Granger Centre
for time series econometrics

Granger Centre Discussion Paper Series

Testing for a unit root against ESTAR stationarity

by

David I. Harvey, Stephen J. Leybourne and
Emily J. Whitehouse

Granger Centre Discussion Paper No. 17/02

University of Nottingham
UK | CHINA | MALAYSIA

Testing for a unit root against ESTAR stationarity

David I. Harvey, Stephen J. Leybourne and Emily J. Whitehouse*

School of Economics, University of Nottingham

October 20, 2016

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.

*Correspondence to: Emily Whitehouse, School of Economics, University of Nottingham, Nottingham, NG7 2RD, UK. E-mail: `emily.whitehouse@nottingham.ac.uk`

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

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

where $\varepsilon_t \sim iid(0, \sigma^2)$ and $y_1 = O_p(1)$, and ρ and γ 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

$$G(\theta, 0) = 0, \qquad \lim_{y_{t-d} \to \pm\infty} G(\theta, y_{t-d}) = 1$$

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* θ 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 θ, the degree of nonlinearity will be relatively small, while relatively large degrees of nonlinearity will be associated with intermediate values of θ.

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.

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

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 $d = \{d_1, ..., d_{\max}\}$. 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 γ is not identified under the null, KSS take the first order approximation of the Taylor expansion of the ESTAR model yielding the auxiliary equation

$$\Delta y_t = \delta y_{t-1}^3 + e_t \tag{2}$$

KSS then test the hypotheses

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

using the test statistic

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

where $\widehat{\delta}$ is the OLS estimate of δ 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

$$\Delta y_t = \delta y_{t-1} + e_t \tag{3}$$

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_μ. Both a non-zero mean and a deterministic (linear) trend can be accommodated by using the demeaned and detrended data $y_{\tau,t} = y_t - \widehat{\mu} - \widehat{\beta}t$ 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_τ. Similarly demeaned and detrended DF tests are denoted DF_μ and DF_τ. We focus on these four tests throughout this paper. Note that to illustrate the core results, our analysis assumes that ε_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 Δy_t (or $\Delta y_{\tau,t}$ 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 θ. In order to prevent the nonlinear ESTAR component from becoming asymptotically negligible, DK show that θ needs to be non-zero, but order T^{-1}. We also introduce a scaling by σ^2, so that the magnitude of the nonlinearity parameter is measured relative to the variance of ε_t, and ensuring that σ^2 does not appear in the limit distributions that follow; specifically, we set $\theta = \frac{g^2}{\sigma^2 T}$.

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

$$\begin{aligned} y_t &= \mu + \beta t + u_t, \qquad t = 2, ..., T & (4) \\ \Delta u_t &= -\frac{c}{T} u_{t-1} \left(1 - \exp\left(-\frac{g^2}{\sigma^2 T} u_{t-1}^2\right)\right) + \varepsilon_t & (5) \end{aligned}$$

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 ε_t and u_t have the following limit distributions

$$\left(\frac{1}{\sigma\sqrt{T}} \sum_{j=1}^{\lfloor rT \rfloor} \varepsilon_j \ ; \ \frac{1}{\sigma\sqrt{T}} u_{\lfloor rT \rfloor}\right) \xrightarrow{d} (W(r) \ ; \ X(r))$$

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

$$dX(r) = -cX(r)\mathcal{G}\left(g^2 X(r)^2\right) dr + dW(r)$$

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

$$\mathcal{G}\left(g^2 X(r)^2\right) = 1 - \exp\left(-g^2 X(r)^2\right).$$

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 κ a finite constant, using the appropriate Pitman drift on the trend coefficient; the scaling by σ 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_μ, KSS_μ, DF_τ and KSS_τ 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 $c \geq 0$,*

$$DF_\mu \xrightarrow{d} \frac{\int_0^1 \left(\kappa\left(r-\frac{1}{2}\right)+X_\mu(r)\right)dW(r) - c\int_0^1 \left(\kappa\left(r-\frac{1}{2}\right)+X_\mu(r)\right)X(r)\mathcal{G}\left(g^2X(r)^2\right)dr}{\sqrt{\int_0^1 \left(\kappa\left(r-\frac{1}{2}\right)+X_\mu(r)\right)^2 dr}}$$

$$KSS_\mu \xrightarrow{d} \frac{\begin{array}{c}\kappa\int_0^1 \left(\kappa\left(r-\frac{1}{2}\right)+X_\mu(r)\right)^3 dr + \int_0^1 \left(\kappa\left(r-\frac{1}{2}\right)+X_\mu(r)\right)^3 dW(r) \\ -c\int_0^1 \left(\kappa\left(r-\frac{1}{2}\right)+X_\mu(r)\right)^3 X(r)\mathcal{G}\left(g^2X(r)^2\right)dr\end{array}}{\sqrt{\int_0^1 \left(\kappa\left(r-\frac{1}{2}\right)+X_\mu(r)\right)^6 dr}}$$

$$DF_\tau \xrightarrow{d} \frac{\int_0^1 X_\tau(r)dW(r) - c\int_0^1 X_\tau(r)X(r)\mathcal{G}\left(g^2X(r)^2\right)dr}{\sqrt{\int_0^1 X_\tau(r)^2 dr}}$$

$$KSS_\tau \xrightarrow{d} \frac{\int_0^1 X_\tau(r)^3dW(r) - c\int_0^1 X_\tau(r)^3X(r)\mathcal{G}\left(g^2X(r)^2\right)dr - 12\int_0^1 \left(r-\frac{1}{2}\right)X(r)dr\int_0^1 X_\tau(r)^3dr}{\sqrt{\int_0^1 X_\tau(r)^6 dr}}$$

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

Proof: See Appendix.

The null limit distributions of the DF_μ, KSS_μ, DF_τ and KSS_τ statistics are obtained as a special case of Lemma 1, on setting $c = 0$, so that $X(r) = W(r)$. Note that when $g^2 = 0$, $\mathcal{G}\left(g^2X(s)^2\right) = 0$, 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_τ 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 $\int_0^1 W_\tau(r)^3dW(r)$, whereas the correct expression is $\int_0^1 W_\tau(r)^3dW_\tau(r)$. The fact that $dW_\tau(r) \neq dW(r)$ (since $\Delta y_{\tau,t} = \Delta y_t - \widehat{\beta}$) is the source of the discrepancy, with Lemma 1 providing the correct result.

We observe that the limiting distribution of the demeaned statistics DF_μ and KSS_μ depend on the trend term via κ, whilst the invariant detrended statistics DF_τ and KSS_τ do not. These results coincide with the behaviour of DF_μ and DF_τ reported in Harvey et al. (2009) under a linear AR DGP. Also, as $g^2 \to \infty$, $\mathcal{G}\left(g^2X(s)^2\right) \to 1$, 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_μ and KSS_μ under a fixed magnitude trend $\beta = \kappa\sigma$ where κ is a finite constant. The limiting distribution of the trend-invariant DF_τ and KSS_τ statistics do not change under a fixed trend. The limiting distributions of DF_μ and KSS_μ under the *iid* assumption for ε_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 $c \geq 0$,*

$$DF_\mu \xrightarrow{d} \frac{\int_0^1 \left(r-\frac{1}{2}\right)dW(r) - c\int_0^1 \left(r-\frac{1}{2}\right)X(r)\mathcal{G}\left(g^2X(r)^2\right)dr}{\sqrt{\left(\kappa^2+1\right)/12}}$$

$$KSS_\mu \xrightarrow{d} \frac{3\int_0^1 \left(r-\frac{1}{2}\right)^2 X_\mu(r)dr + \int_0^1 \left(r-\frac{1}{2}\right)^3 dW(r) - c\int_0^1 \left(r-\frac{1}{2}\right)^3 X(r)\mathcal{G}\left(g^2 X(r)^2\right) dr}{\sqrt{\left(\kappa^2+1\right)/448}}.$$

Proof: See Appendix.

We note that both DF_μ and KSS_μ 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_μ reduces to the Harvey et al. (2009) limit. Under the null hypothesis $c = 0$, the DF_μ 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_μ follows an asymptotic normal distribution with mean zero and variance $1/\left(\kappa^2+1\right)$. This implies that for any non-zero fixed trend, DF_μ 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_μ will follow an asymptotic normal distribution under the null, with mean zero and variance approximately $2.32/\left(\kappa^2+1\right)$ (obtained by simulation). As with DF_μ, 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_μ and KSS_μ tests when no trend is present in the DGP ($\kappa = 0$), and subsequently the DF_τ and KSS_τ 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_μ and KSS_μ 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_μ achieves higher power than DF_μ 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_μ has power of 0.64 whilst DF_μ has power of 0.49. It is clear in these circumstances that a practitioner would want to employ KSS_μ to test for a unit root. In Figure 1(b) where $g^2 = 5$, KSS_μ maintains a power advantage over DF_μ 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_μ now has power of 0.67 whilst DF_μ 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

$c = 20$, at which point the power curves diverge with DF_μ outperforming KSS_μ for these higher values of c. Figure 1(e) shows the case for $g^2 = 50$; the DF_μ test now displays higher power across almost all values of c. Finally, in Figure 1(f), $g^2 = 100$, and DF_μ displays substantially higher power than KSS_μ. For example, when $c = 22$, the power of DF_μ is 0.87 whilst KSS_μ 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_τ and KSS_τ 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 κ used in these simulations is irrelevant. Figure 2(a), where $g^2 = 2$, shows that, for modest magnitudes of the nonlinearity parameter, KSS_τ has power gains over DF_τ, due to a high degree of nonlinearity in the process. For example, for $c = 171$, KSS_τ has power of 0.61 and DF_τ 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_τ still maintains a power advantage over DF_τ, 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_τ to KSS_τ improves, to the point where DF_τ outperforms KSS_τ, in line with the degree of nonlinearity reducing. For example, for $g^2 = 100$ and $c = 30$, the power of DF_τ is 0.90 and the power of KSS_τ 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_μ and KSS_μ (for the case of no trend), and one that combines DF_τ and KSS_τ (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

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

where cv_{DFi}^α and $cv_{KSS_i}^\alpha$ denote the asymptotic null critical values of DF_i and KSS_i respectively for a significance level α. 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 λ_i^α, calculated such that the asymptotic empirical size of U_i equals the nominal size α. The decision rule can also be written as

$$U_i : \text{Reject } H_0 \text{ if } t_{U_i} = \min\left(DF_i, \frac{cv^{\alpha}_{DF_i}}{cv^{\alpha}_{KSS_i}} KSS_i\right) < \lambda^{\alpha}_i cv^{\alpha}_{DF_i}, \qquad i = \mu, \tau.$$

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:

$$t_{U_i} \xrightarrow{d} \min\left(L_{DF_i}, \frac{cv^{\gamma}_{DF_i}}{cv^{\gamma}_{KSS_i}} L_{KSS_i}\right), \qquad i = \mu, \tau$$

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 λ^{α}_i, we can simulate the limit distribution of t_{U_i} and calculate the α level critical value $cv^{\alpha}_{U_i}$. Computing $\lambda^{\alpha}_i = cv^{\alpha}_{U_i}/cv^{\alpha}_{DF_i}$ 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_μ for $g^2 = \{2, 5, 10, 20, 50, 100\}$. Figure 1(a) shows asymptotic power of U_μ for $g^2 = 2$. We see that the power of U_μ dominates that of DF_μ for all values of c, with its power curve lying just underneath that of KSS_μ for approximately $c \leq 108$, and with power rising above both KSS_μ and DF_μ beyond this point. U_μ is therefore able to capture the power advantage offered by KSS_μ in this modest nonlinearity parameter magnitude case. For $g^2 = 5$ in Figure 1(b), we see that, again, U_μ outperforms DF_μ across all values of c, and offers power that is very slightly below that of KSS_μ 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_μ closely tracks that of whichever test, KSS_μ or DF_μ, is superior for a given c. In Figures 1(d)-1(f), as the value of g^2 increases, the power of U_μ now dominates that of KSS_μ and closely tracks that of DF_μ for all values of c. U_μ therefore provides power near the effective envelope offered by KSS_μ and DF_μ across all values of the nonlinearity parameter g^2.

