Economic Issues, Vol.6, Part 2, September, 2001

Finite-sample critical values of the Augmented Dickey-Fuller statistic: a note on lag order

Steven Cook[1]

ABSTRACT

The lag order dependence of finite-sample Augmented Dickey-Fuller (ADF) critical values is examined via a comparison of the response surface specifications of Cheung and Lai (1995) and MacKinnon (1991). Theoretical, Monte Carlo and empirical evidence show failure to incorporate lag order effects reduces the power of the ADF test to reject the unit root null hypothesis.

1. INTRODUCTION

TESTING OF the unit root hypothesis has become a standard feature of applied econometric research, with the Dickey-Fuller (1979) (DF) test commonly employed. Due to the invalidity of the DF statistic in the presence of serial correlation, the test is typically applied in its augmented form. To examine the significance of the resulting Augmented Dickey-Fuller (ADF) statistic, the response surface analysis of MacKinnon (MK) (1991) can be used to calculate critical values for any sample size. This has become standard practice in empirical work, with the MK critical values incorporated in leading econometric software packages such as ECONOMETRIC VIEWS, MICROFIT and PC-GIVE. However, despite the flexibility afforded by the MK analysis, a potential problem exists as the MK critical values were derived on the basis of the estimation of a non-augmented DF test. Therefore, although the degree of augmentation, or lag order, of the test is asymptotically irrelevant it may affect finite-sample critical values. Recognition of this led Cheung and Lai (CL) (1995) to derive response surfaces where critical values are a function of both sample size and lag order.[2] Given the demonstrated superiority of these 'extended' response surfaces provided by the authors, it is surprising that the results of CL have not received widespread adoption. Similarly it is surprising that there has as yet not been a comparison made between the critical values obtained from the alternative response surface specifications. Whether lag adjusted critical values are significantly different to those of MK and whether this has implications for the power of the ADF test and the resulting inferences drawn, are the focus of this paper.

This paper will proceed as follows. In the following section the alternative response surface formulations of MK and CL are presented. From inspection of the alternative response surfaces it is not obvious how the critical values under the two approaches differ for ADF tests conducted over different sample sizes using differing lag lengths. Section three therefore examines CL and MK critical values over a range of sample sizes and lag orders. Via this simple exercise interesting properties of the two approaches are uncovered. In section four Monte Carlo simulation is employed to examine the power of the ADF test against a number of near-integrated alternatives using the alternative critical values. Section five provides an empirical application of the two approaches using the U.S. money stock. Section six concludes.

2 Response surface specifications

The response surfaces of MK allow the calculation of DF critical values for any sample size. Using estimated values of the parameters $\{\beta_\infty, \beta_1, \beta_2\}$, critical values, $cv(\cdot)$, are derived from the response surface:

$$cv(T;\alpha) = \beta_\infty + \beta_1 T^{-1} + \beta_2 T^{-2} + \varepsilon_T \tag{1}$$

where T is the sample size, α is the level of significance, and ε_T is the error term. As $T \rightarrow \infty$ the functions of T tend to zero, leaving the asymptotic critical value β_∞. Despite the apparent flexibility afforded by (1), it is based upon the estimation of a DF test rather than the typically employed ADF test. Although asymptotically irrelevant, finite-sample ADF critical values may be subject to lag order dependence. This has been noted by CL who derive more general response surface specifications incorporating terms involving lag order (p) as given below:

$$cv(T,p;\alpha) = \lambda_\infty + \lambda_1 T^{-1} + \lambda_2 T^{-2} + \phi_1\left(\frac{p}{T}\right) + \phi_2\left(\frac{p}{T}\right)^2 + \varepsilon_{T,p} \tag{2}$$

where $\varepsilon_{T,p}$ is an error term. As with (1), the above specification was selected to ensure that as $T \rightarrow \infty$ an asymptotic critical value (λ_∞) is approached.

There are some important features to note concerning (2). First, as the values of the parameters $\{\beta_i\}$ and $\{\lambda_i\}$ differ, an evaluation of the impact of lag order effects requires a direct comparison between (1) and (2) rather than just a calculation of the sum $\{\phi_1(p/T) + \phi_2(p/T)^2\}$. Secondly, the parameters $\{\phi_i\}$ are of opposite sign, with $\phi_1 > 0$, $\phi_2 < 0$. This implies a possible cancelling effect between the lag order terms in (2).

