Inflation Revisited: New Evidence from Modified Unit Root Tests

Walter Enders and Yu Liu*

University of Alabama in Tuscaloosa and University of Texas at El Paso

---

Abstract: We propose a simple modification to the general-to-specific lag-length selection method typically employed in a standard augmented Dickey-Fuller (ADF) test and apply it to examine the stationarity of OECD countries' inflation rates. Instead of using the entire set of lags selected by the general-to-specific method, we suggest using only the lags that are statistically significant. Our Monte Carlo experiments show that the modified test has superior power to the standard ADF test with only minor size distortions. The power of the Dickey-Fuller generalized least squares (DF-GLS) test can also be improved by discarding insignificant intermediate lags. The appropriate use of the modified test is illustrated by using the inflation rates of OECD countries. Our results point out that shocks to many OECD countries' inflation rates are temporary instead of permanent.

***Keywords*:** Inflation, ADF Test, DF-GLS Test, Lag Length Selection
JEL Classification: E31, C22, C51

---

1. Introduction

Keeping a low steady rate of inflation is one of the government's most important responsibilities. Inflation is also related to the real interest rate, output and unemployment through various economic models. Without any surprise, economists have shown continued interest in this key economic variable. The most important question related to inflation is: Does a unit root exist in inflation? The answer to this question, which has important policy implications, can support or jeopardize the validity of several important economic models.

The seminal work of Nelson and Plosser (1982) finds a unit root in a bunch of U.S. macroeconomic time series, including inflation. Their conclusion arouses a series of following studies to re-examine the dynamic property of inflation rates. One major reason that their results are under scrutinizing originates from their test procedure, which follows Dickey and Fuller (1981). However, it is now well-known that Dickey and Fuller (1981)'s unit root tests have very low power when it encounters highly persistent data. By applying more advanced unit root tests, macroeconomists shed new light on the literature while the empirical results remain mixed. For example, Johansen (1992), Evans and Lewis (1995), Crowder and Wohar (1999), and Bai and Ng (2004) find evidence supporting the presence of a unit root in inflation rates, while Rose (1988), Neusser (1991), Culver and Papell (1997) and Basher and Westerlund (2009) find evidence against a unit root in inflation rates. More recently, Narayan and Popp (2011) employ a new seasonal unit root test to the G7 inflation rates and find mixed results. Tsong, Lee and Lee (2012) find mean reversion in inflation rates for 15 out of 19 OECD countries using panel unit root tests.

The aim of this paper is to examine the stationarity of inflation rates by constructing a more powerful lag selection method in the standard unit root test.[1]

Among a large literature which has investigated the issue of unit root testing, Lopez (1997) shows that the power of the augmented Dickey-Fuller (ADF) test, proposed by Dickey and Fuller (1981), is positively related to sample size and negatively related to the number of autoregressive lags needed in the testing regression. Moreover, Ng and Perron (1995) show that an overly parsimonious model can have large size distortions, but an over-parameterized model can have low power. They suggest a general-to-specific method (GTS) for selecting the lag length such that lags 1 through k are included in the regression if it is possible to reject the null hypothesis that lag k is statistically different from zero. In Section 2 of this paper, we take the Ng and Perron (1995) method one step further by purging any insignificant intermediate lags. After the value of k is selected, we examine the t-statistics from lags 1 to k-1 and include only the lags with significant coefficients. In Section 3, our Monte Carlo experiments show that this modified lag selection method has superior power to the standard ADF test with only minor size distortions. Similar results hold for the Dickey-Fuller generalized least squares (DF-GLS) unit root test proposed by Elliott, Rothenberg, and Stock (ERS) (1996). We also consider the effects of choosing the truncation lags using model selection criteria such as the Akaike information criterion (AIC) and Bayesian information criterion (BIC). We consider this to be important since ERS (1996) suggest that the BIC can provide better size adjusted power than the use of standard hypothesis testing. Section 4 illustrates the potential benefits of our procedure by performing unit root tests on the inflation rates of 30 OECD countries. We find more evidence against a unit root in inflation rates for OECD countries. Section 5 offers brief concluding remarks.

