Model specification, observational equivalence and performance of unit root tests

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

IIIE, International Institute of Islamic Economics

July 2008

MODEL SPECIFICATION, OBSERVATIONAL EQUIVALENCE AND PERFORMANCE OF UNIT ROOT TESTS

Atiq-ur-Rehman[*⊥]
E-mail: ateeqmzd@yahoo.com
Asad Zaman[ϒ]
E-mail: asadzaman@alum.mit.edu

Abstract: In this paper we highlight the necessity of new criteria for evaluation of performance of unit root tests. We suggest focusing directly on the reasons that create ambiguity in unit root test's results. Two reasons for unsatisfactory properties of unit root tests can be found in the literature (i) the model misspecification and (ii) observational equivalence. Regarding first reason, there is immense literature on several components of model specification covering specification techniques, consequence of misspecification and robust methods. However complete model specification involves multiple decisions and most of studies on performance of unit root tests do not address issue of multiple specification decisions simultaneously. The Monte Carlo studies are conditional on some of implicit specification and for Monte Carlo; these specifications are by construction valid. But for real data, the implicit decisions are often not true and specification decisions need to be endogenized. A closer match with real case is possible if multiple specification decisions are endogenized, thus providing more reliable measure of performance of unit root tests. Second problem in differentiating trend and difference stationary process is the observational equivalence between two processes. We suggest exploring data generating processes with different long run dynamics and small sample equivalence so that a researcher should have an idea about other plausible models for a data set for which he has estimated some model.

Keywords: Observational equivalence, model specification, trend stationary, difference stationary

JEL Classification: C01, C15, C22

[*] Correspondence Address: Room No. F-11, Kuwait Hostel, International Islamic University, Islamabad, Pakistan, Contact No: +92-334-5336263
[⊥] PhD Scholar (Econometrics), International Institute of Islamic Economics, International Islamic University, Islamabad
[ϒ] Professor, International Institute of Islamic Economics, International Islamic University, Islamabad

1. Introduction

Perhaps the issue discussed most in the history of econometric literature is the debate on trend versus difference stationarity, initiated by Nelson and Plosser (1982). Due to highly important implication of information about stationarity in the empirical econometric modeling, this issue got attention of hundreds of econometricians during the past quarter century, but consensus on several important issues and implications seems not to appear to date (Libanio, 2005). Huge stock of unit root tests emerged during past quarter century, but the certainty about inference can not be achieved. Consider a typical example of US real GNP; following are views on stationarity of this series presented in different papers:

Difference stationary; Nelson and Plosser (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).

Different views on stationarity of same series are result of adopting different procedures for testing unit root stationarity. So, there is ambiguity on what procedure of testing should be adopted and how much this procedure is reliable. We review the reasons that are responsible for ambiguity in the inference from a unit root test, and propose new procedure for evaluating the performance of unit root test.

Two important reasons that can be traced in literature for unsatisfactory performance of unit root tests are the model misspecification and the observational equivalence.

The performance of unit root tests depends on several specification decisions prior to application of unit root test. The Monte Carlo experiments for evaluating the performance of unit root tests often overestimate the performance of tests because the experiments condition on some implicit specifications and the design of data generating process supports the implicit assumptions, whereas for the real data series, implicit assumptions/arbitrary specification decisions are often unjustifiable and sometimes incompatible with data. To measure the performance of unit root tests in real data sets, the specification decisions should be endogenized.

The second reason for uncertainty in the inference from unit root tests is the problem of observational equivalence between trend and difference stationary process. The trend stationary and difference stationary process have quite different long run dynamics but may have similar small sample distribution, see Spanos and McGuirk (2006) for example. We propose to explore data generating process with different long run dynamics and small sample equivalence, so that an investigator should have an idea of other plausible models for a data series for which he have fitted some model.

Next sections of this paper are organized as follows: In section 2, we discuss nature and strength of the reasons that create ambiguity in the inference from unit root. In section 3, we summarize the discussion and recommend the strategy for more reliable inference.