Figure 2 shows the power performance of U_τ 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_τ dominates that of DF_τ and closely tracks that of KSS_τ. In Figure 2(c), as the power curves of the two individual tests intersect, we see that the power of U_τ 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_τ now dominates that of KSS_τ for all values of c, whilst closely tracking the power curve of DF_τ.

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_μ or U_τ. 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_μ and KSS_μ, 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_τ should be employed, Harvey et al. (2009) show that, in the context of a linear AR DGP, DF_μ has a substantial power advantage over DF_τ 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_μ and U_τ when a local trend is present in the DGP.

Figures 3-5 show the asymptotic power of U_μ and U_τ for six different settings of the local trend coefficient, $\kappa = \{0, 0.25, 0.5, 1, 2, 4\}$ 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_μ and U_τ when $\kappa = 0$ such that no trend is present in the DGP. We see that U_μ significantly outperforms U_τ across all values of c. For example, when $c = 135$, U_μ has power of 0.83 whilst U_τ has power of 0.50. This highlights the potential losses involved in always employing the trend-invariant U_τ when there is uncertainty about the presence or otherwise of a linear trend. In Figure 3(b), $\kappa = 0.25$; the presence of this linear trend has decreased the power of U_μ such that, for $c = 135$, power has dropped to 0.56, only marginally higher than that of U_τ. When the local trend coefficient increases to $\kappa = 0.5$ in Figure 3(c), the power of U_τ now dominates that of U_μ. Figures 3(d)-(f) show that as the trend magnitude continues to increase, the power of U_μ deteriorates dramatically, such that for $\kappa = 1$ in Figure 3(d), U_μ has power lower than size for most c, and for $\kappa = 2$ and $\kappa = 4$ in Figures 3(e)-(f), the size and power of U_μ 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_μ and U_τ by taking a union of rejections of all four individual tests DF_μ, KSS_μ, DF_τ and KSS_τ. This proposed testing strategy can then be written as

$$U_4 : \text{Reject } H_0 \text{ if } DF_\mu < \lambda_4^\alpha cv_{DF_\mu}^\alpha \text{ or } KSS_\mu < \lambda_4^\alpha cv_{KSS_\mu}^\alpha \text{ or } DF_\tau < \lambda_4^\alpha cv_{DF_\tau}^\alpha \text{ or } KSS_\tau < \lambda_4^\alpha cv_{KSS_\tau}^\alpha$$

or

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

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

$$t_{U_4} \xrightarrow{d} \min\left(L_{DF_\mu}, \frac{cv_{DF_\mu}^\alpha}{cv_{KSS_\mu}^\alpha} L_{KSS_\mu}, \frac{cv_{DF_\mu}^\alpha}{cv_{DF_\tau}^\alpha} L_{DF_\tau}, \frac{cv_{DF_\mu}^\alpha}{cv_{KSS_\tau}^\alpha} L_{KSS_\tau} \right).$$

Values of the scaling constant λ_4^α 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 κ used previously. Figure 3(a) displays power in the case where $g^2 = 2$ and no trend is present in the DGP ($\kappa = 0$). The power of U_4 dominates that of U_τ across all values of c. For example, when $c = 135$, the power of U_4 is 0.72, whilst the power of U_τ is 0.50. U_4 tracks the power of U_μ, with power losses of, at most, 0.13 when $c = 117$. In Figure 3(b), $\kappa = 0.25$, and the power curve of U_4 sits slightly above U_τ and slightly below U_μ for all values of c. As the trend coefficient increases in

Figures 3(c)-3(f), U_4 outperforms U_μ for all values of c, whilst tracking underneath the power curve of U_τ. For example, when $\kappa = 4$, in Figure 3(f), U_4 has a maximum power loss of 0.21 relative to U_τ, but a power gain of 0.49 relative to U_μ. 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 κ, U_4 has superior power performance to the worst-performing test of either U_μ and U_τ 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, λ_α also increases to ensure the overall procedure is asymptotically correctly sized. As λ_α 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_μ decreases as the trend magnitude increases, it follows that the DF_μ and KSS_μ 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_μ from our testing strategy and use U_τ 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_τ is appropriate, with little to be gained by also including DF_μ and KSS_μ 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_μ, since a small magnitude trend could still be present in the data (which could have a substantial impact on the power of U_μ), 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_τ and KSS_τ tests as well as the DF_μ and KSS_μ 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

$$\begin{aligned} y_t &= \alpha + \beta t + u_t \\ \Delta u_t &= \varepsilon_t. \end{aligned}$$

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

$$\begin{aligned} \widehat{\varepsilon}_{1t} &= \Delta y_t \\ \widehat{\varepsilon}_{2t} &= \Delta y_t - \frac{\sum_{s=2}^{T} \Delta y_s}{T-1} = \Delta y_t - \frac{y_T - y_1}{T-1}. \end{aligned}$$

The BIC for each model is then

$$
\begin{aligned}
BIC_1 &= \ln\left(\frac{\sum_{t=1}^{T-1}\widehat{\varepsilon}_{1t}^2}{T-1}\right) \\
BIC_2 &= \ln\left(\frac{\sum_{t=1}^{T-1}\widehat{\varepsilon}_{2t}^2}{T-1}\right) + \frac{\ln(T-1)}{T-1}.
\end{aligned}
$$

If $BIC_2 < BIC_1$ then there is evidence that a trend is present in y_t, and therefore U_τ 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_τ 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_μ, KSS_μ, DF_τ and KSS_τ as well as the three union of rejections procedures U_μ, U_τ 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_μ in the presence of large trends (due to under-sizing in DF_μ and KSS_μ). 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_μ, KSS_μ and U_μ. We see that for modest values of g^2 where nonlinearity is most apparent, KSS_μ has a power advantage over DF_μ; see, for example, Figure 6(a). However, for larger values of g^2, DF_μ is shown to have superior power to KSS_μ as the degree of nonlinearity reduces, for example in Figures 6(d)-(f). As in the asymptotic case, the power of U_μ 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_τ, KSS_τ and U_τ are considered. Again, we see that KSS_τ has higher power relative to DF_τ for lower settings of g^2, but that the relative power ranking is reversed for higher settings of g^2. We show that U_τ has power either close to or above the power offered by the best-performing of either DF_τ or KSS_τ for all values of c and g^2.

Figures 8-10 report finite sample power results of U_μ, U_τ 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_μ, U_τ and U_4 is very similar to that observed in the asymptotic simulations. We see that the power of U_μ decreases as β increases, such that in Figure 8(a), where $g^2 = 2$ and $\beta = 0$, U_μ has a substantial power advantage over U_τ, but for higher values of β the power of U_μ tends to zero. U_4 is able to capitalize on the differing power profiles of U_μ and U_τ and captures much of the power gains associated with U_μ for small trends, and U_τ 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_τ 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_τ 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_τ, so that the power of U^* equals that of U_τ, 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_τ.

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_μ, DF_τ, KSS_μ and KSS_τ to annual total energy consumption per capita data obtained for 178 countries across the time period 1980-2006.[3] 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_μ and DF_τ tests and for 71 countries using KSS_μ and KSS_τ 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,

[3] Hasanov 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_μ, DF_τ, KSS_μ and KSS_τ, as well as the four union of rejections procedures considered in this paper, U_μ, U_τ, 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_μ and KSS_τ 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_τ 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_τ (40) is substantially higher than that found by the next most frequently rejecting test, KSS_μ (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_μ. No other individual test rejects, but U_μ 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_τ is the only individual test to reject, but U_τ, 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_μ results in a 0.01 level rejection by U_μ 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_τ 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_μ. 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_μ is able to reject the null for a higher proportion of countries than the individual test DF_μ at all conventional significance levels, and rejects a higher proportion of countries than KSS_μ at a 0.05 significance level. Similarly, U_τ

rejects the null more frequently than either DF_τ or KSS_τ 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_μ at all conventional levels of significance and U_τ at a 0.10 level. Finally, our BIC-based strategy, U^*, rejects the same number of times as U_4 and U_τ 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_τ 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

Akdogan, K. (2014), 'Threshold adjustment in the current account: sustainability for danger zone economies?', *Applied Economics Letters* **21**, 1006–1009.

Baharumshah, A. Z., Liew, V. K. S. and Haw, C. T. (2009), 'The real interest rate differential: international evidence based on non-linear unit root tests', *Bulletin of Economic Research* **61**, 83–94.

Beechey, M. and Osterholm, P. (2008), 'Revisiting the uncertain unit root in GDP and CPI: testing for non-linear trend reversion', *Economics Letters* **100**, 221–223.

Chen, S. W. (2014), 'Smooth transition, non-linearity and current account sustainability: evidence from the European countries', *Economic Modelling* **38**, 541–554.

Christidou, M., Panagiotidis, T. and Sharma, A. (2013), 'On the stationarity of per capita carbon dioxide emissions over a century', *Economic Modelling* **33**, 918–925.

Cook, S. (2008), 'More uncertainty: on the trending nature of real GDP in the US and UK', *Applied Economics Letters* **15**, 667–670.

Demetrescu, M. and Kruse, R. (2012), 'The power of unit root tests against nonlinear local alternatives', *Journal of Time Series Analysis* **34**, 40–61.

Guney, P. O. and Hasanov, M. (2014), 'Real interest rate parity hypothesis in post-Soviet countries: evidence from unit root tests', *Economic Modelling* **36**, 120–129.

Hanck, C. (2012), 'On the asymptotic distribution of a unit root test against ESTAR alternatives', *Statistics and Probability Letters* **82**, 360–364.

Harvey, D. I., Leybourne, S. J. and Taylor, A. M. R. (2009), 'Unit root testing in practice: dealing with uncertainty over the trend and initial condition', *Econometric Theory* **25**, 587–636.

Harvey, D. I., Leybourne, S. J. and Taylor, A. M. R. (2012), 'Testing for unit roots in the presence of uncertainty over both the trend and initial condition', *Journal of Econometrics* **169**, 188–195.

Hasanov, M. and Telatar, E. (2011), 'A re-examination of stationarity of energy consumption: evidence from new unit root tests', *Energy Policy* **39**, 7726–7738.

Kapetanios, G., Shin, Y. and Snell, A. (2003), 'Testing for a unit root in the nonlinear STAR framework', *Journal of Econometrics* **112**, 359–379.

Kisswani, K. M. and Nusair, S. A. (2013), 'Non-linearities in the dynamics of oil prices', *Energy Economics* **36**, 341–353.

Marsh, P. (2007), 'The available information for invariant tests of a unit root', *Econometric Theory* **23**, 686–710.

Michael, P., Nobay, R. A. and Peel, D. A. (1997), 'Transactions costs and nonlinear adjustment in real exchange rates: an empirical investigation', *Journal of Political Economy* **105**, 862–879.

Ng, S. and Perron, P. (1995), 'Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag', *Journal of the American Statistical Association* **90**, 268–281.

Ng, S. and Perron, P. (2001), 'Lag length selection and the construction of unit root tests with good size and power', *Econometrica* **69**, 1519–1554.

Norman, S. (2009), 'Testing for a unit root against ESTAR nonlinearity with a delay parameter greater than one', *Economics Bulletin* **29**, 2152–2173.

Sercu, P., Uppal, R. and Van Hulle, C. (1995), 'The exchange rate in the presence of transaction costs: implications for tests of purchasing power parity', *Journal of Finance* **50**, 1309–1319.

Smyth, R. (2013), 'Are fluctuations in energy variables permanent or transitory? A survey of the literature on the integration properties of energy consumption and production', *Applied Energy* **104**, 371–378.

Sollis, R. (2004), 'Asymmetric adjustment and smooth transitions: a combination of some unit root tests', *Journal of Time Series Analysis* **25**, 409–417.

Su, J. J., Cheung, A. W. K. and Roca, E. (2013), 'Lag selection of the augmented Kapetanios-Shin-Snell nonlinear unit root test', *Journal of Mathematics and Statistics* **9**, 102–111.

Taylor, M. P., Peel, D. A. and Sarno, L. (2001), 'Nonlinear mean-reversion in real exchange rates: towards a solution to the purchasing power parity puzzles', *International Economic Review* **42**, 1015–1042.

Tsong, C. C. and Lee, C. F. (2013), 'Further evidence on real interest rate equalization: panel information, non-linearities and structural changes', *Bulletin of Economic Research* **65**, 85–105.

van Dijk, D., Terasvirta, T. and Franses, P. H. (2002), 'Smooth transition autoregressive models – a survey of recent developments', *Econometric Reviews* **21**, 1–47.

