Are linear models really unuseful to describe business cycle data?

Silva Lopes, Artur C. and Florin Zsurkis, Gabriel

ISEG, Universidade de Lisboa

4 May 2017

Are linear models really unuseful to describe business cycle data? *

Artur Silva Lopes †
ISEG, Universidade de Lisboa

Gabriel Florin Zsurkis
ISEG, Universidade de Lisboa

May 1, 2017

Abstract

We use first differenced logged quarterly series for the GDP of 29 countries and the euro area to assess the need to use nonlinear models to describe business cycle dynamic behaviour. Our approach is model (estimation)-free, based on testing only. We aim to maximize power to detect non-linearities and, simultaneously, we purport avoiding the pitfalls of data mining. The evidence we find does not support some descriptions because the presence of significant non-linearities is observed for 2/3 of the countries only. Linear models cannot be simply dismissed as they are frequently useful. Contrarily to common knowledge, nonlinear business cycle variation does not seem to be an universal, undisputable and clearly dominant stylized fact. This finding is particularly surprising for the U.S. case. Some support for nonlinear dynamics for some further countries is obtained indirectly, through unit root tests, but this can hardly be invoked to support nonlinearity in classical business cycles.

Keywords: business cycles; nonlinear time series models; testing.
JEL codes: C22, C51, E32.

*Previous versions of this paper circulated under the title "Revisiting non-linearities in business cycles around the world".

†Corresponding author. Address: ISEG, Rua do Quelhas 6, Gab. 107, 1200–781 Lisboa, Portugal. Email: asl(at)iseg.utl.pt.

1 Introduction

The notion that business cycles are asymmetric, with state- or phase-dependent dynamics, is now widely accepted. Whatever the angle of view or the instrument used to highlight these asymmetries in aggregate fluctuations, they are typically considered as an almost undisputed stylized fact of modern economies And since they are sometimes presented as substantial and pervasive, *univariate linear* models should fail to explain the data.

Two major approaches are available to investigate the possibility to dismiss univariate linear models as adequate tools to characterize data from macroeconomic fluctuations in favour of nonlinear ones: the simple and traditional (non-) linearity tests and the "features approach". The first is the (obvious) recommended initial step when considering fitting a nonlinear model to the data (see, e.g., Granger, 1993, Teräsvirta, 2004). The second is less well known and consists of assessing the ability of linear and nonlinear (univariate) models to provide simulated data that display the same features that are observed in real (usually GDP) data.

Somewhat surprisingly, this approach has supported the need to resort to nonlinear models much less often than expected. Generally, linear models are not clearly dominated by the nonlinear alternatives that are considered. More precisely, there is little evidence that nonlinear models perform much better than linear ones: a) either no benefit is found (Hess and Iwata, 1997); b) or they are not considered sufficiently adequate to mimic asymmetries (Galvão, 2002); c) or their superiority in some aspects is made at the expense of some undesirable, extreme, features (Harding and Pagan, 2002, and Engel, Haugh and Pagan, 2005). Also, sometimes linear models perform much better than expected to reproduce certain basic features of business cycle data (Morley and Piger, 2006[1]).

The simpler approach of detecting non-linearities in the conditional mean of business cycle data through testing has been rarely and unsystematically pursued. Indeed, first, although research on nonlinear modelling has exploded in the last 20 years, the careful analysis of business cycle dynamics has not been a topic attracting much attention[2]. Second, linearity tests are very frequently

[1] Whom, however, consider that nonlinear models are better than linear ones to reproduce some of these features.

[2] Actually, a great effort has been made recently in topics such as unit root testing against nonlinear alternatives or in the estimation of new models. Empirically, much effort has been directed recently towards nonlinear modelling of interest rates, exchange rates and public finances.

used in a rather restricted framework only, i.e., where only a test against a particular specific nonlinear model is envisaged as the alternative, often to illustrate that particular model. Third, moreover, a frequent departing point is precisely that specific nonlinear model and the linearity hypothesis is relegated to a secondary role from the outset. Fourth, on top of this, it is very often the case that these studies focus almost exclusively on U. S. data (which are used to illustrate a new model). Fifth, finally, even some results contesting the conventional nonlinear wisdom remain largely unnoticed, as occurs with those of Psaradakis and Spagnolo (2002), who also apply a battery of tests on U.S. GNP growth rates but are unable to find any evidence for nonlinearity, in contradiction with previous results by Hamilton (1989) and Hansen (1992).

Contrasting with this picture, our aim here is to use the testing approach systematically, employing a battery of tests carefully selected to maximize detection power over a large dataset consisting of business cycle data for 29 countries and a monetary zone (composed of the first seventeen countries of the euro area, EA17). This makes our study related with the ones of Bradley and Jansen (2000) and Singh (2012). In both cases substantial but really neither unanimous nor overwhelming empirical evidence favouring nonlinearity of business cycles dynamics is found. However, also in both cases and particularly in Singh (2012), the transmitted global image is somewhat distorted in favour of nonlinearity because the non-rejections of the linear null are quickly overstepped, the attention focusing on the details of the nonlinear models. Singh (2012) considers the series of quarterly rates of GDP growth for 10 OECD countries but his major concern is much more focussed on estimation (SETAR and STAR models) and forecasting than in the detection of non-linearities. Bradley and Jansen (2000) consider a more varied sample of 26 countries and find strong evidence for nonlinearity using a somewhat wider, less model driven, set of test statistics. However, their work is mainly directed to study the presence of homo/heterogeneity in the characterization of business cycle dynamics. Therefore, it is also much more oriented towards estimation than ours. Convergence with the approaches in these works is confined to the multi-country perspective and to a few common test procedures. In every other respect we diverge from them. Our study is model-free in the sense that we do not intend to estimate any particular nonlinear model. This deliberate purely testing approach allows us avoiding any model dependencies. Insulating the detection from the estimation stage appears to us as essential to provide a neutral or impartial overall picture.

We focus exclusively on the testing perspective and we will not proceed into the estimation of any model. This allows us to avoid any dependencies on the models considered under the different alternative hypotheses. In any case, since the number of alternative possibilities is virtually infinite, selecting the best model for each of the cases would be also an unfeasible task. Our purpose is restricted to gathering evidence on the presence of (univariate) non-linearities in the conditional mean of business cycle series using a testing framework. This purely testing strategy is also useful to control its overall size; we try escaping the pitfalls of data mining, exhaustively searching for rejections that could justify estimating some particular model. We aim at maximizing general power while controlling overall size, at the cost of some power loss against specific but unknown alternatives. Spurious rejections of linearity are therefore avoided. Furthermore, as a secondary purpose, we also dedicate some attention to level data.

Our results suggest that most previous research containing descriptions with a strong flavour of nonlinearity must be viewed with a critical perspective. Our empirical evidence casts serious doubts on the idea that nonlinearity of business cycles, characterized through the differenced logged series of aggregate output, can be considered as a global stylized fact. While it appears to be rather common, its presence does not seem to be so strong and so pervasive as to deserve such a qualification. Even for the U. S., the source of inspiration for much of previous research, the evidence for nonlinearity at the short- and medium-term frequencies is rather weak. Only indirect inference supports the presence of some nonlinear dynamics, and this is found in fluctuations around a linear trend, not in first differenced data. That is, to find some nonlinear features for the U.S., we have to adopt an indirect approach together with the output gap perspective of cycles, abandoning both direct inference and the classical view.

This indirect approach originates from our interest in GDP level data, which is an extension of the main focus. Besides the methodological framework previously mentioned and the (robust) enlargement to level data, possibly nonstationary, we introduce a few innovations in the way that some tests are performed and, as far as we are aware, we use unit root tests against nonlinear alternatives for the first time as an instrument to collect evidence on the nature of business cycles.

The remainder of the paper is organized as follows. The following section discusses the data, including any transformation that might be required to analyze business cycles. In section 3 we perform a preliminary data analysis. Unit root testing techniques are extensively used and this allows us to obtain the first evidence for nonlinearity, albeit indirect. Some methodological principles that guided our study are presented here. Section 4 is central to the paper. We present further methodological guidelines, provide a brief description of the statistical procedures, and present the most important empirical evidence. The final section contains a brief discussion and the most important conclusions.

2 Data: transformation and sources

As argued forcefully by Harding and Pagan (2002), when one wishes to follow the classical NBER tradition, studying the characteristics of business cycles according to the "alternating-phases definition" (Morley and Piger, 2012), one must utilize $\Delta y_t = \Delta \log(Y_t)$, where Y_t denotes real (quarterly) GDP, i.e., one must focus on the (approximate, quarterly) real growth rate of GDP. As Harding and Pagan (2002) emphasize, it is the behaviour of Δy_t that determines the nature of the business cycle, even when this is viewed according to the classical perspective, as referring to the cycles in the *level* of Y_t (or $\log(Y_t)$). Moreover, at this stage, this transformation should not be viewed as the application of a detrending filter, as a way to obtain deviations to some trend, as in the growth cycle or "output-gap" (Morley and Piger, 2012) definition. Also at this stage, it should not be viewed as a means to obtain stationarity as well, although it may be useful in this regard, particularly to ensure the validity of tests for linearity. Instead, this is because the behaviour of the first differenced series is crucial to characterize the business cycle in terms of the *level* of aggregate activity. For instance, classical cycles are defined by the turning points in the level series and this definition, this dating, is done with the sign of the growth rate of the series, a function of its first difference.

This is also the view that we adopt here joining, *inter alia*, Beaudry and Koop (1993), Bradley and Jansen (1997, 2000), Clements and Krolzig (2003), Crowley, Garcia and Quah (2013), Kose, Otrok and Prasad (2008) and Singh (2012). That is, regardless of the requirement to use stationary data, our main (but not exclusive) focus of attention will be the series Δy_t as defined above. This is also the case because our data are seasonally adjusted. Otherwise, it is not unusual to replace first differencing with seasonal differencing (see e.g., Teräsvirta and Anderson, 1992).