Despite CL graphically presenting critical values for their response surfaces across a range of lag orders, sample sizes and significance levels, the comparison between their critical values and those of MK was limited, with only critical values for $p = 6$ and $T = 50,100$ examined. The impact of lag adjustment on ADF critical values warrants a more thorough examination as CL found lag order regressors to be highly significant, with a range of criteria finding (2) to be a better fitting specification than the commonly used restricted form of (1). More precisely, the extended response surfaces of CL were found to be statistically superior on the basis of the mean and maximum absolute regression error using ε_T in (1) and $\varepsilon_{T,p}$ in (2), estimated equation standard error and squared multiple correlation coefficient.

3. Numerical examples

In this section the critical values for the 'constant and trend' ADF test[3] will be calculated for the following lag orders (p), sample sizes (T) and significance levels (α):

$$p = \{2, 4, 6, 8, 10, 12, 14\}$$

$$T = \{80, 100, 120, 150, 200, 500\}$$

$$\alpha = \{0.01, 0.05, 0.1\}$$

The critical values derived from this analysis are presented in Tables 1 to 3. Features to note from these results are:

- For all sample sizes (including $T = \infty$) and at all levels of significance, the MK critical values are greater in absolute value than those of CL. Consequently the MK values further exaggerate the low power problem associated with ADF tests.
- For any given sample size the MK and CL critical values diverge as the lag order is increased. This is observed at all levels of significance and persists when the sample size is increased. Therefore the possible cancelling out effect due to the opposite signs of ϕ_1 and ϕ_2 does not occur.
- Interestingly, while convergence to the asymptotic critical value as T increases occurs from below for MK, convergence occurs from above in many cases for the CL values. This is particularly apparent for high values of p at the 5 per cent and 10 per cent levels of significance. This reinforces the low power issue noted above.

The results show lag order effects to substantially alter finite-sample ADF critical values, especially for larger values of p. This is particularly relevant as recent studies suggest larger values of p should be applied in practice. First, De Jong et al. (1992) argue that the loss of power associated with larger values of p is outweighed by the size distortion resulting from the use of lower values.[4] Similar arguments have been presented by Banerjee *et al.* (1993), Harris (1992) and Schwert (1989), the latter suggesting the value of p should increase with the sample size employed. Secondly, studies such as Choi and Chung (1995) and Ng (1995) show the use of higher frequency data to increase the power of ADF test.[5] As the lag order of the ADF test typically increases with the frequency of observation, this promotes the use of higher values of p in empirical research.[6] Lag adjustment of the critical values is therefore particularly significant when using high frequency data over a short span.

4. Monte Carlo simulation

In previous sections it has been noted that the lower absolute values obtained from the CL response surfaces will allow the unit root hypothesis to be rejected more frequently. Analysis of the possible difference in power of the ADF test under the alternative critical values can be performed using Monte Carlo simulation. To examine the power of the ADF test against near-integrated alternatives the following data generation process (DGP) was employed:

$$y_t = \rho y_{t-1} + \xi_t \qquad t = 1, \ldots, T \tag{3}$$

The $\{\xi_t\}$ series was generated as pseudo i.i.d. $\mathbf{N}(0, 1)$ random numbers using the RNDNS procedure in the Gauss programming language version 3.2.13. The Monte Carlo experiments were

Table 1: ADF critical values, α=0.1

	Sample size (T)						
	80	100	120	150	200	500	∞
MK	-3.159	-3.153	-3.149	-3.144	-3.140	-3.133	-3.128
Cheung and Lai							
p = 2	-3.138	-3.134	-3.132	-3.130	-3.128	-3.124	-3.128
p = 4	-3.117	-3.117	-3.118	-3.118	-3.119	-3.121	-3.122
p = 6	-3.097	-3.101	-3.104	-3.107	-3.110	-3.117	-3.122
p = 8	-3.077	-3.085	-3.090	-3.096	-3.102	-3.113	-3.122
p = 10	-3.059	-3.069	-3.077	-3.085	-3.093	-3.110	-3.122
p = 12	-3.042	-3.055	-3.064	-3.074	-3.085	-3.106	-3.122
p = 14	-3.026	-3.041	-3.052	-3.064	-3.077	-3.103	-3.122

Table 2: ADF critical values, α=0.05

	Sample size (T)						
	80	100	120	150	200	500	∞
MK	-3.466	-3.455	-3.448	-3.440	-3.433	-3.421	-3.413
Cheung and Lai							
p = 2	-3.439	-3.431	-3.426	-3.422	-3.417	-3.410	-3.406
p = 4	-3.416	-3.412	-3.410	-3.409	-3.408	-3.406	-3.406
p = 6	-3.395	-3.395	-3.395	-3.396	-3.398	-3.402	-3.406
p = 8	-3.376	-3.379	-3.381	-3.385	-3.389	-3.398	-3.406
p = 10	-3.359	-3.364	-3.368	-3.374	-3.380	-3.395	-3.406
p = 12	-3.344	-3.350	-3.356	-3.363	-3.372	-3.391	-3.406
p = 14	-3.330	-3.337	-3.344	-3.353	-3.363	-3.387	-3.406

