OXFORD BULLETIN OF ECONOMICS AND STATISTICS, 57, 3 (1995)
0305-9049 $3.00

PRACTITIONERS CORNER

Lag Order and Critical Values of a Modified Dickey-Fuller Test

Yin-Wong Cheung and Kon S. Lai

I. INTRODUCTION

Unit-root nonstationarity in economic time series has been a hotly contested issue. A widely used unit-root is the augmented Dickey-Fuller or ADF test (Dickey and Fuller (1979)). Derived from an autoregressive AR($p+1$) representation, the test examines the null hypothesis of a unit root against stationary alternatives. Since the null hypothesis maintained is a nonstationary process, empirical failures to find stationarity may reflect the power of the test.

Elliott, Rothenberg, and Stock (1992) devise a new unit-root test with good power. In studying the asymptotic power envelope for various unit-root tests, these authors propose a simple modification of the ADF test such that the modified test, referred to as the DF-GLS test, can nearly achieve the power envelope. The DF-GLS test is shown to be approximately uniformly most power invariant (UMPI) while no strictly UMPI test exists. Monte Carlo results reported indicate that the power improvement from using the modified Dickey-Fuller test can be large. Elliot, Rothenberg, and Stock (1992) derive the limiting distribution of the test with and without a time trend. Approximate finite-sample critical values for the test with a time trend are tabulated for several sample sizes based on $p = 0$.

The purpose of this study is twofold. First, it demonstrates the significant effect of lag order on finite-sample critical values of the DF-GLS test. Empirical applications of the test necessarily deal with finite samples. The result here suggests that a same set of critical values should not be applied to tests with different values of p. Second, the study provides estimates of finite-sample critical values that correct for the effect of lag order. Lag-adjusted critical values can be computed directly from response surface equations. The response surface analysis is useful in that it yields estimates of critical values not only for a few specific sample sizes but a full range of sample sizes.

Response surface analysis has been used by MacKinnon (1991) to obtain approximate finite-sample critical values for the conventional ADF test. In estimating the ADF critical values, p is assumed to be fixed and equal to zero only. Consequently, the potential sensitivity of critical values to the lag order has not been considered and allowed for. Cheung and Lai (1995) extend the response surface analysis and show that, although the asymptotical distribution of the ADF test does not depend on the lag parameter, its finite-sample distribution can be rather sensitive to the parameter. Computationally simple response surface equations are reported, yielding lag-adjusted critical values for the ADF test. In related research, Cheung and Lai (1993) provided finite-sample critical values for Johansen's (1991) reduced rank cointegration tests by taking into account their dependence on the sample size and also the lag order.

The paper is organized as follows: Section II briefly describes the DF-GLS test. Section III discusses the experimental design and reports results of response surface estimation of critical values for the DF-GLS test. Section IV provides some empirical examples illustrating the use of lag-adjusted finite-sample critical values. Section V contains concluding remarks.

II. THE DF-GLS TEST

Elliott, Rothenberg, and Stock (1992) — hereafter referred to as ERS — obtain the asymptotic power envelope for unit-root tests by analyzing the sequence of Neyman-Pearson tests of the unit-root null hypothesis ($\alpha = 1$) against the local alternative of $\bar{\alpha} = 1 + \bar{c}/T$ in the Gaussian AR($p+1$) model, for which T is the sample size and $\bar{c}$ is some constant. Based on asymptotic power calculation, it is shown that a modified Dickey-Fuller test, called the DF-GLS test, can achieve a substantial gain in power over conventional unit-root tests.

Let $\{y_t\}$ be the data process under examination. The DF-GLS$^\tau$ test that allows for a linear time trend is conducted based on the following regression:

$$(1-L)y_t^\tau = \alpha_0 y_{t-1}^\tau + \Sigma_{j=1}^p a_j(1-L)y_{t-j}^\tau + u_t \tag{1}$$

where L is the lag operator; u_t is a white noise error term; and y_t^τ, the locally detrended data process under the local alternative of $\bar{\alpha}$, is given by

