the reasons why. By allocating $(\Gamma_{11} + \Gamma_{12} - \Gamma_{21} - \Gamma_{22})\Delta y_{t-1}^b$ to the ADF error u_t , the ADF regression ignores how $(y_t^a - y_t^b)$ may respond differently to dynamics in y^a and in y^b . In the ADF regression, the exclusion of Δy_{t-1}^b , separate from Δy_{t-1}^a , implies a worse fit and hence (possibly) lower power. Also, the Johansen-based test of convergence incorporates information about the cointegrating vector from the equations for both Δy_t^a and Δy_t^b —equivalently, for both $\Delta(y^a - y^b)_t$ and Δy_t^b —whereas the ADF regression aggregates that information into a single equation, again (possibly) losing valuable information. So, the ADF statistic can lose out on power for two reasons. It doesn't allow Δy_{t-i}^a and Δy_{t-i}^b to enter differentially, and it ignores possible cross-equation restrictions that might be present.

Put somewhat differently, the ADF regression error u_t (in (6)) is predictable from past values of y^a and y^b unless $\Gamma_{11} - \Gamma_{21} = \Gamma_{12} - \Gamma_{22}$. That equality requires that short-run responses between y^a and y^b match their long-run responses, corresponding to a common factor (or comfac) restriction. Empirically and economically, such a restriction seems implausible. If the comfac restriction does not hold, the ADF test is still consistent, but it lacks power relative to the Johansen procedure. In effect, inclusion of stationary variables in the VAR may improve the fit of the VAR and, by doing so, obtain more power from the tests employed. This situation can arise even in the classical regression framework. Adding a variable to a regression equation can improve the fit of the model, even if the added variable is independent of all of the other included variables. With the variable added, the equation standard error falls, t -ratios on coefficients increase, and hence the powers of tests increase. See Hendry and Mizon (1978), Sargan (1980), and Kremers, Ericsson, and Dolado (1992) *inter alia* for previous discussions on common factor restrictions.

Fourth, and relatedly, the Johansen-based test has a potential increased ability to detect *lack* of convergence. Suppose that, empirically, $y^a - y^b$ is comprised of a large stationary component and a “small” I(1) component. For the same reasons discussed for power, that I(1) component might not be detected by an ADF test but might be

detected by the Johansen-based test.

Fifth, the Johansen-based test of convergence accounts for general dynamics in the variables. Doing so may ameliorate or avoid a problem noted by Bernard and Durlauf (1996, p. 171) that unit-root type tests can have poor power properties if the data are far from their steady-states.

Sixth, the Johansen framework does have a potential drawback: it requires the full VAR to be reasonably well specified. Being a system procedure is both the strength and the weakness of the Johansen procedure. As a system procedure, it accounts for the interactions and information across the VAR's equations and across the variables in those different equations. Equally, results from the Johansen procedure can be sensitive to the specification of all of the equations in the system. The ADF statistic relies only on obtaining a good specification for that single variable's univariate representation. However, because the ADF regression is derivable from the VAR, as in (5), mis-specification of the VAR may well imply mis-specification of the ADF regression.

Testing for convergence via the Johansen procedure entails first testing for cointegration and then, provided that y^a and y^b cointegrate, testing whether the cointegrating vector is $(+1, -1)$. Other hypotheses about the cointegrating vector are also testable, and stationarity of y^a (or y^b) is one such hypothesis. Because tests of convergence typically assume that y^a and y^b are nonstationary, some observations on Johansen-based stationarity tests are germane.

First, for bivariate VARs, the restriction that $\beta' = (1, 0)$ corresponds to stationarity of the first variable. In general, a stationary variable implies a trivially simple cointegrating vector, namely, that the coefficient on that variable is unity (or nonzero) and that all of the other coefficients in that cointegrating vector are zero. If the trend is allowed to enter the cointegrating vector freely, then the null hypothesis is trend stationarity.

Second, tests of stationarity are the mirror image of tests of significance for an individual coefficient. The former imposes zero restrictions on all coefficients in β but one, whereas the latter impose a single zero restriction on β , leaving the other coefficients in β unconstrained. Equally, the test of stationarity is equivalent to the test of the joint significance of all but one of the variables in the cointegrating vector.

Third, if π has full rank, then all variables in the VAR are stationary (trend stationary if trend enters the cointegrating vectors freely).

Fourth, the Johansen procedure for testing cointegration thus allows for stationary variables among the variables being analyzed. The validity of the Johansen procedure is not affected by whether the individual variables are $I(1)$ or $I(0)$. Stationary variables do have direct implications for both the number of cointegrating vectors and the characteristics of those cointegrating vectors. Each stationary variable that is present implies an additional cointegrating vector, where that cointegrating vector contains all zeros except for one nonzero element, which corresponds to the stationary variable.

Notably, the Engle–Granger procedure does not allow for the presence of sta-

tionary variables, nor do estimation procedures based on a triangular system, as in Phillips and Loretan (1991). For both the Engle–Granger procedure and estimation of triangular systems, the actual presence of stationary variables in the model being estimated radically alters the finite sample and asymptotic distributions of the associated estimators and test statistics.

2.3 Alternative Numeraires

The discussion above considers the convergence between the output series from two arbitrary countries. In practice, convergence is typically analyzed relative to U.S. output or relative to the aggregated output of the set of countries being examined. *Inter alia*, see Linden (2000) and Cheung and Pascual (2000) on the former and Evans (1998) and Li and Papell (1999) on the latter. The choice of numeraire has both economic and statistical implications for the tests.

With (for example) U.S. output as numeraire, the tests of convergence are testing whether each country’s output is converging to U.S. output, not whether there is output convergence among the countries examined. A similar problem exists if the aggregated output is used as numeraire. To highlight the problem, suppose that French output y^{FR} and German output y^{DE} are converging, but that neither is converging to U.S. output y^{US} . The output pair (y^{FR}, y^{DE}) exhibits convergence, but the pairs (y^{FR}, y^{US}) and (y^{DE}, y^{US}) each do not. Whether or not convergence is detected can depend fundamentally on the choice of numeraire.

One alternative is to vary the numeraire. Thus, for France, Germany, and the United States, the analysis of convergence would examine each of the following output pairs: (y^{FR}, y^{DE}) , (y^{DE}, y^{US}) , and (y^{FR}, y^{US}) . A second alternative, considered in Section 2.4, is to analyze larger numbers of countries simultaneously.

2.4 Multivariate Tests for Convergence

The choice of a numeraire is arbitrary, yet it has consequences for the analysis of convergence. Selection of the numeraire may be avoided by analyzing all relevant countries’ output simultaneously. For instance, in the example above, output for France, Germany, and the United States should have one cointegrating vector, $\beta' = (+1, -1, 0)$, corresponding to convergence between French and German outputs alone.

Groups of countries might well exhibit convergence within a group, but with no convergence across groups. That notion underlies “convergence clubs”, as in Quah (1997). If the output of p countries exhibit convergence, then those p output series have $p - 1$ cointegrating relations. Those cointegrating relations can be written such that each cointegrating vector is filled with zeros, except for two values, $+1$ and -1 , corresponding to the two countries whose output series are converging.

A multivariate Johansen-based test of convergence has several potential advantages over a bivariate Johansen-based test, including potential gains in power, increased ability to detect lack of convergence, and freedom from choice of numeraire. However, the multivariate approach may lose power if the analysis includes output series that are not converging.

