ΣΠΟΥΔΑΙ / SPOUDAI

ΕΤΟΣ 2001	ΙΟΥΛΙΟΣ-ΔΕΚΕΜΒΡΙΟΣ	ΤΟΜΟΣ 51	ΤΕΥΧ. 3-4
YEAR 2001	JULY-DECEMBER	VOL. 51	No 3-4

AN ALTERNATIVE APPROACH FOR SELECTING *TS vs. DS* PROCESSES USING THE NELSON AND PLOSSER TIME SERIES

By

*Christos Agiakloglou**
Department of Economics, University of Piraeus, Greece

Abstract

The initial study of Nelson and Plosser (1982) has been established in the literature as a point of reference and many research papers have worked on their data set trying to determine whether each U.S. series is generated by a trend stationary or by a difference stationary process. The objective of this paper is to re-examine the Nelson and Plosser data set using maximum likelihood estimation and to comment on the results based on the existing testing procedures. (JEL Classification System: C12, C22)

1. Introduction

The ability to determine whether an observed macroeconomic time series is generated by a Trend Stationary *(TS)* process or by a Difference Stationary *(DS)* process, according to the terminology of Nelson and Plosser (1982), has received a considerable attention by many time series analysts. In essence, the challenging feature of this particular analysis is the idiosyncratic structure of most aggregate macroeconomic time series, which are believed to be generated by a unit autoregressive root processes, whereas at the

* I would like to thank Professor Paul Newbold for his useful comments and suggestions and Professor Theodora Sougianni for his support. My thanks also to the anonymous referee for his critical remarks.

same time their magnitude grows over time indicating the presence of a time trend component. Thus, following the initial study of Nelson and Plosser (1982) this type of research has been the main ambition of several interesting studies emphasizing the possibility of identifying, mainly through testing procedures, the dominant characteristic of the generating process using the same 14 annual U.S. series that were originally studied by Nelson and Plosser (1982).

The objective of this paper is to re-examine the Nelson and Plosser (1982) data set using maximum likelihood estimation and to comment on the results based on the existing testing procedures. The main advantage of this approach is the fact that the decision as to what is the structure of the generating behavior of each process is going to be independent of any testing procedure, minimizing in that sense the possibility of obtaining misleading results either because the sample size is small or because the performance of the applied test is not satisfactory.

2. Deterministic versus Stochastic Trend

If the generating behavior of an observed time series X_t, $t=1,2,...,T$, is being characterized as a deterministic function of the time trend and at the same time all fluctuations from the time trend can be expressed as a stationary and invertible *ARMA(p,q)* process, then the process is called trend stationary. On the other hand, if the time trend component has a stochastic rather than a deterministic role in nature, then the process is called difference stationary and it is stationary and invertible *ARIMA(p,1,q)* process.

An observed time series X_t is expressed as a *TS* process if

$$X_t = a + \beta t + u_t \text{ and } \Phi(B)\, u_t = \Theta(B)\, \varepsilon_t \qquad (1)$$

and it is expressed as a *DS* process if

$$\Xi(B)(1 - B)(X_t - \mu) = \Psi(B)\, \varepsilon_t \qquad (2)$$

where ε_t is white noise, B is the backshift operator, μ is the mean of the process and the polynomials $\Phi(B)$ & $\Xi(B)$ and $\Theta(B)$ & $\Psi(B)$ satisfy conditions of stationarity and invertibility with known orders respectively.

Moreover, if an observed time series is characterized as a difference stationary process, then the existence of a unit autoregressive root must be identified for the levels of the series, whereas if the process is characterized as a trend stationary process, then a unit moving average root must be identified in the first differences of the series. Selecting thus one of the above models, (1) or (2), as the best fitted model for a given time series can sometimes become a very challenging issue especially when the magnitude of the autoregressive or the moving average root is near one. This is in fact the reason why many research papers have come to a different conclusion for the same data set based on the statistical methodology that they used. A classical example of this situation is the Nelson and Plosser data set.

It should be pointed out as a concluding remark that the most attractive feature of this analysis is the ability to use the economic theory to explain the generating behavior of an observed time series based on the selected model. In other words, the issue of distinguishing a *TS* process from a *DS* process is very important mainly from the economic theoretical point of view since it determines whether shocks to the economic system persist for a short or for a long period of time respectively. To the contrary, if the objective is to select one of the above models to use it only for making inference about very short-run future behavior, then from the statistical point of view it will make no difference which model will be selected since both types of models will generate very similar, if not identical, forecasts. The challenge however remains in statistics and not in economics and it is concentrated on the effort to successfully develop a test that will reliably determine the generating behavior of an observed time series with a small number of observations.

