A new criterion for lag-length selection in unit root tests

Agunloye, Oluokun Kasali[*]**; Arnab, Raghunath; Shangodoyin, Dahud Kehinde**

Department of Statistics, University of Botswana, Gaborone, Botswana

Email address:

k.agunloye@yahoo.com (Agunloye O. K.)

Abstract: This paper examines lag selection problem in unit root tests which has become a major specification problem in empirical analysis of non-stationary time series data. It is known that the implementation of unit root tests requires the choice of optimal truncation lag for good power proper ties and it is equally unrealistic to assume that the true optimal truncation lag is known a prior to the practitioners and other applied researchers. Consequently, these users rely largely on the use of standard information criteria for selection of truncation lag in unit root tests. A number of previous studies have shown that these criteria have problem of over-specification of truncation lag-length leading to the well-known low power problem that is commonly associated with most unit root tests in the literature. This paper focuses on the problem of over-specification of truncation lag-length within the context of augmented Dickey-Fuller (ADF) and generalized least squares Dickey-Fuller (DF-GLS)unit root tests. In an attempt to address this lag selection problem, we propose a new criterion for the selection of truncation lag in unit root tests based on Koyck distributed lag model and we show that this new criterion avoids the problem of over-specification of truncationlag-length that is commonly associated with standard information criteria.

Keywords: Truncation Lag, Information Criteria, Koyck Distributed Lag Model, Unit Root Test, Low Power, Partial Correlation Coefficient

1. Introduction

Testing for the presence of unit root in time series data is a major precondition in any cointegration analysis and other empirical research using time series data. Determination of appropriate truncation lag is quite a challenging aspect of unit root testing. A number of previous studies such as [1], [2], [3],[4],[5] and [6] have shown that there is a strong connection between truncation lag and the empirical power of unit root tests. Among numerous unit root tests proposed in the literature, the augmented Dickey–Fuller (ADF) test introduced by [7] and the generalized least squares Dickey-Fuller (DF-GLS) test introduced by [8] appears to be the most popular unit root tests among applied researchers. Both ADF and DF-GLS unit root tests are well-formulated to handle any possible serial correlations in the error terms of the Dickey–Fuller regressions by augmenting the regressions with lagged differences of the original series. However, the empirical implementation of the ADF and DF-GLS unit root tests requires the inclusion of appropriate number of lagged differences in the Dickey–Fuller regression, which is commonly referred to as lag selection problem in the literature. In practice, lag-lengths are commonly selected by two different lag selection techniques such as general-to-specific criterion proposed by [9] which is based on some sequential t-test procedure and standard information criteria such as Akaike information criterion (AIC), Final Prediction Error (FPE),Bayesian information criterion(BIC) and Hannan-Quinn information criterion (HQIC) to mention only a few but these criteria are generally well-known to have problem of over-fitting the truncation lags in unit root tests leading to low power in unit root tests.[2],[9] and [10]compared the performance of ADF test under these two lag-selection techniques and showed that sequential t-test outperformed standard information criteria.[11] compared the performance of DF-GLS test under modified Akaike information criteria (MAIC) and sequential t-testing procedure and show that the latter perform better than former. In an attempt to address the well-known problem of low power of ADF unit root test occasioned by over-specification of truncation lag-length, [8] introduced a class of efficient unit root tests called generalized least squares Dickey-Fuller(DF-GLS) test using ADF test based on GLS-detrending. Although, DF-GLS test was shown to be more powerful than the ADF test but it also inherits the lag length selection problem since it is a modification of the original ADF test. Hence, DF-GLS test is also characterized by low power problem due to over-specification of truncation lag-length. Another noticeable problem of

standard information criteria is the fact that different lag selection criteria choose different optimal truncation lag-lengths for the same dataset. This multiple suggestions of possible truncation lags by different information criteria raises a fundamental question as to on which particular information criterion could we rely upon for the choice of optimal truncation lag-length. To date, there exists no operational procedure for selecting optimal truncation lag-length for unit root tests that gives the best results. Hence, this paper focuses on the choice of optimal truncation lag-length for the ADF family of unit root tests and DF-GLS test since these are the most widely used unit root tests in empirical analysis.

