Abstract

In this paper we examine the local power of unit root tests against globally stationary exponential smooth transition autoregressive [ESTAR] alternatives under two sources of uncertainty: the degree of nonlinearity in the ESTAR model, and the presence of a linear deterministic trend. First, we show that the Kapetanios, Shin and Snell (2003, Journal of Econometrics 112, 359–379) [KSS] test for nonlinear stationarity has local asymptotic power gains over standard Dickey-Fuller [DF] tests for certain degrees of nonlinearity in the ESTAR model, but that for other degrees of nonlinearity, the linear DF test has superior power. Second, we derive limiting distributions of demeaned, and demeaned and detrended KSS and DF tests under a local ESTAR alternative when a local trend is present in the DGP. We show that the power of the demeaned tests outperforms that of the detrended tests when no trend is present in the DGP, but deteriorates as the magnitude of the trend increases. We propose a union of rejections testing procedure that combines all four individual tests and show that this captures most of the power available from the individual tests across different degrees of nonlinearity and trend magnitudes. We also show that incorporating a trend detection procedure into this union testing strategy can result in higher power when a large trend is present in the DGP.

Keywords: Nonlinearity; Trend uncertainty; Union of rejections.

JEL classification: C22.
1 Introduction

There exist several areas of economics where economic theory suggests a variable should be stationary, but standard Dickey-Fuller [DF] tests fail to reject the null hypothesis of a unit root. Perhaps the most obvious example of this phenomenon occurring is with exchange rates. If variations in real exchange rates represent deviations from Purchasing Power Parity (PPP), then the finding of nonstationary behaviour in real exchange rates undermines the existence of a long-run tendency to PPP (Taylor et al., 2001). This conflict between economic theory and empirical results has led to concerns that, for the purposes of unit root testing, a linear time series framework might provide an insufficiently rich description of the underlying dynamics of many series.

A number of papers have found theoretical justification for modelling certain economic time series as nonlinear processes. For example, inter alios, Sercu et al. (1995) develop equilibrium models of real exchange rates in the presence of transaction costs, and show that the introduction of transaction costs results in a nonlinear, rather than linear, adjustment process towards PPP. A particular type of nonlinear model, the smooth threshold autoregressive (STAR) model, has received much attention in the literature (for a review, see van Dijk et al. (2002)). The STAR framework allows for a smooth transition between two autoregressive regimes such that the process exhibits a sharper mean reversion when it suffers from larger deviations away from its equilibrium but, for smaller deviations, displays more persistent, unit root type behaviour. Michael et al. (1997) argue that not accounting for STAR nonlinearity in unit root testing may explain the failure of previous studies to reject a unit root null in favour of the PPP hypothesis. Evidence therefore suggests that this form of nonlinearity may provide a more appropriate framework within which to test for unit roots for various economic series.

Kapetanios et al. (2003) [KSS] propose a test of the null hypothesis of a unit root against the alternative of a nonlinear, but globally stationary, exponential STAR (ESTAR) process. Under this alternative hypothesis it has been shown that KSS has power gains over the standard DF test, particularly where the process is highly persistent. The KSS test has subsequently been widely used in empirical applications on a variety of economic series such as, inter alia, real interest rates (Baharumshah et al., 2009; Tsong and Lee, 2013; Guney and Hasanov, 2014), GDP (Beechey and Osterholm, 2008; Cook, 2008) and current account balances (Akdogan, 2014; Chen, 2014).

Demetrescu and Kruse (2012) [DK] compare the KSS test to the linear DF test and show that the size of the error variance has an impact on their relative power performance against local ESTAR alternatives. In situations where the degree of nonlinearity is small relative to the error variance, the DF test can have superior local asymptotic power. In contrast, when the degree of nonlinearity is large relative to the error variance, the KSS test maintains a power advantage over DF. DK therefore argue that combining inference from both DF and KSS tests, along the lines of the Harvey et al. (2009, 2012) union of rejections testing approach, could provide a successful strategy for unit root testing against ESTAR alternatives. In this paper, we formally consider such a union of rejections approach, whereby the null hypothesis is rejected if either of the individual DF or KSS tests reject, subject to a critical value scaling that ensures the empirical size of the overall procedure is equal to nominal size asymptotically. We find that a union of rejections approach is able to capitalize on the respective power advantages exhibited by the DF and KSS tests under different DGPs, and can offer high levels of power across all degrees of nonlinearity.
Most economic time series are not characterized by zero-mean processes, but instead contain non-zero means or possibly non-zero means plus linear deterministic trends. KSS suggest accommodating these deterministic features by prior OLS demeaning or demeaning and detrending of the data. In empirical applications, inference made from KSS unit root tests will often depend on the deterministic specification assumed. For example, Christidou et al. (2013) consider the persistence of carbon dioxide emissions. They employ the no deterministic, demeaned, and demeaned and detrended KSS tests on a set of 36 countries. At least one of these three tests is able to reject the null hypothesis for the vast majority of countries. However, inference depends on whether or not a mean and trend are accommodated in the deterministic specification of the unit root tests. Kisswani and Nusair (2013) investigate the dynamics of real oil prices in seven Asian currencies plus the US dollar, using the same three KSS specifications. Again, they find that they are able to reject the null hypothesis of a unit root in most cases using at least one test, but that the three different KSS specifications provide conflicting results.

In situations such as these, where inference depends on the choice of deterministic specification, a practitioner uncertain about the presence or otherwise of a trend must choose which test to employ. A risk-averse strategy might be to always employ trend-invariant unit root tests. However, Marsh (2007) shows that, in the case of standard DF tests where data is generated by a linear AR process, the Fisher information for a test statistic invariant to a linear trend is zero at the unit root. Consequently, when a trend is absent, the power of a unit root test that is invariant to a trend will be compromised relative to the power of an appropriate demeaned but not detrended test statistic. Harvey et al. (2009) show that these power losses can be substantial, therefore opting to always use the trend-invariant test is a costly strategy. Conversely when a trend is present, the power of a DF test that is demeaned but not detrended is shown to decrease as the magnitude of the trend increases. Motivated by these considerations, in this paper we not only examine the power performance of a union of rejections based on demeaned tests and a union of rejections based on demeaned and detrended tests, but also consider a union of rejections based on all four possible tests (i.e. demeaned DF and KSS tests, and demeaned and detrended DF and KSS tests). This union procedure is shown to achieve attractive power levels across all settings of the trend coefficient.

In a related setting, Harvey et al. (2012) note that use of information regarding the presence of a trend can be applied to reduce the number of tests entering the union of rejections when it is clear that a trend exists, allowing more power to be achieved. We therefore consider a variant procedure where the Bayesian Information Criterion (BIC) is used to determine whether there is evidence of a trend. When the BIC suggests a trend is present, the procedure employs a union of rejections of only the two demeaned and detrended DF and KSS tests, while otherwise all four tests are used in the union of rejections. Finite sample simulations show that this approach is indeed able to improve on the power offered by the simple four-test union of rejections when the magnitude of the trend coefficient is large.

Finally, we apply the union of rejections procedures considered in this paper to energy consumption data across the period 1980-2011 for 180 countries. We find that the union of rejections procedures are able to capitalize on the differing rejections offered by the four individual tests, and at both a 0.10 and 0.05 significance level, the BIC-based union of rejections procedure achieves rejections for more countries than any individual test (or other union of rejections procedures) considered in this paper.

The outline of this paper is as follows. In section 2, we briefly outline the ESTAR
model and the KSS and DF tests for a unit root. Section 3 presents the limit distributions of the four individual tests under a local ESTAR alternative for both local to zero and fixed magnitude trends. In section 4 we present asymptotic power simulations and outline the different union of rejections procedures that we propose. Section 5 considers the finite sample performance of the proposed test procedures, and the empirical application to energy consumption data is given in section 6. Section 7 concludes.

2 Unit root tests and the ESTAR model

We write the univariate STAR model of order 1, i.e. a STAR(1) model, as

\[ y_t = \rho y_{t-1} + \gamma y_{t-d} G(\theta, y_{t-d}) + \varepsilon_t, \quad t = d+1, ..., T \]

where \( \varepsilon_t \sim iid(0,\sigma^2) \) and \( y_1 = O_p(1) \), and \( \rho \) and \( \gamma \) are unknown parameters. KSS adopt an exponential transition function of the form

\[ G(\theta, y_{t-d}) = 1 - \exp \left( -\theta y_{t-d}^2 \right) \]

where \( \theta \geq 0 \). The delay parameter, \( d \), measures the time taken for the transition process to begin reverting an economic variable back to its long run equilibrium. To simplify, we set \( d = 1 \) throughout this paper, as do KSS.\(^1\) The exponential transition function is bounded between 0 and 1, i.e. \( G : \mathbb{R} \to [0,1] \) has the properties

\[ G(\theta, 0) = 0, \quad \lim_{y_{t-d} \to \pm \infty} G(\theta, y_{t-d}) = 1 \]

and is symmetrical around zero. The function also implies that for \( \theta = 0 \) and \( \theta \to +\infty \), \( y_t \) reduces to a linear AR(1) process, while nonlinearity is present for \( 0 < \theta < \infty \).

In what follows, we make the distinction between the magnitude of the nonlinearity parameter \( \theta \) and the degree of nonlinearity manifest in the \( y_t \) process. Since the model for \( y_t \) reduces to a linear process when \( \theta = 0 \) or \( \theta \to +\infty \), an increase in the magnitude of the nonlinearity parameter does not translate to a monotonically increasing degree of nonlinearity in the process. For very small and very large magnitudes of \( \theta \), the degree of nonlinearity will be relatively small, while relatively large degrees of nonlinearity will be associated with intermediate values of \( \theta \).

The exponential STAR (ESTAR) model can then be written as

\[ \Delta y_t = \phi y_{t-1} + \gamma y_{t-1} \left( 1 - \exp \left( -\theta y_{t-1}^2 \right) \right) + \varepsilon_t \]

where \( \phi = \rho - 1 \). Imposing \( \phi = 0 \), i.e.

\[ \Delta y_t = \gamma y_{t-1} \left( 1 - \exp \left( -\theta y_{t-1}^2 \right) \right) + \varepsilon_t \] (1)

implies that \( y_t \) follows either a unit root or globally stationary process, and we consider testing the null hypothesis that \( y_t \) follows a unit root process, given by \( \gamma = 0 \) or \( \theta = 0 \),

---

\(^1\)In practical applications, \( d \) can be chosen to maximise goodness of fit over \( d = \{d_1, ..., d_{\text{max}}\} \). Norman (2009) shows that the asymptotic distribution of the KSS test is unchanged for \( d > 1 \), but that using the true value of \( d \), when it is known with certainty, offers almost uniformly higher power than setting \( d = 1 \).
against the alternative that $y_t$ is nonlinear and globally stationary, i.e. $\theta > 0$, with $-2 < \gamma < 0$ by assumption.\(^2\)

Since $\gamma$ is not identified under the null, KSS take the first order approximation of the Taylor expansion of the ESTAR model yielding the auxiliary equation

$$\Delta y_t = \delta y^3_{t-1} + e_t$$  \hspace{1cm} (2)

KSS then test the hypotheses

$$H_0 : \delta = 0, \quad H_1 : \delta < 0$$

using the test statistic

$$KSS = \frac{\widehat{\delta}}{s.e. \left(\widehat{\delta}\right)}$$

where $\widehat{\delta}$ is the OLS estimate of $\delta$ from (2) and $s.e. \left(\widehat{\delta}\right)$ is the corresponding standard error.

