'Nonlinear causality between crude oil price and exchange rate: A comparative study of China and India' - A failed replication (negative Type 1 and Type 2)

Glauco De Vita[a,*] and Emmanouil Trachanas[b]

[a] Centre for Business in Society, Coventry University, Priory Street, Coventry CV1 5FB, UK.
* Corresponding author. E-mail: glauco.devita@coventry.ac.uk
[b] Department of Accounting, Economics and Finance, Oxford Brookes University, Oxford OX33 1HX, UK. E-mail: etrachanas@brookes.ac.uk

Energy Economics 56, p. 150-160, 2016
http://www.sciencedirect.com/science/article/pii/S0140988316300524

Abstract
Evidence published in this journal by Bal and Rath (2015) purports a bidirectional nonlinear causality between oil price and India's exchange rate and, for China, unidirectional nonlinear causality running from exchange rate to oil price. Their entire testing protocol and ensuing results rest upon claims that all the variables contain a unit root. We raise several critical issues and revisit the order of integration of the series as well as their cointegration and Granger causality properties through a 'pure replication' and a 'reanalysis'. Contrary to Bal and Rath (2015), when we repeat their estimated model with their specification of the Ng and Perron (2001) unit root test on their data, we find that their oil price series (*ROL*) is level stationary (negative replication Type 1), a result which makes all their subsequent results biased and misleading. Our *reanalysis* confirms that *ROL* is *I*(0), linearly as well as nonlinearly. We also find that the basic bivariate model proposed by Bal and Rath (2015) fails to produce statistically robust and stable cointegrating patterns. Nonlinear causality tests confirm the absence of any nonlinear causality for both countries (negative replication Type 2).

JEL classification: C22; C52; C59; F31; Q41; Q43
Keywords: Replication; Causality; Oil price; Exchange rate; Unit root; Cointegration

Acknowledgements: We are grateful to Debi P. Bal and Badri N. Rath for making available their data. We also wish to thank Cees Diks, David Harvey, Robinson Kruse and Paresh K. Narayan for providing us with their estimation codes, Lester Hunt and Dimitris Hatzinikolaou for their comments on an earlier version of this paper, and Dimitrios Asteriou, Bahram Pesaran and Anestis Fachantidis for some technical advice. Finally, we are indebted to three anonymous reviewers and the Editor Richard Tol, for their helpful suggestions. The usual disclaimer applies.

1. Introduction

The validity and reliability of published results are at the heart of scientific rigor, and yet, verification through replication remains, disappointingly, an undervalued endeavor of economic research. Indeed, aside from sporadic 'Comments' or 'Notes', standard, full-length replication articles rarely appear in the pages of top journals.[1] In their strong call for more replication studies, Burman et al. (2010: 788) emphasize that replication is a critical tool for scientific progress and that the absence of such studies "is particularly problematic because empirical economic research is often prone to error." The inherent value of the present study lies in contributing to scientific progress by invalidating Bal and Rath's (2015) research findings.

In a recent article in this journal (2015, 51, 149-156), Bal and Rath (henceforth B-R) investigate the nonlinear Granger causality between crude oil price and the exchange rate for both China and India over the period January 1994 to March 2013. They claim to unveil results indicating that all the variables contain a unit root (UR) when the (linear) Ng and Perron (2001) UR test, and the Narayan and Popp (2010) UR test with two structural breaks are performed, and that, when the Hiemstra and Jones (1994) nonlinear Granger causality test is applied to the VAR residuals, a significant bidirectional nonlinear Granger causality between crude oil prices and exchange rates is found for both countries. They also find that when repeating the Hiemstra and Jones (1994) test on the residuals of a GARCH (1, 1) model to check for robustness, their results show that bidirectional nonlinear Granger causality only holds for India, whilst for China nonlinear causality only runs one way, from exchange rate to oil price.

B-R (2015) raises, in our view, as many questions as it provides answers to, in terms of its ambiguous theoretical premise, the results pertaining to the primary aim of the study, and the econometric procedures applied in pursuit of such aim. We revisit B-R (2015) published results through a 'pure replication' and a 'reanalysis'. Consistent with the harmonizing framework for replications advanced by Clemens (in press), the former exercise is based on *verifying* the original results by *replicating* - using the same model specification, test, and sample –[2] the exact statistical analysis B-R (2015) conducted in the original paper (up to the point where any discrepancies are found). With regard to the latter, our 'reanalysis' can be classified as a 'robustness test' of their unit root, cointegration and nonlinear causality results but one that retains the same data set, sample period and variable and measures specification adopted by B-R (2015), with the only variant being the estimation or testing techniques employed.[3]

---

[1] For a recent exception, see Herndon et al. (2014).

[2] This differs from replication by 'reproduction', which is undertaken using a different sample (see Clemens, in press, Table 1).

[3] In contrast, according to Clemens (in press, p. 2), an 'extension' (robustness) test entails "using new data gathered on a sample representative of a different population, or gathered on the same sample at a substantially different time, or both." We avoid performing an 'extension' since such robustness test would estimate population parameters that are different from those in the original study and hence generate results that would not be identical in expectation.

Arulampalam et al. (1997) draw a similar distinction between the term *replication*, taken to mean using the original data and code to attempt to duplicate exactly the same results as appear in the paper, and *reanalysis*, interpreted as a robustness test that allows for changes in empirical specifications and/or estimation methods. Hamermesh (1997: 107) argues that "The best replication studies [...] will attempt duplication as their starting point, but go far beyond that. They might, for example, [...] try alternative methods and other specifications."

The above framework is broadly consistent with the codes for failed replications proposed by Burman et al. (2010: 789), according to which 'negative Type 1' replications refer to situations where replicating authors "are unable to reproduce the original article's results using the same data, the same specification, and the same econometric software" whereas 'negative Type 2' replications find that the original results are not robust to substantial changes, for example, in terms of functional form or alternative estimation procedures. In the present study we find that B-R (2015) results fail to pass both kinds of replications, Type 1 and Type 2.

Why did we do this? That is, what is the motivation underlying our replication study? Questionable methodological choices and puzzling results aside, what first drew our attention to this paper relates to the ambiguous economic rationale offered by B-R (2015) to investigate the bilateral causality between the real international price of crude oil and exchange rates, notwithstanding the significance that crude oil plays as a form of exhaustible energy tradable in international markets, particularly when examined in relation to countries such as China and India, two of the largest oil-importing countries in the world. Their findings, if proven to be correct, are certainly of importance for the field of energy economics as well as energy finance. Yet it is not immediately obvious why the Indian or Chinese exchange rate should be expected to have an impact on the international price of crude oil, as is their purported finding that both the Indian and Chinese exchange rates have a significant long-run causal effect (nonlinearly). Despite the few empirical studies cited (see their Section 2), their article offers very little in terms of theoretical grounding, leaving the reader puzzled as to what exactly is the theory behind the postulated causal relationship. It is, of course, true that both India and China are relatively large oil importers, but this does not necessarily mean that fluctuations in their national currency or even devaluations could reasonably be expected to impact the oil price in international markets, linearly or nonlinearly.[4]

Additionally, from both an economic and econometric perspective, there appears little justification to assume a simple *bivariate* causal relationship where either the exchange rate is dependent on oil price or *vice versa*; a premise which makes the model estimated by B-R (2015) highly susceptible to omitted variable bias. Oil price is certainly not the main, let alone the unique variable that determines movements in the exchange rate, which – as predicted theoretically – has a number of likely determinants such as inflation, interest rate, public debt,

[4] As pointedly observed by an anonymous reviewer, this is particularly the case when considering that, unlike other commodities, the price of crude oil is mainly set by the large oil exporters or OPEC, at least during the estimation period considered.

etc. The same logic applies to the Indian or Chinese exchange rate taken as the sole factor to have explanatory power in the determination of the international price of crude oil (geopolitically driven oil supply disruptions being a case in point). Evidently, ignoring other main determinants may lead to unreliable results. Indeed, a number of studies in the energy literature (e.g., Narayan and Smyth, 2009) show that conducting such bivariate causality exercises might be misleading and it is well established in the econometrics literature that the omission of causality patterns from other theoretically predicted variables can lead to spurious inferences (see Granger, 1969; Lutkepohl, 1982; Triacca, 1998).

There are other econometric gray areas in B-R (2015) that offer scope for critical and empirical scrutiny (as discussed in the next section). First and foremost, their failure to report the UR test results in both levels and first differences for both UR tests conducted, which is, in itself, most unusual.

The nature and structure of our replication adheres to the excellent guidance provided by Burman et al. (2010) in terms of the ground rules and principles for replications, including the expectation for such studies to be presented as standard, full-length manuscripts, to be submitted for peer-review to the same journal where the original research was published, to provide sufficient detail to show that the replication was done correctly and, finally, to attempt first to replicate exactly the original findings by starting with the same data and specification before testing the robustness of the original research through alternative techniques.

The rest of this paper is organized as follows. Section 2 summarizes B-R (2015) analytical steps alongside the gray areas inherent in their methodological choices. Sections 3 and 4 present and discuss our results obtained from a *pure replication* and a *reanalysis*. The final section concludes.

