University of California, Santa Cruz
Department of Economics
Working Paper #282, February 1994 (revised)
JEL classification: C22, E32, O57
Forthcoming, Oxford Economic Papers

DETERMINISTIC, STOCHASTIC, AND SEGMENTED TRENDS IN AGGREGATE OUTPUT: A CROSS-COUNTRY ANALYSIS

by

Yin-Wong Cheung and Menzie David Chinn
Department of Economics
University of California
Santa Cruz, CA 95064

December 21, 1994

ABSTRACT: This paper examines whether output per capita in 126 countries is better described as trend or difference stationary using formal statistical tests. Appropriate finite-sample critical values are constructed to evaluate the test results. Depending upon whether one uses solely a test with a trend stationary null, or solely one with a difference stationary null, one obtains very different conclusions regarding the long run dynamics of GDP. This outcome suggests that it is useful to consider the tests complementary, rather than competing. Based on results from both tests, we find that when a definite characterization of GDP can be made, it is very likely to indicate a difference stationary process. However, the likelihood of making definite conclusions does vary positively with both the income level and the quality of data.

JEL classification: C22, E32, O57

Acknowledgements: Helpful comments were received from Tim Cogley, David Papell, Doug Steigerwald, Shang-Jin Wei, seminar participants at UC Santa Barbara, Harvard University, and our colleagues at UCSC. We thank Tuan Tran for research assistance, and the Pacific Rim Project of the University of California for financial support. Any remaining errors remain solely ours.

Address for correspondence: Menzie Chinn, Department of Economics, Social Sciences I, University of California, Santa Cruz, CA 95064. Tel: (408) 459-2079. Fax: (408) 459-5900. E-mail: chinn@cats.ucsc.edu

1. INTRODUCTION

One of the most active debates in macroeconomics over the past decade has been whether output contains a unit root. The attention devoted to this topic is well deserved. Knowledge regarding the persistence of GDP is important for several reasons. First, whether GDP is trend or difference stationary has dramatic implications for the long run behavior of output dynamics. A shock to a trend stationary process has a transitory impact, whereas a shock to a difference stationary one permanently shifts the trend. Second, when GDP is used in regressions with other variables, the interpretation of the regression results can depend on whether the variables involved are trend or difference stationary. This phenomenon is related to the "spurious regression" literature due to Granger and Newbold (1974). Third, the issue of output persistence is also relevant to the recent empirical studies examining the hypothesis of convergence across countries (e.g., Bernard and Durlauf, 1991). A necessary, though not sufficient, condition is that the output data have common long run persistence. Recently the finding of a unit root GDP process based on the conventional unit root tests, such as augmented Dickey-Fuller (ADF) and the Phillips-Perron tests, has been re-evaluated using different methodologies. Christiano and Eichenbaum (1990) and Rudebusch (1993), for example, argue that the standard unit root tests do not have much power to differentiate between the trend stationary and difference stationary properties of GNP. Thus these tests may generate spurious unit root results.

Another concern is the effect of structural instability on these unit root tests. Perron (1989b, 1990) shows that the conventional unit root tests tend to misinterpret a trend stationary time series with a structural break as a difference stationary series. For instance, Kormendi and Meguire (1990) find evidence of a unit root in both long historical and post Second World War

GDP data. However, Zelhorst and De Haan (1994) and De Haan and Zelhorst (1993) show that the unit root result may be spurious and induced by structural changes in the data generating process. The issue of structural breaks' effects on unit root tests becomes more prominent when one has to analyze GDP series over long time periods, during which changes in regime may occur.

This study attempts to provide a better insight on the persistence of GDP data in the following manner. First we use different types of tests to examine the unit root property. In addition to the standard ADF test which has the unit root process as the null hypothesis and trend stationarity as the alternative hypothesis, we also subject the data to: (1) a unit root test which has trend stationarity as the null and the unit root process as the alternative, developed by Kwiatkowski, Phillips, Schmidt and Shin (1992, hereafter referred to as "KPSS"), and; (2) the unit root test which has trend stationarity and a structural break as the alternative, developed by Banerjee, Lumsdaine and Stock (1992, hereafter referred to as "BLS"). Second, we assess the test results using the appropriate finite sample critical values. Finally, we examine GDP data across 126 countries to investigate if unit root persistence is a general phenomenon, or a characteristic that only applies to a specific group of countries.

Regarding the first point, consider the case of two econometricians, one using the ADF test, and the other, the KPSS test. Each independently might make quite different conclusions regarding the time series characteristics of GDP. In fact, in the data set used in this analysis, they would. Econometrician "A", using the ADF, would conclude that most GDP series were difference stationary, while econometrician "K", using the KPSS test, would conclude that most series were trend stationary.