Until now, Bernard and Durlauf (1995) appear to be unique in adopting a multivariate approach. They report Johansen-based tests of cointegration for groups of six, eleven, and fifteen countries. For each group, they also test the convergence restrictions on the cointegrating vectors. Bernard and Durlauf (1995, footnote 9) cite (but do not report) some evidence on pair-wise convergence as well.

The remainder of the current paper briefly discusses the datasets and sample periods used and implements the univariate, bivariate, and multivariate approaches for testing output convergence.

3 The Data

Section 4 analyzes convergence among the G-7 countries. Two datasets are used, one from Maddison (1995), and the other from Summers and Heston (1991) in the form of the Penn World Table (denoted PWT). This section briefly describes the data and considers some of their properties. Throughout, abbreviations for the G-7 countries are CA (Canada), FR (France), DE (Germany), IT (Italy), JA (Japan), UK (United Kingdom), and US (United States).

Both datasets are annual. Maddison's dataset spans somewhat over a century, whereas Summers and Heston's starts after World War II. The length of Maddison's dataset is both a potential advantage and potential disadvantage. Tests involving unit roots typically have more power on longer samples. However, as several graphs below show, output in the G-7 countries responded differently to the two world wars; and some data prior to 1940 look distinctly different from data after World War II. These observations led to consideration of three pairings of dataset and sample period:

- Maddison's data (1889–1994),
- Maddison's data (1954–1992), and
- Summers and Heston's data (1954–1992).

The short sample of Maddison's data permits direct comparison with results using data from the Penn World Table. The long sample of Maddison's data obtains the benefit of that sample's long span. See Cheung and Pascual (2000) for additional discussion of similar sample and dataset choices.

Figures 1, 2, and 3 plot the logs of real per capita GDP for the long Maddison sample, the short Maddison sample, and the PWT respectively. Figure 4 plots the

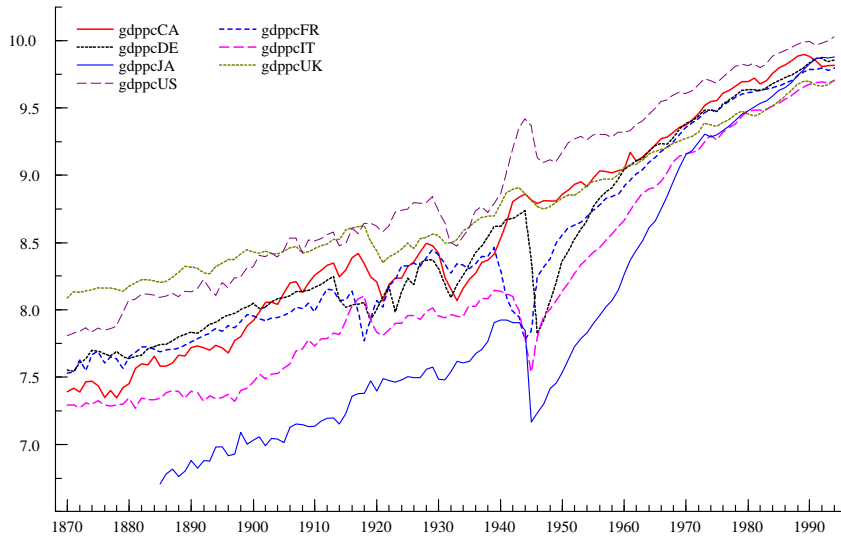


Figure 1: Logs of real per capita GDP from Maddison (1995).

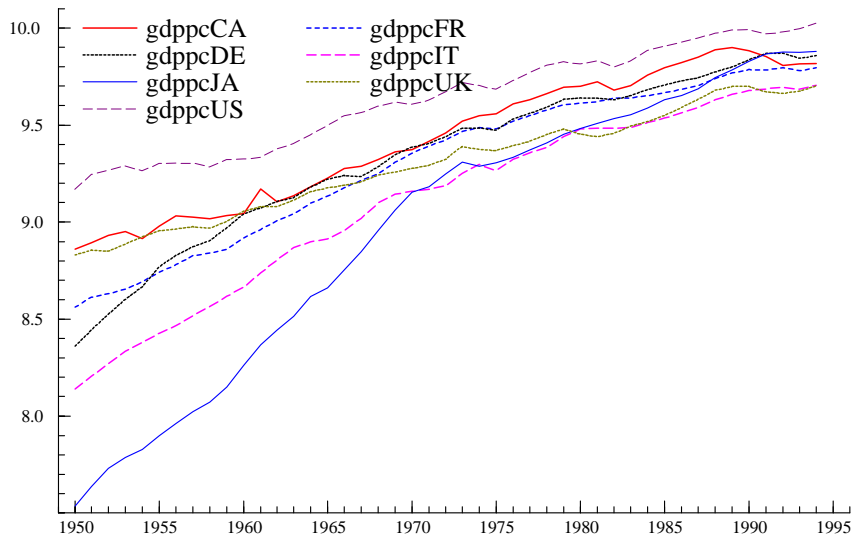


Figure 2: Logs of real per capita GDP from Maddison (1995) (post-war).

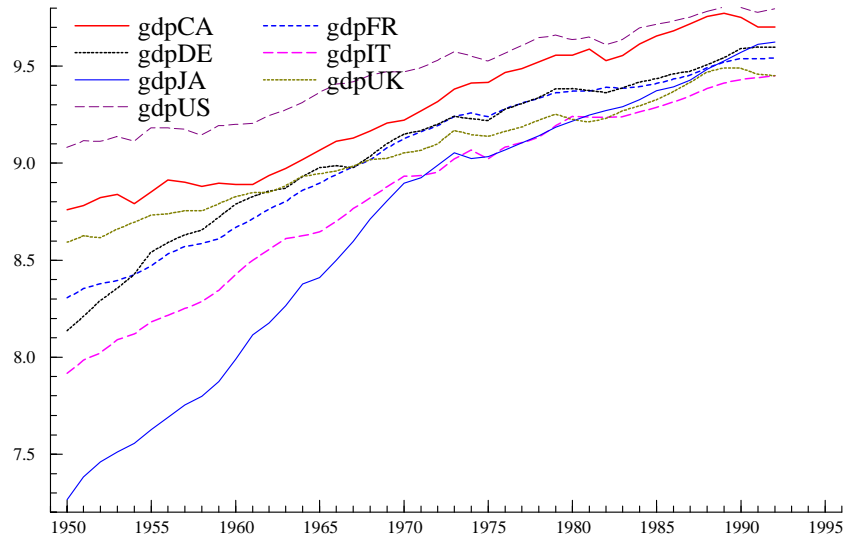


Figure 3: Logs of real per capita GDP from the Penn World Table.

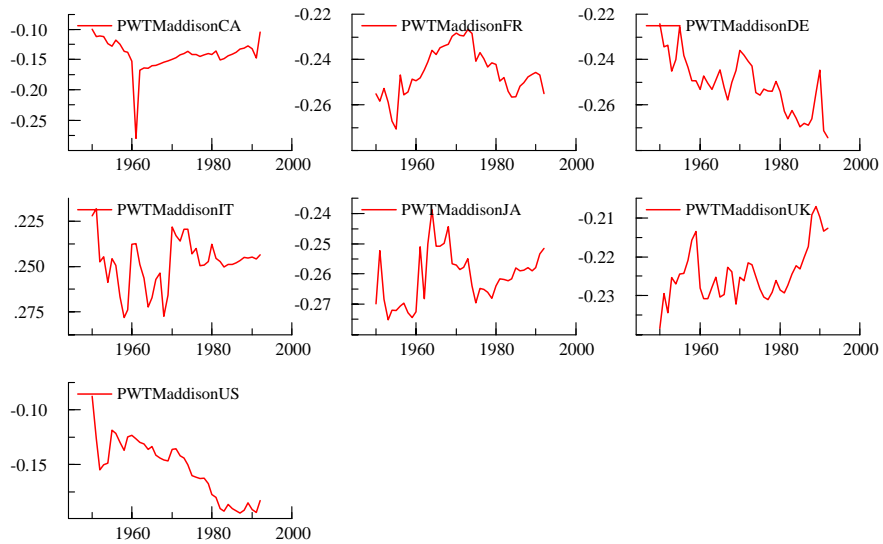


Figure 4: Discrepancies between the logs of real per capita GDP from the Penn World Table and Maddison (1995) (post-war).

differences between the Maddison and PWT measures of real per capita GDP (in logs).

