

# Unit roots in the Nelson-Plosser data: do they matter for forecasting?

Philip Hans Franses<sup>a,\*</sup>, Frank Kleibergen<sup>b</sup>,

<sup>a</sup>Department of Econometrics, Erasmus University, Rotterdam, The Netherlands

<sup>b</sup>Department of Finance, Erasmus University, Rotterdam, The Netherlands

## Abstract

In this paper we compare two univariate time series models, i.e. one with and one without an imposed unit root, in a forecasting experiment for the fourteen annually observed US data analyzed by Nelson and Plosser (1982, *Journal of Monetary Economics* 10, 139–162). Our main result is that the unit root model is regularly preferred. This result holds for a variety of sample sizes and forecast horizons as well as for one-step and multi-step ahead forecasts.

**Keywords:** Unit roots; Forecasting

## 1. Introduction and summary

Nelson and Plosser (1982) investigate fourteen annual US macroeconomic time series for the presence of a unit root using the Dickey–Fuller (1979) method. If a univariate time series contains a unit root, it should be analyzed after transforming it by taking first differences. Such a time series is called a difference-stationary (DS) process. The trend in a DS process with a drift term is partly deterministic and partly stochastic. If a time series does not have a unit root, and it can be described using a time series model which includes a purely deterministic trend, it is called

a trend-stationary (TS) process. One of the main reasons to investigate unit roots in univariate time series is that DS processes assume that shocks have a permanent effect, while such shocks only have a transitory effect for TS processes. Furthermore, given this permanent effect of shocks for the DS process, the multi-step ahead forecast intervals of the DS process are much wider than those of the TS process, especially for long horizons.

Nelson and Plosser (1982) find that only one of the fourteen time series they investigate, i.e. the Unemployment Rate, is not a DS process, that there is weak evidence that Real GNP is a TS process, and hence that most time series analyzed can be described by a DS process. Since this seminal study, many researchers have re-analyzed the Nelson and Plosser (1982) data set, possibly in an extended form. The main

\* Corresponding author at: Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands. Tel. +3110 4081273; fax: +3110 4527746.

motivation for most additional studies is that the Dickey–Fuller test can have low power in small samples. Another drawback of this unit root test may be that it assumes the approximate adequacy of an autoregressive (AR) time series model. A limited selection of relevant studies includes Schotman and van Dijk (1991) where a Bayesian approach is considered, Rudebusch (1992) where a small-sample correction of the original Dickey–Fuller test is given, Lucas (1995) where outlying observations are taken into account, and Perron (1989) where one known structural break is allowed to be present in the time series. As expected, the conclusion of these studies usually is that more than one of the fourteen time series can be described using the TS model, although the results are mixed across the various approaches.

Despite the large focus on testing for unit roots in the Nelson–Plosser data in the econometric literature, the impact of assuming the adequacy of the TS or the DS process on out-of-sample forecasting is not analyzed in the above studies. In this short paper we seek to fill this gap by explicitly analyzing the relative forecasting performance of both models. Our evaluation method is very straightforward. We estimate DS and TS models for various sample sizes and we generate one-step and multi-step ahead forecasts for various horizons. We calculate the mean squared prediction errors (MSPE) and mean absolute percentage errors (MAPE) to compare the relative quality of the forecasts. Non-parametric sign and Wilcoxon signed-rank tests are used to check whether differences between these measures are statistically significant. Note that we do not calculate forecast intervals since the intervals of the DS process in general may be wider than those of the TS process.

The main conclusion of this paper is that, in cases where the differences between the point forecasts are significant, the DS forecasts outperform those of the TS model. Moreover, the DS model produces better forecasts for more time series when the forecast horizon increases. This conclusion holds for one-step and multi-step ahead forecasts.

The outline of this short paper is as follows. In Section 2, we review in more detail the DS and

TS models and highlight some of their specific properties. In Section 3, we discuss the forecast evaluation method and the results we obtain which lead to the overall conclusions. In Section 4, we conclude the paper with some remarks on the practical impact of our findings.