2. Bal and Rath (2015) analytical steps and 'gray areas'

The analysis reported in B-R (2015) begins with presenting plots of the rates of change between crude oil price and the real effective exchange rate (REER) of India and China, the visual inspection of which leads them to infer ('assume' rather) that a nonlinear relation exists between the variables, for both countries.[5] They then perform the Ng and Perron (2001) UR test, and the Narayan and Popp (2010) test with two structural breaks (in level and trend). These *linear*

[5] We use the original data set, which Bal and Rath provided to us. In terms of the definition of the variables, B-R (2015: 152) state: "The real effective exchange rate of India (RIX), obtained from the official website of the Reserve Bank of India published in the Handbook of Statistics on Indian Economy, was used in this study. The real effective exchange rate of China (RCX) was obtained from the CEIC database, a product of the Euromoney Institutional Investor Company. The crude oil price, taken in real terms and deflated by the US consumer price index following Faria et al. (2009), was defined as the spot price of West Texas Intermediate (WTI), a definition obtained from the Energy Information Administration, US Department of Energy. The data for crude oil prices (ROL) and the US consumer price index were obtained from the CEIC database." According to our inspection of these databases, both exchange rate measures would appear to be based on domestic currency in terms of foreign currency and adjusted for relative price levels, with an increase in *RIX* or *RCX* indicating a real appreciation of the domestic currency.

tests are performed on the raw data series rather than their log form, a very uncommon yet potentially legitimate choice since the natural logarithmic transformation may induce a linearization of the raw data. For the Ng-Perron (2001) test, only evidence of the stationarity of the series in first difference is reported (a highly atypical and inadequate way of presenting UR test results), while for the Narayan and Popp (2010) test with two breaks (henceforth N-P), the reported statistics do not reject the null of a unit root for the (level) series of real crude oil price (*ROL*), REER of India (*RIX*), and REER of China (*RCX*).

The first objection to the procedures employed by B-R (2015) - in addition to the omission of the Ng-Perron (2010) results for the series in levels - is that despite the susceptibility of the variables to shocks, both temporary and permanent, and B-R (2015) initial assumption of a nonlinear relation between *ROL* and REER of India and China, they rely solely on the results of linear UR tests to ascertain the order of integration of the variables. Moreover, although the (linear) N-P (2010) test accounts for up to two structural breaks, it has been found to produce misleading results in application to high frequency data. For example, Mishra and Smyth (2014) find that US monthly natural gas consumption is $I(1)$ according to the N-P (2010) test but when the Narayan and Liu (2013) test that accommodates for heteroskedasticity is employed, the series is found to be mean reverting.

The ambiguity surrounding the order of integration of *ROL* in levels is augmented by the fact that there is considerable disagreement in the literature as to whether the finding of the non-stationarity of oil prices may be functional-form dependent, for example, in terms of a deterministic trend, a stochastic trend or structural breaks (for a review see Şimşek, 2014). The modelling of such alternative specifications is also dependent on whether the nature of shocks is a temporary phenomenon that reflects short-term variability or a permanent one that affects the long-run path of oil prices, thereby causing nonlinearities in the evolution of the series (an aspect that cannot be verified through a visual inspection of the plots of the series). The UR testing protocol in B-R (2015) neglects the possible existence of nonlinearities, which cannot be captured, if present, simply by accounting for breaks, an issue that may also affect whether a unit root is found (see, e.g., Aksoy and Leon-Ledesma, 2008). Evidently, more sophisticated UR tests should be used for a more accurate detection of the potential nonlinearities of the series so as to ascertain, with greater confidence, the true order of integration of the variables. For example, the tests developed by Harvey and Leybourne (2007) and Harvey et al. (2008) can be used to pre-test for linearity to help establish whether a linear or nonlinear UR test should be employed. If the series are found to be nonlinear, then nonlinear UR tests such as those proposed by Kapetanios et al. (2003) or Kruse (2011) should be performed.

B-R (2015) then proceed by testing for a linear long-run relationship through the Johansen and Juselius (1990) (henceforth J-J) method, which uncovers cointegration in the case of both China and India (their Table 3 and 4, p. 153). Next, they perform a linear causality test, on both equations (their Table 5 and 6, p. 154). They account for the possible effects of cointegration on the basis of the cointegrating VECM, finding evidence of linear long-run

Granger causality running from oil price to exchange rate for both countries, and no evidence of causality running the other way, in either country. However, they then dismiss the reliability of these estimation results due to low R^2 values and high F-test statistics.

Another gray area in the analytical choices outlined above pertains to the value of performing exclusively a *linear* cointegration test such as J-J (1990) in the context of a study focusing on the nonlinear causality of the long-run relationship between the variables, especially given B-R (2015) initial assumption of a nonlinear relation between *ROL* and REER for both countries. In fact, two main complications may arise from this approach. First, should either of the two variables be found to be stationary nonlinearly, this would invalidate any inferences drawn from the results of the trace and maximum eigenvalue statistics of the J-J (1990) cointegration test. This is because - even in the event in which such a cointegrating relation is linear in nature – application of the J-J (1990) cointegration test requires all the system's variables to be integrated of the same order for valid inference (see, e.g., Harris, 1995; De Vita and Abbott, 2002; and De Vita et al., 2006).[6] The second concern relates to the need to investigate the existence of nonlinear cointegration as well as linear cointegration between the variables in the context of an analysis aiming to investigate the nonlinear causal properties of the relationship in question. On this account, it bears reminding that whilst 'correlation' does not imply 'causality', 'cointegration' between two variables must entail a causal relationship, at least in one direction. Against this backcloth, the finding by B-R (2015: 152) of the existence of, "at most", one *linear* cointegrating vector, if taken to be a true reflection of the underlying data generation process, would suggest the absence of any additional long-run co-movement path shared by the variables from which any nonlinear long-run Granger causality patterns could be associated with. Evidently, in this context, nonlinear cointegration tests such as the Nonlinear ARDL (NARDL) asymmetric cointegration test recently developed by Shin et al. (2014) are called for in order to at least verify the existence of a long-run nonlinear relation between the pairs of variables from which long-run *nonlinear* causality can be established.

Next, following confirmation of the existence of nonlinear dependencies from a basic BDS test [7] on the VAR residuals, B-R (2015) present the results of the Hiemstra and Jones (1994) nonlinear Granger causality test (henceforth H-J). The reported results (their Table 8, p. 154) suggest that, for both India and China, a significant bidirectional nonlinear causality is

[6] For example, Harris (1995) argues that this assumption becomes highly restrictive in the bivariate ($n = 2$) case since, given that a $I(0)$ variable is stationary by itself, it forms a linearly independent cointegrating relation by itself in the cointegrating rank. Rahbek and Mosconi (1999) also demonstrate that in Johansen-type frameworks the likelihood testing procedure for the cointegrating rank can be sensitive to the presence of stationary variables as they can lead to nuisance parameters in the asymptotic distribution of the trace test.

[7] First discussed in Brock, Dechert and Scheinkman (1987), the BDS (acronym) test is a test for independence based upon the correlation dimension. Although the BDS test was not conceived as a leading indicator it can serve as a residual diagnostic to test the null that the residuals of a model subjected to filtering or first-differencing are independent and identically distributed. Rejection of the null implies that the remaining model structure may hide forms of misspecification such as non-linearites or non-stationarity.

observed between *ROL* and the exchange rate. B-R (2015) end their empirical exercise by attempting to test for robustness. With this aim in mind, they repeat the H-J (1994) test on *ROL* and REER series filtered using a GARCH (1, 1) model but only to obtain yet more different results (their Table 10, p. 155). For India the variables now appear to Granger cause each other, nonlinearly (though at lower lag lengths, i.e., lags 1 and 2, the exchange rate does not significantly Granger cause oil price) whereas in the case of China, only unidirectional causality from exchange rate to oil price is detected.

With regard to the above steps, the main objection we raise relates to the *exclusive use* of the H-J (1994) test to establish the nonlinear causal relation between a pair of variables. The H-J (1994) test is simply a modified version of the Baek and Brock (1992) test in which the assumption that the time series are 'mutually and individually independent and identically distributed' is relaxed, thus allowing the series tested for to display *weak* (or *short-term*) temporal dependence. However, it is widely acknowledged (see Bekiros and Diks, 2008a and 2008b; Hassani et al., 2010; Smyth and Narayan, 2015) that this nonparametric causality test suffers from severe drawbacks. For example, the test has been found to be biased toward the rejection of the null hypothesis when it is true (Bekiros and Diks, 2008a; Diks and Panchenko, 2005 and 2006; Hassani et al., 2010). Diks and Panchenko (2005) also show that the H-J test is not consistent, at least against a specific class of alternatives, while Smyth and Narayan (2015) reiterate that this test is not applicable to non-stationary data. It would be useful, therefore, when applying the H-J (1994) test, to – at least – perform concomitantly an additional nonparametric Granger causality test – such as the one developed by Diks and Panchenko (2006) – that overcomes some of these limitations.[8]

Our final concern pertains to the issue of what these nonlinear causality tests should be performed on. Most previous empirical studies apply such tests on the residuals of a VAR model - if the series are $I(0)$, of a VAR in first differences if the series are $I(1)$ but not cointegrated, or, if the series are $I(1)$ and cointegrated, on the VECM residuals.[9] The applicability of the linear BDS and H-J (1994) tests employed by B-R (2015), therefore, also rest upon the confidence that can be placed upon the reliability of the order of integration of the series they reported from the UR test results.

