The Power of Unit Root Tests Against Nonlinear Local Alternatives

Matei Demetrescu and Robinson Kruse

CREATES Research Paper 2012-01

The Power of Unit Root Tests Against Nonlinear Local Alternatives[*]

Matei Demetrescu[a] and Robinson Kruse[b]

[a] University of Bonn[†]   [b] CREATES, Aarhus University and Leibniz University Hannover[‡]

January 4, 2012

Abstract

This article extends the analysis of local power of unit root tests in a nonlinear direction by considering local nonlinear alternatives and tests built specifically against stationary nonlinear models. In particular, we focus on the popular test proposed by Kapetanios et al. (2003, Journal of Econometrics 112, 359-379) in comparison to the linear Dickey-Fuller test. To this end, we consider different adjustment schemes for deterministic terms. We provide asymptotic results which imply that the error variance has a severe impact on the behavior of the tests in the nonlinear case; the reason for such behavior is the interplay of nonstationarity and nonlinearity. In particular, we show that nonlinearity of the data generating process can be asymptotically negligible when the error variance is moderate or large (compared to the "amount of nonlinearity"), rendering the linear test more powerful than the nonlinear one. Should however the error variance be small, the nonlinear test has better power against local alternatives. We illustrate this in an asymptotic framework of what we call persistent nonlinearity. The theoretical findings of this article explain previous results in the literature obtained by simulation. Furthermore, our own simulation results suggest that the user-specified adjustment scheme for deterministic components (e.g. OLS, GLS, or recursive adjustment) has a much higher impact on the power of unit root tests than accounting for nonlinearity, at least under local (linear or nonlinear) alternatives.

Key words: Nonlinear models, Stochastic trend, Near integration, Persistent nonlinearity, Local power

JEL classification: C12 (Hypothesis Testing), C22 (Time-Series Models)

1 Introduction

Nonlinear stationary time series models such as threshold autoregressive [TAR] models (Tong, 1990) or smooth transition autoregressive [STAR] models (Teräsvirta, 1994) have become quite popular in applied time series econometrics. Smooth transition models are particularly successful in this respect: they have been used with a variety of economic time series ranging from real (effective) exchange rates (e.g. Taylor et al., 2001; Sarantis, 1999) real money balances (Sarno et al., 2003), business cycle data (e.g. Teräsvirta and Anderson, 1992) and inflation (e.g. Nobay et al., 2010) to price-dividend ratios, see McMillan (2007). Such series often exhibit nonlinear regime-dependent dynamic behavior, depending on whether they are close to, or away from, some (fundamental) equilibrium value; in particular, their tendency to revert to equilibrium is weak when the series is in the middle regime (i.e. close to equilibrium), but may gain strength in the outer regimes; three-regime TAR and STAR models with a unit root regime in the middle are well-suited to capture this type of nonlinear behavior.

The empirical question of relevance for such series is whether the outer regimes do actually exhibit reversion to equilibrium. If not, the series under consideration, say y_t, is actually integrated; the inner and outer regimes exhibit the same unit root behavior. In contrast, if the series is mean-reverting in the outer regimes, the nonlinear models themselves are globally stationary in spite of the nonstationary of the middle regime, see Kapetanios et al. (2003) and Kapetanios and Shin (2006). The degree of mean-reversion depends (i) on the distance between y_{t-1} and its equilibrium value, and (ii) on the shape of the transition function. The linear Dickey and Fuller (1979) [DF] test is consistent against generic ergodic alternatives, so one might use the usual DF test in order to discriminate between unit root processes and nonlinear processes with mean-reverting outer regimes. Kapetanios et al. (2003) [KSS] and He and Sandberg (2006) suggest unit root tests against specific nonlinear alternatives and show that the power of unit root tests can be improved. Bec et al. (2004), Kapetanios and Shin (2006), Bec et al. (2008) develop tests against threshold autoregressive models and find similar power gains.

The first focus of this paper is on the local power of unit root tests since the power under sequences of local alternatives gives insight into the behavior of test procedures when the alternative, although true, is close to the null. The DF test has nontrivial power against (linear) alternatives in $1/T$ neighborhoods of the unit root (Phillips, 1987). Thus, the question arises: How do tests designed against nonlinear stationary models behave in the presence of such nearly integrated behavior? And, more importantly: how do tests against linear and nonlinear models behave when the data generating process [DGP] does indeed exhibit nonlinearities in addition to near integration?

The second focus is on the removal of deterministic components. In particular, the

power depends largely on the chosen adjustment procedure, especially when the DGP is close to the null of a unit root. Ordinary least squares [OLS] adjustment for instance is known to be inefficient in the linear case (Elliott et al., 1996); it also induces a large bias in the estimator of the autoregressive parameter. The application of efficient adjustment procedures is thus worth considering. Along these lines, Kapetanios and Shin (2008) adopt local-to-unity generalized LS detrending procedures for unit root tests in the nonlinear STAR and TAR frameworks. In the case of a linear DGP, Leybourne et al. (2005) show recursive adjustment of the series (Shin and So, 2001; Taylor, 2002) and forward-reverse application of DF tests (Leybourne, 1995) to be of effectiveness comparable with local-to-unity GLS adjustment. Therefore, we additionally compare the efficiency of such alternative adjustment methods when considering local nonlinear alternatives and possibly influential initial values.

In more detail, our contributions are as follows. We concentrate on the linear DF and nonlinear KSS unit root tests; after describing in Section 2 the setup of the paper, we analyze in Section 3 the local power of the DF and KSS tests against the benchmark of linear nearly integrated alternatives. We find the KSS test to have nontrivial power in the same $1/T$-neighborhoods as the DF test; the DF test is, however, locally more powerful. While this is not surprising given the linear nature of the considered alternatives, we find the same to hold true for nonlinear models with nearly integrated outer regimes as well. In other words, the nonlinearity of the DGP has negligible effect on both the DF and the KSS test when the outer regimes are close to the null. The reason for this counter-intuitive behavior of the tests is that near integration in the outer regimes leads to high unconditional variances of the process and hence to vanishing probability of "hitting" the middle regimes.

Thus, it might seem that unit root tests built against nonlinear alternatives have no advantage over linear unit root tests when the alternative is near integration. This is a fallacy, however, since the argument relies on fixed values of the parameters characterizing the nonlinear behavior. We argue in Section 4 that such fixed-nonlinearity asymptotics do not always describe the data well, as the error variance of nonlinear processes is often found to be small compared to the sample size, see e.g. Taylor et al. (2001). The direct effect of small innovations variance is that the (effective) width of the middle regime is quite large, in fact large enough for the probability of the process to hit the middle regime to be nonzero even in larger samples.[1] Hence, we discuss in Section 4 *persistently* nonlinear alternatives for which the variance of the process y_t is comparable with the width of the middle regime and a certain number of observations belong to this regime with positive probability in the limit. We accomplish this by letting the width of the

---

[1] Small variances also lead to problems with identifying the middle regime even when the model is stationary nonlinear; see Donauer et al. (2010).

middle regime increase with the sample size at an appropriate rate so the error variance is comparatively small; this type of asymptotics ensures that nonlinearity not negligible any longer. Under near integration with persistent nonlinearity, we often confirm the superiority of the nonlinear KSS test compared to the DF test when the same detrending procedure is applied. Most of the power gains, however, come from efficiently removing the deterministic component: e.g. the nonlinear KSS test with OLS detrending is clearly dominated by the DF test with GLS detrending for zero initial values. Since the relative behavior of the tests turns out to also depend on the magnitude of the initial value, we conjecture that combining test statistics following Harvey et al. (2009, 2011) would be the better strategy. Section 5 concludes by addressing some implications of our results for applied work.

Before proceeding to the analysis, let us set some notation. Denote by $W(s)$ the standard Wiener process and by "⇒" weak convergence in a space of càdlàg functions endowed with a suitable norm. Let C be a generic constant whose value may change from equation to equation. Finally, the symbols "$\overset{d}{\to}$" and "$\overset{p}{\to}$" stand for convergence in distribution and convergence in probability.

2 Setup

2.1 Model and test statistics

The observed series y_t is assumed to be generated according to the usual additive component model. This allows one to deal with deterministic components in the same manner under the null and under the alternative; see also Kapetanios and Shin (2008). We consider here the empirically more relevant cases of a constant and of a linear trend. Since the procedures used here to adjust for deterministic components lead to invariant (w.r.t. the trend parameters) test statistics, we may proceed as if the trend coefficients were zero.

Assumption 1 *Let*

$$\Delta y_t = \phi y_{t-1} G\left(y_{t-1}; \gamma\right) + \varepsilon_t, \quad t = 1, \ldots, T,$$

with $\gamma \in \mathbb{R}^+$; the starting value satisfies $y_0 = 0$ for $\phi = 0$ and $y_0 = \kappa/\sqrt{1-(1+\phi)^2}$ for $\phi < 0$, $\kappa \in \mathbb{R}$.

The function G takes values between 0 and 1; in particular $G(y, \cdot) \to 1$ as $y \to \pm\infty$, as is specified in the corresponding assumptions below. This allows for the well-known interpretation of a transition function between regimes. Also, $\lim_{\gamma \to 0} G(\cdot, \gamma) = 0$. Depending on the smoothness properties of G, one obtains either threshold autoregressions or STAR models. The initial condition is specified in the manner familiar from the linear

case. One reason to do so is the lack of a closed-form expression for the variance of nonlinear STAR processes in the general case; another reason is given in Section 3.2. The errors ε_t satisfy standard assumptions in the literature on nonlinear models.

Assumption 2 *Let ε_t be an iid sequence such that $\mathrm{E}(\varepsilon_t) = 0$, $\mathrm{Var}(\varepsilon_t) = \sigma^2$, and $\exists \delta > 0$ with $\mathrm{E}\,|\varepsilon_t|^{4+\delta} < C < \infty$.*

The derivations of this paper hold under more general data generating processes (in particular linear processes with stationary martingale difference [m.d.] innovations and 1-summable coefficients); but the aim of the paper is to give insight into the behavior of unit root tests when nonlinearity interacts with near integration and not to have state-of-the-art assumptions.

The DGP given by Assumptions 1 and 2 exhibits transitions between several regimes: the middle regime ($G = 0$) is characterized by a unit root, while the process is driven a stationary AR(1) process by in the two outer regimes ($G = 1$) as long as $\phi < 0$.[2] In this sense, the process is, roughly speaking, "partially nonstationary" as one of the regimes contains a unit root while the other regimes are stationary.

In the popular exponential STAR [ESTAR] case (where G is a smooth exponential function), Kapetanios et al. (2003) establish stationarity of y_t for fixed parameters ϕ and γ by using the drift condition of Tweedie (1975); de Jong (2009) analyzes the stationarity properties of such models under weaker assumptions about the innovations. Roughly speaking, if the outer regimes are stationary, so is the process y_t itself.

We are interested in modeling situations where the alternative, if true, is close to the null. This takes us away from the fixed-ϕ and fixed-γ setup. So it is required for ϕ throughout the paper that

Assumption 3 *Let $\phi = -\frac{c}{T}$ with $c \geq 0$ fixed.*

We shall write in fact ϕ_T to emphasize the local-to-unity character of the sequence of alternatives examined. It is natural to consider such neighborhoods of the null: the DF test has nontrivial power against this type of local alternatives, so we evaluate the nonlinear KSS test in a similar fashion.

Since the alternative is characterized by both ϕ and γ, there are two dimensions of the local-to-the-null interpretation of a nonlinear model. First, one can define sequences of local alternatives for which the autoregressive parameters in the outer regimes tend to one, thereby implying a unit root in all regimes which are identical in the limiting case; this is the classical approach captured in Assumption 3. Second, if the shape parameter γ of the transition function tends to zero, the model becomes linear in the limit and the

---

[2]The considered model is symmetric, therefore the upper and the lower regime are characterized by the same autoregressive parameter.

nonstationary middle regime dominates. Such reasoning applies to a variety of nonlinear models where one parameter is unidentified under the null.

So we shall analyze three mutually exclusive cases resulting from the type of alternative we consider for γ. In the case of a purely linear DGP (i.e. identical inner and outer regimes), we have

Assumption 4 *Let $G\left(y;\gamma\right) = 1 \ \forall y \in \mathbb{R}$ and $\forall \gamma \in \mathbb{R}^{+}$.*

Assumption 4 can be seen as the limiting case $\gamma \to \infty$. The second case we examine is the usual nonlinear one requiring the following

Assumption 5 *Let $\gamma > 0$ fixed. Let $G\left(y;\gamma\right) \mapsto \left[0;1\right]$ be a piecewise Lipschitz function such that $y^{2}\left(1 - G\left(y;\gamma\right)\right) \to 0$ as $y \to \pm\infty \ \forall \gamma > 0$.*

The condition $y^{2}\left(1 - G\left(y;\gamma\right)\right) \to 0$ implies that $G\left(y;\gamma\right) \to 1$ as $y \to \pm\infty$ at a rate higher than quadratic; it is purely technical and fulfilled by virtually all popular transition functions in TAR or STAR models. The function G is actually asymptotically homogenous of order 0 in the sense of Park and Phillips (1999); see also Park and Phillips (2001) and Chang et al. (2001). This type of behavior ultimately leads to asymptotic negligibility of the nonlinear part of the DGP; cf. the proof of Proposition 2 in Section 3.