Two characteristics dominate output in Figure 1. First, output for each country trends upward overall, with the dispersion across different countries' output being much less numerically in the 1970s–1980s than in the 1890s–1930s. Second, World War II and its aftermath affected each countries' output differently. Output for Italy, France, Germany, and Japan all plummet precipitously around the end of the war, whereas output for Canada, the United Kingdom, and the United States do not. That said, output for the first four countries are back “on trend” or above trend by about 1970. Figures 2 and 3—for the shorter sample—highlight that post-war experience. The post-war output series from Maddison and the PWT are remarkably similar; see Figures 2 and 3. Figure 4 plots the differences between the Maddison and PWT measures, country by country. Aside from a scale factor, the two measures only differ by a few percent overall, and rarely by more than a few tenths of a percent on a year-by-year basis.¹

The initial univariate and bivariate analyses in Sections 4.1 and 4.2 follow much of the literature by using U.S. output as the numeraire. Figures 5–10 capture this graphically for each dataset and sample period. Figure 5 plots each country's output with U.S. output, country by country, for the full Maddison sample. Figure 6 plots the corresponding differentials, i.e., $y^a - y^{US}$ for country a . Univariate tests of convergence are based on that differential. The two figures highlight the different effects that World War II had on output for the G-7 countries, along with the marked change in output behavior after the war for both Italy and Japan.

Figure 7 plots output pairwise for the Maddison subsample, and Figure 8 plots the corresponding differentials. Figure 9 likewise plots output pairwise for the PWT data, and Figure 10 plots the differentials. Visually, the Maddison and PWT data appear very similar for the post-war period. The percentage gap between foreign and U.S. output does narrow over time, the United Kingdom excepted. The next section focuses on whether this narrowing of the gap between the output series constitutes convergence in any of the senses above.

4 Empirical Results

Section 2 discussed different empirical methodologies for testing convergence. The current section applies those empirical methodologies to each of the three data samples and compares the results. Section 4.1 uses ADF tests, while Section 4.2 uses bivariate Johansen-based tests. Section 4.3 examines how the choice of numeraire influences the results, focusing on France, Germany, and Italy. Section 4.4 analyzes

¹One exception is Canada in 1961, where Maddison's published real per capita GDP contain a typographic error, which is corrected in Section 4's numerical analyses.

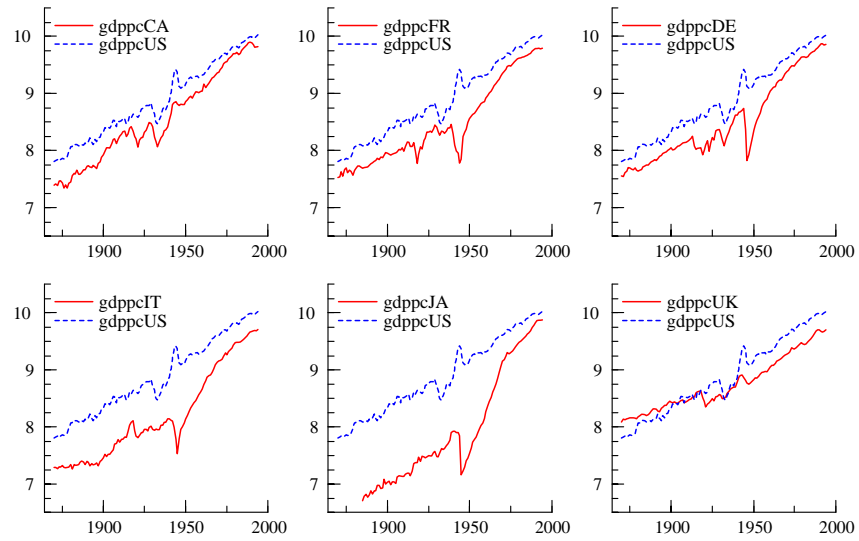


Figure 5: Logs of real per capita GDP from Maddison (1995), compared individually versus real per capita U.S. GDP.

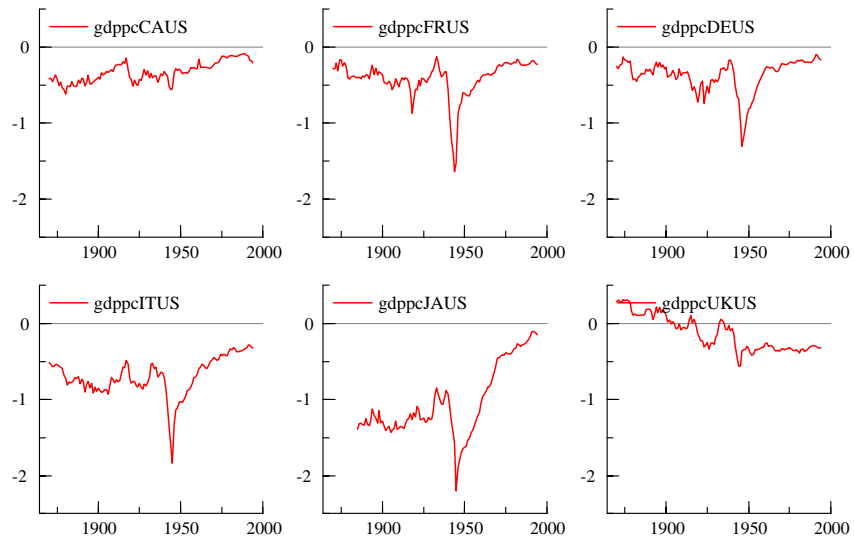


Figure 6: Logs of real per capita GDP from Maddison (1995), all relative to real per capita U.S. GDP.

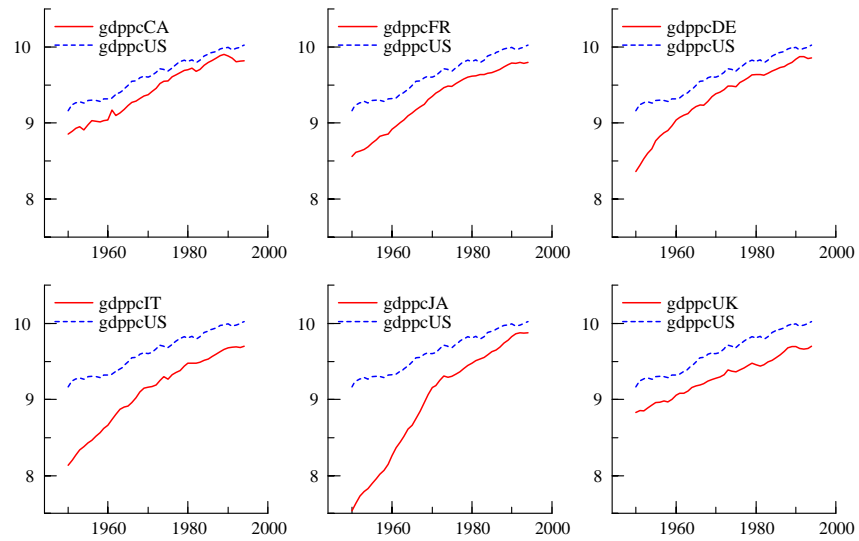


Figure 7: Logs of real per capita GDP from Maddison (1995), compared individually versus real per capita U.S. GDP (post-war).