2. Inference from unit root tests; the reasons of uncertainty

To achieve the goal of greater confidence in the inference from unit root tests, we propose to focus directly on the reasons that create ambiguity in the inference. Two major reasons that are discussed in the literature for undesirable properties of unit root tests are the *model misspecification* and the *observational equivalence*. Brief history of literature on these two reasons is given and their significance is highlighted by (i) evidences found in literature (ii) real life examples and (iii) artificially generated data. We highlight the deficiency in procedure for evaluating performance of unit root test and present our suggestion for improvement. Below we discuss the two reasons separately.

2.1 Model misspecification

Prior to the application of unit root test, the investigator has to make number of specification decisions which might be implicit. These specification decisions play significant role in the final output of a unit root test. Since very beginning of debate on unit root, one can trace the 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 a unit root 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 e.g. the specification of autoregressive lags (Dickey and Fuller, 1981), specification of distribution of innovation (Said and Dickey, 1984), presence of structural breaks (Perron, 1989, 1990) etc. However, putting these specifications altogether and study of the impact of multiple decisions has little literature in its credit. When we have a real data series, we do not have access to its true DGP. Before choosing a test for unit root, we have to make number of specification decisions which may be arbitrary, data based or based on previous information. This is the situation which is quite different from usual Monte Carlo experiment. In Monte Carlo experiments, even when studying the consequences of some endogenized specification decision, we are conditioning on several implicit specifications. For example, Perron (1989), Zivot and Andrew (1992)'s analysis of unit root tests in presence of structural break is conditional on the implicit assumption of homoskedasticity of the innovations. In Monte Carlo, the implicit specification, built in the assumptions of a model is by construction valid. However, when working with real data series, the validity of implicit assumptions should be justified. But there is lack of study on the properties of unit root tests when overall specification of model is data dependent. The importance of data based specification decisions is further enhanced due to number of evidences against validity of implicit specification e.g. evidences collected by Andreou and Spanos* (2003). The inference from a unit root test is ambiguous until we are confident that the model we specified is appropriate. And in our opinion, data dependent specification is the best way to ensure the appropriateness of the specification of model prior to application of unit root test.

It is important to note that no serious problem in properties in properties of unit root test arises if the model is properly specified. The size or power problem of unit root tests arises due to misspecification of testing scenario prior to application of unit root test. Even the second problem, i.e. observational equivalence also arises (at least some times) due to the misspecification of testing scenario e.g. Spanos and McGuirk (2006). We focus on four specification decisions and analyze their role in the inference from unit root tests. These four specification decisions are: (i) choice of deterministic part (ii) choice of autoregressive lags (iii) the structural breaks and (iv) distribution of innovation process. Following are the motivational factors behind choice of these four decisions we mentioned above:

1. Unit root tests are sensitive to each of four specification decisions mentioned above
2. None of existing study on performance of unit root tests focus on all these decisions simultaneously
3. There are evidences of violation of the conventional implicit specification in real data sets.

* In later sections we present more detail of this study

4. For the real data sets, except in few cases, the arbitrary *priori* specification is not justifiable. This necessitates the endogenous decisions about these specifications.

We present few evidences in favor of the statements given above. Let us discuss the four specification decisions separately.

2.1.1 The specification of deterministic part (drift and trend)

As stated earlier, the significance of specification of trend and drift is recognized since very beginning of research on unit root tests. Since then, a lot of literature emerged on various aspects of specification of trend and drift. Hamilton (1994) summarizes the distribution theory of unit root tests statistics under various specifications of trend and drift. However, unlike the lag length selection and the structural change, little literature is there addressing the issue that how to specify drift and trend while applying unit root test to real data sets. Some recommendations in this regard are provided by Perron (1988), Perman (1994), Ayat and Burridge (2000), Elder and Kennedy (2001) and Enders (2004). The two streams of these strategies can be identified from literature. First of these is sequential testing strategy latest advocated by Ender (2004). The other strategy is utilization of priori information on growth properties of underlying time series and is advocated by Elder and Kennedy (2001). Since the significance of appropriate specification of deterministic component is recognized at earliest, so one would expect valuable work on its relevant aspects. But there is lack of empirical studies on comparison of the strategies for specification of deterministic component. A relevant study is due to Hacker and Hatemi (2006), who compare 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 which is subject to serious critique by later authors. What strategy of specification of deterministic component works better in framework of recently recommended tests like Ng and Perron (2001) test; there is no proper guideline available on this question in the literature.

