Applied Economics Letters, 1995, **2**, 1–6

How sensitive are trends to data definitions? Results for East Asian and G5 countries

YIN-WONG CHEUNG, MENZIE DAVID CHINN AND TUAN TRAN

Department of Economics, University of California, Santa Cruz, CA 95064, USA

Received 3 August 1994

This paper examines whether test results characterizing per capita output as either trend or difference stationary are sensitive to whether output is valued in domestic currency terms, or in some international numeraire, such as the Summers-Heston international dollar. Using the conventional ADF test, *and* the Kwiatkowsi *et al.* test with a trend stationary null, we find that for economies such as those of the East Asian countries, the best description of the persistence of the data *does* depend upon the valuation of output. No such discrepancy is found for the output series of the G5 countries. We conclude that researchers should be extremely cautious about making generalizations regarding the time-series properties of output.

I. INTRODUCTION

The issue of whether aggregate output contains a unit root has occupied a central role in macroeconomic debate over the past decade. This concern has risen is tandem with a resurgence in interest in long run growth. As a consequence, the Penn World Table Mark 5 (PWT5) of Summers and Heston (1991) has become a fixture in empirical analyses of growth and convergence.[1] The PWT5 data set attempts to control for differences in systems of national accounts and, more importantly, deviations from purchasing power parity, so that incomes can be compared *across* countries (or 'interspatially') as well as intertemporally.

Recent work has made it apparent that the inferred time series characteristics of per capita output are dependent upon the data quality. Cheung and Chinn (1994) find that the higher the data quality, the more likely one is to be able to characterize data as either trend stationary or difference stationary. Since different imputation procedures are used for high quality data countries as opposed to low ones, it may prove useful to investigate the sensitivity of test results to data definitions.

In this paper we focus on the Pacific Basin countries for two reasons. First, these countries have been the centre of much discussion about the sources of rapid economic growth. Second, we expect the difference between growth rates expressed in domestic currency and international dollar valuations to be most pronounced in countries experiencing rapid productivity growth and rapid changes in relative prices.

We apply two types of tests – a test with a unit root null (the ADF test of Fuller (1976)) and a test with a trend stationarity null (the KPSS test of Kwiatkowski *et al.* (1992)) – to data on the East Asian countries: China, Hong Kong, Indonesia, Korea, Malaysia, Philippines, Singapore, Taiwan and Thailand. We compare the results obtained from the PWT5 data set to those obtained from the conventional country sources. We then repeat the exercise for the G5 countries (USA, UK, Germany, Japan and Canada).

We find that for the East Asian countries we obtain very different statistical results depending upon the data set: more evidence for unit roots is found in the domestic currency series than in the PPP valued series. In contrast, for the developed countries the purchasing power parity imputation does not appear to make any substantial difference for statistical inferences.

This paper is organized as follows. In Section II, we discuss the econometric methodology used. The data and the ADF and KPSS results are presented in Section III. Section IV concludes.

[1]See for instance Romer (1990), Fischer (1991), and Levine and Renelt (1992), in a cross-section context. Riezman and Whiteman (1990) examine the individual time series for all the countries in the PWT5 data set, as well as for a world aggregate.

1350–5851

II. ECONOMETRIC METHODOLOGY

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.

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

Test	H_O:	H_A:
ADF	$I(1)$	$I(0)$
KPSS	$I(0)$	$I(1)$

Hence, if the ADF test rejects the null while the KPSS test fails to reject the null, then this outcome is considered strong evidence in favour of a trend stationary process. If, in contrast, the ADF test fails to reject, while the KPSS test rejects, we have strong evidence in favour of a difference stationary process. The failure of both tests to reject could be attributed to the low power of the tests. Rejection of both null hypotheses could be due to a data generating process more complex than those considered.

The ADF Test

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

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

The unit root null hypothesis is rejected if $\hat{\pi}$ is significantly less than zero. While it is typical to conduct the tests using the standard critical values given in Fuller (1976) which control only for sample size, we choose to use the more appropriate finite sample critical values calculated from simulated distributions which control for both sample size and lag structure (see Appendix A for details.[2] [3]

Trend stationarity tests

Assume that the time series is the sum of a deterministic trend, a random walk and a stationary error. The KPSS test is a Lagrange Multiplier test for the null hypothesis that the error variance in the random walk component of the series is zero.