In the nonlinear case, the most common choice of a transition function is an exponential, which leads to the popular Exponential STAR model

$$G_{1}\left(y_{t-1};\gamma\right) = 1 - e^{-\gamma\left(y_{t-1}-m\right)^{2}}$$

for some fixed shape parameter $\gamma > 0$ and location parameter $m \in \mathbb{R}$. The parameter γ can be interpreted as speed of transition between the inner and outer regimes; an equally relevant interpretation is provided by its proportionality to the width of the middle regime. A more flexible choice is given by the 2nd order logistic function,

$$G_{2}\left(y_{t-1};\gamma\right) = \frac{2}{1 + e^{-\gamma\left(y_{t-1}-m_{1}\right)\left(y_{t-1}-m_{2}\right)}} - 1$$

which has two location parameters $m_{1} \in \mathbb{R}$ and $m_{2} \in \mathbb{R}$ satisfying $m_{1} \leq m_{2}$. While both transition functions fulfill Assumption 5, they also fulfill the low-level Assumption 6 below.

The third case we discuss is motivated by our findings in Section 3, which prompts the use of alternatives for which the width of the middle regime grows with the sample size. Since the effective width is inversely proportional to the transition speed γ, this is equivalent to requiring γ to converge to 0 at an appropriate rate as $T \to \infty$ as described by the following

6

Assumption 6 *Let* $\gamma = \gamma_T = {}^{g^2}/_T$. *Let* $\mathcal{G}(\cdot)$ *be a (uniformly) Lipschitz function with* $\mathcal{G}(0) = 0$ *and* $\mathcal{G}(y) \to 1$ *as* $y \to \pm\infty$ *such that,* $\forall \gamma > 0$,

$$\left| G(y; \gamma) - \mathcal{G}(\gamma y^2) \right| \leq C\gamma (1 + |y|).$$

The condition for G from Assumption 6 enables us to characterize the weak limit of the suitably normalized process y_t; see the proof of Lemma 3. Unlike for Assumption 5, threshold autoregressions are excluded because of their intrinsic discontinuous nature.

Note that the ESTAR transition function satisfies Assumption 6 with $\mathcal{G} = 1 - \exp\{-z\}$, while for the 2nd order logistic function we have $\mathcal{G} = \frac{2}{1+\exp\{-z\}} - 1$. Since a nearly integrated process has at time t variance of magnitude $O(t)$, a strong similarity with the practice of standardizing the parameter γ by the standard deviation of the transition variable arises; see e.g. Teräsvirta (2004, p. 229). As the parameter γ enters the model exponentially, the rate $\gamma_T = O({}^1/_T)$ may appear arbitrary; see Park and Phillips (2001). But the derivations in the Appendix ultimately lead to the conclusion that a lower rate recovers nearly-integrated linearity in the limit, whereas a higher rate recovers the null hypothesis of integration. Thus, ${}^1/_T$-neighborhoods of the origin are the relevant ones for nonlinear behavior, at least in conjunction with the condition from Assumption 6.

2.2 Test procedures

Due to its popularity we focus on a specific unit root test in the ESTAR framework. Namely, the KSS test has been widely and successfully applied to a variety of variables and may thus be viewed as a standard unit root test against stationary self-exciting nonlinear models. Within the ESTAR framework, the null hypothesis of a unit root is parameterized by $\phi = 0$, and the alternative by $\phi < 0$ together with $\gamma > 0$. Without deterministic components, the DF test t_{DF} is based on the t statistic of ϕ from the OLS regression

$$\Delta y_t = \widehat{\varphi}_{DF} \, y_{t-1} + \widehat{\epsilon}_t, \tag{1}$$

while the nonlinear KSS test t_{KSS} relies on the t statistic of ϕ from the OLS regression

$$\Delta y_t = \widehat{\varphi}_{KSS} \, y_{t-1}^3 + \widehat{\epsilon}_t. \tag{2}$$

In (2), the specific choice of y_{t-1}^3 as regressor is due to employing a Taylor series expansion to deal with the lack of identification of γ under the null. The asymptotic null distribution

of the nonlinear KSS test statistic was derived by Kapetanios et al. (2003),

$$t_{KSS} \xrightarrow{d} \frac{\int_0^1 W^3(s)\, \mathrm{d}W(s)}{\sqrt{\int_0^1 W^6(s)\, \mathrm{d}s}}.\tag{3}$$

(Correct) Lag augmentation in case of serial dependence does not affect the asymptotics under the null; Rothe and Sibbertsen (2006) propose a nonparametric correction following Phillips and Perron (1988).

Comparing regressions (2) and (1), one can observe that large values of y_{t-1} are more pronounced, while small ones are reduced; this is useful inasmuch as the middle regime (small $|y_{t-1}|$) exhibits no mean reversion under a nonlinear alternative and there is little evidence in favor of stationarity to be obtained there. The nonlinear KSS test is expected to perform better than the DF test under (ESTAR) nonlinear alternatives; see Kapetanios et al. (2003).

When deterministic terms are considered, the standard procedure is to apply OLS demeaning and detrending procedures onto the series y_t before computing the test statistics. Denote the resulting test statistics by t_{DF}^μ, t_{KSS}^μ, t_{DF}^τ and t_{KSS}^τ. See Kapetanios et al. (2003), Kapetanios and Shin (2006) or Kapetanios and Shin (2008).

But efficient adjustment of deterministic terms when testing for unit roots is quite relevant. Several adjustment schemes have proven to be useful in the purely linear case; see Leybourne et al. (2005). Therefore, we employ the following adjustment schemes for both the linear DF and the nonlinear KSS tests and aim at identifying the most powerful technique.

1. **Local-to-unity GLS demeaning and detrending** as proposed by Elliott et al. (1996) for the DF test and Kapetanios and Shin (2008) for the KSS test. For both the linear and the nonlinear tests, the data are demeaned prior to application of the test in the following way: $y^{\mu;\bar{c}} = [y_1, y_2 - (1-\bar{c}/T)y_1, ..., y_T - (1-\bar{c}/T)y_{T-1}]'$ is regressed on $z^{\mu;\bar{c}} = [1, 1 - (1 - \bar{c}/T), \dots, 1 - (1 - \bar{c}/T)]'$ in order to obtain GLS residuals $\widetilde{y}_t^{\mu;\bar{c}}$; in the linear case $\bar{c} = 7$, while $\bar{c} = 9$ in the nonlinear case. The GLS unit root tests are then based on $\widetilde{y}_t^{\mu;\bar{c}}$. GLS detrending is carried out in analogous way, with $z^{\tau;\bar{c}}$ being $z_t^{\mu;\bar{c}}$ with the additional column $[1, 1 - \bar{c}/T, 1 - 2\bar{c}/T, \dots, 1 - \bar{c}(T-1)/T]'$ (i.e. the quasi-differenced trend); in the linear case $\bar{c} = 13.5$, while $\bar{c} = 17.5$ in the nonlinear case. Denote the resulting local-to-unity GLS unit root test statistics by $t_{DF}^{\mu;\bar{c}}$, $t_{KSS}^{\mu;\bar{c}}$, $t_{DF}^{\tau;\bar{c}}$ and $t_{KSS}^{\tau;\bar{c}}$.

2. **Recursive demeaning and detrending**. For recursive demeaning (So and Shin,

1999), the levels are recursively demeaned, $\widetilde{y}_{t-1}^{\mu} = y_{t-1} - \frac{1}{t}\sum_{j=1}^{t} y_j$, leading to

$$\widetilde{t}_{KSS}^{\mu} = \frac{\sum_{t=2}^{T} \left(\widetilde{y}_{t-1}^{\mu}\right)^3 \Delta y_t}{\widehat{\sigma}\sqrt{\sum_{t=2}^{T}\left(\widetilde{y}_{t-1}^{\mu}\right)^6}}$$

For the test with recursive detrending, one uses recursive detrending, and the differences are demeaned, cf. Taylor (2002) and Rodrigues (2006):

$$\widetilde{t}_{KSS}^{\tau} = \frac{\sum_{t=2}^{T} \left(\widetilde{y}_{t-1}^{\tau}\right)^3 \left(\Delta y_t - \overline{\Delta y}\right)}{\widehat{\sigma}\sqrt{\sum_{t=2}^{T}\left(\widetilde{y}_{t-1}^{\tau}\right)^6}}$$

with $\widetilde{y}_{t-1}^{\tau} = y_{t-1} + \frac{2}{t-1}\sum_{j=1}^{t-1} y_j - \frac{6}{t(t-1)}\sum_{j=1}^{t-1} j\, y_j$. The DF tests are constructed analogously and denoted by $\widetilde{t}_{DF}^{\mu}$ and $\widetilde{t}_{DF}^{\tau}$; see Shin and So (2001) and Taylor (2002). The recursive adjustment schemes reduce the bias of the OLS estimator (when close to the unit root) thereby enhancing power properties; see Shin and So (2001) and Taylor (2002) again.

3. **The maximum of the forward-computed and backward-computed** (i.e. applied to $y_t^* = y_{T+1-t}$.) DF tests with usual demeaning and detrending, as well as the corresponding nonlinear KSS tests. Denote the test statistics by $t_{DF}^{\mu;\max}$, $t_{KSS}^{\mu;\max}$, $t_{DF}^{\tau;\max}$ and $t_{KSS}^{\tau;\max}$. Strictly speaking, it is not a different adjustment scheme; but the max test procedure has (apart from new critical values) power properties comparable to the previous two modifications, at least in the linear case (Leybourne et al., 2005).

3 Nearly integrated STAR processes

Before discussing the nonlinear case, we review the benchmark of a linear local-to-unity alternative.

3.1 Linear local-to-unity alternatives

We compare the local power of linear DF and nonlinear KSS tests under the sequence of local linear alternatives from Assumption 3. Since the work of Phillips (1987) it is known what the weak limit of the suitably normalized y_t under such near integration is; the following lemma summarizes it. To formulate the lemma, denote by $J_c(s)$ the standard Ornstein-Uhlenbeck [OU] process with "mean-reversion" parameter c, $\mathrm{d}J_c(s) =$

$-cJ_c(s)\,\mathrm{d}s + \mathrm{d}W(s)$ with $J_c(0) = 0$ a.s., and let

$$J_{c,\kappa}(s) = \begin{cases} W(s) & c = 0 \\ \frac{\kappa}{\sigma\sqrt{2c}} e^{-cs} + J_c(s) & c > 0 \end{cases}.$$

This is nothing else than the standard OU process with starting value $\frac{\kappa}{\sigma\sqrt{2c}}$.

Lemma 1 *Under Assumptions 1, 2, 3 and 4 (linearity), it holds that*

$$\left(\frac{1}{\sigma\sqrt{T}} \sum_{j=1}^{[sT]} \varepsilon_j ; \frac{1}{\sigma\sqrt{T}} y_{[sT]} \right) \Rightarrow (W(s) ; J_{c,\kappa}(s))$$

as $T \to \infty$.

Proof: *omitted.*

The limiting distribution of the DF test statistic under a sequence of local linear alternatives has already been derived in the relevant literature for all examined methods of removing deterministic components. Although this is not entirely the case for the KSS test (where only OLS and GLS adjustment have been considered sofar), the following proposition gives the asymptotic behavior of the linear DF and nonlinear KSS test statistics only for the case without adjustment. The extensions for deterministic adjustment are actually particular cases of the results we give in Section 4 so there is no loss of generality in not mentioning them at this time.

Proposition 1 *Under the assumptions of Lemma 1 (linearity), it holds as $T \to \infty$ that*

$$t_{DF} \xrightarrow{d} \frac{\int_0^1 J_{c;\kappa}(s)\,\mathrm{d}W(s)}{\sqrt{\int_0^1 J_{c;\kappa}^2(s)\,\mathrm{d}s}} - c\sqrt{\int_0^1 J_{c;\kappa}^2(s)\,\mathrm{d}s};$$

the analogous result holds for the nonlinear KSS test,

$$t_{KSS} \xrightarrow{d} \frac{\int_0^1 J_{c;\kappa}^3(s)\,\mathrm{d}W(s)}{\sqrt{\int_0^1 J_{c;\kappa}^6(s)\,\mathrm{d}s}} - c\frac{\int_0^1 J_{c;\kappa}^4(s)\,\mathrm{d}s}{\sqrt{\int_0^1 J_{c;\kappa}^6(s)\,\mathrm{d}s}}.$$

Proof: *omitted.*

Inspection of the local power curves given in Figure 1 for $T = 1000$ and 20000 Monte Carlo replications reveals that the DF test dominates the nonlinear KSS test for all $c > 0$. This result was to be expected, since the nonlinear KSS test is after all misspecified when the DGP is linear, having replaced y_t with y_t^3 as a regressor. The nonlinear KSS test being
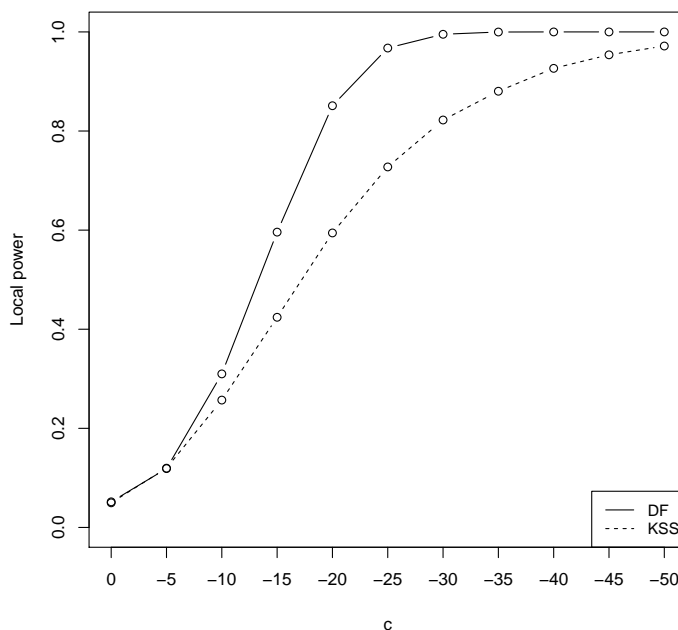
10

Figure 1: Local Power of DF and KSS test against a linear alternative.

explicitly built to take nonlinearities into account, the same reasoning points toward the nonlinear KSS dominating in terms of power when the DGP is nonlinear, say as specified by Assumption 5. This argument fails however in the near-integration setup, as shown in the following subsection.