2. Standard Information Criteria

In this study, we considered four standard information criteria for lag selection in unit root testing. The procedure is to fit autoregressive model on $y_t^{(i)}, i = 1,....,4$ for order ranging from 1 to 9 and subsequently obtaining the quantities to compute the values of the criteria. The following lag selection methods are considered:

i.[12]: Final Prediction Error (FPE)

$$FPE(k) = \ln\left(\hat{\sigma}_k^2\right) + \frac{2(1+k)}{N-k}$$

ii.[13]: Akaike Information Criterion(AIC)

$$AIC(k) = \ln\hat{\sigma}_k^2 + \frac{2(1+k)}{N}$$

iii.[14]:Bayesian Information Criterion (BIC)

$$BIC(k) = \ln\left(\hat{\sigma}_k^2\right) + \frac{(1+k)\ln N}{N}$$

iv.[15]: Hannan-Quinn Information Criteria (HQIC)

$$HQIC(k) = \ln\left(\hat{\sigma}_k^2\right) + \frac{2(1+k)\ln\left(\ln(N)\right)}{N}$$

Where k is the lag-length selected, N is the sample size and $\hat{\sigma}_k^2$ is the MLE of the model residual. Our objective is to ascertain if these conventional lag selection criteria either over-fit or under-fit the truncation lags by comparing the lag-length suggested by these information criteria with lag-length suggested by the new lag selection criterion.

3. A New Approach to Selection of Truncation Lag in Unit Root Tests

We consider a distributed lag representation of ADF and DF-GLS regression models in the form:

$$y_t^{(1)} = \Delta y_t - \rho y_{t-1} = \sum_{j=0}^{k} \gamma_j \Delta y_{t-j} + \varepsilon_{1t}$$

$$y_t^{(2)} = \Delta y_t - \alpha - \rho y_{t-1} = \sum_{j=0}^{k} \gamma_j \Delta y_{t-j} + \varepsilon_{2t}$$

$$y_t^{(3)} = \Delta y_t - \alpha - \beta t - \rho y_{t-1} = \sum_{j=0}^{k} \gamma_j \Delta y_{t-j} + \varepsilon_{3t}$$

$$y_t^{(4)} = \Delta y_t^d - \alpha - \gamma t - \rho y_{t-1}^d = \sum_{j=0}^{k} \delta_j \Delta y_{t-j}^d + \varepsilon_{4t}$$

$$(1)$$

Where $y_t^{(1)}$ is ADF Model I with no constant and no trend

$y_t^{(2)}$ is ADF Model II with constant but no trend

$y_t^{(3)}$ is ADF Model III with both constant and trend

$y_t^{(4)}$ is DF-GLS with both constant and trend

$\varepsilon_{ti}, i = 1,...,4$ are white noise error terms, k is the truncation lag-length to be determined empirically, γ_j and δ_j are coefficients of differenced lagged values. The R.H.S of equation (1) is similar to typical distributed lag models used in econometric modeling. [16] has proposed an ingenious method of specifying the lag in distributed lag model by assuming that the coefficients of γ_i's and δ_i's on the R.H.S of equation (1) are all of the same sign and declines geometrically, then without loss of generality we can assume Koyck postulations as follows:

Let

$$\lambda_k = \left(\delta_k \text{ or } \gamma_k\right) \qquad (2)$$

then we have

$$\lambda_k = \lambda_0 R_{(i)}^k \qquad k = 0,1, \quad i = 1,....,4 \qquad (3)$$

Where R is the partial correlation coefficient between $y_t^{(i)}/i = 1,...,4$ and Δy_{t-j} or Δy_{t-j}^d and measures the rate at which $y_t^{(i)}$ depends on either Δy_{t-j} or Δy_{t-j}^d and for the rate of decline of λ_k's we take $R_{(i)}$ such that $0 < R_{(i)} < 1$ is the indicator of decay of the distributed lag and $1 - R_{(i)}$ is the speed of adjustment[17].