Long-run causality, linearly or nonlinearly, is always best established through cointegration, at least at first, and should the series display a mixed order of integration, the safest testing option would be that of employing the ARDL bounds testing approach (see Pesaran and Shin, 1999; and Pesaran et al., 2001) which unlike the J-J (1990) method does not require all the system's variables to be $I(1)$. A nonlinear extension of the ARDL model (NARDL)

[8] The inconsistency of the H-J (1994) test stems from a bias that cannot be removed simply by data filtering or choosing a smaller bandwidth. The Diks and Panchenko (2006) test reduces this bias by correcting the over-rejection problem of the H-J algorithm under the null hypothesis (see also Yu et al., 2015).

[9] In some cases nonlinear causality tests are also run on GARCH filtered data, to account for time-varying volatilities (e.g., Yu et al., 2015).

is now also available (see Shin et al., 2014) to test for possible nonlinear asymmetries when it is not known with certainty whether the regressor(s) are $I(1)$ or $I(0)$. In the absence of evidence of cointegration, linearly or nonlinearly, for robustness, nonlinear causality tests can still be performed, but only by accounting properly for the order of integration of the underlying series in the specification of the underlying VAR model.

3. Verification by 'pure replication' (negative Type 1)

As noted earlier, we use exactly the same data as B-R (2015), which they provided. For the Ng-Perron (2001) test, they only report the statistics for the variables in first difference, with and without trend (their Table 1, p. 153). Our duplication of the results of the Ng-Perron (2001) test (see Table 1) is able to reproduce *identically* B-R (2015) estimated statistics for the variables in first difference (see shaded area of Table 1 below), evidence which confirms that our replication experiment was done correctly. However, significantly, for the series in levels, our results reveal that *RIX* is actually stationary in the 'constant only' model ('without trend'), and so are *ROL* and *RIX* in the model 'with constant and trend'.

[Tables 1 and 2 here]

Although our replication of the N-P (2010) UR test does not unveil any additional objective or inferential discrepancies (see our Table 2), it is evident that – taken collectively – the results of these two UR tests display, at best, mixed evidence. In the presence of doubts about the order of integration of the variables, Bal and Rath's choice of proceeding to estimation via the J-J (1990) cointegration approach, which requires all the system's variables to be integrated of order one, is evidently misplaced and, as a result, all the subsequent results they obtain on this premise are to be considered unreliable. This is particularly so given the weight B-R (2015:152) place upon inferences drawn from the Ng-Perron (2001) test results:

> Unit root test results from Table 1 indicate that the variables ROL (real oil price), RIX (real exchange rate in India) and RCX (real exchange rate in China) are integrated to order one. Accordingly, the study tested for the existence of a long-term relationship among these variables. For this purpose, the study performed the Johansen and Juselius (1990) cointegration test.

In rationalizing the fundamental discrepancy unveiled by our *verification* of the Ng-Perron (2001) test results (Table 1), we claim that: (i) even in the best case of 'good-faith oversight', we are in the presence of an irrefutable 'misrepresentation by omission'; (ii) there is something indisputably wrong in the inferences and conclusions drawn from the results of this test in the original paper, leading us to conclude that the paper fails the replication test (negative Type 1).

Although we stop reporting our pure replication here, we found additional non-trivial issues in the subsequent analysis reported by B-R (2015). For example, identical results of the J–J (1990) cointegration test for India and China (their Tables 3 and 4, respectively) can only be generated, in both cases, at a lag length equal to one, which contrary to what is stated in their paper (see Notes of Tables 3 and 4, on p. 153), does not reflect any "optimum value" determined by either the Akaike Information Criterion (AIC) or Schwartz Bayesian Criterion (SBC) at any reasonable selection level for the order of the maximum lags. Furthermore, we found that *identical* results to those they reported for what should be the corresponding error correction models (ECMs) presented in their Tables 5 and 6 (p. 154), can only be generated through a specification under the EViews menu option 'intercept (no trend) in cointegrating equation and test VAR' that uses residuals that are actually different from those produced by the long-run cointegrating equations ('intercept and trend in cointegrating equation – no intercept in VAR' option in EViews) they presented in Tables 3 and 4 (p. 153), respectively. This is clearly a flawed procedure evidencing an inconsistent transition from the estimation of the cointegrating long-run models to what should be their *corresponding* ECMs.

4. A reanalysis (negative replication Type 2)

In this *reanalysis* we conduct some robustness tests to investigate further – using the same variables and the original data over the same sample period – the linear and nonlinear unit root properties of the series used by B-R (2015), and the possible existence of any linear and/or nonlinear cointegration or Granger causality by means of alternative testing techniques. The aim here is to establish what results would have been produced had the original researchers adopted a more congruent and accurate testing protocol that also includes novel nonlinear methods.[10]

We begin by probing further the results produced by the N-P (2010) test in the context of B-R (2015). To this end, we run alternative UR tests with two structural breaks, namely the Lee and Strazicich (2003) and Carrion-i-Silvestre et al. (2009) tests.[11] Table 3 presents the results of the former test according to which *RCX* contains a unit root, while *RIX* and *ROL* are level stationary (thus corroborating the *actual* results of the Ng-Perron test with constant and trend we unveiled in Table 1). The results of the more powerful Carrion-i-Silvestre et al. (2009)

[10] According to Clemens (in press, p. 7): "Robustness tests descriptively establish what would have happened if the original researchers had not done *X*; only replication tests normatively claim that the original researcher *should not* have done *X*". We conduct the former now, having performed the latter in the previous section.

[11] We ruled out the Lumsdaine and Papell (1997) test since although it represents an extension of the Zivot and Andrews (1992) UR test by accounting for two breaks, akin to the latter, allows for the break(s) only under the alternative hypothesis. Nunes et al. (1997) show that this assumption can lead to size distortions and Perron (2005) adds that there may be a loss of power. We employ the Lee and Strazicich (2003) and Carrion-i-Silvestre et al. (2009) tests since they are more advanced and powerful as they allow for two breaks in level and trend under both the null and the alternative hypothesis.

test (Table 4) indicate that *RCX* and *RIX* are $I(1)$ while *ROL* is, once again, confirmed to be $I(0)$.

[Tables 3 and 4 here]

Both of the above tests assume that the break magnitudes are fixed. Harvey et al. (2013) (henceforth HLT) argue that these tests can display low finite sample power for the magnitudes of trend breaks. To address this concern they developed a new test that is based on the infimum of the sequence - across all candidate break points - of local GLS detrended augmented Dickey–Fuller-type statistics. HLT (2013) show that this new test has superior power and is robust to any break magnitude.[12] Table 5 presents the results for the minimum DF statistics with one and two breaks. We find that in both the one break in trend case (MDF_1) and the two breaks in trend (MDF_2), the null of a unit root is rejected for *ROL* while it cannot be rejected for *RCX* and *RIX*.[13]

[Table 5 here]

We now move on to testing the linear properties of the individual series, a 'pre-check' exercise that we argue B-R (2015) should have performed at the outset. Table 6 reports the results of both the W^* linearity test statistics of Harvey and Leybourne (2007) and the W_λ linearity test statistic of Harvey et al. (2008).[14] For *ROL* both W_λ and the W^* statistics reject the null of linearity at any reasonable significance level, yet both tests are unable to reject the null for *RCX* and *RIX*. Accordingly, we should proceed by employing a nonlinear UR test for *ROL* since a linear test may lack power if the true process is nonlinear.

[Tables 6 and 7 here]

The most popular nonlinear UR test is the one by Kapetanios et al. (2003) who proposed a modified ADF regression in which the null of a unit root is tested against the alternative of a globally stationary exponential smooth transition autoregression (ESTAR). The test is based on

---

[12] Our purpose in applying this test is to check whether the stationarity of *ROL* proves robust to a model specification that allows for one and two breaks in trend of different magnitudes. Note also that, as recently emphasized by Liddle and Messinis (2015: 281): "while the HLT (2013) test only considers trend shifts, it is based on GLS detrending, and thus, is asymptotically robust to level breaks (or 'slowly evolving trends')".

[13] We should note that although past studies report mixed findings as to the order of integration of the international price of oil, previous evidence pointing to a unit root cannot, of itself, be taken to lend any support to Bal and Rath's (2015) reported results. First, because all of the previous studies that test the international price of oil series, do so in its logarithmic form (which is not the "*ROL*" measure specification tested by B-R 2015), and over different sample periods. Second, because in our re-analysis of the order of integration of the "ROL" series considered by B-R (2015), we employ considerably more advanced linear UR tests with breaks (as well as recently developed nonlinear UR tests) than those typically used in past literature (see, e.g., Amano and van Norden, 1998).

[14] Harvey et al. (2008) extension of the Harvey and Leybourne (2007) test is based upon a data-dependent weighted average of two Wald statistics. The first Wald statistic is efficient when the series are $I(0)$ while the second when the series are $I(1)$. Harvey et al. (2008) show that this weighted statistic has better finite sample size properties and greater power compared to Harvey and Leybourne (2007). For comprehensiveness we apply both.

a function of the type $\Delta y_t = \varphi y_{t-1}\left(1-\exp\left\{-\gamma\left(y_{t-1}-c\right)^2\right\}\right)+\varepsilon_t$, where γ is the smoothness parameter and c is the location parameter, which is assumed to be zero. Kruse (2011) recently extended the Kapetanios et al. (2003) test by relaxing the restrictive assumption of a zero location parameter.[15]

Table 7 presents the results of the Kruse (2011) nonlinear UR test for the *ROL* variable applied on the raw data, the demeaned and the detrended series. Following the SBC, we choose a lag length of one after estimating a linear autoregressive test regression.[16] In addition, given that the lag structure may affect the test statistics, we employ up to 5 lags to avoid potential problems of serial correlation. At different lag lengths the results are quite similar, pointing to the rejection of the null of a unit root. This is an important result since the stationarity of *ROL* raises considerable doubts on the reliability of inferences based on any estimations carried out on regressions that include this series within bivariate model specifications requiring all the variables to be integrated of order one, such as the J-J (1990) method used by B-R (2015).