However, the practical significance of specification of deterministic component; is clear by following example. We apply set of Ng and Perron tests to US real GNP data from 1909 to 1970 obtained from Nelson and Plosser's data set. Logic behind choice of Ng-Perron test is that, it accumulates intellectual heritage of many previous test e.g. Elliot et al. (1996)'s point optimal test etc. The results of two specifications of deterministic components are summarized below.

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)									

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. Remember that for specification of lag length we are using MAIC according to recommendation of Ng and Perron (2001). 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 justification for some arbitrary choice of deterministic component is available in present setup.

One more example we will present by computer generated dataset. Since we know exact data generating process for computer generated data set, it is more suitable to know what is exactly happening. We generate a time series of length 62 (equal to length of US real GNP series used by Nelson and Plosser) with following DGP:

$$y_t = 0.2 + 0.8 y_{t-1} + u_t$$

$$u_t \sim N(0,1) \qquad (1)$$

Coefficient of lag term is 0.8, much below unity and therefore 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 result of estimation is as follows:

Table 2. Ng-Perron test applied to artificial data									
Null Hypothesis: X has a unit root									
Sample:1909-1970									
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)									

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. Remember that the true data generating process is without linear trend. When we apply test in appropriate scenario, the test result provide us right message about stationarity 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 misspecified, unit root test fail to detect right data generating process. But for the real time series, we do not know what scenario is generating data, so we cannot guess whether we are getting right message from the test. Now, if we make endogenous decision of inclusion of trend for both artificial and real time series, a closer resemblance is possible and whatever the result about performance of unit root test we get, will be more worthwhile.

2.1.2 Specification of lag length

There is a consensus on the view that inappropriate specification of autoregressive lag leads to undesirable properties of unit root tests. Therefore the choice of autoregressive lag length got due attention of econometricians (unlike deterministic component). Dickey and Fuller (1981) modify their test to incorporate autoregressive specification in the model. Their modified test is usually called augmented Dickey Fuller test. Said and Dickey (1984) study impact of moving average root on unit root tests. 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 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, different conclusions can be reached. Recall set of four specification decisions we are studying, there are evidences that wrong decision about one specification may result in failure of procedure for specification of another decision. For example, for US real GNP, in augmented DF setup, we analyze choice of lag by MAIC recommended by Ng and Perron. The results are summarized in table 3.

Table 3: Lag length selected for US real GNP in different scenarios

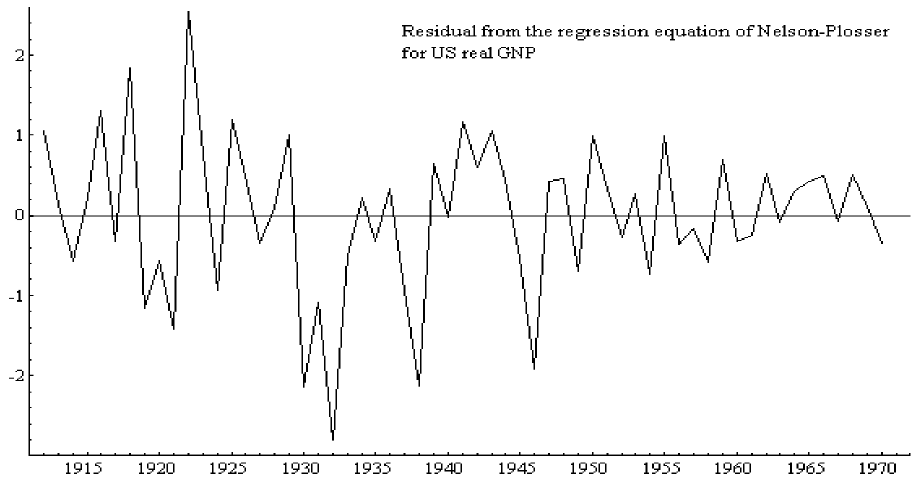
Scenario	Lag length chosen by using MAIC
Without drift and trend	1
With drift	1
With drift and trend	0

The results favor our supposition that misspecification of one member of set of specification discussed above may lead to wrong result in specification for some other member of the same set. Hence the resulting inference from unit root test gains smaller advantage of well designed procedure of selection of lag length.