Equation(2) fundamentally postulates that each λ coefficient is a measure of dependence of $y_t^{(i)}$ on either

Δy_{t-j} or Δy_{t-j}^{d} and each successive λ coefficient is numerically less than each preceding λ, this implies that as one goes back into the distant past the effect of that lag on $y_{t}^{(i)}$ becomes progressively smaller.

To obtain the mean lag, we ensure that the long-run multiplier is finite

That is the sum of λ_{k} is defined as

$$\sum_{k=0}^{\infty} \lambda_k = \lambda_0 \sum R_{(i)}^{k} \qquad (4)$$

but

$$\sum R_{(i)}^{k} = \frac{1}{1-R_{(i)}} \qquad (5)$$

is the inverse of speed of adjustment. Therefore, the Mean lag is

$$\bar{L}_{(i)} = \frac{\sum k \lambda_k}{\sum \lambda_k} = \frac{k \sum \lambda_k}{\lambda_0 \left(1-R_{(i)}\right)^{-1}}$$

$$= \frac{\sum k \lambda_0 R_{(i)}^{k}}{\lambda_0 \left(1-R_{(i)}\right)^{-1}} = \left(1-R_{(i)}\right) \sum k R_{(i)}^{k} \qquad (6)$$

Applying arithmetic-geometric progression rule to (6) we obtain the sum to infinity for the term $k R_{(i)}^{k}$ to be

$$\sum_{k=0}^{\infty} k R_{(i)}^{k} = \frac{R_{(i)}}{\left(1-R_{(i)}\right)^{2}} \qquad (7)$$

So using (7) in (6) gives

$$\bar{L}_{(i)} = \frac{R_{(i)}}{1-R_{(i)}} \quad , i=1,.....,5 \qquad (8)$$

The notation $R_{(i)}$ indicates that we generate different R's for each of the models described in equation (1) and we shall demonstrate the power of test for $R_{(i)}$ for different values of $\rho = 0.001, 0.5, and\ 0.999$.

The following are the immediate consequences of equation (8) :

(i). The closer R is to +1, the slower the rate of decline in λ_{k}, whereas the closer it is to zero, the more rapid the decline in λ_{k}. This ensures that, to have reasonable mean lag $\bar{L}_{(i)}$ we expect the absolute value of R to be in the interval $[0.5, 0.999)$

(ii) The simple linear regression model will be fitted to the L.H.S of equation (1) to generate the parameters needed and to compute $y_{t}^{(i)}$, $\forall\ i=1,.........,4$

(iii) If $R_{(i)} < 0.3$, then the mean lag will be assumed to be zero since $\dfrac{R_{(i)}}{\left(1-R_{(i)}\right)}$ for $R_{(i)} < 0.3$ is a fraction not up to 0.5.

(iv) The partial correlation coefficient $R_{(i)}$ should be computed such that we include all variables and adjusted for maximum lag until it gives a value less than 0.3

4. Power of Test for R

We define power of test for $R_{(i)}$ as follows:

$$H_0 : R_{(i)} = 1 \text{ against } H_1 : R_{(i)} < 1$$

Assuming that $R_{(i)}$ is normally distributed, then we define the test statistic as

$$Z = \frac{\hat{R}_{(i)} - 1}{\sqrt{V\left(\hat{R}_{(i)}\right)}} \qquad (9)$$

We reject the null hypothesis of $R_{(i)} = 1$ and hence leads to undefined mean lag $\bar{L}$ at a specified level of significance α if $Z < -Z_\alpha$. The Power of possible mean lag is the probability of rejecting H_0 when H_1 is true i.e when $R_{(i)} < 1 = R^*$

Thus we have

$$P\left(R^*\right) = P\left\{ Z < -Z_\alpha / R_{(i)} = R^* \right\} \qquad (10)$$