The use of this business cycle representative is not immune to criticism, however. First, the series Δy_t usually contains a larger component of the

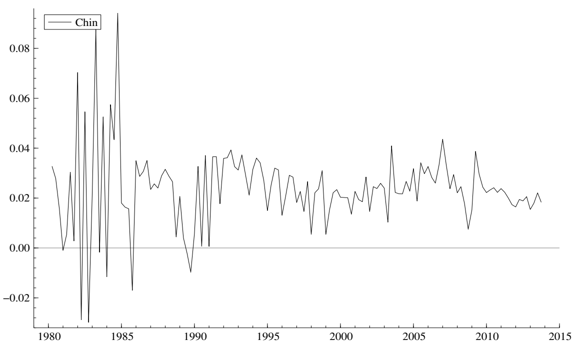
Table 1 – Countries, sample periods and data sources

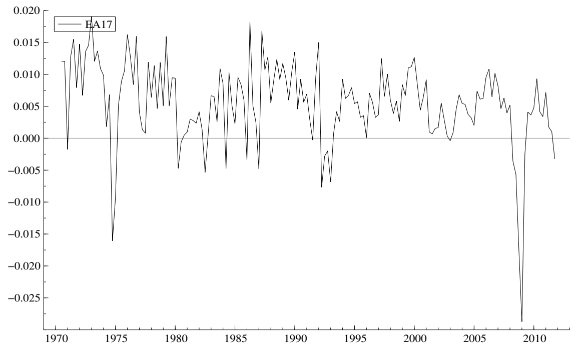
country	sample	source	country	sample	source
Argentina	1980:1–2013:3	Datastream	Italy	1980:1–2013:4	Datastream
Australia	1959:3–2013:4	OECD	Japan	1980:1–2013:4	Datastream
Austria	1980:1–2013:4	Datastream	Mexico	1980:1–2013:4	Datastream
Belgium	1980:1–2013:4	Datastream	Netherlands	1977:1–2013:4	Datastream
Brazil	1980:1–2013:4	Datastream	Norway	1978:1–2013:4	OECD
Canada	1980:1–2013:4	Datastream	Philippinnes	1980:1–2013:4	Datastream
Chile	1980:1–2013:4	Datastream	Portugal	1978:1–2013:4	Ban. de Port.
China	1980:1–2013:4	Datastream	South Africa	1960:1–2013:4	OECD
Denmark	1980:1–2013:4	Datastream	South Korea	1970:1–2013:3	OECD
Finland	1980:1–2013:4	Datastream	Spain	1980:1–2013:4	Datastream
France	1978:1–2013:4	OECD	Switzerland	1980:1–2013:4	Datastream
Germany	1980:1–2013:4	Datastream	Taiwan	1980:1–2013:4	Datastream
Greece	1980:1–2013:4	Datastream	U. K.	1955:1–2013:4	OECD
India	1980:1–2013:4	Datastream	U. S.	1947:1–2013:4	OECD
Ireland	1980:1–2013:4	Datastream	EA17	1970:1–2011:4	EABCN

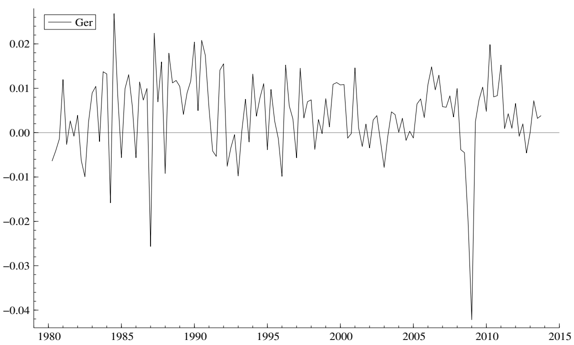
high frequency fluctuations than in some approaches of business cycle analysis, which characterize this phenomenon as corresponding to periods between 2 and 8 years. We do not share this (so long) view of business cycles and we believe that the quarterly rates of GDP growth are the best representative of short- and medium-term aggregate fluctuations, as they are perceived by common economic agents and observers, and they are also the most important indicators to follow in conjunctural analysis. Second, some authors prefer analyzing the growth rates of industrial production because they are more timely and contain more cyclical variation than those of GDP. Actually, some empirical evidence (see Granger and Teräsvirta, 1993) and the simulation study in Granger and Lee (1999) suggest that it is easier to find non-linearities in industrial production than in GDP, particularly when the first variable is observed monthly and because it represents a less aggregated entity from the cross-sectional perspective as well. However, in many of the countries analyzed industrial production currently represents only a minor proportion of economic activity and GDP growth is a much better indicator of aggregate fluctuations.

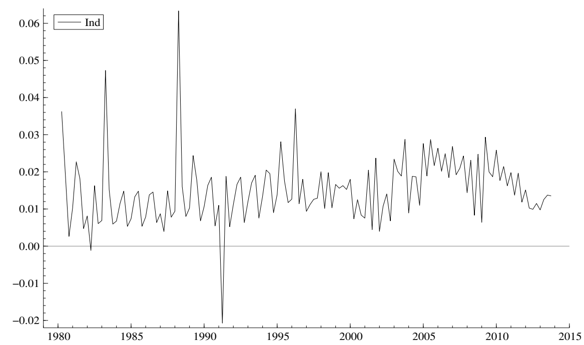
Table 1 contains the list of the 29 countries and the monetary zone (euro area–17, EA17 for short) that we analyze, together with the corresponding sample periods and sources. Our dataset concerns data on Australia, one African country (South Africa), 6 American countries (Argentina, Brazil, Canada, Chile, Mexico and the U.S.), 6 Asian countries (China, India, Japan, Philippines, South Korea and Taiwan), 15 European countries (Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Switzerland and the U.K.) and the EA17.

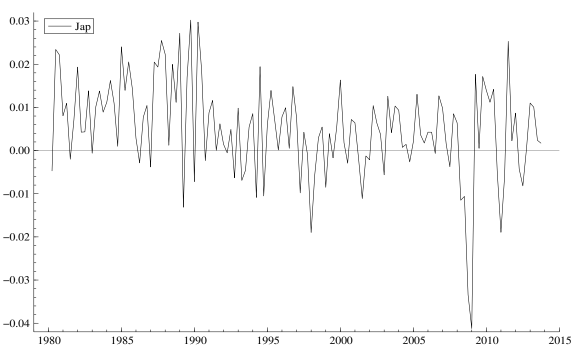
The most frequent sample period begins in 1980:1 and ends in 2013:4, corresponding to a sample size with $T = 136$ observations, which is reasonable for most purposes but may be considered low for the power of many linearity tests to attain a satisfactory level (see Psaradakis and Spagnolo, 2002). We also present the plots for some of the most important countries and for the EA17 of the (approximate) real growth rates. In some of the cases, the beginning of the current Global Crisis is clearly discernible.

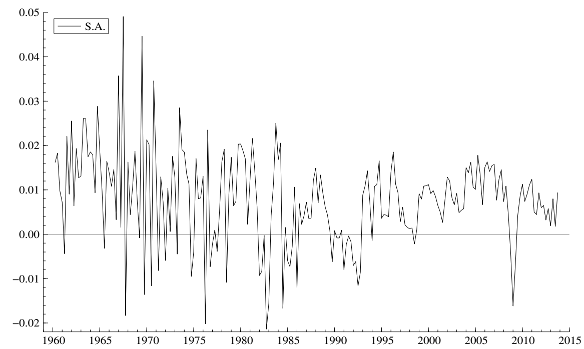












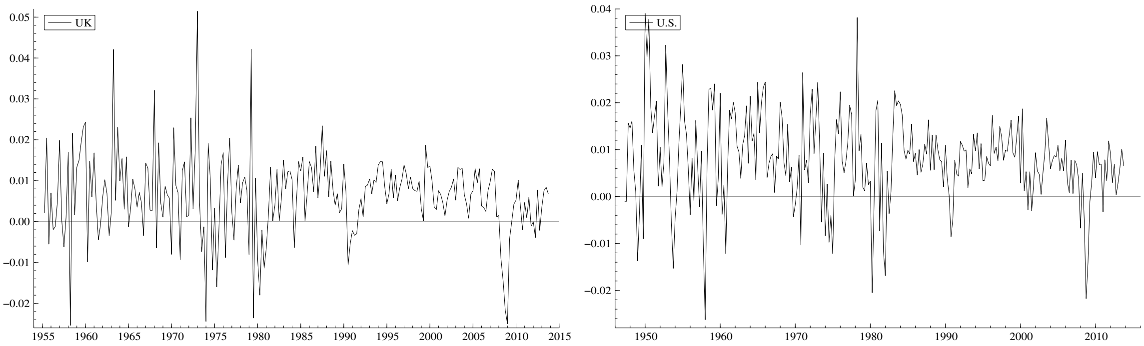


Figure 1. The Δy_t series for some countries.

3 Preliminary data analysis: unit roots

Although our primary interest lies in Δy_t, the first differenced logged series of GDP, the analysis of the non-transformed series, $y_t = \log(\text{GDP}_t)$, with log denoting the natural logarithm, may be also revealing. Regardless of the business cycle definition that one adopts, finding the presence of non-linearities in aggregate macro data is important in its own right, especially when this concerns aggregate output.

Therefore, a preliminary data analysis testing for unit roots is required to avoid spurious inference procedures. In particular, linearity tests usually demand that data series are stationary to be valid. Otherwise the tests may produce spurious evidence against linearity, over-rejecting the true null hypothesis (see Kiliç, 2004). Unit root tests then become useful in another respect: in case the level series is considered trend stationary, unit root tests against nonlinear alternatives become specially relevant from the business cycle perspective because the real object of analysis are the deviations or fluctuations around the trend. In other words, empirical evidence about the level series becomes relevant from the growth cycle or output gap perspective as well. Actually, when the level series is analyzed, it is the properties of deviations from trend that are being investigated.

3.1 Standard unit root tests

In a first stage, we will use standard or conventional, i.e., "linear", unit root tests. We opted to use the popular ADF tests and the more powerful ADF–GLS tests of Elliot, Rothenberg and Stock (1996), employing always as deterministic regressors an intercept and a (linear) trend term. In both cases, trying to get robust results, we choose the number of augmenting lags using three distinct methods: the AIC and the modified AIC (MAIC) criteria of Ng and Perron (2001), and the general-to-specific (GTS) t-sig method; in all the cases we have set at 12 the maximum number of augmenting lags.

We relegate the presentation of the results for these tests to a separate appendix. Table A.1 contains the results of ADF test statistics. As expected, the vast majority of results does not allow the rejection of the unit root null hypothesis, providing empirical support for the I(1) hypothesis. The only cases diverging from this almost unanimity are those of Australia, China and Mexico, and in this case only with one test statistic. We postpone a more detailed analysis until the presentation of ADF-GLS test statistics.

These are presented in table A.2, also for the same 3 different strategies of lag truncation selection. Since ADF-GLS tests are known to be better than ADF tests, particularly in terms of power, in case of dissonant results we tend to give more credit to the results of the former. Therefore, we consider that for Australia the most appropriate order of integration is I(1). Also, the minor evidence of rejection of the unit root null for Brazil, Mexico and South Africa, tends to vanish with ADF-GLS test statistics.

A rather different case is that of China, where the issue initially seems to lie in the number of augmenting lags. Actually, using the MAIC in this case seems to produce an underparametrized test regression because some lags appear to be really needed to capture the dynamics of the series. Hence, it appears that there is no supporting evidence for the unit root. A closer inspection, however, reveals that the main problem is one of heteroskedasticity, already apparent in figure 1. As the decrease in variance occurs relatively early in the sample, this can be a case of a spurious rejection of the unit root. Further, as noted by Kim, Leybourne and Newbold (2002), a simple FGLS transformation will not solve the problem in this case because it introduces a problem of a break in the level and the trend of the series. Therefore, the adequate framework (for the transformed series) is that of Perron's tests and, in particular, his "model C" case (change in level and trend). After having estimated that the break in variance occurs around 1985:1, and using the test regression of section 5 of Kim, Leybourne and Newbold (2002), we get a test statistic of -2.885 which is clearly insufficient to reject the unit root, even at the 10% level. Therefore, it seems that also in the case of China the I(1) hypothesis appears to be better than the I(0).

3.2 Tests allowing for breaks

The case of Switzerland, with two 5% rejections and one at 10% with ADF-GLS statistics, raises the issue of a possible contamination of unit root test statistics by outliers and structural breaks. In fact, the graph of the Δy_t series is very similar to that of Germany (see figure 1), for instance, with what appears to be an outlier in 2008-2009. Correspondingly, as in the case of many other countries, the (log)level series, y_t, presents a marked and abrupt change in level, a *crash*, at the same time, certainly associated with the emergence of the Global Crisis in Europe. Therefore, it is convenient to analyze to what extent the previous results are robust to testing strategies that allow for the presence of such data problems.

