Convergence in International Output: Evidence from Panel Data Unit Root Tests

Mark J. Holmes
Loughborough University

Abstract

This paper investigates international output convergence using methods of panel data unit root testing advocated by Im et al. (1997) and Breuer et al. (1999). Using quarterly data for a sample of OECD economies for the period 1960-98 on GDP differentials, the evidence suggests that power deficiency may be an issue where univariate ADF unit root tests find against convergence with respect to the US or Germany. However, while the Im et al. t-bar test offers strong evidence in favor of convergence, the Breuer et al. SURADF test suggests that this finding may in fact be driven by the rejection of non-stationarity in a small number of cases.

- JEL Codes: C2, C3, F0
- Key Words: Unit Root Testing, Panel Data, Convergence

I. Introduction

A variety of studies (see, inter alia, Serletis and Krichel (1992), Bernard and Durlauf (1995), Greasley and Oxley (1997) and Mills and Holmes (1999)) have measured the extent of international output convergence. Using GDP or index of industrial production data for OECD economies over varying study periods, evidence of long-run convergence is mixed. This paper examines output convergence from a new angle through the application of panel data unit root testing.1

The focus here is on long-run bivariate convergence between a sample of countries

---

and the US and then Germany. This paper offers two important contributions to the literature on output convergence. First, panel data unit root testing offers a means of overcoming problems of low test power associated with the earlier applications of univariate ADF tests. We employ the t-bar panel data unit root tests advocated by Im et al. (1997) on panels of output deviations from a base country. The second contribution is that this study addresses two shortcomings associated with many existing panel data unit root tests. Moreover, rejection of the null of joint non-stationarity in a panel may be due to a single series within the panel being stationary. Also, while the Im et al. procedure has key advantages over early panel data tests, the demeaning procedure associated with the t-bar test does not exploit information in error covariances in an entirely satisfactory manner. For these reasons, this study also employs a technique advocated by Breuer et al. (1999) which involves the estimation of ADF regressions within a seemingly unrelated regression (SUR) framework. The SURADF results are used to qualify the results obtained through the Im et al. procedure.

The paper is structured as follows. The following section discusses some of the fundamental economic issues involved with the convergence of output. The third section discusses the data, methodology and results. While there is little evidence of convergence using univariate ADF tests, the t-bar panel data unit root tests identify strong convergence using either the US or Germany as the base country. Application of the SURADF technique, however, suggests that convergence is present in a small number of cases where convergence with the US is stronger than with Germany. The final section concludes.

II. The International Convergence of Output

Following the original neo-classical growth model proposed by Solow (1956), countries should converge to a balanced growth path where poorer countries grow faster than richer ones. The Solow model implies that the return on capital is lower in countries with more capital per worker and this provides an incentive for capital to flow from richer to poorer countries. Also, lags in the diffusion of knowledge mean that income differences arise because some countries not yet employing the best available technologies. Moreover, income differences will shrink as poorer countries acquire state of the art technologies. Studies by Mankiw et al. (1992), Sala-i-Martin (1996) and others offer evidence in favor of conditional convergence based on movements towards
individual steady states. Moreover, endogenous growth theory (see, inter alia, Romer (1986)) argues that the driving force behind growth is the accumulation of knowledge via Research and Development and that one should take a broader view of capital through incorporating human capital. Divergence in long-term growth can be generated by social increasing returns to scale associated with capital and labor. This literature may suggest that richer countries grow the fastest. However, evidence is mixed and Olson (1996) argues that it is the quality of institutions and economic policies that can push economies towards their production possibility frontiers. For this reason, small subsets of poor and rich countries have typically grown the fastest.

In the context of this study, short-run output linkages are influenced by asymmetric shocks, such as German unification in July 1990, or symmetric shocks, such as the oil price rises in the 1970s. However, structural and institutional factors are crucial in forming the background against which long-run linkages with Germany or the US can exist. This discussion considers a number of these factors. These include the exchange rate regime and capital mobility, international spillovers of investment, trade and economic integration.

