**USING INFLUENCE FUNCTION FOR LAG TRUNCATION IN UNIT ROOT TESTS**

**BY**

**AGUNLOYE1, OLUOKUN KASALI; ARNAB2, RAGHUNATH, & SHANGODOYIN3, DAHUD KEHINDE**

**1DEPARTMENT OF MATHEMATICS,**

**OBAFEMI AWOLOWO UNIVERSITY, ILE-IFE, NIGERIA,**

**2,3DEPARTMENT OF STATISTICS,**

**UNIVERSITY OF BOTSWANA, GABORONE, BOTSWANA.**

**CORRESPONDING AUTHOR: AGUNLOYE, OLUOKUN KASALI**

 **Telephone Number: +2347054933685**

**Email**: k.agunloye@yahoo.com

Abstract

This paper examined the application of influence function as a criterion for lag truncation in unit root tests. Lag selection in unit root tests has been dominated by standard information criteria and application of influence function as a means of determining truncation lag parameter for unit root tests is a new innovation within the context of augmented Dickey-Fuller (ADF) family of unit root tests and generalized least squares Dickey-Fuller (DF-GLS) test. Influence functions were generated for different lag-lengths and the choice of optimal lag-length was based on the particular lag with the largest influence. This methodology uses autocorrelation of time series data to identify the most influential lag among a set of possible truncation lags which is designated as optimal lag for the purpose of lag truncation in unit root tests. We demonstrated that influence function criterion out-performed the standard information in choosing appropriate lag structure for the unit root tests.

**Keywords**: Influence function, autocorrelation, unit root test, optimal truncation lag

**1. Introduction**

An important aspect of unit root testing involves identification of the most influential lag in the unit root test regression. In this paper we consider application of influence function for the identification of optimal truncation lag within the context of ADF family of unit root tests and DF-GLS test introduced by [1] and [2] respectively. The idea of influence function was first mooted by [3] where influence function was used for detecting influential points or outliers. There are various applications of influence function in model selection in time series analysis.[4] developed an ingenious method of outlier detection using a plot of influence function of datum points on the theoretical autocorrelation function.[5] and [6] applied influence function for outlier detection and model order determination in time series data.[7] utilized influence function for model selection in kernel-based regressions, to mention just a few.

The remainder of this paper is organized as follows: Section 2 deals with specification of influence function. Section 3 discusses flowchart for the computation of influence function for lag truncation in unit root tests. Section 4 covers data description and preliminary analysis, Section 5. discusses unit root testing using lag selected by information-based lag selection criteria and influence Function Criterion. Section 6 concludes

**2. Influence Function and Lag Specification**

For a general parameter  expressed as a function of the distribution; the influence function  at  is according to [3] in [5] is given by

 (1)

 and  is a perturbation of  by ,the distribution function for a point mass of one at .[8] has demonstrated that the influence function of  for any univariate distribution with finite second moment is

 (2)

Where  and are standardized forms of variates  and  say. If  and  denote respectively the standardized sum of and difference between  and.

Also if  and . Then equation (2) may be written as

 (3)

 [9] gave a first order approximation to the function and it was noted that a sample analogue of (3) is

 (4)

Where  are the  bivariate observations,  is the correlation based on all but the  observation,  and  are sample analogues of  and .[4] have considered the influence function for the estimation of time series autocorrelation. Extending the work of [4] and [9], [6] considered the use of influence function to detect the presence of outliers and model order determination for time series data. Equation (4) (see [6] ) provides a procedure for model order determination which is particularly useful to check for lag truncation in the ADF and DF-GLS regression models of the form:

 (5)

Where  is ADF model I with no constant and no trend ; is ADF model II with constant but no trend; is ADF model III with both constant and trend and is DF-GLS model with both constant and trend

 are white noise error terms,  and  are coefficients of differenced lagged values, is the truncation lag to be determined empirically using influence function. Suppose that  is autocorrelation function at lag  for a p-periodic data; then the influence function (see [5] ) is

 (6)

Where  is the autocorrelation of lag  for any L.H.S  in equation (5) and

 (7)

The procedure for lag truncation according to [5] is to construct the influence function matrix .Where  is the number of observations and  is a fixed number equals to the periodicity of the data (for quarterly data  and for monthly data ). The construction of this matrix is based on critical value of  .The influence function estimates exceeding the chosen critical value (in magnitude) are designated plus or minus depending on the sign of the estimates, while others are left blank indicating low influence of particular lag. The  with the highest number of blanks and for which cut-off is the possible order of the model.