Appendix

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

Proof of Lemma 1

The DF_μ test statistic is given by

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

where $\widehat{\sigma}^2$ denotes the error variance estimate from the OLS regression of Δy_t on $y_{\mu,t-1}$. Considering first the numerator of DF_μ, we can write

$$\begin{aligned} T^{-1} \sum_{t=2}^{T} y_{\mu,t-1} \Delta y_t &= T^{-1} \sum_{t=2}^{T} (y_{t-1} - \overline{y}) \left(\beta - \frac{c}{T} u_{t-1} G(\theta, u_{t-1}) + \varepsilon_t \right) \\ &= -cT^{-2} \sum_{t=2}^{T} (y_{t-1} - \overline{y}) u_{t-1} G(\theta, u_{t-1}) + T^{-1} \sum_{t=2}^{T} (y_{t-1} - \overline{y}) \varepsilon_t + o_p(1). \end{aligned}$$

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

$$\begin{aligned} & cT^{-2} \sum_{t=2}^{T} (y_{t-1} - \overline{y}) u_{t-1} G(\theta, u_{t-1}) \\ &= cT^{-2} \sum_{t=2}^{T} \{ \beta (t - 1 - \bar{t}) + u_{t-1} - \bar{u} \} u_{t-1} G(\theta, u_{t-1}) \\ &= cT^{-1} \sum_{t=2}^{T} \left\{ \kappa \sigma T^{-1} (t - \bar{t}) + T^{-1/2} (u_{t-1} - \bar{u}) \right\} T^{-1/2} u_{t-1} G(\theta, u_{t-1}) + o_p(1) \\ &\xrightarrow{d} \sigma^2 c \int_0^1 \left(\kappa \left(r - \tfrac{1}{2} \right) + X_\mu(r) \right) X(r) \mathcal{G}\left(g^2 X(r)^2 \right) dr \end{aligned}$$

and

$$\begin{aligned} T^{-1} \sum_{t=2}^{T} (y_{t-1} - \overline{y}) \varepsilon_t &= T^{-1/2} \sum_{t=2}^{T} \left\{ \kappa \sigma T^{-1} (t - \bar{t}) + T^{-1/2} (u_{t-1} - \bar{u}) \right\} \varepsilon_t + o_p(1) \\ &\xrightarrow{d} \sigma^2 \int_0^1 \left(\kappa \left(r - \tfrac{1}{2} \right) + X_\mu(r) \right) dW(r). \end{aligned}$$

In the denominator of DF_μ, it is easily shown that $\widehat{\sigma}^2 \xrightarrow{p} \sigma^2$, and

$$\begin{aligned} T^{-2} \sum_{t=2}^{T} (y_{t-1} - \overline{y})^2 &= T^{-1} \sum_{t=2}^{T} \left(\kappa \sigma T^{-1} (t - \bar{t}) + T^{-1/2} (u_{t-1} - \bar{u}) \right)^2 \\ &\xrightarrow{d} \sigma^2 \int_0^1 \left(\kappa \left(r - \tfrac{1}{2} \right) + X_\mu(r) \right)^2 dr. \end{aligned}$$

The DF_μ test statistic therefore has the limiting distribution

$$DF_\mu \xrightarrow{d} \frac{\int_0^1 \left(\kappa\left(r-\frac{1}{2}\right)+X_\mu(r)\right)dW(r) - c\int_0^1 \left(\kappa\left(r-\frac{1}{2}\right)+X_\mu(r)\right)X(r)\mathcal{G}\left(g^2X(r)^2\right)dr}{\sqrt{\int_0^1 \left(\kappa\left(r-\frac{1}{2}\right)+X_\mu(r)\right)^2 dr}}.$$

The KSS_μ test statistic is given by

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

where $\widehat{\sigma}^2$ denotes the error variance estimate from the OLS regression of Δy_t on $y_{\mu,t-1}^3$. In the numerator,

$$\begin{aligned}
T^{-2}\sum_{t=2}^T y_{\mu,t-1}^3 \Delta y_t &= T^{-2}\sum_{t=2}^T \left(y_{t-1}-\overline{y}\right)^3 \left(\beta - \frac{c}{T}u_{t-1}G\left(\theta,u_{t-1}\right)+\varepsilon_t\right) \\
&= \kappa\sigma T^{-5/2}\sum_{t=2}^T \left(y_{t-1}-\overline{y}\right)^3 - cT^{-3}\sum_{t=2}^T \left(y_{t-1}-\overline{y}\right)^3 u_{t-1}G\left(\theta,u_{t-1}\right) \\
&\quad +T^{-2}\sum_{t=2}^T \left(y_{t-1}-\overline{y}\right)^3 \varepsilon_t \\
&\xrightarrow{d} \kappa\sigma^4 \int_0^1 \left(\kappa\left(r-\tfrac{1}{2}\right)+X_\mu(r)\right)^3 dr \\
&\quad -c\sigma^4 \int_0^1 \left(\kappa\left(r-\tfrac{1}{2}\right)+X_\mu(r)\right)^3 X(r)\mathcal{G}\left(g^2X(r)^2\right)dr \\
&\quad +\sigma^4 \int_0^1 \left(\kappa\left(r-\tfrac{1}{2}\right)+X_\mu(r)\right)^3 dW(r)
\end{aligned}$$

while in the denominator, $\widehat{\sigma}^2 \xrightarrow{p} \sigma^2$ and

$$T^{-4}\sum_{t=2}^T y_{\mu,t-1}^6 \xrightarrow{d} \sigma^6 \int_0^1 \left(\kappa\left(r-\tfrac{1}{2}\right)+X_\mu(r)\right)^6 dr$$

giving the limit for KSS_μ as

$$KSS_\mu \xrightarrow{d} \frac{\begin{array}{c}\kappa\int_0^1 \left(\kappa\left(r-\frac{1}{2}\right)+X_\mu(r)\right)^3 dr + \int_0^1 \left(\kappa\left(r-\frac{1}{2}\right)+X_\mu(r)\right)^3 dW(r) \\ -c\int_0^1 \left(\kappa\left(r-\frac{1}{2}\right)+X_\mu(r)\right)^3 X(r)\mathcal{G}\left(g^2X(r)^2\right)dr\end{array}}{\sqrt{\int_0^1 \left(\kappa\left(r-\frac{1}{2}\right)+X_\mu(r)\right)^6 dr}}.$$

The test statistics based on demeaned and detrended data are invariant to β, hence we set $\beta = 0$ without loss of generality in the remainder of this proof. The DF_τ test statistic is

$$DF_\tau = \frac{\sum_{t=2}^T y_{\tau,t-1}\Delta y_{\tau,t}}{\sqrt{\widehat{\sigma}^2 \sum_{t=2}^T y_{\tau,t-1}^2}}$$

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

$$\begin{aligned}
\Delta y_{\tau,t} &= \Delta y_t - \widehat{\beta} \\
&= -\frac{c}{T} u_{t-1} G\left(\theta, u_{t-1}\right) + \varepsilon_t - \widehat{\beta}
\end{aligned}$$

and

$$\begin{aligned}
T^{1/2}\widehat{\beta} &= \frac{T^{-5/2}\sum_{t=1}^{T}\left(t-\bar{t}\right) y_t}{T^{-3}\sum_{t=1}^{T}\left(t-\bar{t}\right)^2} \\
&\xrightarrow{d} 12\int_0^1 \left(r-\tfrac{1}{2}\right) X(r)dr.
\end{aligned}$$

Also,

$$\begin{aligned}
T^{-1/2} y_{\tau,\lfloor rT\rfloor} &= T^{-1/2} y_{\lfloor rT\rfloor} - T^{-1/2}\widehat{\mu} - T^{1/2}\widehat{\beta} r \\
&= T^{-1/2} y_{\lfloor rT\rfloor} - \left(T^{-1/2}\overline{y} - T^{1/2}\widehat{\beta} T^{-1}\overline{t}\right) - T^{1/2}\widehat{\beta} r \\
&\xrightarrow{d} X(r) - \int_0^1 X(s)ds - 12\left(r-\tfrac{1}{2}\right)\int_0^1 \left(s-\tfrac{1}{2}\right) X(s)ds \\
&= X(r) + (6r-4)\int_0^1 X(s)ds + (6-12r)\int_0^1 sX(s)ds \\
&\equiv X_\tau(r).
\end{aligned}$$

The numerator of DF_τ is then given by

$$\begin{aligned}
T^{-1}\sum_{t=2}^{T} y_{\tau,t-1}\Delta y_{\tau,t} &= T^{-1}\sum_{t=2}^{T} y_{\tau,t-1}\varepsilon_t - cT^{-2}\sum_{t=2}^{T} y_{\tau,t-1}u_{t-1}G\left(\theta, u_{t-1}\right) - T^{1/2}\widehat{\beta} T^{-3/2}\sum_{t=2}^{T} y_{\tau,t-1} \\
&\xrightarrow{d} \sigma^2\int_0^1 X_\tau(r)dW(r) - c\sigma^2\int_0^1 X_\tau(r)X(r)\mathcal{G}\left(g^2X(r)^2\right)dr
\end{aligned}$$

and for the denominator we obtain

$$T^{-2}\sum_{t=2}^{T} y_{\tau,t-1}^2 \xrightarrow{d} \sigma^2\int_0^1 X_\tau(r)^2dr$$

giving

$$DF_\tau \xrightarrow{d} \frac{\int_0^1 X_\tau(r)dW(r) - c\int_0^1 X_\tau(r)X(r)\mathcal{G}\left(g^2X(r)^2\right)dr}{\sqrt{\int_0^1 X_\tau(r)^2dr}}.$$

Finally, the KSS_τ test statistic is

$$KSS_\tau = \frac{\sum_{t=2}^{T} y_{\tau,t-1}^3\Delta y_{\tau,t}}{\sqrt{\widehat{\sigma}^2\sum_{t=2}^{T} y_{\tau,t-1}^6}}$$

where $\widehat{\sigma}^2$ is the error variance estimate from the OLS regression of $\Delta y_{\tau,t}$ on $y^3_{\tau,t-1}$. Now $\widehat{\sigma}^2 \xrightarrow{p} \sigma^2$ as before,

$$\begin{aligned}
T^{-2}\sum_{t=2}^{T} y^3_{\tau,t-1}\Delta y_{\tau,t} &= T^{-2}\sum_{t=2}^{T} y^3_{\tau,t-1}\varepsilon_t - cT^{-3}\sum_{t=2}^{T} y^3_{\tau,t-1}u_{t-1}G(\theta, u_{t-1}) - T^{1/2}\widehat{\beta}T^{-5/2}\sum_{t=2}^{T} y^3_{\tau,t-1} \\
&\xrightarrow{d} \sigma^2 \int_0^1 X_\tau(r)^3 dW(r) - c\sigma^2 \int_0^1 X_\tau(r)^3 X(r)\mathcal{G}\left(g^2 X(r)^2\right) dr \\
&\quad -12\int_0^1 \left(r-\tfrac{1}{2}\right) X(r)dr \int_0^1 X_\tau(r)^3 dr
\end{aligned}$$

and

$$T^{-4}\sum_{t=2}^{T} y^6_{\tau,t-1} \xrightarrow{d} \int_0^1 X_\tau(r)^6 dr$$

so

$$KSS_\tau \xrightarrow{d} \frac{\int_0^1 X_\tau(r)^3 dW(r) - c\int_0^1 X_\tau(r)^3 X(r)\mathcal{G}\left(g^2X(r)^2\right)dr - 12\int_0^1 \left(r-\frac{1}{2}\right)X(r)dr\int_0^1 X_\tau(r)^3 dr}{\sqrt{\int_0^1 X_\tau(r)^6 dr}}.$$