## 2. The models

Consider an annually observed time series  $y_t$ , where  $t = 1, \dots, n$ , and denote  $\varepsilon_t$  as a zero-mean uncorrelated process with constant variance. In this case  $y_t$  can be described by

$$y_t = \delta + \tau t + \beta_1 y_{t-1} + \dots + \beta_{p+1} y_{t-p-1} + \varepsilon_t, \quad (1)$$

i.e. an AR( $p+1$ ) process, and when the characteristic roots of this AR polynomial lie outside the unit circle, it is said that  $y_t$  can be described by a trend-stationary (TS) process. If, however, one of the roots of the AR( $p+1$ ) polynomial is equal to one, the model in (1) can be written as

$$\Delta_1 y_t = \mu + \alpha_1 \Delta_1 y_{t-1} + \dots + \alpha_p \Delta_1 y_{t-p} + \varepsilon_t, \quad (2)$$

where  $\Delta_1$  is defined by  $\Delta_1 y_t = y_t - y_{t-1}$ . Model (2) is called a difference-stationary (DS) process. One may extend (1) and (2) by allowing  $\varepsilon_t$  to be some moving average (MA) process. However, in practice this extension is usually not pursued, also since the standard Dickey–Fuller test only considers pure AR processes. In fact, it is assumed that any MA-type dynamics can be mopped up by increasing the order of the AR polynomial.

To highlight the main differences between (1) and (2), consider the DS process

$$y_t = y_{t-1} + \mu + \varepsilon_t, \quad (3)$$

which is a simple random walk with drift. Via recursively substituting lagged  $y_t$  in (3), it is easy to derive that

$$y_t = y_0 + \mu t + \sum_{i=1}^t \varepsilon_i, \quad (4)$$

where  $y_0$  is the starting-value of  $y_t$ . The expression in (4) indicates that a DS model contains a

deterministic trend component  $\mu t$  and a stochastic trend component  $\sum_{i=1}^t \varepsilon_i$ . Furthermore, the impact of a shock  $\varepsilon_i$  does not die out as  $t$  increases, and hence shocks have a permanent effect on  $y_t$ . It is clear from (1) that this does not apply to the TS process, where the shocks only have a transitory impact. Finally, the models in (1) and (4) indicate that forecasts for the long-run will be dominated by the trend terms. This suggests that the modification of the TS and DS models to include MA terms may yield only small differences in long-run forecasting compared to the models with only AR terms.

### 3. Forecasting

In this section we describe our procedure to evaluate the forecasting performance of the TS and DS models. Note that we abstain from making statements about which model is best within the sample and that we use real-life and not simulated data.

For our forecasting evaluation we consider the fourteen Nelson–Plosser time series, including such variables as Real GNP, Stock Prices, Real Money, and the Unemployment Rate. The original Nelson–Plosser data covered only years until 1970. In this paper we use the data set extended until 1988, which is given in Schotman and van Dijk (1991). The number of annual observations for the various macroeconomic time series now ranges from about 80 to about 120.

The order of the AR polynomial in (1) and (2) is an important issue in practice. In this paper we consider the same model orders as those displayed in Rudebusch (1992, Tables 2 and 3). These model orders were obtained for the sample until 1970. When we experimented with the extended data set, we did not find significant differences between the AR order for the extended and the original data set. Furthermore, some casual investigation of the forecasts from AR models with higher orders did not yield significantly different results either. Finally, the AR order tends to be reasonably constant throughout the entire sample, i.e. in subsamples one may at most want to include a single additional lag. For the extended sample, the esti-

mated models all pass conventional LM-type diagnostic checks for residual autocorrelation. This implies that the models apparently do not display the patterns described in Chan, Hayya and Ord (1977) and Pierce (1975), i.e. there are no obvious indications that the time series are overdifferenced in the DS model or underdifferenced in the TS model. In contrast to the model order, the parameters are not assumed to be constant over all subsamples, and hence (if needed) we re-estimate each model using the relevant subsample observations. Unreported estimated parameter values suggest, however, that usually the hypothesis of parameter constancy cannot be rejected.