3. Testing procedures

Consider first the case of testing the null hypothesis of a difference stationary process H_{DS} against the alternative hypothesis of a trend stationary process H_{TS}. This test is a unit autoregressive root test and it is basically implemented by using the augmented Dickey-Fuller test, known as the *ADF* test, introduced by Said and Dickey (1984) as an extension of the original Dickey and Fuller (1979) test. The structure of the test is based on the approximation of an autoregressive — moving average model of unknown order by an autoregressive process of a sufficiently large order and it is implemented using the following regression:

$$\Delta X_t = \alpha + \beta t + \gamma X_{t-1} + \sum_{j=1}^{k} \delta_j \Delta X_{t-j} + \varepsilon_t \tag{3}$$

where Δ is the first difference operator and the t-statistic for testing the null hypothesis that $\gamma=0$ has the same distribution as that originally tabulated by Fuller (1976). A number of empirical studies, including Schwert (1989) and Agiakloglou & Newbold (1992), have reported simulation evidence showing that the performance of the augmented Dickey-Fuller test is not satisfactory in moderately large samples even for the simplest possible model with one moving average term the ARIMA(0,1,1) process. In fact, this methodology of approximating an ARIMA(p,l,q) process with unknown orders p and *q* to an ARIMA(k,0,0) process is strongly affected not only by the order of the approximating autoregression but also by the magnitude of the moving average component especially if the generating model contains a substantial moving average component.

Our objective however is not simply to detect whether or not each series contains a unit autoregressive root, but to determine whether model (1) or model (2) describes better the generating behavior of each series. Thus, our null hypothesis is a joint hypothesis of testing that the coefficients of equation (3) β and γ are simultaneously zero, i.e., H_o: $\beta=\gamma=0$. This test is implemented by estimating equation (3) for a pre-selected value of *k* and applying the regression F-statistic, known as the Φ_3-statistic, introduced by Dickey and Fuller (1981). Hence, equation (3) is estimated for all 14 U.S. series, which are expressed in natural logs, except for the bond, yield series, and the results are reported on Table 1. Critical values for this test are obtained from Dickey and Fuller (1981). The most appropriate value of *k,* among all values of k=0,l,..,10, is selected by estimating equation (3) using the same sample according to the following three methods: the Akaike Information Criterion *(AIC),* the Schwarz Bayesian Criterion *(SBC)* and the general-to-specific testing approach using a nominal 5% level test (T(5)) proposed by Hall (1994) and Ng & Perron (1995).

The null hypothesis of difference stationary processes, as reported by Nelson and Plosser (1982) and Table 1 shows, is not rejected for all U.S. series except for the unemployment series for which the evidence against the unit root hypothesis is stronger if the *SBC* criterion is used to select the value of *k.* Perhaps, the interesting feature of Table 1 is the fact that the value of *k* selected as the best chosen value for most series using any of the above three methods is the same as that value of *k* selected by Nelson and Plosser (1982), recalling that the value of *k* according to Nelson and Plosser is chosen based on the sample autocorrelations of the series. Moreover, as shown by Newbold *et. al.* (1993), for the U.S. unemployment series for which the null hypothesis of a difference stationary process is rejected, one can obtain stronger evidence against the unit root hypothesis, regardless of the value of *k* if the *ADF* test is applied to the series without the time trend component.

The conclusion that all series except for the U.S. unemployment series are generated by difference stationary processes can also be sustained by two other factors. First by the sample autocorrelations of these series and second by the magnitude of the estimate of the moving average parameter of all series, if ARIMA(0,1,1) models are estimated using maximum likelihood estimation, which indeed is small and negative. The former suggests that these series need to be examined in first differences and the later indicates, that the performance of the unit root test is not going to be affected because of the presence of a substantial moving average component. However, none of all the above findings can witness with certainly that the value of the unit autoregressive root is one. In fact Agiakloglou and Newbold (1996) discuss the issue of the trade-off between size distortions and power loss when the *ADF* test is applied to processes like AR(1) and ARIMA(0,1,1) regardless of the method used to select the order of the approximating regression.

Finally, a recent paper by Leybourne and Newbold (1999) encountered that the non-parametric Phillips-Perron test has serious problems of oversized tests even when the true values of the short and long run variances are used in place of the sample estimates even for the simple ARIMA(0,1,1) model. The authors also indicated that the t-ratio variant of the test performs rather more poorly than the implementable version of the Dickey-Fuller test and this is the reason why the Phillips and Perron (1989) test for a unit autoregressive root was not considered for this study.