Overall, the results from the barrage of UR tests applied, consistently indicate that *ROL* is level stationary, linearly as well as nonlinearly, while both *RCX* and *RIX* appear to be, most likely, integrated of order one (and first-difference stationary).

Having established with greater confidence the order of integration of the series considered by B-R (2015), we can proceed to test for cointegration using, first, the ARDL bounds testing approach (Pesaran and Shin, 1999; Pesaran et al., 2001). This methodology has several virtues. First, unlike the cointegration method employed by B-R (2015), which requires the restrictive assumption that all the system's variables are integrated of order one for valid inference, it allows for the analysis of long-run (level) relationships when it is not known with certainty whether the regressors are $I(1)$ or $I(0)$. Second, as reiterated by Fuinhas and Marques (2012: 512), if the relations investigated prove to be cointegrated, "this assures the presence of causality and its direction" (though this feature is not exclusive to the ARDL bounds testing approach). Finally, the ARDL-based estimator addresses – thanks to the rich set of dynamics of the ARDL specification - the potential endogeneity problem whilst simultaneously correcting for residual serial correlation (see, e.g., Pesaran and Shin, 1999; Panopoulou and Pittis, 2004; Narayan and Smyth, 2006).

---

[15] Kruse (2011) considers the following modified ADF regression: $\Delta y_t = \varphi y_{t-1}\left(1-\exp\left\{-\gamma\left(y_{t-1}-c\right)^2\right\}\right)+\varepsilon_t$. By first-order Taylor approximation, the following auxiliary regression is obtained: $\Delta y_t = \beta_1 y_{t-1}^3 + \beta_2 y_{t-1}^2 + \sum_{i=1}^{p}\rho_i \Delta y_{t-i} + \varepsilon_t$. The Kruse (2011) modified Wald statistic $\tau = t^2_{\beta_2^{\perp}=0} + 1\left(\hat{\beta}_1 < 0\right) t^2_{\beta_1=0}$ is then used to test the null of a unit root against the alternative of a globally stationary ESTAR process:

$$H_0 : \beta_1 = \beta_2 = 0$$
$$H_1 : \beta_1 < 0,\ \beta_2 \neq 0 .$$

[16] In nonlinear STAR modelling it is common to select the optimal lag length by considering a linear AR model.

The standard linear ARDL(p,q) bounds test model with two series y_t and x_t (t = 1, 2, ..., T) has the form: $\Delta y_t = \mu + \rho y_{t-1} + \theta x_{t-1} + \sum_{j=1}^{p-1} \alpha_j \Delta y_{t-j} + \sum_{j=0}^{q-1} \pi_j \Delta x_{t-j} + u_t$, where u_t is an *i.i.d.* stochastic process. The existence of a stable long-run relation can be tested by any of the following three statistics. First, the modified F-test (F_{PSS}) advanced by Pesaran et al. (2001), which tests the joint null hypothesis of no cointegration $\rho = \theta = 0$. Second, a Wald-test (W_{PSS}), which also tests the above joint null. Third, a t-test (t_{BDM}) proposed by Banerjee et al. (1998), which tests the null of no cointegration $\rho = 0$ against $\rho < 0$. The testing procedure uses two critical bounds: upper and lower. If the values of the F_{PSS}, W_{PSS} or t_{BDM} statistics exceed the upper bound, the null hypothesis is rejected. If they lie below the lower critical bound, the null cannot be rejected, and if they lie between the critical bounds, the test is inconclusive.

Table 8, Panel A, presents the ARDL bounds test results for the 'constant only' model. From the F_{PSS}, W_{PSS} and t_{BDM} statistics, at the customary 5% significance level, we find no evidence in support of linear cointegration across the four tested pairs since none of the statistics exceed the upper critical value bound. With the exception of the RCX / ROL pair, the linear ARDL representation also displays poor diagnostics. In search for the possible existence of a statistically robust cointegrating relationship pertaining to the bivariate model posited by B-R (2015), we also consider an ARDL specification 'with constant and trend'. As shown in Table 8, Panel B, we find that the only two pairs of variables displaying some evidence of cointegration, namely ROL / RIX and ROL / RCX , are the ones that fail to pass both the test for normality and homoskedasticity. These diagnostic failures invalidate the reliability of any finding of cointegration from which evidence of long-run causality can be inferred, possibly suggesting an underlying problem of omitted variables in the simple bivariate relationship proposed by B-R (2015). Pesaran and Pesaran (2009) also suggest applying the cumulative sum of recursive residuals (CUSUM) and the CUSUM sum of squares (CUSUMSQ) tests to assess parameter constancy. As can be seen from Fig.1 and Fig.2, the pairs ROL / RIX and ROL / RCX also fail to pass the more powerful CUSUMSQ test since the respective plots exceed the 5% critical bounds.

[Table 8, Fig. 1 and Fig. 2 here]

Next, we investigate whether there may be any reliable evidence of *nonlinear* cointegration characterizing the pairs of variables by applying the NARDL approach of Shin et al. (2014). NARDL is a new technique that allows to model asymmetric effects both in the long-run and the short-run by exploiting partial sum decompositions of the explanatory variable. The general NARDL(p,q) model has the form

$$\Delta y_t = \mu + \rho y_{t-1} + \theta^+ x_{t-1}^+ + \theta^- x_{t-1}^- + \gamma z_t + \sum_{j=1}^{p-1} \alpha_j \Delta y_{t-j} + \sum_{j=0}^{q-1} (\pi_j^+ \Delta x_{t-j}^+ + \pi_j^- \Delta x_{t-j}^-) + u_t ,$$

where $x_t^+ = \sum_{j=1}^{t} \Delta x_{j=1}^+ = \sum_{j=1}^{t} \max(\Delta x_j, 0)$, $x_t^- = \sum_{j=1}^{t} \Delta x_{j=1}^- = \sum_{j=1}^{t} \min(\Delta x_j, 0)$, $\theta^+ = -\rho / \beta^+$ and $\theta^- = -\rho / \beta^-$. β^+ and β^- are the asymmetric long-run parameters associated with positive and negative changes in x_t, respectively, with the long-run equilibrium relation being expressed as: $y_t = \beta^+ x_t^+ + \beta^- x_t^- + u_t$.

The existence of an asymmetric long-run relationship between the levels of the series y_t, x_t^+ and x_t^-, can be tested - similarly to the linear ARDL bounds testing procedure - by the F_{PSS} and W_{PSS} statistics under the joint null of no cointegration $\rho = \theta^+ = \theta^- = 0$, and the t_{BDM} statistic which tests the null of no cointegration $\rho = 0$. Where nonlinear cointegration is confirmed, the next step is to test the null of long-run symmetry ($\beta^+ = \beta^-$) and short-run (*additive*) symmetry ($\sum_{i=0}^{q-1} \pi_i^+ = \sum_{i=0}^{q-1} \pi_i^-$), by using standard Wald tests.

[Table 9 and Fig. 3 here]

Table 9 presents the NARDL tests for all the pairs considered and, where applicable, the symmetry tests and long-run coefficients. For the pair *RIX* to *ROL* the F_{PSS}, W_{PSS} and t_{BDM} statistics reject the null of 'no cointegration' at any reasonable level of significance (Table 9, column 'a').[17] The Wald tests reject the null of long-run symmetry (W_{LR}) but fail to reject the null of short-run symmetry (W_{SR}). Greenwood-Nimmo et al. (2013) argue that in cases where long- or short-run symmetry turns out to be consistent with the data, the general NARDL model should be re-estimated with the respective symmetry condition imposed in order to avoid potential misspecification. Accordingly, we re-estimate the NARDL model for *RIX* to *ROL* with the short-run symmetry condition imposed. These estimates are presented in column 'b' of Table 9. The null of 'no cointegration' is rejected by two out of three test statistics, while the Wald test also indicates the rejection of long-run symmetry. Moving forward, we estimated the long-run asymmetric coefficients (β^+ and β^-) for the pair *RIX* to *ROL* from the optimal NARDL model as indicated by the above process. The estimated coefficients are statistically significant. However, the model fails the functional form, normality and homoskedasticity tests. Such poor diagnostics lead us to conclude that the finding of cointegration (and by implication, causality) is spurious, and most likely symptomatic of a serious problem of omitted variable bias. Furthermore, as shown in Fig. 3, the pair *RIX* to *ROL* also fails to pass the CUSUMSQ test.

For the pair *ROL* to *RIX*, all three test statistics reject the null of 'no cointegration' at any reasonable significance level (Table 9, column 'c'). Additionally, the Wald tests fail to reject the null of long-run symmetry but not the null of short-run symmetry. As per the above procedure, we re-estimate the NARDL model with the long-run symmetry condition imposed (Table 9, column 'd'). Once again, the null of 'no cointegration' is rejected, while the Wald test

[17] Following Shin et al. (2014), we adopt a conservative approach to the choice of critical values (i.e., a higher critical value) by employing $k = 1$.

also indicates the rejection of short-run symmetry. However, the estimated long-run symmetric coefficient (β) is statistically insignificant, thereby denoting no long-run causality.