While Bretton Woods and the Snake during the 1960s and 1970s sought to stabilize nominal exchange rates with respect to the US dollar, the ERM also sought to remove capital controls among its members during the 1980s and 1990s. While the convergence literature has focussed on nominal convergence against a background of exchange rate stability, real convergence is relatively less explored. Economic theory, however, suggests an ambiguous link between exchange rate regime and growth (see, for example, Levy-Yeyati and Sturzenegger (2001)). On the one hand, Friedman (1953) and Poole (1970) argue that flexible regimes are better suited towards insulating the economy against real shocks. The lack of exchange rate adjustments under a peg combined with short-run price rigidities, results in price distortions and the misallocation of resources. Furthermore, the standard Mundell-Fleming model demonstrates that fixed exchange rates combined with perfect capital mobility and asset substitutability removes the scope for autonomous monetary policy as a counter-cyclical mechanism. As a result of this, increases in output volatility may lead to a fall in growth (see, for example, Ramey and Ramey (1995)). Another line of argument is presented by Calvo (1999) and

\[ A \text{ symmetric shock may be reclassified as an asymmetric shock if member countries react to the same shock differently. For example, in the late 1970s the UK was a net exporter of oil and therefore affected differently by the oil price rises of that period compared to other EU members.} \]
others where the need to defend the peg in the event of a large negative shock means that real interest rates are raised thereby potentially harming investment prospects. On the other hand, it can be argued that against a background of risk-averse agents, a peg is likely to stimulate investment and trade and therefore growth (see, for example, Frankel (1999), Rose (2000), Frankel and Rose (2000)). Baldwin (1989) takes the argument further by considering the relationship between exchange rate stability and a subsequent reduction in risk-adjusted discount rates in investment decisions. Using a framework based on the Solow growth model, it can argued that countries participating in a credible fixed exchange rate agreement will move members towards some common discount rate and therefore towards some notion of conditional convergence.

Recent evidence suggests that international spillovers of investment may provide a strong reason for convergence of growth rates over and above the effects of capital mobility, although differences in levels of output between countries may still remain (see, inter alia, Alogoskoufis and van der Ploeg (1991a, 1991b), Grossman and Helpman (1991)). In these models, spillovers of technology cause the marginal productivity of a broad measure of capital in a lower income country to exceed that of a higher income country so the incentive to invest in the former country is higher than in the latter country. If non-tradable and reproducible are used in the production of a tradable commodity, the growth rates of output may differ permanently even if the international mobility of physical capital is perfect. However, there is scope for convergence if there are decreasing returns to capital at the national level but constant returns at the global level and the importance of non-traded factors of production is not large.

Numerous theories predict a positive effect of international trade on the level of income and economic growth. According to this literature, international trade improves performance by promoting specialization in production, technology transfer, learning by doing and competition among firms. However, in may of these models trade can lead to income divergence across countries. Other studies have shown that trade openness promotes income convergence. For example, Sachs and Warner (1995, 1997) find that (conditional) convergence is faster for open economies while Ben-David (1993) studies EU countries and finds that intra-group convergence was more rapid following trade liberalization. Against this background, an important event during this period might have been the creation of

---

3In a more recent paper, Lane (2001) shows how international trade can accelerate convergence through the expansion of access to credit.
the single European market during 1992 removing restrictions on goods and labor mobility.

To summarize, the traditional growth literature argues that poorer countries should grow faster than richer ones as steady state equilibria are achieved. The endogenous growth literature implies that richer countries, who can save more, may grow faster as more resources are channeled into research and development and human capital. A compromise view is offered by Olson (1996) based on efficiency of production. In the case of the EU, there are several factors that one might expect to enhance convergence with Germany. These include exchange rate stability, the relaxation of capital controls and the creation of the single market as key factors in the move towards increased economic integration. The following empirical analysis can shed light on whether integration with Germany is stronger or weaker than with the US.