On the other hand, consider the case of testing the null hypothesis of a trend stationary process H_{TS} against the alternative hypothesis of a difference stationary process H_{DS}. An interesting study by Kwiatkowski *et. al.* (1992) has introduced a testing procedure for testing the above null hypothesis, known as the *KPSS* test. Kwiatkowski *et. al.* (1992) applied their test to the Nelson and Plosser series and they found that for most U.S. series the null hypothesis of a trend stationary process cannot be rejected.

A modification of the *KPSS* test is the Leybourne and McCabe (1994) test which can be viewed as an analogue to the *ADF* test whereas the former can be view as an analogue to the Phillips and Perron test. However, stronger evidence against the null hypothesis of trend stationarity one can obtain by applying the proposed by Arellano and Pantula (1995) test for testing H_{TS} against H_{DS}. This test is implemented by simply regressing the residuals $\hat{\varepsilon}_t$ on 1, t, X_{t-1}, ... , X_{t-p} and $\hat{W}_t = \sum_{j=p}^{t-1} \hat{\varepsilon}_j$ in order to obtain the following two test statistics, $\eta\Delta\hat{\beta}_{\tau,p}$ and $\hat{t}_{\tau,p}$ where $\Delta\hat{\beta}_{\tau,p}$ is the coefficient of $\hat{W}_t$ and $\hat{t}_{\tau,p}$ is the corresponding *t*-statistic for testing the null hypothesis that the coefficient of $\hat{W}_t$ is zero. The order p is selected according to the *AIC* information criterion by regressing X_t on 1, t, X_{t-1}, ... , X_{t-p} for all values of p=0,1,2,...,10 using the same sample size. Thus, as long as the order is selected, the residuals $\hat{\varepsilon}_t$ are obtained by applying the above regression using the right sample size. Critical values for this test can be obtained from Arellano and Pantula (1995) and the results are reported on Table 1. Unfortunately as Table 1 shows for all U.S. series except for the consumer price series the null hypothesis of trend stationarity cannot be rejected.

Lastly, we also need to keep in mind Perron's (1989) analysis under which the unit root hypothesis for most of the series was rejected if a structural change is incorporated to the unit root test imposed to the series. Recall that Perron (1989) has presented a unit root test based on the "augmented" Dickey-Fuller regression model (3) augmented by a time trend term and a set of dummy variables to allow for a deterministic change at a given point of time. Apart from all problems that may typically arise in terms of applying this unit root test, a comment concerning the choice of the break point is strongly needed to be made. As shown by Newbold and Agiakloglou (1992) using as an illustrative example the U.S. common stock prices, this methodology is very sensitive to the choice of the break point. In fact, if the choice of the break point is not *a priori* taken as given but it is allowed to be selected by any usual criteria, then for the best fitted model the hypothesis of a unit autoregressive root could not be rejected at the 10% level.

4. Maximum Likelihood Estimation

The next step is to examine the Nelson and Plosser (1982) data set using the conventional ARIMA analysis. The objective in this case is to determine whether an observed time series is generated by a *TS* process or by a *DS* process based on the estimation of models (1) and (2) and the main advantage of this approach is the fact that the outcome of this analysis is going to be unaffected of any testing procedure.

For this purpose *TS* models of equation (1) and *ARIMA(p,l,q)* models of equation (2) are fitted to each series for all possible combinations of ρ + q 5 where the first observation of the undifferenced series is deleted to ensure comparability. Furthermore, this study also includes Perron's (1989) analysis using the same base year, i.e., $T_B=1929$, to examine whether an observed time series can be characterized as a stationary *(PS)* process according to Perron's methodology. Thus, *PS* models of equation (1) augmented by the appropriate set of dummy variables defined by Perron (1989) are fitted to all U.S. series, except for the unemployment, without the first observation for all possible combinations of p + q 5. In all three cases, estimation was done using SAS program through maximum likelihood estimation and the best-fitted model is selected according to the SBC information criterion. The results are reported on Table 2.

Perhaps, the most astonishing feature of Table 2 is the fact that all series, except for the U.S. unemployment series, are generated as difference stationary processes if the best-fitted model is selected between *TS* and *DS* models. The point estimates along with their standard errors of these *ARlMA(p,l,q)* models are reported on Table 3 and as it can be seen they are all statistical significant except for some cases in which the mean of the process is not. If on the other hand Perron's analysis is incorporated, then for only two series, velocity and industrial production, the model selected as the best fitted model is a *PS* model and for these two series the *DS* model is simply a random walk model with drift.