Instead of viewing the tests as competing, the KPSS test results could be used to corroborate the information obtained from the ADF test, and vice versa. Suppose the ADF test fails to reject the unit root null because of low power. The KPSS test which has trend stationarity as the null should indicate the data have no unit roots. On the other hand, if the KPSS test rejects the trend stationarity null, then we have stronger evidence for unit root persistence. That is, consistent results from ADF and KPSS tests yield more persuasive evidence on data persistence while conflicting results indicate uncertainty associated with the interpretation of the individual test outcomes.

The BLS test is used to determine if the apparent unit root results are attributable to discrete trend breaks in the GDP data. One advantage of the BLS test is that this procedure is not conditional on a pre-assigned break point. By endogenizing the break point date in the procedure, the BLS test avoids the criticism of data-mining and provides an objective evaluation of the unit root and stationary-with-structural-break hypotheses.[1]

Regarding the second point, our study is careful to insure that neither the use of asymptotic KPSS critical values, nor the use of the standard ADF critical values, results in inappropriate inferences. We expect in finite samples that the use of asymptotic critical values will result in over-rejection. Hence, we will report results using finite sample critical values; those results using asymptotic critical values are available from the authors.

Finally, this paper studies the unit root property of GDP data contained in the Summers-Heston data set, which provides the longest consistent output series for a large number of countries. This last attribute is important because we wish to identify patterns of acceptance and rejection that may be missed if we only test "the usual suspects", such as the G-7 countries.[2] The

patterns considered in this study include results one might obtain when examining differences between: (i) developed versus less-developed countries; (ii) different regions, and; (iii) low and high quality data. We are unaware of any other systematic investigation of GDP persistence using these techniques and over such a wide set of countries.[3]

There is one caveat. We do not attempt to fit our empirical findings to any specific economic theory. Rather, the interested reader is encouraged to decide for him- or herself the theoretical reasons underlying the results. In this sense, the current study is an explicitly empirical exercise.

This paper is organized as follows. In Section 2, we discuss the data, and then outline the econometric methodology used. The ADF and KPSS results are presented in Section 3, while in Section 4 we apply the test of BLS to those series which reject the trend stationary hypothesis. Section 5 concludes.

2. DATA AND METHODOLOGY

2.1. Data Description

The data used are the annual Gross Domestic Product (GDP) per capita in real terms using international prices, for 126 countries, obtained from the Summers and Heston (1991) Penn World Tables, Mark 5 (PWT5). In the following discussion, we refer to these log real per capita output as simply "GDP" for convenience.[4]

The use of GDP evaluated at international prices differentiates this study from nearly all previous ones. Typically, cross-country studies rely upon domestically valued GDP, in constant prices. We choose to examine the characteristics of the Summers-Heston measure of output

exactly because it is comparable across countries. Consequently, we abstract from idiosyncratic relative price changes in each country when characterizing the time series properties of output. The relevance of this study is reinforced by the widespread use of the Summers and Heston data set in a number of influential studies of the determinants of long-run growth.

The sample period covers the post-War era for the developed countries, and somewhat shorter time lengths for some of the less-developed countries, for the period up to 1988. The data span at most 39 observations, and at minimum 10 (St. Vincent). These sample sizes may appear small, and hence likely to worsen the power of the ADF test (for example, see Harris (1992)).[5] However as pointed out by Shiller and Perron (1985) and Perron (1989a), the power of the tests for stationarity depend mainly on the length of the data time span and not on the number of observations. That is, the power of the test is essentially the same for both a sample containing 39 annual data points and a sample containing 39 x 12 monthly observations.

2.2. Econometric Methodology

2.2.1. Overview

As mentioned in the introduction, GDP will be subjected to both the ADF and the KPSS tests. The results of these two tests are used to determine the nature of persistence in each GDP series. To investigate the possibility of misinterpreting a structural break as unit root persistence, we apply the BLS test to those series that both the ADF and KPSS tests indicate the presence of a unit root.

The null and alternative hypotheses of the tests can be summarized as follows:

Test	H_0:	H_A:
ADF	I(1)	I(0)
KPSS	I(0)	I(1)
BLS(A)	I(1)	I(0) w/shift in trend
BLS(B)	I(1)	I(0) w/shift in mean

2.2.2. The ADF Test

Let $\{y_t\}$ be the GDP series. The ADF test for unit roots is based on the regression

$$\Delta y_t = c + \mu t + \pi y_{t-1} + \sum_{j=1}^{k} \gamma_j \Delta y_{t-j} + u_t \qquad \textbf{(1)}$$

The unit root null hypothesis is rejected if $\hat{\pi}$ is significantly less than zero according to the finite sample critical values calculated from simulated distributions which control for both sample size and lag structure (Cheung and Lai, forthcoming). We report both sets of results.