Table 3: ADF critical values, α=0.01

	Sample size (T)						
	80	100	120	150	200	500	∞
MK	-4.076	-4.052	-4.037	-4.022	-4.007	-3.981	-3.964
Cheung and Lai							
p = 2	-4.037	-4.018	-4.006	-3.995	-3.985	-3.968	-3.958
p = 4	-4.011	-3.996	-3.987	-3.980	-3.973	-3.963	-3.958
p = 6	-3.989	-3.977	-3.970	-3.965	-3.961	-3.958	-3.958
p = 8	-3.972	-3.961	-3.956	-3.952	-3.951	-3.953	-3.958
p = 10	-3.960	-3.948	-3.943	-3.941	-3.941	-3.948	-3.958
p = 12	-3.953	-3.938	-3.932	-3.930	-3.932	-3.944	-3.958
p = 14	-3.950	-3.931	-3.924	-3.921	-3.923	-3.939	-3.958

designed to relate to the empirical example presented in the following section. A high value of p was therefore considered, leading to the estimation in all experiments of an ADF(13) test with constant and trend as given below:

$$\Delta y_t = \alpha + \beta t + \phi y_{t-1} + \sum_{i=1}^{13} \gamma_i \Delta y_{t-i} + \eta_t \tag{4}$$

The rejection of unit root hypothesis ($\phi = 0$) was noted at the 1 per cent, 5 per cent and 10 per cent levels of significance (α = 0.01, 0.05, 0.1) using MK and CL critical values. To examine power, near-integration was introduced by setting $|\rho| < 1$ in the AR(1) model given by (3). The values selected for ρ were $\rho \in \{0.7, 0.75, 0.8, 0.85, 0.9, 0.95\}$. To allow comparison with the later empirical example, a sample size of T = 214 was chosen. All experiments were performed over 50,000 replications, with the first 100 observations generated subsequently discarded to remove the impact of initial effects. The initial value was set equal to zero ($y_0 = 0$).

The results of the Monte Carlo experiments are given in Table 4. From inspection of Table Four it can be seen that for all values of ρ and at all levels of significance, the power of the ADF test is greater when using CL rather than MK critical values. Therefore as expected the use of CL critical values increases the rejection the false unit root hypothesis. To illustrate this issue, the power of the ADF test at the 5 per cent level of significance for ρ = 0.95 increases by 14.4 per cent when CL rather than MK critical values are employed.

Table 4: Power of the ADF test: Empirical rejection frequencies

ρ	Response surface	α = 0.1	α = 0.05	α = 0.01
0.7	MK	0.799	0.634	0.284
	CL	0.822	0.670	0.323
0.75	MK	0.750	0.575	0.239
	CL	0.777	0.610	0.273
0.8	MK	0.678	0.494	0.185
	CL	0.709	0.533	0.215
0.85	MK	0.572	0.388	0.129
	CL	0.605	0.424	0.152
0.9	MK	0.418	0.257	0.070
	CL	0.451	0.286	0.084
0.95	MK	0.221	0.118	0.024
	CL	0.245	0.135	0.030

5. An Empirical Application

To highlight the significance of lag adjustment empirically, the critical values of MK and CL are applied to examine whether the M3 definition of the US money stock possesses a unit root. The data selected are seasonally adjusted, monthly observations from May 1967 to April 1986 obtained from the Federal Reserve Bank of St. Louis.[7] The results of applying a 'constant and trend' ADF test are presented in Table 5. Given the frequency of the data, a maximum lag order of 13 periods is considered.[8] To ease interpretation of the results presented and allow the MK critical values to remain fixed, a consistent sample size (T = 214) was employed throughout by omitting an initial observation as the lag order decreased. Examining the results for $p = \{11, 12, 13\}$, MK critical values fail to reject the unit root null hypothesis. Conversely, CL critical values reject the null for $p = 11, 13$, although not for $p = 12$. The alternative critical values therefore provide conflicting results, illustrating the non-stationarity bias present in the greater (absolute) values provided by the MK response surface.

Table 5: ADF(p) tests of US M3

p	ADF(p)	Cheung & Lai ($\alpha = 0.05$)	MK ($\alpha = 0.05$)
13	-3.402*	-3.369	-3.432
12	-3.287	-3.373	-3.432
11	-3.380*	-3.377	-3.432

*denotes rejection of the unit root hypothesis using Cheung & Lai critical values.