Proof of Lemma 2

As before, the DF_μ test statistic is given by

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

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

$$\begin{aligned}
T^{-3/2}\sum_{t=2}^{T} y_{\mu,t-1}\Delta y_t &= T^{-3/2}\sum_{t=2}^{T}\left(\sigma\kappa(t-\bar{t}) + u_{t-1} - \bar{u}\right)\left(\sigma\kappa + \varepsilon_t - \frac{c}{T}u_{t-1}G(\theta, u_{t-1})\right) + o_p(1) \\
&= \sigma\kappa T^{-3/2}\sum_{t=2}^{T}(t-\bar{t})\varepsilon_t - \sigma\kappa c T^{-5/2}\sum_{t=2}^{T}(t-\bar{t})u_{t-1}G(\theta, u_{t-1}) + o_p(1) \\
&\xrightarrow{d} \sigma^2\kappa \int_0^1 \left(r-\tfrac{1}{2}\right) dW(r) - \sigma^2\kappa c\int_0^1 \left(r-\tfrac{1}{2}\right) X(r)\mathcal{G}\left(g^2X(r)^2\right) dr
\end{aligned}$$

and

$$\begin{aligned}
T^{-3}\sum_{t=2}^{T} y^2_{\mu,t-1} &= T^{-3}\sum_{t=2}^{T}\left(\beta(t-1-\bar{t}) + u_{t-1} - \bar{u}\right)^2 \\
&= \sigma^2\kappa^2 T^{-3}\sum_{t=2}^{T}(t-\bar{t})^2 + o_p(1) \\
&\xrightarrow{p} \sigma^2\kappa^2/12.
\end{aligned}$$

Also,

$$\begin{aligned}
\widehat{\sigma}^2 &= T^{-1}\sum_{t=2}^{T}\left(\Delta y_t - \frac{\sum_{s=2}^{T}(y_{s-1}-\overline{y})\Delta y_s}{\sum_{s=2}^{T}(y_{s-1}-\overline{y})^2}(y_{t-1}-\overline{y})\right)^2 \\
&= T^{-1}\sum_{t=2}^{T}\left(\Delta y_t - \frac{T^{-3/2}\sum_{s=2}^{T}(y_{s-1}-\overline{y})\Delta y_s}{T^{-3}\sum_{s=2}^{T}(y_{s-1}-\overline{y})^2}T^{-3/2}(y_{t-1}-\overline{y})\right)^2 \\
&= T^{-1}\sum_{t=2}^{T}(\Delta y_t)^2 + o_p(1) \\
&= \sigma^2\kappa^2 + T^{-1}\sum_{t=2}^{T}\varepsilon_t^2 + o_p(1) \\
&\xrightarrow{p} \sigma^2\left(\kappa^2+1\right).
\end{aligned}$$

Therefore the DF_μ test statistic has the limiting distribution

$$DF_\mu \xrightarrow{d} \frac{\int_0^1\left(r-\frac{1}{2}\right)dW(s) - c\int_0^1\left(r-\frac{1}{2}\right)X(r)\mathcal{G}\left(g^2X(r)^2\right)dr}{\sqrt{\left(\kappa^2+1\right)/12}}.$$

For KSS_μ we again have

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

where $\widehat{\sigma}^2$ is the error variance estimate from the OLS regression of Δy_t on $y_{\mu,t-1}^3$. We obtain

$$\begin{aligned}
T^{-7/2}\sum_{t=2}^{T} y_{\mu,t-1}^3\Delta y_t &= T^{-7/2}\sum_{t=2}^{T}\left(\sigma\kappa(t-\overline{t})+u_{t-1}-\overline{u}\right)^3\left(\sigma\kappa - \frac{c}{T}u_{t-1}G\left(\theta,u_{t-1}\right)+\varepsilon_t\right) + o_p(1) \\
&= 3\sigma^3\kappa^3T^{-7/2}\sum_{t=2}^{T}(t-\overline{t})^2(u_{t-1}-\bar{u}) + \sigma^3\kappa^3T^{-7/2}\sum_{t=2}^{T}(t-\overline{t})^3\varepsilon_t \\
&\quad -c\sigma^3\kappa^3T^{-9/2}\sum_{t=2}^{T}(t-\overline{t})^3u_{t-1}G\left(\theta,u_{t-1}\right) + o_p(1) \\
&\xrightarrow{d} 3\sigma^4\kappa^3\int_0^1\left(r-\tfrac{1}{2}\right)^2X_\mu(r)dr + \sigma^4\kappa^3\int_0^1\left(r-\tfrac{1}{2}\right)^3dW(r) \\
&\quad -c\sigma^4\kappa^3\int_0^1\left(r-\tfrac{1}{2}\right)^3X(r)\mathcal{G}\left(g^2X(r)^2\right)dr
\end{aligned}$$

and

$$\begin{aligned}
T^{-7}\sum_{t=2}^{T} y_{\mu,t-1}^6 &= \sigma^6\kappa^6T^{-7}\sum_{t=2}^{T}(t-\overline{t})^6 + o_p(1) \\
&\xrightarrow{p} \sigma^6\kappa^6/448
\end{aligned}$$

together with

$$
\begin{aligned}
\widehat{\sigma}^2 &= T^{-1} \sum_{t=2}^{T} \left(\Delta y_t - \frac{T^{-7/2} \sum_{s=2}^{T} (y_{s-1} - \overline{y})^3 \Delta y_s}{T^{-7} \sum_{s=2}^{T} (y_{s-1} - \overline{y})^6} T^{-7/2} (y_{t-1} - \overline{y})^3 \right)^2 \\
&= T^{-1} \sum_{t=2}^{T} (\Delta y_t)^2 + o_p(1) \\
&\xrightarrow{p} \sigma^2 \left(\kappa^2 + 1 \right)
\end{aligned}
$$

giving the KSS_μ limit distribution

$$
KSS_\mu \xrightarrow{d} \frac{3 \int_0^1 \left(r - \frac{1}{2}\right)^2 X_\mu(r) dr + \int_0^1 \left(r - \frac{1}{2}\right)^3 dW(r) - c \int_0^1 \left(r - \frac{1}{2}\right)^3 X(r) \mathcal{G} \left(g^2 X(r)^2\right) dr}{\sqrt{(\kappa^2 + 1) / 448}}.
$$

Table 1: Asymptotic critical values of KSS_{μ}, DF_{μ}, KSS_{τ}, DF_{τ} and union of rejection scaling values λ_{μ}^{α}, λ_{τ}^{α} and λ_{4}^{α} for significance level α

α	KSS_{μ}	DF_{μ}	KSS_{τ}	DF_{τ}	λ_{μ}^{α}	λ_{τ}^{α}	λ_{4}^{α}
0.10	−2.655	−2.564	−3.118	−3.122	1.063	1.056	1.140
0.05	−2.935	−2.864	−3.396	−3.409	1.059	1.049	1.118
0.01	−3.471	−3.424	−3.939	−3.959	1.053	1.041	1.085

Table 2: Total primary energy consumption per capita: DF_μ, DF_τ, KSS_μ, KSS_τ, U_μ, U_τ, U_4 and U^* unit root test results

	DF_μ	DF_τ	KSS_μ	KSS_τ	U_μ	U_τ	U_4	U^*
Algeria			***	***	***	***	***	***
American Samoa		**	***	***	**	***	***	***
Antigua and Barbuda		***				***	***	***
Argentina				*		*		
Bahamas	*							
Bahrain				*				
Bangladesh		***		**		***	**	***
Bermuda	*		*		*			
Bhutan			***		***		***	
Burkina Faso		*						
Cambodia				**		**	**	**
Cayman Islands			*					
Central African Republic	**	***		*	**	***	***	***
Chad			*		*			
Comoros		*						
Cook Islands			***		***		**	**
Djibouti			*	*				
Dominica		*	*	**		**	*	**
Egypt		**	*	*	*	**	*	**
El Salvador				**		*		*
Faroe Islands	*		***	*	***	*	***	***
Finland			*		*			
French Polynesia	***	**	***	**	***	**	***	***
Greenland			***	***	***	***	***	***
Grenada				**		**	*	*
Guadeloupe		***	***		***	***	***	***
Guinea			**	***	*	***	**	**
Honduras		*		***		***	**	**
Hong Kong		**				**	*	**
India				**		*		*
Jordan			**	*	**		*	*
Kenya	**				**		*	*
Kuwait		**				**	**	**
Lebanon				*				
Lesotho				**		**	*	*
Libya		***	***	***	***	***	***	***
Maldives				**		**	*	**
Mali	***	*	*		**		**	**
Mozambique			*	**	*	**	**	**
Nauru			**	*	**	*	**	**
Netherlands Antilles	***	**	**	***	***	***	***	***
New Caledonia	**	*	**		**	*	*	*
Niger				*				

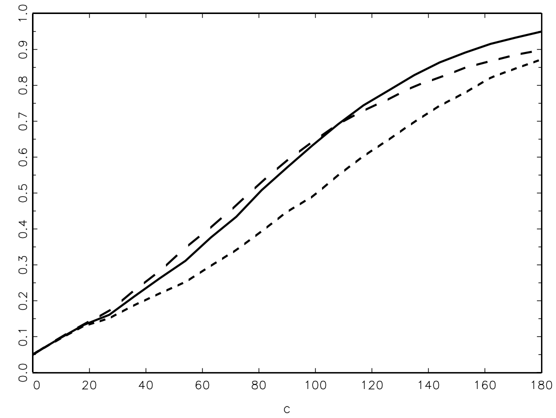
	DF_μ	DF_τ	KSS_μ	KSS_τ	U_μ	U_τ	U_4	U^*
Panama		***				***	***	***
Portugal				***		***	***	***
Puerto Rico			**	**	*	*	*	*
Saint Lucia		**		***		***	***	***
Samoa		***		***		***	***	***
Sao Tome and Principe				*				
Senegal			**	**	*	**	*	*
Seychelles				**		**	*	*
Sierra Leone				***		***	***	***
Solomon Islands		***		***		***	***	***
Suriname	**		*		**		**	**
Swaziland	***	**	***	*	**	*	**	**
Tonga			***	***	***	**	**	**
Trinidad and Tobago				**		*		*
Uganda		***				***	***	***
United Arab Emirates	***	**	***	***	***	***	***	***
Uruguay		*						
Total Rejections at 0.10	12	25	29	40	28	41	43	45
Total Rejections at 0.05	9	18	19	28	20	32	31	34
Total Rejections at 0.01	5	9	12	14	11	19	19	19

*, **, *** indicate rejections at a 0.10, 0.05, 0.01 significance level respectively

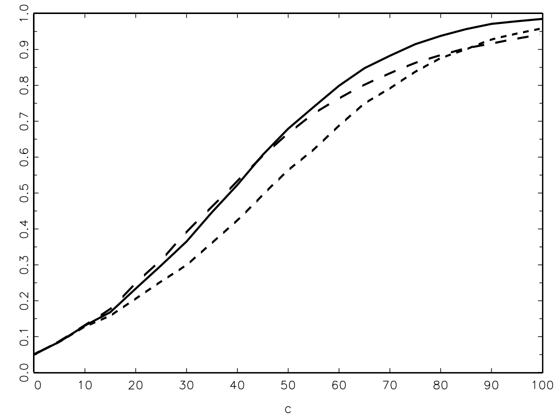
Table 3: Total primary energy consumption per capita: countries for which no unit root test rejections found

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

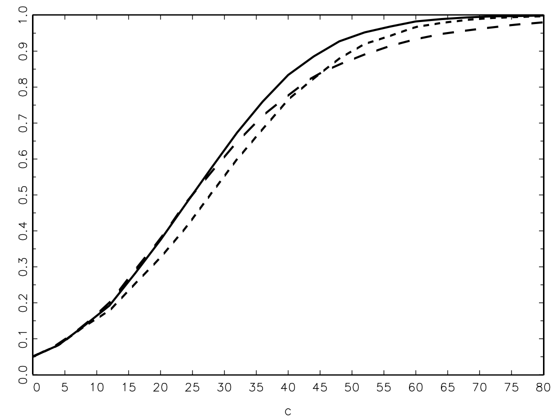
Figure 1: Asymptotic size and local power of KSS_μ, DF_μ and U_μ for fixed g^2 ($\kappa = 0$)



