Impact of Model Specification Decisions on Unit Root Tests

Atiq-ur-Rehman, Atiq-ur-Rehman and Zaman, Asad

IIIE, International Islamic University, Islamabad

2009

Impact of Model Specification Decisions on Unit Root Tests

Atiq-ur-Rehman

IIIE, International Islamic University, Islamabad

Asad Zaman[ϒ]

IIIE, International Islamic University, Islamabad

Abstract:

Performance of unit tests depends on several specification decisions prior to their application e.g., whether or not to include a deterministic trend. Since there is no standard procedure for making such decisions, therefore the practitioners routinely make several arbitrary specification decisions. In Monte Carlo studies, the design of DGP supports these decisions, but for real data, such specification decisions are often unjustifiable and sometimes incompatible with data. We argue that the problems posed by choice of initial specification are quite complex and the existing voluminous literature on this issue treats only certain superficial aspects of this choice. We also show how these initial specifications affect the performance of unit root tests and argue that Monte Carlo studies should include these preliminary decisions to arrive at a better yardstick for evaluating such tests.

Keywords: model specification, trend stationary, difference stationary

JEL Classification: C01, C15, C22

[ϒ] Corresponding Author: Atiq-ur-Rehman, Room No. F-11, Kuwait Hostel, Inetranational Islamic University, Islamabad, Pakistan, Email: ateeqmzd@yahoo.com Contact No: +92-334-5336263

We are thankful to participants of FEMES 2008, Singapore for comments on earlier version of this paper.

1. Model Specification and Reliability of Unit Root Tests

Perhaps the issue discussed most in the history of econometric literature is the debate on trend versus difference stationarity, initiated by Nelson and Plossor (1982). Ignoring stationarity can lead to spurious results and wrong asymptotics for traditional econometric techniques. This has led to a huge amount of research in the past 25 years, but consensus on several important issues and implications has not emerged to date (Libanio, 2005). Even though vast numbers of unit root tests have been proposed and studied, conflicting opinions exist on the simplest of problems. For example, here is a list of the conclusions of authors who have studied the USA annual GNP series:

Difference stationary; Nelson and Plossor (1982),

Trend Stationary; Perron (1989),

Trend Stationary; Zivot and Andrews (1992),

Don't know; Rudebusch (1993),

Trend stationary; Diebold and Senhadji (1996),

Difference stationary; Murray and Nelson (2002), Kilian and Ohanian (2002),

Trend stationary; Papell and Prodan (2003)

There may be several reasons for variety of opinions on dynamics of same series, but one important reason is specification of model before application of unit root tests. Performance of unit root tests depends on several specification decisions prior to application of unit root test: e.g., whether or not to include a deterministic trend and how to choose the order of the included lags in the model. Although there is immense literature on the issue, there is no standard procedure for making such decisions. Therefore practitioners routinely make several arbitrary specification decisions either implicitly or explicitly. Because most Monte Carlo studies take these initial decisions as valid background information, such studies often overestimate the performance of tests on real data. In Monte Carlo, when the experiments condition on some implicit specification, the design of data generating process supports the implicit assumptions. But for the real data series, implicit assumptions/arbitrary specification decisions are often unjustifiable and sometimes incompatible with data. In this paper we show how these initial specifications affect the performance of unit root tests and argue that Monte Carlo studies should include these preliminary decisions to arrive at a better yardstick for evaluating such tests.

While the importance of the initial specification has been noted from the very beginning, and the issue has been subjected to a lot of research, no consensus has emerged on how to choose the initial specification. One feature of this literature is that, researcher focus on some particular decision(s) conditioning on the validity of some other implicit decision(s). Therefore their analysis might lose its validity when data violates these implicit assumption/decisions. For example, consider the study of Perron (1989, 1990), which initiated the debate on structural stability combined with unit root tests. Perron's analysis is a re-specification of models used by Nelson and Plossor (1982). Perron designs a test with better performance for models with structural breaks, but the analysis is conditional on implicit assumption of homoskedasticity. For his Monte Carlo, the design of experiment supports this implicit specification, but for real data, we know nothing about validity of this assumption. Perron's analysis may lose its validity when this assumption is violated. This conjecture is supported by studies of Kim et al. (2002), Cavaliere (2005) etc.