The KSS test for a unit root is clearly similar in form to the DF test, with the DF test statistic the $t$-statistic for testing $\delta = 0$ in the regression

$$\Delta y_t = \delta y_{t-1} + e_t$$  \hspace{1cm} (3)

as opposed to (2).

KSS suggest that a non-zero mean can be accommodated by using the demeaned data $y_{\mu,t} = y_t - \overline{y}$, where $\overline{y} = T^{-1} \sum_{t=1}^{T} y_t$. We refer to a test using this demeaned data as $KSS_\mu$. Both a non-zero mean and a deterministic (linear) trend can be accommodated by using the demeaned and detrended data $y_{t,t} = y_t - \widehat{\mu} - \widehat{\beta}t$ where $\widehat{\mu}$ and $\widehat{\beta}$ are the OLS estimators from the regression $y_t = \mu + \beta t + \eta_t$, $t = 1, ..., T$. We refer to this test as $KSS_r$. Similarly demeaned and detrended DF tests are denoted $DF_\mu$ and $DF_r$. We focus on these four tests throughout this paper. Note that to illustrate the core results, our analysis assumes that $\varepsilon_t$ is not serially correlated; this assumption can be relaxed along the lines of KSS, where additional serial correlation is permitted to enter the model in a linear fashion, provided the usual lag augmentation is applied to the DF and KSS tests, i.e. including lags of $\Delta y_t$ (or $\Delta y_{t,t}$ in the demeaned and detrended case) as additional regressors in (2) and (3).

### 3 Asymptotic properties of tests

To consider the behaviour of the unit root tests discussed in section 2 when data is generated by an ESTAR process, and to examine the effect that a linear deterministic trend has on the power of these tests, we consider a local to unity alternative form of the ESTAR DGP given in (1). DK consider a local alternative by setting $\gamma = -\frac{c}{T}$ ($c > 0$) in (1). This is the natural framework to evaluate local power, since under such a setting, $u_t$ is a local to unit root process, and it is well known that standard DF tests have non-trivial asymptotic power against this type of local alternative, with $T^{-1}$ being the appropriate Pitman drift. However, in the nonlinear model, the alternative DGP is characterized

\(^2\)KSS also consider an extension to the ESTAR model where $\phi > 0$ and $y_t$ is locally explosive and globally stationary under the alternative hypothesis.
in an additional dimension due to the unknown parameter \( \theta \). In order to prevent the nonlinear ESTAR component from becoming asymptotically negligible, DK show that \( \theta \) needs to be non-zero, but order \( T^{-1} \). We also introduce a scaling by \( \sigma^2 \), so that the magnitude of the nonlinearity parameter is measured relative to the variance of \( \varepsilon_t \), and ensuring that \( \sigma^2 \) does not appear in the limit distributions that follow; specifically, we set \( \theta = \frac{\sigma^2}{T} \).

Allowing for a non-zero mean and linear trend, our ESTAR DGP is

\[
y_t = \mu + \beta t + u_t, \quad t = 2, \ldots, T
\]

(4)

\[
\Delta u_t = -\frac{c}{T} u_{t-1} \left(1 - \exp \left(-\frac{g^2}{\sigma^2 T} u_{t-1}^2\right)\right) + \varepsilon_t
\]

(5)

where \( \varepsilon_t \sim iid(0, \sigma^2) \) with finite fourth order moment, and \( u_1 = O_p(1) \). The exponential transition function can be written as

\[
G(\theta, u_{t-1}) = 1 - \exp \left(-\frac{g^2}{\sigma^2 T} u_{t-1}^2\right)
\]

so that

\[
\Delta u_t = -\frac{c}{T} u_{t-1} G(\theta, u_{t-1}) + \varepsilon_t.
\]

Adapting results from DK, the partial sums of \( \varepsilon_t \) and \( u_t \) have the following limit distributions

\[
\left( \frac{1}{\sigma \sqrt{T}} \sum_{j=1}^{[rT]} \varepsilon_j \right) \text{ and } \left( \frac{1}{\sigma \sqrt{T}} \sum_{j=1}^{[rT]} u_{[rT]} \right) \quad \overset{d}{\rightarrow} \quad (W(r) ; X(r))
\]

where \( X(r) \) is the diffusion given by the stochastic differential equation

\[
dX(r) = -cX(r)G(g^2X(r)^2) \, dr + dW(r)
\]

with \( X(0) = 0 \), \( W(r) \) a standard Brownian motion process, and

\[
G(g^2X(r)^2) = 1 - \exp \left(-g^2X(r)^2\right).
\]

3.1 Asymptotic behaviour under a local trend

We now consider the effect of a local linear trend on the unit root tests discussed in section 2 under both the null hypothesis \( H_0 : \gamma = 0 \) and the local alternative hypothesis \( H_1 : \gamma = -c/T < 0 \). We follow Harvey et al. (2009) in setting \( \beta = \kappa \sigma T^{-1/2} \), with \( \kappa \) a finite constant, using the appropriate Pitman drift on the trend coefficient; the scaling by \( \sigma \) means that the trend magnitude is measured relative to the standard deviation of the innovations, and again ensures that this quantity does not appear in the limit expressions. The asymptotic distributions of the four unit root tests: \( DF_\mu \), \( KSS_\mu \), \( DF_r \) and \( KSS_r \) are given by the following lemma.
Lemma 1 Let \( y_t \) be generated according to (4)-(5) with \( \beta = \kappa \sigma T^{-1/2} \). For \( c \geq 0 \),

\[
DF_{\mu} \xrightarrow{d} \frac{\int_0^1 (\kappa (r - \frac{1}{2}) + X_{\mu}(r)) dW_r - c \int_0^1 (\kappa (r - \frac{1}{2}) + X_{\mu}(r)) X(r) G(g^2 X(r)^2) \, dr}{\sqrt{\int_0^1 (\kappa (r - \frac{1}{2}) + X_{\mu}(r))^2 \, dr}}
\]

\[
KSS_{\mu} \xrightarrow{d} \frac{\kappa \int_0^1 (\kappa (r - \frac{1}{2}) + X_{\mu}(r))^3 \, dr + c \int_0^1 (\kappa (r - \frac{1}{2}) + X_{\mu}(r))^3 X(r) G(g^2 X(r)^2) \, dr}{\sqrt{\int_0^1 (\kappa (r - \frac{1}{2}) + X_{\mu}(r))^6 \, dr}}
\]

\[
DF_{\tau} \xrightarrow{d} \frac{\int_0^1 X_{\tau}(r) dW_r - c \int_0^1 X_{\tau}(r) X(r) G(g^2 X(r)^2) \, dr}{\sqrt{\int_0^1 X_{\tau}(r)^2 \, dr}}
\]

\[
KSS_{\tau} \xrightarrow{d} \frac{\int_0^1 X_{\tau}(r)^3 dW_r - c \int_0^1 X_{\tau}(r)^3 X(r) G(g^2 X(r)^2) \, dr - 12 \int_0^1 (r - \frac{1}{2}) X(r) \, dr \int_0^1 X_{\tau}(r)^3 \, dr}{\sqrt{\int_0^1 X_{\tau}(r)^6 \, dr}}
\]

where \( X_{\mu}(r) = X(r) - \int_0^1 X(s) \, ds \) and \( X_{\tau}(r) = X(r) + (6r - 4) \int_0^1 X(s) \, ds + (6 - 12r) \int_0^1 s X(s) \, ds \).

Proof: See Appendix.

The null limit distributions of the \( DF_{\mu} \), \( KSS_{\mu} \), \( DF_{\tau} \) and \( KSS_{\tau} \) statistics are obtained as a special case of Lemma 1, on setting \( c = 0 \), so that \( X(r) = W(r) \). Note that when \( g^2 = 0, G(g^2 X(s)^2) = 0 \), and the limit distributions again reduce to those under the \( c = 0 \) null. Hence, the null can be viewed as either \( c = 0 \) or \( g^2 = 0 \).

The limiting distribution of \( KSS_{\tau} \) under the null corrects the corresponding result in Hanck (2012), which in turn corrected the limits given in KSS. Hanck (2012) writes the limit of the numerator as \( \int_0^1 W_{\tau}(r)^3 dW_r \), whereas the correct expression is \( \int_0^1 W_{\tau}(r)^3 dW_r \). The fact that \( dW_{\tau}(r) \neq dW_r \) (since \( \Delta g_{\tau,t} = \Delta y_t - \beta \)) is the source of the discrepancy, with Lemma 1 providing the correct result.

We observe that the limiting distribution of the demeaned statistics \( DF_{\mu} \) and \( KSS_{\mu} \) depend on the trend term via \( \kappa \), whilst the invariant detrended statistics \( DF_{\tau} \) and \( KSS_{\tau} \) do not. These results coincide with the behaviour of \( DF_{\mu} \) and \( DF_{\tau} \) reported in Harvey et al. (2009) under a linear AR DGP. Also, as \( g^2 \to \infty, G(g^2 X(s)^2) \to 1 \), so the limiting distributions of the two DF tests under an ESTAR alternative will converge to those obtained in Harvey et al. (2009) under a linear local AR alternative.

### 3.2 Asymptotic behaviour under a fixed trend

For completeness, we also consider the behaviour of \( DF_{\mu} \) and \( KSS_{\mu} \) under a fixed magnitude trend \( \beta = \kappa \sigma \) where \( \kappa \) is a finite constant. The limiting distribution of the trend-invariant \( DF_{\tau} \) and \( KSS_{\tau} \) statistics do not change under a fixed trend. The limiting distributions of \( DF_{\mu} \) and \( KSS_{\mu} \) under the iid assumption for \( \epsilon_t \) are given in the following Lemma.