6. Conclusion

The lag order dependence of finite-sample ADF critical values has been considered via a comparison of the alternative response surfaces of MacKinnon (1991) and Cheung and Lai (1995). Via a simple numerical exercise it was found that in many cases the alternative critical values converge upon their respective asymptotic values from different directions. This reinforces the non-stationarity bias associated with the use of non-lag augmented critical values. In many cases it was also found that the critical values could take very different values under the two approaches. This issue was further explored via Monte Carlo simulation where it was found that the use of lag-adjusted critical values increases the power of the ADF to reject a false null of non-stationarity. Finally an empirical example was provided which illustrated how the alternative critical values could lead to a difference in inferences being drawn. In summary the results show that despite practitioners typically using MK response surfaces to calculate ADF critical values, there are alternative, statistically superior response surfaces available which increase the power of the ADF test and can result in different inferences being made. Given the negligible computational burden involved in calculating lag adjusted values, the results suggest that the CL critical values should receive more widespread adoption.

Endnotes

1. Dr Steven Cook, Department of Economics, University of Wales Swansea, Singleton Park, Swansea, SA2 8PP. E-mail: s.cook@swan.ac.uk. I am grateful to Dr Abbott and two referees for numerous useful comments which have helped improve the content and presentation of this paper. However, the usual disclaimer applies.

2. The results of CL relate solely to testing for a unit root, whereas MK also provides critical values for Engle-Granger (1987) two-step tests of cointegration.

3. The 'constant and trend' ADF test is the focus of the present study as this has been suggested by CL (1995, p.280) to be the case where critical values are most sensitive to lag order and is also the form applied when analysing trending macroeconomic time series.

4. The size-power trade-off when considering low and high values of p has led to a large literature on the 'optimal' lag order for ADF tests. Ng and Perron (1995) present and discuss a range of alternative decision rules.

5. Maddala and Kim (1998) provide a useful summary of the 'power and frequency of observation' literature.

6. As an example, the default ADF lag orders of PC-GIVE (see Hendry and Doornik, 1996) are 2 for annual data, 5 for quarterly data and 13 for monthly data.

7. As noted by a referee, the impact of structural breaks should be recognised when performing unit root tests. It should be noted that the series under examination does not experience a structural break in the sample period considered. Details on this issue are available upon request.

8. Although this appears to be a relatively high value of p, the Schwert (1989) criterion suggests a value of 15.

REFERENCES

Banerjee A, Dolado J, Galbraith J and Hendry D. (1993) *Co-integration, Error-correction, and the Econometric Analysis of Non-Stationary Data,* Oxford: Oxford U P.

Cheung L-W and Lai K S (1995) 'Lag order and critical values of the Augmented Dickey-Fuller test', *Journal of Business and Economic Statistics*, 13, 277-280.

Choi I and Chung B (1995) 'Sampling frequency and the power of tests for a unit root: a simulation study', *Economics Letters*, 49, 131-136.

DeJong D, Nankervis J, Savin G and Whiteman C (1992) 'The power problems of unit root tests for time series with autoregressive errors', *Journal of Econometrics*, 53, 323-343.

Dickey D and Fuller W (1979) 'Distribution of the estimators for au-toregressive time series with a unit root', *Journal of the American Statistical Association,* 74, 427-431.

Engle R and Granger C (1987) 'Cointegration and error correction: representation, estimation and testing', *Econometrica*, 55, 251-276.

Harris R (1992) 'Testing for unit roots using the Augmented Dickey-Fuller test: Some issues relating to size, power and lag structure of the test', *Economics Letters*, 38, 381-386.

Hendry D and Doornik J (1996) *Empirical Econometric Modelling using PC-GIVE 9.0 for Windows*, London: International Thomson Business Press.

MacKinnon J (1991) 'Critical values for cointegration tests', in Engle R and Granger C (eds) *Long Run Economic Relationships*, Oxford: Oxford U P.

Maddala G and Kim I-M (1998) *Unit roots, Cointegration and Structural Change*, Cambridge: Cambridge U P.

Ng S (1995) 'Testing for unit roots in flow data sampled at different frequencies', *Economics Letters*, 47, 237-242.

Ng S and Perron P (1995) 'Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag', *Journal of the American Statistical Association*, 90, 268-281.

Schwert G (1989) Tests for unit roots: a Monte Carlo investigation', *Journal of Business and Economic Statistics*, 7, 147-159.