2.1.3 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 Plosser'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-Plosser (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 is continuous 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 Byrne and Perman (2006) and Perron (2006). We do not feel necessity to present 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 considerable disagreement on the issue. However, there was controversary on testing and incorporating techniques of structural breaks in a model. This controversary lead to rapid growth in the literature. Now, there are well documented techniques available for endogenizing structural breaks and testing for unit root in presence of structural breaks however these techniques usually condition on other implicit specification. Again we emphasis that misspecification of one assumption may reveal itself in the specification procedure for the some other assumption. For example it is possible that apparent structural break in deterministic component is due to break in innovation variance. We plot residual from regression equation of Nelson and Plosser (1982) for US real GNP.

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



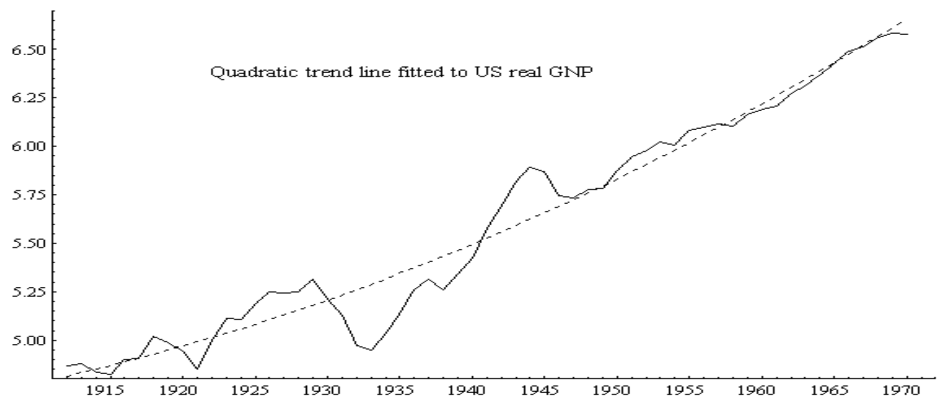
The break in variance of the residual is evident from the above graph. It might be possible that apparent structural break that investigators found in the series is a reflection of break in the innovation variance that we observe in the fig above.

2.1.4 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 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. This assumption needs further attention due to the fact noted by many authors including Pagan and Schwert (1990), Loretan and Phillips (1994), Watson (1999) McConnell and Quiros (2000) and van Dijk et al. (2002) who provide evidences of non-constant variances of many macroeconomic time series. 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. They show that that the locally best invariant (LBI) test of a unit root against level stationarity is robust to heteroskedasticity of any form under the null hypothesis. However, the conclusion of author is valid conditioning on the other implicit assumptions of the model. The cross dependence of this assumptions with violations of other assumptions have not been discussed so far. Below we plot a quadratic trend fitted to US real GNP. The evidence of variability of the variance is quite clear here.

Fig 2. Quadratic trend fitted to US real GNP



So, the problem that arises due to misspecification lead to ambiguous inference from unit root tests. If we summarize the evidences presented so far, following conclusions can be drawn:

1. The specification decisions we selected effect the output of unit root tests
2. The arbitrary/implicit specification in conventional methodology is often opposed by data, so the specification decisions need to be justified
3. Violation of one of assumption/implicit specification may reveal itself in wrong conclusion of some other specification procedure

2.2 Observational Equivalence

The second thing which creates ambiguity in the inference from unit root test is the problem of observational equivalence. Although the trend stationary and difference stationary process have quite different long run dynamics, in finite sample these may be observationally equivalent. Cochrane (1991) argues that in Beveridge-Nelson (1982) decomposition of a time series into transitory and permanent components, the random walk component may have arbitrarily small variance. In that case unit root tests may have arbitrarily low power in small samples. He also shows that there are some difference stationary processes whose likelihood is arbitrarily close to unit root process. Further investigation of observational equivalence between trend and difference stationary process is due to Blough (1992) and Faust (1993). Presenting reasons similar to Cocharane, Blough argues that none of unit root test can have high power for stationary process except having high probability of false rejection for the nearby member of unit root null. However the arguments of Blough does not decrease the importance of unit root testing, because if some component of data generating has arbitrarily small weight/variance, we do not bother to detect it because in the finite time horizon this component with small variance/weight will not create a larger influence. We still need to determine dynamics of dominant portion of data generating process.