Lemma 2 Let \( y_t \) be generated according to (4)-(5) with \( \beta = \kappa \sigma \neq 0 \). For \( c \geq 0 \),

\[
DF_{\mu} \xrightarrow{d} \frac{\int_0^1 (r - \frac{1}{2}) dW_r - c \int_0^1 (r - \frac{1}{2}) X(r) G(g^2 X(r)^2) \, dr}{\sqrt{(\kappa^2 + 1)/12}}
\]
We note that both $DF_\mu$ and $KSS_\mu$ possess well-defined limiting distributions that are dependent on the local alternative parameter $c$. These results correspond with those found by Harvey et al. (2009) in the case of the DF test under a linear AR alternative; indeed, as $g^2 \to \infty$, the limit of $DF_\mu$ reduces to the Harvey et al. (2009) limit. Under the null hypothesis $c = 0$, the $DF_\mu$ limit again coincides with the null limit given in Harvey et al. (2009), and from Remark 5 of that paper we can state that $DF_\mu$ follows an asymptotic normal distribution with mean zero and variance $\frac{1}{(\kappa^2 + 1)}$. This implies that for any non-zero fixed trend, $DF_\mu$ will have a null limiting normal distribution with variance less than one, resulting in an asymptotically under-sized test (maximum size can be calculated to be 0.002 as in Harvey et al. (2009)). Similarly, $KSS_\mu$ will follow an asymptotic normal distribution under the null, with mean zero and variance approximately $\frac{2}{32} \approx \frac{1}{(\kappa^2 + 1)}$ (obtained by simulation). As with $DF_\mu$, this implies an asymptotically under-sized test in the presence of a fixed trend (here, maximum size can be found to be 0.027).

4 Union of rejections strategies

4.1 Union of DF and KSS

To begin our numerical analysis, we abstract from uncertainty regarding the presence or absence of a trend, and evaluate the local asymptotic powers of the $DF_\mu$ and $KSS_\mu$ tests when no trend is present in the DGP ($\kappa = 0$), and subsequently the $DF_\tau$ and $KSS_\tau$ tests which are invariant to any trend. Asymptotic critical values at a nominal 0.05 significance level were first obtained by direct simulation of the limiting distributions given in Lemma 1. The $W(r)$ processes were approximated using iid standard normal random variables and with the integrals approximated by normalized sums of 2,000 steps using 50,000 Monte Carlo replications. These critical values (and those for 0.10 and 0.01 significance levels) are reported in Table 1.

Figures 1(a)-(f) show the local asymptotic powers of both $DF_\mu$ and $KSS_\mu$ unit root tests for six different settings of the nonlinearity parameter $g^2$, when $\kappa = 0$. For each of $g^2 = \{2, 5, 10, 20, 50, 100\}$ we consider values of $c$ such that the power of the tests approaches one. Here and throughout the paper, asymptotic power is simulated using normalized sums of 1,000 steps and 20,000 Monte Carlo replications. Figure 1(a) shows the case for $g^2 = 2$. Here, $KSS_\mu$ achieves higher power than $DF_\mu$ across all values of $c$. This highlights the superiority of KSS tests when the degree of nonlinearity in the process is relatively large via the nonlinearity parameter in the ESTAR model being of modest magnitude. For example, when $c = 99$, $KSS_\mu$ has power of 0.64 whilst $DF_\mu$ has power of 0.49. It is clear in these circumstances that a practitioner would want to employ $KSS_\mu$ to test for a unit root. In Figure 1(b) where $g^2 = 5$, $KSS_\mu$ maintains a power advantage over $DF_\mu$ for almost all settings of $c$, but these power gains are smaller than those seen in Figure 1(a). For example, when $c = 99$, $KSS_\mu$ now has power of 0.67 whilst $DF_\mu$ has power of 0.56. Figure 1(c) shows the power of both tests when $g^2 = 10$. Here, matters are less clear cut, with the power curves intersecting at (approximately) $c = 44$. In Figure 1(d), $g^2$ is increased to 20; the power of both tests is now very similar up until
At which point the power curves diverge with $DF_\mu$ outperforming $KSS_\mu$ for these higher values of $c$. Figure 1(e) shows the case for $g^2 = 50$; the $DF_\mu$ test now displays higher power across almost all values of $c$. Finally, in Figure 1(f), $g^2 = 100$, and $DF_\mu$ displays substantially higher power than $KSS_\mu$. For example, when $c = 22$, the power of $DF_\mu$ is 0.87 whilst $KSS_\mu$ has power of 0.66. These latter figures draw attention to the power gains $DF$ tests can have against $KSS$ tests for large magnitudes of $g^2$; recall that as $g^2 \to \infty$, the ESTAR process reduces to a linear process, hence, compared to modest values of $g^2$, the degree of nonlinearity is reduced for large $g^2$, and it is not surprising that $DF$ power gains become apparent in these cases.

Figures 2(a)-(f) show the asymptotic powers of both $DF_\tau$ and $KSS_\tau$ unit root tests for the same six settings of the nonlinearity parameter $g^2$ and the same range of values for the local alternative parameter $c$. Given that both tests are trend-invariant, the setting of $\kappa$ used in these simulations is irrelevant. Figure 2(a), where $g^2 = 2$, shows that, for modest magnitudes of the nonlinearity parameter, $KSS_\tau$ has power gains over $DF_\tau$, due to a high degree of nonlinearity in the process. For example, for $c = 171$, $KSS_\tau$ has power of 0.61 and $DF_\tau$ has power of 0.51. We note that these power gains are smaller than those seen in the demeaned case in Figure 1; as we would expect, for the same values of $c$, both detrended tests have less power than their demeaned counterparts. In Figure 2(b), where $g^2 = 5$, $KSS_\tau$ still maintains a power advantage over $DF_\tau$, but these gains are diminished compared to the previous figure. Figures 2(c) to 2(f) show that as the value of $g^2$ increases, the relative power performance of $DF_\tau$ to $KSS_\tau$ improves, to the point where $DF_\tau$ outperforms $KSS_\tau$, in line with the degree of nonlinearity reducing. For example, for $g^2 = 100$ and $c = 30$, the power of $DF_\tau$ is 0.90 and the power of $KSS_\tau$ is 0.67. This mirrors the results of the demeaned tests.

The results from Figures 1 and 2 show that the relative powers of the $DF$ and $KSS$ tests are very sensitive to the degree of nonlinearity under the ESTAR alternative, with the $KSS$ test offering greater power for modest magnitudes of the nonlinearity parameter, and the $DF$ test displaying relatively higher power for larger magnitudes. Given that the degree of nonlinearity is almost certainly unknown in practical applications, these results highlight the potential benefits of a testing strategy that combines inference from the $DF$ and $KSS$ tests. We therefore want to employ a testing strategy that can capitalize on the respective power advantages of the $DF$ and $KSS$ tests, regardless of the degree of nonlinearity under the ESTAR alternative. In line with the suggestion of DK, we now consider two union of rejections strategies, cf. Harvey et al. (2009), one that combines $DF_\mu$ and $KSS_\mu$ (for the case of no trend), and one that combines $DF_\tau$ and $KSS_\tau$ (for the trend case).

The union of rejections approach is a simple decision rule where the null hypothesis of a unit root is rejected if either of the individual tests reject. We can write our proposed union of rejections strategies as

$$U_i : \text{Reject } H_0 \text{ if } DF_i < \lambda_{i}^{\alpha} cv_{DF_i}^{\alpha} \text{ or } KSS_i < \lambda_{i}^{\alpha} cv_{KSS_i}^{\alpha}, \quad i = \mu, \tau$$

where $cv_{DF_i}^{\alpha}$ and $cv_{KSS_i}^{\alpha}$ denote the asymptotic null critical values of $DF_i$ and $KSS_i$ respectively for a significance level $\alpha$. Note that if we simply rejected the null hypothesis whenever either $DF_i$ or $KSS_i$ is smaller than its respective (unscaled) critical value, an over-sized procedure would result, hence we incorporate the scaling constant $\lambda_{i}^{\alpha}$, calculated such that the asymptotic empirical size of $U_i$ equals the nominal size $\alpha$. The decision
rule can also be written as

\[ U_i : \text{Reject } H_0 \text{ if } t_{U_i} = \min \left( \frac{DF_i}{\text{cv}_{DF}^i}, \frac{KSS_i}{\text{cv}_{KSS}^i} \right) < \lambda_i^\alpha \text{cv}_{DF}^i, \quad i = \mu, \tau. \]

The limiting distribution of \( t_{U_i} \), under a local trend, is then given by a function of the limiting distributions derived in Lemma 1:

\[ t_{U_i} \xrightarrow{d} \min \left( L_{DF_i}, \frac{\text{cv}_{DF}^\gamma_i}{\text{cv}_{KSS}^\gamma_i} L_{KSS_i} \right), \quad i = \mu, \tau \]

where \( L_{DF_i} \) and \( L_{KSS_i} \) denote the asymptotic distributions of \( DF_i \) and \( KSS_i \), respectively, given in Lemma 1.

To obtain the appropriate value for the scaling constant \( \lambda_i^\alpha \), we can simulate the limit distribution of \( t_{U_i} \) and calculate the level critical value \( \text{cv}_{U_i} \). Computing \( \lambda_i^\alpha = \text{cv}_{U_i}^\alpha / \text{cv}_{DF}^\alpha \) will then give the value for the scaling constant that ensures \( U_i \) is asymptotically correctly sized. Asymptotic scaling constants for nominal 0.10, 0.05 and 0.01 significance levels are given in Table 1.

Figure 1 also displays the power performance of \( U_\mu \) for \( g^2 = \{2, 5, 10, 20, 50, 100\} \). Figure 1(a) shows asymptotic power of \( U_\mu \) for \( g^2 = 2 \). We see that the power of \( U_\mu \) dominates that of \( DF_\mu \) for all values of \( c \), with its power curve lying just underneath that of \( KSS_\mu \) for approximately \( c \leq 108 \), and with power rising above both \( KSS_\mu \) and \( DF_\mu \) beyond this point. \( U_\mu \) is therefore able to capture the power advantage offered by \( KSS_\mu \) in this modest nonlinearity parameter magnitude case. For \( g^2 = 5 \) in Figure 1(b), we see that, again, \( U_\mu \) outperforms \( DF_\mu \) across all values of \( c \), and offers power that is very slightly below that of \( KSS_\mu \) for lower values of \( c \), and slightly above at higher values of \( c \). In Figure 1(c), where \( g^2 = 10 \), the power curve of \( U_\mu \) closely tracks that of whichever test, \( KSS_\mu \) or \( DF_\mu \), is superior for a given \( c \). In Figures 1(d)-(f), as the value of \( g^2 \) increases, the power of \( U_\mu \) now dominates that of \( KSS_\mu \) and closely tracks that of \( DF_\mu \) for all values of \( c \). \( U_\mu \) therefore provides power near the effective envelope offered by \( KSS_\mu \) and \( DF_\mu \) across all values of the nonlinearity parameter \( g^2 \).

Figure 2 shows the power performance of \( U_\tau \) for the same six settings of \( g^2 \). We observe a very similar pattern of behaviour, relative to the DF and KSS tests, to that seen in Figure 1. In Figures 2(a) and 2(b), the power of \( U_\tau \) dominates that of \( DF_\tau \) and closely tracks that of \( KSS_\tau \). In Figure 2(c), as the power curves of the two individual tests intersect, we see that the power of \( U_\tau \) is marginally higher than both of these tests for most values of \( c \). As the nonlinearity parameter, \( g^2 \), increases in Figures 2(d)-(f), the power of \( U_\tau \) now dominates that of \( KSS_\tau \) for all values of \( c \), whilst closely tracking the power curve of \( DF_\tau \).