$$y_t^\tau = y_t - z_t\tilde{\beta} \tag{2}$$

with $z_t = (1, t)$ and $\tilde{\beta}$ being the regression coefficient of $\tilde{y}_t$ on $\tilde{z}_t$, for which $\tilde{y}_t = (y_1, \ (1-\bar{\alpha}L)y_2, \ \ldots, \ (1-\bar{\alpha}L)y_T)'$ and $\tilde{z}_t = (z_1, \ (1-\bar{\alpha}L)z_2, \ \ldots, \ (1-\bar{\alpha}L)z_T)'$. The DF-GLS$^\tau$ test statistic is given by the conventional t-statistic testing $a_0 = 0$ against the alternative of $a_0 < 0$ in regression (1). ERS recommend that the parameter $\bar{c}$, which defines the local alternative through $\bar{\alpha} = 1 + \bar{c}/T$, be set equal to -13.5. For the test without a time trend, denoted by DF-GLS$^\mu$, it involves the same procedure as the DF-GLS$^\tau$ test, except that y_t^τ is replaced by the locally demeaned series y_t^μ and $z_t = 1$. In this case, the use of $\bar{c} = -7$ is recommended. The DF-GLS$^\mu$ test shares the same limiting distribution as the usual ADF test in the no-deterministic case. Approximate finite-sample critical values for the DF-GLS$^\tau$ test statistic are tabulated by ERS (1992, Table 1) using simulation for specific samples sizes: 50, 100, 200, and 500. Using response surface analysis, this study illustrates the roles of both the sample size and the lag order in determining the finite-sample critical values of the two DF-GLS tests.

III. LAG-ADJUSTED FINITE-SAMPLE CRITICAL VALUES

Response surface methodology has been used in many fields of applied statistics (Myers, Khuri, and Carter (1989)). Early studies that use the methodology in econometrics include Hendry (1979), Hendry and Harrison (1974), and Hendry and Srba (1977); later developments are reviewed by

TABLE 1
Response Surface Estimation

Coefficients &Statistics	*DF-GLS$^\mu$ Test*		*DF-GLS$^\tau$ Test*	
	10%	*5%*	*10%*	*5%*
τ_0	−1.624	−1.948	−2.550	−2.838
	(0.001)**	(0.002)**	(0.002)**	(0.003)**
τ_1	−19.888	−17.839	−20.166	−20.328
	(0.314)**	(0.438)**	(0.544)**	(0.560)**
τ_2	155.231	104.086	155.215	124.191
	(9.989)**	(13.943)**	(18.558)**	(18.435)**
ϕ_1	0.709	0.802	1.133	1.267
	(0.063)**	(0.086)**	(0.099)**	(0.114)**
ϕ_2	5.480	5.558	9.808	10.530
	(0.688)**	(0.934)**	(1.124)**	(1.257)**
ϕ_3	−16.055	−18.332	−20.313	−24.600
	(1.809)**	(2.441)**	(3.024)**	(3.432)**
R^2	0.991	0.984	0.981	0.979
$\hat{\sigma}_\varepsilon$	0.012	0.015	0.018	0.020
Mean $\lvert\hat{\varepsilon}\rvert$	0.009	0.012	0.014	0.016
Restricted Regression ($\phi_1=\phi_2=\phi_3=0$):				
R^2	0.805	0.780	0.329	0.350
$\hat{\sigma}_\varepsilon$	0.054	0.056	0.187	0.111
Mean $\lvert\hat{\varepsilon}\rvert$	0.039	0.040	0.072	0.077

Notes:
The response surface regression is given by equation (4). The DF-GLS$^\mu$ test is the test without a time trend; whereas, the DF-GLS$^\tau$ test represents the one with a time trend. Corresponding heteroskedasticity-consistent standard errors for coefficient estimates are in parentheses. Statistical significance is indicated by a double asterisk (**) for the 5% level. $\hat{\sigma}_\varepsilon$ represents the standard error of the regression. Mean $\lvert\hat{\varepsilon}\rvert$ gives the mean absolute error of the response surface predictions versus estimated critical values from simulations.

Henry (1984). Response surface analysis applies in general to a system where the response of some variable depends on a set of other variables that can be controlled and measured in experiments. Simulations are conducted to evaluate the effects on the response variable of designed changes in the control variables. A response surface describing the response variable as a function of the control variables is then estimated.

In our analysis, the response variable is the finite-sample critical value of the DF-GLS test and the control variables are the sample size (T) and the lag order (p). To provide a comprehensive coverage of interactions between the control variables, a factorial experimental design is entertained. The factorial design covers 296 different pairings of (T, p), for which $T = \{28, 30, 31, 33, 34, 36, 37, 39, 40, 42, 43, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 600\}$ and $p = \{0, 1, 2, 3, 4, 5, 6, 8\}$. In each experiment for given (T, p), the 10 percent and 5 percent critical values are computed as corresponding percentiles of the empirical finite-sample distribution constructed based on 25,000 replications. The data generating process (DGP) is specified as

$$x_t = x_{t-1} + e_t \tag{3}$$

with e_t being independently distributed standard normal innovations. Sample series of x_t are generated by setting the initial value x_0 equal to zero and creating $T + 50$ observations, of which the first 50 observations are discarded. The GAUSS programming language and the subroutine RNDN are used to generate random normal innovations.