While some authors view the detection and accommodation of structural breaks in unit root tests as a component of a non-linear analysis, we are skeptical about this view. This is because although their presence always represents a change in regime, and even when such breaks are specified as non-linear, this change is only a function of the time variable. In other words, the transition variable in the nonlinear function, the variable that commands the change in regime, is simply the time variable, a statistically convenient variable but one that is deprived of real economic meaning. Identifying such a change is useful but does not add much to economic knowledge. In particular, neither the size nor the sign of shocks are explicitly considered as potential triggers of the change. Also, neither the level of the series nor the position in the cycle are allowed to perform any role. If a change in regime occurred, the cause must be identified and the transition variable must be specified accordingly. Simply specifying the nonlinear function as driven by the time variable does not seem to be sufficient for a real nonlinear analysis; such a specification appears to be poor. Therefore, we do not attribute a particular significance to the tests of this subsection, even when they change qualitatively the previous evidence. This same reasoning justifies that we do not use the family of unit root tests initiated by Leybourne, Newbold and Vougas (1998).

Also, structural breaks in unit root tests are associated with large shocks that change the long-run behaviour of the series, their trend behaviour. As we are mainly interested in non-linearities in business cycle type fluctuations, we tend to view such low frequency phenomena as marginally interesting only, similar to a nuisance in a preliminary procedure. For our purposes, the main objective of this analysis is to guard against non-valid inferences due to non-stationarities.

Table 2 – Unit root test (LLS $\tau(\widehat{\eta}, p)$) statistics allowing a break in level

country	statistic	date	lagn	country	statistic	date	lagn
Argentina	-1.331	1990:1	10	Italy	-1.212	1991:1	4
Australia	-2.503	1997:2	2	Japan	-0.370	2009:1	0
Austria	-1.117	1988:2	8	Mexico	-2.220	1995:1	2
Belgium	-1.241	1986:4	5	Netherlands	-1.907	1988:1	5
Brazil	-2.553	1991:2	8	Norway	-1.205	1998:3	1
Canada	-2.008	1987:1	1	Phillippinnes	-1.162	1983:3	1
Chile	-2.391	1983:4	7	Portugal	-1.254	1992:1	3
China	-3.985***	1984:4	5	South Africa	-1.085	1969:3	4
Denmark	-2.250	2008:4	1	South Korea	-1.416	1998:1	5
Finland	-3.302**	2009:1	3	Spain	-2.035	1990:4	3
France	-1.989	2009:1	2	Switzerland	-2.187	2008:4	9
Germany	-1.712	2009:1	0	Taiwan	-0.840	2008:4	0
Greece	-1.661	1990:4	5	U.K.	-2.008	2009:1	1
India	-1.247	1988:2	0	U.S.	-2.708	2008:4	3
Ireland	-1.496	2007:4	8	EA17	-1.601	2009:1	1

"Date" denotes the estimated break date and "lagn" denotes the number of augmenting lags. "***", "**", and "*" represent rejections of the (unit root) null hypothesis at the 1%, 5% and 10%, respectively. The asymptotic critical values are -3.55, -3.03 and -2.76, respectively.

To investigate the presence of a unit root while allowing for a break in the level of the series, we resorted to the very flexible approach by Lanne, Lütkepohl and Saikkonen (2003, LLS), also summarized in Lütkepohl (2004), which introduces a rational shift function in the trend, i.e., a rational function in the lag operator applied on a common step dummy variable that is added to the deterministic regressors, permitting general nonlinear changes in level. Besides its flexibility, covering both the classical additive outlier and innovational outlier cases, the approach possesses a further advantage: it is robust to errors in the estimation of the break date.

Further, the testing strategy does not require any previous information concerning this date. In the context of this investigation, with such a wide variety of countries, this is a very useful feature. Actually, this date is estimated in a first step, and the results presented in table 2 illustrate this variety. Although an estimated break date lying in 2007-2009 is the most frequent, as expected, it is far from representing the majority of cases (11 in 30).

A surprising general outcome is the robustness of the previous results. Allowing for a break in level changes the decision only for Finland, now classified as trend stationary. For all the other countries the I(1) hypothesis still remains supported[3]. Therefore, even if one embraces the view of interpreting changes in decision associated with the accommodation of breaks as signalling the presence of non-linearities, the evidence that is found for such presence is extremely feeble.

3.3 Tests against nonlinear alternatives

It is well known that standard unit root tests are designed against linear alternatives and may lack power when a process is nonlinear and globally stationary (see e.g. Pippenger and Goering, 1993). But unit root tests against nonlinear alternatives also possess a more direct relevance for our purposes. As previously mentioned, they allow assessing the behaviour of the deviations from the (traditional, linear) trend, and hence they are useful from the growth cycle or output gap perspective of business cycles.

The two-stage procedure adopted in all these tests — the first stage consisting of the trend removal — makes this point very clear. This is even more clear when a standard test does not allow rejecting the unit root null but a nonlinear one does: since the assumption of a linear trend is common to both procedures, it is the fluctuations around the trend that must be responsible for the rejection; they must contain some nonlinear behaviour that confounds the standard tests. Obviously, some evidence for non-linearity in the level series must be also recorded in this case.

The restriction previously mentioned of neglecting tests against nonlinear structural break models — as those of Leyboune et al. (1998) — allows us to dismiss several tests but still leaves a plethora of available statistics to consider. To further restrain this set we resorted mainly to two criteria:

a) popularity, simplicity and availability of asymptotic critical values for the test statistics, and

b) a good power performance behaviour, documented in Monte Carlo studies, even against alternatives that are different from those that originated the test statistic.

The adoption of these criteria allows us to neglect the tests designed against threshold autoregressive (TAR) models, considering only those tests with smooth transition autoregressive (STAR) models as alternatives. Usually these are

[3] This is the case for China as well, because this test is not robust to the change in the variance of innovations previously detected

simpler and, according to available simulation studies, have reasonable power even against TAR processes, a feature that generally does not occur in reverse. In other words, simulation studies suggest that tests against STAR models encompass tests against TAR alternatives but the reverse does not seem to hold (see e.g. Sollis, 2011). This test selection is also supported by the simulation study of Choi and Moh (2007), whose main conclusion is that the particular type of non-linearity is somewhat irrelevant to explain the power behaviour of these unit root tests; what really matters for their performance is the distance between the unit root process and the alternative model.

Two examples of test statistics that we will not use are those of Enders and Granger (1998) and Bec, Guay and Guerre (2008): though simple, the first one is clearly disappointing in terms of power; considering a set of four tests, the second is the most powerful against TAR alternatives, but in this case the test by Kiliç (2011) is also powerful and besides relatively simpler it is also more powerful than the Bec *et al.* (2008) test for several other DGPs (see Kiliç, 2011). On the contrary, we will use the Kapetanios, Shin and Snell (2003, KSS hereafter) test against the exponential STAR (ESTAR) alternative due to its popularity and simplicity; its power performance in simulation studies is generally poor but surprisingly, according to Choi and Moh (2007), the test is one of the most powerful against the equilibrium-TAR (EQ-TAR) alternative.

Previous to presenting and analyzing the results of these tests, two observations are worth mentioning. First, according to the available simulation studies, to attain a satisfactory power performance the tests usually require samples with at least 150 to 200 observations. Second, in spite of motivating criticisms, DF (OLS) tests are frequently the most powerful to detect stationarity of nonlinear alternatives, particularly for small sample sizes.

In table 3 we present the evidence produced by our preferred tests in conjunction with the AIC method to determine the lag length. In a separate appendix we present a brief description of the test statistics: the t_{NL} statistic of KSS, the Sollis (2009) $F_{AE,t}$ and Shintani's (2013) $\inf -t_{E,\tau}$ tests, and Kiliç's t_{ESTAR} statistic. Although the results obtained with the GTS and the MAIC methods are also available, they do not differ much from those presented here, and the AIC appears to produce the most sensible choices[4]. Note also that although the test regressions are different, the estimated lag truncation parameter rarely changes.

Seen with the light of the previous observations, when the transition vari-

[4]While the GTS t-sig method appears to show a slight tendency to overparametrize in relation to the AIC, the MAIC frequently appears to produce lag lengths that are too short.

able is $\hat{x}_{t-1}$ — which represents the lagged OLS residual of the regression of $\log(GDP_t)$ on a linear trend, not the lagged level of $\log(GDP)$ (see subsection 6.3 in the appendix)—, the results are not surprising:

a) the number of new rejections of the unit root null is very low, i.e., the wide support to the unit root hypothesis gathered through standard tests does not appear to be attributable to the presence of non-linearities;

b) in particular, the rejection evidence for Brazil, Mexico and South Africa is relatively weak and, in the case of China, it is likely that the much stronger rejections are due to the seemingly presence of heteroskedasticity;

c) the only real important new information provided by these tests appears to be the relatively strong rejection evidence for Germany and, to a lesser extent, for Australia.

A rather different picture emerges from the only test that uses $\Delta \hat{x}_{t-1}$ as transition variable, the t_{ESTAR} test of Kiliç (2011):

a) strong rejections now appear for Australia (shared with t_{NL}, however), Chile and Finland;

b) standard, 5% rejection evidence is now found for Austria, Belgium, Brazil, Canada, Ireland, Spain, the U. S. and the EA17;

c) weak rejection evidence, at the 10% level, is obtained for Mexico and South Africa;

d) on the contrary, the previously found evidence for stationary behaviour of the GDP of China and Germany now disappears.

In table A.3 of the appendix we present further evidence on unit root tests against nonlinear alternatives using additional test statistics. Possibly due to the fact that all these tests use a version of $\hat{x}_{t-1}$, not its first difference, as the transition variable, additional rejection information is almost nonexistent. It is worth noting, however, that somewhat weak evidence on stationarity is now provided by the $F_{AE,t}^{GLS}$ test for Argentina and for Greece, and that this same test rejects the unit root null for the U.S. at the 5% level.

Table 3 – Unit root tests against nonlinear alternatives

	t_{NL} (lagn)	$F_{AE,t}$ (lagn)	$\inf -t_{E,\tau}$ (lagn)	t_{ESTAR} (lagn)
Argentina	−2.494 (12)	3.227 (12)	−2.494 (12)	−1.428 (11)
Australia	−3.975 (9)***	7.995 (9)**	−3.587 (8)*	−3.336 (8)***
Austria	−0.351 (9)	1.552 (9)	−1.073 (9)	−2.623 (9)**
Belgium	−0.177 (4)	1.901 (5)	−1.693 (5)	−2.706 (5)**
Brazil	−3.408 (8)**	5.913 (8)*	−3.834 (8)**	−2.991 (8)**
Canada	−2.977 (1)	4.421 (1)	−2.977 (1)	−2.557 (1)**
Chile	−2.640 (12)	4.728 (12)	−2.640 (12)	−3.635 (12)***
China	−5.562 (10)***	15.395 (10)***	−6.392 (12)***	−1.970 (10)
Denmark	0.339 (1)	1.179 (1)	−1.108 (3)	−1.360 (3)
Finland	−2.319 (3)	3.136 (3)	−2.709 (3)	−4.715 (3)***
France	−0.664 (2)	2.339 (2)	−1.631 (2)	−2.114 (2)
Germany	−3.902 (4)**	8.885 (4)***	−2.233 (0)	−1.710 (1)
Greece	−0.481 (5)	0.506 (5)	−2.474 (9)	−1.295 (5)
India	−1.184 (0)	1.551 (0)	−1.184 (0)	−0.617 (0)
Ireland	−0.956 (12)	1.803 (12)	−1.704 (12)	−2.828 (12)**
Italy	0.305 (4)	1.176 (4)	−0.573 (4)	−1.792 (4)
Japan	−1.807 (3)	1.625 (3)	−1.808 (3)	−1.195 (3)
Mexico	−2.953 (3)	6.190 (3)*	−3.043 (3)	−2.333 (3)*
Netherlands	−0.100 (0)	0.920 (0)	−0.421 (0)	−0.750 (1)
Norway	−0.786 (5)	1.202 (5)	−1.159 (5)	−1.286 (5)
Philippines	−0.027 (0)	0.369 (0)	−0.598 (0)	−0.644 (0)
Portugal	−0.684 (4)	0.639 (4)	−0.747 (4)	−0.450 (3)
South Africa	−3.133 (7)*	5.082 (7)	−3.200 (7)*	−2.303 (7)*
South Korea	−0.025 (1)	0.932 (1)	−0.705 (2)	−2.118 (2)
Spain	−0.215 (8)	0.769 (8)	−1.433 (8)	−3.115 (8)**
Switzerland	−2.397 (1)	4.410 (2)	−2.928 (2)	−2.549 (1)
Taiwan	−1.051 (1)	0.549 (1)	−1.051 (1)	−1.152 (0)
U.K.	−1.994 (3)	5.235 (3)	−2.298 (3)	−1.811 (3)
U.S.	−1.691 (2)	3.343 (2)	−2.225 (2)	−2.671 (3)**
EA17	−0.816 (1)	2.499 (1)	−1.966 (2)	−2.886 (5)**