### 4.2 Union of demeaned and demeaned and detrended tests

The above results clearly illustrate the benefit of employing a union of rejections procedure in practice where the degree of nonlinearity is unknown. However, we have so far assumed that it is known whether or not the DGP contains a linear deterministic trend, so that a practitioner would know whether to apply \( U_\mu \) or \( U_\tau \). In practice, there will be uncertainty about the presence or otherwise of a trend in many economic time series. In section 3, we showed that the limiting distributions of \( DF_\mu \) and \( KSS_\mu \), under both local and fixed trends, are not invariant to the trend, suggesting that the power of these tests may be compromised if a linear trend is present in the DGP. Whilst this might suggest that a
conservative strategy of always using $U_r$ should be employed, Harvey et al. (2009) show that, in the context of a linear AR DGP, $DF_\mu$ has a substantial power advantage over $DF_r$ when no trend is present in the DGP, and might also be expected to offer some power gains for very small trend magnitudes. We therefore investigate the power performance of $U_\mu$ and $U_r$ when a local trend is present in the DGP.

Figures 3-5 show the asymptotic power of $U_\mu$ and $U_r$ for six different settings of the local trend coefficient, $\mu = \{0, 0.25, 0.5, 1, 2, 4\}$ and three different settings of the nonlinearity parameter $g^2 = \{2, 10, 50\}$. In Figure 3, we set $g^2 = 2$; Figure 3(a) shows the power of $U_\mu$ and $U_r$ when $\mu = 0$ such that no trend is present in the DGP. We see that $U_\mu$ significantly outperforms $U_r$ across all values of $c$. For example, when $c = 135$, $U_\mu$ has power of 0.83 whilst $U_r$ has power of 0.50. This highlights the potential losses involved in always employing the trend-invariant $U_r$ when there is uncertainty about the presence or otherwise of a linear trend. In Figure 3(b), $\mu = 0.25$; the presence of this linear trend has decreased the power of $U_\mu$ such that, for $c = 135$, power has dropped to 0.56, only marginally higher than that of $U_r$. When the local trend coefficient increases to $\mu = 0.5$ in Figure 3(c), the power of $U_r$ now dominates that of $U_\mu$. Figures 3(d)-(f) show that as the trend magnitude continues to increase, the power of $U_\mu$ deteriorates dramatically, such that for $\mu = 1$ in Figure 3(d), $U_\mu$ has power lower than size for most $c$, and for $\mu = 2$ and $\mu = 4$ in Figures 3(e)-(f), the size and power of $U_\mu$ approach zero. Almost identical patterns of behaviour are observed for $g^2 = 10$ and $g^2 = 50$ in Figures 4 and 5 respectively. These results demonstrate the potential losses that a practitioner can make if they fail to choose the appropriate deterministic specification for unit root testing.

Given that, in practice, the presence or otherwise of a linear deterministic trend will often be unknown, we propose a further union of rejections procedure that combines information from $U_\mu$ and $U_r$ by taking a union of rejections of all four individual tests $DF_\mu$, $KSS_\mu$, $DF_r$ and $KSS_r$. This proposed testing strategy can then be written as

\[ U_4 : \text{Reject } H_0 \text{ if } DF_\mu < \lambda_4^{\alpha}cv_{DF_\mu}^{\alpha} \text{ or } KSS_\mu < \lambda_4^{\alpha}cv_{KSS_\mu}^{\alpha} \text{ or } DF_r < \lambda_4^{\alpha}cv_{DF_r}^{\alpha} \text{ or } KSS_r < \lambda_4^{\alpha}cv_{KSS_r}^{\alpha} \]

or

\[ U_4 : \text{Reject } H_0 \text{ if } t_{U_4} = \min \left( DF_\mu, \frac{cv_{DF_\mu}^{\alpha}}{cv_{KSS_\mu}^{\alpha}} KSS_\mu, \frac{cv_{DF_\mu}^{\alpha}}{cv_{DF_r}^{\alpha}} DF_r, \frac{cv_{DF_r}^{\alpha}}{cv_{KSS_r}^{\alpha}} KSS_r \right) < \lambda_4^{\alpha}cv_{DF_\mu}^{\alpha}. \]

The limiting distribution of $t_{U_4}$, under a local trend, is then given by

\[ t_{U_4} \xrightarrow{d} \min \left( L_{DF_\mu}, \frac{cv_{DF_\mu}^{\alpha}}{cv_{KSS_\mu}^{\alpha}} L_{KSS_\mu}, \frac{cv_{DF_\mu}^{\alpha}}{cv_{DF_r}^{\alpha}} L_{DF_r}, \frac{cv_{DF_r}^{\alpha}}{cv_{KSS_r}^{\alpha}} L_{KSS_r} \right). \]

Values of the scaling constant $\lambda_4^{\alpha}$ required to control the asymptotic size of $U_4$ are given in Table 1 for 0.10, 0.05 and 0.01 significance levels.

Figures 3-5 display asymptotic power results for our proposed union, $U_4$, for the same three settings of $g^2$ and six settings of $\mu$ used previously. Figure 3(a) displays power in the case where $g^2 = 2$ and no trend is present in the DGP ($\mu = 0$). The power of $U_4$ dominates that of $U_r$ across all values of $c$. For example, when $c = 135$, the power of $U_4$ is 0.72, whilst the power of $U_r$ is 0.50. $U_4$ tracks the power of $U_\mu$, with power losses of, at most, 0.13 when $c = 117$. In Figure 3(b), $\mu = 0.25$, and the power curve of $U_4$ sits slightly above $U_r$ and slightly below $U_\mu$ for all values of $c$. As the trend coefficient increases in
Figures 3(c)-3(f), $U_4$ outperforms $U_\mu$ for all values of $c$, whilst tracking underneath the power curve of $U_\tau$. For example, when $\kappa = 4$, in Figure 3(f), $U_4$ has a maximum power loss of 0.21 relative to $U_\tau$, but a power gain of 0.49 relative to $U_\mu$. A similar pattern of results emerges when considering $g^2 = 10$ in Figure 4 and $g^2 = 50$ in Figure 5. Across all settings of $\kappa$, $U_4$ has superior power performance to the worst-performing test of either $U_\mu$ and $U_\tau$ whilst tracking underneath the power of the best-performing test. These results demonstrate that $U_4$ provides a strategy for unit root testing against ESTAR alternatives which offers decent power levels across different trend magnitudes, including the case of no trend, as well as across the degree of nonlinearity under the ESTAR alternative.

4.3 Use of trend detection

A potential drawback of combining all four individual tests is that as the number of tests within the union of rejections procedure increases, the asymptotic scaling value, $\lambda_\alpha$, also increases to ensure the overall procedure is asymptotically correctly sized. As $\lambda_\alpha$ increases, so too does the chance that the union of rejections procedure will fail to reject the null hypothesis of a unit root even if one or more of the individual tests is able to reject at the same level of significance. In a related setting, Harvey et al. (2012) note that use of information regarding the presence of a trend can be applied to reduce the number of tests entering the union of rejections in certain cases, reducing the extent of the required critical value scaling. Given that the asymptotic size and power of $U_\mu$ decreases as the trend magnitude increases, it follows that the $DF_\mu$ and $KSS_\mu$ tests are redundant in $U_4$ for large trends. This implies that $U_4$ is correctly sized for $\beta = 0$ but under-sized elsewhere, resulting in a loss of power. Therefore, for large enough values of the trend, we want to eliminate $U_\mu$ from our testing strategy and use $U_\tau$ rather than $U_4$.

Following in the spirit of Harvey et al. (2012), we propose using a hybrid testing strategy where a BIC procedure is used to detect the presence of a trend. If evidence of a trend is apparent, we can be confident that a trend of reasonably large magnitude is present in the data, and hence that use of $U_\tau$ is appropriate, with little to be gained by also including $DF_\mu$ and $KSS_\mu$ in the union of rejections. However, if no trend is detected by the BIC procedure, it would be unwise to confidently infer that no trend exists and use only $U_\mu$, since a small magnitude trend could still be present in the data (which could have a substantial impact on the power of $U_\mu$), but its magnitude not sufficient to have triggered detection by the BIC procedure. Consequently, if no trend is detected by the BIC, we default to the risk-averse approach of using $U_4$, retaining the $DF_\tau$ and $KSS_\tau$ tests as well as the $DF_\mu$ and $KSS_\mu$ tests in the union of rejections. This approach therefore uses the information from the BIC in an indicative manner, rather than as a standard model selection procedure, consistent with the approach of Harvey et al. (2012).

Our procedure is devised as follows. Consider the null DGP

$$y_t = \alpha + \beta t + u_t$$

$$\Delta u_t = \varepsilon_t.$$

If $\beta = 0$ such that no trend is present, then $\Delta y_t = \varepsilon_t$, while if $\beta \neq 0$, then $\Delta y_t = \beta + \varepsilon_t$. Treating these two possibilities as regression models, we can select between them on the basis of the BIC. Specifically, consider the two corresponding residual series

$$\hat{\varepsilon}_{1t} = \Delta y_t$$

$$\hat{\varepsilon}_{2t} = \Delta y_t - \frac{\sum_{s=2}^{T} \Delta y_s}{T-1} = \Delta y_t - \frac{y_T - y_1}{T-1}.$$
The BIC for each model is then

\[
BIC_1 = \ln \left( \frac{\sum_{t=1}^{T-1} s_{1t}}{T-1} \right)
\]

\[
BIC_2 = \ln \left( \frac{\sum_{t=1}^{T-1} s_{2t}}{T-1} \right) + \ln \left( \frac{T-1}{T-1} \right).
\]

If \( BIC_2 < BIC_1 \) then there is evidence that a trend is present in \( y_t \), and therefore \( U_\tau \) should be applied in place of \( U_4 \). We can write this testing procedure formally as

\[
U^* = \begin{cases} 
U_\tau & \text{if } BIC_2 < BIC_1 \\
U_4 & \text{otherwise}
\end{cases}
\]

When no trend is present in the DGP, \( U^* \) will select \( U_4 \) with probability one in the limit. When a fixed magnitude trend is present, it is \( U_\tau \) that will be selected asymptotically with probability one. Hence \( U^* \) will be asymptotically correctly sized under no trend or a fixed trend, and would be expected to achieve greater power than simply using \( U_4 \) whenever a trend of large magnitude is present in the data. In the next section we consider the finite sample performance of \( U^* \) to quantify the potential gains that this approach can offer.

5 Finite sample simulations

In this section we consider the extent to which asymptotic power is a good indicator of finite sample performance by simulating the four individual test statistics \( DF_\mu, KSS_\mu, DF_\tau \) and \( KSS_\tau \) as well as the three union of rejections procedures \( U_\mu, U_\tau \) and \( U_4 \) using a sample of \( T = 100 \) observations. In addition, we examine the finite sample power of our trend detection-based \( U^* \) procedure. We use the same values of \( g^2 \) and \( c \) used in our asymptotic simulations. As before, we use 20,000 Monte Carlo simulations and the asymptotic critical values and scaling constants given in Table 1.