In applying this test, one must determine the lag length parameter, k. One possibility is to select an arbitrary, large k, to capture any possible serial correlation in the data. Recently, Alastair Hall (1994) has shown that a general to specific modeling strategy that uses the data to determine k can improve both the size and power of the ADF test. Hence in this study, the Akaike Information Criterion and the Schwartz Bayesian Information Criterion (AIC and SBC respectively) are used to determine the lag parameter. In many cases the two criteria yield similar inferences and so in order to conserve space, we only report the results based on the AIC.

2.2.3. Trend Stationarity Tests

The KPSS test assumes that the time series is the sum of either a mean or a deterministic trend, a random walk, and a stationary error. It is a Lagrange Multiplier test for the null hypothesis that the error variance in the random walk component of the series is zero. In this study, we consider the case where the null is a trend stationary process.

To conduct the test, we first obtain the residual e_t from the regression of y_t on a constant and a trend. The KPSS statistics $\hat{\eta}_t$ is then given by

$$\hat{\eta}_t = T^{-2}\sum S_t^2 / s^2(\ell) \qquad \textbf{(2)}$$

where S_t is the partial sum process defined by

$$S_t = \sum_{i=1}^{t} e_i , \quad t=1,2,...,T. \qquad \textbf{(3)}$$

$$s^2(\ell) = T^{-1}\sum_{t=1}^{T} e_t^2 + 2T^{-1}\sum_{s=1}^{\ell} w(s,\ell) \sum_{t=s+1}^{T} e_t e_{t-s} \qquad \textbf{(4)}$$

and $s^2(\ell)$ is the serial correlation and heteroskedasticity consistent variance estimator given by $w(s,\ell)$ is an optimal weighting function corresponding to the choice of a spectral window.[6]

The null of trend stationarity is rejected in favor of the unit root alternative if the KPSS statistic is larger than the asymptotic critical values provided by KPSS. Based on simulation results, KPSS assert that their test has good size and power characteristics. As in the case of the ADF tests, to avoid spurious conclusions resulting from the use of asymptotic critical values, we also employ finite sample critical values (Cheung, Lai and Tran (1994)). ℓ in equation (4) is a choice parameter to be determined. Following KPSS' suggestion we adopt the $\ell 8$ rule, which sets

ℓ=INT[8(T/100)$^{1/4}$].[7]

2.2.4. Structural Breaks

BLS propose several recursive and sequential tests that allow for structural breaks under the trend stationary alternative. Given the limited number of observations available, we use only the sequential tests.[8]

Consider the following model:

$$\Delta y_t = \mu_0 + \mu_1 \tau_{1t}(p) + \mu_2 t + \alpha y_{t-1} + \beta(L)\Delta y_{t-1} + e_t \qquad \textbf{(5)}$$

where τ_{1t}(p) is a dummy variable. When a shift in the trend at time p is considered under the alternative hypothesis, then τ_{1t}(p) is set equal to (t-p)1(t>p), where 1(.) is an indicator function. Alternatively, when a shift in the mean at time p is considered, τ_{1t}(p) is set equal to 1(t>p). These two situations conform to BLS' Case A and B, respectively.

For case A, three sequential statistics are computed:

$$\begin{aligned} i)\quad & F_{A,\max} \equiv MAX_{p_0 \le p \le T-p_0} F(p) \\ ii)\quad & t_A(\hat{p}), \text{ where } F(\hat{p}) = F_{A,\max} \\ iii)\quad & t_{A,\min} \equiv MIN_{p_0 \le p \le T-p_0} t_A(p) \end{aligned}$$

where p_0 is the number of observations reserved by the trimming parameter δ_0, such that $p_0=\delta_0 T$. We set δ_0 = 0.15, as suggested by BLS.

F(p) and t_A(p) are, respectively, the F-statistic for the hypothesis of μ_1 = 0 and the D-F t-statistic computed from the whole sample with τ_{1t}(p) defined as above. For case B, we compute similar statistics and label them as $F_{B,max}$, $t_B(\hat{p})$, and $t_{B,min}$.

The ß(L) polynomial serves to model the serial correlation in the data, a purpose

analogous to the differenced terms in the ADF regression. To facilitate the comparison between the ADF and BLS tests, the lag orders chosen by the AIC used in the ADF regressions are also used to conduct the BLS tests.

The asymptotic distributions and the finite sample critical values of these sequential tests are given in BLS (1992). The smallest sample size considered by these authors is 100, which is relatively large compared to the sample sizes of our data set. To guard against inappropriate inferences caused by small sample effects, the tests conducted in this portion of the study are based on simulated critical values for the sample sizes 35 and 25.[9]

3. UNIT ROOT AND TREND STATIONARITY TEST RESULTS

3.1. Overview

The specific country results are contained in Appendix 1. The overall results are summarized in the following format:

	KPSS accept	KPSS reject	
ADF accepts	cell 1	cell 2	row total
ADF rejects	cell 3	cell 4	row total
	column total	column total	total

Cell 1 reports the number and the percentage of cases that both the ADF and KPSS tests fail to reject (or more concisely and inaccurately, "accept")[10] their respective null hypotheses. Cell 2

reports the number and the percentage of cases that the ADF accepts the unit root null while the KPSS test reject the trend stationary null in favor of the unit root hypothesis. Cell 3 reports the number and the percentage of cases that the ADF rejects the unit root null in favor of the stationary alternative and the KPSS test accepts the trend stationary null. Cell 4 reports the number and the percentage of cases that both the ADF and KPSS tests reject their respective null hypotheses. The row total gives the total number and percentage from the ADF test while the column total gives those from the KPSS. The acceptance and rejection are based on the 5% significance level, and in general we will focus on the finite sample critical value results.