In all the cases the lag length ("lagn") was estimated using the AIC statistic. For the KSS test the asymptotic critical values are −3.13, −3.40, and −3.93 at the 10%, 5% and 1% levels, respectively. For the $F_{AE,t}$ statistic of Sollis (2009) they are 5.372, 6.292 and 8.344. For the Shintani test statistic, $\inf -t_{E,\tau}$, they are −3.35, −3.64 and −4.18, and for the Kiliç t_{ESTAR} test they are −2.23, −2.57 and −3.19.

4 Testing for (non-)linearity

In this section we first present a brief description of the tests employed to detect the presence of non-linearities, and subsequently we focus on the empirical evidence.

4.1 The selected tests: a brief description

Broadly speaking, the available tests can be classified in two groups: a) tests against an unspecified alternative, which are designed without a particular nonlinear alternative model in mind, and b), tests designed to distinguish linearity from a specific nonlinear model. *A priori*, we prefer the first class of tests because the power of those of the second group may be low in many circumstances. However, since there are not many general tests, we will resort to tests from both classes. Moreover, we have selected them using the same criteria that in subsection 3.3. The only exception is the CDR ("current depth of recession") test, which is not very popular and whose power properties are not well known; instead, its relevance here stems from the fact that it is designed specifically for business cycle data.

4.1.1 General tests

Simulation studies such as those of Lee, White and Granger (1993, LWG) and Psaradakis and Spagnolo (2002) are useful to select tests because a wide spectra of alternatives are considered. LWG show that White tests are generally powerful, and that the RESET test outperforms most of the other popular general tests for many of the alternatives. Psadarakis and Spagnolo (2002) concluded that these two tests also have a good power performance against MSAR (Markov-switching AR) models. While the White tests are generally powerful for the seven DGPs used in the experiments, RESET tests are more powerful in the presence of switching autoregressive dynamics. In this paper, we will use a version of each of these two tests.

RESET is very well-known and the version that we use employs the squared and cubed terms. Although not so popular, the White test that we chose has a long tradition; it is called the White dynamic information matrix test, it appears in the study by LWG under the heading of "White3" and it seems to be even more powerful than the White test based on artificial neural network (ANN) models. As far as we know, it was first proposed in White (1982).

4.1.2 A test for threshold nonlinearity

Although specifically designed against self-exciting TAR (SETAR) models, the Tsay (1989) test is sufficiently general to deserve special attention. It makes use of arranged autoregressions and recursive estimation, and although it has been rarely employed, its characteristics make it attractive as a general specification test.

For a SETAR(p) model for y_t, the observations can be arranged in ascending order of the threshold variable, y_{t-d}, as $\{y_{\pi 1}, y_{\pi 2}, \ldots, y_{T-d-h+1}\}$, $h = \max\{1, p+1-d\}$ and π_i denoting the index of the ith smallest observation. An arranged autoregression can be written as

$$y_{\pi_i+d} = \phi_0 + \sum_{v=1}^{p} \phi_v \, y_{\pi_i+d-v} + a_{\pi_i+d},$$

a_{π_i+d} denoting the error term. Recursive regressions can be performed beginning with b observations, making available $T-d-b-h+1$ one-step predictive residuals, $\hat{a}_{\pi_i+d}$. Then, threshold nonlinearity is tested by verifying the orthogonality property between the predictive residuals and the regressors, $\{y_{\pi_i+d-v} \,|\, v = 1, \ldots, p\}$, because it will not hold in case the true model is a nonlinear SETAR. Hence, the global F statistic in the regression

$$\hat{e}_{\pi_i+d} = \omega_0 + \sum_{v=1}^{p} \omega_v \, y_{\pi_i+d-v} + \epsilon_{\pi+d}, \quad i = b+1, \ldots, T-d-h+1,$$

where $\hat{e}_{\pi_i+d}$ denote the standardized predictive residuals, is used to test the orthogonality conditions. Under the null of linearity, it follows approximately an F distribution with $p+1$ and $T-d-b-p-h$ degrees of freedom. As usual with other tests, we performed this test only with $d = 1$, i.e., with y_{t-1} (Δy_{t-1} in our case) as transition variable.

4.1.3 The LM-STAR test

Although it is not strictly model-free, the LM-STAR test is usually considered one of the best tests to detect non-linearities, one of the most powerful for a wide range of alternatives. Consider the STAR(p) model

$$y_t = \boldsymbol{\phi}' \mathbf{w}_t + \boldsymbol{\theta}' \mathbf{w}_t \, G(y_{t-d}, \gamma, c) + u_t, \quad \gamma > 0, \quad t = 1, 2, \ldots, T,$$

where $\boldsymbol{\phi} = (\phi_0, \phi_1, \ldots, \phi_p)'$, $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_p)'$, $G(y_{t-d}, \gamma, c)$ is the transition function, c is the switch parameter and y_{t-d} is the transition variable. Its two most popular versions are the logistic, LSTAR, and the exponential, ESTAR, when the transition function is the logistic and the exponential function, respectively. The first order LSTAR model is capable of characterizing asymmetric behavior, i.e., different dynamics for small and large values of y_{t-d}, and it is therefore considered particularly adequate to describe business cycle data.

When $\gamma = 0$ the model becomes a linear AR(p) and hence testing linearity can be expressed as testing $H_0 : \gamma = 0$ *vs.* $H_1 : \gamma > 0$; this also makes the LM or score principle particularly attractive. The problem, the Davies problem — once again — is that the model becomes unidentified under H_0; in particular, with this null hypothesis the parameters c and $\boldsymbol{\theta}$ become unidentified. A solution to circumvent this problem was proposed by Luukkonen, Saikkonen and Teräsvirta (1988); it consists of replacing the transition function by a suitable Taylor series approximation around $\gamma = 0$. In the most general case, that of the LSTAR model, a third order expansion is used to produce the auxiliary test regression

$$y_t = \beta_0' \mathbf{w}_t + \sum_{j=1}^{3} \boldsymbol{\beta}_j' \, \tilde{\mathbf{w}}_t \, y_{t-d}^j + e_t,$$

where $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{jp})$, see e.g. Teräsvirta (1994). Testing linearity now entails testing $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \boldsymbol{\beta}_3 = \mathbf{0}$ and the LM statistic is, as usual, asymptotically $\chi^2_{(3p)}$ under H_0. However, as this test can be severely oversized in small samples, Teräsvirta (2004) recommends using instead the corresponding F-statistic approximation. It is worth noting that a rejection may also imply the presence of ESTAR-type nonlinearity because a first order expansion of the transition function around $\gamma = 0$ in this case produces the previous equation with $\boldsymbol{\beta}_3 = \mathbf{0}$.

Since our purpose is only to detect non-linearities, not to build a STAR model, and since we also employ several other test statistics, trying to prevent serious over-rejection problems of the null of linearity we perform this test only with $d = 1$.

4.1.4 The CDR test

Beaudry and Koop (1993) introduced a model specifically designed to capture asymmetric persistence in GDP according to the business cycle phase. Nonlinearity is generated augmenting the autoregressive representation of the process with the inclusion of the current depth of recession (CDR) variable,

$$CDR_t = \max\{y_{t-j}\}_{j\geq 0} - y_t,$$

where $y_t = \log(GDP_t)$, in order to examine whether the dynamics of the process in recessions differ from those in expansions. This variable measures the distance between the level of current output and its previous peak, how deep the recession is, and is nonzero when the economy is in recession or in the recovery phase. The CDR model is thus defined by

$$\Delta y_t = \phi_0 + \sum_{i=1}^{p} \phi_i \, \Delta y_{t-i} + \sum_{j=1}^{q} \delta_j \, CDR_{t-j} + \epsilon_t, \tag{1}$$

To facilitate the interpretation, consider the simplest case with $q = 1$. The model then contains two regimes with endogenous switching. The "floor regime", when CDR_{t-1} is nonzero, is activated when output falls and remains activated until it grows back to its pre-recession level. Notice however that, unlike threshold models, the transition variable is not fixed. If, as expected, $\delta_1 > 0$, output growth is greater when CDR_{t-1} is nonzero, and the economy tends to recover quickly from a recession. This is the case where the effects of negative shocks tend to be mainly temporary, less persistent than the effects of positive shocks. Beaudry and Koop (1993) and Bradley and Jansen (2000) found evidence for the presence of this "bounce-back" effect for the U.S. real GNP and for the real GDP of some countries (U.S. included), respectively.

As there are no nuisance parameters under the null hypothesis of AR linearity, the CDR model is estimated and the linearity test consists simply on testing the joint significance of the CDR terms by means of the usual F-statistic.

4.1.5 A test for linearity robust to stationarity issues

We now abandon temporarily our main purpose and focus our attention on a test designed to detect nonlinear behaviour in y_t, the level series, which is robust to stationarity issues. Harvey, Leybourne and Xiao (2008, HLX) designed such a test against STAR-type nonlinearity. It allows investigating the presence of nonlinear dynamics either in the business cycle component or in the deviations from trend without requiring any knowledge on the long-run properties of the (level) series. It consists of a data-dependent weighted average of the Wald test statistics from two linearity tests. While one of them assumes that the series is I(0), the other considers that it must be differenced. A function of a unit root test statistic and of a nonparametric stationary statistic, taking values between zero and one, is then used to assign a weight to each Wald statistic. The weight assigned to the nonstationary (stationary) statistic tends to one (zero) if there is strong evidence for a unit root in the series, and tends to zero (one) when the series appears to be stationary.