We will argue that the problems posed by choice of initial specification are quite complex, and the existing voluminous literature on this issue treats only certain superficial aspects of this choice. Specifically, we consider four specification decisions i.e. (i) choice of deterministic part (ii) choice of lag length (iii) distribution of innovations and (iv) structural stability. To understand the complex relationship between these specification decisions and output of unit root tests, following facts regarding these decisions are important to realize. First fact is Relevance of Specification Decisions, i.e. the decisions discussed have significant impact on final output of unit root tests. Section 2 is dedicated to for this discussion. Second fact is related to Validity and compatibility of Specification Decisions i.e. conventional implicit/arbitrary specifications and probabilistic assumptions are often invalid/incompatible with data. Section 3 gives brief discussion on this issue. Third fact is related to Interdependence of Specification Decisions i.e. these decisions are mutually dependent. The endogenized decision of one specification depends on other specification decisions and therefore a powerful criterion for choice of one decision may not perform well because of wrong conditioning on some other decision. This issue is addressed in Section 4 of our study.

We provide evidences from literature, real life examples and artificially generated data, to explain complex relationship between these decisions and output of unit root tests. Than in Section 5, we summarize the study and discuss implication of these facts in measuring performance of unit root tests.

2. Relevance

The four specification decisions mentioned in previous section have significant impact on final output of unit root tests

Prior to the application of unit root test, the investigator has to make number of specification decisions. Since very beginning of debate on unit root, one can trace the recognition of significance of proper specification in unit root testing e.g. Dickey and Fuller (1979) presented different tests for unit root with three specifications of deterministic part in model, i.e. unit root without drift and linear trend, with drift and with drift plus linear trend. Afterward the significance of several other types specifications were discussed in the literature and their impact on output of unit root tests was analyzed e.g. the specification of autoregressive lags (Dickey and Fuller, 1981), distribution of innovation (Said and Dickey, 1984), presence of structural breaks (Perron, 1989, 1990) etc. As we mentioned above, we are focusing specifically on four specification decisions and for these decisions mentioned above, we will briefly discuss existing evidences of enormous impact of decision on output of unit root test and will provide other evidences if necessary.

2.1 Specification of deterministic part

The deterministic part in a model to be tested for unit root may be any deterministic function of time index of the stochastic process under investigation. However, in literature, drift and linear trend are often discussed instead of more general form of deterministic trend. We focus here on the same, i.e. drift and linear trend.

The significance of specification of trend and drift on output of unit root tests is recognized since very beginning of research on unit root tests. In routine practice, three types of models are used while testing for unit root and these models are:

M1	Without drift, trend	$y_t = \delta \; y_{t-1} + \varepsilon_t$
M2	With drift, but no trend	$y_t = \alpha + \delta \; y_{t-1} + \varepsilon_t$ (1)
M3	With drift and trend	$y_t = \alpha + \beta \, t + \delta \; y_{t-1} + \varepsilon_t$

Where $\varepsilon_t \sim IIDN(0, \sigma^2)$

The distribution of test statistics used for testing unit root hypothesis depends on (i) model that generated the data i.e. DGP and also (ii) the model used for testing unit root hypothesis. Dickey and Fuller (1979) and MacKinnon (1991) provide critical values for unit root test when any one of these three models is used for testing unit root hypothesis. Computation of these critical values assumes a match between DGP and the model used for testing unit root, but in practical situation, we do not know about true form of DGP and the distribution of test statistics varies a lot when there is disagreement between (i) and (ii). Hamilton (1994) provides a discussion of distribution in such a situation of disagreement. Hamilton's analysis gives us the idea of distribution of Dickey-Fuller test statistics in miss-specified models, but it does not provide a guideline about how to choose model for testing the hypothesis when we have no idea of true DGP.