3.2   Nonlinear local-to-unity alternatives

Consider now a fixed nonlinear local alternative, i.e. some fixed transition function which, according to Assumption 5, is not equal to 1 almost everywhere. Its shape is however fixed and does not depend on the sample size T. In spite of the nonlinearity of this DGP, Lemma 2 below shows that the limiting behavior of the suitably normalized y_t is the *same* as in the linear case (Lemma 1).

Lemma 2 *Under Assumptions 1, 2, 3 and 5 (fixed nonlinearity), it holds that*

$$\left(\frac{1}{\sigma\sqrt{T}} \sum_{j=1}^{[sT]} \varepsilon_j ; \frac{1}{\sigma\sqrt{T}} y_{[sT]} \right) \Rightarrow \left(W\left(s\right) ; J_{c,\kappa}\left(s\right) \right)$$

as $T \to \infty$.

Proof: *see the Appendix.*

The explanation for this counter-intuitive result is as follows. With y_t being nearly integrated in the outer regimes, the levels y_t have a relatively large variance even in small samples and they increasingly seldom reach the middle regime. The nonlinearity is consequently rendered asymptotically irrelevant for the levels when keeping the shape of the transition function fixed. Because of the negligibility of fixed nonlinearity in the near-integrated case, the initial value is plausibly specified in Assumption 1 as $y_0 = \kappa / \sqrt{1-(1+\phi)^2}$ even in the nonlinear case.

But the worst is yet to come: although the nonlinearity still enters the DF and KSS test statistics via Δy_t, the asymptotic distribution of neither of the two tests is affected by the nonlinearity of the DGP either. In other words, the limiting behavior of the test statistics under local-to-unity nonlinear alternatives is identical to that of the test statistics under the usual nearly-integrated linear alternative (Proposition 1). Like for the case of a linear DGP in Subsection 3.1, we prove the result for the case without deterministic terms only: considering removal of deterministic components may change the shape of the limiting distributions, but not the fact that the distributions are not affected by the (fixed) nonlinearity.

Proposition 2 *Under the assumptions of Lemma 2 (fixed nonlinearity), it holds as $T \to \infty$ that*

$$t_{DF} \xrightarrow{d} \frac{\int_0^1 J_{c;\kappa}(s)\, \mathrm{d}W(s)}{\sqrt{\int_0^1 J_{c;\kappa}^2(s)\, \mathrm{d}s}} - c\sqrt{\int_0^1 J_{c;\kappa}^2(s)\, \mathrm{d}s};$$

the analogous result holds for the nonlinear KSS test,

$$t_{KSS} \xrightarrow{d} \frac{\int_0^1 J_{c;\kappa}^3(s)\, \mathrm{d}W(s)}{\sqrt{\int_0^1 J_{c;\kappa}^6(s)\, \mathrm{d}s}} - c\frac{\int_0^1 J_{c;\kappa}^4(s)\, \mathrm{d}s}{\sqrt{\int_0^1 J_{c;\kappa}^6(s)\, \mathrm{d}s}}.$$

Proof: *see the Appendix.*

The results in Proposition 2 nicely confirm the findings of Choi and Moh (2007), who found by extensive Monte Carlo simulations that the power of unit root tests built against nonlinear alternatives is mostly influenced by the distance to the null of integration, i.e. by the parameter ϕ, and that the type of fixed nonlinearity does not matter too much. Our work complements their findings by theoretical results.

The key insight is that nonlinearity as specified by Assumption 5 is ultimately negligible compared to the variance of the (even nonlinear) nearly integrated process. This can also be seen as an identification problem as the middle regime is simply ignored asymptotically. See also Donauer et al. (2010). This problem stems from the DGP and not from the KSS test itself, since lack of identification of γ under the null is circumvented by using a Taylor series approximation.
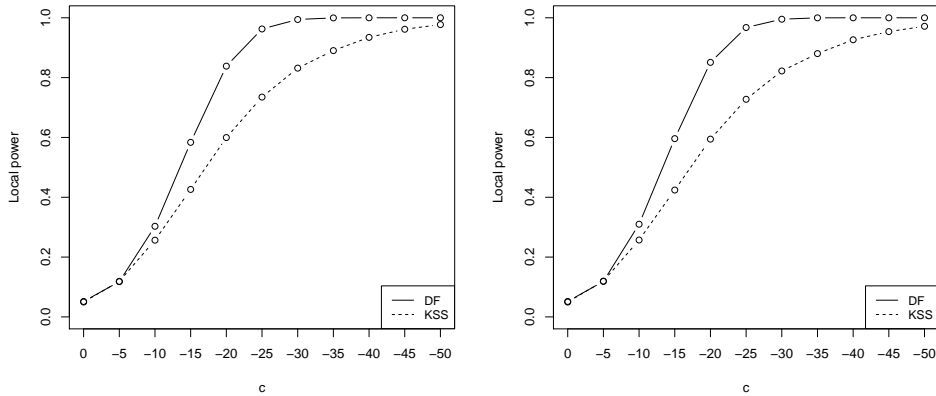
Figure 2: Local Power of DF and KSS test against a nonlinear (left) and linear ESTAR (right) alternative.

To illustrate the implications of the proposition we present a plot of the local power curves under a DGP with fixed nonlinearity in Figure 2 (left-hand panel). Again, the DF test dominates the nonlinear KSS test for all $c > 0$. Moreover, the curves are virtually identical to those in the linear case (right-hand panel of Figure 2).

Note that the findings apply to the TAR model as well, since Assumption 5 only requires a piecewise Lipschitz transition function. Also, several inner/middle regimes in STAR or TAR models are permitted; see also Kapetanios and Shin (2006).

Should however the variance of the errors ε_t be small, the overall variability of y_t is itself reduced; hitting the middle regime becomes more likely in this case and nonlinear dynamics will become relevant under the local alternative as well. Moreover, error variances estimated in empirical work are actually often rather small; see e.g. Taylor et al. (2001). Hence, the type of quasi-linear asymptotics with fixed nonlinearity specified by Assumption 5 may not be appropriate to capture the salient data features.

For this reason we study in Section 4 asymptotics for which the middle regime is not overlooked. We accomplish this by modeling the middle regime as having width proportional to the variance of y_t in order to mimic the effect of a small error variance. We call processes implied by Assumptions 3 and 6 persistently nonlinear. The following section deals with the local power of unit root tests under such a DGP.

4 Persistently nonlinear processes

4.1 Asymptotic results

We now examine the asymptotic behavior of the partial sums of Δy_t under the persistent nonlinearity assumption. To characterize it, let $X(s)$ be the diffusion given by the stochastic differential equation

$$\mathrm{d}X(s) = -cX(s)\,\mathcal{G}\left(g^2\sigma^2 X^2(s)\right)\mathrm{d}s + \mathrm{d}W(s) \tag{4}$$

with starting value $X(0) = \frac{\kappa}{\sigma\sqrt{2c}}$. We then have the following lemma.

Lemma 3 *Under Assumptions 1, 2, 3 and 6 (persistent nonlinearity), it holds that*

$$\left(\frac{1}{\sigma\sqrt{T}}\sum_{j=1}^{[sT]}\varepsilon_j;\ \frac{1}{\sigma\sqrt{T}}\,y_{[sT]}\right) \Rightarrow (W(s);X(s))$$

as $T \to \infty$.

Proof: *see the Appendix.*

Note that the limiting process $X(s)$ depends on the variance σ^2 even when the starting value is $y_0 = 0$; this is an important difference to the linear case.

The distributions of the linear DF and nonlinear KSS tests under the sequence of local alternatives specified by Assumption 6 is then as follows.

Proposition 3 *Under the assumptions of Lemma 3 (persistent nonlinearity), it holds that*

$$t_{DF} \xrightarrow{d} \frac{\int_0^1 X(s)\,\mathrm{d}W(s)}{\sqrt{\int_0^1 X^2(s)\,\mathrm{d}s}} - c\,\frac{\int_0^1 X^2(s)\,\mathcal{G}\left(g^2\sigma^2 X^2(s)\right)\mathrm{d}s}{\sqrt{\int_0^1 X^2(s)\,\mathrm{d}s}}$$

and

$$t_{KSS} \xrightarrow{d} \frac{\int_0^1 X^3(s)\,\mathrm{d}W(s)}{\sqrt{\int_0^1 X^6(s)\,\mathrm{d}s}} - c\,\frac{\int_0^1 X^4(s)\,\mathcal{G}\left(g^2\sigma^2 X^2(s)\right)\mathrm{d}s}{\sqrt{\int_0^1 X^6(s)\,\mathrm{d}s}}$$

as $T \to \infty$.

Proof: *see the Appendix.*

The extension to usual demeaning or detrending is straightforward and involves demeaned or detrendend versions of W and X. But Proposition 3 allows one to compare the behavior of the studied tests under the different assumptions about the nonlinearity: examining the results for linearity and fixed nonlinearity, we observe the DF and KSS tests to have nontrivial local power in the same type of neighborhoods of the null.

14

Since the power functions are not available in closed form, Subsection 4.2 compares the local power properties of the two tests via Monte Carlo simulation. The comparison will take different ways of adjusting for deterministic components into account. To this end, the following proposition derives the asymptotic behavior of the KSS when the data is adjusted for deterministic components as discussed in Subsection 2.2. The results for the DF test can be obtained analogously. To state the proposition, define the recursively demeaned version of the diffusion $X(s)$,

$$\widetilde{X}^{\mu}(s) = \begin{cases} 0 \text{ a.s.} & s = 0 \\ X(s) - \frac{1}{s}\int_0^s X(r)\,\mathrm{d}r & s > 0 \end{cases};$$

the recursively detrended version is

$$\widetilde{X}^{\tau}(s) = \begin{cases} 0 \text{ a.s.} & s = 0 \\ X(s) + \frac{2}{s}\int_0^s X(r)\,\mathrm{d}r - \frac{6}{s^2}\int_0^s rX(r)\,\mathrm{d}r & s > 0 \end{cases}$$

and let

$$\overline{M}_{\overline{c}} = \frac{(1+\overline{c})X(1) + \overline{c}^2\int_0^1 sX(s)\,\mathrm{d}s}{1 + \overline{c} + \frac{\overline{c}^2}{3}}.$$

We then have the following

Proposition 4 *Under the assumptions of Lemma 3 (persistent nonlinearity), it holds as $T \to \infty$:*
a) for local-to-unity GLS adjustment

$$t_{KSS}^{\mu,\overline{c}} \xrightarrow{d} \frac{\int_0^1 X^3(s)\,\mathrm{d}W(s)}{\sqrt{\int_0^1 X^6(s)\,\mathrm{d}s}} - c\frac{\int_0^1 X^4(s)\,\mathcal{G}(g^2\sigma^2 X^2(s))\,\mathrm{d}s}{\sqrt{\int_0^1 X^6(s)\,\mathrm{d}s}}$$

and

$$t_{KSS}^{\tau,\overline{c}} \xrightarrow{d} \frac{\int_0^1 \left(X(s) - s\overline{M}_{\overline{c}}\right)^3 \mathrm{d}W(s) - \overline{M}_{\overline{c}}\int_0^1 \left(X(s) - s\overline{M}_{\overline{c}}\right)^3 \mathrm{d}s}{\sqrt{\int_0^1 \left(X(s) - s\overline{M}_{\overline{c}}\right)^6 \mathrm{d}s}}$$
$$-c\frac{\int_0^1 \left(X(s) - s\overline{M}_{\overline{c}}\right)^3 X(s)\,\mathcal{G}(g^2\sigma^2 X^2(s))}{\sqrt{\int_0^1 \left(X(s) - s\overline{M}_{\overline{c}}\right)^6 \mathrm{d}s}};$$

b) for recursive adjustment

$$\widetilde{t}_{KSS}^{\mu} \xrightarrow{d} \frac{\int_0^1 \left(\widetilde{X}^{\mu}(s)\right)^3 \mathrm{d}W(s)}{\sqrt{\int_0^1 \left(\widetilde{X}^{\mu}(s)\right)^6 \mathrm{d}s}} - c\frac{\int_0^1 \left(\widetilde{X}^{\mu}(s)\right)^3 X(s)\,\mathcal{G}(g^2\sigma^2 X^2(s))\,\mathrm{d}s}{\sqrt{\int_0^1 \left(\widetilde{X}^{\mu}(s)\right)^6 \mathrm{d}s}}$$