Our empirical forecast evaluation method is as follows. The models (1) and (2) are estimated using the observations until 1976, 1970 and 1952. The first sample is chosen in order to capture the oil crisis in 1973 in the estimation period. The second sample corresponds to the Nelson–Plosser data set, and the third sample is chosen such that one may assume that the World War II effects have become negligibly small. Unreported results for other samples indicate that the forthcoming results are fairly robust with respect to the choice of the sample. Based on the estimated models, one-step ahead and multi-step ahead forecasts are generated (using the conventional methods as given in time series textbooks) for 1977–1988, 1971–1988 and 1953–1988, resulting in 12, 18 and 36 pairs of forecast errors to evaluate. For each of the models we calculate the MSPE and MAPE, and the differences between the ratios for the TS and DS models. For the one-step ahead forecasts we use the sign test and the Wilcoxon signed-rank test to investigate whether these differences are significant, see Flores (1989) for an exposition of the second test. Even though the number of forecasts may be small, we use the normal approximation formulas for these two nonparametric test statistics. Our experience is that this approximation does not yield results very different from those reported below. For the multi-step ahead forecasts we do not calculate the nonparametric test statistics since the forecast errors are correlated.

In Table 1, we report the results for the one-step ahead forecasts. Each cell of Table 1 con-

Table 1  
One-step ahead forecasts

Variable	Forecast horizon					
	12		18		36	
	MSPE	MAPE	MSPE	MAPE	MSPE	MAPE
Real GNP	–	–	–	–	+ <sup>a</sup>	+ <sup>a</sup>
Nominal GNP	+ <sup>a,b</sup>	+ <sup>a,b</sup>	+ <sup>a,b</sup>	+ <sup>a,b</sup>	+ <sup>a,b</sup>	+ <sup>a,b</sup>
Real GNP per capita	–	–	–	–	+	+
Industrial Production	+	–	–	–	–	–
Employment	+	+	+	+	+	–
GNP deflator	+ <sup>a,b</sup>	+ <sup>a,b</sup>	+ <sup>a,b</sup>	+ <sup>a,b</sup>	+ <sup>a</sup>	+ <sup>a</sup>
Consumer Prices	+ <sup>a,b</sup>	+ <sup>a,b</sup>	+ <sup>a,b</sup>	+ <sup>a,b</sup>	+ <sup>a,b</sup>	+ <sup>a,b</sup>
Wages	+ <sup>a,b</sup>	+ <sup>a,b</sup>	+ <sup>a,b</sup>	+ <sup>a,b</sup>	+ <sup>a,b</sup>	+ <sup>a,b</sup>
Real Wages	+ <sup>a,b</sup>	+ <sup>a,b</sup>	+ <sup>a,b</sup>	+ <sup>a,b</sup>	+ <sup>b</sup>	+ <sup>b</sup>
Money Stock	+ <sup>a</sup>	+ <sup>a</sup>	+ <sup>a,b</sup>	+ <sup>a,b</sup>	+ <sup>a</sup>	+ <sup>a</sup>
Velocity	–	–	+ <sup>a,b</sup>	+ <sup>a,b</sup>	+ <sup>a,b</sup>	+ <sup>a,b</sup>
Bond Yield	– <sup>b</sup>	– <sup>b</sup>	+	–	+ <sup>a,b</sup>	+ <sup>a,b</sup>
Stock Prices	+	+	+	+	+ <sup>a,b</sup>	+ <sup>a,b</sup>
Unemployment	+	–	+	–	–	–

Notes: <sup>a</sup> The differences between the criteria are significantly different at a 5% level using the nonparametric sign-test.

<sup>b</sup> The differences between the criteria are significantly different at a 5% level using the Wilcoxon signed-rank test, see Flores (1989).

tains either a “+” or a “–”, where the first means that the MSPE or MAPE for the TS model exceeds that for the DS model. Hence a “+” means that the DS model outperforms the TS model in forecasting. Furthermore, we indicate whether this outperformance is significant using the sign test (<sup>a</sup>) and the Wilcoxon signed-rank test (<sup>b</sup>). For the one-step ahead forecasts we observe that in 22 of the 42 cases the DS model significantly outperforms the TS model (where significance here means that either the sign or the signed-rank test indicates significance), that in only one case the TS model significantly outperforms the DS model (Bond Yield for 12 forecasts), and that there are no significant differences for the other 19 cases. Furthermore, the differences between DS and TS increase when the forecast horizon increases, i.e. for 12 periods ahead the DS model is better for 6 of the 14 variables while for 36 periods ahead this applies to 10 variables.