In fact the literature on 'how to choose deterministic part' is much smaller than other decisions of similar nature like choice of lag length. Since Model M1 & M2 are nested in M3, a researcher may use General to Simple Type Strategy to specify model prior to application of unit root test, but as we will show later in this section, General to Simple Strategy is not feasible for the problem under consideration. There are some recommendations in literature for specification of deterministic trend by Perron (1988), Holden & Perman (1994), Ayat and Burridge (2000), Elder and Kennedy (2001) and Enders (2004). These recommendations formulate two types of strategies; first is sequential testing strategy advocated by Enders (2004). The other strategy is utilization of priori information on growth properties of underlying time series and is advocated by Elder and Kennedy (2001). However surprisingly there is lack of studies on comparison of the strategies for specification of deterministic component. A relevant unpublished study is due to Hacker and Hatemi-J (2006), comparing Ender's strategy to Elder and Kennedy's strategy with the conclusion that the later strategy is superior. However the study is restricted to Dickey-Fuller environment. A feasible strategy for choice of deterministic part working better for more recent tests like Ng-Perron's test is still to be explored.

We now demonstrate the practical significance of specification of deterministic component by various examples. First example uses real data; we apply set of Ng and Perron tests to US real GNP data from 1909 to 1970 obtained from Nelson and Plossor's data set. Logic behind choice of Ng-Perron (2001) test is that, it accumulates intellectual heritage of number of previous tests e.g. Elliot et al. (1996)'s point optimal test etc. The results of two specifications of deterministic components are reported in Table 1. When working without linear trend, the unit root hypothesis is nowhere rejected and when working with linear trend, all four tests reject unit root at 10% significance level. The reason for such result is perhaps, Ng-Perron's emphasis is to adjust unit root tests for specification of lag length, so they designed a test optimal in choosing lag length, but the test thus designed is more sensitive to specification of deterministic part. Researcher is again facing similar problem-how to choose deterministic part? No satisfactory answer to this question is there in literature.

We further illustrate the significance of problem by a simple Monte Carlo experiment. We generated time series of length 100 using model M1 described in (1) as data generating process with $\delta = 0.85$ and these artificial series were tested for unit root in all three scenarios described by M1:M3. The results of experiment are reported in column 1 of Table 2. Remember that true DGP is trend stationary ($\delta < 1$). This process was repeated using M2 and M3 as data generating process and results are reported in column 2 and 3 of Table 2 respectively. It is obvious from column 1 of Table 2 that maximum power is attained when testing scenario matches with true DGP and unit root hypothesis is rejected 97% of times. A mismatch of two models leads false acceptance of unit root hypothesis more often. Similar conclusion can be drawn form column 2 & 3 of the Table. The crucial thing to be noted is that, if a researcher starts from ignorance, she may reach any conclusion she wants by switching the choice of deterministic trend. Thus reliability of inference is at stake.

Recently developed tests like Ng-Perron test are not different from classical tests in this regard. We present one more example to illustrate significance of specification of deterministic trend for Ng-Perron test. We generated a time series of length 62 (equal to length of US real GNP series used by Nelson and Plossor) with following DGP:

$$y_t = 0.2 + 0.8 y_{t-1} + u_t$$
$$u_t \sim N(0,1) \qquad (2)$$

Coefficient of lag term is 0.8, much below unity thus generated time series is trend stationary. We apply unit root tests to this series in two scenarios; (i) drift without linear trend (ii) linear trend plus drift. The results of testing are reported in Table 3. For the same series, in one setup, all four members of set of Ng-Perron test reject null of unit root at 5% nominal significance level, while in other setup all four members are unable to reject unit root. We did not selected series for this type of output; this output is for the first data series generated by Microsoft Excel and three more consecutive attempts yield similar results. When we apply test in appropriately specified scenario, the test result provide us right message about dynamics of series and unit root is rejected at 5% significance level. But when the scenario is misspecified, set of Ng-Perron test fails to reach the right conclusion. For the artificial data series, we know what the true data generating process is, and we found that if pre-test scenario is miss-specified, unit root test fail to detect true dynamic structure. But for the real time series, we do not know what scenario is generating data, so we cannot guess whether or not we are getting right message from the test. However, one thing that we observe is that, a researcher may derive any conclusion she wants, playing with different choices of deterministic trend.