15

and

$$\widetilde{t}^{\,\tau}_{KSS} \;\overset{d}{\to}\; \frac{\int_0^1 \left(\widetilde{X}^\tau(s)\right)^3 \mathrm{d}W(s) - W(1)\int_0^1 \left(\widetilde{X}^\tau(s)\right)^3 \mathrm{d}s}{\sqrt{\int_0^1 \left(\widetilde{X}^\tau(s)\right)^6 \mathrm{d}s}}$$

$$-c\,\frac{\int_0^1 \left(\widetilde{X}^\tau(s)\right)^3 X(s)\,\mathcal{G}\left(g^2\sigma^2 X^2(s)\right)\mathrm{d}s - \int_0^1 \left(\widetilde{X}^\tau(s)\right)^3 \mathrm{d}s \int_0^1 X(s)\,\mathcal{G}\left(g^2\sigma^2 X^2(s)\right)\mathrm{d}s}{\sqrt{\int_0^1 \left(\widetilde{X}^\tau(s)\right)^6 \mathrm{d}s}}.$$

c) for the max test *without adjustment*

$$t^{\max}_{KSS} \;\overset{d}{\to}\; \max\left\{ t_{KSS};\, -t_{KSS} - 3\frac{\int_0^1 X^2(s)\,\mathrm{d}s}{\sqrt{\int_0^1 X^6(s)\,\mathrm{d}s}} \right\}$$

with t_{KSS} from Proposition 3 being expressed in terms of the same diffusion $X(s)$; for the max test with OLS adjustment, replace the process $X(s)$ with its demeaned and detrended versions.

Proof: *see the Appendix.*

The behavior of the KSS test under GLS demeaning is the same as in the case without deterministics even in the nonlinear case, which is not true under detrending; see also Kapetanios and Shin (2008). In contrast, the distributions of the nonlinear KSS test with recursive adjustment and of the max KSS test change for both demeaning and detrending. The following subsection compares the local power of the DF and the KSS unit root tests from Proposition 4 and the respective effect of the employed adjustment procedures.

4.2   Local power curves

Since local power functions are not available in closed form in our case, we approximate them by means of Monte Carlo simulation. This section provides a summary of the findings. The following setup is considered for the analysis of local power. The sample size equals $T = 1,000$ and the data generating process is given by

$$\Delta y_t = \phi y_{t-1}(1 - \exp(-\gamma y_{t-1}^2)) + \varepsilon_t, \quad t = 2,...,T, \quad y_0 = \xi\sigma_y,$$

with $\phi = -c/T$, $\gamma = g^2/T$, $\varepsilon_t \sim N(0,\sigma^2)$ and $\xi = \{-4,-2,0,2,4\}$. The initial condition for y_0 is generated as the product of a parameter ξ of our choice and the average sample standard deviation σ_y of y_t (averaged over $20,000$ samples of size T generated with zero starting value). Values of ± 4 stands for a large initial value and ± 2 for moderate ones,
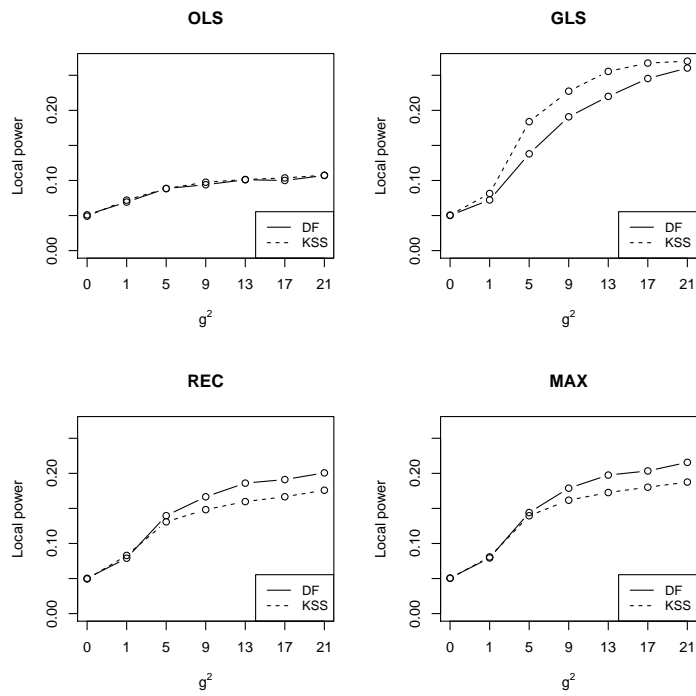
Figure 3: Local Power against ESTAR, $\sigma^2 = 1$, $c = 5$, $\xi = 0$.
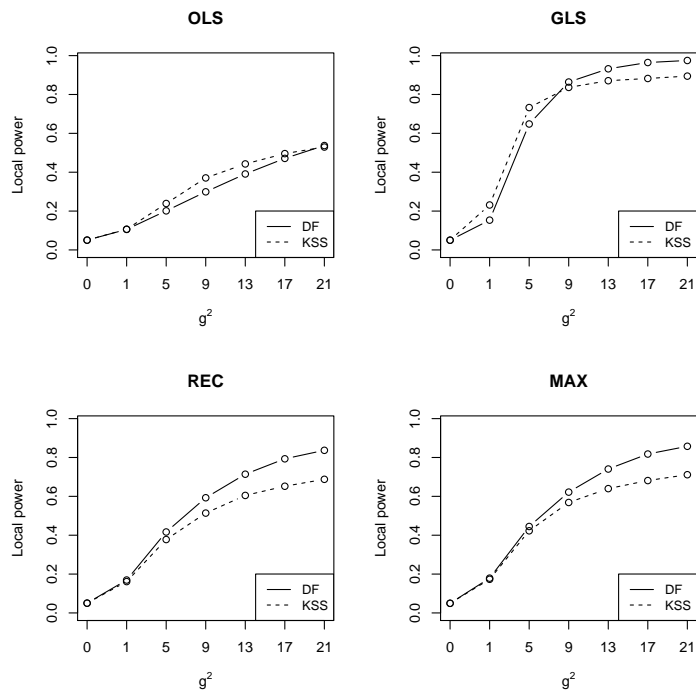

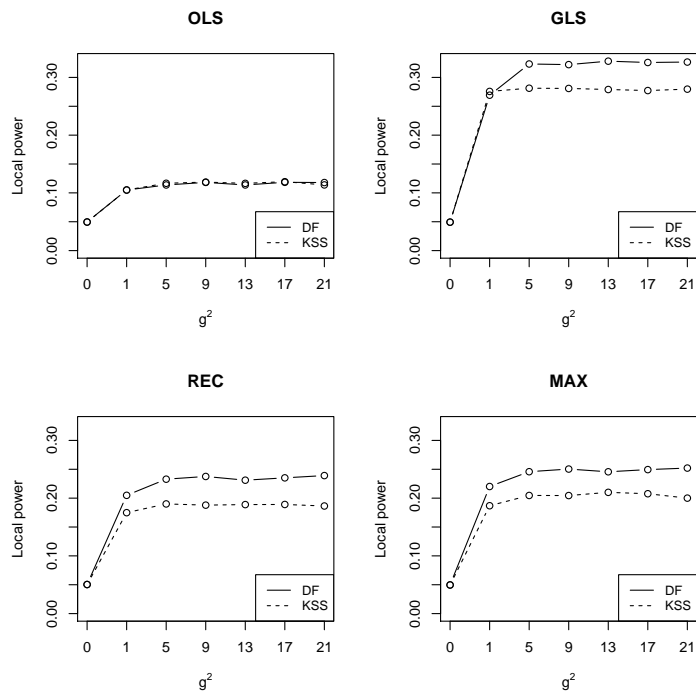Figure 4: Local Power against ESTAR, $\sigma^2 = 1$, $c = 20$, $\xi = 0$.

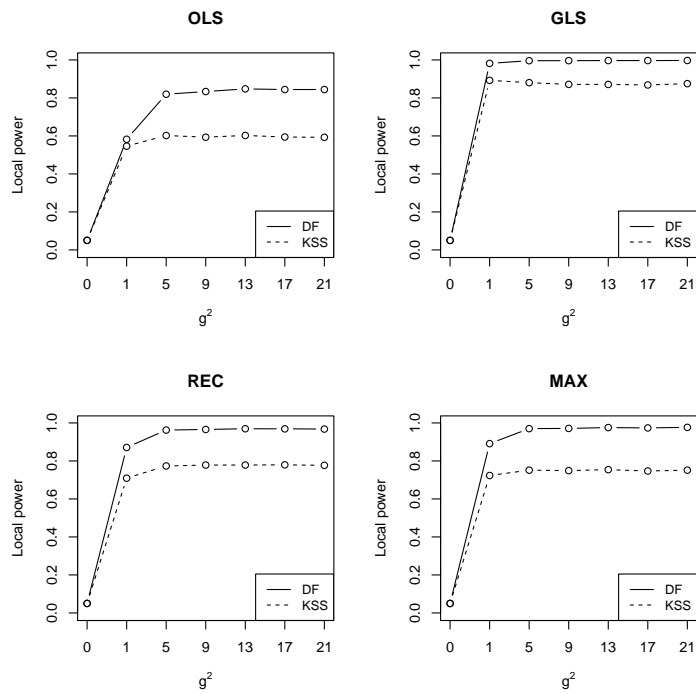Figure 5: Local Power against ESTAR, $\sigma^2 = 5$, $c = 5$, $\xi = 0$.

Figure 6: Local Power against ESTAR, $\sigma^2 = 5$, $c = 20$, $\xi = 0$.

while $\xi = 0$ reduces the initial value to 0. The sample standard deviation of y_t does not converge in probability in the near-integrated case, in fact we have the weak convergence

$$\frac{1}{\sqrt{T}}\sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \overline{y})^2} \Rightarrow \sqrt{\sigma^2 \int_0^1 \left(X(s) - \int_0^1 X(s)\mathrm{d}s\right)^2 \mathrm{d}s}$$

as $T \rightarrow \infty$. Under regularity conditions (in particular the existence of the expectation of the r.h.s.), however, the average sample deviation does converge,

$$\frac{\sigma_y}{\sqrt{T}} \xrightarrow{p} \sigma \, \mathrm{E}\left(\sqrt{\int_0^1 X^2(s)\mathrm{d}s - \left(\int_0^1 X(s)\mathrm{d}s\right)^2}\right)$$

so σ_y is proportional to $\sqrt{T}$ for suitable choice of κ as required by Assumption 1.[3]

The size of the unit root tests is considered whenever $g^2 = 0$, i.e. $\gamma = 0$, and the power of tests is considered whenever $g^2 > 0$ ($\gamma > 0$). Regarding local alternatives, we specify $c = \{1, 2, ..., 50\}$ and $g^2 = \{0, 1, 2, ..., 20\}$. Figures 3-14 cover the whole range of g^2 while c is fixed and takes either the value 5 or 20.[4] Regarding the variance of the errors, we consider the following scenarios: $\sigma^2 = \{0.1, 0.25, 0.5, 1, 2, 5\}$.

Following our asymptotic analysis, we compare the KSS and the DF test with four different adjustment schemes for deterministic components: OLS, GLS recursive (REC) and the maximum-procedure (MAX).[5] In the following, we only discuss results for the case of de-meaning as they are similar to the ones obtained for de-trending which we do not report here to save space. The number of replications is 5,000 for each parameter configuration.

Figures 3-14 highlight certain findings by selecting some specific parameter settings. Each figure shows four different mean adjustment schemes for two different tests, namely the DF (straight line) and the KSS test (dashed line). In particular, Figures 3-6 summarize the findings for a zero initial condition ($\xi = 0$). The remaining Figures cover the cases where ξ equals either 2 or 4. The results for negative values are very similar and therefore not reported.

In Figures 3 and 4, the error variance is fixed at $\sigma^2 = 1$, while c takes values 5 and 20, respectively. A comparison between the linear DF and the nonlinear KSS test for each single adjustment scheme reveals that there is no clear ranking. When different adjustment schemes are compared, it becomes evident that the GLS approach is most

---

[3]The simulation results for σ_y can be found in Table 1 located in the Appendix.

[4]The remaining figures are available from the authors upon request.

