LAG SELECTION OF THE AUGMENTED KAPETANIOS-SHIN-SNELL NONLINEAR UNIT ROOT TEST

[1]Jen-Je Su, [2]Adrian Wai-Kong Cheung and [1]Eduardo Roca

[1]Department of Accounting, Finance and Economics, Griffith University, Brisbane, Australia
[2]Department of Finance and Banking, School of Economics and Finance, Curtin University, Perth, Australia

ABSTRACT

We provide simulation evidence that shed light on several size and power issues in relation to lag selection of the augmented (nonlinear) KSS test. Two lag selection approaches are considered-the Modified AIC (MAIC) approach and a sequential General to Specific (GS) testing approach Either one of these approaches can be used to select the optimal lag based on either the augmented linear Dickey Fuller test or the augmented nonlinear KSS test, resulting in four possible selection methods, namely, MAIC, GS, NMAIC and NGS. The evidence suggests that the asymptotic critical values of the KSS test tends to result in over-sizing if the (N) GS method is used and under-sizing if the (N) MAIC method is utilised. Thus, we recommend that the critical values should be generated from finite samples. We also find evidence that the (N) MAIC method has less size distortion than the (N) GS method, suggesting that the MAIC-based KSS test is preferred. Interestingly, the MAIC-based KSS test with lag selection based on the linear ADF regression is generally more powerful than the test with lag selection based on the nonlinear version.

Keywords: (Nonlinear) Unit Root Test, Augmentation Lag, Modified AIC, General-To-Specific Sequential T-Test, Monte Carlo Simulation

1. INTRODUCTION

There has been growing concern that the Augmented Dickey-Fuller (ADF) test, which is derived under a linear setting, may not possess good power when applied to non-linear but stationary time series that are appropriate to characterize some economic and/or financial time series (Michael *et al.*, 1997; Taylor, 2001; and Sollis, 2009). To respond to this concern, a range of unit-root tests have been developed under a variety of nonlinear frameworks (Enders and Granger (1998); Kapetanios *et al.* (2003) (hereafter called KSS), Bec *et al.* (2004); Sollis (2009) and Kilic (2011)). Among these, Kapetanios *et al.* (2003) is probably the most widely recognized and applied (According to Hanck (2012) and Kapetanios *et al.* (2003) "as evidenced by e.g., a Scopus

citation count that is close to that of the most heavily cited Journal of Econometrics paper of the past five years". In addition, the KSS paper has been cited more than 200 times according to the ISI Web of Science by the end of March 2012). Kapetanios *et al.* (2003) propose a unit-root test using an auxiliary regression model that approximates the Exponential Smooth Transition Autoregressive (ESTAR) process by Taylor (2001) series. The nonlinear KSS test is shown, in general, to be more powerful than the linear DF test under the alternative of a globally stationary ESTAR process.

In the case of the linear DF test, serial correlation in innovations is, as suggested by Said and Dickey (1984), approximated by an augmented autoregression with a truncated lag k. An important issue is the choice of the truncated lag (k) which has vital size and power

Corresponding Author: Jen-Je Su, Department of Accounting, Finance and Economics, Griffith University, Brisbane, Australia

implications. Choosing an unnecessarily large k may reduce the power of the test while if k is inaptly set too small, considerable size distortion arises. Schwert (1989) found that in order to reduce size distortion, which is often due to a large negative Moving Average (MA) root in the innovations of the examined series, one has to choose a large autoregression lag. However, a large k can result in a nontrivial loss of power (DeJong *et al.*, 1992). Using such popular Information Criteria (IC) such as the Akaike Information Criterion (AIC) and the Schwartz or Bayesian Information Criterion (BIC) may not work out satisfactory size and power properties (Ng and Perron, 1995; 2001) as these criteria are not appropriate for modelling highly persistent (integrated) processes.

Ng and Perron (1995) suggested an improved data-driven lag selection procedure using a General-to-Specific (GS) approach based on sequential testing for the significance of coefficients on the highest auto regression lag. The GS approach yields a test with a better control in size but it tends to be over-parameterized in some occasions, leading to unwanted power loss. Ng and Perron (2001), on the other hand, developed a class of modified information criteria with an additional data-driven penalty factor that is apposite for integrated time series. In particular, Ng and Perron (2001) demonstrated via simulations that the Modified AIC (MAIC) is superior to the conventional AIC and BIC and the modified BIC, in controlling size when MA errors are present. Besides, unlike the GS procedure, the MAIC is shown to exhibit better power properties as it is less inclined to over-parameterization.

Like the DF-type test, in practice, the KSS test has to handle serial correlation in innovations. Kapetanios *et al.* (2003) assumes the serial correlation to be in a linear fashion and suggested an augmented test that is similar to the augmented DF test. Kapetanios *et al.* (2003) alleged that "standard model selection criteria or significance testing procedure (can) be used for this purpose because under the null of a linear model, the properties of these criteria are well understood." In applied works, different lag selection methods are used. They include the AIC and BIC (Pesaran *et al.*, 2009), the BIC (Ghoshray, 2010), the AIC (Cuestas and Garratt, 2011), the Modified AIC (MAIC) (Yau and Nieh, 2009) and a sequential-testing GS procedure proposed by (Ng and Perron, 1995; Kapetanios *et al.*, 2003; Chortareas and Kapetanios, 2004; Bahmani-Oskooee *et al.*, 2007; 2008; Baharumshah *et al.*, 2009) (It appears that empirical works applying the KSS test tend to favour the MAIC over other lag selection criteria).