Above discussion is sufficient to argue that, the choice of deterministic part is worth considering and we should not blindly stake on some particular choice of deterministic trend. Furthermore, in a situation where we have nested models, it is natural to start with general to simple type modelling to specify the model. But unfortunately, to specify deterministic component in a model to be tested for unit root/stationary, this strategy fails to work. The evidence is due to study of Nelson and Kang (1984) who generated 1000 simple random walks ($y_t = y_{t-1} + \varepsilon_t$) with 100 observations in each. These series were than regressed on a linear time trend. They found that the (true) null hypothesis that coefficient of trend is zero was rejected in 87% of the regressions at 5% nominal level of significance. This is surprising because we know that there is actually no predictable relationship between time path of a random walk and the linear trend. This is very clear evidence for the fact that, to test presence of linear trend, the distribution of usual t-statistics is not similar to student's t-distribution when series is generated by a random walk. Another evidence is due study of Hacker and Hatemi (2006) who investigate performance of sequential testing strategy which is similar to general to simple strategy. They conclude that sequential testing strategy does not provide a solution to problem of choice of deterministic trend in the problem of unit root testing.

So, testing for presence of unit root is a circular testing problem, to test whether or not there is unit root depends on information about presence of linear trend in the model whereas, to test presence of linear trend in a model depends on the information about stationarity of time series. Given the importance of choice deterministic trend while testing for unit root, we need deep investigation of a procedure that would give us optimal solution to this circular testing problem.

2.2 Specification of lag length

There is a consensus on the view that inappropriate choice of autoregressive lag leads to undesirable properties of unit root tests. Therefore the choice of autoregressive lag length got due attention of econometricians. Dickey and Fuller (1981) ‘augmented’ their test to incorporate autoregressive specification in the model. Said and Dickey (1984) study impact of moving average root on unit root tests and they suggest including sufficient lags in autoregressive specification in augmented Dickey Fuller test which can approximate any moving average process. However, the appropriate choice of lag length remains an important question for econometricians. Several criteria have been recommended in literature for choosing appropriate lag length e.g. AIC, BIC, SIC SBC etc. Ng and Perron (2001) summarize literature on this issue. Ng and Perron’s study reveals that modified Akaike information criterion (MAIC) outperform other procedures for choice of appropriate lag length. To this point, the goal specification of lag length looks to be achieved but when combined with other specifications, it looks over-optimistic to draw such a conclusions. Recall set of four specification decisions we are studying, as we will argue later in this paper, there are evidences that wrong decision about one specification may result in failure of procedure for specification of another decision. We will discuss this issue in greater detail in later sections.

2.3 Structural breaks

The debate of structural breaks in macroeconomic time series has been a major area in unit root research. Perron (1989) suggested that Nelson and Plossor’s strong evidence in support of the unit root hypothesis rested on a failure to account for structural change in the data, and demonstrated this through incorporating an exogenous structural break for the 1929 crash. In doing so he reversed the Nelson-Plossor (1982) conclusions for 10 of the 13 series. Perron's study can be regarded as an attempt to respecify the model of Nelson and Plossor; however his method for incorporating structural breaks is based on knowledge of historical events and is not a data based respecification. At the beginning of the 1990s, Banerjee *et al*. (1992), Christiano (1992) and Zivot and Andrews (1992) argued that selecting the structural break *a priori* based on an *ex post* examination or knowledge of the data could lead to an over-rejection of the unit root hypothesis. To address this issue, these studies incorporated a single endogenous structural break. Endogenizing structural breaks, Zivot and Andrews (1992) were unable to reject unit root for four series that Perron concluded to be stationary. This debate continues to date and many methods for endogenizing, selecting and testing structural breaks have been developed and analyzed so far. Recent survey of literature on this issue is provided by Perron (2006). We do not feel necessity to present more evidences for loss of desirable properties of unit root tests due to misspecification of structural break, because literature contains so many evidences and there is no disagreement on the view that decision of structural stability is crucial in final output of unit root test. However, there was controversary on testing and incorporating techniques of structural breaks in a model.