The interpretation of results from cell 2 and 3 is quite straightforward. Series that fall in cell 2 show a strong evidence of the presence of a unit root in the data while those in cell 3 show a strong evidence of trend stationarity. The cell 1 classification can be explained by the low power of both tests so that neither null is rejected: the data do not contain sufficient information to discriminate between the trend stationary and difference stationary hypotheses. On the other hand, the rejection of both the unit root and trend stationary null hypotheses, as in cell 4, cannot be attributed to the low power of one or both of the tests. One possible interpretation of such cases is that the data is governed by a more complex data generating process (DGP) than the standard trend and difference stationary classification.[11]

3.2. Results for All Countries

The results for the trend- and difference-stationarity tests for 126 countries are presented in Table 1.[12] For 111 (88%) of these series, the ADF test fails to reject the unit root null, and for

15 (12%), the test rejects (using a 5% marginal significance level). In contrast, 86 (68%) fail to reject the trend stationary null, and 40 (32%) reject, using the KPSS test. Hence, the econometrician relying solely on the ADF test might conclude that approximately 90% of all series contained unit roots, while the econometrician relying on the KPSS test would conclude that only one-third did.

When both test results are combined, the number of series that appear to be difference stationary is reduced considerably from that implied using solely the ADF. Admittedly, the most common outcome is "Accept-Accept", which implies some degree of ambiguity (57%, or 72 series, occupy cell 1). However, the next most common outcome, 31% (39 series), is in cell 2, corresponding to the "Accept ADF/Reject KPSS" outcome. 11% (14 series) "Reject ADF/Accept KPSS", which corresponds to a trend stationary view of GDP: Bahamas, Burundi, Canada, Congo, Cyprus, Hong Kong, Hungary, Iraq, Lesotho, Malta, Pakistan, the Philippines, South Korea, and Uganda. Only one series (Papua New Guinea) rejects both tests.

Two prominent overall results are illustrated in Table 1. For a large number of countries, no definite characterization can be made. However, for those countries that a definite characterization can be made, the difference stationary series outnumber the trend stationary series by about 3:1. This pattern holds true for most of the groupings examined below.

3.3. Geographic Differences

One of the advantages of the wide selection availed by using the Summers and Heston data set is that we can stratify the sample by various criteria. One stratification we choose is by geographic location, or specifically, by continent. One motivation behind this choice is the recent

trend of trading blocs in Europe, America, and Asia. Countries in a region linked by similar historical and cultural characteristics usually have similar economic conditions. This in turn makes it possible for these countries to cooperate and promote mutually beneficial economic goals. The results are presented in Table 2.

A consistent finding is that the proportion of failures to reject the ADF is always about 80%, rising to 100% for the case of Oceania (with, however, only five series). On the other hand, there is little variation in the tendency to reject the KPSS, except for the European region, where approximately one half of the series reject. Consequently, in all regions a large proportion fall into the ambiguous "Accept ADF/Accept KPSS" category, save in Europe, where 50% fall into the "Accept ADF/Reject KPSS" category. One common characteristic of the European countries is their relatively high income level. This suggests an alternative grouping.

3.4. Level of Development

A plausible explanation for the previous results is that the DGP governing GDP is different, depending upon the level of development. In this view, less developed, primarily rural-agrarian societies are likely to have output series with very different time series properties than those from highly urban, industrialized economies. To examine this possibility, we segmented the results into three categories -- low, medium and high income countries, corresponding to the World Bank's (1989) definitions. Low income countries are those with GNP per capita of $480 or less in 1987; middle income countries are those with GNP per capita of more than $480 and less than $6000; high income countries are those with GNP per capita in excess of $6000.[13]

The results are tabulated by level of development in Table 3. The results for low- and

middle- income countries are relatively similar: roughly 65% of the entries are "Accept ADF/Accept KPSS", roughly 24% are "Accept/Reject". The proportions change dramatically when one moves from these two categories to the high income category, primarily because then the KPSS test rejects trend stationarity more often. Then the majority of cases fall into the "Accept ADF/Reject KPSS" cell (16 out of 29 cases, or 55%), while a smaller proportion fall into the "Accept/Accept" cell.