Turning our attention to the cointegration statistics for the pairs *RCX* to *ROL* and *ROL* to *RCX* (Table 9, columns 'e' and 'f'), all three and two out of three statistics, respectively, do not reject the null of 'no cointegration', with good diagnostics in the case of the pair $RCX \,/\, ROL^{+}\ ROL^{-}$, evidence which, once again, suggests the absence of nonlinear causality.

As a final robustness test, unable to make use of residuals from any robust linear or nonlinear cointegrating equation for the variables considered, we apply both the H-J (1994) and the Diks and Panchenko (2006) (henceforth D-P) nonlinear causality tests on a properly specified VAR that by accounting for the actual integration properties of the individual series (i.e., *ROL* as an $I(0)$ variable and *RIX* and *RCX* as $I(1)$), ensures the stationarity of the residuals. The nonlinear causal linkages between the variables are investigated in two ways. First, both nonlinear causality tests are applied on delinearized series within the properly specified VAR model (see 'Panel A' of Table 10). This process ensures that any causality identified is solely nonlinear in nature. Second, following Bekiros and Diks (2008b), to account for time varying volatilities we repeat both tests on GARCH (1, 1) filtered VAR-residuals ('Panel B' of Table 10).[18]

[Table 10 here]

As shown in Table 10, in the case of both Panel A and Panel B, we find univocal evidence pointing to the complete absence of nonlinear Granger causality in three out of the four pairs of variables, namely *RIX* to *ROL*, *ROL* to *RCX* and *RCX* to *ROL*. Furthermore, for the *ROL* to *RIX* pair both the H-J (1994) and the D-P (2006) tests reject the null of 'no nonlinear Granger causality' only when three lags are considered and, even then, only at the 10% significance level. At all other lag lengths, also for this pair, both tests and under both the VAR and GARCH (1, 1) model specifications provide consistent evidence of no nonlinear causality. Overall, the observed 'non rejections' of the null, do not lend support to the hypothesis of nonlinear Granger causal relations between the variables in any direction, for both countries (negative replication Type 2).

[18] Whilst some of the older papers that applied only the H-J (1994) test did so on the GARCH filtered series, in implementing these nonparametric/nonlinear test procedures we follow closely the developers of the more advanced Diks and Panchenko (2006) methodology by accounting for time varying volatilities using GARCH (1, 1) filtered VAR-residuals. This second moment filtering procedure applied to filtered VAR-residuals (a process which ensures the delinearization and stationarity of the series tested for), is in line with the one employed by Diks and Panchenko (2006: 1660), Bekiros and Diks (2008a: 2682), Bekiros and Diks (2008b: 1647), and Bampinas and Panagiotidis (2015: 6).

5. Concluding remarks

We revisit the recent evidence published in this journal by B-R (2015) on the nonlinear causality between the global (real) oil price and the exchange rate for both India and China first, by conducting a 'pure replication' and then, after raising a number of critical issues, using a battery of additional linear and nonlinear tests on the same data set to re-analyze the order of integration of the variables as well as their cointegration and causality properties.

Contrary to what was reported by B-R (2015), when we repeat their estimated model with their method on their data we find that *ROL* (their measure of real crude oil price) is stationary in levels using the Ng-Perron (2001) unit root test they themselves used, a result which makes all the subsequent results of their analysis biased and misleading (negative replication Type 1).

Our *reanalysis* of the same data but with more congruent testing techniques appears to confirm that *ROL* is level stationary, linearly as well as nonlinearly. B-R (2015) also purport to unveil bilateral nonlinear Granger causality between the Indian exchange rate and the international price of oil and, for China, unidirectional causality running from exchange rate to oil price. Notwithstanding our concerns regarding the simple bivariate model specification they consider, since cointegration must entail causality, we tested the robustness of their findings through ARDL and NARDL methods that, unlike the J-J (1990) approach, do not require the restrictive assumption that all the variables are integrated of the same order. We find that the bivariate specifications proposed by B-R (2015) fail to produce statistically robust and stable cointegrating patterns from which any linear or nonlinear long-run causality can be inferred, pointing to a serious problem of omitted variable bias. Furthermore, application of the H-J (1994) and D-P (2006) nonlinear causality tests on the residuals of a properly specified VAR model to account for the order of integration of the series, and on GARCH (1, 1) filtered VAR-residuals, confirm the absence of nonlinear causality in any direction, for both countries (negative replication Type 2).

Why does it all matter? To return to the energy economics / energy finance rationale that motivated our interest in B-R (2015), invalidating their results matters because it nullifies the validity of the economic implications that would flow from their findings, namely that in modeling and forecasting the international price of oil, there is a need to take into account the exchange rate of India and China. The determination of global oil prices is an important economic issue of considerable interest to policy makers in countries around the globe in terms, for example, of aspects related to oil demand inventories and oil risk management. Oil producers and investors as well as financial portfolio managers are, therefore, advised to take note of the findings of our replication evidence when modeling oil price determinants in their crude oil price projections.

Our replication study also serves as a powerful reminder that in empirical economic research properly conceived and theory-based model building, the adoption of congruent testing procedures, and accurate statistical inference, are critical steps to obtain reliable estimates from which relevant findings and associated implications about policy and market behavior can be drawn. The increasingly popular strand of applied econometrics energy literature concerned with linear and nonlinear unit root testing, causality and cointegration, should be no exception.

References

Aksoy, Y., Leon-Ledesma, M.A., 2008. Non-linearities and unit roots in G7 macroeconomic variables. The B.E. J. Macroecon. 8, 1–44.

Amano, R.R., van Norden, S. 1998. Oil prices and the rise and fall of the US real exchange rate. J. Int. Mon. Fin. 17, 299–316.

Arulampalam, W., Hartog, J., MaCurdy, T., Theeuwes, J., 1997. Replication and re-analysis. Lab. Econ. 4, 99–105.

Baek, E., Brock, W.A., 1992. A general test for nonlinear Granger causality: bivariate model. Working Paper. Iowa State University of Wisconsin, Madison.

Bal, D.P., Rath, B.N., 2015. Nonlinear causality between crude oil price and exchange rate: a comparative study of China and India. Energy Econ. 51, 149–156.

Bampinas, G., Panagiotidis, T. 2015. On the relationship between oil and gold before and after financial crisis: linear, nonlinear and time-varying causality testing. Stud. Nonlinear Dyn. Econ. 19, 1–12.

Banerjee, A., Dolado, J., Mestre, R., 1998. Error-correction mechanism tests for cointegration in a single-equation framework. J. Time Series Anal. 19, 267–283.

Bekiros, S.D., Diks, C.G.H., 2008a. The relationship between crude oil spot and futures prices: cointegration, linear and nonlinear causality. Energy Econ. 30, 2673–2685.

Bekiros, S.D., Diks, C.G.H., 2008b. The nonlinear dynamic relationship of exchange rates parametric and nonparametric causality testing. J. Macroecon. 30, 1641–1650.

Brock, W.A., Dechert, W., Scheinkman, J., 1987. A test for independence based on the correlation dimension. Working paper, University of Winconsin at Madison, University of Houston, and University of Chicago.

Burman, L.E., Reed, W.R., Alm, J., 2010. A call for replication studies. Public Finan. Rev. 38, 787–793.

Carrion-i-Silvestre, J.L., Kim, D., Perron, P., 2009. GLS-based unit root tests with multiple structural breaks under both the null and the alternative hypotheses. Econom. Theory 25, 1754–1792.

Clemens, M.A., in press. The meaning of failed replications: a review and proposal. J. Econ. Surv. doi: 10.1111/joes.12139.

De Vita, G., Abbott, A., 2002. Are saving and investment cointegrated? An ARDL bounds testing approach. Econ. Lett. 77, 293–299.

De Vita, G., Endresen, K., Hunt, L.C., 2006. An empirical analysis of energy demand in Namibia. Energy Policy 34, 3447–3463.

Dergiades, T., Martinopoulos, G., Tsoulfidis, L. 2013. Energy consumption and economic growth: parametric and non-parametric causality testing for the case of Greece. Energy Econ. 36, 686–697.

Diks, C., Panchenko, V., 2005. A note on the Hiemstra–Jones test for Granger noncausality. Stud. Nonlinear Dyn. Econom. 9 (2, Article 4).

Diks, C., Panchenko, V., 2006. A new statistic and practical guidelines for nonparametric Granger causality testing. J. Econ. Dyn. Control 30, 1647–1669.

Fuinhas, J.A., Marques, A.C., 2012. Energy consumption and economic growth nexus in Portugal, Italy, Greece, Spain and Turkey: an ARDL bounds test approach (1965–2009). Energy Econ. 34, 511–517.

Granger, C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, 424–438.

Greenwood-Nimmo, M., Shin, Y., 2013. Taxation and the asymmetric adjustment of selected retail energy prices in the UK. Econ. Lett. 121, 411–416.

Greenwood-Nimmo, M., Shin, Y., Van Treeck, T., Yu, B., 2013. The decoupling of monetary policy from long-term rates in the US during the great moderation. Working Paper.

Hamermesh, D.S., 1997. Some thoughts on replications and reviews. Lab. Econ. 4, 107–109.

Harris, D., McCabe, B.P., Leybourne, S.J., 2003. Some limit theory for autocovariances whose order depends on sample size. Econom. Theory 19, 829–864.