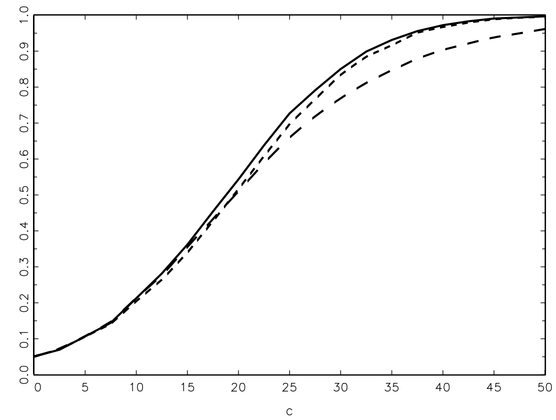
(a) $g^2 = 2$



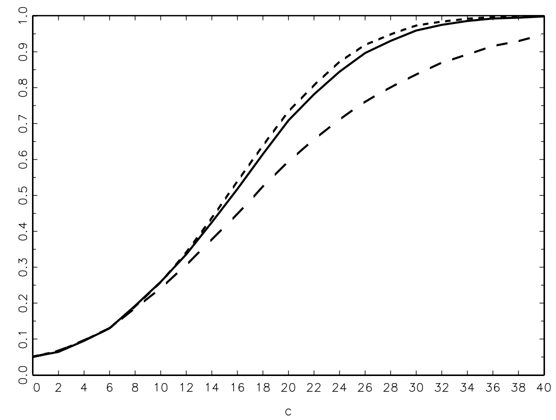
(b) $g^2 = 5$



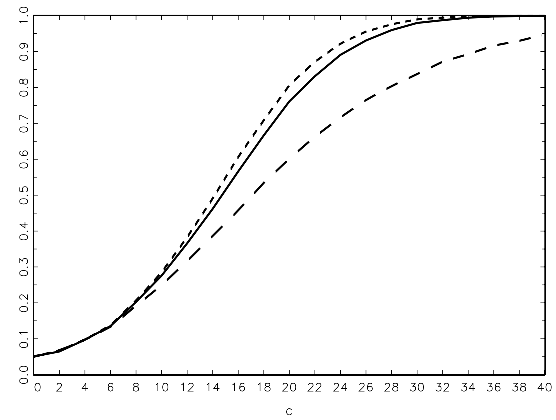
(c) $g^2 = 10$



(d) $g^2 = 20$



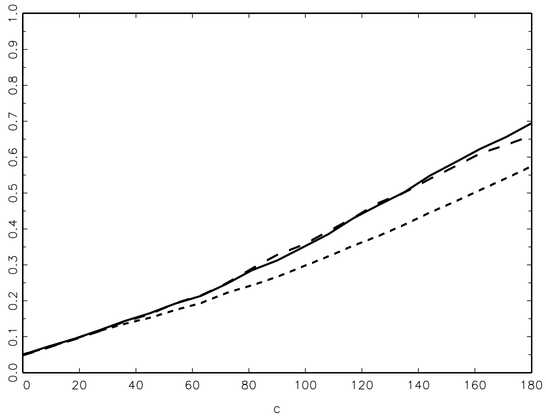
(e) $g^2 = 50$



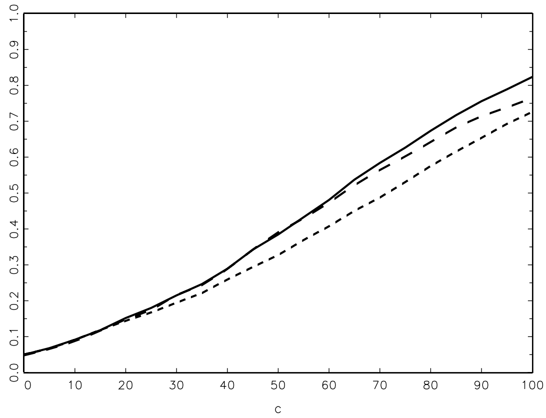
(f) $g^2 = 100$

$- -\ KSS_\mu$, - - - DF_μ, — U_μ

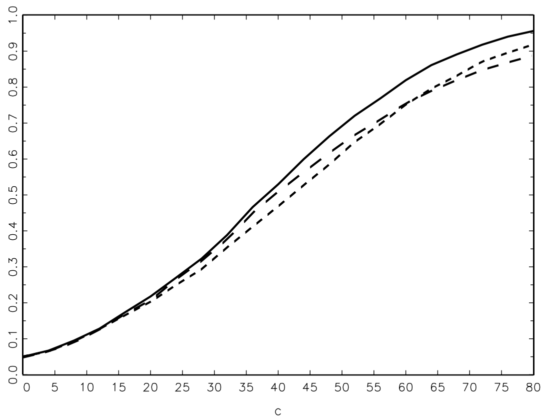
Figure 2: Asymptotic size and local power of KSS_τ, DF_τ and U_τ for fixed g^2



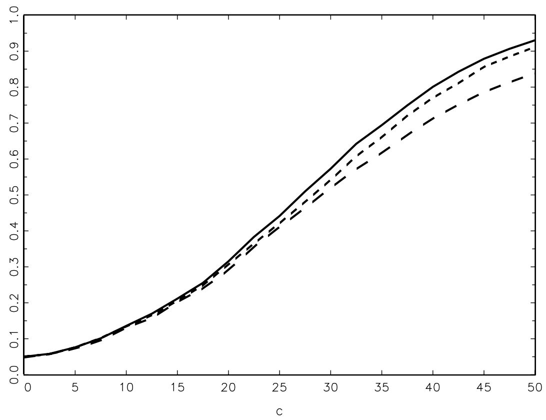
(a) $g^2 = 2$



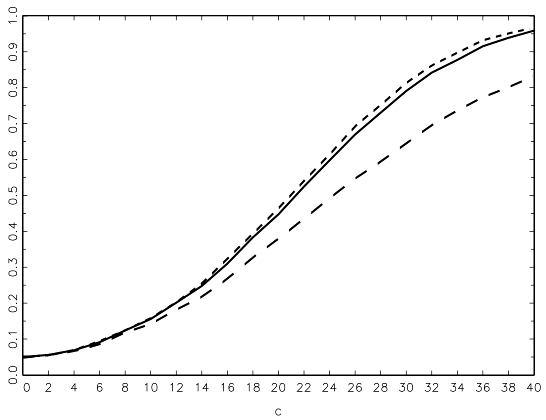
(b) $g^2 = 5$



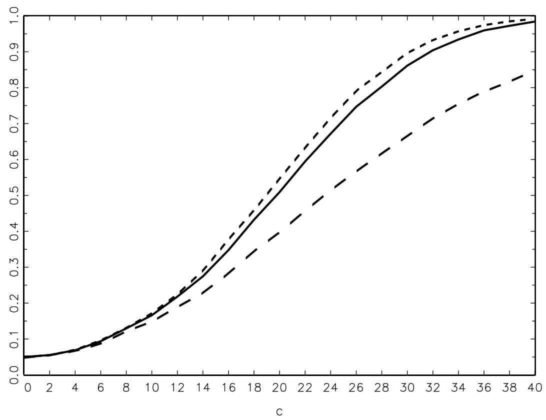
(c) $g^2 = 10$



(d) $g^2 = 20$



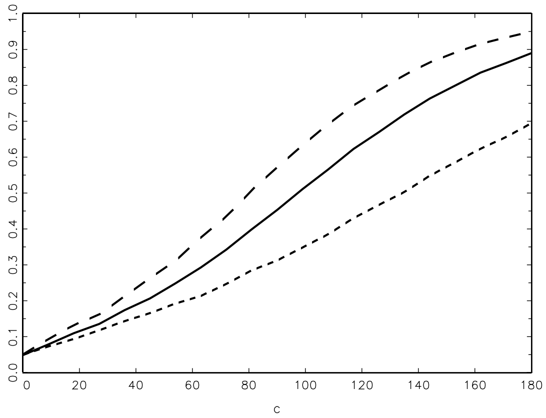
(e) $g^2 = 50$



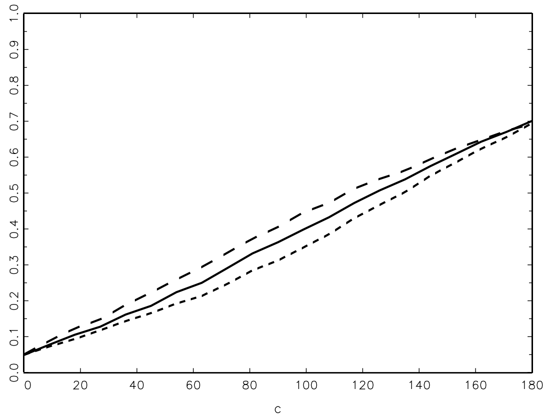
(f) $g^2 = 100$

$- -\ KSS_\tau$, - - - DF_τ, — U_τ

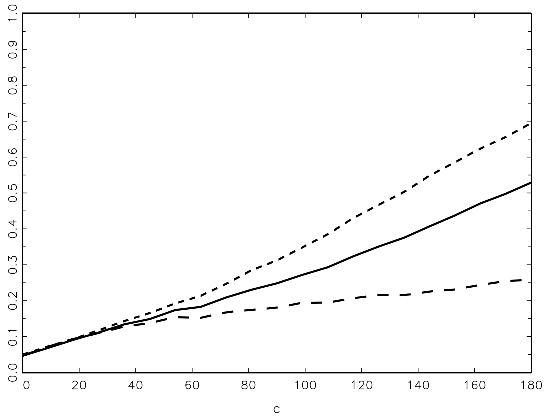
Figure 3: Asymptotic size and local power of U_μ, U_τ and U_4 for fixed $g^2 = 2$