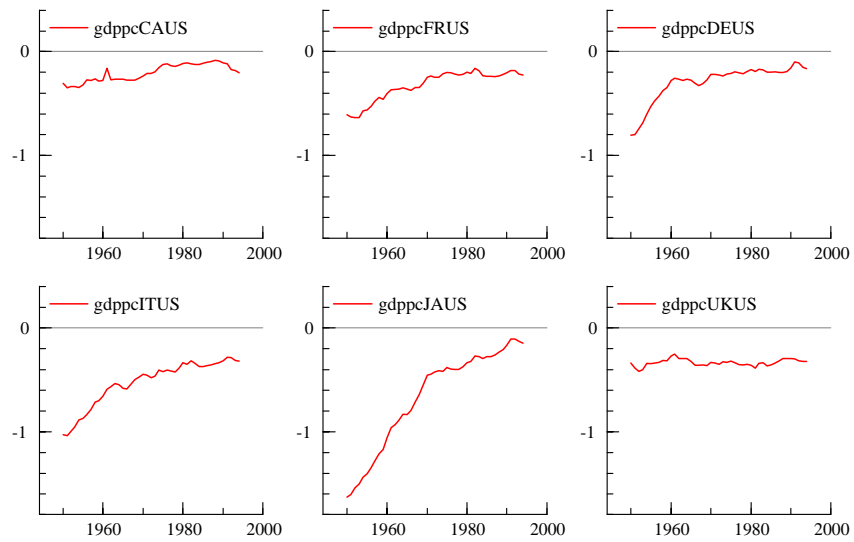


Figure 8: Logs of real per capita GDP from Maddison (1995), all relative to real per capita U.S. GDP (post-war).

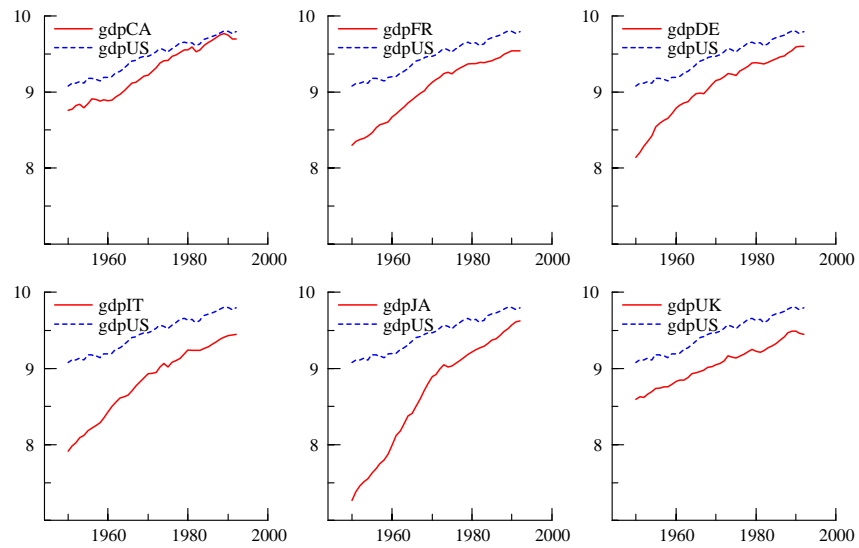


Figure 9: Logs of real per capita GDP from the Penn World Table, compared individually versus real per capita U.S. GDP (post-war).

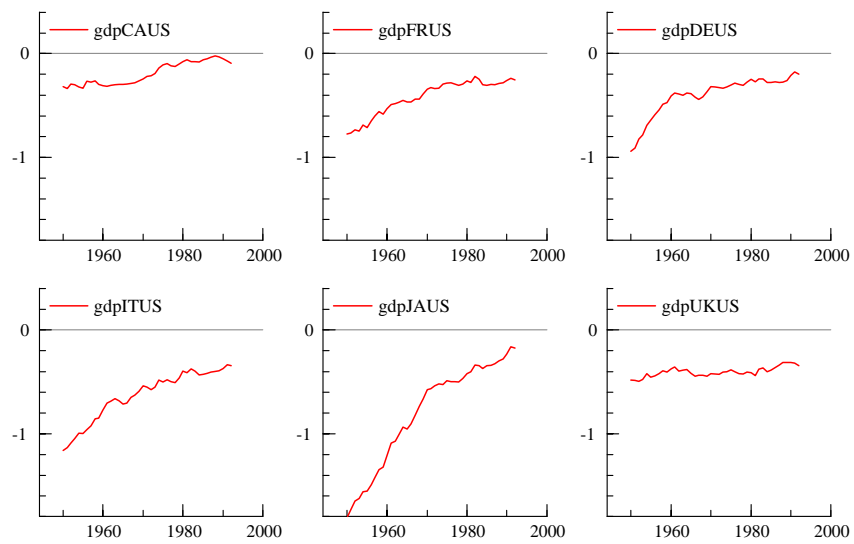


Figure 10: Logs of real per capita GDP from the Penn World Table, all relative to real per capita U.S. GDP (post-war).

convergence of those three countries' output in a multivariate (trivariate) cointegrated VAR. Throughout, rejection of the null hypothesis is denoted by $^+$, $*$, and $**$ for the 10%, 5%, and 1% levels.

4.1 Univariate Tests for Convergence

Table 1 lists ADF statistics and related calculations for the three data samples. For country a in a given data sample, the variable analyzed is $y^a - y^{US}$, which is the difference between the log of country a 's real per capita GDP and the log of U.S. real per capita GDP. The lag length of the ADF regression was based on minimizing the AIC, starting with a maximum of four lags.

German and U.S. output appear to converge for both post-war datasets: the ADF statistic rejects the null hypothesis of no convergence at the 1% level. However, the trend is significant in each of the corresponding regressions, having t -ratios of 2.35 and 2.64 for the Maddison and PWT datasets respectively. Thus, there is stochastic convergence in output, but not deterministic convergence, and hence not zero-mean convergence. The null of no convergence can be rejected at the 10% level for France (full Maddison sample) and the United Kingdom (post-war Maddison sample). Overall, the ADF statistics detect little evidence of convergence among the G-7 countries.

4.2 Bivariate Tests for Convergence

Tables 2, 3, and 4 list the bivariate cointegration analyses and related calculations for the full Maddison sample, the post-war Maddison sample, and the PWT data respectively. All VARs include an intercept and linear trend, restricting the latter to enter the cointegration space.² For country b in a given data sample, the variables analyzed are y^{US} and y^b , with the first half of the table reporting the eigenvalues, max and trace statistics, and estimates of β and α conditional on there being one cointegrating vector.