[5]Critical values simulated for large T for the newly developed KSS-REC and KSS-MAX tests at the nominal significance level of five percent are (-2.15, -2.64) and (-1.97, -3.17) for the case of de-meaning and de-trending, respectively.
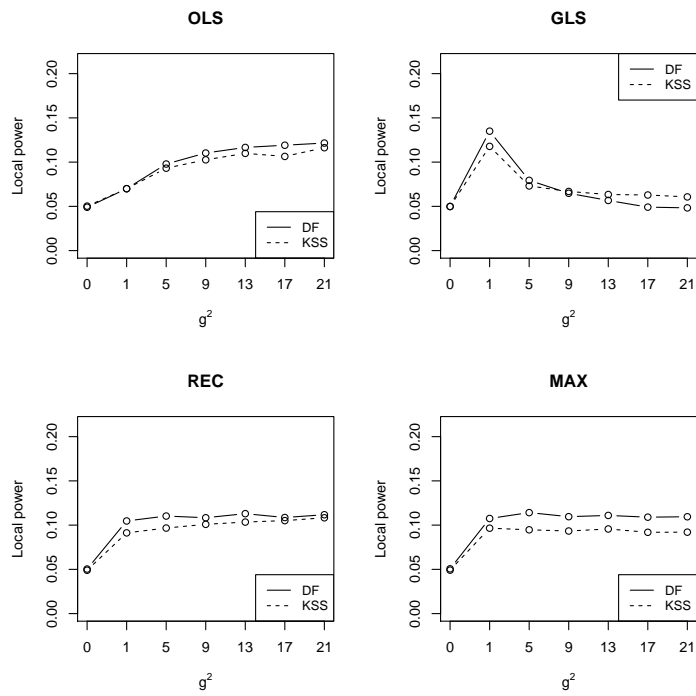
Figure 7: Local Power against ESTAR, $\sigma^2 = 1$, $c = 5$, $\xi = 2$.
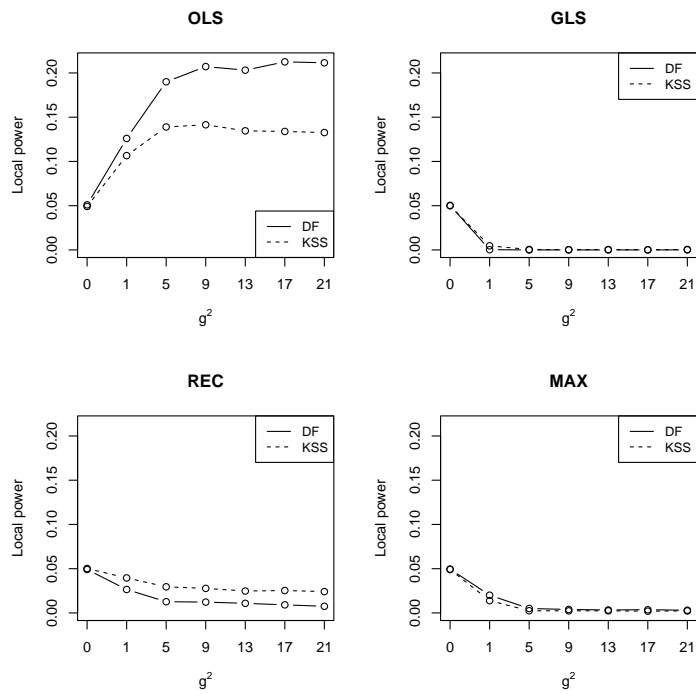

Figure 8: Local Power against ESTAR, $\sigma^2 = 1$, $c = 5$, $\xi = 4$.

Figure 9: Local Power against ESTAR, $\sigma^2 = 1$, $c = 20$, $\xi = 2$.

Figure 10: Local Power against ESTAR, $\sigma^2 = 1$, $c = 20$, $\xi = 4$.

Figure 11: Local Power against ESTAR, $\sigma^2 = 5$, $c = 5$, $\xi = 2$.

Figure 12: Local Power against ESTAR, $\sigma^2 = 5$, $c = 5$, $\xi = 4$.
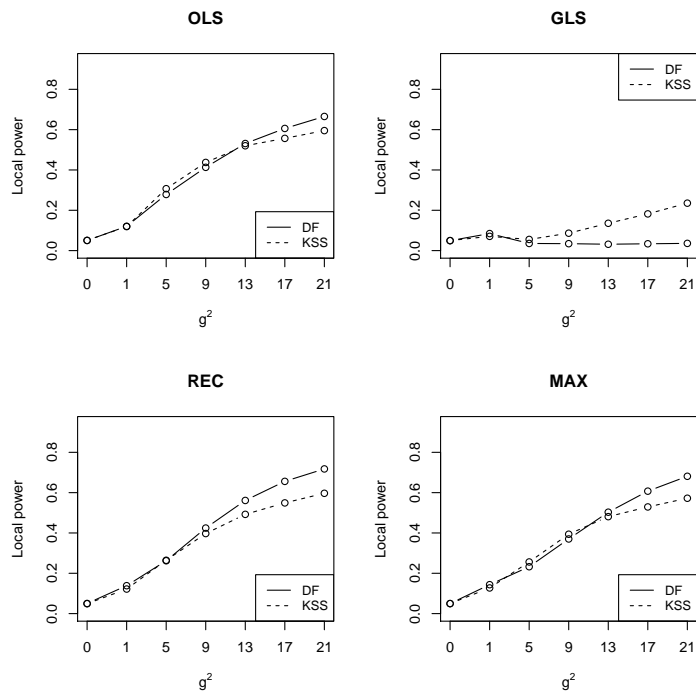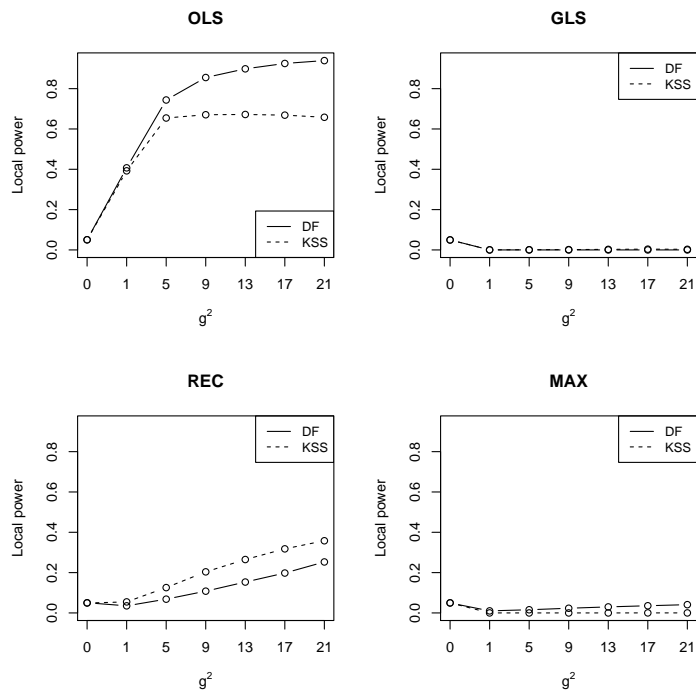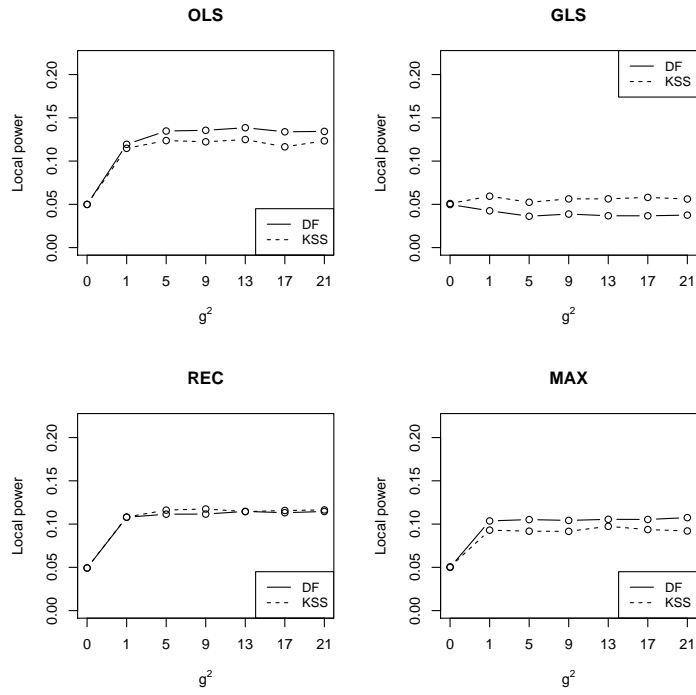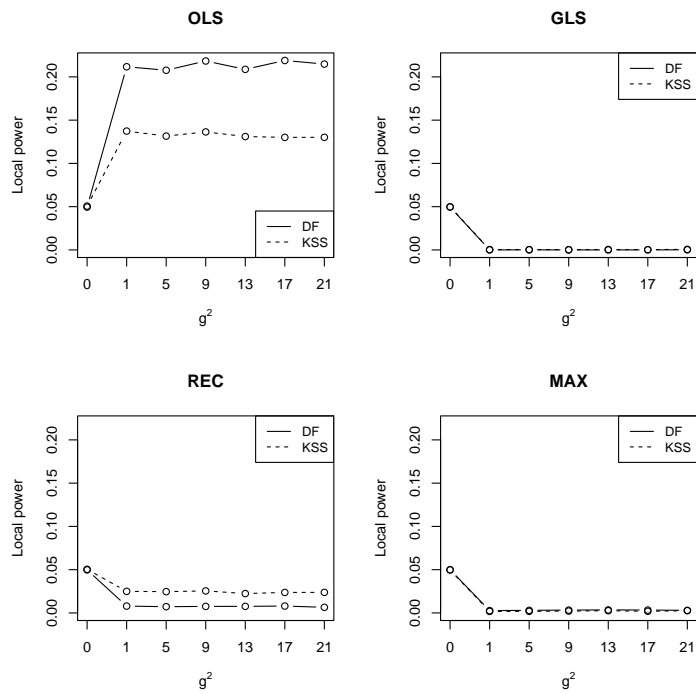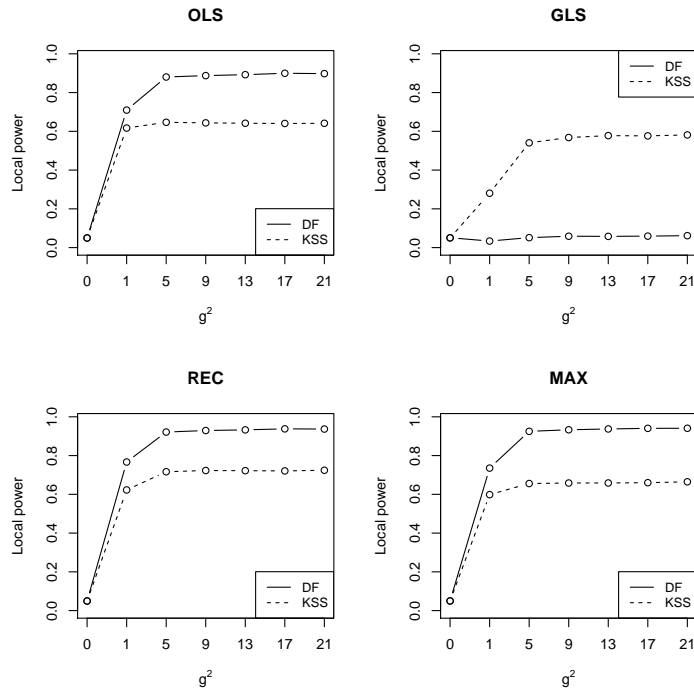
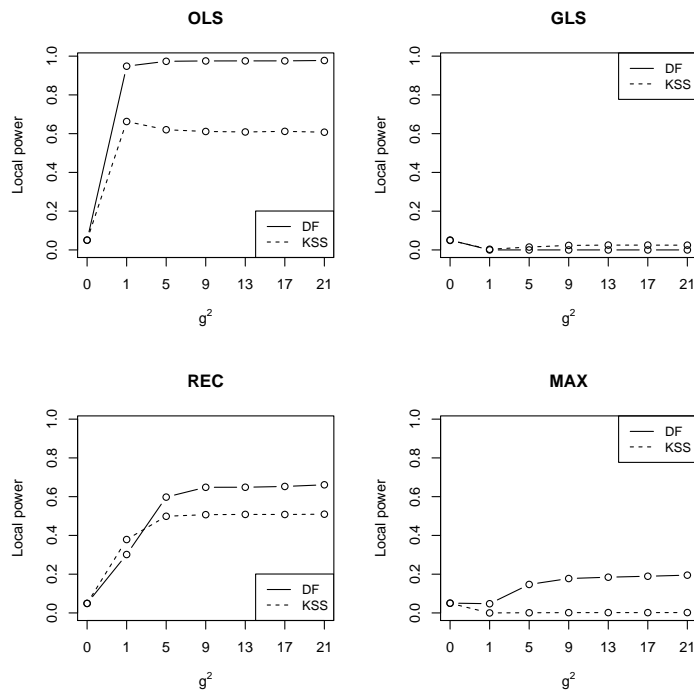Figure 13: Local Power against ESTAR, $\sigma^2 = 5$, $c = 20$, $\xi = 2$.