The lag order considered in the regression equation (1) is more general than that implied by the DGP given by equation (3). Higher-order DGPs, for which e_t is autocorrelated, is allowed for in the test provided that the lag order p is large enough to capture the dependence. If the lag order used is too small relative to the true lag order, the error term u_t in the regression will no longer be white noise. In this case, the DF-GLS test can be seriously biased, making estimates of critical values inaccurate.

Although a response surface can assume different functional forms, it needs to satisfy some restriction in the case here. Intuitively, with a given sample size, the choice of lag order can affect the ADF test by determining the effective number of observations available and the number of parameters to be estimated in the test. As the sample size increases to infinity, nonetheless, the effect of p on critical values is expected to diminish to zero. To account for the intuition, together with the asymptotic restriction, the following response surface equation of a general polynomial form is fitted:

$$CR_{T,p} = \tau_0 + \sum_{i=1}^{r} \tau_i (1/T)^i + \sum_{j=1}^{s} \phi_j (p/T)^j + \varepsilon_{T,p} \tag{4}$$

where $CR_{T,p}$ is the critical value estimate for a sample size T and lag p; r and s are the respective polynomial orders for the variables $1/T$ and p/T; and $\varepsilon_{T,p}$ is a random error term. The second summation term captures the incremental contribution from the lag order. Note that the p/T factor diminishes to zero as the value of T goes to infinity. Since both $1/T$ and $p/T \rightarrow 0$ as $T \rightarrow \infty$, the intercept term gives an estimate of the asymptotic critical value.

Table 1 contains the results of response surface regressions. Since tests with and without a linear trend are conducted at both 10 percent and 5 percent significance levels, four response surface regressions are run. A range of different values of r and s have been considered in estimating equation (4). For both the DF-GLS$^{\mu}$ and DF-GLS$^{\tau}$ tests, $r = 2$ and $s = 3$ fit the data particularly well. Both the $1/T$ and p/T variables show up to be significant in all regressions at the 5 percent level. When higher-order polynomial terms were included, they were found to add little to the explanatory power. Various measures of data fit are also computed, including the squared multiple correlation coefficient (R^2), standard error of regression ($\hat{\sigma}_{\varepsilon}$), and mean absolute error (mean $|\hat{\varepsilon}|$). The results given in Table 1 indicate that the ability of the response surface equations to fit the data is very good, in view of the high goodness of fit as measured by R^2. Both measures of $\hat{\sigma}_{\varepsilon}$ and mean $|\hat{\varepsilon}|$ are also fairly small in all four regressions.

To check the explanatory power of the p/T variable, moreover, restricted regressions excluding the terms involving the variable are fitted. As evidenced by the results reported in Table 1, excluding the p/T variable can substantially reduce the fit of response surfaces, resulting in a much lower R^2 as well as much larger values of $\hat{\sigma}_{\varepsilon}$ and mean $|\hat{\varepsilon}|$.

Some finite-sample critical values have been estimated by ERS (1992) for the DF-GLS$^{\tau}$ test based on $p = 0$. It is interesting to compare those estimates directly with the response surface estimates of critical values obtained here, as displayed in Table 2. The estimates provided by ERS are given in the third column. The last four columns contain response surface estimates for $p = 0$, 4, 6, and 8. Not surprisingly, the response surface estimates for $p = 0$ match very closely with the ERS estimates. This is not the case, however, when other values of p are considered. The critical values for $p = 4$, 6, and 8 clearly differ from the ERS estimates. The differences in estimates magnify as the sample size decreases.

IV. ILLUSTRATIVE APPLICATIONS

To illustrate how lag-adjusted critical values can be used in practice, the DF-GLS$^{\tau}$ test is applied to two sets of economic time series. In applying the unit-root test, researchers are sometimes interested in examining the robustness of test results to lag specifications. Since finite-sample critical values are shown to be sensitive to the lag order, critical values that correct for the lag effect should be used so that reliable inferences can be made.