This study provides simulation evidence to shed light on several unanswered issues regarding the lag selection in the augmented KSS test. First, while it is sensible to expect that under the common null (unit root) hypothesis the statistical properties of the lag selection methods should be much alike across linear and nonlinear unit-root testing schemes (Kapetanios *et al.*, 2003), it is less clear if the statistical properties would remain similar under the alternatives (as the ADF test assumes a linear alternative but the KSS test assumes a nonlinear one). Second, for practitioners, it is imperative to know if the MAIC, as implemented in the ADF test, is also the preferred lag selection method when the KSS test is applied. Specifically, would the MAIC outperform the seemingly more popular GS procedure for the KSS test? Third, though it is natural to construct the lag selection criteria based on the non-linear auxiliary KSS regression when the KSS is applied (as the majority of practitioners do), there is no priori ground ruling out other approaches that may produce better results. For example, it is possible that a "hybrid" approach-namely, establishing the optimal lag first within the linear ADF scheme and then applying the selected lag to the KSS test, may achieve better size and power properties than lags associated with the nonlinear KSS regression (A hybrid approach is considered in Perron and Qu (2007), but in a different context; To the best of our knowledge, Gustavsson and Osterholm (2007) is the only paper that takes this approach)? Fourth, there is a concern in the literature that the critical values from the asymptotic DF distribution might be distant from the critical values based on small samples and the issue can be exacerbated when the unit root tests are lag-augmented. In particular, using the asymptotic critical values may distort size and power (Cheung and Lai, 1995; Cook and Manning, 2002; Wu, 2010). The same issue arises in the case of the KSS test as well where there is a need to get lag-based finite-sample critical values; however, these critical values are not available in the literature. This study therefore sets out to examine and resolve these four issues by Monte Carlo simulation.

The study proceeds as follows. Materials and Methods gives a brief review on the augmented lag selection in unit root tests. Results reports the Monte Carlo results comparing the MAIC and the GS procedure in the KSS test. Discussion summarizes the findings and gives an application example, followed by conclusion.

2. MATERIALS AND METHODS

2.1. Unit Root Tests and Augmentation Lag Selection

Let y_t, t = 1, 2, ..., T, be an observed time series. Our aim is to distinguish if y_t is a unit root or a stationary process. The ADF test due to Dickey and Fuller (1979) and Said and Dickey (1984) is a t-test for the null hypothesis $\beta = 0$ (unit root) against the alternative $\beta < 0$ (stationary) in the autoregression Equation 1:

$$\Delta y_t = \beta y_{t-1}^d + \sum_{j=1}^{k} \phi_j \Delta y_{t-j} + \varepsilon_t \qquad (1)$$

where, y_t^d is the detrended (demeaned) series: $y_t^d = y_t - d_t$ with $d_t = \sum_{i=0}^{p} \lambda_i t^i$. Throughout the study, we consider two cases of the trend function d_t, Case A (level): p=0 and Case B (trend): p = 1. A two-step approach is used: the trend function d_t is first estimated by OLS and then the detrended series y_t^d is used for the unit root testing (Strictly speaking, the ADF test is based on the regression $\Delta y_t = d_t + \beta y_{t-1} + \sum_{j=1}^{k} \phi_j \Delta y_{t-j} + \varepsilon_t$ instead of the 2-step approach considered in this study. The two approaches share the same asymptotic results, though there are minor differences in finite samples. The 2-step approach is considered because it allows the trend function to be handled by both the ADF and the KSS test in the same way; in this study, we focus only on the unit root tests that are based on the OLS detrending. It is well-known that the power of the DF test can be much improved when the GLS-detrending is used, if the initial value is set zero (Elliott *et al.*, 1996). However, with large initial deviations, the GLS-based test suffers dramatic power loss and is dominated by the OLS-based test. Harvey *et al.* (2009; 2012) propose a "Union of Rejection (UR)" testing strategy that is able to produce a power that traces the higher of the DF tests based on different way of detrending. This is certainly a new avenue of for the nonlinear unit root tests. Su and Nguyen (2012) suggest a modified KSS test based on the UR strategy but they assume zero lag. A research combining the issues with different methods of detrending and lag augmentation for the KSS research is currently pursued by the authors).

In practice, one may choose the lag k based on standard information criteria such as the AIC and BIC as earlier studies have done, but neither achieves robust and satisfactory results (Ng and Perron, 1995). Ng and Perron (1995) instead suggested using a data-driven procedure, where the highest lag is sequentially tested. Specifically, the most general model with a maximum lag, k_{max}, is chosen and the coefficient of the highest lag (ϕk_{max}) is tested. If ϕk_{max} is significantly different from zero, the optimal lag is set as k_{max}; otherwise, a model with $k_{max}-1$ is considered and the significance of the coefficient of its highest lag ϕk_{max-1} is tested. The procedure continues until a significant highest lag is found (and hence the optimal lag is established). Since this procedure searches for the optimal lag with the most general setting (with k_{max} lags) and winds down to a more specific one (with less lags), it is often referred to as a General-to-Specific (GS) procedure. The sequential-testing GS procedure yields a unit root test with improved size (comparing with the tests based on the AIC and BIC) but it tends to over-parameterize, thereby resulting in a loss of power.