**3. Flowchart for the Computation of Influence Function for Lag**

 **Truncation in Unit Root Tests**

The following steps are used in preparation of R-package code for the computation of influence function for different lags

Step1: Find the first difference of the series under investigation and represent it as 

Step 2: Compute the autocorrelation for series at lag  for =1,2,……,12

Step 3: Compute  and  for all =1,2,……,12

Step 4 : Compute for  where  and  are the mean and standard deviation for the series  respectively.

Step 5: Compute  for all =1,2,……,12

Step 6: Compute  and 

Step 7: Compute  and  for all =1,2,……,12

Step 8: Compute the critical value 

Step 9: Compute 

Step 10: Compute the influence function denoted by I ,where 

 Insert tables 1 &2 here

Tables 1 and 2 present the summary of lag selection by influence function criterion (IFC) for USMGS and USMTB series respectively. The optimal lags suggested by IFC for USMGS and USMTB series are 5 and 10 respectively being the most influential lags .The critical values for the optimal lags are 0.03996622 and 0.0399610 for USMGS and USMTB respectively.

 Insert tables 3&4 here

Tables 3 and 4 present the summary of lag selection by influence function criterion (IFC) for simulated data 1 and 2 respectively. The optimal lags suggested by IFC for simulated data I and II are 1 and 2 respectively being the most influential lags .The critical values for the optimal lags are 0.03160693 and 0.03160685 for simulated data I and 2 respectively.

**4. Empirical Evaluation of Unit Root Testing using Lag selected by Information-based**

 **lag selection criteria**

We run a battery of unit root tests using the various optimal lag lengths suggested by different information-based lag selection criteria such as AIC and FPE proposed by [10] and [11] as well as BIC and HQIC introduced by [12] and [13] respectively. For USMGS series, the optimal lag-lengths suggested by AIC, MAIC, BIC and HQIC are 11, 0, 1 and 1 respectively. Similarly, for USMTB series, the optimal lag-lengths suggested by AIC, MAIC, BIC and HQIC are 12, 0, 12 and 12 respectively. The empirical results are presented in tables 5, 6, 7 and 8 below:

 Insert tables 5,6,7 &8 here

Table 5 through table 8 present the outcome of unit root tests conducted for testing the stationarity properties of the level of USMGS and USMTB series using the optimal lag-lengths suggested by conventional lag selection criteria. The null hypotheses of unit root for the level of USMGS and USMTB cannot be rejected across the various optimal truncation lag-lengths considered since the test statistic is greater than the critical value at 5% level of significance for the three versions of ADF tests considered as well as DF-GLS test.These results indicate that both USMGS and USMTB series are non-stationary at level

 Insert tables 9,10,11 & 12 here

Table 9 through table 12 present the outcome of unit root tests conducted for testing the stationarity properties of the first difference of USMGS and USMTB series using the optimal lag-lengths suggested by conventional lag selection criteria. The null hypotheses of unit root for the first difference of USMGS and USMTB were rejected across the various optimal truncation lag-lengths considered since the test statistic is less than the critical value at 5% level of significance for the three versions of ADF tests considered as well as DF-GLS test.This indicates that both USMGS and USMTB series are stationary after first difference indicating that each series is integrated of order 1

**5. Empirical Evaluation of Unit Root Testing Using Lag Selected by IFC**

Based on influence function criterion (IFC), the optimal truncation lag required for the implementation of unit root test for USMGS is 5 being the most influential lag among a set of candidate truncation lags. We run unit root test on USMGS series for lag 1 and lag 5 to evaluate possibility of discrepancy in the outcome of unit testing under these two influential lags as shown in table 1 above. Similarly, the optimal truncation lag suggested by IFC for USMTB series is 10 being the most influential lag. Consequently we run unit root test on USMTB for lag 1, lag 9 and lag 10 to evaluate the behavior of the test across these three influential lags and the empirical results are presented in tables 13 and 14 below:

 Insert tables 13 &14 here

Tables 13 and 14 present the outcome of unit root tests conducted for testing the stationarity properties of the first difference of USMGS and USMTB series using the optimal lag-lengths suggested by IFC. The null hypotheses of unit root for the level of USMGS and USMTB cannot be rejected series across the various optimal truncation lag-lengths considered since the test statistic is greater than the critical value at 5% level of significance for the three versions of ADF tests considered as well as DF-GLS test.This indicates that both USMGS and USMTB series are non-stationary at level.