Considering a nonlinear AR(1) model for an I(0) time series y_t and assuming that additional dynamics enter linearly, one obtains the auxiliary regression

$$y_t = \beta_0 + b\,t + \sum_{i=1}^{3} \beta_i\, y_{t-1}^i + \sum_{j=1}^{p} \beta_{4j}\, \Delta y_{t-j} + \varepsilon_t.$$

Under the null of linearity, $H_{0,0} : \beta_2 = \beta_3 = 0$, the Wald standard statistic W_0 is asymptotically $\chi^2_{(2)}$. Now consider the corresponding regression assuming that y_t is I(1):

$$\Delta\, y_t = \lambda_0 + \sum_{i=1}^{3} \lambda_i \Delta y_{t-1}^i + \sum_{j=1}^{p} \lambda_{4j}\, \Delta y_{t-j} + \varepsilon_t.$$

Standard large sample theory assures that the Wald statistic W_1 is asymptotically distributed as $\chi^2_{(2)}$ under $H_{0,1} : \lambda_2 = \lambda_3 = 0$. The HLX statistic asymptotically selects W_0 when the data are stationary and W_1 when the series contains a unit root, using a weighted average,

$$W_\lambda = (1 - \lambda)\, W_0 + \lambda\, W_1,$$

with $\lambda = \exp\left[-g(U/S)^2\right]$, where g is some finite positive constant — HLX recommend $g = 0.1$ —, U is the usual ADF test statistic, and the S statistic is given by

$$S = \frac{T^{-1/2} \sum_{t=k+1}^{T} \tilde{y}_t\, \tilde{y}_{t-k}}{\widehat{\omega}\{\tilde{y}_t \tilde{y}_{t-k}\}},$$

where $\tilde{y}_t$ denote the OLS residuals of a regression on a linear trend, $\tilde{y}_t = y_t - \hat{a} - \hat{b}\,t$, and $\widehat{\omega}^2\{a_{t,k}\}$ is the Bartlett kernel-based estimator of the long-run variance of a sequence of variables $a_{1,k}, \ldots, a_{T,k}$ [5]. Under the null of either I(0) or I(1) linearity, W_λ selects the efficient, adequate test in the limit and it is

[5] Recall that it is defined by $\widehat{\omega}^2\{a_{t,k}\} = \hat{\gamma}_0\{a_{t,k}\} + 2\sum_{j=1}^{l}(1 - \frac{j}{l})\hat{\gamma}_j\{a_{t,k}\}$, $\hat{\gamma}_j\{a_{t,k}\} = T^{-1}\sum_{t=j+k+1}^{T} a_{t,k}a_{t-j,k}$, with $a_{t,k} = \tilde{y}_y\tilde{y}_{t-k}$, $k = (2T)^{1/2}$ and $l = 12(T/100)^{1/4}$ rounded to the nearest integer.

asymptotically $\chi^2_{(2)}$.

4.2 Empirical evidence

We first concentrate on our main purpose, the analysis of business cycle data, and then proceed to the detection of nonlinear behaviour in level data.

4.2.1 Non-linearities in business cycles

The first five tests just described were used to detect the presence of nonlinearities in business cycle data, i.e., in the Δy_t series (regardless of the order of integration defined for y_t). A linear autoregressive model had to be selected and estimated first for each of the 30 cases. A defensive strategy was followed to select the autoregressive order for each case: as the SIC criterion could lead to overly parsimonious models, possibly with insufficient dynamics, potentially leading to spurious evidence for nonlinearity, we adopted instead the AIC criterion, considering a maximum lag length (p_{MAX}) of 12 lags for all the five tests. But the nature of our testing strategy, designed to control overall size as strictly as possible, is also apparent in the following features:

a) we have used a single version for each test statistic,

b) and we have selected it *a priori* on the basis of a plausibility criterion only.

While this appears to be a common practice for the RESET test, it is rather unusual for the LM-STAR or the CDR tests, where a search for the delay parameters producing the most favorable outcomes for nonlinearity is typically carried out. Actually, as far as we know, ours is the only empirical study where these two tests were performed along these lines. For the LM-STAR and the Tsay tests we have fixed $d = 1$, and for the CDR test we considered $q = 2$ only and tested the joint significance of the two terms.

In table 4 we present a qualitative synthesis of the empirical results for all the tests and in table 5 we present the numerical results for the growth data; these assume the p-value form, both to save space and to allow a simple and clear reading [6].

[6] We adopt the view of the p-value as representing "*a continuous measure of the compatibility between the data and the entire model used to compute it, ranging from 0 to complete incompatibility to 1 for perfect compatibility, and in this sense (it) may be viewed as measuring the fit of the model to the data*", Greenland *et al.*, 2016, p. 3.

Table 4 – A synthesis of results for the linearity tests

	RESET	White	Tsay	LM-STAR	CDR	HLX
Argentina	—	5	—	1	5	1
Australia	—	1	1	1	10	—
Austria	10	1	5	1	—	1
Belgium	—	5	5	1	—	1
Brazil	1	1	1	1	—	5
Canada	—	—	—	—	—	5
Chile	1	10	5	1	—	10
China	—	1	10	1	—	10
Denmark	—	5	—	—	—	—
Finland	1	1	1	5	—	5
France	—	5	—	1	—	1
Germany	—	10	—	—	—	—
Greece	—	5	5	—	—	5
India	—	—	—	—	—	—
Ireland	—	1	—	1	—	—
Italy	5	1	—	1	—	1
Japan	—	10	—	—	—	—
Mexico	5	5	1	5	—	—
Netherlands	10	5	—	5	—	5
Norway	—	—	—	—	—	10
Philippines	1	10	5	1	—	1
Portugal	—	—	5	—	—	—
South Africa	10	1	5	1	1	1
South Korea	—	—	—	—	—	—
Spain	5	1	1	1	1	1
Switzerland	—	—	—	—	—	—
Taiwan	10	—	—	—	5	5
U.K.	5	1	1	1	—	—
U.S.	—	—	—	—	—	—
EA17	—	5	1	5	—	1

"1", "5" and "10" mean that the test rejects the null hypothesis at the 1%, 5% and 10%, respectively. A "—" means that the p-value for the test statistic is larger than 0.10.

Another distinctive feature of our conservative approach concerns the interpretation of the test results. With so many and diversified tests, we consider that a union of rejections strategy is not admissible, as it would inflate overall size far above the usual nominal 5%. In other words, a single rejection is deemed insufficient to proclaim nonlinear behaviour, particularly if it occurs at the 10% level only. Instead, we have considered that every country could be classified into one of four groups, according to the number and strength of the rejections of the null of linearity. The first group is formed by those countries whose evidence for nonlinearity is very weak, simply nonexistent or with only one rejection at the 10% level. These are Canada, Germany, India, Japan, Norway, South Korea, Switzerland and the U.S. . Notice that four of the G7 countries are in this group — Canada, Germany, Japan and the U.S. —, whose business cycle dynamics seems to dispense completely a description based on a nonlinear model.

Then, we considered a very small group of (small) countries which present stronger evidence against linearity, but only marginally, i.e., only one of the 5 tests rejects it at the usual 5% level: Denmark, Portugal and Taiwan, but in this last case there is also a further rejection at the 10% level. For a third group, the number of rejections of the linearity null is two or three at 5% or lower, suggesting that a nonlinear model is really needed to explain asymmetric behaviour. This group includes Argentina, Belgium, China, France, Greece, Ireland, Italy, Netherlands and the EA17. Notice that some of these rejections occur already at the 1% significance level: one for Argentina, Belgium, France and the EA17, and two for China, Ireland and Italy. We could have further split this group according to the number (and/or strength) of the rejections but what appears to be relevant is that we consider that beginning with this third group, i.e., for 19 of the 30 cases — 63.3(3)% — there is clear evidence that a simple linear model is not satisfactory to describe business cycle dynamic behaviour, and therefore that some kind of nonlinear model is required to perform this role satisfactorily.

Finally, for a fourth group of countries the evidence for nonlinearity is either very strong or even overwhelming, with at least 4 of the 5 linearity tests rejecting the null. These countries are the remaining 10: Australia, Austria, Brazil, Chile, Finland, Mexico, Philippines, South Africa, Spain and the U. K. . This is a very diversified group of countries, containing only one of the G7 countries (U. K.) but also the large economies of Australia and Brazil, or 5 of the G20 countries[7]: Australia, Brazil, Mexico, South Africa and the U. K. .

With at least three rejections at the 1% level, for six countries the inadequacy of the linear autoregression seems especially conspicuous; these are Australia, Finland, South Africa, and the U.K. with three rejections, and Brazil and Spain with four. However, there is no single country for which all the five tests produce no evidence for linearity.

In summary, although substantial, our evidence in favour of non-linearity does not conform with some detailed descriptions of business cycles. It is neither as generalized nor so strong as to permit dismissing linear autoregressions as useful instruments to describe them in many cases.

As the case of the U. S. motivated most research both about business cycles and about nonlinear models, the results for this country are particularly interesting. Somewhat surprisingly, none of the five tests detects significant nonlinear effects in the conditional mean, not even the CDR test, specially designed to detect the post-recession bounce-back effect found by Beaudry and Koop (1993) and confirmed, *inter alia*, by Bradley and Jansen (1997, 2000). With the largest available data sample, from 1947:1 to 2013:4, even the "usual suspect" — the low power of the tests originated by small sample sizes — does not appear to have a strong alibi here. This is not, however, neither a completely new nor a totally surprising finding. Actually, our results are consistent with those of Psaradakis and Spagnolo (2002), who do not find also any significant evidence for nonlinearity in U. S. real GNP with a different set of tests (but containing the RESET and the White tests in common with us) and a much shorter sample, from 1953:2 to 1984:4. They are also broadly consistent with the findings of the features approach briefly described in the introduction section, where the relative failure of nonlinear models is reported.

Further investigation of these results is beyond the purposes of this study. We conjecture that they might be related with the special characteristics of the last three recessions in the U. S., all originating in the financial sector and all followed by slow recoveries [8]. Furthermore, notice also that the robust HLX test does not detect any trace of nonlinear dynamic behaviour in the level of the series. Therefore, the only support for some nonlinearity is indirect and

[7] Notice that we could not find data for several countries of this group: Indonesia, Russia, Saudi Arabia and Turkey.

[8] In a recent investigation, Bec, Bouabdallah and Ferrara (2015) sucessfully specify and estimate a substantially modified version of Hamilton's (1989) Markov-Switching model; one of the most important modifications consists of allowing the bounce-back effect to appear only with some delay after the trough, which our conservative testing strategy did not allowed. See also Gadea et al. (2017).

Table 5 – Linearity tests for business cycle data (p-values)

	lagn(p)	RESET	White	Tsay	LM-STAR	CDR
Argentina	11	0.675	0.045	0.211	0.009	0.041
Australia	8	0.260	0.001	0.001	0.001	0.060
Austria	9	0.084	0.001	0.047	0.000	0.562
Belgium	4	0.107	0.027	0.045	0.001	0.794
Brazil	8	0.000	0.002	0.000	0.000	0.744
Canada	1	0.529	0.516	0.725	0.733	0.268
Chile	7	0.000	0.076	0.033	0.005	0.731
China	10	0.829	0.009	0.065	0.000	0.176
Denmark	1	0.183	0.017	0.162	0.332	0.377
Finland	3	0.000	0.000	0.000	0.010	0.305
France	2	0.305	0.040	0.198	0.002	0.399
Germany	1	0.773	0.062	0.537	0.916	0.416
Greece	5	0.991	0.016	0.026	0.175	0.254
India	1	0.850	0.605	0.742	0.955	0.422
Ireland	8	0.953	0.001	0.256	0.001	0.720
Italy	4	0.019	0.000	0.118	0.000	0.345
Japan	1	0.664	0.079	0.276	0.303	0.962
Mexico	3	0.015	0.011	0.008	0.011	0.212
Netherlands	1	0.053	0.034	0.296	0.013	0.214
Norway	1	0.475	0.589	0.789	0.406	0.417
Philippinnes	1	0.009	0.064	0.022	0.002	0.542
Portugal	3	0.124	0.345	0.016	0.117	0.760
South Africa	7	0.077	0.001	0.019	0.003	0.002
South Korea	2	0.113	0.395	0.343	0.186	0.196
Spain	8	0.032	0.000	0.000	0.000	0.000
Switzerland	1	0.415	0.179	0.670	0.448	0.669
Taiwan	1	0.093	0.106	0.102	0.141	0.032
U.K.	3	0.038	0.000	0.008	0.006	0.331
U.S.	1	0.806	0.601	0.402	0.920	0.276
EA17	1	0.123	0.017	0.003	0.033	0.440

"Lagn" or "p" now denotes the order of the autoregression which serves as the basis for the calculation of the test statistics.

comes from the unit root tests, particularly from the contradictory evidence provided by the classical, linear tests and those against nonlinear alternatives; in this case it is the Kiliç (2011) t_{ESTAR} and the Su and Nguyen (2013) $F^{GLS}_{AE,t}$ tests that suggest trend stationary, with nonlinear fluctuations around the deterministic trend, contradicting the comfortable evidence for the unit root hypothesis provided by ADF and ADF-GLS tests. Besides corresponding to the output gap view of cycles, not to the classical view, this is only an indirect indication, which is left open to explore in the future.