Ng and Perron (2001), on the other hand, developed an information criterion that is adequate for highly persistent (integrated) series called Modified AIC (MAIC) (A modified form of the BIC (MBIC) is considered in Ng and Perron (2001) as well. However, since the MAIC outperforms the MBIC as far as size is concerned, the MBIC will not be considered in this study). The approach estimates the lag as follows Equation 2:

$$\hat{k}_{MAIC} := \underset{0 \leq k \leq k_{max}}{\arg\min} \, MAIC(k)$$
$$MAIC(k) = \ln(\hat{\sigma}_{\varepsilon,k}^2) + 2\left(\tau_T(k) + k\right)/\left(T - k_{max}\right) \qquad (2)$$

where, $\hat{\sigma}_{\varepsilon,k}^2 = (T - k_{max})^{-1} \sum_{t=k_{max}+1}^{T} \hat{\varepsilon}_{k,t}^2$, $\hat{\varepsilon}_{k,t}$ is the residual from (1) and $\tau_T(k) = \hat{\sigma}_{\varepsilon,k}^{-2} \hat{\beta}^2 \sum_{t=k_{max}+1}^{T} (y_{t-1}^d)^2$ with $\hat{\beta}$ being the OLS estimate of β (For the purpose of meaningful comparison, following Ng and Perron (2001), MAIC(k) is computed using the same number (T-k_{max}) of residuals over different k. See also the discussion in Ng and Perron (2005) but in a different context). Note that the MAIC is the same as the AIC except that it includes an additional penalty term $\tau_T(k)$ which can better capture the relevant cost among different lag selections in finite sample. Ng and Perron (2001) demonstrated via simulations that the MAIC yields considerable size improvements over the standard criteria for the purpose of unit root testing, especially in the presence of a large negative moving-average component (Note that, as a

matter of fact, Ng and Perron (2001) examined the effect of the GLS-detrended ADF test only. Our simulation results, however, demonstrate that the good properties of the MAIC still hold when the test is OLS-detrending based. Perron and Qu (2007) for a similar discussion on this issue). Besides, since the MAIC is less inclined to over-parameterize, it is expected to produce more powerful testing results than when the GS procedure is implemented (Ng and Perron (2001) did not include the GS procedure in their simulations. Besides, results from Ng and Perron (1995; 2001) are not directly comparable as different detrending methods are used (OLS in Ng and Perron (1995) versus GLS in Ng and Perron (2001)). However, our simulation results show, in general, that the power of the MAIC-based tests is superior to those based on the GS approach).

The nonlinear unit root test of Kapetanios et al. (2003) is motivated by the observation that many time series (such as real exchange rates) exhibit local nonstationarity (unit root) but are stationary globally (Michael et al., 1997)) and nonlinear models, such as the Exponential Smooth Transition Auto Regression (ESTAR) model, could produce a better fit of these series than the linear models. Taylor (2001) shows with Monte Carlo evidence that the linear DF test does not have good power against a stationary ESTAR process. Kapetanios et al. (2003) consider the following ESTAR model Equation 3:

$$\Delta y_t = \rho y_{t-1} + \gamma y_{t-1}\left\{1 - \exp\left[-\theta(y_{t-1}^d)^2\right]\right\} + u_t \qquad (3)$$

where, u_t is a stationary innovation. When $\rho = \theta = 0$, y_t is a unit root process $\Delta y_t = u_t$. On the other hand, with $\theta > 0$, y_t follows a nonlinear but globally stationary process, provided that $-2 < \rho + \gamma < 0$. Using first-order Taylor series approximation (imposing that $\rho = 0$), KSS propose a unit root test based on the following auxiliary equation with lag-augmentation Equation 4:

$$\Delta y_t = \delta(y_{t-1}^d)^3 + \sum_{j=1}^{h} \eta_j \Delta y_{t-j} + v_t \qquad (4)$$

where, $\delta = -\gamma\theta$ with the null hypothesis $\delta = 0$ against the alternative: $\delta_{<0}$. Lag order h, as suggested by Kapetanios et al. (2003), can be chosen by the standard IC or the GS procedure of Ng and Perron (1995). As pointed out in the Introduction, applied works with the KSS test seem to favour the GS procedure. Given that the standard IC is not suitable for integrated series and the GC procedure tends to over-parameterize as far as the ADF test is concerned, the same issues are expected for the KSS test as well. In addition, an often neglected issue is, whenever the GS procedure is used, there are two possible lag selection approaches. The first and natural one is to base the lag selection on the auxiliary KSS regression (4), while the second one is to use a lag chosen in the linear ADF setting (1) for the KSS test (To the best of our knowledge, no applied works (using the GS procedure for the KSS test) set the lag based on the ADF regression). It is an open question whether these two approaches will generate similar results or not. For clarity, we shall denote GS as the GS procedure based on (1) and N-GS as that based on (4).