The second half of each table reports a battery of test statistics, all also conditional on there being one cointegrating vector. The hypotheses tested are homogeneity (i.e., $\beta' = (+1, -1)$), both with a trend and without a trend in the cointegrating vector; the significance of each coefficient β_i in the cointegrating vector β ; and the significance of each coefficient α_i in α . Homogeneity with a trend corresponds to stochastic convergence, and homogeneity without a trend corresponds to deterministic convergence. The hypothesis $\beta_{US} = 0$ (where β_{US} is the cointegrating coefficient

²The lag length in each VAR was initially four and was shortened, based on the AIC, the Schwarz criterion, the Hannan-Quinn criterion, and the F -statistic for the omitted lags. If the criteria differed in their outcomes, some judgement was exercised in selecting the final lag length, based on the variation of the individual criteria over the lag length. For each country, the final lag length was two using the full Maddison sample and one for the post-war samples.

Table 1: ADF statistics of the differential of the log of GDP ($y^a - y^{US}$), by country, by dataset, and by sample period.

Country a	lag k	$t_{ADF(k)}$	$\hat{\rho}$	$\hat{\sigma}$ (%)	t -prob (%)	F -prob (%)	AIC
Full Maddison dataset (1889–1994, $T = 106$)							
Canada	0	−3.07	0.830	4.827	–	60.07	−6.034
France	1	−3.31 ⁺	0.875	9.543	0.00	50.51	−4.662
Germany	3	−2.76	0.897	7.851	1.68	–	−5.034
Italy	1	−2.79	0.900	8.502	0.08	81.55	−4.893
Japan	0	−1.52	0.955	9.783	–	51.08	−4.621
UK	1	−2.95	0.858	5.379	6.66	18.44	−5.808
Subsample of the Maddison dataset (1954–1992, $T = 39$)							
Canada	1	−0.97	0.887	1.704	4.13	96.23	−8.047
France	0	−2.68	0.840	2.113	–	77.61	−7.640
Germany	1	−4.40 ^{**}	0.797	1.897	1.44	68.75	−7.833
Italy	0	−2.77	0.837	2.433	–	50.02	−7.358
Japan	1	−1.53	0.946	2.828	7.51	71.31	−7.034
UK	3	−3.33 ⁺	0.602	1.941	4.51	–	−7.743
Penn World Table dataset (1954–1992, $T = 39$)							
Canada	1	−1.97	0.830	1.881	2.01	51.29	−7.850
France	0	−2.00	0.862	2.503	–	80.92	−7.301
Germany	1	−4.37 ^{**}	0.761	2.085	9.39	47.11	−7.644
Italy	0	−2.09	0.841	2.916	–	79.73	−6.996
Japan	0	−1.22	0.947	3.370	–	62.43	−6.707
UK	3	−2.58	0.657	2.153	3.72	–	−7.536

Notes:

1. Third-order ADF regressions were initially estimated, and the final lag length was selected to minimize the Akaike Information Criterion (AIC). For each variable examined for a given dataset and sample period, the columns report the name of the country whose GDP is being compared to the United States, the selected lag length k , the ADF statistic on the simplified regression ($t_{ADF(k)}$), the estimated coefficient on the lagged level that is being tested for a unit value ($\hat{\rho}$), the regression's residual standard error ($\hat{\sigma}$), the tail probability of the t -statistic on the longest lag of the final regression (t -prob, in %), the tail probability of the F -statistic for the lags dropped (F -prob, in %), and the AIC.

2. All of the ADF regressions include both an intercept and a linear trend. MacKinnon's (1991) approximate finite-sample critical values for the corresponding ADF statistics are −3.15 (10%), −3.45 (5%), and −4.05 (1%) for $T = 106$, and −3.19 (10%), −3.53 (5%), and −4.21 (1%) for $T = 39$.

Table 2: Bivariate cointegration analyses of the log of real per capita US GDP (y^{US}) and the log of real per capita foreign GDP (y^b), by country, on the full-period Maddison dataset (1889–1994).

Country	Eigen-values	Statistic		Estimate of β'			Estimate of α'	
		max	trace	y^{US}	y^b	t	Δy^{US}	Δy^b
Canada	0.183 0.058	20.67* 6.06	26.72* 6.06	1	-0.419 (0.100)	-0.0097 (0.0022)	-0.324 (0.071)	-0.106 (0.072)
France	0.160 0.035	17.73+ 3.65	21.38 3.65	1	-0.222 (0.119)	-0.0139 0.0026	-0.165 (0.045)	0.163 (0.056)
Germany	0.148 0.060	16.37 6.32	22.70 6.32	1	0.053 (0.115)	-0.0201 (0.0025)	-0.142 (0.048)	-0.265 (0.071)
Italy	0.117 0.042	12.66 4.38	17.04 4.38	1	-0.168 (0.131)	-0.0147 (0.0032)	-0.177 (0.049)	0.008 (0.053)
Japan	0.141 0.032	15.47 3.34	18.81 3.34	1	-0.044 (0.077)	-0.0179 (0.0024)	-0.146 (0.050)	-0.236 (0.079)
UK	0.153 0.047	16.90+ 4.91	21.81 4.91	1	-0.246 (0.189)	-0.0157 (0.0027)	-0.183 (0.054)	-0.104 (0.032)
Test statistics for:								
	homogeneity		significance of β_i			significance of α_i		
	with t	without t	y^{US}	y^b	t	Δy^{US}	Δy^b	
Canada	12.92**	17.19**	12.46**	3.93*	10.39**	14.27**	1.63	
France	6.77**	8.19*	12.68**	2.91+	5.77*	10.55**	6.84**	
Germany	10.44**	11.37**	8.22**	0.12	10.43**	5.49*	8.49**	
Italy	5.77*	9.52**	8.34**	1.18	4.52*	8.56**	0.01	
Japan	12.54**	15.30**	12.26**	0.27	11.29**	6.71**	7.15**	
UK	8.57**	12.97**	11.23**	0.93	10.89**	8.05**	7.73**	

Table 3: Bivariate cointegration analyses of the log of real per capita US GDP (y^{US}) and the log of real per capita foreign GDP (y^b), by country, on the subsample Maddison dataset (1954–1992).

Country	Eigen- values	Statistic		Estimate of β'			Estimate of α'	
		max	trace	y^{US}	y^b	t	Δy^{US}	Δy^b
Canada	0.321 0.035	14.30 1.30	15.61 1.30	1	0.096 (0.248)	-0.0253 (0.0070)	-0.141 (0.090)	0.084 (0.108)
France	0.493 0.186	25.11** 7.63	32.74** 7.63	1	-0.393 (0.155)	-0.0137 (0.0048)	-0.078 (0.064)	0.155 (0.037)
Germany	0.588 0.128	32.84** 5.07	37.91** 5.07	1	-1.477 (0.277)	0.0152 (0.0084)	-0.038 (0.029)	0.116 (0.023)
Italy	0.350 0.168	15.91 6.79	22.70 6.79	1	-0.537 (0.204)	-0.0064 (0.0074)	-0.088 (0.060)	0.160 (0.053)
Japan	0.346 0.235	15.72 9.92	25.64* 9.92	1	0.130 (0.515)	-0.0003 (0.0291)	-0.004 (0.012)	-0.056 (0.013)
UK	0.248 0.077	10.54 2.96	13.50 2.96	1	-1.580 (0.436)	0.0111 (0.0092)	-0.102 (0.100)	0.170 (0.086)
Test statistics for:								
		homogeneity		significance of β_i			significance of α_i	
		with t	without t	y^{US}	y^b	t	Δy^{US}	Δy^b
Canada		11.57**	11.72**	4.41*	0.06	11.20**	2.39	0.60
France		3.71 ⁺	4.42	5.61*	3.66 ⁺	4.40*	1.22	11.19**
Germany		0.70	3.17	3.01 ⁺	20.79**	1.40	1.64	18.48**
Italy		1.01	1.33	2.90 ⁺	4.14*	0.44	1.43	5.27*
Japan		0.13	0.62	0.07	0.02	0.00	0.06	5.90
UK		1.31	2.06	4.86*	7.30**	1.12	0.83	3.00 ⁺

Table 4: Bivariate cointegration analyses of the log of real per capita US GDP (y^{US}) and the log of real per capita foreign GDP (y^b), by country, on the Penn World Table dataset (1954–1992).