This particular pattern illustrates the importance of casting the net wider. The relatively unambiguous support in favor of unit roots (against the alternative of a simple linear trend) in GDP appears to be more a rich-country phenomenon, than low or middle income one. In fact, among the OECD countries, only one country -- Canada -- rejects the unit root null.

One is tempted to judge this pattern of results as support for the view that high income (and especially developed[14]) economies behave differently than low and medium income ones. Caution in interpretation is necessary because high income countries also tend to be countries with high quality data, and the tendency to find more apparently trend stationary, or at least ambiguous, cases in the lower income countries may be due to the manner in which GDP series are calculated by these countries' statistical agencies. Alternatively, the valuation of GDP at international prices may be more precisely calculated for higher income countries.[15] In order to evaluate this possibility we stratify the GDP series by the quality of national income data.

3.5. Data Quality

Summers and Heston (1991) conveniently provide a "grade", ranging from D (worst) to A (best) for each country's data. While there is an apparent correlation between level of

development and data quality, the correlation is not exact. All low income countries have grade D and C quality data, but middle income Greece and Portugal have grade A quality data. In contrast, high income countries such as the Bahamas and Singapore have D and C quality data. Moreover, the Middle East oil exporting nations are also classified as high income countries, but have D quality data. Overall, the correlation between income and data quality is 0.71.[16]

The results stratified by data quality are presented in Table 4. For the low quality data the results are similar across grades C and D: roughly 23% are in the "Accept ADF/Reject KPSS" category, with 65% in the ambiguous "Accept/Accept" cell. When one moves to the grade A and B quality data, one finds that a majority of the results indicate unit root processes, since 16 out of 28 series Accept ADF and Reject KPSS.

A priori, there is no particular reason to believe that poorly constructed data will appear more trend stationary than well constructed data. However, it is conceivable that LDCs with limited statistics-gathering resources may use methods of calculating GDP involving linear interpolation and smoothing which generate data that look like trend stationary processes. For instance, the output of the traditional, or subsistence, sector is often assumed to grow with population, which is in turn interpolated between infrequent population censuses (Blades and Marczewski, 1974; UNRISD, 1977). Since the traditional sector often constitutes a large proportion of total output, this can be a significant source of measurement error. At the limit, the reported per capita GDP figures would be either a constant or a deterministic trend reflecting assumed productivity growth.[17]

4. STRUCTURAL BREAK TEST RESULTS

It is possible that the apparent finding of unit roots is due to the presence of structural breaks in otherwise trend stationary processes. We apply the BLS sequential test to GDP series for which both the ADF and KPSS tests indicate the presence of a unit root (detailed individual country results are contained in Appendix 3). Series for four countries appear to have significant breaks in the trend per capita GDP [break years in brackets]: Gambia [1978], Jamaica [1973], Spain [1975] and Yugoslavia [1980]. The Japanese [1974] series also rejects the null hypothesis of a unit root at the 10% significance level, along with trend breaks essentially the same as those obtained by BLS. No series reject the null of no mean shift.[18]

In sum, for the 39 series for which we found evidence of a unit root, only four can be explained by the presence of structural breaks. Hence, discrete shifts do not explain the tendency to find a relatively large number of series that fail to reject the ADF and reject the KPSS.

5. CONCLUSIONS

We have found that for a large proportion of GDP series, it is not possible to make definitive conclusions regarding their time series properties. However, for those series that a conclusion can be made (either "Accept/Reject" or "Reject/Accept"), we find evidence in support of unit roots as opposed to simple linear trends, in the ratio of approximately 3:1. This conclusion is based on the application of the ADF test, which has the unit root process as a null hypothesis, in conjunction with the KPSS test, which has a null of trend stationarity. Moreover we assess whether the apparent acceptance of the unit root view is due to the presence of structural breaks, of the nature examined by Banerjee, Lumsdaine and Stock (1992), and find that the cases of structural break account for a definite minority of the apparent unit root processes. In all these

tests, we have controlled for possible biases due to the use of inappropriate asymptotic critical values.

Finally, we have also found that there is substantial variation in how strongly the "accept ADF/reject KPSS" result holds. In particular, low incomes and poor data quality appear to be associated with greater ambiguity.

Endnotes

1. Recently, Cheung (1994) has applied this test the GDP data of to the G-7 countries, for the period 1900-1990.

2. Cogley (1990), Banerjee, Lumsdaine and Stock (1992), for example. Another difference between the current work and previous studies is the sample period. Some previous studies examine long historical data that span up to 100 years or more. Because of the consistency of GDP data across a large number of countries, we choose the Heston-Summers data. The trade-off is the possible reduction in the test power.

3. Although Riezman and Whiteman (1990) do apply Bayesian unit root tests to an almost equally large number of countries, they do not interpret the individual country results, nor do they analyze how the results vary over certain county groupings.