Figure 14: Local Power against ESTAR, $\sigma^2 = 5$, $c = 20$, $\xi = 4$.

promising. Recursive adjustment and the MAX procedure are performing very similar and both are better in terms of local power than their OLS counterparts. The OLS adjustment leads to the lowest local power overall. These findings imply, amongst other things, that a KSS test with OLS adjustment is dominated by a DF test with any other adjustment scheme than OLS. Hence, the results suggest that it may be more important to carefully treat the deterministic part of the model than to account for nonlinearity when it comes to unit root testing. This conclusion is further supported by the circumstance that, for a given adjustment scheme, the local power curve of the two tests are often quite close to each other. This means that even if the KSS test is performing better than the DF test in some situations (e.g. Figure 4 upper left cell, OLS adjustment), the gains in terms of local power coming from taking nonlinearity into account are small. The relative importance of efficient demeaning/detrending was also highlighted by Westerlund (2011) in a different setup.

Figures 5 and 6 are similar to the previous two, the difference is an error variance of 5 instead of 1. The simulation results support our theoretical findings and further indicate that for a fixed value of c, the degree of nonlinearity (measured by g^2) does not have a large impact on the local power. In fact, after a certain small threshold the curves become flat. In addition, the results suggest that the DF test dominates the KSS test in nearly all cases. One exception is found in Figure 5 (upper left cell, OLS adjustment) where both tests perform similar. In line with the previous setup with $\sigma^2 = 1$, the GLS approach turns out to be most efficient. Recursive adjustment and the MAX procedure are ranked second, while the OLS adjustment performs rather poorly in these settings as well. Moreover, it is evident from the results that the local power is, after a certain increase, hardly influenced by a change in the shape parameter of the transition function. A change in the autoregressive parameter appears to be more important. These findings nicely corroborate the results of Choi and Moh (2007), illustrating that the local power depends rather on c than on g^2.

We now turn to the case of a non-zero initial condition. Figures 7-10 cover the case of $\sigma^2 = 1$, while Figures 11-14 present the results for $\sigma^2 = 5$. In Figures 7 and 8 the specifications $\sigma^2 = 1, c = 5$ and $\xi = \{2, 4\}$ are considered. The results indicate that OLS adjustment performs best which comes along with monotonically increasing local power curves. For all other adjustment schemes we find either flat (see Figure 7 with $\xi = 2$) or decreasing (see Figure 8 with $\xi = 4$) local power curves. Only for the case of a large initial condition we find a clear ranking of the two competing tests: the linear DF test dominates the KSS test when OLS adjustment is applied. Figures 9 and 10 present results when c equals 20. For a moderate non-zero initial condition (see Figure 9 with $\xi = 2$) we find a similar performance of OLS, REC and MAX adjustment schemes while GLS is ranked last. Moreover, there is no clear distinction between the KSS and DF tests. Things are

different for $\xi = 4$ (Figure 10), though, where a similar picture to that of Figure 8 emerges (similar specification, but $c = 5$ instead of $c = 20$).

Finally, we summarize the findings for the case of $\sigma^2 = 5$ when the initial condition is non-zero. The results in Figure 11 ($c = 5, \xi = 2$) are very similar to the ones in Figure 3 ($c = 5, \xi = 0$) with the following important differences: the local power is lower in general and the OLS adjustment scheme provides the best performance followed by REC, MAX and GLS adjustment. We only find the KSS test to be slightly more powerful than the DF test for all considered values of g^2 if GLS adjustment is applied. The results in Figure 12 ($\xi = 4$) indicate once more the clear superiority of the DF test with OLS adjustment. By considering the results shown in Figure 13 it becomes clear that OLS, REC and MAX adjustment are performing pretty well and similar. For all three adjustment schemes, the DF test appears to be more powerful than the KSS test. This conclusion is reversed for the case of GLS adjustment. But still, the KSS test with GLS adjustment is performing worse than all of its competitors, i.e. KSS with either OLS, REC or MAX adjustment. In the last Figure 14, we find a clear ranking which we observe in other situations where the initial condition is large as well: the DF test with OLS adjustment is locally most powerful, followed by the KSS test with OLS demeaning. Notably, the recursive adjustment procedure (with a similar ranking of tests) is performing clearly better than its two remaining competitors, namely MAX and GLS adjustment.

We find that the initial condition has a strong impact on the ranking of adjustment schemes. See Müller and Elliott (2003) for the explanation in the linear case. While the GLS approach is most promising in the case of a zero initial condition, OLS adjustment is recommended for non-zero initial conditions. When OLS adjustment is considered under non-zero initial conditions, we observe that the DF test performs better than the KSS test in nearly all situations. In contrast, the ranking is less clear when GLS adjustment is applied under a zero initial condition. As a general conclusion, taking into account the outcomes of the studied tests and several different adjustment schemes is worthwhile to consider. Such a combination, as proposed in the linear case by Harvey et al. (2009, 2011) would ensure robustness against an unknown value of y_0.

5   Concluding remarks

This paper examines the local power of linear and nonlinear unit root tests under sequences of nearly integrated alternatives. They are either linear, nonlinear, or persistently nonlinear. The need to investigate persistently nonlinear alternatives arises from the interplay of near integration and nonlinearity: under fixed nonlinearity, the near-integrated behavior of the outer regimes dominates asymptotically unless the variance of the process is comparable to the width of the middle (unit root) regime.

We derive asymptotic distributions of various unit root test statistics and study their corresponding local power curves. Under linear and fixed nonlinear nearly-integrated alternatives, the linear DF test dominates the nonlinear KSS test in terms of local power. In contrast, under persistent nonlinearity, capturing e.g. small variance of the errors, the nonlinear KSS test can be more powerful. Furthermore, we find that the efficient removal of deterministic components is typically more important (in terms of local power) than using a specifically nonlinear test. The answer to the question which adjustment procedure is better depends to a large extent on the magnitude of the initial value; for instance, local-to-unity GLS should be preferred for negligible initial values while OLS appears to work better provided that the initial value is significant. These findings are in line with the knowledge of the linear case. But irrespective of the specific initial condition, we find that a DF test with the best available detrending procedure typically dominates a standard KSS test. Considering our results, it would be a more successful strategy to combine linear and nonlinear tests and several methods of detrending, following the suggestion of Harvey et al. (2009, 2011), rather than to set exclusively on the DF or the KSS test. We leave this to further work.

While the paper focuses on nonlinear models where the transition between regimes depends on lagged values of the process itself (SETAR/STAR), it is likely that our findings hold for other nonlinear DGPs exhibiting near integration. The findings of Choi and Moh (2007) support this conjecture to hold for other nonlinear DGPs as well. Furthermore, Balke and Fomby (1997) and Kapetanios et al. (2006) among others address the issue of nonlinear cointegration. The findings of this paper are expected to extend to nonlinear cointegration tests as well, as long as nonlinearity enters the system trough the error correction mechanism and not via the cointegrating relations only.

Appendix

Auxiliary results

Lemma 4 *Under Assumptions 1, 2, 3 and 6, it holds as $T \to \infty$ that*

1.
$$\frac{1}{T} \sum_{t=2}^{T} (\Delta y_t)^4 \overset{p}{\to} \mu_4 = \mathrm{E}\left(\varepsilon_t^4\right).$$

2.
$$\frac{1}{T^4} \sum_{t=2}^{T} y_{t-1}^6 \overset{d}{\to} \sigma^6 \int_0^1 X^6(s)\,\mathrm{d}s$$

3.
$$\frac{1}{T^2} \sum_{t=2}^{T} y_{t-1}^3 \varepsilon_t \overset{d}{\to} \sigma^4 \int_0^1 X^3(s)\,\mathrm{d}W(s)$$

4.
$$\frac{1}{T^2} \sum_{t=2}^{T} y_{t-1}^2 (\Delta y_t)^2 \overset{d}{\to} \sigma^4 \int_0^1 X^2(s)\,\mathrm{d}s$$

5.
$$\frac{1}{T^2} \sum_{t=2}^{T} y_{t-1} (\Delta y_t)^3 \overset{p}{\to} 0.$$

Note that, under a fixed nonlinear alternative, the results hold with $X(s)$ replaced by $J_{c;\kappa}(s)$.

Proof of Lemma 4

1. The item follows with $y_t = O_p\left(\sqrt{T}\right)$ (cf. Lemma 3), the sequence of equalities

$$\frac{1}{T} \sum_{t=2}^{T} (\Delta y_t)^4 = \frac{1}{T} \sum_{t=2}^{T} (\varepsilon_t + \phi_T y_{t-1} G(y_{t-1}; \gamma_T))^4 = \frac{1}{T} \sum_{t=2}^{T} \varepsilon_t^4 + o_p(1),$$

   and the Law of Large Numbers.

2. Follows from Lemma 3 with an application of the continuous mapping theorem [CMT].

3. Follows with Theorem 2.2 from Kurtz and Protter (1991), whose conditions (in particular the martingale difference property of $y_{t-1}\varepsilon_t$) are easily checked to hold given Lemma 3.

4. Write

$$\frac{1}{T^2} \sum_{t=2}^{T} y_{t-1}^2 (\Delta y_t)^2 = \frac{1}{T^2} \sum_{t=2}^{T} y_{t-1}^2 ((\Delta y_t)^2 - \sigma^2) + \sigma^2 \frac{1}{T^2} \sum_{t=2}^{T} y_{t-1}^2.$$

Should the first term on the r.h.s. vanish as $T \to \infty$, the result follows with the CMT and Lemma 3. Now,

$$\frac{1}{T^2} \sum_{t=2}^{T} y_{t-1}^2 ((\Delta y_t)^2 - \sigma^2)$$

$$= \frac{1}{T^2} \sum_{t=2}^{T} y_{t-1}^2 ((\varepsilon_t + \phi_T y_{t-1} G(y_{t-1}; \gamma_T))^2 - \sigma^2)$$

$$= \frac{1}{T^2} \sum_{t=2}^{T} y_{t-1}^2 (\varepsilon_t^2 - \sigma^2) + \frac{2\phi_T}{T^2} \sum_{t=2}^{T} y_{t-1}^3 G(y_{t-1}; \gamma_T) \varepsilon_t + \frac{\phi_T^2}{T^2} \sum_{t=2}^{T} y_{t-1}^4 G^2(y_{t-1}; \gamma_T);$$

note that $y_{t-1}^2 (\varepsilon_t^2 - \sigma^2)$ is an m.d. sequence with variance $t(\mu_4 - \sigma^4)$ under the *iid* assumption, so the first summand on the r.h.s. vanishes as $T \to \infty$, as the second and the third summands do thanks to the fact that $\mathcal{G}$ is bounded; the result follows. (If ε_t is a m.d. sequence with conditional heteroskedasticity, cumulant summability conditions are required to establish that $\mathrm{E}\left(\left(\frac{1}{T^2} \sum_{t=2}^{T} y_{t-1}^2 (\varepsilon_t^2 - \sigma^2)\right)^2\right) \to 0$; see e.g. Demetrescu, 2010.)

5. Note that

$$\frac{1}{T^2} \sum_{t=2}^{T} y_{t-1} (\Delta y_t)^3 = \frac{1}{T^2} \sum_{t=2}^{T} y_{t-1} (\varepsilon_t^3 - \mathrm{E}(\varepsilon_t^3)) + \mathrm{E}(\varepsilon_t^3) \frac{1}{T^2} \sum_{t=2}^{T} y_{t-1};$$

the second term on the r.h.s. vanishes in probability. The result follows with the *iid* property of the errors implying $\mathrm{E}|y_{t-1}(\varepsilon_t^3 - \mathrm{E}(\varepsilon_t^3))| = \mathrm{E}|y_{t-1}| \mathrm{E}|\varepsilon_t^3 - \mathrm{E}(\varepsilon_t^3)|$ where $\mathrm{E}|y_{t-1}| < C\sqrt{T} \ \forall 1 \le t \le T$ because $\mathrm{E}(y_{t-1}^2) \le Ct \ \forall t$, as can easily be checked by induction.