Seen from a rather different angle, table 4 suggests that the general White test is possibly the most powerful, with 22 rejections, conforming with simulation studies. The case for a size problem seems weak because for 3 countries only does the test produce the single rejection. Notwithstanding our rather conservative strategy, the stronger rejections, however, are those from the LM-STAR test: 14 at the 1% level. This is an expected outcome, according well with previous research. On the other hand, a very small number of rejections is produced by the CDR test, only 5, suggesting that bounce-back effects occurring after recessions are much less frequent and/or much weaker then was previously identified in business cycles [9].

4.2.2 Analyzing level data

In Table 6 we present the results for the HLX test, detailing the last column of table 5. Eighteen (60%) rejections of the linearity null are obtained for the level series, independently of their order of integration, and it is worth noting that for fifteen of these countries strong evidence for nonlinearity at business cycle frequencies had already been detected. These are Argentina, Austria, Belgium, Brazil, Chile, China, Finland, France, Greece, Italy, Netherlands, Philippines, South Africa, Spain and the EA17. It thus appears that this test is really helpful detecting nonlinear dynamics regardless of the long-run properties of the data.

Only for three countries — Australia, Ireland and the U. K. — does strong evidence for nonlinearity at the short- and medium-term frequencies does not translate into a rejection by the HLX test. For the cases of Australia and the U. K. this appears to be due to the presence of a strong linear trend, leaving only a small role to fluctuations around that trend, which therefore represent only a minor variation of the series. This is exemplified through the plots of

[9] However, this is not a completely new finding; Bradley and Jansen (1997) did not find evidence for asymmetry with the CDR test for Canada, France and Japan.

the (logged) series for Australia and for South Africa, together with their fitted values of a simple regression on a linear deterministic trend term. The case of South Africa was chosen as a basis of comparison with that of Australia due to the similarity of the available samples. This explanation does not seem to adhere to the case of Ireland, however, where fluctuations around the linear trend are relatively important when compared to those of Australia and the U. K. .

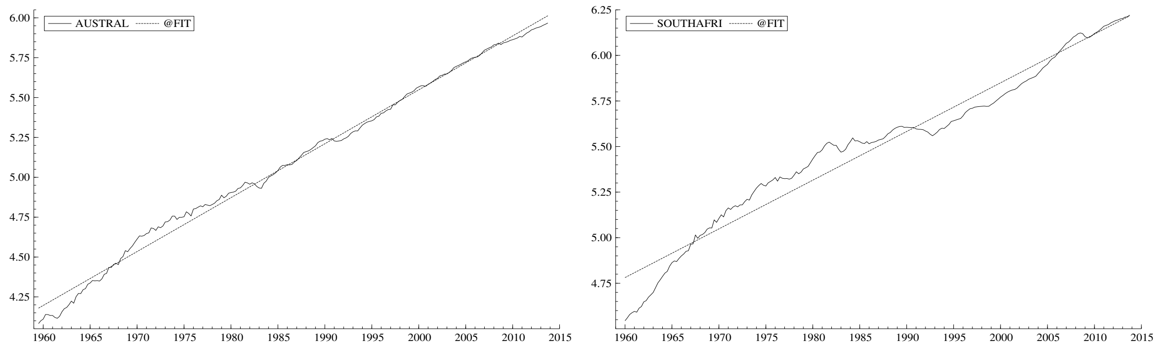


Figure 2. Logged GDP and the fitted linear trend for Australia and South Africa.

On the other hand, the HLX test detects nonlinear behaviour for three countries, Canada, Norway and Taiwan, where it was previously (almost) unnoticed. However, for one of them, Norway, the rejection occurs only at the 10% level.

5 Concluding remarks

Using quarterly data for 29 countries and the euro zone, we adopted a purely testing approach to assess the need to resort to nonlinear models to describe business cycle data. Linear autoregressive models are our departure base and to maximize power to detect non-linearities we use several tests, carefully selected, and as model-free as possible. Simultaneously, a neutral or impartial standing requires that our approach must be also conservative or cautious in size terms; we do not purchase power at any price (size). For instance, we consider inadequate a simple union of rejections strategy, as this would increase overall size well above the usual nominal 5%. Instead, we take into consideration the number and strength of rejections of the linear null, and we do not follow a data mining procedure, one of aggressively searching for rejections. Therefore, we manage to control overall size inside reasonable limits.

Table 6 – Results for the HLX linearity test

	W_0	W_1	λ	$W_\lambda p$-val.		W_0	W_1	λ	$W_\lambda p$-val.
Argentina	12.54	7.51	0.551	0.008	Italy	2.14	20.45	0.997	0.000
Australia	0.09	0.213	0.260	0.940	Japan	13.91	2.12	0.901	0.193
Austria	12.84	30.41	0.880	0.000	Mexico	4.88	1.57	0.169	0.115
Belgium	14.27	4.37	0.187	0.002	Netherlan.	7.06	7.33	0.996	0.026
Brazil	8.05	3.15	0.046	0.020	Norway	12.64	3.06	0.731	0.060
Canada	6.27	1.62	0.000	0.043	Philippin.	23.46	10.27	0.977	0.005
Chile	25.05	1.27	0.802	0.051	Portugal	7.68	0.60	1.000	0.741
China	0.03	7.22	0.721	0.074	South Af.	4.52	16.26	0.816	0.001
Denmark	3.74	2.45	0.976	0.290	South Ko.	19.74	4.45	1.000	0.108
Finland	7.22	22.96	0.000	0.027	Spain	5.18	25.14	0.428	0.001
France	10.73	0.95	0.017	0.005	Switzerl.	0.97	1.28	0.397	0.578
Germany	19.36	0.54	0.847	0.181	Taiwan	16.24	7.61	0.994	0.022
Greece	7.15	1.14	0.000	0.028	U. K.	1.33	6.07	0.066	0.440
India	14.27	0.34	0.957	0.627	U. S.	8.11	0.442	0.916	0.580
Ireland	10.71	0.10	0.871	0.480	EA17	9.50	3.79	0.000	0.009

Following Harvey et al. (2008), in this case the lag augmentation of the ADF statistics was based on the general-to-specific t-sig procedure.

Though finding substantial evidence for nonlinear dynamics, our results cast serious doubts on the common belief that business cycles are clearly and everywhere nonlinear, as would be the case if linear models always and noticeably fail. Actually, we consider that for almost 2/3 of the cases there is clear evidence that simple autoregressions are not totally adequate to describe short- and medium-term fluctuations of aggregate output. But for a rather significant group of 8 countries — including Canada, Germany, India, Japan and the U. S. —, evidence for nonlinearity at those frequencies is very weak or simply inexistent. Neither our general purpose tests nor the specific current depth of recession test corroborate strong nonlinear features previously found for the U.S..

True, our study is limited in several ways: a) we do not analyze the duration and amplitude of cycles and their phases; b) data could have a higher frequency, and c) could be less aggregated to increase power, and d) the sample size is also a serious limitation for the power of the tests, particularly for those cases where it begins in 1980:1. Notwithstanding these limitations, our evidence is far from providing full support to some descriptions, which appear to us as exaggerating some characteristics of business cycles. Our conjecture is that they are somewhat period- and/or country-specific; linear models maintain some usefulness for many cases, and cannot be simply dismissed. In short, our conclusions broadly agree with those of the "features approach".

A rather different but likely limitation concerns the issue of the robustness of these results in relation to heteroskedasticity. This issue was previously addressed in the context of the unit root test for the (log level of) output of China and a simple look at some of the plots for the differenced series — particularly for the cases of India, South Africa (S.A.), the U.K. and the U.S. — clearly suggests that it may influence the results for the linearity tests. It is widely known that it tends to inflate the size of these tests but a thorough analysis is clearly beyond the purposes of this paper. Actually, the results previously presented appear already sufficient to demonstrate the usefulness of linear models, making a lengthy discussion dispensable. Furthermore, in cases such as ours, where the purpose is only to detect non-linearities, robustification "*cannot be recommended*" (van Dick et al., 2002, p. 16, our italics) because it tends to substantially reduce the power of the tests. None the less, we have calculated the heteroskedasticity robust versions of the two general tests — RESET and White statistics — and the seemingly most powerful specific test — the LM test statistic for STAR non-linearity — and, as expected, the results became (much) more favorable to our thesis. While with the RESET test the evidence for non-linearity is only marginally weaker, large differences are found for the White and (particularly) for the LM tests: the first now produces (5% level) evidence for non-linearity for 12 series only (down from 18), and the second detects non-linearity (again at 5%) for two countries only (Brazil and Ireland)[10].

Therefore, we may conclude this small discussion stating that, in what concerns heteroskedasticity issues, the evidence that we have found for non-linearity appears to perform the role of an upper bound, i.e., the maximum of the possible evidence that one can get. When the likely presence of heteroskedasticity is accommodated, linear autoregressive models appear to become even more attractive.

On the other hand, using our unit root test based approach, we have also found some evidence for nonlinear dynamics in level data for 3 of the 8 countries mentioned above, namely Canada, Germany and the U.S.; it is strong only for the European country and rather feeble for the two American economies. Although this evidence is only indirect, it may be validly interpreted in terms of business cycle dynamics as well. However, as it refers to the fluctuations around a linear trend, it is the output gap, not the classical view of business cycles

[10] The complete results are available from the authors upon request.

that can be invoked to justify such an interpretation. Hence, its relevance for our main purpose is limited.

6 Appendix

6.1 Standard unit root test results

In this appendix we present the results for the standard unit root tests.