Andreou and Spanos (2003) recommend using the 'statistical adequacy' as a guide to specify model while testing for unit root test. According to the authors, by statistical adequacy they mean empirical validity of the probabilistic assumptions associated with the model. The statistical adequacy provides sufficient condition for validity of the model in that there is no violation of the assumption of the model found in the data. They test models of Nelson and Plosser (1982) and Perron (1989) for statistical adequacy and found that most of their models are not adequate. After re-specification of the models, they found result different from both Nelson-Plosser and Perron.

Spanos and McGuirk (2006) note an important shortcoming of strategy of Andreou and Spanos (2003) by noting that two different types of models i.e. difference stationary and trend stationary, when fitted to same data set, may be both statistically adequate in that their respective probabilistic assumptions are not opposed by data. In particular they note if true data generating process is random walk with drift, a trend stationary model with drift plus trend may provide a statistically adequate approximation. The authors consider following three scenarios:

$$
\begin{array}{lll}
\text{S1} & y_t = \delta y_{t-1} + \varepsilon_t & \\
\text{S2} & y_t = \alpha + \delta y_{t-1} + \varepsilon_t & \varepsilon_t \sim IIDN(0,\sigma^2) \quad (2) \\
\text{S3} & y_t = \alpha + \beta t + \delta y_{t-1} + \varepsilon_t &
\end{array}
$$

The authors note that if data is generated by S2 with δ = 1, (random walk with drift) and fitted model belong to S3, one may get a trend stationary statistically adequate approximation to the data.

The study of Spanos and McGuirk can be the basis for new direction of the research in unit root. Following conclusion can be drawn from the study of Spanos and McGuirk:

1. For a real data series, there may be more than one statistically adequate models with quite opposite long run dynamics
2. At least in some cases, the observational equivalence can be observed due to misspecification of scenario that we use for testing unit root

Spanos and McGuirk present just one example of observational equivalence between models with different long run dynamics. Now it is plausible to assume that some other types of trend and difference stationary models may be observationally equivalent in that their respective probabilistic assumptions are not opposed by the data. If it is so, investigator needs to know what other types of models may be supported by the series for which he has fitted some particular model. We present here an example of observational equivalence between integrated ARIMA(0,1,1) model and stationary AR(1) model. We had a Monte Carlo experiment in which data was generated by following data generating process:

$$y_t = y_{t-1} + u_t$$

$$u_t = ku_{t-1} + e_t$$

$$e_t \sim N(0,1) \qquad (3)$$

Dickey Fuller unit root test was applied to the generated data series. The decision about inclusion of trend and drift was based on the significance of their estimated coefficient starting from general model and reducing successively. Final selection of model was verified for statistical adequacy by testing for the assumptions i.e. autocorrelation and homoskedasticity of residuals. We test for first, second and third order autocorrelation of residual to ensure no evidence of correlation is there since moving averages should reveal them in autocorrelation. The true data generating process is difference stationary. For a data series of length 100, following percentage of statistically adequate stationary model was obtained. 2000 simulations were carried out for each value of *k*.

Table 4: Percentage rejection of unit root null when true DGP is according to eq. (3)

Value of *k*	% of stationary models
-0.5	29
-0.6	37
-0.7	45
-0.8	62
-0.9	84

This is one more evidence of observational equivalence between trend and difference stationary. Similarly if we look at other types of data generating processes one can search the observationally equivalent process in the class with different long run dynamics. Knowledge of model with observational equivalence will provide a guide to investigator, about other plausible models for a data set for which he have found some statistically adequate model.

3. Measuring strength of reasons of uncertainty; a step forward to more reliable inference

The results and evidences presented so far to can be summarized into following:

1. Each specification decisions discussed in the study effect the output of unit root test

2. The arbitrary/implicit specification decisions are often incompatible or unjustifiable for the real data

3. Invalid decision on one specification may effect other specification decisions negatively

4. Even with data supported specifications, there may be more than one models with quite opposite long run dynamics but having compatibility with data

The results 1-3 motivate us to revisit criteria for evaluation of procedures for testing stationarity/unit root in real data sets. The performance of procedure with endogenized specification decisions has a closer resemblance with real data sets. So the more reliable measure of performance of unit root tests is possible when multiple specification decisions are endogenized. Whatever performance of a testing procedure is measured via this criterion would be more reliable and more realistic.