III. Data, Methodology and Results

Quarterly GDP data are employed for eleven countries- Austria, Belgium, Canada, Denmark, France, Germany, Italy, Japan, Netherlands, UK and US- for the period 1960Q1-98Q4. The data are obtained from the OECD Main Economic Indicators.\(^4\) Output differentials are defined with respect to the US and then Germany to compare the extent of convergence allowing for the possibility that convergence with Germany may be the more relevant for the EU economies. Whether or not long-run output convergence prevails between the natural logarithms of domestic (non-US or non-German) and base (US or German) real output (respectively denoted as \(y_{it}\) and \(y_{jt}\)) depends on the time series properties of \(x_{it}\), which is computed as

\[
x_{it} = y_{it} - y_{jt}
\]

where \(i = 1, 2, \ldots, N\) non-base countries, \(j = US,\) Germany and \(i = 1, 2, \ldots, T\) time periods. Table 1 reports the results from univariate ADF tests. At the 5% significance level, non-stationarity is rejected in only four cases: Germany-US, UK-US, Japan-Germany and UK-Germany. These initial tests therefore suggest that long-run bivariate convergence only occurs in a small number of cases.

\(^4\)Real GDP data for Belgium are only available on an annual basis. Quarterly values have been interpolated using the index of industrial production.
However, the univariate ADF tests can suffer from low power and for this reason may be unable to reject the null of non-stationarity. Panel data unit root testing, on the other hand, utilizes more observations where the cross-country variations of the data in estimation are exploited.

Following Im et al. (1997), suppose \( x_{it} \) is generated by a first order autoregressive process

\[
\Delta x_{it} = \alpha_i + \phi_i x_{i,t-1} + \epsilon_{it}
\]  

(2)

where \( \epsilon_{it} \) is a disturbance term. The null hypothesis is \( H_0: \phi_i \leq 0, \forall i \) and the alternative is \( H_0: \phi_i < 0, i = 1, 2, ..., N_i, \phi_i = 0, i = N_i + 1, N_i + 2, ..., N \). The disturbances

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
\( y_{it} - y_{USt} \) & No Trend & Trend \\
\hline
Austria & -2.251 & -2.295 \\
Belgium & -1.674 & -2.149 \\
Canada & -1.800 & -0.329 \\
Denmark & -0.232 & -2.022# \\
France & -0.846 & -0.957 \\
Germany & -2.908** & -3.258* \\
Italy & -1.695 & -0.629 \\
Japan & -2.483 & -0.169 \\
Netherlands & -1.459 & -1.826 \\
UK & -1.744 & -3.510**# \\
\hline
\( y_{it} - y_{GERt} \) & & \\
Austria & -2.048 & -2.089 \\
Belgium & -0.918 & -2.018 \\
Canada & -2.208 & -2.298 \\
Denmark & -0.956 & -2.038 \\
France & -1.927 & -1.623 \\
Italy & -2.126 & -1.158 \\
Japan & -3.027** & -0.737 \\
Netherlands & -2.578* & -2.493 \\
UK & -2.211 & -3.470**# \\
\hline
\end{tabular}
\end{table}

For each augmented Dickey Fuller (ADF) regression, the lag length was chosen using Said and Dickey’s (1984) \( T^{1/3} \) rule. In all cases, the residuals were free of serial correlation. The conclusions were qualitatively unaffected by the employment of alternative procedures for lag length selection. ** and * indicate rejection of the null of non-stationarity at the 5 and 10% levels of significance respectively. # denotes the significance of the time trend in the ADF regression at the 5% level. For regressions excluding a trend, relevant critical values taken from Fuller (1976) are \(-2.89\) and \(-2.58\), while for regressions including a trend, these are \(-3.45\) and \(-3.15\) respectively.
across the panel may be correlated. Indeed, wrongly assuming identically and independently distributed disturbances can have dramatic implications for statistical size and power to the extent that the null may not be correctly accepted or rejected. To address this issue, let \( \varepsilon_{it} = \theta_t + u_{it} \) where \( \theta_t \) is a time-specific common effect that allows for a degree of dependency across the series and \( u_{it} \) is an idiosyncratic random effect that is independently distributed across groups. To remove the effect of \( \theta_t \) subtract the cross-section means from both sides of (2) to obtain the following demeaned regression