Table A.1 – ADF unit root tests

	ADF_{AIC}	lagn	ADF_{GTS}	lagn	ADF_{MAIC}	lagn
Argentina	−2.037	12	−1.636	11	−1.448	10
Australia	−3.408*	8	−3.408*	8	−3.408*	8
Austria	−0.346	9	−0.760	8	−0.346	9
Belgium	−0.744	4	−1.360	5	−0.744	4
Brazil	−3.218*	8	−3.218*	8	−2.223	5
Canada	−2.389	1	−2.390	1	−2.389	1
Chile	−1.758	12	−1.758	12	−1.637	7
China	−4.056***	12	−4.056***	12	−2.524	0
Denmark	−0.262	1	−1.267	10	−0.262	1
Finland	−2.540	3	−2.540	3	−2.540	3
France	−1.486	2	−1.164	9	−0.825	4
Germany	−1.300	0	−2.104	4	−1.230	0
Greece	−0.644	5	−1.906	8	−0.644	5
India	−0.995	0	−1.323	10	−0.995	0
Ireland	−1.195	12	−1.195	12	−1.185	8
Italy	+0.295	4	+0.295	4	+0.295	4
Japan	−1.665	0	−2.001	9	−1.665	0
Mexico	−3.045	3	−3.427**	2	−3.045	3
Netherlands	−0.356	0	−0.918	12	−0.356	0
Norway	−0.578	1	−0.417	12	−0.578	1
Philippinnes	−0.529	0	−0.640	9	−0.530	0
Portugal	−0.377	3	−0.403	12	−0.377	3
South Africa	−3.059	7	−3.059	7	−3.059	7
South Korea	−0.311	1	0.0923	8	−0.311	1
Spain	−0.838	8	−0.838	8	−0.838	8
Switzerland	−2.907	2	−2.949	9	−2.627	1
Taiwan	−0.436	6	−0.531	5	−0.436	6
U.K.	−2.068	3	−2.068	3	−1.589	2
U.S.	−1.938	2	−1.037	12	−1.595	1
EA17	−1.588	1	−1.714	5	−1.714	5

"Lagn" denotes the number of augmenting lags. "***", "**", and "*" represent rejections of the (unit root) null hypothesis at the 1%, 5% and 10%, respectively. The asymptotic critical values are −3.96, −3.41 and −3.13, respectively.

Table A.2 – ADF-GLS unit root tests

	ADF-GLS$_{AIC}$	lagn	ADF-GLS$_{GTS}$	lagn	ADF-GLS$_{MAIC}$	lagn
Argentina	-1.419	12	-1.183	11	-1.183	11
Australia	-1.060	8	-1.060	8	-1.060	8
Austria	-1.535	9	-1.535	9	-1.535	9
Belgium	-1.541	5	-1.541	5	-1.286	4
Brazil	-2.393	8	-2.393	8	-2.393	8
Canada	-2.428	1	-2.428	1	-2.428	1
Chile	-2.694*	12	-2.694*	12	-1.470	7
China	-4.008***	12	-4.008***	12	-2.560	0
Denmark	-1.325	3	-1.803	10	-0.713	1
Finland	-2.637*	3	-2.637 *	3	-2.637*	3
France	-1.590	2	-1.446	9	-1.590	2
Germany	-1.404	0	-2.199	4	-1.404	0
Greece	-1.524	5	-2.007	8	-1.524	5
India	-0.541	0	-1.173	10	-0.541	0
Ireland	-1.898	12	-1.898	12	-1.688	8
Italy	-0.404	4	-0.777	11	-0.404	4
Japan	-0.513	3	-0.816	9	-0.341	1
Mexico	-2.810*	3	-2.301	10	-2.810*	3
Netherlands	-0.776	0	-1.395	12	-0.776	0
Norway	-1.002	5	-0.908	12	-0.520	1
Philippinnes	-0.698	0	-1.124	9	-0.698	0
Portugal	-0.698	4	-0.719	3	-0.719	3
South Africa	-0.876	7	-1.003	6	-0.876	7
South Korea	-0.064	2	-0.231	9	-0.064	2
Spain	-1.731	8	-1.731	8	-1.731	8
Switzerland	-2.918**	2	-2.994**	9	-2.637*	1
Taiwan	-0.342	0	-0.309	5	-0.342	0
U.K.	-2.105	3	-2.314	6	-2.105	3
U.S.	-1.314	2	-0.672	12	-1.063	1
EA17	-0.535	1	-0.783	2	-0.535	1

"Lagn" denotes the number of augmenting lags. "***", "**", and "*" represent rejections of the (unit root) null hypothesis at the 1%, 5% and 10%, respectively. The asymptotic critical values are -3.48, -2.89 and -2.57, respectively.

6.2 Overview of unit root tests against nonlinear alternatives

To understand the mechanics of some of the tests consider a zero-mean stochastic process $\{x_t\}$. To build a unit root test, instead of the usual linear autoregression let us consider the following nonlinear dynamic model:

$$\Delta x_t = \phi\, x_{t-1}\, G(z_{t-d}; \gamma) + u_t,\ t = 1, \ldots, T,$$

where $G(.)$ is the transition function, a nonlinear function taking values between 0 and 1, z_{t-d} is the transition variable, $d(\geq 1)$ is the delay parameter, and u_t is a stationary and invertible zero-mean process. When $G = 0$ the process is in the middle regime and contains a unit root. On the other hand, when the function G satisfies the condition that it approaches 1 when $z_{t-d} \to \pm\infty$, as the exponential, provided that $\phi < 0$ the process is globally stationary and it is in the outer regime in that case, showing a tendency to revert to its mean.

Adopting the most popular transition function, the exponential function, and making $z_t \equiv x_t$ and $d = 1$, the process becomes the exponential STAR (ESTAR)

$$\Delta x_t = \phi\, x_{t-1}\, [1 - \exp(-\gamma\, x_{t-1}^2)] + u_t,\ \gamma \geq 0, \tag{2}$$

where γ is a parameter controlling the smoothness of the function G. In this context, the test for a unit root is the test of

$$H_0 : \phi = 0 \quad vs. \quad H_1 : \phi < 0 \ \ (\text{and } \ \gamma > 0).$$

This model can be easily extended to more empirically relevant cases, with a non-zero constant mean or a linear deterministic trend. In the first case x_t is replaced with $x_t^* = x_t - \mu_x$, where μ_x represents the constant mean of x_t, and in the second, which is the relevant one for our purposes, the original observed time series, y_t, is detrended, i. e. x_t is replaced with

$$x_t^* = y_t - (\alpha + \beta\, t),$$

where α and β are parameters to be estimated, usually by OLS, producing $\hat{x}_t$ (where we have dropped the asterisk to simplify the notation).

The problem with testing the previous hypothesis, the so-called "Davies problem", is that the parameter γ is not identified under the null [11]. To

[11] The fact that testing for the unit root may also be formulated as $H_0 : \gamma = 0$ $vs.$ $H_1 : \gamma > 0$, as in KSS, is also a manifestation of this problem; in this case it is the parameter ϕ that is not identified under H_0.

circumvent it two main approaches have been used so far:

a) to employ a first-order Taylor series expansion of the nonlinear model around $\gamma = 0$ and to formulate a test in terms of the corresponding parameters of the new (linear) model;

b) to construct a test statistic based on an extremum over the parameter space of the original nonlinear model.

The well KSS test follows the first route and owes its popularity to the simplicity of the auxiliary test regression. The second route is followed in the test of Shintani (2013), who extends the work of Park and Shintani (2005) to trending data. Shintani uses the parametrization $\gamma = \theta^2$ and proposes running the regressions

$$\Delta \hat{x}_t = \phi\, \hat{x}_{t-1} \left[1 - \exp(-\theta^2 \hat{x}_{t-1}^2)\right] + \sum_{i=1}^{k} \alpha_i\, \Delta \hat{x}_{t-i} + \varepsilon_t,$$

for all $\theta \in \Theta_n = [10^{-1},\, 10] \times P_n$, where $P_n = (\sum \hat{x}_{t-1}^2 / T)^{-1/2}$. The test statistic is the infimum of the t-ratios of $\hat{\phi}(\theta)$ over Θ_n, i.e.,

$$\inf -t_{E,\tau} = \inf_{\theta \in \Theta_n} \frac{\hat{\phi}(\theta)}{se(\hat{\phi}(\theta))}.$$

In the same vein, Kiliç (2011) proposes using the infimum of the t-ratios of ϕ in the auxiliary regressions

$$\Delta \hat{x}_t = \phi\, \Delta \hat{x}_{t-1} \left[1 - \exp(-\gamma\, \Delta \hat{x}_{t-1}^2)\right] + \sum_{i=1}^{k} \beta_i\, \Delta \hat{x}_{t-i} + \epsilon_t,$$

over all the possible values of γ, that is, the transition variable is the lagged difference of the (detrended) variable, not its lagged level. The test statistic is defined as

$$t_{ESTAR} = \inf_{\gamma \in \Gamma_T} \frac{\hat{\phi}(\gamma)}{se(\hat{\phi}(\gamma))},$$

where $\gamma \in \Gamma_T = [\frac{1}{100 s_{zT}}; \frac{100}{s_{zT}}]$, s_{zT} representing the sample standard deviation of the transition variable, $\Delta \hat{x}_{t-1}$, which Kiliç finds having good power properties for several DGPs, even when they follow not a smooth transition model but a threshold one.

The Sollis (2009) $F_{AE,t}$ test adopts the Taylor series expansion approach but departs from a model that generalizes the ESTAR, permitting asymmetric behaviour in the adjustment towards the mean under the (globally stationary) alternative. The extended model is called asymmetric ESTAR (AESTAR) and combines both an exponential and a logistic transition function, i.e., instead of (2) the model becomes

$$\Delta x_t = G_t(\gamma_1, x_{t-1})[S_t(\gamma_2, x_{t-1})]\,\rho_1 + [1 - S_t(\gamma_2, x_{t-1})\,\rho_2]x_{t-1} + \epsilon_t,$$

where

$$G_t(\gamma_1, x_{t-1}) = 1 - \exp[-\gamma_1(x_{t-1}^2)], \gamma_1 \geq 0, \quad \text{and}$$
$$S_t(\gamma_2, x_{t-1}) = [1 + \exp(-\gamma_2 x_{t-1})]^{-1}, \gamma_2 \geq 0.$$

Taking several Taylor series expansions, Sollis shows that the test regression is

$$\Delta\,\hat{x}_t = \phi_1\,\hat{x}_{t-1}^3 + \phi_2\,\hat{x}_{t-1}^4 + \sum_{i=1}^{k} \kappa_i\,\Delta\,\hat{x}_{t-1} + \eta_t,$$

where testing for the unit root amounts to testing $H_0 : \phi_1 = \phi_2 = 0$

6.3 Further nonlinear unit root test results

Besides using the previous test statistics, we have gathered more evidence through further unit root tests against nonlinear alternatives. These are all based on the Taylor series expansion approach and are: i) the GLS version of the KSS test, as proposed by Kapetanios and Shin (2008); b) the $F_{s,ct}$ statistic of Sollis (2011) derived against a stationary STAR model that resorts to a second-order logistic transition function (replacing the usual exponential) and that nests a three-regime TAR model; iii) the GLS version of the $F_{AE,t}$ statistic, proposed by Su and Nguyen (2013); iv) and the F_{ABG} test of Addo, Billio and Guégan (2014, ABG), which is derived against a MT-STAR stationary alternative model that allows asymmetric adjustment towards equilibrium (see ABG for details).

References

[1] Addo, P. M., Billio, M. and Guégan, D. (2014), The univariate MT-STAR model and a new linearity and unit root test procedure, *Computational Statistics and Data Analysis*, 76(C), 4-19.

[2] Beaudry, P. and Koop, G. (1993), Do recessions permanently change output?, *Journal of Monetary Economics*, 31, 149-64.