On the other hand, the augmentation lag selection can be determined by the MAIC (the preferred criterion for the ADF test) when the KSS test is used. The issue, once again (like in the case of the GS procedure), is that we do not have a clue as to whether the MAIC should be constructed based on the linear ADF setting or the nonlinear KSS setting and whether they would produce different results. If the former is the preferred one, one can simply apply the lag obtained from the ADF test to the KSS test. However, if it is the latter, then we need to build a nonlinear version of MAIC. To this end, we suggest, along the line of the construction of the MAIC based on the DF regression (1), to construct a nonlinear MAIC (N-MAIC) on the basis of the KSS regression (4) as follows Equation 5:

Table 1. Finite sample critical values of the augmented KSS tests

	T = 100				T = 200				T = 500			
	MAIC	GS	N-MAIC	N-GS	MAIC	GS	N-MAIC	N-GS	MAIC	GS	N-MAIC	N-GS
A level												
10%	-3.185	-3.698	-3.328	-3.716	-3.247	-3.600	-3.346	-3.616	-3.346	-3.523	-3.384	-3.486
5%	-2.698	-3.058	-2.768	-3.082	-2.765	-3.004	-2.812	-3.016	-2.792	-2.920	-2.816	-2.908
1%	-2.456	-2.751	-2.499	-2.768	-2.518	-2.712	-2.543	-2.723	-2.511	-2.611	-2.523	-2.603
B Trend												
10%	-3.675	-4.217	-3.760	-4.234	-3.683	-4.082	-3.775	-4.087	-3.768	-4.021	-3.822	-4.027
5%	-3.138	-3.578	-3.194	-3.583	-3.194	-3.506	-3.242	-3.506	-3.273	-3.454	-3.290	-3.459
1%	-2.884	-3.256	-2.919	-3.257	-2.942	-3.209	-2.971	-3.210	-3.017	-3.170	-3.026	-3.170

$$\hat{h}_{N-MAIC} := \underset{0 \le h \le h_{max}}{\arg\min} \ N\text{–}MAIC(h)$$

$$N\text{–}MAIC(h) = \ln(\hat{\sigma}^2_{v,h}) + 2\left(\tau^*_T(h) + h\right)/\left(T - h_{max}\right) \tag{5}$$

where, $\hat{\sigma}^2_{v,h} = (T - h_{max})^{-1} \sum_{t=h_{max}+1}^{T} \hat{v}^2_{h,t} \hat{v}_{h,t}$ is the residual from (4) and $\tau^*_T(h) = \hat{\sigma}^{-2}_{v,k} \hat{\delta}^2 \sum_{t=h_{max}+1}^{T} (y^d_{t-1})^6$ with $\hat{\delta}$ the OLS estimate of δ (Note that (5) is the same as (4) except for the term $\tau^*_T(h)$. It is trivial to prove that N-MAIC or Equation (5) holds for the approximate ESTAR model (i.e., Equation (4)). The proof is available from the authors upon request).

In this study, we are particularly interested in the performance of the augmented KSS test when the augmentation lag selection is based on one of the following four different ways: MAIC, N-MAIC, GS and N-GS. While lag selection strategies will not affect the asymptotic distribution of the KSS statistic, the finite-sample distribution may be rather different across different lag-selection rules and far apart from the asymptotic distribution. As a consequence, using critical values from the asymptotic distribution in small sample simulations may lead to erroneous conclusions (see a similar issue in the ADF context discussed in Cook and Manning (2002) and Wu (2010)). To resolve the issue, we use the critical values generated from finite (small) samples and the results are shown in **Table 1**.

In **Table 1**, via simulations, we tabulate the critical values of the KSS test based on each of the lag selection strategies in finite samples (The critical values are obtained from simulation using GAUSS with 100,000

replications). By comparing the critical values in **Table 1** with those of the asymptotic distribution given in Kapetanios *et al.* (2003, **Table 1**), it can be seen that the finite-sample critical values are distant from the asymptotic one, especially when the sample size (T) is small, despite the fact that there is a tendency for the KSS test to converge under different lag selection procedures. Taking T = 100 and Case A (level) as an example, the finite-sample critical values at 5% level are -2.698, -3.058, -2.768 and -3.082 for the KSS tests based on the MAIC, GS, N-MAIC and N-GS, respectively, while the corresponding asymptotic critical value is -2.93. This means that using the asymptotic value may make the test to have a tendency to be over-sized with (N) GS and under-sized with (N) MAIC.

3. RESULTS

3.1. Monte Carlo Results

We report the results of Monte Carlo simulations designed to investigate the size and power performance of the KSS test incorporated with the four different augmentation lag selection strategies, namely, MAIC, N-MAIC, GS, N-GS. For the purpose of comparison, we also report the results from the ADF test (but only linear lag selection criteria, MAIC and GS, are considered) (Note that the critical values for the ADF test are obtained in a way similar to those of the KSS test. To save space, they are not reported in this study but available upon request).