Country	Eigen-values	Statistic		Estimate of β'			Estimate of α'	
		max	trace	y^{US}	y^b	t	Δy^{US}	Δy^b
Canada	0.340 0.117	15.38 4.61	19.99 4.61	1	-0.068 (0.169)	-0.0172 (0.0048)	-0.218 (0.120)	0.087 (0.143)
France	0.491 0.162	24.95** 6.54	31.49** 6.54	1	-0.359 (0.105)	-0.0113 (0.0033)	-0.120 (0.093)	0.230 (0.057)
Germany	0.545 0.131	29.15** 5.19	34.34** 5.19	1	-1.031 (0.190)	0.0061 (0.0057)	-0.040 (0.051)	0.190 (0.043)
Italy	0.290 0.134	12.67 5.32	17.99 5.32	1	-0.417 (0.153)	-0.0066 (0.0056)	-0.151 (0.099)	0.226 (0.090)
Japan	0.305 0.241	13.47 10.19	23.66+ 10.19	1	-0.109 (0.167)	-0.0048 (0.0095)	-0.056 (0.045)	-0.189 (0.048)
UK	0.192 0.099	7.87 3.86	11.73 3.86	1	-1.805 (0.536)	0.0184 (0.0114)	-0.050 (0.104)	0.166 (0.082)
Test statistics for:								
		homogeneity		significance of β_i			significance of α_i	
		with t	without t	y^{US}	y^b	t	Δy^{US}	Δy^b
Canada		10.98**	14.58**	7.39**	0.06	8.11**	2.50	0.30
France		6.71**	7.04*	8.45**	5.69*	6.75**	1.41	11.13**
Germany		0.01	4.69+	5.14*	18.54**	0.66	0.59	14.54**
Italy		2.59	2.94	4.15*	4.29*	1.03	1.52	3.78+
Japan		0.41	0.76	0.34	0.08	0.08	0.45	3.42+
UK		1.23	2.91	2.26	4.13*	1.53	0.14	2.20

on y^{US}) corresponds to a significance test of y^{US} in the cointegrating vector. The hypothesis $\beta_{US} = 0$ also corresponds to the (null) hypothesis that y^b (not y^{US}) is *trend* stationary. Likewise, the hypothesis that $\beta_b = 0$ generates a significance test of y^b in the cointegrating vector and corresponds to the (null) hypothesis that y^{US} is trend stationary.³

As Table 2 shows, little evidence exists for convergence (or even cointegration) between a given country's output and U.S. output over the full Maddison sample. From Table 3 for the post-war Maddison data, output from each of France and Germany exhibit deterministic convergence to U.S. output. For each of these bivariate analyses, a single cointegrating vector is found, and it can be restricted as $\beta' = (+1, -1)$, with the trend coefficient being zero. From Table 4 for the PWT data, French output and German output are each cointegrated with U.S. output, but stochastic convergence is rejected for French output.

Comparison with the ADF results in Table 1 confirms the gains in power from the bivariate Johansen-based tests of convergence. The ADF statistics for French relative to U.S. output on the post-war datasets are -2.68 and -2.00 , which are approximately the first-quartile and median values for the ADF statistic under the null hypothesis of no convergence. By contrast, the corresponding max and trace statistics reject the null hypothesis of no cointegration at the 1% level. The evidence on German output in Tables 3 and 4 confirms the corresponding results in Table 1.

4.3 Alternative Numeraires

As Section 2.3 discussed, with U.S. output as numeraire, the tests of convergence are testing whether each country's output is converging to U.S. output, not whether there is output convergence among the countries examined. Testing output convergence among non-U.S. countries requires analyzing those countries' output together. In principle, output convergence among any of the seven G-7 countries could be examined. However, graphs and geography suggest focusing on three countries: France, Germany, and Italy.

Figure 11 plots output for France, Germany, and Italy from the three data samples. Without series from the other four countries, the similarities in these three countries' output paths are striking. French and German output move closely together throughout, with Italian output tending to systematically track below French and German output. Thus, for each data sample, this subsection tests for cointegra-

³The test statistics in the lower half of each table, along with the estimates of α and β , are conditional on accepting the presence a single cointegrating vector in the bivariate system. In several bivariate systems, the cointegration evidence from the max and trace statistics does not support cointegration, in which case these auxiliary estimates and test statistics should be ignored. They are included in part for completeness and in part because the choice of critical values for the max and trace statistics (as with any statistic) is an open issue.

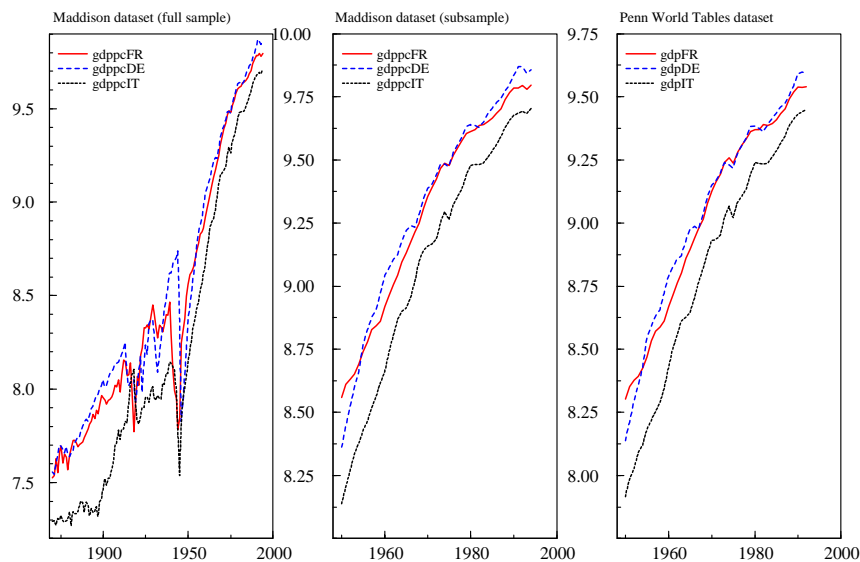


Figure 11: Logs of real per capita GDP for France, Germany, and Italy from the Maddison and Penn World Table datasets.

tion and convergence between each possible pair of these three output series.

Table 5 lists the cointegration and convergence results. Over the full Maddison sample, French and German output satisfy deterministic convergence, whereas French and Italian output only satisfy stochastic convergence. Rejection of deterministic convergence for French and Italian output reflect the relatively precise estimates of the cointegrating coefficients. When homogeneity is imposed, the trend coefficient is only 0.00343 but has a standard error of 0.00078, implying a t -ratio of 4.4 on a coefficient representing about one-third of one percent growth per annum. Cointegration between German and Italian output is less evident, but if cointegration holds, then these output series satisfy stochastic and (possibly) deterministic convergence.⁴

For the post-war samples, French and German output exhibit two cointegrating vectors ($r = 2$), implying that the series are stationary. From the first column of statistics in the lower half of the table, the linear trend is statistically significant for that sample, implying that French and German output are trend stationary. That result reflects the very smooth upward movement of both series in the second two graphs of Figure 11. Analysis of German and Italian output also indicates stationarity, with the evidence for stationarity being less strong on the French–Italian pairing. While the post-war results for the three pairings are not entirely consistent with each

⁴As in Section 4.2, selection of each VAR’s lag length is based on the AIC, the Schwarz criterion, the Hannan-Quinn criterion, and the F -statistic. For the full Maddison sample, the VAR for French and German output is fourth order, whereas the other two VARs are second order. All VARs on post-war data are first order.