In Table 2, the outcomes for the multi-step forecasts are displayed. When one generates forecasts for 1 through 12 periods ahead, based on a model estimated until 1976, the DS model will outperform the TS model in 8 of the 14 cases, while the TS model seems better in the

case of Consumer Prices and Velocity. For 18 forecasts ahead the DS model is better 8 times, while the TS model outperforms the DS model 4 times. Comparable results emerge for 36 forecasts. Hence, for the 42 cases it implies that DS is better 26 times while the TS is better 10 times. Note that these results should be interpreted with care, since we cannot substantiate the results with test statistics which are significant at a 5% significance level.

We can conclude from the results in Tables 1 and 2 that when the differences between the forecasts from the DS and TS models are significant, the DS model usually outperforms the TS model in forecasting the fourteen annual macro-economic series for the US economy. This result seems to be consistent across sample sizes, forecast horizons and one- and multi-step forecasts.

Finally, in Table 3 we report the results of constructing one-step ahead forecasts from recursive regressions. For this purpose we estimate the models until 1952, generate a forecast for 1953, re-estimate the model until 1953, generate a forecast for 1954, and so on. This results in 36 forecasts which are based on re-estimated regression models. This recursive analysis allows for

Table 2  
Multi-step ahead forecasts

Variable	Forecast horizon					
	12		18		36	
	MSPE	MAPE	MSPE	MAPE	MSPE	MAPE
Real GNP	+	+	+	–	+	+
Nominal GNP	+	+	+	+	+	+
Real GNP per capita	+	+	+	–	+	+
Industrial Production	–	–	–	–	–	–
Employment	+	+	+	+	+	+
GNP deflator	+	+	+	+	+	+
Consumer Prices	–	–	–	–	–	–
Wages	+	+	+	+	+	+
Real Wages	+	+	+	+	+	+
Money Stock	+	+	–	–	–	–
Velocity	–	–	–	–	+	+
Bond Yield	+	–	+	–	+	+
Stock Prices	+	+	+	+	+	+
Unemployment	+	+	–	–	–	–

parameter nonconstancy, if there is any. Since we consider one-step ahead forecasts, we may again use the nonparametric test statistics as in Table 1. The results in Table 3 indicate that for 6 of the 14 series, the DS model significantly outperforms the TS model, while only for one case the reverse holds. For 7 series there are no significant differences. In sum, it seems that the

results in Table 3 broadly correspond to those obtained in Tables 1 and 2.

#### 4. Remarks

Obviously, the results in the previous section do not automatically imply that a DS process is the “true” underlying data generating process for the fourteen time series under consideration. Neither do they imply that the DS process is the most likely model, nor that the same results will emerge for other time series variables for the US economy or for similar variables for other economies.

In fact, it seems most appropriate to interpret the outcomes in terms of structural or temporal breaks in the forecast evaluation period, i.e. not in the period for which the model is estimated but in the out-of-sample period. Hence, these breaks cannot be considered in the estimation process since these are unknown beforehand. This is most easily understood as follows. A change in the trend parameter in the post-estimation sample of a true TS process can yield one-step ahead forecast errors for a TS process which are larger than those from an “incorrect” DS model. And, the reverse result does *not* hold. Hence, a DS process is able to adapt more

Table 3  
One-step ahead forecasts from rolling regressions

Variable	MSPE	MAPE
Real GNP	+	+
Nominal GNP	+ <sup>a,b</sup>	+ <sup>a,b</sup>
Real GNP per capita	–	–
Industrial Production	–	–
Employment	+	+
GNP deflator	+ <sup>a,b</sup>	+ <sup>a,b</sup>
Consumer Prices	– <sup>a,b</sup>	– <sup>a,b</sup>
Wages	+ <sup>a,b</sup>	+ <sup>a,b</sup>
Real Wages	+ <sup>a,b</sup>	+ <sup>a,b</sup>
Money Stock	–	–
Velocity	+ <sup>a,b</sup>	+ <sup>a,b</sup>
Bond Yield	+	+
Stock Prices	+ <sup>a,b</sup>	+ <sup>a,b</sup>
Unemployment	–	–

Note: <sup>a</sup> The differences between the criteria are significantly different at a 5% level using the nonparametric sign-test.