 Insert tables 15 &16 here

Tables 15 and 16 present the outcome of unit root tests conducted for testing the stationarity properties of the first difference of USMGS and USMTB series using the optimal lag-lengths suggested by IFC. The null hypotheses of unit root for the first difference of USMGS and USMTB were rejected across the various optimal truncation lag-lengths considered since the test statistic is less than the critical value at 5% level of significance for the three versions of ADF tests considered as well as DF-GLS test.This indicates that both USMGS and USMTB series are stationary after first difference indicating that each series is integrated of order 1

**Conclusion**

This paper has examined application of influence function as an alternative criterion for lag specification in ADF and DF-GLS unit root tests. This methodology provided a hierarchical order of influence for different candidate optimal lag-lengths and this hierarchical structure could serve as standard guide for applied researchers in avoiding the problem of over-estimation and under-estimation of truncation lag-length that is commonly associated with conventional lag selection criteria.

REFERENCES

[1] 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.

[2] Elliott, G. Rothenberg, T. J. and Stock, J. H. (1996) .Efficient Tests for an Autoregressive

 Unit Root.*Econometrica*, 64, 4, 813-836

[3] Hampel, F.R., (1974). “The Influence Curve and its Role in Robust Estimation”,

 *Journal of the American Statistical Association,* vol. 69, no.346, 383-393.

[4] Chernick, M.R, Downing, D.I and Pike, D.H (1982): Detecting outliers in Time series data.

 *Journal of American Statistical Association.* Vol. 17, No. 380, pp 743-747.

[5]Shangodoyin, D.K.,(1994). *On the specification of time series models in the presence of*

 *aberrant observations.* Unpublished PhD Thesis. University of Ibadan, Ibadan,

 Nigeria.

[6] Shangodoyin, D. K., (2000).Specification of dynamic distributed lag model connecting two

 outlier infested time series. *Journal of Science Research* , Vol.6(2),97-103

[7]Debruyne, M.,Hubert,M.and Suykens,J.A.K.(2008): Model Selection in Kernel-based

 Regression Using the Influence Function.*Journal of Machine Learning*

 *Research,*9,2377-2400

[8] Miller, R.G. (1974). An unbalanced Jackknife, *The Annals of Statistics*, 2, 880-891.

[9] Gnanadesikan, R., and Kettenring, J. R. (1972). Robust estimates, residuals and outlier

 detection with multiresponse data. *Biometrics*, *28*, 81-124.

 [10]Akaike, H., (1973). Information theory and an extension of the maximum likelihood

 principle.2nd International Symposium on Information Theory. In: B.N. Petrov and

 F. Csáki, (Eds) (Budapest: Académiai Kiadó), 267–281.

[11] Ng,S. & Perron(2005).Lag Length Selection and Construction of Unit Root Tests With

 Good Size and Power.*Econometrica* 69,1519-1554.

[12]Schwarz, G., 1978, Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464.

[13]Hannan, E. J. and Quinn, B. G. (1978). “The determination of the order of an

 autoregression”.*Journal of Royal Statistical Society*, 41, 190 – 195.

Table1: Lag selection for USMGS Data

|  |  |  |  |
| --- | --- | --- | --- |
| Lag | No of Observations | Critical Value | IFC |
| 1 | 623 | 0.03996788 | 385 |
| 2 | 622 | 0.03996763 | 362 |
| 3 | 621 | 0.03996726 | 363 |
| 4 | 620 | 0.03996679 | 366 |
| 5 | 619 | 0.03996622 | 390\*\* |
| 6 | 618 | 0.03996553 | 346 |
| 7 | 617 | 0.03996474 | 337 |
| 8 | 616 | 0.03996384 | 358 |
| 9 | 615 | 0.03996282 | 371 |
| 10 | 614 | 0.03996170 | 375 |
| 11 | 613 | 0.03996046 | 354 |
| 12 | 612 | 0.03995910 | 353 |

**\*\*** indicates optimal lag

Table 2: Lag selection for USMTB data

|  |  |  |  |
| --- | --- | --- | --- |
| Lag | No of Observations | Critical Value | IFC |
| 1 | 623 | 0.03996788 | 431 |
| 2 | 622 | 0.03996763 | 389 |
| 3 | 621 | 0.03996726 | 378 |
| 4 | 620 | 0.03996679 | 400 |
| 5 | 619 | 0.03996622 | 422 |
| 6 | 618 | 0.03996553 | 398 |
| 7 | 617 | 0.03996474 | 389 |
| 8 | 616 | 0.03996384 | 426 |
| 9 | 615 | 0.03996282 | 444 |
| 10 | 614 | 0.03996170 | 448\*\* |
| 11 | 613 | 0.03996046 | 425 |
| 12 | 612 | 0.03995910 | 375 |