Table 5: Bivariate cointegration analyses of the log of real per capita GDP for France (FR), Germany (DE), and Italy (IT) for all pairs, by country pair and sample period.

Countries Sample	Eigen- values	Statistic		Estimate of β'			Estimate of α'	
		max	trace	y^a	y^b	t	Δy^a	Δy^b
FR, DE								
1889–1994 (Maddison)	0.253 0.038	28.54** 3.76	32.30** 3.76	1	–1.032 (0.104)	0.0009 (0.0022)	–0.062 (0.053)	0.305 (0.053)
1954–1992 (Maddison)	0.466 0.396	23.20* 18.68**	41.88** 18.68**	1	–2.942 (0.571)	0.0435 (0.0174)	0.028 (0.012)	0.076 (0.014)
1954–1992 (PWT)	0.464 0.393	23.09* 18.45**	41.54** 18.45**	1	–1.743 (0.170)	0.0202 (0.0051)	–0.048 (0.048)	0.179 (0.063)
DE, IT								
1889–1994 (Maddison)	0.135 0.039	14.75 4.02	18.77 4.02	1	–0.896 (0.157)	0.0006 (0.0038)	–0.193 (0.049)	0.013 (0.042)
1954–1992 (Maddison)	0.478 0.285	24.06** 12.42*	36.48** 12.42*	1	–0.350 (0.193)	–0.0117 (0.0070)	–0.207 (0.036)	–0.105 (0.045)
1954–1992 (PWT)	0.444 0.308	21.70* 13.65*	35.34** 13.65*	1	–0.585 (0.106)	–0.0064 (0.0039)	–0.385 (0.087)	–0.009 (0.101)
FR, IT								
1889–1994 (Maddison)	0.177 0.037	19.82* 3.82	23.64+ 3.82	1	–1.007 (0.110)	0.0036 (0.0027)	0.040 (0.059)	0.206 (0.045)
1954–1992 (Maddison)	0.464 0.143	23.04* 5.72	28.76* 5.72	1	–0.989 (0.174)	0.0104 (0.0063)	–0.161 (0.029)	–0.144 (0.049)
1954–1992 (PWT)	0.407 0.158	19.31* 6.35	25.66* 6.35	1	–1.055 (0.139)	0.0113 (0.0051)	–0.233 (0.049)	–0.159 (0.079)
Test statistics for:								
	<i>trend</i> ($r = 2$)	homogeneity		significance of β_i			significance of α_i	
		with t	without t	y^a	y^b	t	Δy^a	Δy^b
FR, DE								
1889–1994 (M)	4.18	0.08	0.18	24.57**	25.30**	0.14	1.32	26.79**
1954–1992 (M)	12.48**	2.73+	15.49**	0.46	1.63	1.56	1.30	4.73*
1954–1992 (P)	13.90**	4.80*	12.45**	4.22*	4.80*	4.80*	0.27	1.88
DE, IT								
1889–1994 (M)	3.99	0.32	5.10+	10.96**	8.98**	0.02	11.11**	0.07
1954–1992 (M)	3.61	6.42*	7.83*	2.20	0.41	2.71+	12.12**	3.01+
1954–1992 (P)	2.87	8.35**	8.36**	5.38*	2.98+	2.82+	6.56*	0.00
FR, IT								
1889–1994 (M)	5.39+	0.00	11.13**	15.32**	16.07**	1.52	0.39	15.88**
1954–1992 (M)	4.29	0.00	11.37**	3.23+	3.33+	2.56	18.16**	6.54*
1954–1992 (P)	6.58*	0.16	8.70*	3.48+	3.49+	3.60+	12.86**	2.99+

other, a resolution may be feasible by analyzing all three series jointly.

The results for all three data samples illustrate how important the choice of numeraire can be. These long-run relations are empirically undetected and conceptually undetectable from analyzing pairs of series with U.S. output as numeraire.

4.4 Multivariate Tests for Convergence

To illustrate multivariate Johansen-based tests for convergence, this subsection examines output for France, Germany, and Italy over the full Maddison sample. The estimated VAR is second-order.

Table 6 reports the cointegration statistics and various corresponding convergence test statistics. The format of Table 6 differs somewhat from previous tables, due to the number of variables involved. The max and trace statistics clearly reject the null hypothesis of no cointegration in favor of at least one cointegrating vector.

If a single cointegrating vector is assumed ($r = 1$), the coefficients on German and Italian output are virtually the same (-0.507 and -0.538) and sum to approximately minus the coefficient on French output; see the lower half of Table 6. That cointegrating vector can be restricted to satisfy deterministic convergence between French and German output: $\chi^2(3) = 5.43$ [0.143], where the p -value appears in brackets. However, the hypothesis of either stochastic or deterministic convergence between French and Italian output is rejected: $\chi^2(2) = 6.58$ [0.0373] and $\chi^2(3) = 21.25$ [0.0001] respectively.

This result is puzzling in light of Table 5, where stochastic convergence was detected between French and Italian output for the full sample. A resolution exists by considering the trivariate analysis in Table 6 under the assumption of two cointegrating vectors. While the max and trace statistics do not reject the hypothesis of at most one cointegrating vector ($r \leq 1$) at standard levels, the second largest eigenvalue is relatively large, being equal to 0.133. The bottom two blocks of Table 6 thus report estimates assuming $r = 2$. The first set of estimates are just identified. Convergence between French and German output and between German and Italian output appears numerically sensible: corresponding estimates in β are $(+1, -1.106, 0)$ and $(0, +1, -0.899)$. Stochastic convergence is acceptable statistically: $\chi^2(2) = 0.48$ [0.7848]. However, deterministic convergence is rejected at the 5% level, albeit not at the 1% level: $\chi^2(4) = 13.02$ [0.0112]. The culprit appears to be a numerically small but statistically significant trend in German (or French) output relative to Italian output. The final (restricted) block in Table 6 reports the estimated β with deterministic convergence imposed for French and German output but only stochastic convergence imposed for German and Italian output: $\chi^2(3) = 0.72$ [0.870]. Those restrictions are statistically acceptable, and the estimated trend in French or German output relative to Italian output is 0.00357, virtually the same as that obtained from the bivariate analysis of French and Italian output in Section 4.3. Thus, over the full Maddison

Table 6: A trivariate cointegration analyses of the logs of real per capita GDP for France, Germany, and Italy on the full-period Maddison dataset (1889–1994).

rank of π	$r = 0$	$r \leq 1$	$r \leq 2$	$r \leq 3$	
Eigenvalue	–	0.258	0.133	0.035	
Statistic	Null hypothesis				
	$r = 0$	$r \leq 1$	$r \leq 2$		
max statistic	29.89*	14.24	3.51		
trace statistic	47.64*	17.75	3.51		
	Eigenvalues β'				
Variable	y^{FR}	y^{FR}	y^{IT}	<i>trend</i>	
	1	–0.508	–0.538	0.0031	
	0.443	1	–1.339	0.0022	
	2.307	0.352	1	–0.0898	
	Adjustment coefficients α				
y^{FR}	0.084	–0.003	–0.014		
y^{DE}	0.159	–0.135	0.005		
y^{IT}	0.226	0.065	–0.002		
	Assumption: $r = 1$				
Estimate of β' (unrestricted)	1	–0.507 (0.132)	–0.538 (0.134)	0.0031 (0.0020)	
Estimate of β' (restricted)	1	–1	0	0	$\chi^2(3) = 5.43$ [0.143]
	Assumption: $r = 2$				
Estimate of β' (unrestricted)	1	–1.106 (0.103)	0	0.0027 (0.0023)	
	0	1	–0.899 (0.121)	0.0007 (0.0030)	
Estimate of β' (restricted)	1	–1	0	0	$\chi^2(3) = 0.72$ [0.870]
	0	1	–1	0.00357 (0.00074)	

sample, French, German, and Italian output exhibit convergence, albeit of different forms.