<sup>b</sup> The differences between the criteria are significantly different at a 5% level using the Wilcoxon signed-rank test, see Flores (1989).

rapidly to structural parameter changes, at least for one-step ahead forecasts. Notice that this corresponds to the findings in Table 3. Given that the forecast periods considered in the present paper incorporate possible break-points, it may not be surprising that the DS model seems to outperform the TS model.

Of course, when a TS model is the DGP and there are no breaks, forecasts from a TS model outperform those of the incorrect DS model, at least theoretically. Therefore, it seems worthwhile to investigate the presence of structural breaks within the sample while selecting between a DS and a TS process, and also to investigate whether the economic process at hand is susceptible to structural shifts once in a while. Recent developments in this area are Perron and Vogelsang (1992) and Zivot and Andrews (1992). Naturally, it remains uncertain whether such breaks will indeed also occur in the forecasting sample. It seems then of interest to analyze time series forecasting models which are able to adapt to changes in parameters or even to changes in the model structure.

In the meantime, the practitioner who wants to select between a DS and a TS model may add the out-of-sample forecasting performance to the set of selection criteria. Indeed, confidence in a certain model is gained when it persistently outperforms its rival model over a range of forecast horizons. Furthermore, even though the standard and regularly used Dickey–Fuller method does not have much power in small samples, this method often indicates that the DS model is appropriate within sample. Given that our results suggest that the same model is often useful for out-of-sample forecasting, the Dickey–Fuller test remains worthwhile to consider in practice.

### Acknowledgements

The first author thanks the Royal Netherlands Academy of Arts and Sciences for its financial support. Some parts of this paper were presented by the first author at the 12th International Symposium on Forecasting in Wellington, New Zealand, August 1992. We are indebted to the

editor, an associate editor and three anonymous referees for helpful comments. The data were made available to us by Herman van Dijk.

### References

- Chan, K.H., J.C. Hayya and J.K. Ord, 1977, A note on trend removal methods: The case of polynomial regression versus variate differencing, *Econometrica*, 45, 737–744.
- Dickey, D.A. and W.A. Fuller, 1979, Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association*, 74, 427–431.
- Flores, B.E., 1989, The utilization of the Wilcoxon test to compare forecasting methods: A note, *International Journal of Forecasting*, 5, 529–535.
- Lucas, A., 1995, An outlier robust unit root test with application to the extended Nelson–Plosser data, *Journal of Econometrics*, 66, 153–173.
- Nelson, C.R. and C.I. Plosser, 1982, Trends and random walks in macroeconomic time series: Some evidence and implications, *Journal of Monetary Economics*, 10, 139–162.
- Perron, P., 1989, The great crash, the oil price shock, and the unit root hypothesis, *Econometrica*, 57, 1361–1401.
- Perron, P. and T.J. Vogelsang, 1992, Nonstationarity and level shifts with an application to purchasing power parity, *Journal of Business and Economic Statistics*, 10, 301–320.
- Pierce, D.A., 1975, On trend and autocorrelation, *Communications in Statistics*, 4, 163–175.
- Rudebusch, G.R., 1992, Trends and random walks in macroeconomic time series: A re-examination, *International Economic Review*, 33, 661–680.
- Schotman, P.C. and H.K. van Dijk, 1991, On Bayesian routes to unit roots, *Journal of Applied Econometrics*, 6, 387–401.
- Zivot, E. and D.W.K. Andrews, 1992, Further evidence on the great crash, the oil price shock and the unit root hypothesis, *Journal of Business and Economic Statistics*, 10, 251–270.

**Biographies:** Philip Hans FRANCES is affiliated with the Econometric Institute of the Erasmus University Rotterdam as a fellow of the Royal Netherlands Academy of Arts and Sciences. One of his research interests is model selection in time series. On this topic he has published in such journals as *Econometric Reviews*, *Journal of Econometrics* and *International Journal of Forecasting*.

Frank KLEIBERGEN holds a research position at the Department of Finance of the Erasmus University Rotterdam. His research focuses on cointegration and Bayesian methods as well as on financial models. On these topics he has published in *Econometric Theory*, *Journal of Econometrics* and the *Journal of Applied Econometrics*.