2. Unit Root Tests

Consider the standard regression equation used in the augmented Dickey-Fuller (ADF) test:

$$\Delta y_t = \alpha + \gamma t + \rho y_{t-1} + \sum_{i=1}^{k} \delta_i \Delta y_{t-i} + \varepsilon_t \qquad (1)$$

The null hypothesis of a unit root can be rejected if it is possible to conclude that $-2 < \rho < 0$. Ng and Perron (1995) show the importance of using the proper lag length, k, when performing the test. Too few lags result in size distortions and too many lags reduce the test's power. They use the general to specific (GTS) method, starting with a large value of k and examining the t-statistic for the null hypothesis $\delta_k = 0$. If the null hypothesis is rejected, (1) is estimated using all k lags. Otherwise, k is reduced by 1 and the procedure is repeated until the null is rejected or until $k = 0$. Note that lags 1 through k are included in the testing regression if the null hypothesis $\delta_k = 0$ is rejected. The number of autoregressive truncation lags can also be chosen by minimizing an objective function that trades off a reduction in the sum of squares of the residuals for a more parsimonious model. The two most widely used model selection criteria are the AIC and BIC.

ERS (1996) argue that the Dickey-Fuller test applied to a locally demeaned or detrended time series, using a data-dependent lag length selection procedure, has the best overall performance in terms of small-sample size and power. Specifically, ERS (1996) demean or detrend the data as follows:

$$y_t^d = y_t - \beta' z_t \tag{2}$$

where $z_t=(1,t)$` is used for detrending and $z_t=(1)$`for demeaning. The demeaned or detrended series, y_t^d, is then employed in the ADF regression without any deterministic regressors:

$$\Delta y_t^d = \rho y_{t-1}^d + \sum_{i=1}^{k} \delta_i \Delta y_{t-i}^d + \varepsilon_t \tag{3}$$

For the DF-GLS test, ERS (1996) suggest that the BIC could provide better size adjusted power than the GTS method. [2]

Though the ADF and DF-GLS are two most widely used unit root tests and the DF-GLS provides better performance in terms of small-sample size and power than the ADF, the low power problem remains for the stationary series with a relatively large autoregressive root. One possible reason for the inefficiency of the above unit root tests could come from their lag selection method. For instance, many of the economic series we use are monthly data. It is likely that the true data generating process (DGP) includes only lags 1, 2, 12 and 24 of the dependent variable. If we use the traditional lag selection method, a k equal to or higher than 24 means too many lags are included and the regression is inefficient, while a k less than 24 leads to size distortion.

Hence, we propose an easy modification to the ADF and DF-GLS tests for their lag selection method. Instead of using all of the lags determined by the standard lag selection process, we examine the *t*-statistics of every possible lag and only the lags with significant coefficients are included in the final regression.

3. Monte Carlo Experiments

Although the ADF and DF-GLS tests are probably the most widely used unit root tests, both can have low power for the sample sizes typically found in economic applications. Clearly, the power can be improved if unnecessary coefficients are eliminated from the estimating equations. With monthly data, for example, the true data generating process (DGP) might contain only lags 1, 12 and 24. If we use the traditional GTS method, a value of k = 24 means that too many lags are included. Of course, excluding lag 24 can result in a size distortion. As such, once k has been determined, it seems plausible to examine the *t*-statistics of the intermediate lags and retain only the lags with significant coefficients. Of course, similar remarks hold for non-seasonal data.

In order to assess that size and power of such a procedure, we perform a series of Monte Carlo experiments. We begin by simulating the following autoregressive process:

$$\Delta y_t = \alpha + \rho y_{t-1} + \delta_i \Delta y_{t-i} + \varepsilon_t ; \quad t = 1, \ldots , T \tag{4}$$

where, in order to save space, we report results for $\alpha = \delta = 0.5$, ρ = -0.2, -0.1, -0.05 or 0, and sample size $T = 100$. Our experiments use 5,000 replications and we report results using 5% critical values.[3] Although the true DGP does not include a trend, we report results with a constant and results with a trend and a constant in the testing regression.[4]

Table 1 reports the size and power of the ADF test and our modified ADF test when k is determined using the AIC, BIC, and GTS methods. Note that *All* refers to the traditional method of including lags 1 through k and *Modified* refers to the deletion of insignificant intermediate lags.

Our key findings for the ADF test are:

1. Our modification greatly improves the power of the ADF test for all lag selection methods. For example, for the GTS method with no trend and ρ = -0.10, the power of the standard ADF test is 38% and the power of the modified test is 59%. The gain in power is 16 percentage points for the same case with a trend.

2. For the ADF test with a constant, the size loss for the modified test is about one percentage point. In the presence of a constant and a trend, there is a mild size distortion of a few percentage points.