2.4 The distributional assumptions

Among the distributional assumption, the most common assumption is normality of innovation process. However many authors note that violation of this assumption does not have serious impact on unit root tests. Another assumption is that the innovations should be serially independent. Several authors have documented that if innovations are serially dependent i.e. innovation creates moving average roots, and the moving average roots can be approximated by sufficient number of autoregressive lags. Hence this problem is not serious because we have well documented procedures for selection of autoregressive lags. A rejection of the normality assumption could be due to some other factors, in particular due to outliers. In that case, it is also well documented that the presence of outliers induces a strong finite sample bias towards not rejecting the unit root too often. This is because outliers produce large moving average roots. This problem can be handled by appropriate choice of lag length. For detail of discussion on these assumptions see Perron (2003) and references cited there.

However another distributional assumption is homoskedasticity of innovation process. Recent studies reveal that non-constant variances can both inflate and deflate the rejection frequency of the commonly used unit root tests, both under the null and under the alternative. Kim et al (2002) shown that change in the innovation variance of an integrated process can generate spurious rejections of the unit root null hypothesis in routine applications of Dickey Fuller tests. They develop and investigate modified test statistics, based on unit root tests of Perron (1989) which are applicable when there is a change in innovation variance of an unknown magnitude at an unknown location. Cavaliere (2004) show that non-constant variances can both inflate and deflate the rejection frequency of the unit root tests, with early negative and late positive variance changes having the strongest impact.

So far, we presented some evidences from literature showing that heteroskedasticity effect the output of unit root test. Here are some Monte Carlo evidences of relationship between validity of assumption of homoskedasticity and output of unit root tests. Consider two data generating processes:

DGP1	$y_t = y_{t-1} + \varepsilon_t$	$\varepsilon_t \sim IIDN(0, \sigma^2)$	Homoskedastic
DGP2	$y_t = y_{t-1} + \varepsilon_t$	$\varepsilon_t \sim N(0, (1+kt)\sigma^2)$	Heteroskedastic

The first DGP is homoskedastic random walk process, whereas second is heteroskedastic process. We generate time series of length 100 by two processes and apply Dickey-Fuller test (without drift and trend) to both processes. For the homoskedastic process, there is no problem with size of unit root test, since it is standard random walk process and Dickey-Fuller test works fine for such process. But for second DGP, for different values 'k' following rejection of unit root was observed.

Value of k	Rejection of unit root
0.05	8%
0.10	9%
0.15	10%
0.20	11%

From this simple experiment, it is evident that violation of assumption of homoskedasticity leads to over-rejection of unit root. Therefore validity of this assumption must be taken into account when analyzing dynamics of a real time series. For further detail, reader is referred to Kim et al (2002) and Cavaliere (2004).

3. Validity and compatibility

Conventional implicit/arbitrary specifications and probabilistic assumptions are often invalid/incompatible with data

Until now we have presented evidences that model assumption/prior specification plays vital role in determining final output of unit root tests. Next we discuss validity of such assumption/specifications in routine applications. It is important to note that like other scientific theories; we are unable to prove some particular assumption/specification decision however we can often find evidences to disprove some decision using misspecification testing, and if so, than there is internal evidence of incompatibility of model under consideration with data.

The job of misspecification testing is highly technical compared to other hypothesis testing situations like Neyman Pearson testing (see Andreou and Spanos, 2003), however sometimes there are straight forward evidences against validity of some assumptions. We are not going to present here evidences of misspecification of lag length and structural stability because size of literature on these issues is self speaking evidence of ambiguity in decision making. If for some particular time series, decisions about choice of lag length prior to application of unit root test are different in two studies, obviously both decisions can not be true, this simple argument makes us realize how many number of times we make wrong decision in choosing lag length. However we present here evidences of violation assumptions about distribution of innovation in the models used for unit root testing. In Fig

1, we plot residuals from regression equation of Nelson and Plossor (1982) for US real GNP. Variation in innovation variation is obvious from this figure.