Harris, R., 1995. Using Cointegration Analysis in Econometric Modelling. Prentice Hall, London.

Harvey, D. I., Leybourne, S. J., 2007. Testing for time series linearity. Econom. J. 10, 149–165.

Harvey, D. I., Leybourne, S.J., Xiao, B., 2008. A powerful test for linearity when the order of integration is unknown. Stud. Nonlinear Dyn. Econom. 12 (Article 2).

Harvey, D.I., Leybourne, S.J., Taylor, A.M.R., 2013. Testing for unit roots in the possible presence of multiple trend breaks using minimum Dickey–Fuller statistics. J. Econom. 177, 265–284.

Hassani, H., Zhigljavskya, A.A., Patterson, K., Soof, A.S., 2010. A comprehensive causality test based on the singular spectrum analysis. In: Illary, P.M., Russo, F., Williamson, J. (Eds.), Causality in Science. Oxford University Press, Oxford.

Herndon, T., Ash, M., Pollin, R., 2014. Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff. Cambridge J. Econ. 38, 257–279.

Hiemstra, C., Jones, J.D., 1994. Testing for linear and nonlinear Granger causality in the stock price volume relation. J. Financ. 49, 1639–1664.

Johansen, S., Juselius, K., 1990. Maximum likelihood estimation and inference on cointegration with applications to the demand for money. Oxf. Bull. Econ. Stat. 52, 169–210.

Kapetanios, G., Shin, Y., Snell, A., 2003. Testing for a unit root in the nonlinear STAR framework. J. Econom. 112, 359–379.

Kruse, R., 2011. A new unit root test against ESTAR based on a class of modified statistics. Stat. Papers 52, 71–85.

Lee, J., Strazicich, M.C., 2003. Minimum LM unit root test with two structural breaks. Rev. Econ. Stat. 85, 1082–1089.

Lee, J., Strazicich, M.C., 2013. Minimum LM unit root test with one structural break. Econ. Bull. 33, 2483–2492.

Liddle, B., Messinis, G., 2015. Revisiting sulfur Kuznets curves with endogenous breaks modeling: substantial evidence of inverted-Us/Vs for individual OECD countries. Econ. Modell. 49, 278–285.

Lumsdaine, R.L., Papell, D.H., 1997. Multiple trend breaks and the unit root hypothesis. Rev. Econ. Stat. 79, 212–218.

Lutkepohl, H., 1982. Non-causality due to omitted variables. J. Econom. 19, 367–378.

Mishra, V., Smyth, R., 2014. Is monthly US natural gas consumption stationary? New evidence from a GARCH unit root test with structural breaks. Energy Policy 69, 258–262.
models. Econom. J. 2, 76–91.

Narayan, P.K., Liu, R., 2013. New evidence on the weak-form efficient market hypothesis. Working Paper. Centre for Financial Econometrics, Deakin University.

Narayan, P.K., Popp, S., 2010. A new unit root test with two structural breaks in level and slope at unknown time. J. Appl. Stat. 37, 1425–1438.

Narayan, P.K., Smyth, R., 2006. Dead man walking: an empirical reassessment of the deterrent effect of capital punishment using the bounds testing approach to cointegration. Appl. Econ. 38, 1975–1989.

Narayan, P.K., Smyth, R., 2009. Multivariate Granger causality between electricity consumption, exports and GDP: evidence from a panel of Middle Eastern countries. Energy Policy 37, 229–236.

Nazlioglu, S. 2011. World oil and agricultural commodity prices: evidence from nonlinear causality. Energy Policy 39, 2935–2943.

Ng, S., Perron, P., 2001. Lag selection and the construction of unit root tests with good size and power. Econom. 69, 1519–1554.

Nunes, L., Newbold, P., Kuan, C., 1997. Testing for unit roots with breaks: evidence on the great crash and the unit root hypothesis reconsidered. Oxford Bull. Econ. Stat. 59, 435–448.

Panopoulou, E., Pittis, N., 2004. A comparison of autoregressive distributed lag and dynamic OLS cointegration estimators in the case of a serially correlated cointegration error. Econom. J. 7, 585–617.

Perron, P., 2005. Dealing with structural breaks. Mimeo. Boston University.

Pesaran, B., Pesaran, M. H., 2009. Time Series Econometrics using Microfit 5.0. Oxford University Press, New York.

Pesaran, M.H., Shin, Y., 1999. An autoregressive distributed lag modeling approach to cointegration analysis. In: Strøm, S. (Ed.), Econometrics and Economic Theory in the

20th Century: The Ragnar Frisch Centennial Symposium. Cambridge University Press, Cambridge.

Pesaran, M.H., Shin, Y., Smith, R.J., 2001. Bounds testing approaches to the analysis of level relationships. J. Appl. Econom. 16, 289–326.

Rahbek, A., Mosconi, R., 1999. Cointegration rank inference with stationary regressors in VAR

Shin, Y., Yu, B.C., Greenwood-Nimmo, M., 2014. Modelling asymmetric cointegration and dynamic multipliers in a nonlinear ARDL framework. In: Sickles, R.C., Horrace, W.C., (Eds.), The Festschrift in Honor of Peter Schmidt: Econometric Methods and Applications. Springer, New York.

Şimşek, N., 2014. Long Run Evaluation of Crude Oil Prices: Structural Breaks, Long Memory or Nonlinearity? Eurasian Research Institute, Almaty.

Smyth, R., Narayan, P.K., 2015. Applied econometrics and implications for energy economics research. Energy Econ. 50, 351–358.

Triacca, U., 1998. Non-causality: the role of the omitted variables. Econ. Lett. 60, 317–320.

Yu, L., Li, J., Tang, L., Wang, S., 2015. Linear and nonlinear Granger causality investigation between carbon market and crude oil market: a multi-scale approach. Energy Econ. 51, 300–311.

Zivot, E., Andrews, D.W.K., 1992. Further evidence on the great crash, the oil price shock and the unit root hypothesis. J. Bus. Econ. Stat. 10, 251–270.

Table 1
Ng-Perron (2001) unit root tests.

Ng-Perron statistics (constant only)

Variable	MZ_a	MZ_t	*MSB*	*MPT*	Inference
ROL	-3.01125 (1)	-1.02234 (1)	0.33951 (1)	7.77943 (1)	Non-stationary
RCX	1.00726 (1)	0.79733 (1)	0.79159 (1)	46.4714 (1)	Non-stationary
RIX	-19.5325*** (1)	-3.06554*** (1)	0.15695*** (1)	1.47010*** (1)	**Stationary**
ΔROL	-99.9854*** (0)	-7.06663*** (0)	0.07068*** (0)	0.25249*** (0)	Stationary
ΔRCX	-105.636*** (0)	-7.26706*** (0)	0.06879*** (0)	0.23294*** (0)	Stationary
ΔRIX	-102.239*** (0)	-7.10117*** (0)	0.06946*** (0)	0.33085*** (0)	Stationary

Critical values (constant only)

1%	-13.800	-2.580	0.174	1.780
5%	-8.100	-1.980	0.233	3.170
10%	-5.700	-1.620	0.275	4.450

Ng-Perron statistics (constant and trend)

Variable	MZ_a	MZ_t	*MSB*	*MPT*	Inference
ROL	-26.1898*** (1)	-3.61857*** (1)	0.13817*** (1)	3.48010*** (1)	**Stationary**
RCX	-4.20585 (1)	-1.43566 (1)	0.34135 (1)	21.5225 (1)	Non-stationary
RIX	-22.0725** (1)	-3.31093** (1)	0.15000** (1)	4.19754** (1)	**Stationary**
ΔROL	-100.586*** (0)	-7.08641*** (0)	0.07045*** (0)	0.92639*** (0)	Stationary
ΔRCX	-105.739*** (0)	-7.27029*** (0)	0.06876*** (0)	0.86494*** (0)	Stationary
ΔRIX	-108.871*** (0)	-7.36663*** (0)	0.06766*** (0)	0.87898*** (0)	Stationary

Critical values (constant and trend)

1%	-23.800	-3.4200	0.1430	4.0300
5%	-17.300	-2.9100	0.1680	5.4800
10%	-14.200	-2.6200	0.1850	6.6700

Notes: Δ is the first difference operator. The optimal lag structure is chosen using the SBC and displayed in parentheses. The critical values are from Ng and Perron (2001). The estimation and tests were conducted using EViews 9.0. *** and ** denote the rejection of the null of a unit root at the 1 and 5% significance level, respectively.

Table 2
Narayan and Popp (2010) unit root tests with two structural breaks.

Two breaks in level and slope

Variable	Test statistic	Break dates	φ	k
ROL	-2.502	2003M08; 2007M11	-0.2609	2
RCX	-4.059	2001M02; 2006M08	-0.4299	2
RIX	-3.013	1999M11; 2001M08	-0.3362	5
ΔROL	-5.115**	2006M03; 2007M07	-1.140	5
ΔRCX	-4.793**	2006M03; 2006M12	-0.9693	5
ΔRIX	-6.110***	1999M09; 1999M12	-1.401	5

Notes: Δ is the first difference operator. φ denotes the autoregressive coefficient and k the optimal lag order. The 1, 5 and 10% critical values are −5.138, -4.741 and -4.430, respectively. The critical values are from Narayan and Popp (2010). The estimation and tests were conducted using a program code written in GAUSS that was produced by Narayan and Popp (2010). *** and ** denote rejection of the null of a unit root at the 1 and 5% significance level, respectively.