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

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

where S_t is the partial sum process defined by

$$S_t = \sum_{i=1}^{t} e_i,\ t = 1, 2, \ldots, T \tag{3}$$

and $s^2(\ell)$ is the serial correlation and heteroscedasticity consistent variance estimator given by

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

$w(s, \ell)$ is an optimal weighting function corresponding to the choice of a spectral window.[4]

The null of trend stationarity is rejected in favour of the unit root alternative if the KPSS statistic is larger than the critical values provided by KPSS. On the basis of simulation results, KPSS assert that their test has good size and power characteristics. Since the KPSS test is a relatively new tool, we focus on the results based on the finite sample critical values (see Appendix A for details of how these values were generated). Following KPSS's suggestion, we adopt the $\ell 8$ rule, which sets $\ell = \mathrm{INT}[8(T/100)^{1/4}]$ in Equation 4.

III. DATA AND TEST RESULTS

Data description

We use two types of output data. The first type is real per capita GNP or GDP, valued in real domestic currency terms. These data are drawn from either the IMF's *International Financial Statistics* or directly from the domestic statistical agencies.[5] Such series are appropriate for intertemporal comparisons.

The second type is GDP per capita in real 1985 international dollar terms obtained from the Summers and Heston (1991) Penn World Tables, Mark 5. The variable (*RGDPCH*, using the Summers and Heston mnemonic) is calculated using a chain index to minimize base year problems.

Since the specifics of the Summers and Heston methodology for imputing international prices are not well known, it is useful to review the mechanics of the procedure. For the countries involved in the UN's International Comparison Project (ICP), prices of hundreds of identically specified goods and services are collected. The resulting price parities and subaggregate and total

[2]Cheung and Lai (1993) have shown in Monte Carlo studies that both parameters are important.
[3]The lag parameter k is selected by the Schwartz Bayesian information criterion (SIC). Alastair Hall (1994) has shown that a general-to-specific modelling strategy that uses the data to determine k can improve both the size and power of the ADF test.
[4]We use a Bartlett window, $w(s, \ell) = 1 - s / (\ell + 1)$, as suggested by KPSS.
[5]The Hong Kong data are from the Hong Kong Census and Statistics Department, *Estimates of Gross Domestic Product: 1966–1992* (March 1993). The Taiwanese data are from the Council for Economic Planning and Development, Republic of China *Taiwan Statistical Data Book* (June 1988). The Chinese domestic currency series is national income, rather than GDP, in 1985 yuan.

aggregate levels are used to convert the countries' national currency expenditures into a common numeraire. This common numeraire is defined in such a way that US GDP in 1985 US$ and in 1985 international dollars are the same. In principle, this makes units of GDP comparable across countries as well as across time.[6]

The sample period covers the post-war era for the developed countries, and somewhat shorter time periods for some of the less-developed countries, for the period up to 1988. The data span at most 39 observations, and at minimum 19 (Malaysia). These sample sizes may appear small, and hence likely to worsen the power of the ADF test. However, as pointed out by Schiller and Perron (1985), the power of the tests for stationarity depends 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, say 39 × 12 monthly observations.

Overview of econometric results

The summary results are presented in the following format:

	Fail to reject KPSS	Reject KPSS
Fail to reject ADF	Cell 1	Cell 2
Reject ADF	Cell 3	Cell 4

Cell 1 reports the cases that both the ADF and KPSS tests fail to reject their respective null hypotheses. Cell 2 reports the cases that the ADF fails to reject the unit root null while the KPSS test rejects the trend stationary null in favour of the unit–root hypotheses. Cell 3 reports the cases that the ADF rejects the unit root null in favour of the stationary alternative and the KPSS test fails to reject the trend-stationary null. Cell 4 reports the cases where both the ADF and KPSS tests reject their respective null hypotheses. The rejections are based on the 5% significance level.

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 are governed by a more complex data-generating process than either a deterministic trend or a unit root process.