Table A.3 – Further unit root tests against nonlinear alternatives

	t_{NL}^{GLS}(lagn)	$F_{s,ct}$ (lagn)	$F_{AE,t}^{GLS}$ (lagn)	ABG (lagn)
Argentina	−2.873 (12)	3.154 (12)	4.547 (12)*	2.313 (12)
Australia	−3.511 (8)**	7.869 (9)**	1.131 (8)	5.327 (9)**
Austria	−2.494 (9)	0.499 (9)	3.048 (9)	1.143 (9)
Belgium	−1.818 (5)	1.461 (5)	1.804 (5)	1.776 (5)
Brazil	−2.859 (8)	6.464 (8)*	5.154 (8)*	5.157 (8)**
Canada	−2.747 (1)	4.616 (1)	3.829 (1)	4.364 (1)
Chile	−2.870 (12)	3.464 (12)	4.168 (12)	2.400 (12)
China	−6.697 (12)**	16.026 (10)***	22.677(12)***	10.566 (10)***
Denmark	−0.869 (3)	1.441 (3)	1.077 (1)	0.988 (3)
Finland	−2.551 (3)	3.377 (3)	3.297 (3)	2.607 (3)
France	−1.171 (2)	1.723 (2)	2.134 (2)	1.459 (2)
Germany	−3.680 (4)**	10.042 (4)***	8.412 (4)***	6.315 (4)**
Greece	−1.561 (5)	0.718 (5)	4.581 (9)*	0.715 (5)
India	−0.928 (0)	1.406 (0)	1.276 (0)	0.755 (0)
Ireland	−2.115 (12)	1.607 (12)	2.196 (12)	1.245 (8)
Italy	−0.018 (4)	0.378 (4)	1.413 (4)	0.886 (4)
Japan	−1.173 (3)	1.880 (3)	0.854 (3)	2.013 (3)
Mexico	−3.107 (3)**	4.974 (3)	4.913 (3)*	3.002 (3)
Netherlands	−1.135 (0)	0.937 (0)	1.191 (3)	0.933 (0)
Norway	−1.175 (5)	0.688 (5)	1.448 (5)	0.385 (5)
Philippines	−0.172 (0)	0.845 (0)	0.446 (0)	0.132 (0)
Portugal	−0.966 (4)	0.232 (4)	0.837 (4)	2.017 (4)
South Africa	−1.726 (7)	4.900 (7)	2.536 (7)	3.413 (7)
South Korea	−0.137 (2)	0.444 (2)	1.635 (1)	0.614 (2)
Spain	−1.362 (8)	2.103 (9)	1.089 (8)	0.630 (8)
Switzerland	−2.461 (1)	4.436 (2)	4.511 (2)	3.531 (1)
Taiwan	−0.455 (0)	0.549 (1)	0.125 (0)	0.433 (1)
U.K.	−2.609 (3)	2.400 (3)	3.734 (3)	1.903 (3)
U.S.	−1.117 (2)	1.998 (2)	6.001 (2)**	2.234 (2)
EA17	−1.965 (2)	2.724 (2)	4.020 (2)	1.955 (2)

In all the cases the lag length ("lagn") was estimated using the AIC statistic. For the t_{NL}^{GLS} test the 5% asymptotic critical value, the only one made available by Kapetanios and Shin (2008) is −2.93. For the $F_{s,ct}$ statistic the asymptotic critical values are 5.727, 6.717 and 8.617. For the $F_{AE,t}^{GLS}$ statistic they are 4.531, 5.373 and 7.286 and for the F_{ABG} test they are 4.444, 5.132 and 6.602, respectively.

[3] Bec, F., Guay, A. and Guerre, E. (2008), Adaptive consistent unit-root tests based on autoregressive threshold model, *Journal of Econometrics* 142, 94–133.

[4] Bec, F., Bouabdallah, O. and Ferrara, L. (2015), Comparing the shape of recoveries: France, the UK and the US, *Economic Modelling*, 44, 327-34.

[5] Bradley, M. D. and Jansen, D. W. (1997), Nonlinear business cycle dynamics: cross-country evidence on the persistence of aggregate shocks, *Economic Inquiry*, 35, 495-509.

[6] Bradley, M. D. and Jansen, D. W. (2000), Are business cycle dynamics the same across countries? Testing linearity around the globe, *Studies in Nonlinear Dynamics and Econometrics*, 4(2), 51-71.

[7] Choi, C.-Y. and Moh, Y.-K. (2007), How useful are tests for unit-root in distinguishing unit-root processes from stationary but non-linear processes? *Econometrics Journal*, 10, 82-112.

[8] Clements, M. P. and Krolzig, H.-M., (2003), Business cycle asymmetries: characterization and testing based on Markov-switching autoregressions, *Journal of Business & Economic Statistics*, 21 (1), 196-211.

[9] Crowley, P., Garcia, H. and Quah, C. H. (2008), Is Europe growing together or growing apart? *Research Discussion Papers* 33/2013, Bank of Finland.

[10] Elliott, G., Rothenberg, T. J. and Stock, J. H. (1996), Efficient tests for an autoregressive unit root, *Econometrica*, 64 (4), 813-36.

[11] Enders, W. and Granger, C. W. J. (1998), Unit-root tests and asymmetric adjustment with an example using the term structure of interest rates, *Journal of Business & Economic Statistics*, 16, 304-11.

[12] Engel, J., Haugh, D. and Pagan, A., (2005), Some methods for assessing the need for non-linear models in business cycle analysis, *International Journal of Forecasting*, 21 (4), 651-62.

[13] Gadea, M. D., Gomez-Loscos, A. and Perez-Quiros, G. (2017), Dissecting US recoveries, *Economis Letters*, 154, 59-63.

[14] Galvão, A. B. C. (2002), Can non-linear time series models generate US business cycle asymmetric shape? *Economics Letters*, 77, 187-94.

[15] Granger, C. W. J. (1993), Strategies for modelling nonlinear time-series relationships, *The Economic Record*, 69, 233-8.

[16] Granger, C. W. J. and Lee, T.-H. (1999), Effect of aggregation on nonlinearity, *Econometric Reviews*, 18, 259-69.

[17] Granger, C. W. J. and Teräsvirta, T. (1993), *Modelling Nonlinear Economic Relationships*, Oxford University Press.

[18] Greenland, S. J., Senn, K. J., Rothman, J. B., Carlin, J. B., Poole, C., Goodman, S. N. and Altman, D. G. (2016), Statistical tests, P-values, confidence intervals, and power: a guide to misinterpretations. *The American Statistician, Online Supplement, 1-12.*

[19] Hamilton, J. (1989), A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, 57, 357-84.

[20] Hansen, B. E. (1992), The likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP, *Journal of Applied Econometrics*, 7, S61-S82 (Erratum: 11, 195-198).

[21] Harding, D. and Pagan, A. R., (2002), Dissecting the cycle: a methodological investigation, *Journal of Monetary Economics*, 49, 365-81.

[22] Harvey, D. I., Leybourne, S. J. and Xiao, B., (2008), A powerful test for linearity when the order of integration is unknown, *Studies in Nonlinear Dynamics & Econometrics*, 12 (3), article 2.

[23] Hess, G. D. and Iwata, S., (1997), Measuring and comparing business-cycle features, *Journal of Business & Economic Statistics*, 15, 432-44.

[24] Kapetanios, G. and Shin, Y. (2008), GLS detrending-based unit root tests in nonlinear STAR and SETAR models, *Economics Letters*, 100, 377-80.

[25] Kapetanios, G., Shin, Y. and Snell, A. (2003), Testing for a unit root in the nonlinear STAR framework, *Journal of Econometrics*, 112, 359-79.

[26] Kiliç, R. (2004), Linearity tests and stationarity, *Econometrics Journal*, 7, 55-62.

[27] Kiliç, R. (2011), Testing for a unit root in a stationary ESTAR process, *Econometric Reviews*, 30 (3), 274-302.

[28] Kose, M. A., Otrok, C. and Prasad, E. (2008), Global business cycles: convergence or decoupling?, *IZA Discussion Paper* no. 3442.

[29] Lanne, M., Lütkepohl, H. and Saikkonen, P. (2003), Test procedures for unit roots in time series with level shifts at unknown time, *Oxford Bulletin of Economics and Statistics*, 65, 91-115.

[30] Lee, T.-H., White, H. and Granger, C. W. J. (1993), Testing for neglected nonlinearity in time series models: a comparison of neural network methods and alternative testes, *Journal of Econometrics*, 56, 269-90.

[31] Leybourne, S.J., Newbold, P. and Vougas, D. (1998), Unit roots and smooth transitions, *Journal of Time Series Analysis*, 19, 83-97.

[32] Lim, T.-H., Leybourne, S. and Newbold, P. (2002), Unit root tests with a break in innovation variance, *Journal of Econometrics*, 109, 365-387.

[33] Lütkepohl, H. (2004), Univariate time series models, *in* Lütkepohl, H. and Krätzig, M. (eds.), *Applied Time Series Econometrics*, Cambridge University Press, 8-85.

[34] Luukkonen, R., Saikkonen, P. and Teräsvirta, T. (1988), Testing linearity against smooth transition autoregressive models, *Biometrika*, 75, 491-99.

[35] Morley, J. and Piger, J. (2006), The importance of nonlinearity in reproducing business cycle features, *in* Milas, C., Rothman, P. and van Dijk, D. (eds.), *Nonlinear Time Series Analysis of Business Cycles*, Elsevier, 75- 94.

[36] Morley, J. and Piger, J. (2012), The asymmetric business cycle, *The Review of Economics and Statistics*, 94 (1), 208-21.

[37] Park, J. Y. and Shintani, M. (2005), Testing for a unit root against transitional autoregressive models, Working Paper, Vanderbilt University.

[38] Pippenger, M. K. and Goering, G. E. (1993), A note on the empirical power of unit root tests under threshold processes, *Oxford Bulletin of Economics and Statistics*, 55, 473-81.

[39] Psaradakis, Z. and Spagnolo, N. (2002), Power properties of nonlinearity tests for time series with Markov regimes, *Studies in Nonlinear Dynamics & Econometrics*, 6(3), article 2.

[40] Shintani, M. (2013), The inf-t test for a unit root against asymmetric exponential smooth transition autoregressive models, *Japanese Economic Review*, 64(1), 3-15.

[41] Singh, T. (2012), Testing nonlinearities in economic growth in the OECD countries: an evidence from SETAR and STAR models, *Applied Economics*, 44, 3887-3908.

[42] Sollis, R. (2009), A simple unit root test against asymmetric STAR nonlinearity with an application to real exchange rates in Nordic countries, *Economic Modelling*, 26, 118-25.

[43] Sollis, R. (2011), Testing the unit root hypothesis against TAR nonlinearity using STAR-based tests, *Economics Letters*, 112, 19–22.

[44] Su, J.-J. and Nguyen, J. K. (2013), GLS detrending in Sollis nonlinear unit root tests, *Applied Economics Letters*, 20 (13), 1259-62.

[45] Teräsvirta, T. (1994), Specification, estimation, and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association*, 89, 208-18.

[46] Teräsvirta, T. (2004), Smooth transition regression modelling, *in* Lütkepohl, H. and Krätzig, M. (eds.), *Applied Time Series Econometrics*, Cambridge university Press, 222-41.

[47] Teräsvirta, T. and Anderson, H. A. (1992), Characterising non-linearities in business cycles using smooth transition autoregressive models, *Journal of Applied Econometrics*, 7, S119-S136.

[48] Tsay, R. S. (1989), Testing and modeling threshold autoregressive models, *Journal of the American Statistical Association*, 84, 231-40.

[49] van Dijk, D., Teräsvirta, T. and Franses, P. H. (2002), Smooth transition autoregressive models – a survey of recent developments, *Econometric Reviews*, 21(1), 1-47.

[50] White, H. (1982), Maximum likelihood estimation of misspecified models, *Econometrica*, 50, 1-25.