TABLE 2
Lag Order and Finite-Sample Critical Values

Sample Size	Sig. Level	ERS Estimate	Response Surface Estimates of Critical Values $p=0$	$p=4$	$p=6$	$p=8$
50	10%	−2.89	−2.891	−2.748	−2.649	−2.542
	5%	−3.19	−3.195	−3.039	−2.934	−2.823
75	10%	N/A	−2.791	−2.706	−2.648	−2.583
	5%	N/A	−3.087	−2.993	−2.931	−2.862
100	10%	−2.74	−2.736	−2.676	−2.637	−2.593
	5%	−3.03	−3.029	−2.963	−2.920	−2.873
150	10%	N/A	−2.678	−2.641	−2.619	−2.592
	5%	N/A	−2.968	−2.927	−2.902	−2.874
200	10%	−2.64	−2.647	−2.621	−2.605	−2.587
	5%	−2.93	−2.937	−2.907	−2.890	−2.871
500	10%	−2.59	−2.590	−2.580	−2.574	−2.569
	5%	−2.89	−2.878	−2.867	−2.861	−2.855

Notes:
The finite-sample critical values tabulated are for the DF-GLS$^{\tau}$ Test. The column beneath ERS gives the estimates of critical values provided by Elliott, Rothenberg, and Stock (1992); these finite-sample critical values are computed based on $p=0$.

A. *Persistence in Output Dynamics*

An issue concerns the persistence in U.S. output fluctuations, witness studies by Campbell and Mankiw (1987), Cochrane (1988), Diebold and Rudebusch (1989), and Watson (1986), among others. In this example, historical output data for the U.S. as well as France, Germany, and Italy are examined for trend stationarity. The data consist of annual series of real per capita gross domestic product for the period of 1870–1989 (120 observations). The data through 1979 are taken from Maddison (1982) and updated to 1989 based on data from the OECD *Main Economic Indicators*. All data series are expressed in natural logarithms.

Table 3A reports the results of the DF-GLS$^{\tau}$ test, along with the corresponding critical values for $p=0$, 2, 4, 6, and 8. For France, Germany, and Italy, the null of a unit root cannot be rejected at any standard level of significance. In contrast, significant evidence of trend stationarity can be found for the U.S. output series. These results are shown to be robust with respect to the choice of lag order. The findings on the whole suggest that the U.S. experience is unique in showing much less output persistence than the other countries.

B. *Purchasing Power Parity*

According to the purchasing power parity theory, currencies are valued for the goods they can purchase and, in equilibrium, the exchange rate between two countries' currencies should be equal to their relative price levels. A testable implication is that real exchange rates should display mean reversion at least in the long run. In the present example, three currencies against the U.S. $ are examined, those of French franc/$, German mark/$, and Italian lira/$. The data under study are annual real exchange rate data constructed from nominal exchange rates and national price levels measured by consumer prices indices (CPIs). The data sample covers the period from 1900 through 1990 (91 observations). Data on annual exchange rates are obtained from Lee (1978), and annual CPIs from 1900 to 1989 are drawn from Maddison (1991). These data series are updated and extended to 1990 using annual data taken from the IMF's *International Financial Statistics* data tape. All the series of real exchange rates are transformed using natural logarithms.

Table 3B contains the results of the DF-GLS$^{\tau}$ test for real exchange rates. In all cases under examination, the null hypothesis of trend nonstationarity can be rejected. Accordingly, significant evidence of mean reversion is found

TABLE 3
Illustrative Applications of the DF-GLS$^{\tau}$ Test

	Lag Used in the Test				
	$p=0$	$p=2$	$p=4$	$p=6$	$p=8$
A. Real Per Capita GDP ($T=120$):					
France	−0.232	−1.196	−1.144	−0.632	−0.760
Germany	−1.585	−1.351	−1.284	−1.149	−0.954
Italy	−0.475	−0.381	−0.465	−0.305	−0.420
U.S.	−2.915*	−3.918**	−3.162**	−3.270**	−3.222**
Critical Value (10%)	−2.707	−2.686	−2.659	−2.629	−2.594
Critical Value (5%)	−2.999	−2.975	−2.946	−2.912	−2.875
B. Real Exchange Rates ($T=91$):					
France/U.S.	−3.674**	−4.124**	−3.836**	−2.661*	−2.641*
Germany/U.S.	−11.699**	−6.702**	−5.103**	−4.405**	−3.466**
Italy/U.S.	−3.470**	−3.398**	−3.028**	−2.724*	−2.626*
Critical Value (10%)	−2.753	−2.723	−2.686	−2.641	−2.591
Critical Value (5%)	−3.046	−3.014	−2.972	−2.924	−2.870

Notes:
Lag-adjusted finite-sample critical values for the DF-GLS$^{\tau}$ test applied are computed from the response surface equations given in Table 1. Statistical significance in indicated by an asterisk (*) for the 10% level or a double asterisk (**) for the 5% level.

in real exchange rates, in accordance with purchasing power parity. Note that the findings are shown to be not sensitive to the lag order. Consider the Italy/U.S. case closely. The DF-GLS$^{\tau}$ test statistics are given by -2.724 for $p=6$ and -2.626 for $p=8$. These statistics would not be significant at the 10 percent level if critical values not adjusted for the lag effect — i.e., the critical values for $p=0$ — were used. A similar situation applies to the France/U.S. case. The DF-GLS$^{\tau}$ test statistics for $p=6$ and $p=8$ would not be significant at the 10 percent level if non-lag-adjusted critical values were used. Such situation highlights the virtue of using lag-adjusted critical values.