This reduces to

$$P\left(R^*\right) = P\left\{ \frac{\hat{R}_{(i)} - 1}{\sqrt{V\left(\hat{R}_{(i)}\right)}} < -Z_\alpha / R_{(i)} = R^* \right\} \qquad (11)$$

$$= P\left\{ \frac{\hat{R}_{(i)}}{\sqrt{V\left(\hat{R}_{(i)}\right)}} < -Z_\alpha + \frac{1}{\sqrt{V\left(\hat{R}_{(i)}\right)}} / R_{(i)} = R^* \right\}$$

$$= P\left\{ \frac{\hat{R}_{(i)} - R^*}{\sqrt{V\left(\hat{R}_{(i)}\right)}} < -Z_\alpha + \frac{1}{\sqrt{V\left(\hat{R}_{(i)}\right)}} - \frac{R^*}{\sqrt{V\left(\hat{R}_{(i)}\right)}} / R_{(i)} = R^* \right\}$$

$$= P\left\{ Z < -Z_\alpha + \frac{(1-R^*)}{\sqrt{V\left(\hat{R}_{(i)}\right)}} \right\} = P\left\{ Z < -Z_\alpha - \frac{(R^*-1)}{\sqrt{V\left(\hat{R}_{(i)}\right)}} \right\}$$

$$= P\left\{ Z > Z_\alpha + \frac{(R^*-1)}{\sqrt{V\left(\hat{R}_{(i)}\right)}} \right\} \qquad (12)$$

After some algebraic manipulation of Equation (11) we have Equation (12) which defines the power of test for $R_{(i)}$

5. Empirical Illustration

For empirical illustration, we fit simulated data to the distributed lag specifications of Dickey-Fuller regression models of ADF and DF-GLS unit root tests as follows:

For ADF MODEL I , $y_t^{(1)} = \Delta y_t - \rho y_{t-1} = \sum_{j=0}^{k} \gamma_j \Delta y_{t-j} + \varepsilon_{t1}$

We set $\rho = 0.001, 0.5$ and 0.999 in $y_t^{(1)}$ to have the following representations:

$$y_t^{(1)} = \begin{cases} \Delta y_t - 0.001 * y_{t-1} = a_1 \ , & \rho = 0.001 \\[2mm] \Delta y_t - 0.5 * y_{t-1} = a_2 \ , & \rho = 0.5 \\[2mm] \Delta y_t - 0.999 * y_{t-1} = a_3 \ , & \rho = 0.999 \end{cases}$$

We generate series a_1 , a_2 and a_3 for different values of $\rho = 0.001, 0.5$ and 0.999 .We compute partial correlation coefficients between : a_1 and different choices of variables from the set of independent variables $\Delta y_t, \Delta y_{t-1}, \Delta y_{t-2}, ..., \Delta y_{t-12}$ while controlling for the effects of other remaining independent variables. We also repeat the same procedure for a_2 and a_3

For ADF MODEL II , $y_t^{(2)} = \Delta y_t - \alpha - \rho y_{t-1} = \sum_{j=0}^{k} \gamma_j \Delta y_{t-j} + \varepsilon_{t2}$,

we set $\rho = 0.001, 0.5$ and 0.999 in $y_t^{(2)}$ to have the following representations:

$$y_t^{(2)} = \begin{cases} \Delta y_t - \alpha_1 - 0.001 * y_{t-1} = b_1 \ , & \rho = 0.001 \\[2mm] \Delta y_t - \alpha_2 - 0.5 * y_{t-1} = b_2 \ , & \rho = 0.5 \\[2mm] \Delta y_t - \alpha_3 - 0.999 * y_{t-1} = b_3 \ , & \rho = 0.999 \end{cases}$$

where α_1 , α_2 and α_3 are obtained by fitting of a regression model $y_t^{(2)}$ equals to a constant for different values of ρ .With values of α_1 , α_2 and α_3 known ,we therefore proceed to generate the series for b_1 , b_2 and b_3 .Thereafter, we compute partial correlation coefficient between : b_1 and different choices of variables from the set of independent variables $\Delta y_t, \Delta y_{t-1}, \Delta y_{t-2}, ..., \Delta y_{t-12}$ while controlling for the effects of other remaining independent variables. We also repeat the same procedure for b_2 and b_3

