Journal of Business & Economic Statistics, July 1995, Vol. 13, No. 3

Lag Order and Critical Values of the Augmented Dickey–Fuller Test

Yin-Wong Cheung
Department of Economics, University of California, Santa Cruz, CA 95064

Kon S. Lai
Department of Economics and Statistics, California State University, Los Angeles, CA 90032

Response surface analysis is used to obtain approximate finite-sample critical values for the augmented Dickey–Fuller (ADF) test. Previous studies estimating the critical values for the test have generally ignored their possible dependence on the lag order. This study shows that the lag order, in addition to the sample size, can affect the finite-sample behavior of the test. The result points to the importance of correcting for the effect of lag order in applying the ADF test.

KEY WORDS: Finite-sample critical value; Monte Carlo; Response surface.

The augmented Dickey–Fuller (or ADF) test is a commonly used unit-root test. Fitting an (autoregressive) AR(k) model, this test examines the null hypothesis of an (autoregressive integrated moving average) ARIMA(p, 1, 0) process against the stationary ARIMA(p+1, 0, 0) alternative. Dickey and Fuller (1979) derived the limiting distribution of the ADF test when $p \leq k - 1$. Approximate critical values for the test with $k = 1$ were tabulated by Fuller (1976) for specific sample sizes. For the case of $k > 1$, the ADF test has the same limiting distribution as when $k = 1$, provided that the condition $p \leq k - 1$ holds. Although this is an asymptotic result, the critical values tabulated by Fuller (1976) have often been applied to tests with arbitrary values of k in finite samples.

Using response surface analysis, MacKinnon (1991) provided finite-sample critical values for the ADF test. The analysis yields estimates of critical values not for only a few sample sizes but any sample size. These critical values, like those of Fuller (1976), are based on the ADF test with $k = 1$ only. Through extensive response surface estimation of quantiles, MacKinnon (1994) further showed that an approximate asymptotic distribution function for the test can be obtained.

Although the distribution of the ADF test statistic does not depend on the lag order asymptotically, it can be sensitive to the lag order in finite samples, with which empirical applications necessarily deal. This study examines the individual roles of the sample size and the lag order in determining the finite-sample critical values of the ADF test. By properly accounting for the effect of lag order, this study extends MacKinnon's (1991) and provides improved estimates of critical values of the ADF test. MacKinnon's (1994) analysis may check the accuracy of results on asymptotic distributions derived analytically, and the lag order k is controlled and fixed at unity in his analysis. In contrast, the present study focuses the analysis on finite-sample distributions and demonstrates that critical values previously available are biased for ignoring their dependence on the lag order.

1. THE AUGMENTED DICKEY–FULLER TEST

Let x_t be a time series. Deriving from an AR(k) representation, the ADF test involves the following regression:

$$\Delta x_t = \mu + \gamma t + \alpha x_{t-1} + \sum_{j=1}^{k-1} \beta_j \Delta x_{t-j} + u_t, \tag{1}$$

where Δ is the difference operator and u_t is a white-noise innovation. The test examines the negativity of the parameter α based on its regression t ratio. Dickey and Fuller (1979) derived the asymptotic distribution of the statistic. Hall (1994) showed that the asymptotic distribution is unaffected by data-based model selection using standard information criteria. To the extent that the distribution can be sensitive to the lag order in finite samples, there remains the problem of applying appropriate lag-adjusted critical values.

2. FINITE–SAMPLE CRITICAL VALUES

Response surface methodology has been used in many fields of applied statistics (Myers, Khuri, and Carter 1989). Early studies using the methodology in econometrics include Hendry (1979), Hendry and Harrison (1974), and Hendry and Srba (1977); references for later work were provided by Ericsson (1991) and Hendry (1984). Cheung and Lai (1993a) estimated finite-sample critical values for reduced-rank cointegration tests by taking into account their dependence on the lag order. Computationally simple response surface equations were obtained, yielding critical values that correct for the lag-order effect (see also Cheung and Lai in press).

Response surface analysis applies in general to a system in which the response of some variable depends on a set of other variables that can be controlled and measured in experiments. Simulations are conducted to assess 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 ones is then estimated.