\[
\Delta \tilde{x}_{it} = \tilde{\alpha}_t + \tilde{\phi}_t \tilde{x}_{i,t-1} + \tilde{\xi}_{it}
\]

(3)

For a heterogeneous panel with serially correlated errors (3) may be rewritten as

\[
\Delta \tilde{x}_{it} = \tilde{\alpha}_t + \tilde{\phi}_t \tilde{x}_{i,t-1} + \sum_{j=1}^{q_i} \rho_{ik} \Delta \tilde{x}_{i,t-k} + \tilde{\xi}_{it}
\]

(4)

Equation (4) forms the basis of the t-bar test for output convergence. Using data for \( x_{it} = y_{it} - y_{jt} \) it is estimated for a sample involving the major EU economies where the t-bar statistic is calculated using the average value of the individual ADF statistics based on each \( \tilde{\phi}_t \).5

The t-bar test results for both the US and Germany differentials are reported in Table 2. In the both cases, the null of joint non-stationarity is strongly rejected at

### Table 2. IPS Panel Data Unit Root Tests

<table>
<thead>
<tr>
<th>y_{it} - y_{US}</th>
<th>t-bar</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 Countries</td>
<td>−2.266***</td>
<td>−2.145</td>
<td>−1.967</td>
<td>−1.869</td>
</tr>
<tr>
<td>8 Countries</td>
<td>−2.418***</td>
<td>−2.224</td>
<td>−2.03</td>
<td>−1.913</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>y_{it} - y_{GER}</th>
<th>t-bar</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 Countries</td>
<td>−2.161***</td>
<td>−2.145</td>
<td>−1.967</td>
<td>−1.869</td>
</tr>
<tr>
<td>7 Countries</td>
<td>−2.305***</td>
<td>−2.315</td>
<td>−2.078</td>
<td>−1.951</td>
</tr>
</tbody>
</table>

These are t-bar tests based on equation (4). The lag length, \( q_i \), is chosen using Said and Dickey’s (1984) \( T^{1/3} \) rule. In all cases, the residuals were free of serial correlation. The conclusions were qualitatively unaffected by the employment of alternative procedures for lag length selection. *** denotes rejection of the null of non-stationarity at the 1% level of significance. Critical values are simulated with 10,000 replications.

---

5The t-bar test requires that the t-bar statistic follows an asymptotic normal distribution as both \( N \rightarrow \infty \) and \( T \rightarrow \infty \) with \( (N/T) \rightarrow k \) where \( k \) is a finite positive constant.

5Using the average estimates of \( \tilde{\phi}_t \), the half life of a random shock in the case of the US (German) differentials is computed as 11.696 (12.995) quarters.
the 1% significance level for the full panel of 10 countries.\textsuperscript{6} It is possible that such a strong rejection is being driven by the inclusion of the univariate-stationary series. These differentials are therefore excluded from their respective panels to form the groups of ‘8 countries’ and ‘7 countries’. In each case, we have panels that comprise univariate non-stationary series only. Again, the null of joint non-stationarity is strongly rejected at the 1% significance level indicating real convergence.