For ADF MODEL III,

$$y_t^{(3)} = \Delta y_t - \alpha - \beta t - \rho y_{t-1} = \sum_{j=0}^{k} \gamma_j \Delta y_{t-j} + \varepsilon_{t3} ,$$

we set $\rho = 0.001, 0.5$ and 0.999 in $y_t^{(3)}$ to have the following representations:

$$y_t^{(3)} = \begin{cases} \Delta y_t - \alpha_1 - \beta_1 t - 0.001 * y_{t-1} = d_1 \ , & \rho = 0.001 \\[2mm] \Delta y_t - \alpha_2 - \beta_2 t - 0.5 * y_{t-1} = d_2 \ , & \rho = 0.5 \\[2mm] \Delta y_t - \alpha_3 - \beta_3 t - 0.999 * y_{t-1} = d_3 \ , & \rho = 0.999 \end{cases}$$

We estimate the regression equation $y_t^{(3)} = \alpha + \beta * t + \Delta y_t + \rho * y_{t-1}$, for values of $\rho = 0.001, 0.5$ and 0.999 .Thereafter, we proceed to generate series for d_1 , d_2 and d_3 .We compute partial correlation coefficients between: d_1 and different choices of variables from the set of independent variables $\Delta y_t, \Delta y_{t-1}, \Delta y_{t-2}, ..., \Delta y_{t-12}$ while controlling for the effects of other remaining independent variables. We also repeat the same procedure for d_2 and d_3

For DF-GLS unit root test,

$$y_t^{(4)} = \Delta y_t^d - \alpha - \gamma t - \rho y_{t-1}^d = \sum_{j=0}^{k} \delta_j \Delta y_{t-j}^d + \varepsilon_{t4} ,$$

we set $\rho = 0.001, 0.5$ and 0.999 in $y_t^{(4)}$ to have the following representations:

$$y_t^{(4)} = \begin{cases} \Delta y_t^d - \alpha_1 - \gamma_1 t - 0.001 * y_{t-1}^d = e_1 \ , & \rho = 0.001 \\[2mm] \Delta y_t^d - \alpha_2 - \gamma_2 t - 0.5 * y_{t-1}^d = e_2 \ , & \rho = 0.5 \\[2mm] \Delta y_t^d - \alpha_3 - \gamma_3 t - 0.999 * y_{t-1}^d = e_3 \ , & \rho = 0.999 \end{cases}$$

We estimate the regression equation $y_t^{(4)} = \alpha + \gamma * t + \Delta y_t^d + \rho * y_{t-1}^d$, for values of $\rho = 0.001, 0.5$ and 0.999 .Thereafter, we proceed to generate series for e_1 , e_2 and e_3 .We compute partial correlation coefficients between: e_1 and different choices of variables from the set of independent variables $\Delta y_t, \Delta y_{t-1}, \Delta y_{t-2}, ..., \Delta y_{t-12}$ while controlling for the effects of other remaining independent variables. We also repeat the same procedure for e_2 and e_3

6. Power of Test for ADF and DF-GLS Unit Root Tests under the New Lag Selection Criterion

In this section, we discuss the power of test for R for the ADF and DF-GLS unit root tests based on the optimal truncation lag selected by the new criterion.

Table 1. Power of Test for R When $\rho = 0.001$

TABLE 1A

Unit Root Test	Lag	R	p-value	Power of Test
ADF MODEL I	0	0.652	0.000	0.6118
ADF MODELII	0	0.652	0.000	0.6118
ADF MODEL III	0	0.652	0.000	0.6118
DF-GLS	0	0.752	0.000	0.7563

TABLE 1B

Unit Root Test	Lag	R	p-value	Power of Test
ADF MODEL I	1	0.054	0.094	0.3423
ADF MODELII	1	0.054	0.094	0.3423
ADF MODEL III	1	0.054	0.094	0.3423
DF-GLS	1	0.374	0.014	0.4475

From Tables 1A and 1B above, it is obvious that the partial correlation coefficient denoted by R is significant at zero lag but not significant at lag 1.Hence,for $\rho = 0.001$, lag 1may be pre-specified as maximum lag for these unit root tests but optimal truncation lag is 0 because the power of test for R is higher at lag 0 than at lag 1.From both tables it is also seen that the power of test for R under DF-GLS test is higher compared with ADF tests at lag 0 and lag 1 respectively.