Thus, for five series, nominal GNP, real GNP, real per capita GNP, money stock and GNP deflator, the best-fitted model is an ARIMA(1,1,0) model. For other four series, wages, employment, common stock prices and consumer prices, the best-fitted model is an ARIMA(0,1,1) model and for the bond yield series is an ARIMA(2,1,0) model. For the last three series, real wages, velocity and industrial production, the best fitted model among

TS and *DS* models is a random walk model with drift, although for the velocity series the estimate of the drift is not statistical significant. In addition, for two of these three series the best-fitted model according to the SBC criterion is generated as a *PS* process and although all estimates are statistical significant the estimate of the AR(1) parameter is 0,961 and 0,734 for the velocity and for the industrial production series respectively.

Finally, for the U.S. unemployment series the estimate of the coefficient of the time trend component of the *TS* model, which is selected as the best fitted model, is statistical insignificant meaning that the series needs to be examined without trend. Furthermore, the best fitted model among all *DS* models, the ARIMA(1,1,2) model, has a unit moving average root indicating the case of over-differencing. Fitting now ARIMA(p,0,q) models for all possible combinations of $p + q$ 5 we find that the best-fitted model is an ARIMA(1,0,1) model. This model is very close to the ARIMA(1,1,2) model but it does not contain a unit autoregressive root. However as discussed in Newbold *et. al.* (1993) the unit moving average root could very well be a manifestation of the 'pile-up' effect and whether we have near or exact cancellation it is hard to say. The truth is that we select the ARIMA(1,0,1) model simply because we like simple models.

5. Conclusion

So far it has been demonstrated that it is very difficult to declare with certainty the generating behavior of each series based on the existing testing procedures. It is even more difficult to state at this stage whether the data are not sufficiently informative or the performance of each test is not reliable and to what extent because the sample size is small. The truth is that the outcome based on a statistical approach is very sensitive to the applied testing procedure.

Using on the other hand the conventional ARIMA analysis, one can find the answer to this particular dilemma of model selection more easily. Of course, *SBC* cannot decisively determine whether or not each series contains a unit root neither can successfully distinguish two similar models. *SBC* is used only for model selection and it is preferred from *AIC* because it has the tendency to select small models. Hence, applying this methodology to Nelson and Plosser data set this study finds that all series, except for the unemployment series, are more likely to be generated as difference stationary processes rather that anything else.

References

Agiakloglou, C. and Newbold, P. (1996). The balance between size and power in Dickey-Fuller tests with data-dependent rules for the choice of truncation lag. *Economic Letters,* 52, 229-234.

Agiakloglou, C. and Newbold, P. (1992). Empirical Evidence on Dickey — Fuller — type tests. *Journal of Time Series Analysis,* 13, 471-483.

Arellano, C. and Pantula, S. G. (1995). Testing for trend stationarity versus difference stationarity. *Journal of Time Series Analysis,* 16, 147-164.

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

Dickey, D. A. and Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica,* 49, 1057-1072.

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

Hall, A. (1994). Testing for a unit root in time series with pretest data based model selection. *Journal of Business and Economic Statistics,* 12, 461-470.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative a unit root. *Journal of Econometrics,* 54, 159-178.

Leybourne, S. and McCabe, B. P. M. (1994). A consistent test for a unit root. *Journal of Business and Economics Statistics,* 14, 435-446.

Leybourne, S. and Newbold, P. (1999). On the size properties of Phillips-Perron tests. *Journal of Time Series Analysis,* 20, 51-61.

Nelson, C. R. and Plosser, C. I. (1982). Trends and random walks in macroeconomic time series. *Journal of Monetary Economics,* 10, 139-162.

Newbold, P. and Agiakloglou, C. (1992). U.S. Common stock process, 1871-1970: Playing with dummies. *Journal of Quantitative Finance and Accounting, 2,* 215-220.

Newbold, P., Agiakloglou, C. and Miller, J. (1993). Long-term inference based on short-term forecasting models. *In Developments in Time Series Analysis, In Honor of M.B. Priestley,* edited by T.S. Rao. London: Chapman & Hall.

Ng, S. and Perron, P. (1995). Unit root tests in ARMA models with data-dependent method for the selection of the truncation lag. *Journal of the American Statistical Association,* 90, 268-281.

Perron, P. (1989). The great crash, the oil price shock and the unit root hypothesis. *Econometrica,* 57, 1361-1401.

Phillips, P. C. B. and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika,* 75, 335-346.

Said, S. E. and Dickey, D. A. (1984). Testing for unit roots in autoregressive — moving average models of unknown order. *Biometrika,* 71, 599-607.

Schwert, G. W. (1989). Tests for unit roots: A monte carlo investigation. *Journal of Business and Economic Statistics,* 7, 147-160.