Empirical results

The summary results for the trend- and difference-stationarity tests for the East Asian countries are presented in Table 1; the detailed results are in Table 2. We first discuss the conventional (domestic currency) valuation series. The most common outcome is 'fail to reject–fail to reject': Indonesia, Malaysia, Philippines, Singapore and Taiwan. Three series fail to reject the ADF null and reject the KPSS null (China, Korea and Thailand). Only the Hong Kong series falls into the apparently trend stationary category.

We might expect that the time–series characteristics of output series would be the same regardless of the valuation method – that is a shock to output is permanent regardless of whether it is

Table 1. *ADF and KPSS results for East Asian Countries*

	Fail to reject KPSS	Reject KPSS
Domestic currency		
Fail to reject ADF	Indonesia Malaysia Philippines Singapore Taiwan	China Korea Thailand
Reject ADF	Hong Kong	
Summers and Heston		
Fail to reject ADF	China Indonesia Korea Malaysia Singapore Taiwan	
Reject ADF	Hong Kong Philippines Thailand	

NOTES: Results using the 5% significance level. 'Fail to reject ADF and fail to reject KPSS' indicates failure to reject both the unit–root null and the trend–stationary null hypotheses. 'Fail to reject ADF and reject KPSS' indicates the failure to reject unit–root null, but rejection of trend–stationary null. 'Reject ADF and fail to reject 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.

[6]One problem is that not all of the East Asian countries participated in the ICP. Hong Kong, Malaysia, Philippines, South Korea and Thailand did (as did all of the G5 countries), while Indonesia and Taiwan did not. China participated in a quasi-benchmarking procedure with the USA. For those countries that did not participate in the ICP, survey data from the capital cities were used instead.

Table 2. *Results for East Asian countries*

Pacific countries (with ADF SIC, KPSS *l*8)

			IFS				Summer-Heston			
			ADF			Decision	ADF			Decision
Country	Period	IFS-series	Lag	Statistic	KPSS statistic		Lag	Statistic	KPSS statistic	
China	1960–88	NI 85	2	−3.075	0.153	AR	2	−3.1072	0.142	AA
Hong Kong	1966–88	GDP 80*	2	−4.067	0.104	RA	2	−3.9343	0.099	RA
Indonesia	1964–88	GDP 85	5	−0.142	0.110	AA	2	−1.2232	0.096	AA
Korea	1953–88	GDP 85	1	−1.721	0.151	AR	1	−1.9185	0.105	AA
Malaysia	1970–88	GDP 85	1	−1.262	0.144	AA	2	−1.9984	0.136	AA
Philippines	1950–88	GDP 85	2	−2.018	0.115	AA	2	−3.6009	0.137	RA
Singapore	1960–85	GDP 85	5	−0.4	0.086	AA	5	−1.8755	0.079	AA
Taiwan	1952–86	GNP 81**	2	−2.547	0.123	AA	2	−2.6744	0.117	AA
Thailand	1950–88	GDP 85	8	−3.316	0.141	AR	5	−5.6834	0.097	RA

NOTES: *Data from Census and Statistics Department – Hong Kong (March 1993). **Data from Council for Economic Planning and Development – Republic of China (June 1988). ADF statistics are based on SIC lags. KPSS statistics are based on the *l*8 lag rule: $l = \text{INT}(8(T/100)^{\wedge}0.25)$. Decisions are based on finite-sample critical values at 5% significance level.

valued in domestic currency terms or international dollars. When using the Summers and Heston data set, we find that *four out of the nine series have switched classification.* Six series fall into the ambiguous 'fail to reject–fail to reject' category, including China and Korea, while the Philippines and Thailand now fall into the 'Reject–fail to reject' category.

The results become more compelling when the conclusions are restated: if the two test results agree, then in the domestic currency series a difference–stationary process is a more common finding, while in the Summers and Heston data set, the *only* unambiguous finding is that of trend stationarity.

To examine whether this contrast in results was common, we repeated the exercise for the G5 countries (US, UK, Japan, Germany and France) which presumably have the best price and quantity data. The results from this exercise are reported in Table 3. We find that the results do not differ between the two output series.