Table 3
Lee and Strazicich (2003) unit root tests with two structural breaks.

Variable	Model A		Model C	
	LM test Statistic	Break dates	LM test Statistic	Break dates
ROL	-4.509** (2)	2004:09; 2009:06	-6.023*** (4)	2004:12; 2009:06
RCX	-1.511 (1)	1998:10; 2005:02	-4.622 (1)	1997:09; 2003:11
RIX	-4.598*** (8)	2003:06; 2010:06	-6.096*** (8)	1999:12; 2008:07
ΔROL	-9.796*** (0)	2006:02; 2008:11	-9.957*** (0)	2003:03; 2008:12
ΔRCX	-11.431*** (0)	1997:12; 2004:11	-11.456*** (0)	2002:02; 2009:03
ΔRIX	-4.187** (5)	2000:10; 2007:05	-12.789*** (0)	2007:11; 2010:12

Notes: Δ is the first difference operator. Model A allows for a change in level while Model C allows for changes in level and slope of the trend. The optimal lag structure, reported in parentheses, is chosen following a *general-to-specific* approach starting with max 12 lags (see Lee and Strazicich, 2013). The critical values are from Lee and Strazicich (2003). The estimation and tests were conducted using RATS 8.0. *** and ** denote rejection of the null of a unit root at the 1 and 5% significance level, respectively.

Table 4
Carrion-i-Silvestre et al. (2009) unit root tests with two structural breaks.

Variable	P_T^{GLS}	MP_T^{GLS}	MZ_a^{GLS}	MSB^{GLS}	MZ_t^{GLS}	Break dates
ROL	4.572**	3.882**	-44.268**	0.106**	-4.704)**	2000:11; 2008:07
	(6.037)	(6.037)	(-28.116)	(0.136)	(-3.722)	
RCX	9.095	8.877	-25.796	0.137	-3.556	1997:12; 2004:12
	(7.439)	(7.439)	(-30.050)	(0.128)	(-3.852)	
RIX	11.712	11.373	-18.078	0.166	-3.003	1996:02; 2008:01
	(7.306)	(7.306)	(-28.418)	(0.131)	(-3.759)	
ΔROL	3.208**	2.426**	-61.215**	0.090**	-5.521**	2006:04; 2008:10
	(5.590)	(5.590)	(-26.019)	(0.141)	(-3.580)	
ΔRCX	3.010**	2.920**	-76.542**	0.080**	-6.182**	1998:01; 2008:02
	(7.548)	(7.548)	(-29.416)	(0.129)	(-3.820)	
ΔRIX	3.039**	2.741**	-75.975**	0.080**	-6.148**	1996:02; 2008:05
	(7.236)	(7.236)	(-28.186)	(0.132)	(-3.745)	

Notes: Δ is the first difference operator. For the P_T^{GLS} , MP_T^{GLS} and MSB^{GLS} tests, the null hypothesis is rejected in favor of stationarity when the estimated value is smaller than the critical value. The 5% critical values, obtained from simulations using 1000 steps to approximate the Wiener process and 10000 replications, are displayed in parentheses. The optimal lag structure is chosen based on the modified AIC, starting with max 6 lags. The estimation and tests were conducted using a program code written in GAUSS that was produced by Carrion-i-Silvestre. ** denotes rejection of the null of a unit root at the 5% significance level.

Table 5
Harvey et al. (2013) unit root tests with one and two structural breaks

Variable	MDF_1	MDF_2
ROL	-4.247**	-4.884**
RCX	-1.582	-3.971
RIX	-3.063	-3.481
ΔROL	-7.996***	-8.010***
ΔRCX	-11.653***	-11.734***
ΔRIX	-4.293**	-4.343*
Critical values		
1%	-4.40	-5.10
5%	-3.85	-4.58
10%	-3.57	-4.30

Notes: Δ is the first difference operator. The critical values are from Harvey et al. (2013). The estimation and tests were conducted using a program code written in GAUSS that was produced by Harvey et al. (2013). ***, ** and * denote rejection of the null of a unit root at the 1, 5 and 10% significance level, respectively.

Table 6
Harvey and Leybourne (2007) and Harvey et al. (2008) linearity tests.

Variable	$W^*_{10\%}$	$W^*_{5\%}$	$W^*_{1\%}$	W_λ
ROL	13.77***	13.91***	14.18***	7.76**
RCX	4.01	4.03	4.09	1.97
RIX	5.55	5.57	5.60	1.74

Notes: The W_λ statistic follows the χ^2_2 distribution and the relevant critical values are 9.21 (1%), 5.99 (5%) and 4.60 (10%). The W^* statistic follows the χ^2_4 distribution and the relevant critical values are 13.27 (1%), 9.48 (5%) and 7.77 (10%). The estimation and tests were conducted using a program code written in GAUSS that was produced by Harvey et al. (2008). *** and ** denote the rejection of the null of linearity at the 1 and 5% significance level, respectively.

Table 7
Kruse (2011) nonlinear unit root test on oil price.

Lag(s)	Level series	Demeaned series	Detrended series
0	7.33	4.65	5.37
1[a]	11.75**	9.23*	11.31*
2	14.04***	11.39**	14.81**
3	13.33***	10.62**	14.59**
4	12.78**	10.10*	14.27**
5	11.57**	8.95*	11.68*
Critical values			
1%	13.15	13.75	17.10
5%	9.53	10.17	12.82
10%	7.85	8.60	11.10

Notes: The critical values are from Kruse (2011). [a] denotes the optimal lag length selected by the SBC. The estimation and tests were conducted using a program code written in 'R' that was produced by Kruse. ***, ** and * denote the rejection of the null of a unit root at the 1, 5 and 10% significance level, respectively.

Table 8
ARDL cointegration tests

Panel A: Constant only models

ARDL model	*ROL / RIX*	*RIX / ROL*	*ROL / RCX*	*RCX / ROL*
Specification	(2,0)	(2,0)	(2,1)	(2,1)
F_{PSS}	1.99	5.58	0.98	4.30
W_{PSS}	3.98	11.16	1.97	8.60
t_{BDM}	-1.90	-3.24	-1.19	-2.86
Diagnostics				
SC	14.09 [0.295]	20.26* [0.062]	14.05 [0.297]	11.70 [0.470]
FF	3.88** [0.049]	0.46 [0.497]	2.23 [0.135]	1.50 [0.220]
NOR	51.03*** [0.000]	24.59*** [0.000]	31.49*** [0.000]	2.67 [0.262]
HET	28.35*** [0.000]	0.35 [0.549]	27.46*** [0.000]	1.78 [0.181]

Panel B: Constant and time trend models

ARDL model	*ROL / RIX*	*RIX / ROL*	*ROL / RCX*	*RCX / ROL*
Specification	(2,0)	(2,0)	(2,0)	(2,1)
F_{PSS}	9.33**	5.55	9.47**	4.85
W_{PSS}	18.66**	11.10	18.95**	9.71
t_{BDM}	-4.38***	-3.19	-4.48***	-1.50
Time trend	0.018*** [0.000]	-0.001 [0.634]	0.021*** [0.000]	-0.003 [0.289]
Diagnostics				
SC	12.44 [0.411]	19.77* [0.071]	13.08 [0.363]	16.49 [0.169]
FF	0.20 [0.650]	0.41 [0.521]	1.68 [0.195]	1.94 [0.163]
NOR	26.74*** [0.000]	25.64*** [0.000]	31.29*** [0.000]	2.56 [0.277]
HET	31.32*** [0.000]	0.22 [0.636]	33.02*** [0.000]	2.21 [0.137]

Notes: The choice of the optimal linear ARDL specifications is based on the SBC, starting with max $q =$ max $p = 18$. At the 5% significance level, the pair of critical values (bounds) for the F_{PSS} and the W_{PSS} statistics are 5.02 to 5.79 and 10.05 to11.58, respectively, for the constant only models, and 6.75 to 7.39 and 13.50 to 14.79, respectively, for the constant and time trend models. The 5% critical value (bounds) for the F and W statistics are computed by stochastic simulations using 20,000 replications in Microfit 5.0 (Pesaran and Pesaran, 2009). At the 5% significance level, the pair of critical values bounds for the t_{BDM} statistic are -2.86 to -3.53, for the constant only models, and -3.41 to -3.95, for the constant and time trend models (taken from Pesaran et al., 2001). *SC*, *FF*, *NOR* and *HET* denote LM tests for serial correlation, functional form, normality and homoskedasticity, respectively. p-values are reported in square brackets. ***, ** and * denote statistical significance at the 1, 5 and 10% level, respectively.

Table 9

NARDL cointegration tests, symmetry tests and long-run coefficients.