4. The variable used is RGDPCH (the Summers-Heston mnemonic), which is real GDP per capita evaluated in 1985 international prices calculated using a Chain index. The use of international prices converts GDP data from different countries to a common denomination and facilitate cross-country comparison. The specifics of the data series are detailed in Summers and Heston (1991), to which the interested reader is referred.

5. We report the results for all the series below. The overall pattern of rejections and non-rejections is virtually the same if we omit series with below 20 observations (there are only five such cases), or, surprisingly, even if we omit series with below 30 observations. However, in this latter case, the number of GDP series available is reduced from 126 to 75, or about 40%.

6. We use a Bartlett window, $w(s,\ell) = 1 - s/(\ell+1)$, as suggested by KPSS.

7. Recently, Leybourne and McCabe (1994) develop a unit root test that also has stationarity as its null hypothesis. However, the alternative process is an I(1) process with an MA(1) component. This is in contrast with the KPSS test which has an I(0) process under the null and an I(1) process under the alternative. Further the use of finite-sample critical values that account for both sample-size and truncation-lag effects should minimize the possible size-distortion of the KPSS test in small samples.

8. In the literature there are other methods to detect the break-point. As Zelhorst and De Haan (Forthcoming) and Cheung (1994) indicate that various methods to determine the break date yield similar results. Hence the use of the BLS method should yield reasonable results. The inclusion of only a shift in trend or mean in the regression regression can increase the power of the test. The reason is as follows. If there is only a shift in the trend, the inclusion of the shift in mean term lowers the power. Next consider there are shifts in both the trend and the mean. If only one term term is included in the regression, the term has to capture both break effects and gives a higher chance to reject the unit root null.

9. The simulated critical values are based on 10,000 replications. The computer code is available from the authors upon request.

10. For the sake of convenience, we will use the shorthand term "accept" in place of the more cumbersome phrase "fail to reject", while remaining cognizant of the difference.

11. However, if the true series was trend stationary with a structural break, this would be insufficient to explain the results, as ADF tests are biased against rejection of the unit root null in such cases. Another possibility is that GDP is persistent, but not completely, as in fractionally integrated processes (e.g., Diebold and Rudebusch, 1989). The small samples we have precludes examining this possibility; for the Geweke and Porter-Hudak (1983) test, there would be only six effective observations. The presence of nonlinear trends is another possible explanation.

12. Results for samples restricted to series with 19+, and 29+, observations are reported in Appendix 2.

13. It would have been preferable to use groupings based on incomes closer to the midpoint of the sample (e.g., 1970). However, earlier issues of the World Bank World Development Report grouped countries according to other development criteria besides income (see World Bank, 1989: x). Countries outside of the World Bank classification scheme but in our sample are: (low) Gambia. Cape Verde. Guyana, Angola; (middle) Taiwan, Swaziland, St. Lucia, St. Vincent, Fiji, Suriname, Cyprus, Malta, Barbados; (high) Bahrain, Bahamas, Iceland, Luxembourg.

14. The distinction is important, since some high income countries are oil exporters.

15. The PWT5 data set covers a large number of countries for which detailed price data are not available. For these, usually less developed, countries, the valuation calculations are based on a survey of capital city prices.

16. The correlation coefficient was calculated by assigning numerical scores to the data quality indices, such as, 3.3 for B+, and correlating this with the income category (1 for low, 3 for high).

17. An alternative data-based explanation is that higher quality data also tends to span longer periods. As the sample size grows the power of the KPSS test increases so that (given that the true processes are difference stationary) more entries fall into the "Accept/Reject" cell. In fact, all grade A quality series span the entire 39 years. Grade C series range from 26 to 39 years, while grade D series are as short as 10 years.

18. It is interesting that the indicated break points coincide with significant political events. For instance, Spain's and Yugoslavia's implied breaks match with the deaths of Franco and Tito, respectively, while the break in Jamaica coincides with Manley's attempted restructuring of the economy.

References

Bannerjee, Anindya, Robin Lumsdaine, and James Stock, "Recursive and Sequential tests of the Unit Root and Trend Break Hypotheses: Theory and International Evidence." Journal of Business and Economic Statistics, 1992, 10, 271-287.

Bernard, Andrew and Steven Durlauf, "Convergence of International Output Movements," NBER Working Paper #3717, 1991.

Blades, Derek and Witold Marczewski, "National Accounts Priorities for Low Income Countries from a Statistician's Point of View." in OECD Development Centre, National Accounts and Development Planning in Low Income Countries, Proceedings of a Study Session, Paris 1972, Paris: OECD, 1974.

Cheung, Yin-Wong, "Aggregate Output Dynamics in the 20th Century," Economics Letters, 1994, 45, 15-22.

_____, and Kon S. Lai, "Lag Order and Critical Values of the Augmented Dickey-Fuller Test," Journal of Business and Economics Statistics, forthcoming.