3. The three lag selection methods usually exhibit similar size and power properties.

The DF-GLS test is becoming quite popular because local-to-unity detrending enhances the power of the ADF test. Table 2 shows the results of our method applied to the DF-GLS test for the case in which k is selected by the BIC. As expected, the power of the DF-GLS test is always higher than the power of the ADF test. The key point is that our modification further improves the power of the DF-GLS test. For example, in the absence of a trend, when ρ = -0.10, the power of the DF-GLS test increases from 52% to 64%. In the presence of a trend, the power of the DF-GLS rises from 26% to 39%. There is a small size cost in that our modification adds 1 or 2 percentage points to the nominal size of the test.

Table 3 provides some intuition as to why our modification can improve power with only a small size reduction. The mean estimate of ρ has a smaller downward bias and/or smaller standard error when the intermediate lags are excluded from the testing regression.

4. Empirical Example

Although it is hard to believe that inflation rates actually contain a unit root, the empirical evidence is mixed. To further examine the issue, we calculate the inflation rates of 30 OECD countries using their seasonally unadjusted monthly consumer price indices (obtained from the *International Financial Statistics*). The series vary in length from 155 to 635 observations, and cover different time periods from January 1957 to December 2009.

We perform the ADF and DF-GLS unit root tests on each inflation rate using a constant, but no trend, in the testing equation. Table 4 reports the number of countries with a unit root null rejected by the traditional lag length selection methods and our modified method.

Note that the null hypothesis of a unit root is rejected more frequently when the intermediate lags are purged from the testing regression. For example, when the standard GTS method is employed in the ADF test, the null of a unit root is rejected for eight countries. The null is rejected for fourteen countries with our modified method. Similarly, the DF-GLS test (using the BIC to select the lag length) finds that seven inflation rates are stationary whereas our modified method finds ten stationary rates.

5. Conclusion

This paper has examined the stationarity of OECD countries' inflation rates by proposing a modification to the commonly used lag selection method. Standard lag selection methods can overfit the data by including too many lags in the testing regression. It is possible to improve the power of the ADF and DF-GLS unit root tests by omitting insignificant intermediate lags. This modification is reasonable for analyzing the quarterly or monthly time series data. The improvement comes at the expense of a small size distortion. Our lag length selection method finds stronger evidence of stationarity in OECD inflation rates than the standard lag length selecting method. Note that other methods of deleting intermediate lags (such as using the *F*-test, AIC, or BIC on groups of lags) can be employed.

Our contributions here are twofold: First, we show that simple methods can enhance the power of the standard ADF and DF-GLS unit root tests. This easily applied modification to unit root tests may be of interest to empirical economists. Second, our results indicate that inflation rates in many OECD countries are stationary, which means that shocks to inflation rates have a finite life for those countries.

Endnotes

[*] Walter Enders, Department of Economics, University of Alabama, Tuscaloosa, AL 35487.

Yu Liu, Corresponding author, Room 239, College of Business, University of Texas at El Paso, El Paso, Texas 79968, USA. Tel: 1-915-747-7728; Email: yliu4@utep.edu.

1. Please note that this paper is not intended to be a full-fledged discussion about modeling inflation. We focus on testing the unit root in inflation rates assuming that the variable follows a linear process. For the possibility that inflation fluctuation process may be nonlinear, see Henry and Shields (2004) for modeling inflation rates as a two-regime threshold unit root process, Becker, Enders and Hern (2001) for modeling inflation with a Fourier-Series approximation, and Zhou (2013) for modeling inflation rates as a smooth transition autoregressive model.

2. The modified AIC (MAIC) lag selection method, proposed by Ng and Perron (2001), can be used to correct the size distortions when there is a large negative MA root in the differenced series. We do not investigate this issue and only focus on the data with normal positive MA terms.

3. Results using other values of i, δ_i and T are available from the authors upon request.

4. We do not investigate the case of a large negative moving average term in the data generating process.

References

Bai, J. and S. Ng. 2004. “A PANIC Attack on Unit Roots and Co-integration,” *Econometrica*, 76, 1127-1177.

Basher, S. A. and J. Westerlund. 2009. “Panel Co-integration and the Monetary Exchange Rate Model,” *Economic Modelling*, 26, 506–513.

Becker, R., W. Enders, and S. Hurn. 2001. “Modeling Structural Change in Money Demand Using a Fourier-Series Approximation,” University of Technology Sydney, *Research Paper Series 67*.

Crowder, W. and M. Wohar. 1999. “Are Tax Effects Important in the Long-Run Fisher Relationship? Evidence from the Municipal Bond Market,” *Journal of Finance*, 54, 307-317.

Culver, S. E. and D. H. Papell. 1997. “Is There a Unit Root in the Inflation Rate? Evidence from Sequential Break and Panel Data Models,” *Journal of Applied Econometrics*, 12, 435-444.