Andreou and Spanos (2003) revisit models of Nelson and Plossor (1982) and Perron (1989) for verification of various probabilistic assumptions. We report their finding about assumption of homoskedasticity. Their study reveal that homoskedasticity assumption is rejected for 5 out of 14 series in Nelson-Plossor's study and for 8 out of 14 in Perron's study. Pagan and Schwert (1990) and Loretan and Phillips (1994) provide evidences against constancy of variances in empirical exchange rate and international stock market data. Watson (1999) shows change in variance of US long/short term interest rates. Many more evidences of non constant variances of macroeconomic time series can be seen in Cavalier (2004) and references cited there.

4. Interdependence

The specification decisions are mutually dependent; therefore a powerful criterion for choice of one decision may not perform well because of wrong conditioning on some other decision

We will present just one evidence to show interdependence of specification decisions i.e. the interdependence of choice of lag length and deterministic trend. Ng and Perron (2001) devise Modified Information Criterion (MIC) for the choice of lag length in autoregressive time series and they provide simulation evidences for the nice properties of this criterion. We use same criterion to choose lag length for series in Nelson Plossor data set, for three different models M1:M3 discussed in section 2. The results of estimation are reported in table 4.

Except for Nominal GNP and GNP deflator, different models lead to different choice of autoregressive of lag length. Therefore we conclude that, appropriate choice of lag length depends on choice of deterministic trend.

5. Summary and Conclusion

Now we summarize the discussion presented so far in the study. In section 1 of the study, we discuss the controversial issue of long run dynamic of econometric time series. Despite lot of professional interest in this issue, the problem of approximating long run dynamics of econometric time series is still unresolved. One reason of controversy over the issue is choice of various specification decisions prior to application of unit root tests. We argue that Monte Carlo studies analyzing performance of unit root tests are invalid for real data because these are conditional on some specification decisions supported by data generating process but real data set does not necessarily support such specification decisions. We discuss that the problem of model specification prior to application of unit root tests is more complex than it is treated in usual Monte Carlo studies. This is due to interdependence of specification decisions on each other. We choose four specification decisions i.e. choice of deterministic part, choice of lag length, structural stability and distribution of innovations to illustrate this complex relationship between specification of model and output of unit root tests. Section 2 consists of evidences that these decisions really affect the output of unit root tests. Second and third specification decision have lot of literature in their credit which is self speaking evidence of impact of these decisions on unit root tests. But decision of deterministic trend and assumption of homoskedasticity attracted little professional interest than they deserve. We present number of evidences of enormous impact of first and fourth specification decision on unit root tests. In section 3, we discuss validity of conventional implicit/explicit choices of four types of decisions in real data sets. We claim that implicit/explicit choices of these four decisions are often incompatible with real econometric time series. This means that serious attention is needed to verify validity of specification of model to be tested for unit root. Section 4 presents evidence that choice of such specification decisions is interdependent and a powerful criterion may fail to work because of wrong conditioning on some other decision.

This study has very important theoretical and practical implications. It is obvious from the study that there is little if any resemblance between Monte Carlo and real application of unit root tests. This gap may be reduced by paying deeper attention to the implicit specification decisions. Focusing on any single decision would not solve the problem, rather multiple specification decisions should all be considered simultaneously, because all these decisions effect output of unit root tests and there validity is often questionable for real data sets. Furthermore any powerful criterion for choice of one decision would not provide reliable results unless we are confident about validity of other implicit specification. A better yardstick for measuring performance of unit root tests would essentially treat all multiple specification decisions simultaneously.

References:

Andreou, E. and A. Spanos, 2003, Statistical Adequacy and the Testing of Trend versus Difference Stationarity, Econometric Reviews 223, 217 – 237.

Ayat, L. and P. Burridge, (2000), Unit root tests in the presence of uncertainty about the nonstochastic trend, Journal of Econometrics 95, 71-96.

Banerjee, A., R. Lumsdaine and J. Stock, 1992. Recursive and sequential tests of unit-root and the trend break hypotheses: theory and international evidence. Journal of Business Economics and Statistics, 10, 271-287.