From Tables 2A and 2B above, it is obvious that the partial correlation coefficient denoted by R is significant at zero lag but not significant at lag 1.Hence,for $\rho = 0.5$，lag 1may be pre-specified as maximum lag for these unit root tests but optimal truncation lag is 0 because the power of test for R is higher at lag 0 than at lag 1.From both tables it is also seen that the power of test for R under DF-GLS test

is higher compared with ADF tests at lag 0 and lag 1 respectively.

Table 2.POWER OF TEST FOR R WHEN $\rho = 0.5$

TABLE 2A

Unit Root Test	Lag	R	p-value	Power of Test
ADF MODEL I	0	0.785	0.000	0.9535
ADF MODELII	0	0.785	0.000	0.9535
ADF MODEL III	0	0.785	0.000	0.9535
DF-GLS	0	0.815	0.000	0.9607

TABLE 2B

Unit Root Test	Lag	R	p-value	Power of Test
ADF MODEL I	1	0.062	0.051	0.3771
ADF MODELII	1	0.062	0.051	0.3771
ADF MODEL III	1	0.062	0.051	0.3771
DF-GLS	1	0.412	0.022	0.6237

Table 3.POWER OF TESTFOR R WHEN $\rho = 0.999$

TABLE 3A

Unit Root Test	Lag	R	p-value	Power of Test
ADF MODEL I	0	0.850	0.000	0.9871
ADF MODELII	0	0.850	0.000	0.9871
ADF MODEL III	0	0.850	0.000	0.9871
DF-GLS	0	0.883	0.000	0.9924

TABLE 3B

Unit Root Test	Lag	R	p-value	Power of Test
ADF MODEL I	1	0.066	0.038	0.4317
ADF MODELII	1	0.066	0.038	0.4317
ADF MODEL III	1	0.066	0.038	0.4317
DF-GLS	1	0.312	0.013	O.5978

From Tables 3A and 3B above, it is obvious that the partial correlation coefficient denoted by R is significant at zero lagbut not significant at lag 1.Hence,for $\rho = 0.999$, lag 1may be pre-specified as maximum lag for these unit root tests but optimal truncation lag is 0 because the power of test for R is higher at lag 0 than at lag 1.From both tables it is also seen that the power of test for R under DF-GLS test is higher compared with ADF tests at lag 0 and lag 1 respectively.

7. Lag-Selection by Conventional Information Criteria

In order to evaluate the performance of standard information criteria, we fitted auto regression of order 1 to 9 to our simulated data and these information criteria select the following truncation lag as represented in table 4 below:

Table 4. *LAG-SELECTION BY CONVENTIONAL INFORMATION CRITERIA*

MODEL	RMSE	$\hat{\sigma}_k^2$	AIC	FPE	BIC	HQIC
$AR(1)$	2.38	5.6787	1.7496	1.7496	1.7736	1.7592
$AR(2)$	2.36	5.5649	1.7358	1.7360	1.7720	1.7503
$AR(3)$	2.34	5.4943	1.7296	1.7299	1.7779	1.7489
$AR(4)$	2.34	5.4897	1.7353	1.7358	1.7959	1.7595
$AR(5)$	2.35	5.5037	1.7445	1.7452	1.8174	1.7736
$AR(6)$	2.32	5.4242	1.7366	1.7375	1.8218	1.7706
$AR(7)$	2.33	5.4382	1.7459	1.7471	1.8435	1.7849
$AR(8)$	2.33	5.4010	1.7458	1.7474	1.8558	1.7898
$AR(9)$	2.33	5.4103	1.7543	1.7563	1.8769	1.8033

Table 4 has autoregressive (AR) models of order 1 to 9 in the first column and root mean square error (RMSE) for each model in the second column. The optimal truncation lag for a particular criterion is the lag that minimizes the value of that criterion. Hence, the optimal truncation lag chosen by AIC, FPE and HQIC is 3 respectively whilst BIC picked 2.This is clearly an over-specification of truncation lag when compared with truncation lag suggested by our new criterion.