Dickley, D. A. and W. A. Fuller. 1981. “Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root,” *Econometrica* 49, 1057-1072.

Elliott, G., T. J. Rothenberg, and J. H. Stock. 1996. “Efficient Tests for an Autoregressive Unit Root,” *Econometrica*, 64, 813-836.

Evans, M. D. and K. K. Lewis. 1995. “Do Expected Shifts in Inflation Affect Estimates of the Long-Run Fisher Relation,” *Journal of Finance*, 50, 225-253.

Henry, O. T. and K. Sheilds. 2004. “Is There a Unit Root in Inflation?” *Journal of Macroeconomics*, 26, 481-500.

Johansen, S. 1992. “Testing Weak Exogeneity and Order of Co-integration in UK Money Demand Data,” *Journal of Policy Modeling*, 14, 313-334.

Lopez, J. H. 1997. "The Power of the ADF Test," *Economics Letters*, 57, 5-10.

Narayan, P. K., and S. Popp. 2011. “An Application of a New Seasonal Unit Root Test to Inflation,” *International Review of Economics & Finance*, 20, 707-716.

Nelson, C. R. and C. I. Plosser. 1982. “Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications,” *Journal of Monetary Economics*, 10, 139-162.

Neusser, B. 1991. “Testing the Long-Run Implication of the Neoclassical Growth Model,” *Journal of Monetary Economics*, 27, 3-37.

Ng, S., and P. Perron. 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.

Ng, S. and P. Perron. 2001. "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power," *Econometrica*, 69, 1519–1554.

Rose, A. 1988. "Is the Real Interest Rate Stable?" *Journal of Finance*, 43, 1095-1112.

Tsong, C. C., C. F. Lee, and C. C. Lee. 2012. "A Revisit to the Stationarity of OECD Inflation: Evidence from Panel Unit Root Tests and the Covariate Point Optimal Test," *Japanese Economic Review*, 63, 380-396.

Zhou, S. 2013. "Nonlinearity and Stationarity of Inflation Rates: Evidence from the Euro-Zone Countries," *Applied Economics*, 45, 849-856.

Table 1: Power and Size of the Standard and Modified ADF Test

			Constant		*Constant + Trend*	
			All	***Modified***	***All***	***Modified***
Power	ρ = -0.20	GTS	0.63	0.95	0.44	0.87
		AIC	0.62	0.95	0.44	0.86
		BIC	0.71	0.9	0.5	0.79
	ρ = -0.10	GTS	0.38	0.59	0.26	0.42
		AIC	0.38	0.59	0.26	0.42
		BIC	0.35	0.49	0.23	0.34
	ρ = -0.05	GTS	0.21	0.26	0.14	0.19
		AIC	0.21	0.26	0.15	0.19
		BIC	0.17	0.21	0.11	0.15
Size	ρ = -0.00	GTS	0.06	0.07	0.07	0.1
		AIC	0.06	0.07	0.07	0.1
		BIC	0.06	0.07	0.06	0.08

Table 2: Power and Size of the Standard and Modified DF-GLS Test

		Constant		*Constant + Trend*	
		All	***Modified***	***All***	***Modified***
ρ = -0.20	ADF	0.71	0.9	0.5	0.79
	DF-GLS	0.78	0.91	0.56	0.82
ρ = -0.10	ADF	0.35	0.49	0.23	0.34
	DF-GLS	0.52	0.64	0.26	0.39
ρ = -0.05	ADF	0.17	0.21	0.11	0.15
	DF-GLS	0.29	0.35	0.13	0.17
ρ = -0.00	ADF	0.06	0.07	0.06	0.08
	DF-GLS	0.05	0.06	0.04	0.06

Table 3: Average Estimates of ρ (with standard errors) in the ADF Test

Actual	*All*	*Modified*	*All*	*Modified*
ρ = -0.20	-0.29 (0.09)	-0.26 (0.09)	-0.25 (0.07)	-0.24 (0.07)
ρ = -0.10	-0.16 (0.07)	-0.14 (0.06)	-0.15 (0.06)	-0.14 (0.06)
ρ = -0.05	-0.11 (0.06)	-0.09 (0.05)	-0.11 (0.05)	-0.09 (0.05)
ρ = -0.00	-0.10 (0.06)	-0.04 (0.05)	-0.10 (0.06)	-0.04 (0.05)

Table 4: Instances of Stationarity in the OECD Inflation Rates

	ADF: GTS	*ADF: AIC*	*ADF: BIC*	*DF-GLS*
All	8	4	14	7
Modified	14	11	16	10