Figures 6-10 report finite sample size and power at a nominal 0.05 significance level. First, it can be observed that none of the procedures display much in the way of size distortions, with the exception of under-sizing for \( U_\mu \) in the presence of large trends (due to under-sizing in \( DF_\mu \) and \( KSS_\mu \)). Next, we find that the power results obtained are qualitatively very similar from those observed in the asymptotic case. Figures 6(a)-(f) show results for \( DF_\mu, KSS_\mu \) and \( U_\mu \). We see that for modest values of \( g^2 \) where nonlinearity is most apparent, \( KSS_\mu \) has a power advantage over \( DF_\mu \); see, for example, Figure 6(a). However, for larger values of \( g^2 \), \( DF_\mu \) is shown to have superior power to \( KSS_\mu \) as the degree of nonlinearity reduces, for example in Figures 6(d)-(f). As in the asymptotic case, the power of \( U_\mu \) dominates that of the worst-performing test for all settings of \( g^2 \) and all values of \( c \). It either closely tracks or has slightly higher power than the best-performing test across all values of \( g^2 \) and \( c \). These results are mirrored in Figures 7(a)-(f) when finite sample power results for \( DF_\tau, KSS_\tau \) and \( U_\tau \) are considered. Again, we see that \( KSS_\tau \) has higher power relative to \( DF_\tau \) for lower settings of \( g^2 \), but that the relative power ranking is reversed for higher settings of \( g^2 \). We show that \( U_\tau \) has power either close to or above the power offered by the best-performing of either \( DF_\tau \) or \( KSS_\tau \) for all values of \( c \) and \( g^2 \).

Figures 8-10 report finite sample power results of \( U_\mu, U_\tau \) and \( U_4 \) for three settings of the mean reversion parameter, \( g^2 = \{2, 10, 50\} \). We set \( \beta = \{0, 0.025, 0.05, 0.1, 0.2, 0.4\} \).
such that for $T = 100$ these values coincide with the local trend values considered in the previous asymptotic analysis. The behaviour of $U_{\mu}$, $U_{\tau}$ and $U_4$ is very similar to that observed in the asymptotic simulations. We see that the power of $U_{\mu}$ decreases as $\beta$ increases, such that in Figure 8(a), where $g^2 = 2$ and $\beta = 0$, $U_{\mu}$ has a substantial power advantage over $U_{\tau}$, but for higher values of $\beta$ the power of $U_{\mu}$ tends to zero. $U_4$ is able to capitalize on the differing power profiles of $U_{\mu}$ and $U_{\tau}$ and captures much of the power gains associated with $U_{\mu}$ for small trends, and $U_{\tau}$ for large trends. An almost identical pattern of behaviour is seen for $g^2 = 10$ in Figure 9 and $g^2 = 50$ in Figure 10.

Figures 8-10 also report results for the $U^*$ procedure which selects between $U_{\tau}$ and $U_4$ on the basis of the BIC trend detection procedure. The power of $U^*$ equals that of $U_4$ for smaller trend magnitudes, i.e. $\beta \leq 0.1$, as here the trend is not detected by the BIC. In Figure 8(e), we see that when $\beta = 0.2$, $U^*$ is selecting $U_{\tau}$ in some replications due to detection of a trend by the BIC, and correspondingly, the power of $U^*$ is marginally higher than that of $U_4$. In Figure 8(f), when $\beta = 0.4$, the trend is now detectable in all replications, and $U^*$ always selects $U_{\tau}$, so that the power of $U^*$ equals that of $U_{\tau}$, and therefore exceeds that of $U_4$ by a substantial margin. The same pattern of results is seen in Figures 9 and 10, where higher values of $g^2$ are considered. These results show that the overall power performance of $U_4$ can be improved by incorporating a trend detection procedure. In our simulations, $U^*$ never has lower power than $U_4$, but when a large trend is present in the data, $U^*$ achieves higher power than $U_4$ by defaulting to $U_{\tau}$.

6 Empirical example

To demonstrate the practical use of our proposed union of rejection procedures, we undertake an empirical application of the unit root tests considered in this paper to energy consumption data. Economists are keen to understand whether shocks to energy consumption have permanent or transitory effects. If energy consumption follows a unit root process, then shocks to world energy markets have a permanent effect. This is likely to have implications for policymakers who wish to set government targets relating to energy consumption. Additionally, an understanding of the statistical properties of energy consumption is required in order to reliably model the relationship between energy consumption and other macroeconomic variables. In recent years, a growing number of studies have been dedicated to understanding the integration properties of energy consumption. For a review of this literature see Smyth (2013). In a recent study, Hasanov and Telatar (2011) apply the four individual unit root tests $DF_{\mu}$, $DF_{\tau}$, $KSS_{\mu}$ and $KSS_{\tau}$ to annual total energy consumption per capita data obtained for 178 countries across the time period 1980-2006. At a 0.10 significance level they are able to reject the null hypothesis of a unit root for 55 out of the 178 countries using $DF_{\mu}$ and $DF_{\tau}$ tests and for 71 countries using $KSS_{\mu}$ and $KSS_{\tau}$ tests. Combining results from these different test procedures, they suggest that the null hypothesis of a unit root can be rejected in many of the countries considered. Implicitly, this strategy of testing for a unit root is equivalent to a union of rejections procedure that does not apply a scaling constant to the critical values of the individual tests. As discussed in section 4, without this scaling constant the union of rejections procedure will be over-sized. Caution must therefore be exercised over any inference made from the combination of results provided by these individual tests.

---

3Hasanov and Telatar (2011) also apply the ST-TAR unit root test of Sollis (2004) that allows for an asymmetric transition function and a gradual trend break in the data.
as the high number of rejections could be due to over-sizing.

We obtain total energy consumption per capita data from the Energy Information Administration as Hasanov and Telatar (2011) do. Our sample covers 180 countries and a slightly longer time series from 1980-2011. We apply the four individual unit root tests, $DF_\mu$, $DF_\tau$, $KSS_\mu$ and $KSS_\tau$, as well as the four union of rejections procedures considered in this paper, $U_\mu$, $U_\tau$, $U_4$ and $U^*$ to this data. Serial correlation in the data is accounted for by augmenting the test regressions with lagged differences of the dependent variable. The optimal number of lagged difference terms is chosen using the MAIC procedure of Ng and Perron (2001), with a maximum of 6. Given the small number of observations available per country, we want to minimise any size concerns that might arise with using asymptotic critical values. Su et al. (2013) show that in finite samples the $KSS_\mu$ and $KSS_\tau$ unit root test suffers from smaller size distortions when MAIC is used to select the optimal number of lags rather than the sequential downwards testing method of Ng and Perron (1995).

Results from our unit root testing exercise are given in Tables 2 and 3. For the 60 countries where at least one of the four individual tests rejects at a 0.10 significance level, the countries are listed in Table 2, detailing the significance levels at which rejections are obtained by the different tests. Table 3 lists the remaining countries for which we failed to reject the null hypothesis of a unit root using any of the four individual tests. We note that of the four individual test procedures, $KSS_\tau$ is able to reject the null hypothesis of a unit root more often than the other procedures. Indeed, at a 0.10 level of significance the difference between the number of rejections found by $KSS_\tau$ (40) is substantially higher than that found by the next most frequently rejecting test, $KSS_\mu$ (29). The pattern of rejections found by these four individual tests allows us to demonstrate the advantages of our union of rejections procedures. For example, a rejection of the null is found for Kenya (see Figure 11(c) for a plot of this series) at a 0.05 significance level using $DF_\mu$. No other individual test rejects, but $U_\mu$ picks up this rejection at a 0.05 level, and $U_4$ and $U^*$ are able to reject at a 0.10 level. For Panama (see Figure 11(d)), $DF_\tau$ is the only individual test to reject, but $U_\tau$, $U_4$ and $U^*$ also pick up this rejection, all at a 0.01 significance level. A 0.01 level rejection for the Cook Islands (see Figure 11(b)) by $KSS_\mu$ results in a 0.01 level rejection by $U_\mu$ and 0.05 level rejections by $U_4$ and $U^*$. Finally, a 0.01 level rejection found for Portugal (see Figure 11(e)) by only $KSS_\tau$ is picked up by the three relevant union procedures at the same significance level. We can also highlight countries where rejections are found only in the two demeaned tests (Suriname, see Figure 11(f)), or only in the two detrended tests (Bangladesh, see Figure 11(a)). Again, these rejections are then picked up by the relevant union procedures.

It is also to be expected from the simulation results that there will be cases where the union of rejections procedures fail to pick up some rejections found by individual tests, and we can see examples of this in the results, for example there are a number of countries for which one of the individual tests rejects at the 0.10 level, but $U^*$ does not reject at conventional significance levels. There is also one case (Bhutan), where $U^*$ fails to reject when $U_4$ rejects at the 0.01 level; here, the BIC is suggesting the presence of a trend, but the rejection is lost since it only arises from application of $KSS_\mu$. Overall, however, the results clearly demonstrate the ability of the union of rejections procedures to capitalize on the differing rejections offered by the four individual unit root tests. In total, the union procedure $U_\mu$ is able to reject the null for a higher proportion of countries than the individual test $DF_\mu$ at all conventional significance levels, and rejects a higher proportion of countries than $KSS_\mu$ at a 0.05 significance level. Similarly, $U_\tau$.
rejects the null more frequently than either $DF_r$ or $KSS_r$ at all conventional levels of significance. Our proposed $U_4$ procedure is able to further capitalize on the differing rejections offered by demeaned and detrended unit root tests, and rejects for a higher proportion of countries than $U_4$ at all conventional levels of significance and $U_r$ at a 0.10 level. Finally, our BIC-based strategy, $U^*$, rejects the same number of times as $U_4$ and $U_r$ at a 0.01 level of significance and a higher proportion of the time at both 0.05 and 0.10 levels. Therefore we conclude that the highest number of rejections across all 8 unit root testing procedures considered in this exercise was achieved by $U^*$ at 0.05 and 0.10 levels and by $U^*, U_4$ and $U_r$ at a 0.01 level. This provides clear justification for the use of our proposed union of rejections procedures in empirical applications. We can also conclude that, compared with Hasanov and Telatar (2011), at a 0.10 significance level we find evidence of stationarity in, at most, 45 countries out of 180. This suggests that nonstationarity in energy consumption may be less prevalent than previously thought.