Proofs

Proof of Lemma 2

The case of $c = 0$ is trivial. For $c > 0$, define the auxiliary process $\widetilde{y}_t$ by

$$\Delta \widetilde{y}_t = \phi_T \widetilde{y}_{t-1} + \varepsilon_t$$

with $\widetilde{y}_0 = y_0$ and note that the weak limit of $\widetilde{y}_{[sT]}/\sqrt{\sigma^2 T}$ is precisely $J_{c,\kappa}(s)$ since

$$\widetilde{y}_t = \sum_{j=1}^{t} (1 + \phi_T)^{t-j} \varepsilon_j + \widetilde{y}_0 (1 + \phi_T)^t$$

and thus

$$\frac{1}{\sigma\sqrt{T}} \widetilde{y}_t = \frac{1}{\sigma\sqrt{T}} \sum_{j=1}^{t} (1 + \phi_T)^{t-j} \varepsilon_j + \frac{\kappa}{\sigma\sqrt{2c}} \left((1 + \phi_T)^{\frac{1}{\phi_T}} \right)^{-c\frac{t}{T}} + o(1).$$

It then suffices to show that

$$y_t = \widetilde{y}_t + O_p(1)$$

uniformly in t to establish the desired result. Begin by writing

$$|y_t - \widetilde{y}_t| \leq |y_{t-1} - \widetilde{y}_{t-1}| + |\Delta y_t - \Delta \widetilde{y}_t| ;$$

and note that

$$\Delta y_t = \phi_T \widetilde{y}_{t-1} + \phi_T (y_{t-1} - \widetilde{y}_{t-1}) - \phi_T y_{t-1} (1 - G(y_{t-1}; \gamma)) + \varepsilon_t$$

or

$$|\Delta y_t - \Delta \widetilde{y}_t| \leq |\phi_T| |y_{t-1} - \widetilde{y}_{t-1}| + |\phi_T y_{t-1} (1 - G(y_{t-1}; \gamma))| ,$$

hence

$$|y_t - \widetilde{y}_t| \leq (1 + |\phi_T|) |y_{t-1} - \widetilde{y}_{t-1}| + |\phi_T y_{t-1} (1 - G(y_{t-1}; \gamma))| .$$

Considering the assumption on the tail behavior of $1 - G$, it follows that

$$\sup_{1 \leq t \leq T} |\phi_T y_{t-1} (1 - G(y_{t-1}; \gamma))| \leq \frac{c \sup_{\gamma \in \mathbb{R}^+, y \in \mathbb{R}} |y(1 - G(y; \gamma))|}{T} = \frac{C}{T}$$

uniformly in t for a suitable constant C. With $y_0 = \widetilde{y}_0$, we then have for all t that

$$|y_t - \widetilde{y}_t| \leq \sum_{j=1}^{t} (1 + |\phi_T|)^j \frac{C}{T}$$

or

$$|y_t - \widetilde{y}_t| \leq \sum_{j=1}^{t} \left((1 + |\phi_T|)^{\frac{1}{|\phi_T|}} \right)^{j|\phi_T|} \frac{C}{T}.$$

Now, $(1 + |\phi_T|)^{\frac{1}{|\phi_T|}} \to e$ for $|\phi_T| \to 0$ so $(1 + |\phi_T|)^{\frac{1}{|\phi_T|}}$ is bounded; considering that $j\,|\phi_T| \le c$, we finally obtain

$$|y_t - \widetilde{y}_t| \le \sum_{j=1}^{t} \frac{C}{T},$$

as required for the result.

Proof of Lemma 3

Weak convergence of standardized partial sums of ε_t to Wiener process is a standard result in the literature under this paper's assumptions.

To prove the remaining results, note that the function $\mathcal{G}(\cdot)$ is (uniformly) Lipschitz so the solution $X(s)$ of the stochastic differential equation (4) exists uniquely.

Then,

$$\Delta y_t = \phi_T y_{t-1} \mathcal{G}\left(g^2 \sigma^2 \frac{y_{t-1}^2}{\sigma^2 T}\right) + \varepsilon_t + \phi_T y_{t-1}\left(G - \mathcal{G}\right);$$

since $|G - \mathcal{G}| \le \frac{C}{T}\left(1 + |y_{t-1}|\right),$ we have that

$$\Delta y_t = \phi_T y_{t-1} \mathcal{G}\left(g^2 \sigma^2 \frac{y_{t-1}^2}{\sigma^2 T}\right) + \varepsilon_t + O_p\left(T^{-1}\right)$$

where the O_p term is uniform in t. Hence y_{t-1} will have the same weak limit as $\widetilde{y}_t$ given by

$$\frac{\Delta \widetilde{y}_t}{\sigma \sqrt{T}} = c \frac{\widetilde{y}_{t-1}}{\sigma \sqrt{T}} \mathcal{G}\left(g^2 \sigma^2 \frac{\widetilde{y}_{t-1}^2}{\sigma^2 T}\right) \frac{1}{T} + \frac{\varepsilon_t}{\sigma \sqrt{T}}$$

with $\widetilde{y}_0 = y_0$.

Weak convergence of standardized $\widetilde{y}_{[sT]}$ to $X(s)$ (jointly with $T^{-0.5}\sum_{j=1}^{[sT]} \varepsilon_j$) follows from the work of Gikhman and Skorokhod (1969); see Kushner (1974) for the same convergence under similar conditions for $\mathcal{G}(\cdot)$ or, more recently, Kurtz and Protter (1991).

Proof of Proposition 2

Under fixed nonlinearity, we have for the DF test that

$$t_{DF} = \frac{\sum_{t=2}^{T} y_{t-1}\left(\varepsilon_t - \frac{c}{T} y_{t-1} G\left(y_{t-1}; \gamma\right)\right)}{\widehat{\sigma}\sqrt{\sum_{t=2}^{T} y_{t-1}^2}};$$

since $\widehat{\sigma}$ is consistent (the proof is straightforward, cf. the proof of Proposition 3), it follows that

$$t_{DF} = \frac{\frac{1}{\sigma^2 T} \sum_{t=2}^{T} y_{t-1}\varepsilon_t}{\sqrt{\frac{1}{\sigma^2 T^2} \sum_{t=2}^{T} y_{t-1}^2}} - c\sqrt{\frac{1}{\sigma^2 T^2} \sum_{t=2}^{T} y_{t-1}^2} - c\frac{\frac{1}{\sigma^2 T^2} \sum_{t=2}^{T} y_{t-1}^2 \left(1 - G\left(y_{t-1};\gamma\right)\right)}{\sqrt{\frac{1}{\sigma^2 T^2} \sum_{t=2}^{T} y_{t-1}^2}} + o_p\left(1\right).$$

Since $y^2\left(1 - G\left(y;\cdot\right)\right)$ is bounded at $\pm\infty$, the Lipschitz assumption implies that it is bounded over $\mathbb{R}$. The third term on the r.h.s. hence vanishes as $T \to \infty$. The proof is analogous for the nonlinear KSS test.

Proof of Proposition 3

Begin with the DF test and show $\widehat{\varphi}_{DF}$,

$$\widehat{\varphi}_{DF} = \frac{\sum_{t=2}^{T} y_{t-1}\Delta y_t}{\sum_{t=2}^{T} y_{t-1}^2}, \tag{5}$$

to be superconsistent. For $\gamma_T = g^2/T$, we have along the lines of the proof of Lemma 3 that

$$\sum_{t=2}^{T} y_{t-1}\Delta y_t = \sum_{t=2}^{T} y_{t-1}\varepsilon_t - c\frac{1}{T}\sum_{t=2}^{T} y_{t-1}^2 \mathcal{G}\left(\frac{g^2}{T}y_{t-1}^2\right) + O_p\left(\frac{1}{T^2}\sum y_{t-1}^3\right);$$

furthermore, $\mathcal{G}$ is bounded, $y_t = O_p\left(\sqrt{T}\right)$ and $y_{t-1}\varepsilon_t$ is a m.d. sequence, so the numerator of the r.h.s. of (5) is $O_p(T)$. The denominator in (5) is of exact order $O_p\left(T^2\right)$, establishing superconsistency of $\widehat{\varphi}_{DF}$ and consistency of $\widehat{\sigma}^2 = T^{-1}\sum_{t=2}^{T}(\Delta y_t - \widehat{\varphi}_{DF}y_{t-1})^2$. Then,

$$t_{DF} = \frac{\sum_{t=2}^{T} y_{t-1}\Delta y_t}{\widehat{\sigma}\sqrt{\sum_{t=2}^{T} y_{t-1}^2}} = \frac{\frac{1}{\sigma^2 T} \sum_{t=2}^{T} y_{t-1}\varepsilon_t}{\sqrt{\frac{1}{\sigma^2 T^2} \sum_{t=2}^{T} y_{t-1}^2}} - c\frac{\frac{1}{\sigma^2 T^2} \sum_{t=2}^{T} y_{t-1}^2 \mathcal{G}\left(g^2\sigma^2\frac{y_{t-1}^2}{\sigma^2 T}\right)}{\sqrt{\frac{1}{\sigma^2 T^2} \sum_{t=2}^{T} y_{t-1}^2}} + o_p\left(1\right).$$

Convergence follows with the CMT and Lemma 4 item 4. For the nonlinear KSS test, the result follows similarly.

Proof of Proposition 4

a) local-to-unity GLS   We have from Elliott et al. (1996) that

$$\widetilde{y}_t^{\mu} = y_t - \widehat{\mu}$$

with $\widehat{\mu}$ the GLS estimator of μ under an assumed linear local alternative $\rho = 1 - \frac{\bar{c}}{T}$. This leads to

$$\widetilde{y}_t^\mu = y_t - \frac{y_1 + \frac{\bar{c}}{T}\left(y_T - y_1\right) + \frac{\bar{c}^2}{T^2}\sum_{t=2}^T y_{t-1}}{1 + \frac{\bar{c}^2}{T^2}\left(T-1\right)} = y_t + O_p\left(T^{-0.5}\right).$$

Moreover, $\Delta \widetilde{y}_t^\mu = \Delta y_t$, so the asymptotics of the GLS-demeaned KSS test are the same as that of the KSS test without deterministics.

In the case of detrending, we obtain on the one hand

$$\begin{aligned}
\widetilde{y}_t^\tau &= y_t - \frac{t}{T\left(1 + \bar{c} + \frac{\bar{c}^2}{3}\right)}\left(y_T - y_1 + \frac{\bar{c}}{T}\sum_{t=2}^T (t-1)\Delta y_t + \frac{\bar{c}}{T}\sum_{t=2}^T y_{t-1} + \frac{\bar{c}^2}{T^2}\sum_{t=2}^T (t-1)y_{t-1}\right) \\
&\quad + O_p\left(T^{-0.5}\right)
\end{aligned}$$

leading with $\sum_{t=2}^T (t-1)\Delta y_t = (T-1)\left(y_T - y_1\right) - \sum_{t=2}^T y_{t-1}$ and the CMT to

$$\frac{1}{\sigma\sqrt{T}}\widetilde{y}_t^\tau \Rightarrow X(s) - s\frac{(1+\bar{c})X(1) + \bar{c}^2\int_0^1 sX(s)\,\mathrm{d}s}{1 + \bar{c} + \frac{\bar{c}^2}{3}} = X(s) - s\overline{M_{\bar{c}}}.$$

On the other hand we have that

$$\Delta \widetilde{y}_t^\tau = \Delta y_t - \frac{(1+\bar{c})\left(y_T - y_1\right) + \frac{\bar{c}^2}{T^2}\sum_{t=2}^T (t-1)y_{t-1}}{T\left(1 + \bar{c} + \frac{\bar{c}^2}{3}\right)} + o_p\left(T^{0.5}\right)$$

The result follows again with the CMT, since $\Delta y_t = \phi y_{t-1}\mathcal{G}\left(\gamma y_{t-1}^2\right) + O_p\left(T^{-1}\right)$.

b) recursive detrending Like in the proof of Proposition 3, it follows from Assumption 6 that

$$\widetilde{t}_{NL}^\mu = \frac{\sum_{t=2}^T \left(\widetilde{y}_{t-1}^\mu\right)^3 \varepsilon_t}{\widehat{\sigma}\sqrt{\sum_{t=2}^T \left(\widetilde{y}_{t-1}^\mu\right)^6}} - \frac{c}{T}\frac{\sum_{t=2}^T \left(\widetilde{y}_{t-1}^\mu\right)^3 y_{t-1}\mathcal{G}\left(\gamma_T y_{t-1}^2\right)}{\widehat{\sigma}\sqrt{\sum_{t=2}^T \left(\widetilde{y}_{t-1}^\mu\right)^6}} + o_p(1).$$

Using Proposition 2 in Born and Demetrescu (2011) together with Lemma 3, we have that

$$\frac{1}{\sigma\sqrt{T}}\widetilde{y}_{[sT]}^\mu \Rightarrow \widetilde{X}^\mu(s)$$

jointly with weak convergence of the normalized partial sums of ε_t. Proposition 2.2 in Kurtz and Protter (1991) and the CMT lead to the desired result.

For the test with recursive detrending, we have

$$\widetilde{t}_{NL}^\tau = \frac{\sum_{t=2}^T \left(\widetilde{y}_{t-1}^\tau\right)^3\left(\varepsilon_t - \bar{\varepsilon}\right)}{\widehat{\sigma}\sqrt{\sum_{t=2}^T \left(\widetilde{y}_{t-1}^\tau\right)^6}} - \frac{c}{T}\frac{\sum_{t=2}^T \left(\widetilde{y}_{t-1}^\tau\right)^3\left(y_{t-1}\mathcal{G}\left(\gamma_T y_{t-1}^2\right) - \overline{y\,\mathcal{G}\left(\gamma_T y^2\right)}\right)}{\widehat{\sigma}\sqrt{\sum_{t=2}^T \left(\widetilde{y}_{t-1}^\tau\right)^6}} + o_p(1);$$

the arguments used for the demeaning case apply analogously to complete the proof.

c) max test   The reverse nonlinear KSS test is given by

$$t^r_{KSS} = \frac{\sum_{t=T-1}^{1} y_{t+1}^3 (y_t - y_{t+1})}{\widehat{\sigma} \sqrt{\sum_{t=T-1}^{1} y_{t+1}^6}}$$

with $\widehat{\sigma}^2 = T^{-1} \sum_{t=T-1}^{1} \left(-\Delta y_{t+1} - \widehat{\varphi}^r_{DF} y_{t+1}^3\right)^2$ and $\widehat{\varphi}^r_{DF} = \frac{\sum_{t=T-1}^{1} y_{t+1}^3 (y_t - y_{t+1})}{\sum_{t=T-1}^{1} y_{t+1}^6}$.