In our analysis, the response variable is the finite-sample critical value of the ADF test, and the control variables are the sample size (N) and the lag order (k). A factorial experimental design is employed, covering 228 different pairings of $N = \{18, 20, 22, 25, 27, 30, 33, 36, 39, 42, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 250, 300, 350, 400, 500\}$ and $k = \{1, 2, 3, 4, 5, 6, 7, 9\}$. For $N \leq 25$, $k \leq 5$ is used. The data-generating process (DGP) is specified as

$$x_t = x_{t-1} + e_t, \quad (2)$$

with e_t being independently distributed standard normal innovations. Setting the variance of e_t equal to unity is without loss of generality for determining critical values. Sample series of x_t are generated by setting the initial value x_0 equal to 0 and creating $N + 50$ observations, of which the first 50 observations are discarded. Pseudorandom normal variates are generated using the GAUSS subroutine RNDN (the simulation program is available on request). For a given pairing of (N, k), all of the 10%, 5%, and 1% critical values are computed as quantiles directly using 30,000 replications (40,000 for $N \leq 30$) in an experiment, though different sets of random numbers are used across experiments.

The regression model given by (1) is more general than the DGP considered. Higher-order DGP's, for which e_t is autocorrelated, will contain additional nuisance parameters. Our experimental design generalizes MacKinnon's (1991) by including k but still omits those other nuisance parameters. Finite-sample correction for the latter—albeit desirable—is hard to make, given the potential size of the parameter space of these unknown parameters. Cheung and Lai (1993b) examined the sensitivity of finite-sample critical values of eight different DGP's with AR or MA dependence in e_t. Consistent with asymptotic results of Said and Dickey (1984), Cheung and Lai (1993b) found that size distortion can be small, provided that k is large enough to capture the dependence.

To check the accuracy in estimating finite-sample critical values, standard errors of the Monte Carlo-estimated critical values are calculated based on Rohatgi (1984, pp. 496–500). In general, and as expected, tests with larger samples tend to have smaller estimated standard errors (ESE). For the 10%, 5%, and 1% test, the ranges of ESE are [.006, .013], [.008, .018], and [.012, .037], respectively.

Selecting the functional form for the response surface is not entirely arbitrary and may need to satisfy some restrictions. In our case, the effects of k on critical values should diminish to 0 as N goes to infinity. After much experimentation with alternative functional forms, a second-order polynomial equation—which nests MacKinnon's (1991) and satisfies the asymptotic restriction—is found to fit the simulation data well. The response surface equation is given as follows:

$$CR_{N,k} = \tau_0 + \sum_{i=1}^{2} \tau_i (1/T)^i + \sum_{j=1}^{2} \phi_j [(k-1)/T]^j + \epsilon_{N,k}, \quad (3)$$

Table 1. Response Surface Estimation of Critical Values for the ADF Statistic

Coefficients & statistics	No constant or trend 10%	5%	1%	Constant, no trend 10%	5%	1%	Constant and trend 10%	5%	1%
τ_0	−1.609	−1.931	−2.564	−2 566	−2.857	−3.430	−3.122	−3.406	−3 958
	(.001)*	(.001)*	(.003)*	(.001)*	(.002)*	(.004)*	(.002)*	(.003)*	(.005)*
τ_1	−.285	−1.289	−2.906	−1.319	−2.675	−4.959	−2.850	−4.060	−7.448
	(.143)*	(162)*	(.242)*	(.205)*	(.269)*	(.540)*	(.373)*	(.464)*	(.886)*
τ_2	−4.090	−5.719	−29 773	−15.086	−23.558	−72.303	−15.813	−40.552	−104.947
	(2.273)+	(2.434)*	(4.112)*	(3.900)*	(4.923)*	(10.133)*	(7.200)*	(9.122)*	(16.919)*
ϕ_1	.321	.380	.599	.667	.748	.842	.907	1.021	1.327
	(.024)*	(.032)*	(.061)*	(.035)*	(.046)*	(.072)*	(.058)*	(.069)*	(150)*
ϕ_2	−.525	−.722	−1.580	−.650	−1.077	−2.090	−.804	−1.501	−3.753
	(.066)*	(.082)*	(.168)*	(.093)*	(.128)*	(.227)*	(.171)*	(.207)*	(.584)*
R^2	.647	.770	.894	.877	.924	.968	.866	.930	.969
$\hat{\sigma}_\epsilon$	.009	.012	.026	.012	.015	.026	.020	.023	.039
Mean $\|\hat{\epsilon}\|$	.007	.009	.021	.009	.011	.020	.015	.018	.029
Max $\|\hat{\epsilon}\|$	.022	.038	.087	.046	.049	.096	.068	.086	.105
*Mean $\|\hat{\epsilon}\|$	.006	.009	.020	.008	.010	.019	.012	.014	.025
*Max $\|\hat{\epsilon}\|$	.021	.036	.086	.029	.045	.070	.042	.045	.099
				Restricted regression (MacKinnon 1991)					
R^2	.100	.572	.864	.305	.753	.915	.417	.812	.903
$\hat{\sigma}_\epsilon$	.014	.016	.029	.029	.027	.042	.041	.038	.068
Mean $\|\hat{\epsilon}\|$	.011	.013	.022	.022	.021	.028	.031	030	.040
Max $\|\hat{\epsilon}\|$	.039	.054	.130	.081	.079	.279	.115	.104	.557
*Mean $\|\hat{\epsilon}\|$	.011	.013	.021	.021	.021	.023	.029	.028	.030
*Max $\|\hat{\epsilon}\|$	.039	.054	.102	.081	.079	.129	.115	.091	.189