Campbell, J., and P. Perron, 1991, Pitfalls and opportunities: what macroeconomists should know about unit roots? NBER Macroeconomics Annual 6, 141–201.

Cati, R., M. Garcia and P. Perron, 1999, Unit roots in the presence of abrupt governmental interventions with an application to Brazilian data. Journal of Applied Econometrics 14, 27–56.

Cavaliere, G., 2005, Unit Root Tests under Time-Varying Variances. Econometric Reviews, 23, 259 – 292.

Christiano, L., 1992, Searching for a break in GNP. Journal of Business Economics and Statistics, 10, 237-250.

Cocharane., H. 1987., "A Critique of the Application of Unit Root Tests," mimeo, University of Chicago.

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

Dickey, D. and W. Fuller, 1981. Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root, Econometrica, 49, 1057-1072.

Diebold, F. and A. Senhadji, 1996. The Uncertain Root in Real GNP: Comment. American Economic Review, 86, 1291-1298.

Elder J. and P. Kennedy, 2001, Testing for Unit Roots: What Should Students Be Taught? Journal of Economic Education, 322, 137-46.

Elliott, G., T. Rothenberg, and J. Stock, J. H. (1996), Efficient Tests for an Autoregressive Unit Root, Econometrica 64, 813–836.

Enders, W., 2004, Applied Econometric Time Series, Second Edition. John Wiley & Sons, United States.

Hacker. R. and . A. Hatemi-J, 2006, The Properties of Procedures Dealing with Uncertainty about Intercept and Deterministic Trend in Unit Root Testing, unpublished manuscript

Hamilton J., 1994, Time Series Analysis. Princeton University Press, Princeton, New Jersey.

Holden, D. and Perman, R. (1994) “Unit roots and cointegration for the economist.” In B. B. Rao, (Ed) Cointegration for the Applied Economist, New York, pp. 47-112

Kilian, L. and L. Ohanian, 2002, Unit Roots, Trend Breaks and Transitory Dynamics: A Macroeconomic Perspective. Macroeconomic Dynamics, 6, 614-631.

Kim, T., S. Leybourne and P. Newbold, 2002. Unit root tests with a break in innovation variance. Journal of Econometrics, 109, 365–387.

Libanio, G. 2005, Unit roots in macroeconomic time series: theory, implications and evidence. nova Economia Belo Horizonte15, 3145-176

Loretan, M. and P. Phillips, 1994, Testing covariance stationarity under moment condition failure with an application to common stock returns. Journal of Empirical Finance 1, 211–248

Lumsdaine, R. and D. Papell, 1997. Multiple trend breaks and the unit root hypothesis. Review of Economics and Statistics, 79, 212-218.

MacKinnon, J., (1991), Critical Values for Cointegration Tests, in: Engle, R. and Granger, C. (Eds), Long-run Economic Relationships, Readings in Cointegration, Oxford University Press, pp 267-276.

McConnell, M. and P. Quiros, 2000. Output fluctuations in the United States, what has changed since the early 1980s? American Economic Review. 90, 1464–1476.

Murray, C. and C. Nelson, 2000. The uncertain trend in U.S. GDP. Journal of Monetary Economics 46, 79–95.

Nelson, C.R & Kang, H., 1984. 'Pitfalls in the Use of Time as an Explanatory Variable in Regression', Journal of Business & Economic Statistics, American Statistical Association, vol. 2(1), pages 73-82, .

Nelson & Kang, Journal of Business & Economic Statistics, 1984, pp 73-82

Nelson, C. R. and C. I. Plossor, 1982, Trend sand random walks in macroeconomics time series: some evidence and implications. Journal of Monetary Economics 10, 139–162.

Ng, S. and P. Perron, 2001, Lag length selection and the construction of unit root tests with good size and power. Econometrica 69, 1519–1554.

Nunes, L., C. Kuan and P. Newbold 1997. Spurious regression. Econometric Theory, 11, 736-49.

Pagan, A. R., Schwert, G. W. (1990). Testing for covariance stationarity in stock market data. Economics letters 33, 165–170.