TABLE 1

Testing Hypotheses

	H_{DS} against H_{TS}			H_{TS} against H_{DS}
	Criterion	Order	Φ_3	$\eta\Delta\hat{\beta}_{\tau,p}$ / $\hat{t}_{\tau,p}$
Nominal GNP	ALL	2	2.77	-12.45 / -2.57
Real GNP	ALL	2	4.62	-11.80 / -2.37
Real per capita GNP	ALL	2	4.76	-12.63 / -2.47
Bond yield	ALL-	1	4.45	-18.83 / -2.98
	N&P	3	1.61	
Wages	AIC, SBC	2	3.30	-11.02 / -2.32
	N&P	3	2.65	
	T(5)	7	3.47	
Real Wages	ALL	2	4.95	-9.66 / -2.12
Employment	ALL-	2	4.96	-17.47 / -3.12
	N&P	3	3.55	
Unemployment	SBC	2	7.69*b*	-12.24 / -2.46
	ALL-	4	6.38*c*	
Money Stock	ALL	2	4.74	-17.85 / -2.99
GNP deflator	ALL	2	3.52	-11.14 / -2.37
Common stock prices	ALL-	2	3.98	-8.25 / -2.02
	N&P	3	2.86	
Velocity	ALL	1	2.93	-16.10 / -2.99
Industrial production	AIC, SBC	1	4.78	-10.36 / -2.25
	T(5), N&P	6	3.40	
Consumer prices	ALL-	3	2.13	-32.85*a* / -3.62*b*
	N&P	4	4.12	

Notes: 1) Order=k+1 according to Nelson and Plosser notation where k=0,1,...,10.
2) ALL denotes the order chosen by all criteria and ALL- denotes the order chosen by most criteria except the reported ones.
3) N&P denotes the order chosen by Nelson and Plosser.
4) Letters *a, b* and c denote statistical significance at the 1%, 5% and 10% level respectively.
5) Critical values for the Φ_3 test statistic are obtained from Dickey and Fuller (1981).
6) Critical values for the $\eta D \beta_{\tau,\rho}$ and $t_{\tau,\rho}$ test statistics are obtained from Arellano and Pantula (1995).

TABLE 2

Maximum Likelihood Estimation
Model selection according to SBC criterion

	TS	*DS*	*PS*
Nominal GNP	(2,0,0)	(1,1,0)@,*	(1,0,1)
Real GNP	(2,0,0)	(1,1,0)@,*	(2,0,0)
Real per capita GNP	(2,0,0)	(1,1,0)@,*	(2,0,0)
Bond yield	(3,0,0)	(2,1,0)@,*	(3,0,0)
Wages	(2,0,0)	(0,1,1)@,*	(2,0,0)
Real Wages	(2,0,0)	(0,1,0)@,*	(2,0,1)
Employment	(1,0,1)	(0,1,1)@,*	(1,0,1)
Unemployment	(1,0,1)@	(1,1,2)	
Money Stock	(2,0,0)	(1,1,0)@,*	(2,0,0)
GNP deflator	(2,0,0)	(1,1,0)@,*	(2,0,0)
Common stock prices	(1,0,1)	(0,1,1)@,*	(1,0,1)
Velocity	(1,0,0)	(0,1,0)@,*	(1,0,0)*
Industrial production	(1,0,0)	(0,1,0)@,*	(1,0,0)*
Consumer prices	(1,0,1)	(0,1,1)@,*	(1,0,1)

Note: Symbol @ denotes best fitted model selected between *TS* and *DT* models and symbol * denotes best fitted model selected among *TS, DT* and *PS* models.

TABLE 3

Estimates of the best-fitted models selected as *DS* models

	Mean	AR(1)	AR(2)	MA(1)
Nominal GNP	0.055	0.432		
	(0.020)	(0.117)		
Real GNP	0.030	0.336		
	(0.012)	(0.123)		
Real per capita GNP	0.016	0.326		
	(0.012)	(0.123)		
Bond yield	0.080	0.173	0.355	
	(0.071)	(0.124)	(0.128)	
Wages	0.040			-0.469
	(0.011)			(0,107)
Real Wages	0.018			
	(0.004)			
Employment	0.016			-0.391
	(0.006)			(0.105)
Money Stock	0.058	0.616		
	(0.014)	(0.088)		
GNP deflator	0.020	0.432		
	(0.009)	(0.101)		
Common stock prices	0.029			-0.310
	(0.021)			(0.097)
Velocity	-0.012			
	(0.007)			
Industrial production	0.043			
	(0.010)			
Consumer prices	0.013			-0.686
	(0.007)			(0.072)

Note: Figures in parentheses are the reported standard errors.