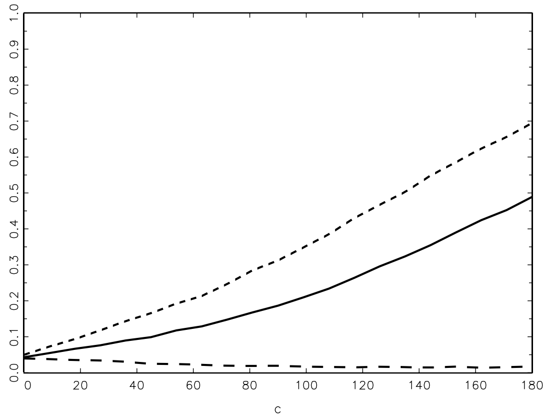
(a) $\kappa = 0$



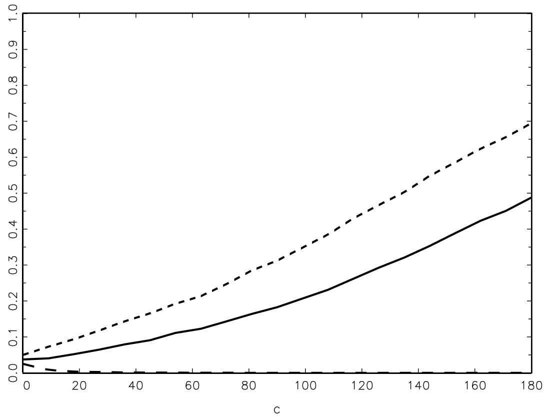
(b) $\kappa = 0.25$



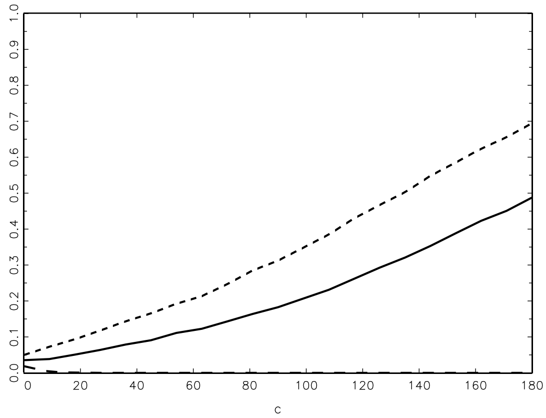
(c) $\kappa = 0.5$



(d) $\kappa = 1.0$



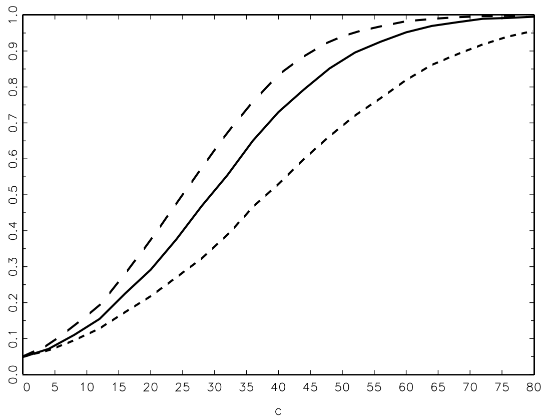
(e) $\kappa = 2.0$



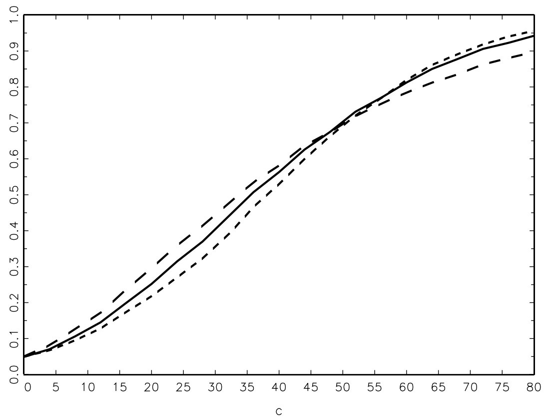
(f) $\kappa = 4.0$

$- -$ U_μ, - - - U_τ, — U_4

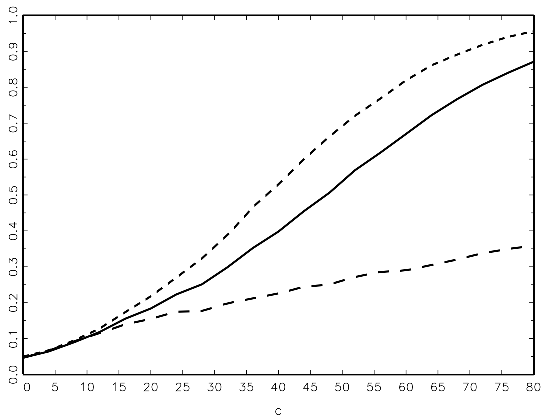
Figure 4: Asymptotic size and local power of U_μ, U_τ and U_4 for fixed $g^2 = 10$



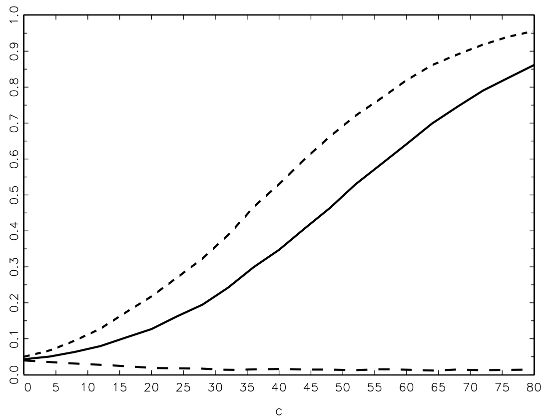
(a) $\kappa = 0$



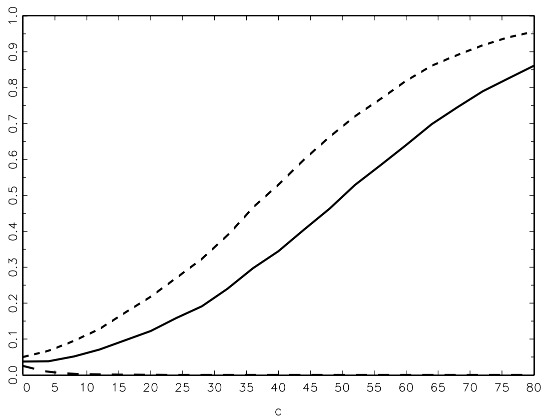
(b) $\kappa = 0.25$



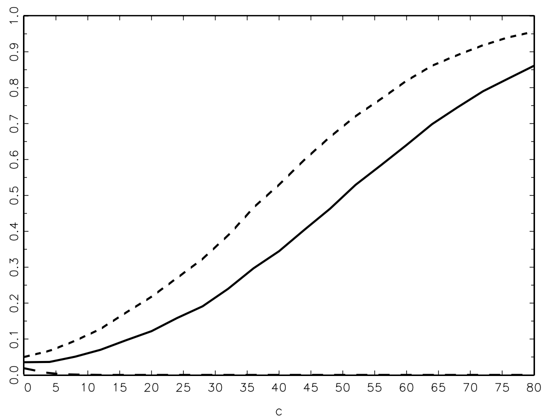
(c) $\kappa = 0.5$



(d) $\kappa = 1.0$



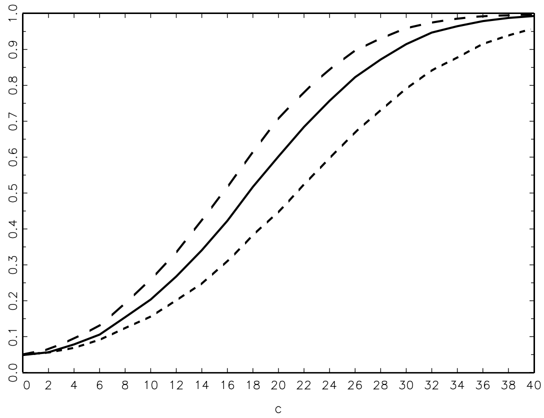
(e) $\kappa = 2.0$



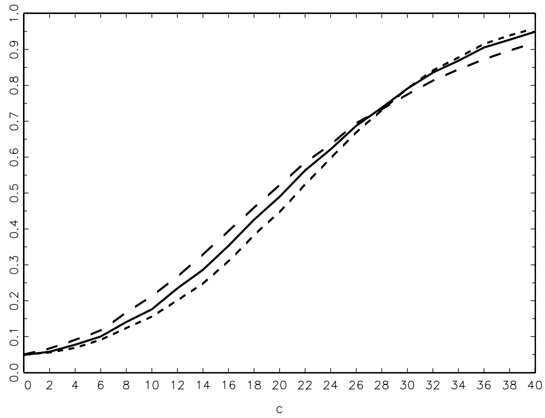
(f) $\kappa = 4.0$

-- U_μ, - - - U_τ, — U_4

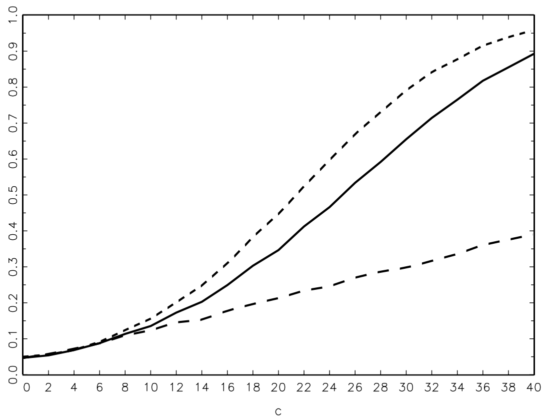
Figure 5: Asymptotic size and local power of U_μ, U_τ and U_4 for fixed $g^2 = 50$



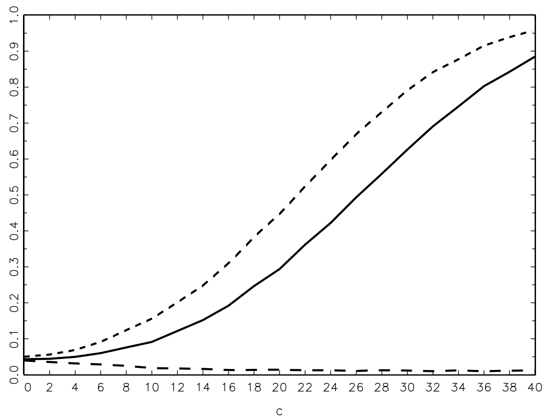
(a) $\kappa = 0$



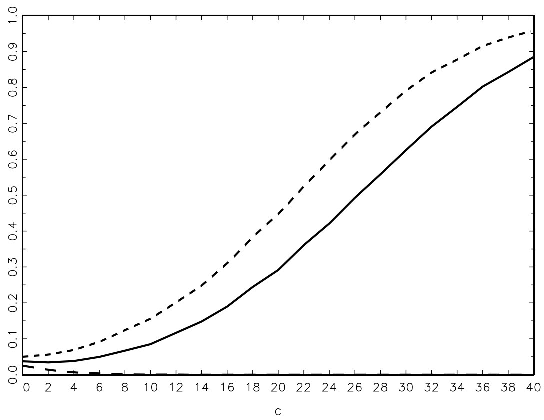
(b) $\kappa = 0.25$



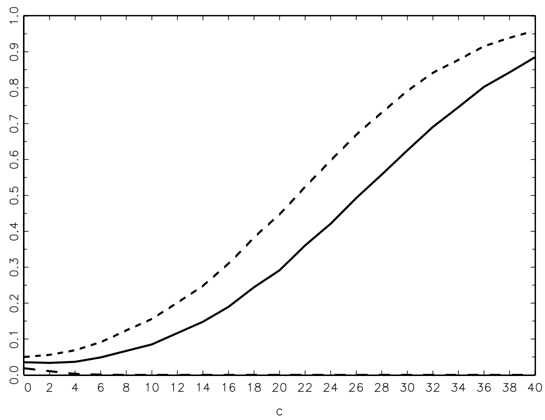
(c) $\kappa = 0.5$



(d) $\kappa = 1.0$



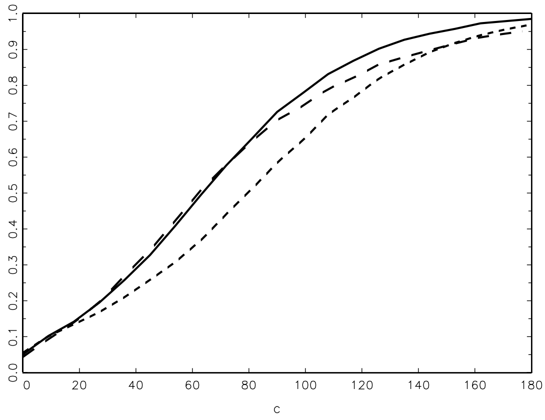
(e) $\kappa = 2.0$



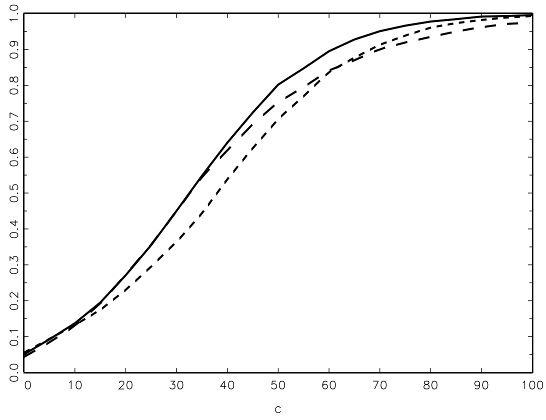
(f) $\kappa = 4.0$

$- -\ U_\mu$, $- - -\ U_\tau$, — U_4

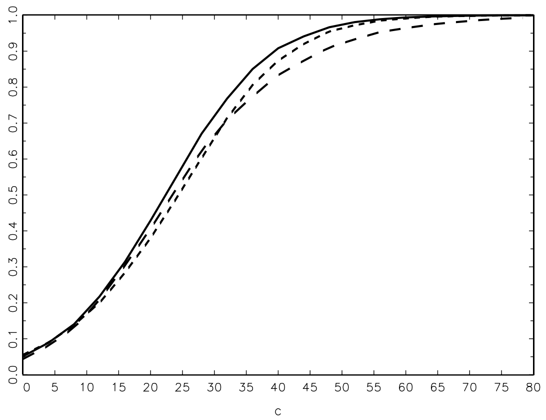
Figure 6: Finite sample size and local power of KSS_μ, DF_μ and U_μ for fixed g^2 ($\beta = 0$)