Papell, D. H. and R. Prodan, 2003. Restricted Structural Change and the Unit Root Hypothesis. Working paper, University of Houston

Perron, P., 1988, Trends and random walks in macroeconomic time series. Journal of Economic Dynamics and Control, 12 (12): 297-332.

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

Perron, P., 1990, Testing for a unit root in a time series with a changing mean. Journal of Business & Economic Statistics (8) 153-162.

Perron, P., 2003. Comment on "Statistical Adequacy and the Testing of Trend Versus Difference Stationarity" by Andreou and Spanos (Number 1., Econometric Reviews, 22, 3, 239-245

Perron, P. 2006, Dealing with Structural Breaks, Mimeo forthcoming in the Handbook of Econometrics, Econometric Theory.

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

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

Spanos A. and Anya McGuirk, 2006. Revisiting the foundations of unit root testing: why the AR(1). model does not nest the unit root, Virginia Tech working paper series

van Dijk, D., D.R. Osborn, and M. Sensier, 2002. Changes in variability of the business cycle in the G7 countries, Erasmus University Rotterdam, Econometric Institute Report EI 2002-28.

Watson, M. W. 1999. Explaining the increased variability in long-term interest rates. Richmond Economic Quarterly, 85, 71–96.

Zivot, E. and D. 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.

Table 1. Output of Ng-Perron test applied to US real GNP

Null Hypothesis: LUSGNP_R has a unit root

Sample:1909-1970

Exogenous		Constant				Constant + trend			
		MZa	MZt	MSB	MPT	MZa	MZt	MSB	MPT
Ng-Perron test statistics*		**1.34**	**1.01**	**0.75**	**45.64**	**-15.5**	**-2.771**	**0.177**	**5.967**
Asymptotic Critical Values**	1%	-13.80	-2.58	0.17	1.78	-23.80	-3.42	0.14	4.03
	5%	-8.10	-1.98	0.23	3.17	-17.30	-2.91	0.16	5.48
	10%	-5.70	-1.62	0.27	4.45	-14.20	-2.62	0.18	6.67

*Automatic Lag Length Selection Procedure: Spectral GLS-detrended AR based on MAIC, MAXLAG=10

**Ng-Perron (2001, Table 1)

Table 2: Percentage Rejection of Null Hypothesis of Unit Root Using Dickey Fuller Unit Root Test

		True DGP		
		M1	M2	M3
Model used for testing	M1	97	46	0
	M2	62	70	0
	M3	38	39	41

Table 3. Ng-Perron test applied to artificial data									
Null Hypothesis: X has a unit root									
Exogenous		Constant				Constant + trend			
		MZa	MZt	MSB	MPT	MZa	MZt	MSB	MPT
Ng-Perron test statistics*		**-10.55**	**-2.29**	**0.21**	**2.33**	**-12.04**	**-2.44**	**0.20**	**7.62**
Asymptotic Critical Values**	1%	-13.80	-2.58	0.17	1.78	-23.80	-3.42	0.14	4.03
	5%	-8.10	-1.98	0.23	3.17	-17.30	-2.91	0.16	5.48
	10%	-5.70	-1.62	0.27	4.45	-14.20	-2.62	0.18	6.67
*Automatic Lag Length Selection Procedure: Spectral GLS-detrended AR based on MAIC, MAXLAG=10									
**Ng-Perron (2001, Table 1)									

Table 4 : Choice of autoregressive lag for Nelson Plossor data

	None	Intercept	Drift + trend
REAL GNP	1	1	0
NOM GNP	1	1	1
GNP DEFL	1	1	1
PER CAP RGNP	1	1	0
INDUST PROD	1	5	0
EMPLOYMENT	1	2	0
UNEMPLOY	4	2	2
CPI	3	3	2

WAGES	1	2	2
REAL WAGE	1	1	0
MONEY	1	2	2
VELOCITY	1	1	5
BOND YIELD	2	2	0
S&P500	1	2	5

Fig 1. Residuals from regression equation of Nelson and Plossor for US real GNP