NOTE The response surface regression is given by Equation (3) Heteroscedasticity-consistent standard errors are in parentheses Significance is indicated by + for the 10% level and by * for the 5% level $\hat{\sigma}_\epsilon$ indicates the estimated standard error of the regression Mean $|\hat{\epsilon}|$ gives mean absolute error of the response surface predictions versus estimated critical values, whereas max $|\hat{\epsilon}|$ gives maximum absolute error The no-constant-or-trend, constant-no-trend, and constant-and-trend statistics for the ADF test were respectively referred to as $\hat{\tau}$, $\hat{\tau}_\mu$, and $\hat{\tau}_\tau$ by Fuller (1976)

* Computed from residual errors corresponding to $T \geq 30$ (too small T can lead to large errors $\hat{\epsilon}$)

where $CR_{N,k}$ is the critical value estimate for a sample size N and lag k, $T = N - k$ indicates the effective number of observations, and $\epsilon_{N,k}$ is the error term. When higher-order polynomial terms are included, they add little to the explanatory power. Equation (3) includes MacKinnon's (1991) as a special case in which k is fixed and equal to 1. Note that the $(k - 1)/T$ factor diminishes to 0 as the value of N increases to infinity. Because both $1/T$ and $(k-1)/T \rightarrow 0$ as $N \rightarrow \infty$, the intercept term (τ_0) provides an estimate for the asymptotic critical value.

Response surface regressions are reported in Table 1 for three versions of the ADF test at the 10%, 5%, and 1% levels. Various measures of data fit are computed, including the squared multiple correlation coefficient (R^2), standard error of regression ($\widehat{\sigma}_\epsilon$), mean absolute error (mean $|\widehat{\epsilon}|$), and maximum absolute error (max $|\widehat{\epsilon}|$). In all the cases, both the $1/T$ and $(k - 1)/T$ variables are statistically significant. Excluding the $(k - 1)/T$ terms, as in MacKinnon's (1991) restricted regression, generally reduces the fit of the response surface. The reduction can be substantial in some cases. The Reinsel–Ahn approximation for the response surface, explored by Cheung and Lai (1993a), has also been considered but it fails to yield an improved fit over Equation (3).

It is interesting to compare directly the estimates of critical values here with those reported by MacKinnon (1991) and Fuller (1976). In general, they are found to match closely with one another when $k = 1$. Consider the test with a time trend, for example. Our estimated critical values are −3.187 (10%), −3.506 (5%), and −4.154 (1%) for a sample size of 50 and −3.152 (10%), −3.451 (5%), and −4.044 (1%) for a sample size of 100. Those from MacKinnon (1991) are −3.180 (10%), −3.502 (5%), and −4.154 (1%) for a sample size of 50 and −3.153 (10%), −3.455 (5%), and −4.053 (1%) for a sample size of 100. Both sets of estimates are very close to Fuller's (1976)— −3.18 (10%), −3.50 (5%), and −4.15 (1%) for a sample size of 50 and −3.15 (10%), −3.45 (5%), and −4.04 (1%) for a sample size of 100.

The close match disappears, however, when $k > 1$ and lag-adjusted critical values are needed. Consider again the ADF test with a time trend but with $k = 6$. The estimates of lag-adjusted critical values are −3.102 (10%), −3.423 (5%), and −4.079 (1%) for a sample size of 50 and −3.108 (10%), −3.404 (5%), and −3.989 (1%) for a sample size of 100. These estimates clearly differ from those from MacKinnon (1991) and Fuller (1976) with no lag adjustments. Note that the differences between the estimates become smaller in larger sample sizes. This follows from the property that the response surface is a function of $(k - 1)/T$.