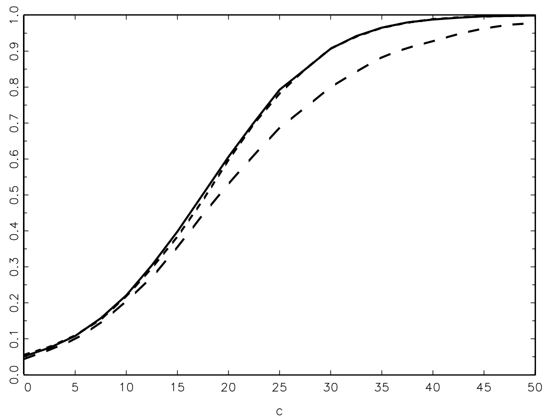
(a) $g^2 = 2$



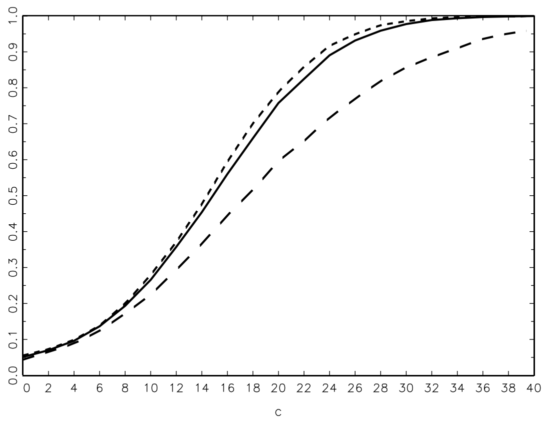
(b) $g^2 = 5$



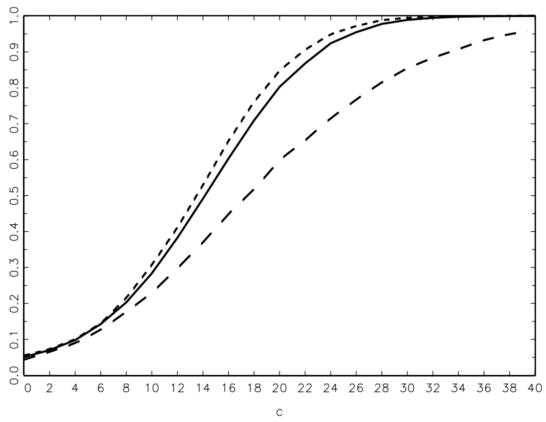
(c) $g^2 = 10$



(d) $g^2 = 20$



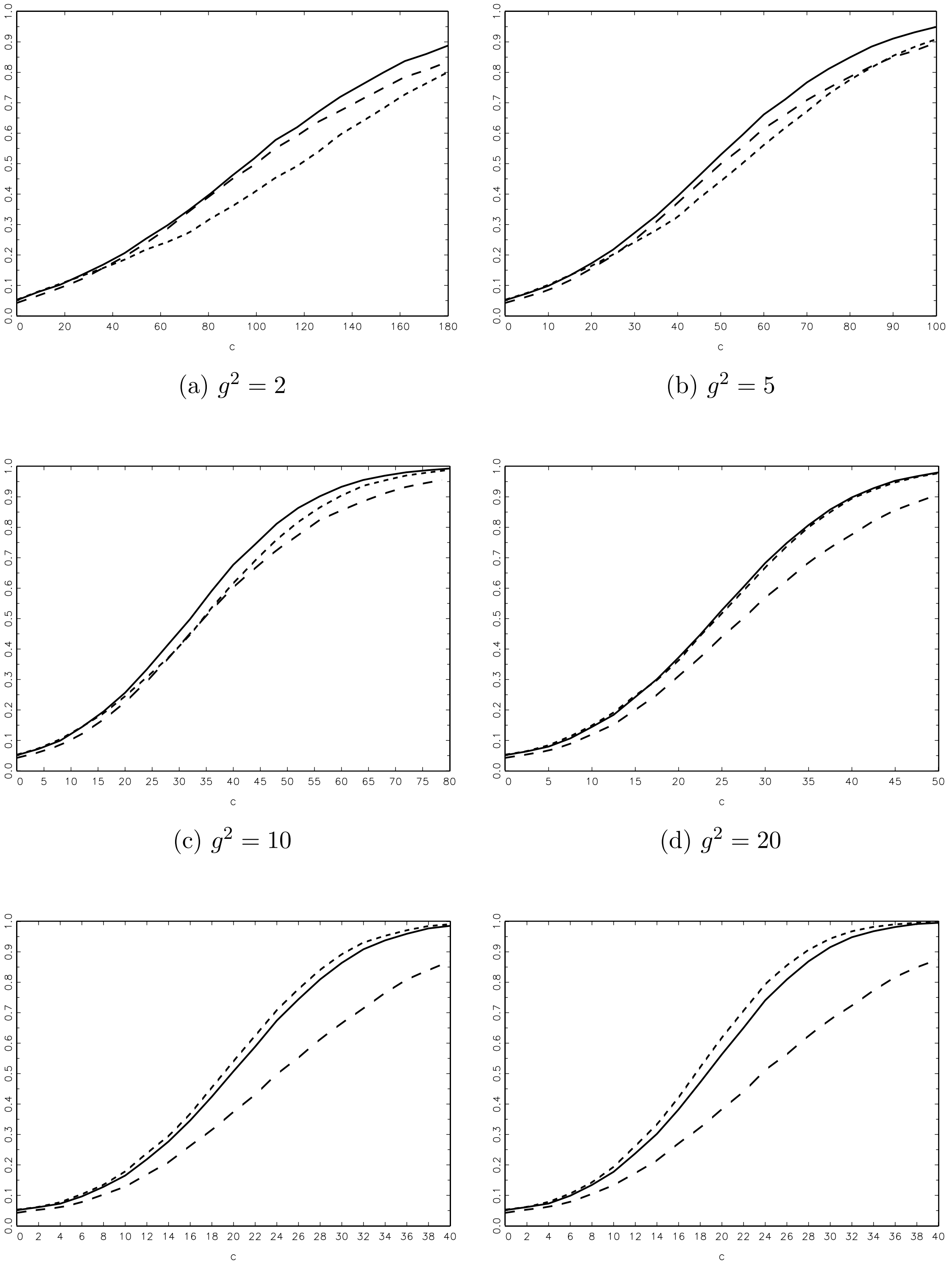
(e) $g^2 = 50$



(f) $g^2 = 100$

$- -$ KSS_μ, - - - DF_μ, — U_μ

Figure 7: Finite sample size and local power of KSS_τ, DF_τ and U_τ for fixed g^2



(a) $g^2 = 2$

(b) $g^2 = 5$

(c) $g^2 = 10$

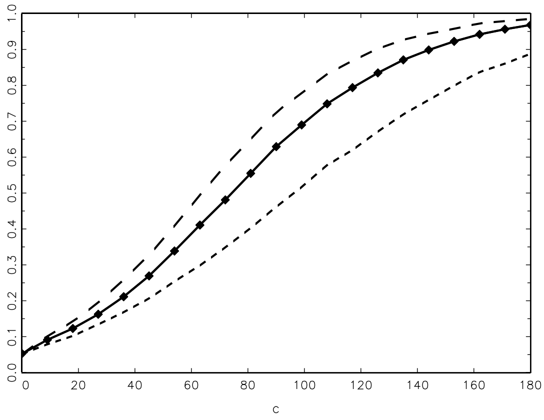
(d) $g^2 = 20$

(e) $g^2 = 50$

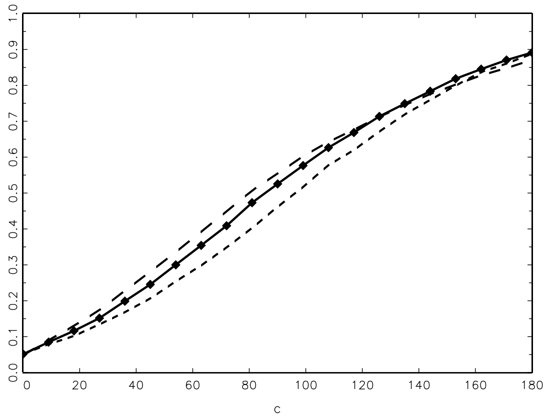
(f) $g^2 = 100$

$- - KSS_\tau$, $\mathbf{- - -}\ DF_\tau$, $— U_\tau$

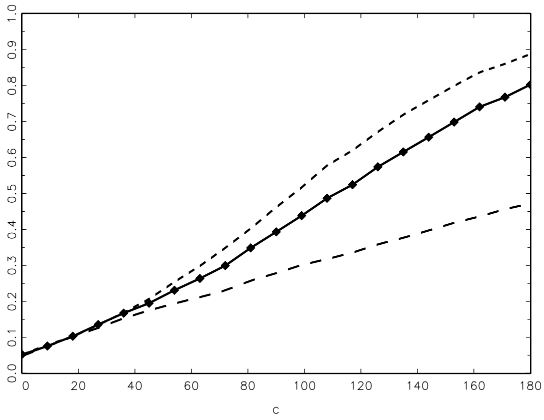
Figure 8: Finite sample size and local power of U_μ, U_τ, U_4 and U^* for fixed $g^2 = 2$



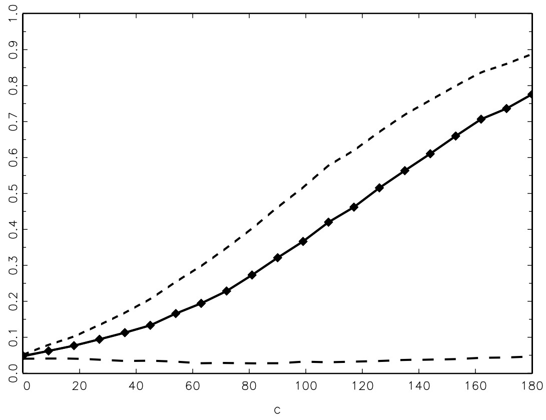
(a) $\beta = 0$



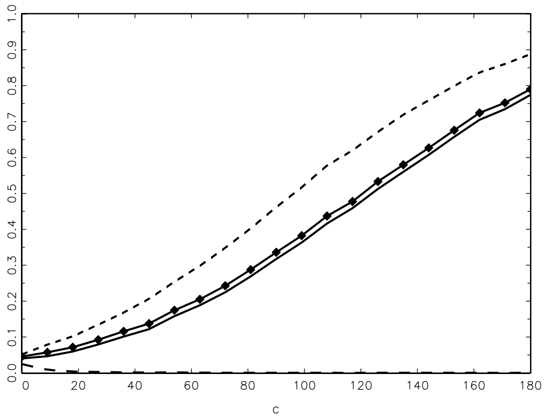
(b) $\beta = 0.025$



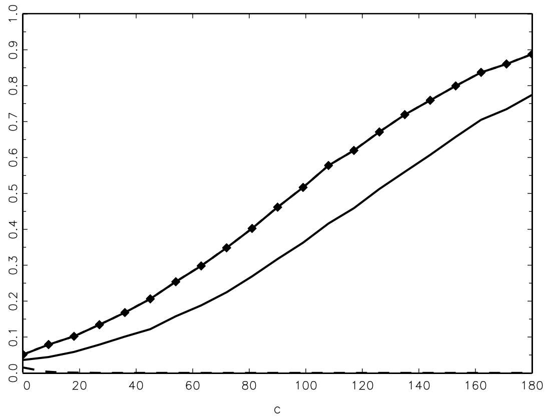
(c) $\beta = 0.05$



(d) $\beta = 0.1$



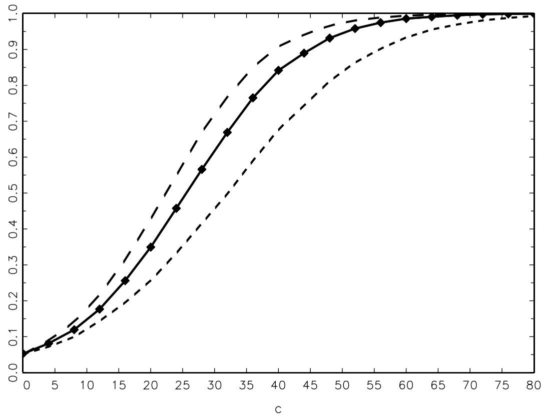
(e) $\beta = 0.2$



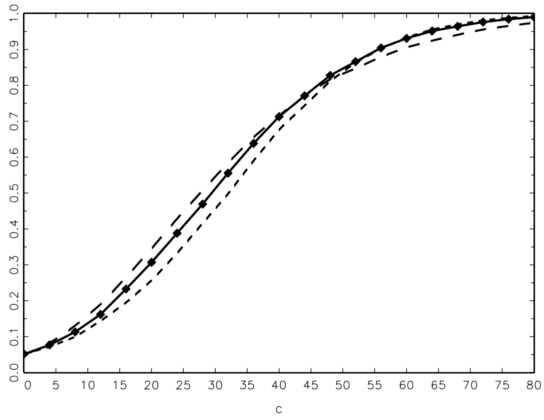
(f) $\beta = 0.4$

$- - U_\mu$, $\mathbf{- - -}\ U_\tau$, — U_4, —◆— U^*

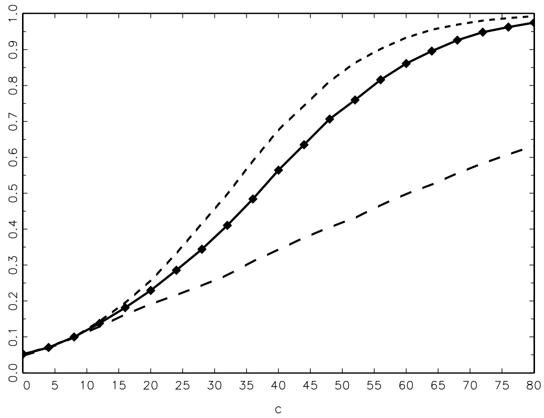
Figure 9: Finite sample size and local power of U_μ, U_τ, U_4 and U^* for fixed $g^2 = 10$