The result 4 motivate us to explore the models with different long run dynamics and small sample equivalence so that an investigator would be able to guess what other types of model may be plausible for a data set for which he has fitted some model.

Bibliography

[1] Andreou, E. and Spanos, A. (2003). Statistical Adequacy and the Testing of Trend Versus Difference Stationarity, Econometric Reviews, 22(3), 217 – 237.

[2] Banerjee, A., Lumsdaine, R. and Stock, J.H. (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.

[3] Byrne J. P. and Roger Perman (2006). Unit Roots and Structural Breaks, A Survey of the Literature

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

[5] Cati, R. C., Garcia, M. G. P., Perron, P. (1999). Unit roots in the presence of abrupt governmental interventions with an application to Brazilian data. Journal of Applied Econometrics 14, 27–56.

[6] Cavaliere, G. (2005). Unit Root Tests under Time-Varying Variances. Econometric Reviews, 23, 3, 259 – 292.

[7] Christiano, L.J. (1992). Searching for a break in GNP. Journal of Business Economics and Statistics, 10, 237-250.

[8] Cocharane., H. (1987)., "A Critique of the Application of Unit Root Tests," mimeo, University of Chicago.

[9] 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.

[10] Dickey, D.A. and Fuller, W.A. (1981). Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root, *Econometrica*, 49, 1057-1072.

[11] Diebold, F. X., and Senhadji, A. S. (1996). The Uncertain Root in Real GNP: Comment. *American Economic Review*, 86, 1291-1298.

[12] Elder J. and Kennedy P. E. (2001). Testing for Unit Roots: What Should Students Be Taught? Journal of Economic Education, 32(2), 137-46.

[13] Enders, W. (2004). Applied Econometric Time Series, Second Edition. John Wiley & Sons, United States.

[14] Hamilton J. D. (1994). Time Series Analysis. Princeton University Press, Princeton, New Jersey.

[15] Kilian, L. and Ohanian, L. (2002), Unit Roots, Trend Breaks and Transitory Dynamics: A Macroeconomic Perspective. *Macroeconomic Dynamics*, 6, 614-631.

[16] Kim, T. H., Leybourne, S., Newbold, P. (2002). Unit root tests with a break in innovation variance. Journal of Econometrics, 109, 365–387.

[17] Libanio, G. A. (2005). Unit roots in macroeconomic time series: theory, implications and evidence. nova Economia Belo Horizonte15 (3)145-176

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

[19] Lumsdaine, R.L. and Papell, D.H. (1997). Multiple trend breaks and the unit root hypothesis. Review of Economics and Statistics, 79, 212-218.

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

[21] Murray, C. J. and Nelson, C. R. (2000). The uncertain trend in U.S. GDP. Journal of Monetary Economics 46, 79–95.

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

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

[24] Nunes, L.C., C-M. Kuan and P. Newbold (1997). Spurious regression. Econometric Theory, 11, 736-49.

[25] Papell, D. H. and Prodan, R. (2003). Restricted Structural Change and the Unit Root Hypothesis. *Working paper, University of Houston*

[26] Perron P. (1988). Trends and Random Walks in Macroeconomic Time Series: Further Evidence from a New Approach. Journal of Economic Dynamics and Control, 12, 297-332.

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

[28] 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

[29] Perron, P. (2005). Dealing with Structural Breaks. Mimeo forthcoming in the Handbook of Econometrics, Econometric Theory.

[30] Rudebusch, G. D. (1993), The Uncertain Unit Root in Real GNP. *The American Economic Review*, 83, 264-272.

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

[32] Scott, H. R. and Hatemi-J. A. (2006). The Properties of Procedures Dealing with Uncertainty about Intercept and Deterministic Trend in Unit Root Testing.

[33] 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

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

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

[36] Zivot, E. and Andrews, D.W.K. (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.