**\*\*** indicates optimal lag

Table 3: Lag selection for simulated data 1

|  |  |  |  |
| --- | --- | --- | --- |
| Lag | No of Observations | Critical Value | IFC |
| 1 | 998 | 0.03160693 | 548\*\* |
| 2 | 997 | 0.03160685 | 536 |
| 3 | 996 | 0.03160674 | 553 |
| 4 | 995 | 0.03160659 | 539 |
| 5 | 994 | 0.03160642 | 546 |
| 6 | 993 | 0.03160621 | 533 |
| 7 | 992 | 0.03160597 | 540 |
| 8 | 991 | 0.03160569 | 535 |
| 9 | 990 | 0.03160538 | 544 |
| 10 | 989 | 0.03160504 | 534 |
| 11 | 988 | 0.03160467 | 526 |
| 12 | 987 | 0.03160426 | 547 |

**\*\*** indicates optimal lag

Table 4: Lag selection for simulated data 2

|  |  |  |  |
| --- | --- | --- | --- |
| Lag | No of Observations | Critical Value | IFC |
| 1 | 998 | 0.03160693 | 552 |
| 2 | 997 | 0.03160685 | 568\*\* |
| 3 | 996 | 0.03160674 | 524 |
| 4 | 995 | 0.03160659 | 546 |
| 5 | 994 | 0.03160642 | 548 |
| 6 | 993 | 0.03160621 | 544 |
| 7 | 992 | 0.03160597 | 541 |
| 8 | 991 | 0.03160569 | 522 |
| 9 | 990 | 0.03160538 | 534 |
| 10 | 989 | 0.03160504 | 537 |
| 11 | 988 | 0.03160467 | 541 |
| 12 | 987 | 0.03160426 | 544 |

**\*\*** indicates optimal lag

##### Table 5 : Unit root test for the level of USMGS and USMTB series using ADF model I

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | ADF I |  | ADF I |  | ADF I |  | ADF I |  |
| USMGS | -0.8024 | 11 | -0.7003\* | 0 | -0.8553\* | 1 | -0.8553 | 1 |
| USMTB | -1.0598\* | 12 | -0.7003\* | 0 | -1.0598\* | 12 | -1.0598 | 12 |

\*null hypothesis rejected at 0.05 level of significance

Table 6 : Unit root test for the level of USMGS and USMTB series using ADF model II

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | ADF II |  | ADF II |  | ADF II |  | ADF II |  |
| USMGS | -0.9913\* | 11 | -0.5683\* | 0 | -1.0509\* | 1 | -1.0509\* | 1 |
| USMTB | -1.7072\* | 12 | -1.5458\* | 0 | -1.7072\* | 12 | -1.7072\* | 12 |

\*null hypothesis rejected at 0.05 level of significance

Table 7 : Unit root test for the level of USMGS and USMTB series using ADF model III

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | ADF III |  | ADF III |  | ADF III |  | ADF III |  |
| USMGS | -1.7504\* | 11 | -1.3415\* | 0 | -1.7353\* | 1 | -1.7353\* | 1 |
| USMTB | -2.5131\* | 12 | -2.2999\* | 0 | -1.5131\* | 12 | -2.5131\* | 12 |

\*null hypothesis rejected at 0.05 level of significance

Table 8 : Unit root test for the level of USMGS and USMTB series using DF- GLS Test

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | DF-GLS |  | DF-GLS |  | DF-GLS |  | DF-GLS |  |
| USMGS | -0.8522 | 10 | -0.8522 | 10 | -0.8522 | 10 | -0.8522 | 10 |
| USMTB | -1.5292\* | 12 | -1.4060\* | 0 | -1.5292\* | 12 | -1.5292\* | 12 |

\*null hypothesis rejected at 0.05 level of significance

##### Table 9 : Unit root test for the first difference of USMGS and USMTB series using ADF model I

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | ADF I |  | ADF I |  | ADF I |  | ADF I |  |
| USMGS | -6.3316 | 11 | -16.7993 | 0 | -17.0449 | 1 | -17.0449 | 1 |
| USMTB | -6.1472 | 12 | -17.6117 | 0 | -6.1472 | 12 | -6.1472 | 12 |

\*null hypothesis rejected at 0.05 level of significance

Table 10 : Unit root test for the first difference of USMGS and USMTB series using ADF model II

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | ADF II |  | ADF II |  | ADF II |  | ADF II |  |
| USMGS | -6.3377\* | 11 | -16.7919\* | 0 | -17.0384\* | 1 | -17.0384\* | 1 |
| USMTB | -6.1459\* | 12 | -17.5988 | 0 | -6.1459\* | 12 | -6.1459\* | 12 |