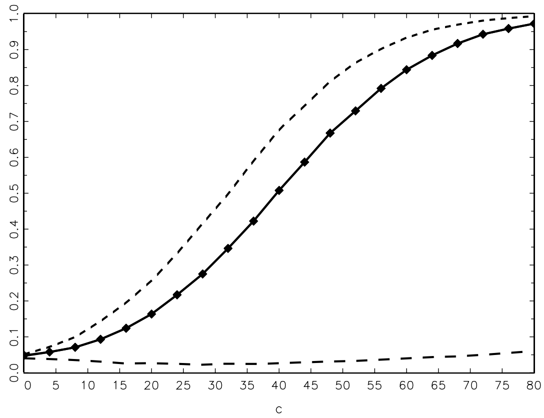
(a) $\beta = 0$



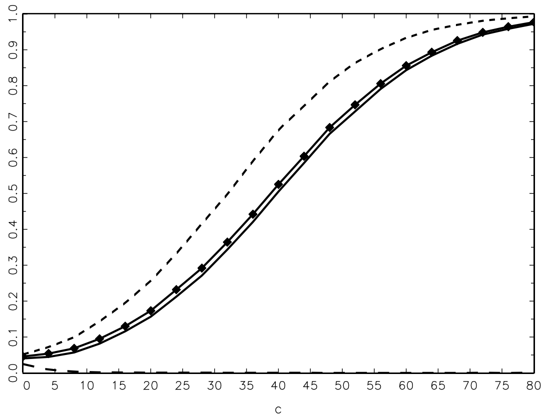
(b) $\beta = 0.025$



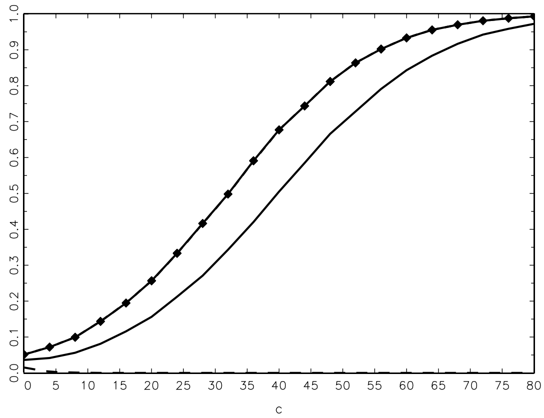
(c) $\beta = 0.05$



(d) $\beta = 0.1$



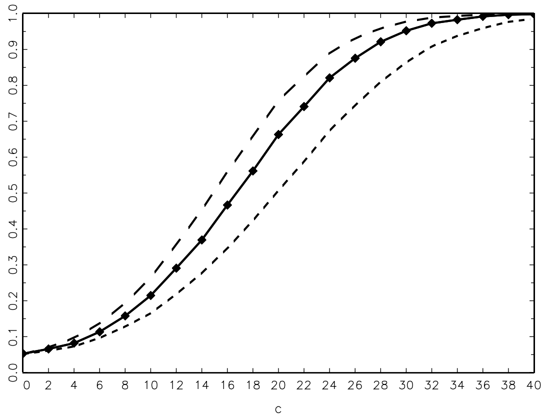
(e) $\beta = 0.2$



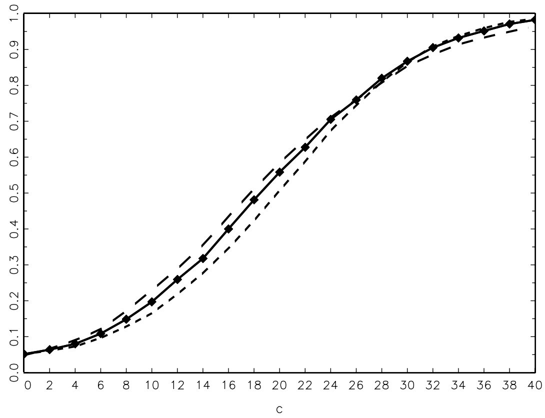
(f) $\beta = 0.4$

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

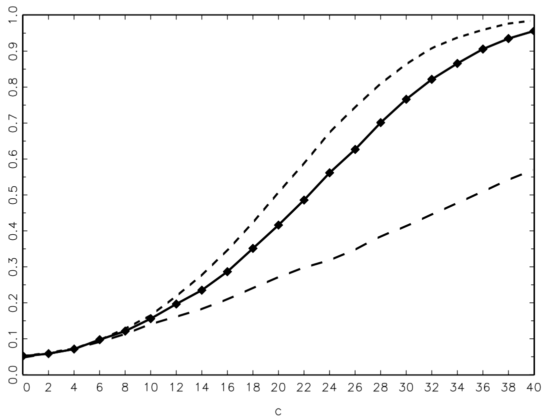
Figure 10: Finite sample size and local power of U_μ, U_τ, U_4 and U^* for fixed $g^2 = 50$



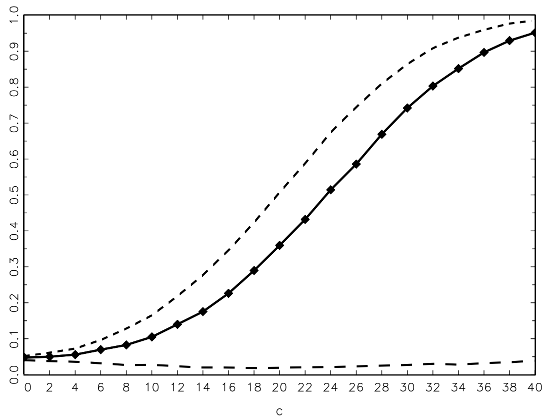
(a) $\beta = 0$



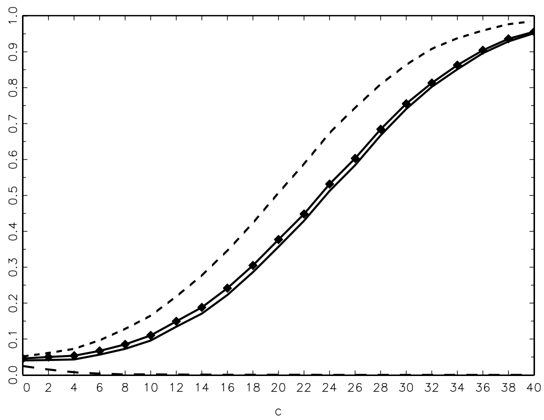
(b) $\beta = 0.025$



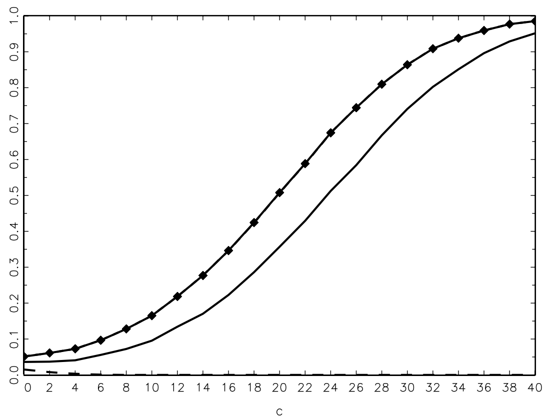
(c) $\beta = 0.05$



(d) $\beta = 0.1$



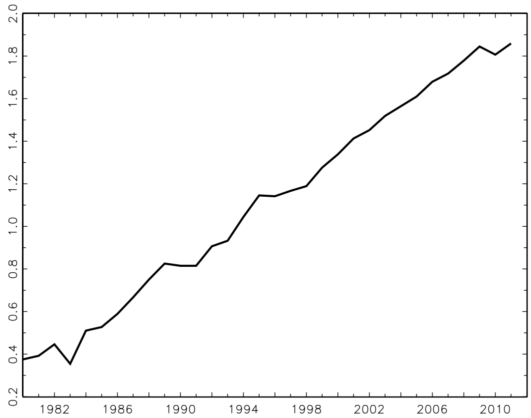
(e) $\beta = 0.2$



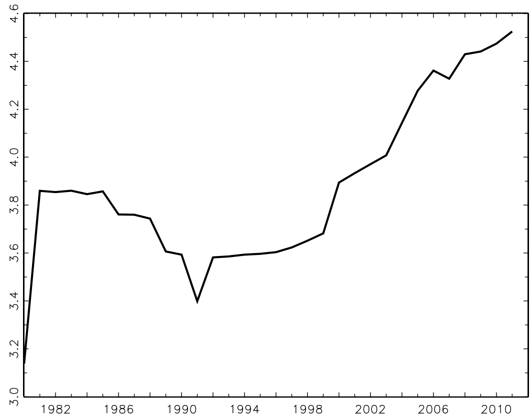
(f) $\beta = 0.4$

$- - U_\mu$, $\mathbf{- - -}\ U_\tau$, — U_4, —◆— U^*

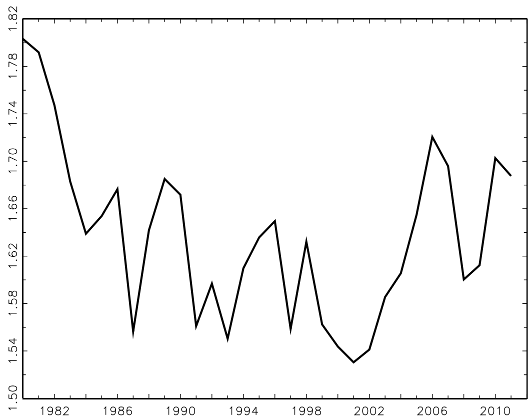
Figure 11: Total primary energy consumption per capita (logs of million BTU per person)



(a) Bangladesh

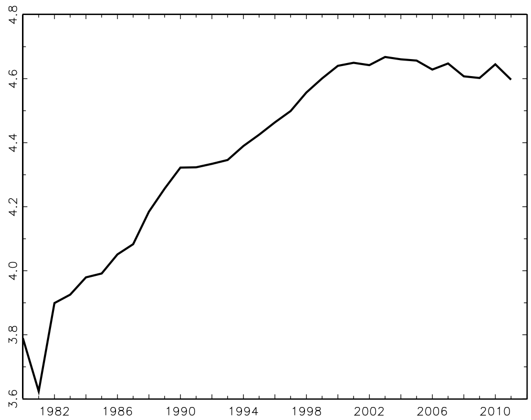


(b) Cook Islands

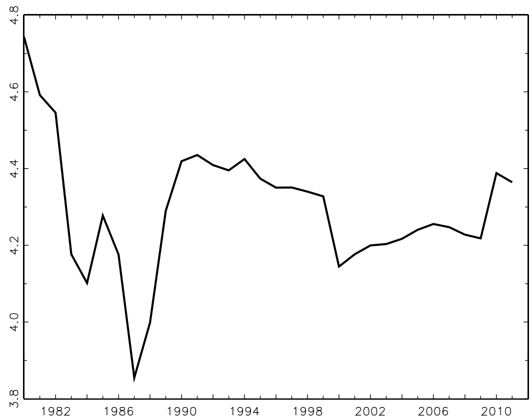


(c) Kenya

(d) Panama



(e) Portugal



(f) Suriname