Bearing in mind the discussion in the previous section, these panel results in favor of convergence would appear to contradict EU experience. Upto the early 1990s, the ERM had mixed success in achieving its aims. While Artis and Taylor (1988) find evidence of stabilized nominal exchange rates during its early years, the permitted fluctuations in nominal exchange rates were set at ±2.25% around a central parity and there were several realignments within the ERM.\textsuperscript{7} Speculative crises in the early 1990s resulted in the exit of Italy and the UK in September 1992 and the subsequent widening of the permitted bands of exchange rate fluctuations to ±15% for the remaining members in August 1993. Also, there has been a diversity of experience with regard to the use of capital controls. As documented by Ungerer \textit{et al.} (1990), these controls have been gradually relaxed over the period of the ERM with the removal of all controls for most countries by May 1990. However, some EU experience is consistent with these results. This is particularly the case during 1996-8 as prospective members of the EU single currency endeavored to satisfy the Maastricht convergence criteria concerning interest rates, inflation, exchange rate stability, debt-income and budget deficit-income ratios. The Maastricht convergence criteria was adopted against the background of the creation of the single market in 1992.

It might, however, be excessive to conclude that convergence holds for all members of each panel. The t-bar test does not inform us how many or which members of the panel contain are stationary. Also, a demeaning procedure has been employed to deal with contemporaneous correlation but this does not deal with the presence of idiosyncratic shocks to $\epsilon_i$ in a satisfactory manner.\textsuperscript{8} To

\textsuperscript{6}See Artis (1990) for an account of the early realignments. Italy was allowed a larger band fluctuation of until the narrower band was followed during 1990-92. The UK was a formal member of the ERM for only 1990-92 though shadowed the German Mark during 1987-88. During its brief membership, the UK also adhered to the limit.

\textsuperscript{7}While O’Connell (1998) argues that a demeaning approach is preferable to the use of time dummies, the relative merits of controlling for contemporaneous correlation through the estimation of the covariance matrix for are stressed.
address these problems, Breuer et al. (1999) advocate a panel data unit root test which involves estimating ADF regressions in a seemingly unrelated regression (SUR) framework and then testing for individual unit roots within the panel. The SURADF test is more powerful than independently estimated single equation ADF tests. Earlier SUR-based tests of Abuaf and Jorion (1990), O’Connell (1998) and Taylor and Sarno (1998) have an ‘all or nothing’ characteristic to their tests insofar as the null hypothesis is that all series are non-stationary (or stationary) against the alternative that at least one series is stationary (or non-stationary). In this respect, these tests are vulnerable to the criticism that rejection of the null could be attributable the behavior of a single series. The Breuer et al. SURADF procedure, however, allows \( \phi_i \) to differ across the series under the alternative

---

Note: The Taylor and Sarno SUR-based test is accompanied by a Johansen likelihood test of the null that the long-run matrix of the series is less than full rank. This can consume considerable degrees of freedom as the panel expands and can still leave the researcher unable to infer the breakdown between stationary and non-stationary series.

---

### Table 3. SURADF Analysis

<table>
<thead>
<tr>
<th>Series</th>
<th>Critical Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1%</td>
</tr>
<tr>
<td>( y_{it} - y_{US} )</td>
<td>SURADF</td>
</tr>
<tr>
<td>Austria</td>
<td>-3.036*</td>
</tr>
<tr>
<td>Belgium</td>
<td>-2.823*</td>
</tr>
<tr>
<td>Canada</td>
<td>-2.388</td>
</tr>
<tr>
<td>Denmark</td>
<td>-0.878</td>
</tr>
<tr>
<td>France</td>
<td>-2.244</td>
</tr>
<tr>
<td>Italy</td>
<td>-4.039***</td>
</tr>
<tr>
<td>Japan</td>
<td>-3.522*</td>
</tr>
<tr>
<td>Netherlands</td>
<td>-2.934</td>
</tr>
<tr>
<td>( y_{it} - y_{GER} )</td>
<td>SURADF</td>
</tr>
<tr>
<td>Austria</td>
<td>-2.590</td>
</tr>
<tr>
<td>Belgium</td>
<td>-2.416</td>
</tr>
<tr>
<td>Canada</td>
<td>-2.335</td>
</tr>
<tr>
<td>Denmark</td>
<td>-2.269</td>
</tr>
<tr>
<td>France</td>
<td>-2.785</td>
</tr>
<tr>
<td>Italy</td>
<td>-3.625*</td>
</tr>
<tr>
<td>Netherlands</td>
<td>-3.319*</td>
</tr>
</tbody>
</table>