7 Conclusion

In this paper, we have examined the power performance of DF and KSS unit root tests against an alternative hypothesis of ESTAR stationarity. Our analysis focused on prior demeaned, and demeaned and detrended test statistics. We highlighted the sensitivity of these tests to both the degree of nonlinearity under the ESTAR alternative and to the presence or otherwise of a linear deterministic trend. In practice, these are likely to be sources of uncertainty in unit root testing. We therefore proposed four union of rejections strategies that attempt to mitigate against this uncertainty while capitalizing on the power available from each individual DF and KSS test under different degrees of nonlinearity and trend magnitudes. Both asymptotic and finite sample power simulations were undertaken, and we found that a union of rejections procedure that combines all four individual tests offers attractive levels of power across different nonlinearity and trend settings. Also, a hybrid approach that uses information from a BIC approach to trend detection to select between this four-way union of rejections and a union of rejections based on just the trend based DF and KSS variants, was found to improve power still further when the magnitude of the trend is large. An empirical example using energy consumption data highlighted the potential usefulness of this testing strategy in practice.

References


Appendix

Due to the invariance of all statistics to $\mu$ in (4), we set $\mu = 0$ in what follows, without loss of generality.

Proof of Lemma 1

The $DF_{\mu}$ test statistic is given by

$$DF_{\mu} = \frac{\sum_{t=2}^{T} y_{\mu,t-1} \Delta y_t}{\sqrt{\hat{\sigma}^2 \sum_{t=2}^{T} y_{\mu,t-1}^2}}$$

where $\hat{\sigma}^2$ denotes the error variance estimate from the OLS regression of $\Delta y_t$ on $y_{\mu,t-1}$.

Considering first the numerator of $DF_{\mu}$, we can write

$$T^{-1} \sum_{t=2}^{T} y_{\mu,t-1} \Delta y_t = T^{-1} \sum_{t=2}^{T} (y_{t-1} - \bar{y}) \left( \beta - \frac{c}{T} u_{t-1} G (\theta, u_{t-1}) + \varepsilon_t \right)$$

$$= -cT^{-2} \sum_{t=2}^{T} (y_{t-1} - \bar{y}) u_{t-1} G (\theta, u_{t-1}) + T^{-1} \sum_{t=2}^{T} (y_{t-1} - \bar{y}) \varepsilon_t + o_p(1).$$

Evaluating each term separately, and defining $\bar{t} = T^{-1} \sum_{t=1}^{T} t$, $\bar{u} = T^{-1} \sum_{t=1}^{T} u_t$,

$$cT^{-2} \sum_{t=2}^{T} (y_{t-1} - \bar{y}) u_{t-1} G (\theta, u_{t-1})$$

$$= cT^{-2} \sum_{t=2}^{T} \{ \beta (t - 1 - \bar{t}) + u_{t-1} - \bar{u} \} u_{t-1} G (\theta, u_{t-1})$$

$$= cT^{-1} \sum_{t=2}^{T} \{ \kappa \sigma T^{-1} (t - \bar{t}) + T^{-1/2} (u_{t-1} - \bar{u}) \} T^{-1/2} u_{t-1} G (\theta, u_{t-1}) + o_p(1)$$

$$\stackrel{d}{\rightarrow} \sigma^2 c \int_{0}^{1} \left( \kappa \left( r - \frac{1}{2} \right) + X_\mu(r) \right) X(r) G \left( g^2 X(r)^2 \right) dr$$

and

$$T^{-1} \sum_{t=2}^{T} (y_{t-1} - \bar{y}) \varepsilon_t = T^{-1/2} \sum_{t=2}^{T} \{ \kappa \sigma T^{-1} (t - \bar{t}) + T^{-1/2} (u_{t-1} - \bar{u}) \} \varepsilon_t + o_p(1)$$

$$\stackrel{d}{\rightarrow} \sigma^2 \int_{0}^{1} \left( \kappa \left( r - \frac{1}{2} \right) + X_\mu(r) \right) dW(r).$$

In the denominator of $DF_{\mu}$, it is easily shown that $\hat{\sigma}^2 \overset{p}{\rightarrow} \sigma^2$, and

$$T^{-2} \sum_{t=2}^{T} (y_{t-1} - \bar{y})^2 = T^{-1} \sum_{t=2}^{T} \left( \kappa \sigma T^{-1} (t - \bar{t}) + T^{-1/2} (u_{t-1} - \bar{u}) \right)^2$$

$$\stackrel{d}{\rightarrow} \sigma^2 \int_{0}^{1} \left( \kappa \left( r - \frac{1}{2} \right) + X_\mu(r) \right)^2 dr.$$
The $DF_{\mu}$ test statistic therefore has the limiting distribution
\[
DF_{\mu} \xrightarrow{d} \frac{\int_0^1 \left( \kappa \left( r - \frac{1}{2} \right) + X_{\mu}(r) \right) \, dW(r) - c \int_0^1 \left( \kappa \left( r - \frac{1}{2} \right) + X_{\mu}(r) \right) X(r)G \left( g^2 X(r)^2 \right) \, dr}{\sqrt{\int_0^1 \left( \kappa \left( r - \frac{1}{2} \right) + X_{\mu}(r) \right)^2 \, dr}}.
\]

The $KSS_{\mu}$ test statistic is given by
\[
KSS_{\mu} = \frac{\sum_{t=2}^{T} y_{\mu,t-1}^3 \Delta y_t}{\sqrt{\hat{\sigma}^2 \sum_{t=2}^{T} y_{\mu,t-1}^6}}
\]
where $\hat{\sigma}^2$ denotes the error variance estimate from the OLS regression of $\Delta y_t$ on $y_{\mu,t-1}^3$.

In the numerator,
\[
T^{-2} \sum_{t=2}^{T} y_{\mu,t-1}^3 \Delta y_t = T^{-2} \sum_{t=2}^{T} (y_{t-1} - \bar{y})^3 \left( \beta - \frac{c}{T} u_{t-1} G (\theta, u_{t-1}) + \varepsilon_t \right)
\]
\[
= \kappa \sigma T^{-3/2} \sum_{t=2}^{T} (y_{t-1} - \bar{y})^3 - c T^{-3} \sum_{t=2}^{T} (y_{t-1} - \bar{y})^3 u_{t-1} G (\theta, u_{t-1})
\]
\[
+ T^{-2} \sum_{t=2}^{T} (y_{t-1} - \bar{y})^3 \varepsilon_t
\]
\[
\xrightarrow{d} \kappa \sigma^4 \int_0^1 \left( \kappa \left( r - \frac{1}{2} \right) + X_{\mu}(r) \right)^3 \, dr
\]
\[
- c \sigma^4 \int_0^1 \left( \kappa \left( r - \frac{1}{2} \right) + X_{\mu}(r) \right)^3 X(r)G \left( g^2 X(r)^2 \right) \, dr
\]
\[
+ \sigma^4 \int_0^1 \left( \kappa \left( r - \frac{1}{2} \right) + X_{\mu}(r) \right)^3 dW(r)
\]
while in the denominator, $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$ and
\[
T^{-4} \sum_{t=2}^{T} y_{\mu,t-1}^6 \xrightarrow{d} \sigma^6 \int_0^1 \left( \kappa \left( r - \frac{1}{2} \right) + X_{\mu}(r) \right)^6 \, dr
\]
giving the limit for $KSS_{\mu}$ as
\[
KSS_{\mu} \xrightarrow{d} \frac{\kappa \int_0^1 \left( \kappa \left( r - \frac{1}{2} \right) + X_{\mu}(r) \right)^3 \, dr + \int_0^1 \left( \kappa \left( r - \frac{1}{2} \right) + X_{\mu}(r) \right)^3 dW(r)}{\sqrt{\int_0^1 \left( \kappa \left( r - \frac{1}{2} \right) + X_{\mu}(r) \right)^6 \, dr}}.
\]

The test statistics based on demeaned and detrended data are invariant to $\beta$, hence we set $\beta = 0$ without loss of generality in the remainder of this proof. The $DF_r$ test statistic is
\[
DF_r = \frac{\sum_{t=2}^{T} y_{r,t-1} \Delta y_{r,t}}{\sqrt{\hat{\sigma}^2 \sum_{t=2}^{T} y_{r,t-1}^2}}
\]
where $\hat{\sigma}^2$ denotes the error variance estimate from the OLS regression of $\Delta y_{r,t}$ on $y_{r,t-1}$. Here,

$$
\Delta y_{r,t} = \Delta y_t - \hat{\beta}
$$

$$
= -\frac{c}{T} u_{t-1} G(\theta, u_{t-1}) + \varepsilon_t - \hat{\beta}
$$

and

$$
T^{1/2} \hat{\beta} = \frac{T^{-5/2} \sum_{t=1}^{T} (t - \bar{t}) y_t}{T^{-3} \sum_{t=1}^{T} (t - \bar{t})^2}
$$

$$
d \rightarrow 12 \int_{0}^{1} \left( r - \frac{1}{2} \right) X(r) dr.
$$

Also,

$$
T^{-1/2} y_{r,[rT]} = T^{-1/2} y_{rT} - T^{-1/2} \hat{\mu} - T^{-1/2} \hat{\beta} r
$$

$$
= T^{-1/2} y_{rT} - \left( T^{-1/2} \bar{y} - T^{1/2} \hat{\beta} T^{-1/2} \bar{t} \right) - T^{1/2} \hat{\beta} r
$$

$$
d \rightarrow X(r) - \int_{0}^{1} X(s) ds - 12 \left( r - \frac{1}{2} \right) \int_{0}^{1} \left( s - \frac{1}{2} \right) X(s) ds
$$

$$
= X(r) - (6r - 4) \int_{0}^{1} X(s) ds + (6 - 12r) \int_{0}^{1} s X(s) ds
$$

$$
= X_r (r).
$$

The numerator of $DF_r$ is then given by

$$
T^{-1} \sum_{t=2}^{T} y_{r,t-1} \Delta y_{r,t} = T^{-1} \sum_{t=2}^{T} y_{r,t-1} \varepsilon_t - c T^{-2} \sum_{t=2}^{T} y_{r,t-1} u_{t-1} G(\theta, u_{t-1}) - T^{1/2} \hat{\beta} T^{-3/2} \sum_{t=2}^{T} y_{r,t-1}
$$

$$
d \rightarrow \sigma^2 \int_{0}^{1} X_r (r) dW(r) - c \sigma^2 \int_{0}^{1} X_r (r) X(r) G\left(g^2 X(r)^2\right) dr
$$

and for the denominator we obtain

$$
T^{-2} \sum_{t=2}^{T} y_{r,t-1}^2 \rightarrow \sigma^2 \int_{0}^{1} X_r (r)^2 dr
$$

giving

$$
DF_r \rightarrow \frac{\int_{0}^{1} X_r (r) dW(r) - c \int_{0}^{1} X_r (r) X(r) G\left(g^2 X(r)^2\right) dr}{\sqrt{\int_{0}^{1} X_r (r)^2 dr}}.
$$