V. CONCLUSION

Usual practice of applying the Dickey-Fuller-type test in empirical work has largely ignored the effect of lag order. This is often justified by the asymptotic result that the limiting distribution of the test is free of the lag parameter. The practice may not be appropriate, however, because the test can be sensitive to the lag order in finite samples, with which empirical applications always deal.

This study examines the application of a new modified Dickey-Fuller test, the DF-GLS test, which has been shown to be more powerful than conventional unit-root tests. The study shows that the lag order can significantly affect the critical values of the test in finite samples. This points to the importance of correcting for the lag order effect in implementing the DF-GLS test. Approximate lag-adjusted finite-sample critical values for the DF-GLS test are provided, which can be straightforwardly computed from response surface equations.

University of California, Santa Cruz
California State University, Los Angeles

Date of Receipt of Final Manuscript: June 1994

REFERENCES

Campbell, J. Y. and Mankiw, N. G. (1987). 'Are Output Fluctuations Transitory?', *Quarterly Journal of Economics*, Vol. 102, pp. 857–80.

Cheung, Y. W. and Lai, K. S. (1993). 'Finite-Sample Sizes of Johansen's Likelihood Ratio Tests for Cointegration', BULLETIN, Vol. 55, pp. 313–28.

Cheung, Y. W. and Lai, K. S. (1995). 'Lag Order and Critical Values of the Augmented Dickey-Fuller Test', *Journal of Business and Economic Statistics*, Vol. 13 (forthcoming).

Cochrane, J. H. (1988). 'How Big is the Random Walk in GNP?', *Journal of Political Economy*, Vol. 96, pp. 893–920.

Diebold, F. X. and Rudebusch, G. D. (1989). 'Long Memory and Persistence in Aggregate Output', *Journal of Monetary Economics*, Vol. 24, pp. 189–209.

Dickey, D. A. and Fuller, W. A. (1979). 'Distribution of the Estimators for Autoreggresive Time Series with a Unit Root', *Journal of American Statistical Association*, Vol. 74, pp. 427–31.

Elliott, G., Rothenberg, T. J. and Stock, J. H. (1992). 'Efficient Tests for an Autoregressive Unit Root', National Bureau of Economic Research Technical Working Paper No. 130.

Hendry, D. F. (1979). 'The Behaviour of Inconsistent Instrumental Variables Estimators in Dynamic Systems with Autocorrelated Errors', *Journal of Econometrics*, Vol. 9, pp. 295–314.

Hendry, D. F. (1984). 'Monte Carlo Experimentation in Econometrics', in Griliches, Z. and Intriligator, M. D. (eds.), *Handbook of Econometrics*, Vol. 2, North-Holland, Amsterdam, pp. 937–76.

Hendry, D. F. and Harrison, R. W. (1974). 'Monte Carlo Methodology and the Small Sample Behaviour of Ordinary and Two-Stage Least Squares', *Journal of Econometrics*, Vol. 2, pp. 151–74.

Hendry, D. F. and Srba, F. (1977). 'The Properties of Autoregressive Instrumental Variables Estimators in Dynamic Systems', *Econometrica*, Vol. 45, pp. 969–90.

Johansen, S. (1991). 'Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models', *Econometrica*, Vol. 59, pp. 1551–80.

Lee, M. K. (1978). *Purchasing Power Parity*, Marcel Dekker, New York.

MacKinnon, J. G. (1991). 'Critical Values for Cointegration Tests', in Engle, R. F. and Granger, C. W. J. (eds), *Long-Run Economic Relationships: Readings in Cointegration*, Oxford University Press, New York.

Maddison, A. (1982). *Phases of Capitalist Development*. Oxford University Press, New York.

Maddison, A. (1991). *Dynamic Forces in Capitalist Development*, Oxford University Press, New York.

Myers, R. H., Khuri, A. I. and Carter, W. H. (1989). 'Response Surface Methodology: 1966–1988', *Technometrics*, Vol. 31, pp. 137–57.

Watson, M. W. (1986). 'Univariate Detrending Methods with Stochastic Trends', *Journal of Monetary Economics*, Vol. 18, pp. 49–75.