Next we examined whether the change in results for the Asian countries was due to the sensitivity of one or the other tests. The ADF switches from 'fail to reject' to 'reject' for the Philippines and Thailand when moving from domestic data to PWT5 data. In the latter case, the SIC chooses a different lag. In the former, the same lag length is chosen, and yet the ADF statistics are very different (–2.02 versus –3.61). The KPSS switches from 'reject' to 'fail to reject' in the cases of China, South Korea and Thailand (using the same lag window). Thus both tests tend to provide more evidence of trend stationarity when output is valued at international prices.

Summers and Heston (1991) point out that the difference in calculated growth rates using either domestic currency or international dollars is most pronounced when there are drastic changes in relative prices. East Asian countries appear to be likely candidates for this effect. However, it is surprising that the mode of valuation affects the *persistence* characteristics of output. If indeed the Summers and Heston method does provide an adequate measure of 'quantities' of production, this suggests that findings of difference–stationary output in LDCs are likely as a result of persistent relative price changes.

Table 3. *Results for G5 countries*

G5 countries (with ADF SIC, KPSS *l*8)

			IFS				Summer-Heston			
			ADF			Decision	ADF			Decision
Country	Period	IFS-series	Lag	Statistic	KPSS statistic		Lag	Statistic	KPSS statistic	
France	1950–88	GDP 85	1	−2.038	0.171	AR	1	0.89377	0.164	AR
Germany	1960–88	GNP 85	1	−1.386	0.157	AR	1	−1.665	0.159	AR
Japan	1952–88	GNP 85	2	−0.557	0.166	AR	2	−0.5569	0.166	AR
UK	1950–88	GDP 85	1	−2.194	0.129	AA	1	−2.7026	0.130	AA
USA	1950–88	GDP 85	2	−2.582	0.086	AA	1	−2.3781	0.079	AA

NOTES: ADF statistics are based on SIC lags. KPSS statistics are based on the *l*8 lag rule: $l = \text{INT}(8(T/100)^{\wedge}0.25)$. Decisions are based on finite-sample critical values at 5% significance level.

IV. CONCLUSIONS

In this paper we have assessed the persistence in output for nine East Asian countries, using a test with a unit–root null, as well as a test with a trend–stationary null. We use output data in both domestic real terms, as well as in international dollar terms.

We obtain several interesting results. First, for about two-thirds of the East Asian countries, the tests have inadequate power to reject their respective null hypotheses. Second, and more importantly, the finding of evidence in support of trend stationarity versus difference stationarity depends importantly on the type of data used. In general, output measured in a common international numeraire appears more trend stationary than its domestic–currency counterpart. Both tests appear to detect this effect.

This set of results suggests that researchers should be careful about extrapolating results from one series to another. Moreover, it suggests caution in indiscriminately using either time series. For instance, while the Summers and Heston series may have the desirable attribute of providing a 'real' quantity comparable across countries, it is important to remember that agents in these economies do not face 'international' prices, and instead may base their decisions on domestic-currency prices.

ACKNOWLEDGEMENTS

We thank the Pacific Rim Project of the University of California for financial support.

REFERENCES

Cheung, Y.-W. and Chinn, M. D. (1994) Deterministic, stochastic, and segmented trends in aggregate output: a cross-country analysis, *GICES/Dept. of Economics Working Paper No. 282*, University of California, Santa Cruz, February.

Cheung, Y.-W. and Lai, K. S. (1993) Lag order and the finite sample behavior of the augmented Dickey–Fuller test, *GICES/Dept. of Economics Working Paper No. 269*, University of California, Santa Cruz, September.

Fischer, S. (1991) Growth, macroeconomics and development, in *NBER Macroeconomics Annual, 1991*, ed. O. J. Blanchard and S. Fischer, MIT Press, Cambridge, pp. 329–63.

Fuller, W. (1976) *Introduction to Statistical Time Series*, John Wiley, New York.

Hall, A. (1994) Testing for a unit root in time series with pretest data based model selection, *Journal of Business and Economic Statistics*, April.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. and Shin, Y. (1992) 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*, **54**, 159–78.

Levine, R. and Renelt, D. (1992) A sensitivity analysis of cross-country growth regressions, *American Economic Review*, **82**, 942–63.

Riezman, R. G. and Whiteman, C. H. (1990) Worldwide persistence, business cycles, and economic growth, *Social Science Working Paper No. 719*, California Institute of Technology, February.