Cheung, Yin-Wong, Kon S. Lai and Tuan Tran, 1994, Finite-Sample Critical Values of the KPSS Test: A Response Surface Approach, Mimeograph, University of California, Santa Cruz, 1994.

Cogley, Timothy, "International Evidence on the Size of the Random Walk in Output," Journal of Political Economy, 1990, 98, 501-518.

Christiano, Lawrence J. and Martin Eichenbaum, "Unit Roots in Real GNP: Do We Know, and Do We Care?" in Karl Brunner and Allen Meltzer, eds., Carnegie-Rochester Conference Series on Public Policy,1990,32,7-62.

Diebold, Francis X. and Glenn D. Rudebusch, "Long Memory and Persistence in Aggregate Output," Journal of Monetary Economics, 1989, 24, 189-209.

Geweke, John and S. Porter-Hudak, "The Estimation and Application of Long Memory Time Series Models," Journal of Time Series Analysis, 1983, 4, 221-238.

Granger, C.W.J. and Paul Newbold, "Spurious Regressions in Econometrics," Journal of Econometrics, 1974, 2, 111-120.

De Haan, Jakob and Dick Zelhorst, "Does Output have a Unit Root? New International Evidence," Applied Economics, 1993, 25, 953-960.

Hall, Alastair, "Testing for a Unit Root in Time Series with Pretest Data-based Model Selection,"

Journal of Business and Economics Statistics, 1994, 12, 461-470.

Harris, R.I.D., "Small Sample Testing for Unit roots," Oxford Bulletin of Economics and Statistics, 1992, 54 615-625.

Kormendi, Roger C. and Philip Meguire, "A Multicountry Characterization of the Nonstationarity of Aggregate Output," Journal of Money, Credit, and Banking, 1990, 22, 77-93.

Kwiatkowski, Denis, P.C.B. Phillips, Peter Schmidt and Yongcheol Shin, "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root?" Journal of Econometrics, 1992, 54, 159-178.

Leybourne, S.J. and B.P.M. McCade, "A Consistent Test for a Unit Root," Journal of Business & Economic Statistics, 1994, 12, 157-166.

Perron, Pierre, "Testing for a Unit Root in a Time Series with a Changing Mean," Journal of Business and Economic Statistics, 1990, 8, 153-162.

_____, "Testing for a Random Walk: A Simulation Experiment When the Sampling Interval Is Varied," in B. Raj, ed., Advances in Econometrics and Modeling, Kluwer Academic Publishers, 1989a.

_____, "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," Econometrica, 1989b, 57, 1361-1401.

Riezman, Raymond G. and Charles H. Whiteman, "Worldwide Persistence, Business Cycles, and Economic Growth," Social Science Working Paper #719, California Inst. of Technology, February 1990.

Rudebusch, Glenn D., "The Uncertain Unit Root in Real GNP," American Economic Review. 1993, 83, 264-272.

Shiller, Robert J., and Pierre Perron, "Testing the Random Walk Hypothesis: Power versus Frequency of Observation," Economics Letters, 1985, 18: 381-386.

Summers, Robert and Alan Heston, "The Penn World Table (Mark 5): An Expanded Set of International Comparisons," Quarterly Journal of Economics, May 1991.

UN Research Institute for Social Development, Research Data Bank of Development Indicators: Volume IV Notes on the Indicators,
Report 77.2. Geneva: United Nations, 1977.

World Bank, World Development Report, 1989, Oxford: Oxford University Press, 1989.

Zelhorst, Dick and Jakob De Haan, "Testing for a Break in Output: New International Evidence," Oxford Economic Papers, forthcoming.

_____, and _____, "The Nonstationarity of Aggregate Output: Some Additional Evidence," Journal of Money, Credit and Banking, 1994, 26, 9-22.

TABLE 1
RESULTS FOR ALL COUNTRIES

		KPSS test		
		ACCEPT	REJECT	
ADF	ACCEPT	72 57%	39 31%	111 88%
test	REJECT	14 11%	1 1%	15 12%
		86 68%	40 32%	126 100.0%

Notes: Results using the 5% significance level and finite sample critical values. "Accept ADF and accept KPSS" indicates failure to reject both the unit root null and the trend stationary null hypotheses. "Accept ADF and reject KPSS" indicates the failure to reject unit root null, but rejection of trend stationary null. "Reject ADF and accept KPSS" indicates rejection of unit root null and failure to reject the trend stationary null. "Reject ADF and Reject KPSS" indicates rejection of both the unit root and trend stationary null hypotheses. KPSS results refer to use of $\ell 8$ rule.