SURADF is the augmented Dickey Fuller statistic obtained through seemingly unrelated regression estimation. The lag length, \( q_i \), is chosen using Said and Dickey’s (1984) \( T^{1/3} \) rule. In all cases, the residuals were free of serial correlation. The conclusions were qualitatively unaffected by the employment of alternative procedures for lag length selection. Critical values specific to each series in the panels are simulated using 10000 replications based on the estimated covariance matrix of the system, \( N \) and \( T \).
hypothesis while exploiting information in error covariances to produce efficient estimators with potentially powerful test statistics. The test statistics from the SUR model feature nonstandard distributions with critical values that must be derived through simulations.

Table 3 reports the SURADF estimates for the ‘8 countries’ and ‘7 countries’ panels that exclude the univariate-stationary series along with 1, 5 and 10% critical values tailored to each ADF statistic that have been specifically generated using Monte Carlo simulations. With regard to US-based output deviations, the null of non-stationarity is rejected at the 5% significance level in the case of Italy. However, at the 10% significance level it is also possible to include Austria, Belgium and Japan as stationary series within the panel. It is surprising that the null of non-stationarity is accepted in the case of Canada. Despite the close trade links between these economies, they are insufficient to generate long-run bivariate convergence. In the case of German-based output differentials, the null of stationarity can only be rejected in the cases of Italy and the Netherlands at the 10% significance level. Within the EU, these economies have had contrasting experiences. Although both countries have been members of the EU for the full sample period, the Netherlands has a much stronger record of exchange rate stability particularly during the ERM. The absence of long-run convergence between France and Germany is consistent with the difficulties experienced by France in maintaining ERM parities during the early 1990s while Denmark has decided not to proceed towards economic and monetary union. On the other hand, the absence of Austria and Belgium from the list of stationary output differentials is surprising. Moreover, Austria has very close trading links with Germany while Belgium has a reasonably stable track record of ERM membership. An explanation for this result might be that Table 1 confirmed bivariate convergence between Germany and the US according to the univariate ADF unit root tests. Table 3 confirms bivariate convergence between the US and Austria and the US and Belgium. We therefore have indirect evidence of convergence between Germany and Austria and Germany and Belgium that operates through convergence between Germany and the US. Moreover, the results reported in Tables 1 and 3 suggest that any perceived convergence within the EU occurs with the US playing a central role through its convergence with Austria, Belgium, Germany, Italy and the UK.
IV. Summary and Conclusion

Using quarterly data for 1960-98, univariate augmented Dickey Fuller unit root tests offer limited evidence that output differentials defined against the US or Germany are stationary. Such a finding may be attributable to low test power as the outcome is dramatically modified to one of strong convergence if the univariate non-stationary data are used to create a panel and the t-bar test is applied. Here the null of joint non-stationarity is strongly rejected. Further analysis is based on univariate augmented Dickey Fuller unit root tests by seemingly unrelated regression estimation. These tests suggest that this rejection of the joint null of non-stationarity is in fact driven by a small number of countries and that convergence with the US is greater than with Germany. Despite the periods of exchange rate stability, relaxed capital controls and mobility of labor in Europe, convergence with Germany is limited. Avenues for future research could examine how these relationships may change following the introduction of the single currency in the EU.

Acknowledgement

I am grateful for the helpful comments provided by an anonymous referee, research assistance from Kate Morrison and the ESRC (Award No. L1382511013) for their financial support. Any remaining errors are my own.

Received 9 April 2001, Accepted 22 October 2001

References

Cambridge, pp. 185-206.
Frankel, J. (1999), “No single currency regime is right for all countries or at all times”, NBER, Working Paper No. 7338.