Table 2. Size (I): AR errors

		KSS				ADF	
	(ρ,η)	MAIC	GS	N-MAIC	N-GS	MAIC	GS
A level							
T = 100	(0.5,0)	0.051	0.054	0.053	0.053	0.051	0.051
	(0.8,0)	0.051	0.062	0.054	0.062	0.058	0.059
	(-0.5,0)	0.045	0.043	0.043	0.042	0.046	0.049
	(-0.8,0)	0.037	0.032	0.034	0.031	0.043	0.049
T = 200	(0.5,0)	0.049	0.051	0.049	0.051	0.046	0.048
	(0.8,0)	0.048	0.051	0.050	0.052	0.054	0.052
	(-0.5,0)	0.047	0.045	0.045	0.045	0.047	0.048
	(-0.8,0)	0.039	0.037	0.037	0.036	0.044	0.050
B Trend							
T = 100	(0.5,0)	0.057	0.057	0.055	0.058	0.050	0.050
	(0.8,0)	0.061	0.063	0.062	0.066	0.053	0.056
	(-0.5,0)	0.041	0.037	0.037	0.037	0.044	0.049
	(-0.8,0)	0.030	0.025	0.025	0.025	0.041	0.047
T = 200	(0.5,0)	0.053	0.054	0.051	0.055	0.049	0.047
	(0.8,0)	0.056	0.061	0.056	0.061	0.054	0.053
	(-0.5,0)	0.043	0.040	0.041	0.038	0.048	0.048
	(-0.8,0)	0.030	0.026	0.029	0.025	0.044	0.046

Table 3. Size (II): MA errors

	KSS				ADF	
(ρ,η)	MAIC	GS	N-MAIC	N-GS	MAIC	GS
A level						
(0,0.5)	0.047	0.060	0.049	0.061	0.042	0.055
(0,0.8)	0.042	0.059	0.043	0.058	0.034	0.057
(0,-0.5)	0.057	0.079	0.054	0.071	0.064	0.101
(0,-0.8)	0.063	0.222	0.038	0.096	0.112	0.323
(0,0.5)	0.045	0.052	0.045	0.051	0.044	0.051
(0,0.8)	0.043	0.055	0.041	0.055	0.038	0.051
(0,-0.5)	0.051	0.064	0.051	0.063	0.056	0.079
(0,-0.8)	0.041	0.107	0.036	0.069	0.082	0.208
B Trend						
(0,0.5)	0.050	0.066	0.045	0.068	0.038	0.062
(0,0.8)	0.043	0.063	0.036	0.062	0.022	0.058
(0,-0.5)	0.056	0.110	0.045	0.080	0.079	0.150
(0,-0.8)	0.098	0.377	0.024	0.095	0.202	0.499
(0,0.5)	0.045	0.054	0.043	0.055	0.039	0.053
(0,0.8)	0.042	0.053	0.037	0.053	0.031	0.052
(0,-0.5)	0.052	0.070	0.045	0.065	0.065	0.105
(0,-0.8)	0.036	0.167	0.026	0.062	0.111	0.337

Table 4. Power against stationary ESTAR processes

	γ	θ	KSS				ADF	
			MAIC	GS	N-MAIC	N-GS	MAIC	GS
Case A level								
T = 100	-0.1	0.01	0.097	0.084	0.087	0.083	0.087	0.083
		0.05	0.155	0.125	0.133	0.123	0.126	0.121
		0.10	0.191	0.154	0.167	0.147	0.156	0.148
	-0.5	0.01	0.246	0.190	0.214	0.187	0.156	0.148
		0.05	0.626	0.568	0.564	0.553	0.476	0.447
		0.10	0.730	0.710	0.634	0.686	0.606	0.670
	-1.0	0.01	0.435	0.354	0.393	0.343	0.268	0.233
		0.05	0.815	0.811	0.725	0.783	0.634	0.722
		0.10	0.873	0.882	0.767	0.843	0.685	0.789
T = 200	-0.1	0.01	0.162	0.146	0.145	0.143	0.133	0.135
		0.05	0.397	0.340	0.367	0.329	0.318	0.292
		0.10	0.479	0.422	0.447	0.403	0.467	0.423
	-0.5	0.01	0.629	0.566	0.597	0.558	0.459	0.415
		0.05	0.905	0.903	0.854	0.880	0.788	0.835
		0.10	0.931	0.936	0.870	0.911	0.824	0.884
	-1.0	0.01	0.827	0.806	0.787	0.786	0.679	0.701
		0.05	0.968	0.972	0.936	0.959	0.830	0.899
		0.10	0.987	0.987	0.968	0.980	0.858	0.933
Case B trend								
T = 100	-0.1	0.01	0.072	0.064	0.067	0.062	0.066	0.066
		0.05	0.098	0.076	0.089	0.077	0.094	0.077
		0.10	0.115	0.091	0.101	0.088	0.110	0.087
	-0.5	0.01	0.144	0.110	0.128	0.112	0.116	0.095
		0.05	0.443	0.337	0.382	0.335	0.325	0.226
		0.10	0.606	0.515	0.493	0.505	0.507	0.414
	-1.0	0.01	0.255	0.191	0.230	0.189	0.170	0.136
		0.05	0.717	0.661	0.597	0.647	0.547	0.521
		0.10	0.826	0.790	0.660	0.756	0.661	0.715
T = 200	-0.1	0.01	0.101	0.088	0.093	0.088	0.097	0.089
		0.05	0.222	0.180	0.199	0.177	0.199	0.165
		0.10	0.294	0.235	0.264	0.226	0.288	0.236
	-0.5	0.01	0.398	0.316	0.362	0.311	0.281	0.228
		0.05	0.818	0.789	0.728	0.760	0.706	0.725
		0.10	0.880	0.860	0.753	0.821	0.779	0.811
	-1.0	0.01	0.668	0.603	0.614	0.596	0.516	0.450
		0.05	0.937	0.927	0.858	0.902	0.785	0.828
		0.10	0.969	0.963	0.912	0.944	0.817	0.874