TABLE 2
RESULTS BY GEOGRAPHICAL REGION

African Countries: n=43

		KPSS test: ACCEPT	KPSS test: REJECT	
ADF test	ACCEPT	25 58%	14 33%	39 91%
	REJECT	4 9%	0 0%	4 9%
		29 67%	14 33%	43 100.0%

Asian Countries: n=26

	KPSS test: ACCEPT	KPSS test: REJECT	
ACCEPT	17 65%	4 15%	21 81%
REJECT	5 19%	0 0%	5 19%
	22 85%	4 15%	26 100.0%

European Countries: n=23

		KPSS test: ACCEPT	KPSS test: REJECT	
ADF test	ACCEPT	8 35%	12 52%	20 87%
	REJECT	3 13%	0 0%	3 13%
		11 48.0%	12 52.0%	23 100.0%

North American Countries: n=17

	KPSS test: ACCEPT	KPSS test: REJECT	
ACCEPT	12 71%	3 18%	15 88%
REJECT	2 12%	0 0%	2 12%
	14 82%	3 18%	17 100.0%

Oceania: n=5

		KPSS test: ACCEPT	KPSS test: REJECT	
ADF test	ACCEPT	2 40%	2 40%	4 80%
	REJECT	0 0%	1 20%	1 20%
		2 40.0%	3 60.0%	5 100.0%

South American Countries: n=12

	KPSS test: ACCEPT	KPSS test: REJECT	
ACCEPT	8 67%	4 33%	12 100%
REJECT	0 0%	0 0%	0 0%
	8 67%	4 33%	12 100.0%

Notes: Results using 5% significance level. See notes to Table 1.

TABLE 3
RESULTS BY INCOME LEVEL

Low Income Countries: n=38

		KPSS test: ACCEPT	KPSS test: REJECT	
ADF test	ACCEPT	25 66%	9 24%	34 89%
	REJECT	4 11%	0 0%	4 11%
		29 76%	9 24%	38 100.0%

Middle Income Countries: n=59

		KPSS test: ACCEPT	KPSS test: REJECT	
ADF test	ACCEPT	37 63%	14 24%	51 86%
	REJECT	7 12%	1 2%	8 14%
		44 75%	15 25%	59 100.0%

High Income Countries: n=29

		KPSS test: ACCEPT	KPSS test: REJECT	
ADF test	ACCEPT	10 34%	16 55%	26 90%
	REJECT	3 10%	0 0%	3 10%
		13 45%	16 55%	29 100.0%

Notes: Results using 5% significance level. See notes to Table 1.

TABLE 4
RESULTS BY DATA QUALITY

D and C Quality Data: n=98

		KPSS test: ACCEPT	KPSS test: REJECT	
ADF test	ACCEPT	64 65%	23 23%	87 89%
	REJECT	10 10%	1 1%	11 11%
		74 76%	24 25%	98 100.0%

B and A Quality Data: n=28

		KPSS test: ACCEPT	KPSS test: REJECT	
ADF test	ACCEPT	8 29%	16 57%	24 86%
	REJECT	4 14%	0 0%	4 14%
		12 43%	16 57%	28 100.0%

Notes: Results using the 5% significance level. See notes to Table 1.

APPENDIX 1
Detailed Unit Root Test Results

Notes:

NOB is the number of observations. QLT is the data quality assigned by Summers and Heston (1991). ALAG is the lag length chosen by the AIC. ASTAT is the ADF statistic. AA95 and AF95 refer to whether the ADF test rejects at the 5% significance level, using either the standard critical values, controlling for only sample size (AA95), or the critical values controlling for both sample size and lag structure (AF95). An "A" entry denotes failure to reject, and "R" entry denotes rejection of the null. ℓ8 denotes the KPSS statistic using the ℓ8 rule. The next two columns denote failure to reject, or rejection, based on the KPSS ℓ8 rule and asymptotic (KA895) or finite sample (KF895) critical values.

The next two columns presents summaries of combinations of results for the ADF and KPSS tests; for instance "AR" denotes failure to reject the ADF, and rejection of the KPSS, tests. AKA895 refers to standard ADF/asymptotic KPSS critical values and KPSS ℓ8 rule; AKF895 refers to ADF critical values controlling for sample size and lag structure, finite sample KPSS critical values and KPSS ℓ8 rule.

More detailed results, reporting the lag length, data quality, etc., are available from the authors upon request.

APPENDIX 2
Patterns in Test Results
Using Restricted Samples

Notes:

These tables provide information analogous to that provided in Tables 1-4, for restricted samples. On the left hand side are tables for samples restricted to series with 20 and above observations; on the right hand side, those with 30 and above.

APPENDIX 3
Detailed Structural Break Test Results

Notes:

AICLAG is the lag length chosen by the AIC.

GSMXF is the maximum F statistic for a shift in the trend (change in slope).

GSYR is the indicated break year.

GSFDF is the ADF statistic for the residuals assuming a break at GSYR.

GSMID is the minimum ADF statistic for the residuals.

GBMXF is the maximum F statistic for a shift in the mean of the trend.

GBYR is the indicated break year.

GBFDF is the ADF statistic for the residuals assuming a break at GBYR.

GBMID is the minimum ADF statistic for the residuals.