Finally, the Monte Carlo-estimated critical values, CR's, are plotted in Figure 1 for various ADF tests. These three-

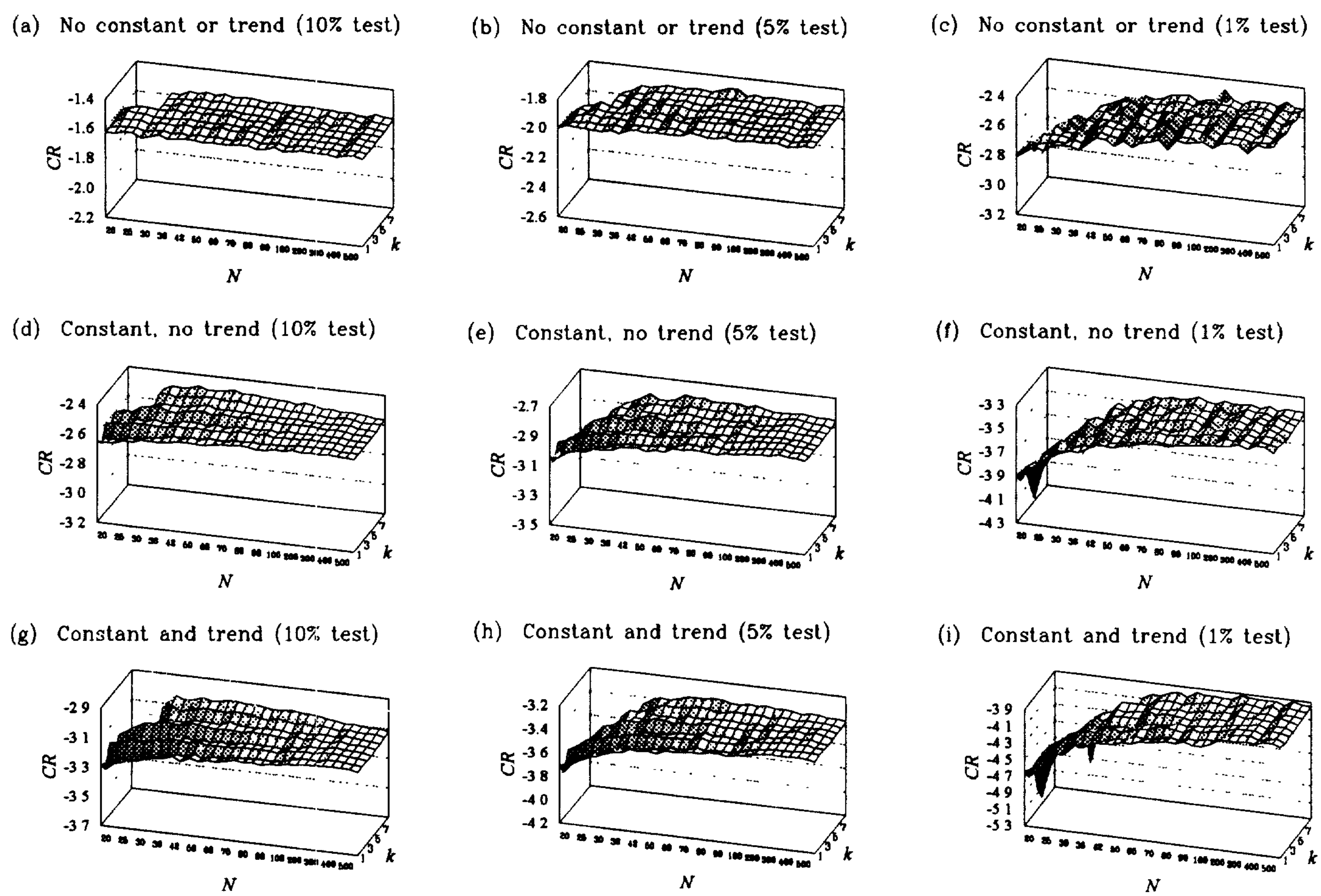

Figure 1. Plots of Monte Carlo-Estimated Critical Values for Various ADF Tests. N is the sample size, and k is the lag order parameter. In each graph, the vertical axis gives the Monte Carlo-estimated critical values corresponding to different combinations of N and k.