Now,

$$\sum_{t=T-1}^{1} y_{t+1}^3 (y_t - y_{t+1}) = -\sum_{t=2}^{T} y_t^3 \Delta y_t = -\sum_{t=2}^{T} (y_{t-1} + \Delta y_t)^3 \Delta y_t$$

so

$$\sum_{t=T-1}^{1} y_{t+1}^3 (y_t - y_{t+1}) = -\sum_{t=2}^{T} \left(y_{t-1}^3 \Delta y_t + 3 y_{t-1}^2 (\Delta y_t)^2 + 3 y_{t-1} (\Delta y_t)^3 + (\Delta y_t)^4\right).$$

Then, by making use of Lemma 4 items 1 and 5, we obtain that

$$\widehat{\varphi}^r_{DF} = -\frac{\sum_{t=2}^{T} y_t^3 \varepsilon_t}{\sum_{t=2}^{T} y_t^6} + \frac{c}{T} \frac{\sum_{t=2}^{T} y_t^4 \mathcal{G}\left(\gamma_T y_{t-1}^2\right)}{\sum_{t=2}^{T} y_t^6} - \frac{\sum_{t=2}^{T} 3 y_{t-1}^2 (\Delta y_t)^2}{\sum_{t=2}^{T} y_t^6} + o_p(1);$$

note that the 2nd term on the r.h.s. vanishes since hence, by using Lemma 4 again, we obtain

$$T^2 \widehat{\varphi}^r_{DF} \xrightarrow{d} \frac{-\int_0^1 X^3(s) \, dW(s) + c \int_0^1 X^4(s) \mathcal{G}(g^2 \sigma^2 X^2(s)) \, ds - 3 \int_0^1 X^2(s) \, ds}{\sigma^2 \int_0^1 X^6(s) \, ds}.$$

As a direct consequence of this convergence rate, the residuals and the residual variance estimator are consistent. Moving on to the t statistic, we obtain

$$t^r_{KSS} = -\frac{\frac{1}{\sigma^4 T^2} \sum_{t=2}^{T} y_t^3 \Delta y_t}{\frac{\widehat{\sigma}}{\sigma} \sqrt{\frac{1}{\sigma^6 T^4} \sum_{t=2}^{T} y_t^6}}$$

and correspondingly

$$t^r_{KSS} \xrightarrow{d} \frac{-\int_0^1 X^3(s) \, dW(s) + c \int_0^1 X^4(s) \mathcal{G}(g^2 \sigma^2 X^2(s)) \, ds - 3 \int_0^1 X^2(s) \, ds}{\sqrt{\int_0^1 X^6(s) \, ds}}.$$

For the max test with OLS adjustment, simply replace the processes with their demeaned and detrended versions to complete the result.

Table

Table 1: **Simulated values for σ_y^2**

	(σ^2, c)			
g^2	$(1,5)$	$(1,20)$	$(5,5)$	$(5,20)$
1	9.35	7.96	39.01	26.40
5	8.38	6.47	38.39	23.92
9	8.07	5.96	38.12	23.59
13	7.97	5.69	38.24	23.52
17	7.89	5.51	38.22	23.48
21	7.83	5.37	38.18	23.41

Nonlinear data generating process is given by: $\Delta y_t = \phi y_{t-1}(1 - \exp(-\gamma y_{t-1}^2)) + \varepsilon_t$, with $\phi = -c/T$, $\gamma = g^2/T$, $\varepsilon_t \sim N(0, \sigma^2)$. Reported values are Monte Carlo averages of the sample variance of y_t taken over 20,000 replications. The sample size T equals 1,000.

References

Balke, N. S. and T. B. Fomby (1997). Threshold Cointegration. *International economic review 38*(3), 627–645.

Bec, F., M. Ben Salem, and M. Carrasco (2004). Tests for Unit-Root versus Threshold Specification with an Application to the Purchasing Power Parity Relationship. *Journal of Business and Economic Statistics 22*(4), 382–395.

Bec, F., A. Guay, and E. Guerre (2008). Adaptive Consistent Unit-Root Tests Based on Autoregressive Threshold Model. *Journal of Econometrics 142*(1), 94–133.

Born, B. and M. Demetrescu (2011). Recursive Adjustment for General Deterministic Components and Improved Tests for the Cointegration Rank. Working paper, Department of Economics, University of Bonn.

Chang, Y., J. Y. Park, and P. C. B. Phillips (2001). Nonlinear Econometric Models with Cointegrated and Deterministically Trending Regressors. *The Econometrics Journal 4*(1), 1–36.

Choi, C. Y. and Y. K. Moh (2007). How Useful are Tests for Unit-Root in Distinguishing Unit-Root Processes from Stationary but Non-Linear Processes? *The Econometrics Journal 10*(1), 82–112.

Demetrescu, M. (2010). On the Dickey-Fuller Test with White Standard Errors. *Statistical Papers 51*(1), 11–25.

Dickey, D. A. and W. A. Fuller (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association 74*(366), 427–431.

Donauer, S., F. Heinen, and P. Sibbertsen (2010). Identification Problems in ESTAR Models and a New Model. Diskussionspapiere der Wirtschaftswissenschaftlichen Fakultät, Universität Hannover.

Elliott, G., T. J. Rothenberg, and J. H. Stock (1996). Efficient Tests for an Autoregressive Unit Root. *Econometrica 64*(4), 813–836.

Gikhman, I. I. and A. V. Skorokhod (1969). *Introduction to the Theory of Stochastic Processes*. Saunders, Philadelphia.

Harvey, D. I., S. J. Leybourne, and A. M. R. Taylor (2009). Unit Root Testing in Practice: Dealing with uncertainty over the trend and initial condition. *Econometric Theory 25*(3), 587–636.

Harvey, D. I., S. J. Leybourne, and A. M. R. Taylor (2011). Unit Root Testing Under a Local Break in Trend. *Journal of Econometrics forthcoming*.

He, C. and R. Sandberg (2006). Dickey–Fuller Type of Tests against Nonlinear Dynamic Models. *Oxford Bulletin of Economics and Statistics 68*(S1), 835–861.

de Jong, R. M. (2009). Nonlinear Time Series Models and Weakly Dependent Innovation. Working paper, Department of Economics, The Ohio State University.

Kapetanios, G. and Y. Shin (2006). Unit Root Tests in Three-Regime SETAR Models. *Econometrics Journal 9*(2), 252–278.

Kapetanios, G. and Y. Shin (2008). GLS Detrending-Based Unit Root Tests in Nonlinear STAR and SETAR Models. *Economics Letters 100*(3), 377–380.

Kapetanios, G., Y. Shin, and A. Snell (2003). Testing for a Unit Root in the Nonlinear STAR Framework. *Journal of Econometrics 112*(2), 359–379.

Kapetanios, G., Y. Shin, and A. Snell (2006). Testing for Cointegration in Nonlinear Smooth Transition Error Correction Models. *Econometric Theory 22*(02), 279–303.

Kurtz, T. G. and P. Protter (1991). Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations. *Annals of Probability 19*(3), 1035–1070.

Kushner, H. J. (1974). On the Weak Convergence of Interpolated Markov Chains to a Diffusion. *The Annals of Probability 2*(1), 40–50.

Leybourne, S., T. H. Kim, and P. Newbold (2005). Examination of Some More Powerful Modifications of the Dickey-Fuller Test. *Journal of Time Series Analysis 26*(3), 355–369.

Leybourne, S. J. (1995). Testing for Unit Roots Using Forward and Reverse Dickey-Fuller Regressions. *Oxford Bulletin of Economics and Statistics 57*(4), 559–571.

McMillan, D. G. (2007). Bubbles in the Dividend-Price Ratio? Evidence from an asymmetric exponential smooth-transition model. *Journal of Banking & Finance 31*(3), 787–804.

Müller, U. K. and G. Elliott (2003). Tests for Unit Roots and the Initial Condition. *Econometrica 71*(4), 1269–1286.

Nobay, B., I. Paya, and D. A. Peel (2010). Inflation Dynamics in the U.S.: Global but not local mean reversion. *Journal of Money, Credit and Banking 42*(1), 135–150.

Park, J. Y. and P. C. B. Phillips (1999). Asymptotics for Nonlinear Transformations of Integrated Time Series. *Econometric Theory 15*(3), 269–298.

Park, J. Y. and P. C. B. Phillips (2001). Nonlinear Regressions with Integrated Time Series. *Econometrica 69*(1), 117–161.

Phillips, P. C. B. (1987). Towards a Unified Asymptotic Theory for Autoregression. *Biometrika 74*(3), 535–547.

Phillips, P. C. B. and P. Perron (1988). Testing for a Unit Root in Time Series Regression. *Biometrika 75*(2), 335–346.

Rodrigues, P. M. M. (2006). Properties of Recursive Trend-Adjusted Unit Root Tests. *Economics Letters 91*(3), 413–419.

Rothe, C. and P. Sibbertsen (2006). Phillips-Perron-Type Unit Root Tests in the Nonlinear ESTAR Framework. *Allgemeines Statistisches Archiv 90*(3), 439–456.

Sarantis, N. (1999). Modeling Non-Linearities in Real Effective Exchange Rates. *Journal of International Money and Finance 18*(1), 27–45.

Sarno, L., M. P. Taylor, and D. A. Peel (2003). Nonlinear Equilibrium Correction in U.S. Real Money Balances, 1869-1997. *Journal of Money, Credit and Banking 35*(5), 787–799.

Shin, D. W. and B. S. So (2001). Recursive Mean Adjustment for Unit Root Tests. *Journal of Time Series Analysis 22*(5), 595–612.

So, B. S. and D. W. Shin (1999). Recursive Mean Adjustment in Time-Series Inferences. *Statistics & Probability Letters 43*(1), 65–73.

Taylor, A. M. R. (2002). Regression-Based Unit Root Tests With Recursive Mean Adjustment for Seasonal and Nonseasonal Time Series. *Journal of Business & Economic Statistics 20*(2), 269–281.

Taylor, M. P., D. A. Peel, and L. Sarno (2001). Nonlinear Mean-Reversion in Real Exchange Rates: Toward a solution to the Purchasing Power Parity puzzles. *International Economic Review 42*(4), 1015–1042.

Teräsvirta, T. (1994). Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models. *Journal of the American Statistical Association 89*(425), 208–218.

Teräsvirta, T. (2004). Smooth Transition Regression Modeling. In H. Lütkepohl and M. Krätzig (Eds.), *Applied Time Series Econometrics*. Cambridge University Press.

Teräsvirta, T. and H. M. Anderson (1992). Characterizing Nonlinearities in Business Cycles Using Smooth Transition Autoregressive Models. *Journal of Applied Econometrics 7*(S1), S119–S136.

Tong, H. (1990). *Non-linear time series: a dynamical system approach.* Oxford University Press.

Tweedie, R. L. (1975). Sufficient Conditions for Ergodicity and Recurrence of Markov Chains on a General State Space. *Stochastic Processes and their Applications 3*(4), 385–403.

Westerlund, J. (2011). A Simple Information-Intensive Unit Root Test. Working paper, University of Gothenburg.

Research Papers
2012

2011-40: Søren Johansen and Bent Nielsen: Asymptotic theory for iterated one-step Huber-skip estimators

2011-41: Luc Bauwens, Arnaud Dufays and Jeroen V.K. Rombouts: Marginal Likelihood for Markov-switching and Change-point Garch Models

2011-42: Manuel Lukas: Utility-based Forecast Evaluation with Multiple Decision Rules and a New Maxmin Rule

2011-43: Peter Christoffersen, Ruslan Goyenko, Kris Jacobs, Mehdi Karoui: Illiquidity Premia in the Equity Options Market

2011-44: Diego Amaya, Peter Christoffersen, Kris Jacobs and Aurelio Vasquez: Do Realized Skewness and Kurtosis Predict the Cross-Section of Equity Returns?

2011-45: Peter Christoffersen and Hugues Langlois: The Joint Dynamics of Equity Market Factors

2011-46: Peter Christoffersen, Kris Jacobs and Bo Young Chang: Forecasting with Option Implied Information

2011-47: Kim Christensen and Mark Podolskij: Asymptotic theory of range-based multipower variation

2011-48: Christian M. Dahl, Daniel le Maire and Jakob R. Munch: Wage Dispersion and Decentralization of Wage Bargaining

2011-49: Torben G. Andersen, Oleg Bondarenko and Maria T. Gonzalez-Perez: Coherent Model-Free Implied Volatility: A Corridor Fix for High-Frequency VIX

2011-50: Torben G. Andersen and Oleg Bondarenko: VPIN and the Flash Crash

2011-51: Tim Bollerslev, Daniela Osterrieder, Natalia Sizova and George Tauchen: Risk and Return: Long-Run Relationships, Fractional Cointegration, and Return Predictability

2011-52: Lars Stentoft: What we can learn from pricing 139,879 Individual Stock Options

2011-53: Kim Christensen, Mark Podolskij and Mathias Vetter: On covariation estimation for multivariate continuous Itô semimartingales with noise in non-synchronous observation schemes

2012-01: Matei Demetrescu and Robinson Kruse: The Power of Unit Root Tests Against Nonlinear Local Alternatives