8. Conclusion

In this paper, we have highlighted the lag selection problem in the context of ADF and DF-GLS tests based on standard information criteria which we have shown to over-estimate the truncation lag leading to the well-known problem of low-power associated with unit root tests. Given the persistent problem of over-specification of truncation lag by data-dependent standard information criteria, we introduced a new lag selection criterion based on Koyck distributed lag model where truncation lag is specified as a deterministic function of the partial correlation coefficient between dependent variable and different choices of independent variables of distributed lag specifications of ADF and DF-GLS unit root tests. This new procedure was shown to avoid the problem of over-specification that is commonly associated with standard information criteria that are commonly used by applied researchers for lag-selection.

References

[1] Schwert, G.W. (1989) "Tests for Unit Roots: A Monte Carlo Investigation," Journal of Business and Economic Statistics, 7, 147-160.

[2] Campbell, J. C. and Perron, P. (1991) "Pitfall and Opportunities: What Macroeconomists shouldknow about Unit Roots," NBER Technical Working Paper # 100

[3] Xiao, Z. and Phillips, P. C. B. (1997) "An ADF Coefficient Test for a Unit Root in ARMAModels of Unknown Order with Empirical Applications to the U.S. Economy," CowlesFoundation Discussion Paper # 1161,

[4] Maddala, G. S. and Kim, I. M. (1998) Unit Roots, Cointegration and Time Series, Cambridge University Press

[5] Cavaliere, G. (2012) "Lag-length Selection for a Unit Root test in the presence of non-stationaryVolatility," Cowles Foundation Discussion Paper # 1844.

[6] Dufour, J. M and King, M. L. (1991) "Optimal Invariant Tests for the Autocorrelation Coefficient in Linear Regressions with Stationary or Non-stationary AR(1) errors," Journal of Econometrics,47, 115-143

[7] Said, E. S. and Dickey, D. A. (1984) "Testing for a Unit Root in Autoregressive Moving Average Models of Unknown Order," Biometrika, 71, 3, 599-607.

[8] Elliott, G. Rothenberg, T. J. and Stock, J. H. (1996) "Efficient Tests for an Autoregressive UnitRoot," Econometrica, 64, 4, 813-836

[9] Hall, A. (1994) "Testing for a Unit Root in Time Series with Pretest Data-based ModelSelection," Journal of Business and Economic Statistics, 12, 4, 461-470

[10] Ng, S. and Perron, P. (2001) "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power," Econometrica, 69, 6, 1519-1554

[11] Shaowen Wu (2010)"Lag Length Selection In DF-GLS Unit Root Tests",Communication in Statistics-Simulation and computation,39:8,1590-1604

[12] Akaike, H., 1969. Fitting Autoregressive Models for Prediction. Annalsof The Institute of Statistical Mathematics, 21(2), 243–247.

[13] Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F., 2ndInternational Symposium on Information Theory. AkademiaiKiado`, Budapest, pp. 267–281.

[14] Schwarz, G. (1978) " Estimating the dimension of a model". Annals of Statistics, 6, 461 –464.

[15] Hannan, E. J. and Quinn, B. G. (1978). "The determination of the order of an autoregression". Journal of Royal Statistical Society, 41, 190 – 195.

[16] Koyck, L.M. (1954), Distributed Lags and Investment Analysis, Amsterdam: North-Holland.

[17] Gujarati, D.(2005),Essentials of Econometrics, McGraw-Hill School Education Group.