Table 5. Unit root tests for 17 OECS countries (1870-2008)

	KSS				ADF	
	MAIC	GS	N-MAIC	N-GS	MAIC	GS
(a) Unit root tests (bold: significant at 5% level)						
Austria	-2.908	-2.908	-2.908	-3.690	-1.786	-1.786
Australia	-2.716	-2.613	-2.381	-2.716	-1.439	-1.384
Belgium	-2.766	-2.830	-2.394	-2.830	-2.267	-2.379
Canada	-4.164	-4.164	-4.164	-4.164	-2.311	-2.311
Denmark	-3.257	-3.830	-3.106	-3.830	-1.768	-2.019
Finland	-4.156	-4.731	-2.688	-4.731	-2.676	-3.053
France	-4.522	-4.449	-3.353	-4.522	-2.219	-2.095
Germany	-5.532	-5.989	-4.602	-5.950	-2.031	-2.347
Italy	-2.576	-2.576	-2.576	-3.262	-2.005	-2.005
Japan	-2.698	-2.367	-2.698	-2.367	-2.434	-2.414
Norway	-1.081	-0.982	-1.081	-0.982	-1.019	-0.634
New Zealand	-1.426	-1.931	-1.426	-1.931	-0.783	-1.273
Spain	-2.717	-2.636	-2.717	-2.636	-1.808	-1.990
Sweden	-1.366	-1.366	-1.366	-1.366	-2.029	-2.029
Switzerland	-1.413	-1.340	-1.472	-1.340	-1.998	-2.247
UK	-1.882	-1.049	-1.049	-1.049	-2.957	-2.702
USA	-1.924	-1.924	-1.924	-1.924	-1.686	-1.686
(b) Lag Selection						
Austria	0.000	0.000	0.000	1.000		
Australia	3.000	1.000	2.000	3.000		
Belgium	3.000	0.000	2.000	0.000		
Canada	4.000	4.000	4.000	4.000		
Denmark	6.000	13.000	2.000	13.000		
Finland	2.000	1.000	0.000	1.000		
France	1.000	11.000	0.000	1.000		
Germany	2.000	1.000	0.000	3.000		
Italy	0.000	0.000	0.000	3.000		
Japan	0.000	9.000	0.000	9.000		
Norway	4.000	11.000	4.000	11.000		
New Zealand	1.000	5.000	1.000	5.000		
Spain	3.000	13.000	3.000	13.000		
Sweden	2.000	2.000	2.000	2.000		
Switzerland	9.000	8.000	10.000	8.000		
UK	0.000	9.000	9.000	9.000		
USA	0.000	0.000	0.000	0.000		

For each simulation, we compute the rejection frequency of the null hypothesis at the 5% level and the sample size is considered for T = 100, 200 under Case A (level) and Case B (trend), respectively. In line with the literature, we set the trend function in the simulations equal to zero since all the tests considered are similar. To alleviate the initial effects, additional 500 observations in each simulated path are generated first but they are not used. All simulations are performed in GAUSS with 20,000 replications.

We first examine the size based on two sets of integrated simulated paths $y_t = y_{t-1} + u_t$, $t = 1,...,T$, one with AR errors: $u_t = \rho u_{t-1} + \varepsilon_t$ and the other with MA errors $u_t = \varepsilon_t + \eta \varepsilon_{t-1}$, assuming that ε_t is i.i.d. N(0,1). We report the results with AR errors in **Table 2**. The results with MA errors are presented in **Table 3**.

We next turn to the power performance the KSS test against the stationary ESTAR models Equation 6:

$$y_t = y_{t-1} + \gamma y_{t-1}\left\{1 - \exp\left(-\theta y_{t-1}^2\right)\right\} + \varepsilon_t \tag{6}$$

where, $\varepsilon_t \sim N(0,1)$. Similar to Kapetanios et al. (2003), the values of γ and θ considered in this study are (-0.1, -0.5, -1.0) and (0.01, 0.05, 0.10), respectively. We report the simulation results in **Table 4**.

3.2. An Application to Gross Domestic Product (GDP)

Finally, we report in **Table 5** an empirical example where the KSS tests with different augmented lag selection rules are applied on yearly GDP of 17 OECD countries. The data covers the period 1870 to 2008 (The data are available from the lwebsite: http://www.ggdc.net/MADDISON/oriindex.htm. Panel A of **Table 5** shows the unit root test results while Panel B reports the results on optimal lag length selection.

4. DISCUSSION

As shown in **Table 2**, the augmented KSS test is subject to moderate size distortion with AR errors (generally, over-sized with a positive AR coefficient (ρ) and under-sized with a negative one). Besides, size distortion is relatively larger in Case B (trend) than in Case A (level) and the distortion becomes less noticeable as the sample size (T) increases. Among the four different lag selection strategies, the MAIC and N-MAIC tend to show a bit better control in size (closer to the nominal size, 5%) than the GC and N-GC procedures. The size distortion of the ADF test is similar to that of the KSS test.

Table 3 can be summarized as follows. When the MA parameter η is positive, the KSS test is slightly under-sized if the MAIC or N-MAIC is implemented and slightly over-sized if the GS or N-GS procedures are used. On the other hand, when the MA parameter is negative, size distortion becomes harder to control: size distortion is more serious with a larger MA coefficient (in magnitude) and/or with a trend (i.e., Case B) but is less so as T increases.