dimensional graphs provide a sense of the numerical fluctuations in the critical values as a function of the sample size N and the lag order k. Following Tufte's (1983, 1990) idea of "small multiples," the nine graphs are arranged in a 3×3 matrix to allow efficient comparison across types of tests and across test sizes. The graphs show the presence of an unambiguously signed sample-size correction to the asymptotic values when $k = 1$. This is consistent with the fact that both τ_1 and τ_2, which determine the pure sample-size effect, are negatively signed in all response surfaces. The effect of additional lags may be ambiguously signed, however: $\phi_1 > 0$, but $\phi_2 < 0$ in all response surfaces, implying some balancing effects from $(k-1)/T$ and $[(k-1)/T]^2$. This is apparent in many of the graphs.

A comparison between graphs also shows the different speeds at which finite-sample critical values can approach their asymptotic levels. With a given test size, critical values for the test with constant and trend approach their asymptotic limits most slowly, whereas those with no constant or trend approach most rapidly. For the latter, critical values for a 10% test with no constant or trend are, indeed, nearly invariant to N and k. This contrasts most sharply with the case of a 1% test with constant and trend, for which critical values can be very sensitive to N and k.

3. CONCLUSION

In this study, response surface analysis is used to obtain approximations to the finite-sample critical values for the ADF test. Previous studies estimating the critical values for the ADF test have largely ignored their possible dependence on the lag order. This study shows that the lag order, in addition to the sample size, can affect the finite-sample behavior of the ADF test. Proper correction for the lag effect in implementing the ADF test is therefore desirable. Because appropriate critical values for the ADF test can be easily computed with reasonable accuracy from response surface equations for any sample size and lag length, the analysis here should be useful for researchers in practical applications.

ACKNOWLEDGMENTS

This article has benefited from valuable comments from Alastair Hall, two anonymous referees, an associate editor, and the editor of this journal. All remaining errors are ours. The research was supported by the GICES at the University of California at Santa Cruz.

[*Received August 1993. Revised September 1994.*]

REFERENCES

Cheung, Y.-W., and Lai, K. S. (1993a), "Finite-Sample Sizes of Johansen's Likelihood Ratio Tests for Cointegration," *Oxford Bulletin of Economics and Statistics*, 55, 313–328.

——— (1993b), "Lag Order and the Finite-Sample Behavior of the Augmented Dickey–Fuller Test," Working Paper 269, University of California at Santa Cruz, Dept. of Economics.

——— (in press), "Lag Order and Critical Values of a Modified Dickey–Fuller Test," *Oxford Bulletin of Economics and Statistics*, 57

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.

Ericsson, N. R. (1991), "Monte Carlo Methodology and the Finite Sample Properties of Instrumental Variables Statistics for Testing Nested and Non-nested Hypotheses," *Econometrica*, 59, 1249–1277

Fuller, W. A. (1976), *Introduction to Statistical Time Series*, New York John Wiley.

Hall, A. (1994), "Testing for a Unit Root in Time Series With Pretest Data-Based Model Selection," *Journal of Business & Economic Statistics*, 12, 461–470.

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

——— (1984), "Monte Carlo Experimentation in Econometrics," in *Handbook of Econometrics* (Vol. 2), eds. Z. Griliches and M D Intriligator, Amsterdam: North-Holland, pp 937–976.

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*, 2, 151–174.

Hendry, D. F., and Srba, F. (1977), "The Properties of Autoregressive Instrumental Variables Estimators in Dynamic Systems," *Econometrica*, 45, 969–990.

MacKinnon, J G. (1991), "Critical Values for Cointegration Tests," in *Long-Run Economic Relationships: Readings in Cointegration*, eds. R. F. Engle and C. W. J. Granger, New York: Oxford University Press, pp. 266–276

——— (1994), "Approximate Asymptotic Distribution Functions for Unit-Root and Cointegration Tests," *Journal of Business & Economic Statistics*, 12, 167–176.

Myers, R. H., Khuri, A. I, and Carter, W. H. (1989), "Response Surface Methodology: 1966–1988," *Technometrics*, 31, 137–157.

Rohatgi, V. K. (1984), *Statistical Inference*, New York: John Wiley.

Said, S., and Dickey, D. A. (1984), "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," *Biometrika*, 71, 599–607.

Tufte, E. R. (1983), *The Visual Display of Quantitative Information*, Cheshire, CT: Graphics Press.

——— (1990), *Envisioning Information*, Cheshire, CT: Graphics Press.