\*null hypothesis rejected at 0.05 level of significance

Table 11 : Unit root test for the first difference of USMGS and USMTB series using ADF model III

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | ADF III |  | ADF III |  | ADF III |  | ADF III |  |
| USMGS | -6.5958\* | 11 | -16.8841\* | 0 | -17.1743\* | 1 | -17.1743\* | 1 |
| USMTB | -6.2115\* | 12 | -17.6114\* | 0 | -6.2115\* | 12 | -6.2115\* | 12 |

\*null hypothesis rejected at 0.05 level of significance

Table 12 : Unit root test for the first difference of USMGS and USMTB series using DF-GLS Test

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | DF-GLS |  | DF-GLS |  | DF-GLS |  | DF-GLS |  |
| USMGS | -2.0461 | 10 | -2.0461 | 10 | -2.0461 | 10 | -2.0461 | 10 |
| USMTB | -6.2048\* | 12 | -17.6336\* | 0 | -6.2048\* | 12 | -6.2048\* | 12 |

\*null hypothesis rejected at 0.05 level of significance

Table 13: Unit root test for the level of USMGS using IFC

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ADF I |  | ADF II |  | ADF III |  | DF-GLS |  |
| -0.8553\* | 1 | -1.0509\* | 1 | -1.7353\* | 1 | -1.0109\* | 1 |
| -0.7820\* | 5 | -0.8515\* | 5 | -1.5991\* | 5 | -0.8393\* | 5 |

\*null hypothesis rejected at 0.05 level of significance

Table 14: Unit root test for the level of USMTB using IFC

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ADF I |  | ADF II |  | ADF III |  | DF-GLS |  |
| -1.3901\* | 1 | -2.3995\* | 1 | -3.0979\* | 1 | -2.1273\* | 1 |
| -1.2015\* | 9 | -2.0511\* | 9 | -2.8207\* | 9 | -1.8112\* | 9 |
| -1.1525\* | 10 | -1.9260\* | 10 | -2.7011\* | 10 | -1.7106\* | 10 |

\*null hypothesis rejected at 0.05 level of significance

Table 15: Unit root test for the first difference of USMGS using IFC

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ADF I |  | ADF II |  | ADF III |  | DF-GLS |  |
| -17.0449\* | 1 | -17.0384\* | 1 | -17.1743\* | 1 | -9.4135\* | 1 |
| -9.7895\* | 5 | -9.7901\* | 5 | -9.9824\* | 5 | -4.1234\* | 5 |

\*null hypothesis rejected at 0.05 level of significance

Table 16: Unit root test for the first difference of USMTB using IFC

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ADF I |  | ADF II |  | ADF III |  | DF-GLS |  |
| -17.9198\* | 1 | -17.9072\* | 1 | -17.9345\* | 1 | -17.9538\* | 1 |
| -7.3429\* | 9 | -7.3403\* | 9 | -7.4022\* | 9 | -7.3982\* | 9 |
| -6.6370\* | 10 | -6.6348\* | 10 | -6.6964\* | 10 | -6.6909\* | 10 |

\*null hypothesis rejected at 0.05 level of significance

APPENDIX

 R-Package Code for the Computation of Influence Function for Lag Truncation in Unit Root Tests

The following R code was used for the computation of influence function for different lags using both real and simulated data :

IFC=function(y,k){

 delta\_y=diff(y)

 z=(delta\_y-mean(delta\_y))/sd(delta\_y)

 rho=acf(delta\_y, lag.max=k, plot=F)$acf[-1]

 r\_1=sqrt(1+rho)

 r\_2=sqrt(1-rho)

 r [1] = 1-rho^2

 n = rep(0,k)

 c.pt = rep(0,k)

 IF = rep(0,k)

 for (i in 1:k){

a = (z+lag(z,i))/r\_1[i] , b = (z-lag(z,i))/r\_2[i]

V\_1=(a+b)/2; V\_2=(a-b)/2;

IF=V= r[1]\*V\_1\*V\_2

V = c(V)

n[i]=length(V)

c.pt[i]=sqrt(n[i]-i)/(n[i]\*(n[i]+2)))

 for (j in 1:n[i]) {

 if(V[j]<c.pt){V[j]=1}

 else {V[j]=0}

 }

IF[i] =sum(V)

 }

Critical .pt=c.pt; IFC=IF

max.IFC=max(IFC)

for(i in 1:length(IF)){

 if (IF[i]=max.IFC){IF[i]=”\*\*\*”}

 else{IF[i]=””}

 }

result = data.frame(n,critical.pt,IFC,optimal.lag=IF)

return(result)

}