Among the four lag selection rules, the GS procedure appears to be the worst (with $\eta = -0.8$ and T = 100, the size is as large as 0.222 and 0.377 in Case A and Case B, respectively), followed by the N-GS procedure. The MAIC and N-MAIC tend to work better in size control and they are competitive with each other. Interestingly, with a negative MA parameter, the MAIC tends to associate with over-sizing while the N-MAIC with

under-sizing. Comparing with the KSS test, in general, the DF test is less size distorted when $\eta > 0$ but more over-sized when $\eta < 0$.

We comment on the results of **Table 4** as follows. First, the KSS test gains power as θ gets larger (given γ) and as $|\gamma|$ gets larger (given θ) and as expected, the test is relatively more powerful in Case A (level) with a larger sample size (T = 200). Second, the KSS test with lag selection based on the linear ADF regression is generally more powerful than its nonlinear counterparts (MAIC Vs. N-MAIC and GS vs. N-GS) (Occasionally, the N-MAIC outperforms the GS procedure in terms of power). Third, the MAIC appears to outperform the GS procedure. The MAIC-based KSS test achieves the highest power in most cases in Case A (particularly, when θ is small) and all cases in Case B (As pointed in Kapetanios et al. (2003), when θ grows larger, the series becomes less persistent. This implies that the MAIC-based KSS test is more powerful than its GS-based counterpart but may not be so when the examined series is less persistent). Fourth, the augmented KSS test, in general, has considerable power advantages over the ADF tests in testing against the stationary ESTAR alternatives.

Thus, the KSS test is more powerful than the DF test against the stationary ESTAR models, not only in the special case that sets lag equal to zero (Kapetanios et al., 2003) but also in the cases when the augmented lag selection is implemented. Interestingly, compared to the simulation results of **Table 3** in Kapetanios et al. (2003), we notice that there is a power loss in both the KSS and ADF tests with lags. In particular, the ADF test appears to suffer somewhat larger power loss than the KSS test. As a consequence, there are cases where the ADF test is more powerful than the KSS test when the lag is set at zero; but once augmentation lags are considered, the result turns opposite. For example, T = 100 with Case B and $(\gamma, \theta) = (-1, 0.05)$, in which the rejection rates are 0.910 (KSS) and 0.934 (DF) when the lag is set at zero (Case 3 of **Table 3** in Kapetanios et al. (2003)), the rates become 0.717 (KSS) and 0.547 (ADF) when the MAIC is implemented.

As for the empirical example, Panel B of **Table 5** reveals that the optimal lag length of GDP varies very considerably across different OECD countries. It also depends on what lag length selection method is used. As the MAIC-based KSS test is found to have good size and power, we focus our discussion on its result. In particular, MAIC suggests that GDP is nonstationary in most of the countries. The exceptions are Canada, Denmark, Finland, France and Germany where GDP is consistent with a globally stationary ESTAR process. For the purpose of comparison, we also report the ADF test

results. In contrast, the ADF tests do not reject the unit root null hypothesis for any of the countries considered.

5. CONCLUSION

In this study, we provide simulation evidence that shed light on several size and power issues in relation to lag selection of the augmented (nonlinear) KSS test. Two lag selection approaches are considered-the Modified AIC (MAIC) approach and a sequential General to Specific (GS) testing approach. One may use either one of these approaches to select the optimal lag based on either the augmented linear ADF test or the augmented nonlinear KSS test, resulting in four selection methods, namely, MAIC, GS, NMAIC and NGS. The evidence suggests that the asymptotic critical values of the KSS test tends to be over-sized if (N-) GS method is used and under-sized if (N-)MAIC method is used. Thus, we recommend that the critical values should be generated from finite samples. We also find evidence that (N-) MAIC method has less size distortion than (N-)GS method, suggesting that the MAIC-based KSS test is preferred. Interestingly, we also find that the MAIC-based KSS test with lag selection based on the linear ADF regression is generally more powerful than the test with lag selection based on the nonlinear version.

6. REFERENCES

Baharumshah, A.Z., V.K.S. Liew and C.T. Haw, 2009. The real interest rate differential: International evidence based on non-linear unit root tests. Bull. Econ. Res., 61: 83-94. DOI: 10.1111/j.1467-8586.2008.00288.x

Bahmani-Oskooee, M., A.M. Kutan and S. Zhou, 2007. Testing PPP in the non-linear STAR framework. Econ. Lett., 94: 104-110. DOI: 10.1016/j.econlet.2006.08.008

Bahmani-Oskooee, M., A.M. Kutan and S. Zhou, 2008. Do real exchange rates follow a nonlinear mean reverting process in developing countries. Southern Econ. J., 74: 1049-1062.

Bec, F., M.B. Salem and M. Carrasco, 2004. Tests for unit-root versus threshold specification with an application to the purchasing power parity relationship. J. Bus. Econ. Statist., 22: 382-395. DOI: 10.1198/073500104000000389

Cheung, L.W. and K.S. Lai, 1995. Lag order and critical values of the augmented Dickey-Fuller test. J. Econ. Bus. Statist., 13: 277-280.