Finally, the $KSS_r$ test statistic is

$$
KSS_r = \frac{\sum_{t=2}^{T} y_{r,t-1}^2 \Delta y_{r,t}}{\sqrt{\sigma^2 \sum_{t=2}^{T} y_{r,t-1}^6}}
$$
where $\hat{\sigma}^2$ is the error variance estimate from the OLS regression of $\Delta y_{r,t}$ on $y_{r,t-1}$. Now $\hat{\sigma}^2 \overset{P}{\rightarrow} \sigma^2$ as before,

$$T^{-2} \sum_{t=2}^{T} y_{r,t-1}^3 \Delta y_{r,t} = T^{-2} \sum_{t=2}^{T} y_{r,t-1}^3 \varepsilon_t - cT^{-3} \sum_{t=2}^{T} y_{r,t-1}^3 u_{t-1} G (\theta, u_{t-1}) - T^{1/2} \beta T^{-5/2} \sum_{t=2}^{T} y_{r,t-1}^3$$

and

$$\frac{d}{d} \sigma^2 \int_{0}^{1} X_r(r)^3 dW(r) - c \sigma^2 \int_{0}^{1} X_r(r)^3 X(r) G (g^2 X(r)^2) \, dr - 12 \int_{0}^{1} \left( r - \frac{1}{2} \right) X(r) dr \int_{0}^{1} X_r(r)^3 \, dr$$

so

$$KSS_r \overset{d}{\rightarrow} \int_{0}^{1} X_r(r)^3 dW(r) - c \int_{0}^{1} X_r(r)^3 X(r) G (g^2 X(r)^2) \, dr - 12 \int_{0}^{1} \left( r - \frac{1}{2} \right) X(r) dr \int_{0}^{1} X_r(r)^3 \, dr \sqrt{\int_{0}^{1} X_r(r)^6 \, dr}$$

**Proof of Lemma 2**

As before, the $DF_\mu$ test statistic is given by

$$DF_\mu = \frac{\sum_{t=2}^{T} y_{\mu,t-1} \Delta y_t}{\sqrt{\hat{\sigma}^2 \sum_{t=2}^{T} y_{\mu,t-1}^2}}$$

where $\hat{\sigma}^2$ is the error variance estimate from the OLS regression of $\Delta y_t$ on $y_{\mu,t-1}$. We find

$$T^{-3/2} \sum_{t=2}^{T} y_{\mu,t-1} \Delta y_t = T^{-3/2} \sum_{t=2}^{T} \left( \sigma \kappa (t - \bar{t}) + u_{t-1} - \bar{u} \right) \left( \sigma \kappa + \varepsilon_t - \frac{c}{T} u_{t-1} G (\theta, u_{t-1}) \right) + o_p (1)$$

$$= \sigma \kappa T^{-3/2} \sum_{t=2}^{T} (t - \bar{t}) \varepsilon_t - \sigma \kappa c T^{-5/2} \sum_{t=2}^{T} (t - \bar{t}) u_{t-1} G (\theta, u_{t-1}) + o_p (1)$$

$$\overset{d}{\rightarrow} \sigma^2 \kappa \int_{0}^{1} \left( r - \frac{1}{2} \right) dW(r) - \sigma^2 \kappa c \int_{0}^{1} \left( r - \frac{1}{2} \right) X(r) G (g^2 X(r)^2) \, dr$$

and

$$T^{-3} \sum_{t=2}^{T} y_{\mu,t-1}^2 = T^{-3} \sum_{t=2}^{T} \left( \beta (t - 1 - \bar{t}) + u_{t-1} - \bar{u} \right)^2$$

$$= \sigma^2 \kappa^2 T^{-3} \sum_{t=2}^{T} (t - \bar{t})^2 + o_p (1)$$

$$\overset{P}{\rightarrow} \sigma^2 \kappa^2 / 12.$$
Also,

\[ \hat{\sigma}^2 = T^{-1} \sum_{t=2}^{T} \left( \Delta y_t - \frac{\sum_{s=2}^{T} (y_{s-1} - \bar{y}) \Delta y_s}{\sum_{s=2}^{T} (y_{s-1} - \bar{y})^2} (y_{t-1} - \bar{y}) \right)^2 \]

\[ = T^{-1} \sum_{t=2}^{T} \left( \Delta y_t - \frac{T^{-3/2} \sum_{s=2}^{T} (y_{s-1} - \bar{y}) \Delta y_s}{T^{-3} \sum_{s=2}^{T} (y_{s-1} - \bar{y})^2} T^{-3/2} (y_{t-1} - \bar{y}) \right)^2 \]

\[ = T^{-1} \sum_{t=2}^{T} (\Delta y_t)^2 + o_p(1) \]

\[ = \sigma^2 \kappa^2 + T^{-1} \sum_{t=2}^{T} \varepsilon_t^2 + o_p(1) \]

\[ \xrightarrow{p} \sigma^2 (\kappa^2 + 1) . \]

Therefore the \( DF_\mu \) test statistic has the limiting distribution

\[ DF_\mu \xrightarrow{d} \int_0^1 \left( r - \frac{1}{2} \right) dW(s) - c \int_0^1 \left( r - \frac{1}{2} \right) X(r)G (g^2 X(r)^2) dr \]

\[ \sqrt{(\kappa_0^2 + 1) / 12} . \]

For \( KSS_\mu \) we again have

\[ KSS_\mu = \frac{\sum_{t=2}^{T} y_{\mu,t-1}^3 \Delta y_t}{\sqrt{\hat{\sigma}^2 \sum_{t=2}^{T} y_{\mu,t-1}^6}} \]

where \( \hat{\sigma}^2 \) is the error variance estimate from the OLS regression of \( \Delta y_t \) on \( y_{\mu,t-1}^3 \). We obtain

\[ T^{-7/2} \sum_{t=2}^{T} y_{\mu,t-1}^3 \Delta y_t = T^{-7/2} \sum_{t=2}^{T} \left( \sigma \kappa (t - \bar{t}) + u_{t-1} - \bar{u} \right)^3 \left( \sigma \kappa - \frac{C}{T} u_{t-1} G(\theta, u_{t-1}) + \varepsilon_t \right) + o_p(1) \]

\[ = 3 \sigma^3 \kappa^3 T^{-7/2} \sum_{t=2}^{T} (t - \bar{t})^2 (u_{t-1} - \bar{u}) + \sigma^3 \kappa^3 T^{-7/2} \sum_{t=2}^{T} (t - \bar{t})^3 \varepsilon_t \]

\[ - c \sigma^3 \kappa^3 T^{-9/2} \sum_{t=2}^{T} (t - \bar{t})^3 u_{t-1} G(\theta, u_{t-1}) + o_p(1) \]

\[ \xrightarrow{d} 3 \sigma^4 \kappa^3 \int_0^1 (r - \frac{1}{2})^2 X_\mu(r) dr + \sigma^4 \kappa^3 \int_0^1 (r - \frac{1}{2})^3 dW(r) \]

\[ - c \sigma^4 \kappa^3 \int_0^1 (r - \frac{1}{2})^3 X(r)G (g^2 X(r)^2) dr \]

and

\[ T^{-7} \sum_{t=2}^{T} y_{\mu,t-1}^6 = \sigma^6 \kappa^6 T^{-7} \sum_{t=2}^{T} (t - \bar{t})^6 + o_p(1) \]

\[ \xrightarrow{p} \sigma^6 \kappa^6 / 448 \]

22
together with

\[ \sigma^2 = T^{-1} \sum_{t=2}^{T} \left( \Delta y_t - \frac{T^{-7/2} \sum_{s=2}^{T} (y_{s-1} - \bar{y})^3 \Delta y_s T^{-7/2} (y_{t-1} - \bar{y})^3}{T^{-7} \sum_{s=2}^{T} (y_{s-1} - \bar{y})^6} \right)^2 \]

\[ = T^{-1} \sum_{t=2}^{T} (\Delta y_t)^2 + o_p(1) \]

\[ \xrightarrow{p} \sigma^2 (\kappa^2 + 1) \]

giving the \(KSS_\mu\) limit distribution

\[ KSS_\mu \xrightarrow{d} 3 \int_0^1 (r - \frac{1}{2})^2 X_\mu(r) dr + \int_0^1 (r - \frac{1}{2})^3 dW(r) - c \int_0^1 (r - \frac{1}{2})^3 X(r) \mathcal{G}(g^2 X(r)^2) dr \]

\[ \frac{1}{\sqrt{(\kappa^2 + 1)/448}} \]

23
Table 1: Asymptotic critical values of $KSS_\mu$, $DF_\mu$, $KSS_\tau$, $DF_\tau$ and union of rejection scaling values $\lambda^\alpha_\mu$, $\lambda^\alpha_\tau$ and $\lambda^\alpha_k$ for significance level $\alpha$

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$KSS_\mu$</th>
<th>$DF_\mu$</th>
<th>$KSS_\tau$</th>
<th>$DF_\tau$</th>
<th>$\lambda^\alpha_\mu$</th>
<th>$\lambda^\alpha_\tau$</th>
<th>$\lambda^\alpha_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>-2.655</td>
<td>-2.564</td>
<td>-3.118</td>
<td>-3.122</td>
<td>1.063</td>
<td>1.056</td>
<td>1.140</td>
</tr>
<tr>
<td>0.05</td>
<td>-2.935</td>
<td>-2.864</td>
<td>-3.396</td>
<td>-3.409</td>
<td>1.059</td>
<td>1.049</td>
<td>1.118</td>
</tr>
<tr>
<td>0.01</td>
<td>-3.471</td>
<td>-3.424</td>
<td>-3.939</td>
<td>-3.959</td>
<td>1.053</td>
<td>1.041</td>
<td>1.085</td>
</tr>
</tbody>
</table>

T.1
Table 2: Total primary energy consumption per capita: $DF_\mu$, $DF_\tau$, $KSS_\mu$, $KSS_\tau$, $U_\mu$, $U_\tau$, $U_4$ and $U^*$ unit root test results