Romer, P. (1990) Capital, labor, productivity, *Brookings Papers on Economic Activity: Microeconomics 1990*, 337–69.

Shiller, R. J. and Perron, P. (1985) Testing the random walk hypothesis: power versus frequency of observation, *Economics Letters*, **18**, 381–6.

Summers, R. and Heston, A. (1991) The Penn World Table (Mark 5): an expanded set of international comparisons, *Quarterly Journal of Economics*, May.

APPENDIX A: DESCRIPTION OF CALCULATION OF FINITE SAMPLE CRITICAL VALUES

ADF Test Statistics

In generating the critical values controlling for both sample size and lag structure, a response surface analysis was used. This term applies to a system where the response of some variable depends on a set of other variables that can be controlled and measured in experiments. The surface is usually fitted by regression analysis.

Cheung and Lai (1993) report details of the simulation for the ADF critical values. The control variables are the sample size (N) and the lag parameter (k). A factorial experimental design is used, with 200 total combinations of $N = \{27, 30, 33, 36, 39, 42, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 250, 300, 350, 400, 500\}$ and $k = \{1, 2, 3, 4, 5, 6, 7, 9\}$. The DGP is specified as:

$$x_t = x_{t-1} + e_t \tag{A1}$$

with e_t distributed n.i.d. with mean zero. Each experiment contains 30 000 replications.

The following response surface is fitted:

$$CR_{N,k}^{ADF} = \tau_0 + \sum_{i=1}^{r} \tau_i (1/T)^i + \sum_{j=1}^{s} \phi_j (k/T)^j + w_{N,k} \tag{A2}$$

where $CR_{N,k}$ is the finite sample simulation estimate of the critical value for a sample size N and lag order k, $T = N - k$, the effective number of observations, and $w_{N,k}$ the random error term.

The results are reported in Table A1.

The KPSS Finite Sample Test Statistics

The generation of the KPSS response surface follows in principle that set forth above for the ADF. The two variables controlled in the experiments are the total number of observations, N, and the lag *truncation* parameter, ℓ. The DGP is specified as:

$$x_t = \varepsilon_t \tag{A3}$$

where ε is an error n.i.d. with mean zero. There are 30 000 replications in each experiment.

The following response surface is fitted:

$$CR_{N,\ell}^{KPSS} = \delta + \sum_{i=1}^{r} \theta_i (1/T)^i + \sum_{j=1}^{s} \psi_j (\ell/T)^j + V_{n,\ell} \tag{A4}$$

Using the effective number of observations, T, does not make a

Table A1. *Response surface estimation of critical values*

Reg coefficient	ADF with trend 10%	ADF with trend 5%	Reg coefficient	KPSS 10%	KPSS 5%
τ_0	–3.1219*	–3.4013*	δ	0.1198*	0.1489*
	(0.0021)	(0.0025)		(0.0001)	(0.0001)
τ_1	–4.5243*	–6.8786*	Θ_1	–0.0397*	–0.1986*
	(0.2304)	(0.2675)		(0.0165)	(0.0245)
			Θ_2	–3.8261*	–6.0331*
				(0.5040)	(0.7467)
ϕ_1	1.0301*	1.2123*	ψ_1	–0.0211*	–0.1320*
	(0.0544)	(0.0614)		(0.0037)	(0.0055)
ϕ_2	–0.9518*	–1.7218*	ψ_2	0.2665*	0.2459*
	(0.1454)	(0.1562)		(0.0392)	(0.0581)
			ψ_3	1.2195*	1.9767*
				(0.1099)	(0.1629)
$\bar{R}^2$	0.8230	0.8701		0.9886	0.9485
MSE	0.0003	0.0004		4.9×10^{-7}	1.1×10^{-6}

NOTES: The response surface for the ADF with trend is given by Equation A2. The response surface for the KPSS is given by Equation A4. Estimates for critical values are obtained from simulation with 30 000 replications. The numbers in parentheses are heteroscedasticity-consistent standard errors. Statistical significance at the 5% level is indicated by an asterisk (*). MSE is mean squared error.

substantial difference in the results. The estimated response surface is reported in Table A1.