Chortareas, G. and G. Kapetanios, 2004. The yen real exchange rate may be stationary after all: Evidence from non-linear unit-root tests. Oxford Bull. Econ. Statist., 66: 113-131. DOI: 10.1046/j.0305-9049.2003.00080.x

Cook, S. and N. Manning, 2002. Lag optimisation and finite-sample size distortion of unit root tests. Econ. Lett., 84: 267-274. DOI: 10.1016/j.econlet.2004.02.010

Cuestas, J.C. and D. Garratt, 2011. Is real GDP per capita a stationary process? Smooth transitions, nonlinear trends and unit root testing. Empirical Econ., 41: 555-563. DOI: 10.1007/s00181-010-0389-0

DeJong, D., J. Nankervis, N. Savin and C. Whiteman, 1992. The power problems of unit root test in time series with autoregressive errors. J. Econ., 53: 323-343. DOI: 10.1016/0304-4076(92)90090-E

Dickey, D.A. and W.A. Fuller, 1979. Distribution of the estimators for autoregressive time series with a unit root. J. Am. Statist. Assoc., 74: 427-431. DOI: 10.1080/01621459.1979.10482531

Elliott, G., T.J. Rothenberg and J.H. Stock, 1996. Efficient tests for an autoregressive unit root. *Econometrica*, 64: 813-836.

Enders, W. and C.W.J. Granger, 1998. Unit-root tests and asymmetric adjustment with an example using the term structure of interest rates. J. Bus. Econ. Statist., 16: 304-311. DOI: 10.1080/07350015.1998.10524769

Ghoshray, A., 2010. The extent of the world coffee market. Bull. Econ. Res., 62: 97-107. DOI: 10.1111/j.1467-8586.2009.00318.x

Gustavsson, M. and P. Osterholm, 2007. Does unemployment hysteresis equal employment hysteresis. Econ. Record, 83: 159-173. DOI: 10.1111/j.1475-4932.2007.00391.x

Hanck, C., 2012. On the asymptotic distribution of a unit root test against ESTAR alternatives. Statist. Probab. Lett., 82: 360-364. DOI: 10.1016/j.spl.2011.11.001

Harvey, D.I., S.J. Leybourne and A.M.R. Taylor, 2009. Unit root testing in practice: Dealing with uncertainty over the trend and initial condition. Econ. Theory, 25: 587-636. DOI: 10.1017/S026646660809018X

Harvey, D.I., S.J. Leybourne and A.M.R. Taylor, 2012. Testing for unit roots in the presence of uncertainty over both the trend and initial condition. J. Econ., 169: 188-195. DOI: 10.1016/j.jeconom.2012.01.018

Kapetanios, G., A. Snell and Y. Shin, 2003. Testing for a unit root in the nonlinear STAR framework. J. Econ., 112: 359-379. DOI: 10.1016/S0304-4076(02)00202-6

Kilic, R., 2011. Testing for a unit root in a stationary ESTAR process. Econ., Rev., 30: 274-302. DOI: 10.1080/07474938.2011.553511

Michael, P., R.A. Nobay and D.A. Peel, 1997. Transactions costs and nonlinear adjustment in real exchange rates: An empirical investigation. J. Political Econ., 105: 862-879. DOI: 10.1086/262096

Ng, S. and P. Perron, 1995. Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. J. Am. Statist. Assoc., 90: 268-281. DOI: 10.1080/01621459.1995.10476510

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. DOI: 10.1111/1468-0262.00256

Ng, S. and P. Perron, 2005. A note on the selection of time series models. Oxford Bull. Econ. Statist., 67: 115-134. DOI: 10.1111/j.1468-0084.2005.00113.x

Perron, P. and Z. Qu, 2007. A simple modification to improve the finite sample properties of Ng and Perron's unit root tests. Econ. Lett., 94: 12-19. DOI: 10.1016/j.econlet.2006.06.009

Pesaran, M.H., R.P. Smith, T. Yamagata and L. Hvozdyk, 2009. Pairwise tests of purchasing power parity. Econ. Rev., 28: 495-521. DOI: 10.1080/07474930802473702

Said, S.E. and D.A. Dickey, 1984. Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika, 71: 599-607. DOI: 10.1093/biomet/71.3.599

Schwert, G.W., 1989. Tests for unit roots: A monte carlo investigation. J. Bus. Econ. Statist., 7: 147-160.

Sollis, R., 2009. A simple unit root test against asymmetric STAR nonlinearity with an application to real exchange rates in Nordic countries. Econ. Model., 26: 118-125. DOI: 10.1016/j.econmod.2008.06.002

Su, J.J. and D.T. Nguyen, 2012. On maintaining the power of the Kapetanios-Snell-Shin (KSS) test under uncertain initial conditions. Applied Econ. Lett., 19: 891-896. DOI: 10.1080/13504851.2011.607120

Taylor, A.M., 2001. Potential pitfalls for the purchasing-power-parity puzzle? Sampling and specification biases in mean-reversion tests of the law of one price. Econometrica, 69: 473-498. DOI: 10.1111/1468-0262.00199

Wu, S., 2010. Lag length selection in DF-GLS unit Root tests. Commun. Statist. Simulat. Comput., 39: 1590-1604. DOI: 10.1080/03610918.2010.508143

Yau, H.Y. and C.C. Nieh, 2009. Testing for cointegration with threshold effect between stock prices and exchange rates in Japan and Taiwan. Japan World Econ., 21: 292-300. DOI: 10.1016/j.japwor.2008.09.001