<table>
<thead>
<tr>
<th></th>
<th>$DF_\mu$</th>
<th>$DF_\tau$</th>
<th>$KSS_\mu$</th>
<th>$KSS_\tau$</th>
<th>$U_\mu$</th>
<th>$U_\tau$</th>
<th>$U_4$</th>
<th>$U^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algeria</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>American Samoa</td>
<td>**</td>
<td>***</td>
<td>***</td>
<td>**</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Antigua and Barbuda</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Argentina</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Bahamas</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bahrain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bangladesh</td>
<td>***</td>
<td></td>
<td>**</td>
<td></td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Bermuda</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bhutan</td>
<td></td>
<td></td>
<td>***</td>
<td>***</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burkina Faso</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cambodia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Cayman Islands</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central African Republic</td>
<td>**</td>
<td></td>
<td>***</td>
<td>*</td>
<td>**</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Chad</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comoros</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cook Islands</td>
<td>***</td>
<td>***</td>
<td></td>
<td></td>
<td>**</td>
<td></td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Djibouti</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dominica</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Egypt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>El Salvador</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Faroe Islands</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Finland</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>French Polynesia</td>
<td>***</td>
<td></td>
<td>***</td>
<td>***</td>
<td>**</td>
<td>**</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Greenland</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>**</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Grenada</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Guadeloupe</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Guinea</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>***</td>
<td>*</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Honduras</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>***</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Hong Kong</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>India</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Jordan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
<td>*</td>
<td>**</td>
<td>*</td>
</tr>
<tr>
<td>Kenya</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Kuwait</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
<td></td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Lebanon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Lesotho</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Libya</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Maldives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Mali</td>
<td>***</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Mozambique</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Nauru</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Netherlands Antilles</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>New Caledonia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
<td>**</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Niger</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Country</td>
<td>$DF_\mu$</td>
<td>$DF_\tau$</td>
<td>$KSS_\mu$</td>
<td>$KSS_\tau$</td>
<td>$U_\mu$</td>
<td>$U_\tau$</td>
<td>$U_4$</td>
<td>$U^*$</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>----------</td>
<td>-----------</td>
<td>-----------</td>
<td>------------</td>
<td>---------</td>
<td>---------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>Panama</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Portugal</td>
<td></td>
<td></td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Puerto Rico</td>
<td>**</td>
<td>**</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Saint Lucia</td>
<td>**</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Samoa</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Sao Tome and Principe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senegal</td>
<td>**</td>
<td>**</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Seychelles</td>
<td>**</td>
<td>**</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Sierra Leone</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Solomon Islands</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Suriname</td>
<td>**</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Swaziland</td>
<td>***</td>
<td>**</td>
<td>***</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Tonga</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Trinidad and Tobago</td>
<td>**</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Uganda</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>United Arab Emirates</td>
<td>***</td>
<td>**</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Uruguay</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Rejections at 0.10</td>
<td>12</td>
<td>25</td>
<td>29</td>
<td>40</td>
<td>28</td>
<td>41</td>
<td>43</td>
<td>45</td>
</tr>
<tr>
<td>Total Rejections at 0.05</td>
<td>9</td>
<td>18</td>
<td>19</td>
<td>28</td>
<td>20</td>
<td>32</td>
<td>31</td>
<td>34</td>
</tr>
<tr>
<td>Total Rejections at 0.01</td>
<td>5</td>
<td>9</td>
<td>12</td>
<td>14</td>
<td>11</td>
<td>19</td>
<td>19</td>
<td>19</td>
</tr>
</tbody>
</table>

*, **, *** indicate rejections at a 0.10, 0.05, 0.01 significance level respectively
Table 3: Total primary energy consumption per capita: countries for which no unit root test rejections found

<table>
<thead>
<tr>
<th>Afghanistan</th>
<th>Gibraltar</th>
<th>Pakistan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albania</td>
<td>Greece</td>
<td>Papua New Guinea</td>
</tr>
<tr>
<td>Angola</td>
<td>Guam</td>
<td>Paraguay</td>
</tr>
<tr>
<td>Australia</td>
<td>Guatemala</td>
<td>Peru</td>
</tr>
<tr>
<td>Austria</td>
<td>Guinea-Bissau</td>
<td>Philippines</td>
</tr>
<tr>
<td>Barbados</td>
<td>Haiti</td>
<td>Poland</td>
</tr>
<tr>
<td>Belgium</td>
<td>Hungary</td>
<td>Qatar</td>
</tr>
<tr>
<td>Belize</td>
<td>Iceland</td>
<td>Reunion</td>
</tr>
<tr>
<td>Benin</td>
<td>Indonesia</td>
<td>Romania</td>
</tr>
<tr>
<td>Bolivia</td>
<td>Iran</td>
<td>Rwanda</td>
</tr>
<tr>
<td>Botswana</td>
<td>Iraq</td>
<td>Saint Kitts and Nevis</td>
</tr>
<tr>
<td>Brazil</td>
<td>Ireland</td>
<td>Saint Pierre and Miquelon</td>
</tr>
<tr>
<td>British Virgin Islands</td>
<td>Israel</td>
<td>Saint Vincent/Grenadines</td>
</tr>
<tr>
<td>Brunei</td>
<td>Italy</td>
<td>Saudi Arabia</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>Jamaica</td>
<td>Singapore</td>
</tr>
<tr>
<td>Burma (Myanmar)</td>
<td>Japan</td>
<td>Somalia</td>
</tr>
<tr>
<td>Burundi</td>
<td>Kiribati</td>
<td>South Africa</td>
</tr>
<tr>
<td>Cameroon</td>
<td>Laos</td>
<td>South Korea</td>
</tr>
<tr>
<td>Canada</td>
<td>Liberia</td>
<td>Spain</td>
</tr>
<tr>
<td>Cape Verde</td>
<td>Luxembourg</td>
<td>Sri Lanka</td>
</tr>
<tr>
<td>Chile</td>
<td>Macau</td>
<td>Sudan and South Sudan</td>
</tr>
<tr>
<td>China</td>
<td>Madagascar</td>
<td>Sweden</td>
</tr>
<tr>
<td>Colombia</td>
<td>Malawi</td>
<td>Switzerland</td>
</tr>
<tr>
<td>Congo (Brazzaville)</td>
<td>Malaysia</td>
<td>Syria</td>
</tr>
<tr>
<td>Congo (Kinshasa)</td>
<td>Malta</td>
<td>Taiwan</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>Martinique</td>
<td>Tanzania</td>
</tr>
<tr>
<td>Cote d’Ivoire</td>
<td>Mauritania</td>
<td>Thailand</td>
</tr>
<tr>
<td>Cuba</td>
<td>Mauritius</td>
<td>Togo</td>
</tr>
<tr>
<td>Cyprus</td>
<td>Mexico</td>
<td>Tunisia</td>
</tr>
<tr>
<td>Denmark</td>
<td>Mongolia</td>
<td>Turkey</td>
</tr>
<tr>
<td>Dominican Republic</td>
<td>Montserrat</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>Ecuador</td>
<td>Morocco</td>
<td>United States</td>
</tr>
<tr>
<td>Equatorial Guinea</td>
<td>Nepal</td>
<td>United States Virgin Islands</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>Netherlands</td>
<td>Vanuatu</td>
</tr>
<tr>
<td>Fiji</td>
<td>New Zealand</td>
<td>Venezuela</td>
</tr>
<tr>
<td>France</td>
<td>Nicaragua</td>
<td>Vietnam</td>
</tr>
<tr>
<td>French Guiana</td>
<td>Nigeria</td>
<td>Western Sahara</td>
</tr>
<tr>
<td>Gabon</td>
<td>North Korea</td>
<td>Yemen</td>
</tr>
<tr>
<td>Gambia</td>
<td>Norway</td>
<td>Zambia</td>
</tr>
<tr>
<td>Ghana</td>
<td>Oman</td>
<td>Zimbabwe</td>
</tr>
</tbody>
</table>
Figure 1: Asymptotic size and local power of $KSS_\mu$, $DF_\mu$ and $U_\mu$ for fixed $g^2 (\kappa = 0)$

(a) $g^2 = 2$

(b) $g^2 = 5$

(c) $g^2 = 10$

(d) $g^2 = 20$

(e) $g^2 = 50$

(f) $g^2 = 100$

$KSS_\mu$, $DF_\mu$, $U_\mu$
Figure 2: Asymptotic size and local power of $KSS_\tau$, $DF_\tau$ and $U_\tau$ for fixed $g^2$

(a) $g^2 = 2$

(b) $g^2 = 5$

(c) $g^2 = 10$

(d) $g^2 = 20$

(e) $g^2 = 50$

(f) $g^2 = 100$

--- $KSS_\tau$, --- $DF_\tau$, $U_\tau$
Figure 3: Asymptotic size and local power of $U_\mu$, $U_\tau$ and $U_4$ for fixed $g^2 = 2$

(a) $\kappa = 0$

(b) $\kappa = 0.25$

(c) $\kappa = 0.5$

(d) $\kappa = 1.0$

(e) $\kappa = 2.0$

(f) $\kappa = 4.0$

$\mu$, $\tau$, $U_4$
Figure 4: Asymptotic size and local power of $U_\mu$, $U_\tau$ and $U_4$ for fixed $g^2 = 10$

(a) $\kappa = 0$

(b) $\kappa = 0.25$

(c) $\kappa = 0.5$

(d) $\kappa = 1.0$

(e) $\kappa = 2.0$

(f) $\kappa = 4.0$

---

$U_\mu$, $U_\tau$, $U_4$
Figure 5: Asymptotic size and local power of $U_\mu$, $U_\tau$ and $U_4$ for fixed $g^2 = 50$

(a) $\kappa = 0$

(b) $\kappa = 0.25$

(c) $\kappa = 0.5$

(d) $\kappa = 1.0$

(e) $\kappa = 2.0$

(f) $\kappa = 4.0$

$-- U_\mu$, $-- U_\tau$, $- U_4$
Figure 6: Finite sample size and local power of $KSS_{\mu}$, $DF_{\mu}$ and $U_{\mu}$ for fixed $g^2 (\beta = 0)$
Figure 7: Finite sample size and local power of $KSS_\tau$, $DF_\tau$ and $U_\tau$ for fixed $g^2$

(a) $g^2 = 2$

(b) $g^2 = 5$

(c) $g^2 = 10$

(d) $g^2 = 20$

(e) $g^2 = 50$

(f) $g^2 = 100$

--- $KSS_\tau$, --- $DF_\tau$, --- $U_\tau$
Figure 8: Finite sample size and local power of $U_\mu$, $U_\tau$, $U_4$ and $U^*$ for fixed $g^2 = 2$

(a) $\beta = 0$

(b) $\beta = 0.025$

(c) $\beta = 0.05$

(d) $\beta = 0.1$

(e) $\beta = 0.2$

(f) $\beta = 0.4$

--- $U_\mu$, --- $U_\tau$, — $U_4$, ♦ $U^*$

F.8
Figure 9: Finite sample size and local power of $U_{\mu}$, $U_{\tau}$, $U_4$ and $U^*$ for fixed $g^2 = 10$

(a) $\beta = 0$

(b) $\beta = 0.025$

(c) $\beta = 0.05$

(d) $\beta = 0.1$

(e) $\beta = 0.2$

(f) $\beta = 0.4$

--- $U_{\mu}$, --- $U_{\tau}$, --- $U_4$, ♦ $U^*$

F.9
Figure 10: Finite sample size and local power of $U_\mu$, $U_\tau$, $U_4$ and $U^*$ for fixed $g^2 = 50$

(a) $\beta = 0$

(b) $\beta = 0.025$

(c) $\beta = 0.05$

(d) $\beta = 0.1$

(e) $\beta = 0.2$

(f) $\beta = 0.4$

$U^*$

--- $U_\mu$, --- $U_\tau$, --- $U_4$, • $U^*$
Figure 11: Total primary energy consumption per capita (logs of million BTU per person)

(a) Bangladesh
(b) Cook Islands
(c) Kenya
(d) Panama
(e) Portugal
(f) Suriname