5 Conclusions

Whether output converges or not has substantive economic consequences for the countries involved. Convergence also has possible implications for policy and for the class of model relevant to empirical modeling. Tests of convergence in a time series framework have historically been formulated as univariate tests of stationarity of the output differential. However, convergence is a multivariate hypothesis, and a univariate test of a multivariate hypothesis entails a loss of information. Conversely, a multivariate framework is natural for testing hypotheses of convergence. The cointegrated VAR is such a framework, with the Johansen procedure providing the statistical basis for tests of convergence. This approach yields novel results relative to those obtained from the univariate approach. For instance, the multivariate approach detects output convergence among France, Germany, and Italy over the last century, whereas the univariate approach fails to detect any such convergence.

Several general remarks are germane, and each suggests possible extensions to the current analysis. First, the order of integration for output series need not remain the same over time, perhaps explaining the apparent trend stationarity of French and German output post-war. For long stretches of data, the general constancy of the output process is an open issue and should be examined further. Second, multivariate VARs including all seven countries' output could be estimated and tested for convergence. Third, the VAR could include variables other than output, such as country-specific variables that would permit interpreting convergence as conditional convergence. Such variables may introduce an additional complication—cointegration among variables of a given country; see Ericsson, Irons, and Tryon (2001). Fourth, the multivariate approach discussed is applicable to economic issues well beyond that of convergence. Conceptually similar restrictions arise when testing (e.g.) purchasing power parity, or price convergence across regions; see Culver and Papell (1999) and Rogers (2001) for examples.

Data Appendix

This appendix describes the two datasets utilized in this paper. Each dataset contains annual series for each of the G-7 countries (Canada, France, Germany, Italy, Japan, the United Kingdom, and the United States).

Series from Maddison (1995, Appendix D) are real per capita GDP in 1990 Geary-Khamis dollars. The published 1961 value for Canadian per capita output is incorrect,

due to a typographic error in Canadian population for that year, which should be 18.269 million, not 16.269 million. The data analyzed in Section 4 are corrected for that typo.

Series from Summers and Heston (1991) are chain-weighted real GDP per capita in constant dollars (expressed in international prices, base 1985). This series is RGDPCH from Summers and Heston's Mark 5.6a Penn World Table, available at www.pwt.econ.upenn.edu on the WorldWide Web.

References

- Bernard, A. B., and S. N. Durlauf (1995) "Convergence in International Output", *Journal of Applied Econometrics*, 10, 97–108.
- Bernard, A. B., and S. N. Durlauf (1996) "Interpreting Tests of the Convergence Hypothesis", *Journal of Econometrics*, 71, 161–173.
- Carlino, G. A., and L. O. Mills (1993) "Are U.S. Regional Incomes Converging?", *Journal of Monetary Economics*, 32, 335–346.
- Cheung, Y.-W., and A. G. Pascual (2000) "Testing for Output Convergence: A Re-examination", CESifo Working Paper No. 319, Center for Economic Studies, Ludwig Maximilians Universität, Munich, Germany, July.
- Culver, S. E., and D. H. Papell (1999) "Long-run Purchasing Power Parity with Short-run Data: Evidence with a Null Hypothesis of Stationarity", *Journal of International Money and Finance*, 18, 5, 751–768.
- Dickey, D. A., and W. A. Fuller (1979) "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", *Journal of the American Statistical Association*, 74, 366, 427–431.
- Dickey, D. A., and W. A. Fuller (1981) "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root", *Econometrica*, 49, 4, 1057–1072.
- Doornik, J. A., and D. F. Hendry (1996) *PcGive Professional 9.0 for Windows*, International Thomson Business Press, London.
- Doornik, J. A., and D. F. Hendry (2001) *PcGive Version 10 for Windows*, Timberlake Consultants Press, London.
- Ericsson, N. R., J. S. Irons, and R. W. Tryon (2001) "Output and Inflation in the Long Run", *Journal of Applied Econometrics*, 16, 3, 241–253.
- Evans, P. (1998) "Using Panel Data To Evaluate Growth Theories", *International Economic Review*, 39, 2, 295–306.
- Evans, P., and G. Karras (1996) "Do Economies Converge? Evidence from a Panel of U.S. States", *Review of Economics and Statistics*, 78, 3, 384–388.

- Hendry, D. F., and G. E. Mizon (1978) “Serial Correlation as a Convenient Simplification, Not a Nuisance: A Comment on a Study of the Demand for Money by the Bank of England”, *Economic Journal*, 88, 351, 549–563.
- Johansen, S. (1991) “Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models”, *Econometrica*, 59, 6, 1551–1580.
- Johansen, S. (1995) *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.
- Kremers, J. J. M., N. R. Ericsson, and J. J. Dolado (1992) “The Power of Cointegration Tests”, *Oxford Bulletin of Economics and Statistics*, 54, 3, 325–348.
- Lee, K., M. H. Pesaran, and R. Smith (1997) “Growth and Convergence in a Multi-country Empirical Stochastic Solow Model”, *Journal of Applied Econometrics*, 12, 4, 357–392.
- Li, Q., and D. H. Papell (1999) “Convergence of International Output Time Series Evidence for 16 OECD Countries”, *International Review of Economics and Finance*, 8, 3, 267–280.
- Linden, M. (2000) “Testing Growth Convergence with Time Series Data—A Non-parametric Approach”, *International Review of Applied Economics*, 14, 3, 361–370.
- Maddison, A. (1995) *Monitoring the World Economy: 1820–1992*, Organisation for Economic Co-operation and Development, Paris.
- Phillips, P. C. B., and M. Loretan (1991) “Estimating Long-run Economic Equilibria”, *Review of Economic Studies*, 58, 3, 407–436.
- Quah, D. T. (1997) “Empirics for Growth and Distribution: Stratification, Polarization, and Convergence Clubs”, *Journal of Economic Growth*, 2, 27–59.
- Rassekh, F. (1998) “The Convergence Hypothesis: History, Theory, and Evidence”, *Open Economies Review*, 9, 85–105.
- Rogers, J. H. (2001) “Price Level Convergence, Relative Prices, and Inflation in Europe”, International Finance Discussion Paper No. 699, Board of Governors of the Federal Reserve System, Washington, D.C., (March, available on the WorldWide Web at www.federalreserve.gov/pubs/ifdp/2001/699/default.htm).
- Sargan, J. D. (1980) “Some Tests of Dynamic Specification for a Single Equation”, *Econometrica*, 48, 4, 879–897.
- Summers, R., and A. Heston (1991) “The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950–1988”, *Quarterly Journal of Economics*, 106, 2, 327–368.
- Temple, J. (1999) “The New Growth Evidence”, *Journal of Economic Literature*, 37, 1, 112–156.