NARDL model	*ROL* / *RIX*⁺ *RIX*⁻		*RIX* / *ROL*⁺ *ROL*⁻		*ROL* / *RCX*⁺ *RCX*⁻	*RCX* / *ROL*⁺ *ROL*⁻
Statistic	LR asymmetry and SR asymmetry (a)	LR asymmetry and SR symmetry (b)	LR asymmetry and SR asymmetry (c)	LR symmetry and SR asymmetry (d)	LR asymmetry and SR asymmetry (e)	LR asymmetry and SR asymmetry (f)
F_{PSS}	9.71***	5.41	7.82**	11.16***	3.14	4.59
W_{PSS}	29.15***	16.23***	23.46***	26.65***	9.43	13.78**
t_{BDM}	-4.12***	-3.63**	-4.38***	-4.42***	-2.65	-2.16
W_{LR}	64.12*** [0.000]	56.82*** [0.000]	1.36 [0.242]	-	-	-
W_{SR}	0.003 [0.955]	-	4.93** [0.026]	5.07** [0.024]	-	-
β	-	-	-	0.06 [0.420]	-	-
β^{+}	-	1.48*** [0.001]	-	-	-	-
β^{-}	-	1.16*** [0.007]	-	-	-	-
Diagnostics						
SC	10.43 [0.578]	10.85 [0.542]	8.30 [0.761]	8.01 [0.784]	13.34 [0.344]	19.95* [0.068]
FF	23.78*** [0.000]	18.97*** [0.000]	1.12 [0.289]	1.15 [0.282]	15.77*** [0.000]	0.09 [0.761]
NOR	0.97 [0.615]	10.32*** [0.006]	8.30** [0.016]	8.33** [0.015]	13.61*** [0.001]	1.29 [0.522]
HET	21.61*** [0.000]	12.41*** [0.000]	2.68 [0.101]	3.30* [0.069]	19.56*** [0.000]	0.01 [0.917]

Notes: β is the estimated symmetric long-run coefficient defined by $\hat{\beta} = -\hat{\theta}/\hat{\rho}$. β^{+} and β^{-} are the estimated asymmetric long-run coefficients defined by $\hat{\beta}^{+} = -\hat{\theta}^{+}/\hat{\rho}$ and $\hat{\beta}^{-} = -\hat{\theta}^{-}/\hat{\rho}$, respectively. Following Greenwood-Nimmo and Shin (2013) and Shin et al. (2014), the lag length in each case was selected through a *general-to-specific* approach, starting with $max\ q = max\ p = 18$ and then dropping insignificant regressors with a 5% unidirectional decision rule. For $k = 1$ and at the 1% (5%) significance level, the pair of critical values (bounds) for the F_{PSS}, W_{PSS} and t_{BDM} statistics are 6.84 to 7.84 (4.94 to 5.73), 14.11 to 15.63 (9.86 to 11.52) and -3.43 to -3.82 (-2.86 to -3.22), respectively. The critical values are from Pesaran et al. (2001) and Pesaran and Pesaran (2009). *SC*, *FF*, *NOR* and *HET* denote LM tests for serial correlation, functional form, normality and homoskedasticity, respectively. *p*-values are reported in square brackets. ***, ** and * denote significance at the 1, 5 and 10% level, respectively.

Table 10
Heimstra and Jones (1994) and Diks and Panchenko (2006) nonlinear Granger causality tests.

Panel A: with VAR filtered residuals

	India				China			
	ROL ≠> *RIX*		*RIX* ≠> *ROL*		*ROL* ≠> *RCX*		*RCX* ≠> *ROL*	
Lx = *Ly*	HJ	DP	HJ	DP	HJ	DP	HJ	DP
1	-0.432	-0.435	-0.700	-0.811	0.789	0.850	-1.524	-1.509
	[0.667]	[0.668]	[0.758]	[0.791]	[0.214]	[0.197]	[0.936]	[0.934]
2	1.219	1.115	-0.602	-0.806	-0.157	-0.047	-1.701	-1.545
	[0.111]	[0.132]	[0.726]	[0.790]	[0.562]	[0.519]	[0.955]	[0.938]
3	1.559*	1.608*	0.436	0.228	-1.571	-1.565	-2.395	-2.263
	[0.059]	[0.053]	[0.331]	[0.409]	[0.941]	[0.941]	[0.991]	[0.988]
4	1.151	1.167	0.476	0.255	-1.171	-0.480	-2.278	-1.942
	[0.124]	[0.121]	[0.316]	[0.399]	[0.879]	[0.684]	[0.988]	[0.973]
5	1.243	1.134	0.334	0.006	-0.482	0.120	-1.780	-1.411
	[0.106]	[0.128]	[0.369]	[0.497]	[0.685]	[0.451]	[0.962]	[0.921]
6	0.994	0.901	-0.052	-0.148	-0.231	0.553	-1.917	-1.304
	[0.159]	[0.183]	[0.520]	[0.559]	[0.591]	[0.290]	[0.972]	[0.903]
7	0.809	0.630	-1.331	-1.014	-0.933	-0.364	-1.922	-1.268
	[0.209]	[0.264]	[0.908]	[0.844]	[0.824]	[0.642]	[0.972]	[0.897]
8	0.810	0.667	-1.611	-1.090	-0.831	-0.315	-2.156	-1.515
	[0.208]	[0.252]	[0.946]	[0.862]	[0.797]	[0.623]	[0.984]	[0.935]

Panel B: with GARCH[1,1] filtered VAR-residuals

	India				China			
	ROL ≠> *RIX*		*RIX* ≠> *ROL*		*ROL* ≠> *RCX*		*RCX* ≠> *ROL*	
Lx = *Ly*	HJ	DP	HJ	DP	HJ	DP	HJ	DP
1	-0.404	-0.487	-0.702	-1.003	0.713	0.658	-1.396	-1.412
	[0.656]	[0.687]	[0.758]	[0.842]	[0.237]	[0.255]	[0.918]	[0.921]
2	0.999	0.909	-0.610	-0.952	-0.156	-0.055	-1.457	-1.303
	[0.158]	[0.181]	[0.729]	[0.829]	[0.562]	[0.522]	[0.927]	[0.903]
3	1.405*	1.442*	0.585	0.336	-1.322	-1.343	-2.419	-2.285
	[0.079]	[0.074]	[0.279]	[0.368]	[0.907]	[0.910]	[0.992]	[0.988]
4	1.032	1.076	0.853	0.728	-0.711	-0.011	-2.066	-1.469
	[0.150]	[0.140]	[0.196]	[0.233]	[0.761]	[0.504]	[0.980]	[0.929]
5	1.192	1.141	0.483	0.205	-0.179	0.408	-1.485	-0.927
	[0.116]	[0.126]	[0.314]	[0.418]	[0.571]	[0.341]	[0.931]	[0.823]
6	0.974	0.762	0.132	-0.020	-0.324	0.317	-1.531	-0.708
	[0.164]	[0.222]	[0.447]	[0.507]	[0.627]	[0.375]	[0.937]	[0.760]
7	0.768	0.497	-1.261	-1.022	-1.104	-0.367	-1.863	-1.233
	[0.221]	[0.309]	[0.896]	[0.846]	[0.865]	[0.643]	[0.968]	[0.891]
8	0.642	0.357	-1.319	-1.036	-0.817	-0.418	-1.848	-1.338
	[0.260]	[0.360]	[0.906]	[0.850]	[0.793]	[0.662]	[0.967]	[0.909]

Notes: HJ and DP denote the Heimstra and Jones (1994) and the Diks and Panchenko (2006) test, respectively. The values reported in Table 10 are the test statistics (T-values) and the p-values of the HJ and DP tests (see, e.g., Diks and Panchenko, 2006; Nazlioglu, 2011; Dergiades et al., 2013). Parameter C for the bandwidth is 8, the theoretical optimal rate β is 2/7, and the optimal bandwidth ε_n is 1.5 (like the bandwidth selected by Bal and Rath, 2015). The VAR lag order was selected using the SBC starting with max 12 lags. The selected VAR lag order is equal to two (univocal choice by SBC and AIC). p-values are reported in square brackets. The estimation and tests were conducted using a program code written in C language provided by Cees Diks. Superscript * denotes rejection of the null at the 10% significance level.

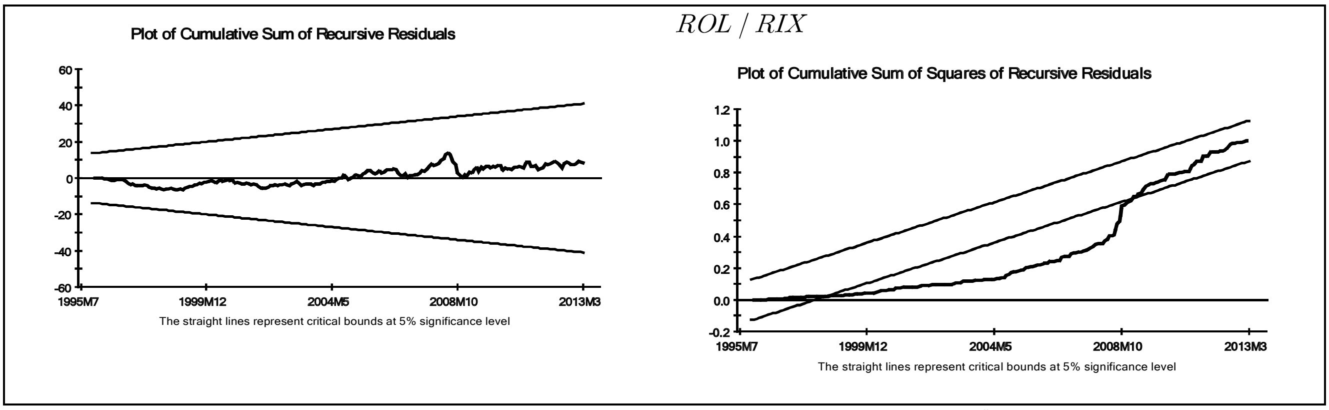

Fig. 1. CUSUM and CUSUMSQ tests on ARDL model ()

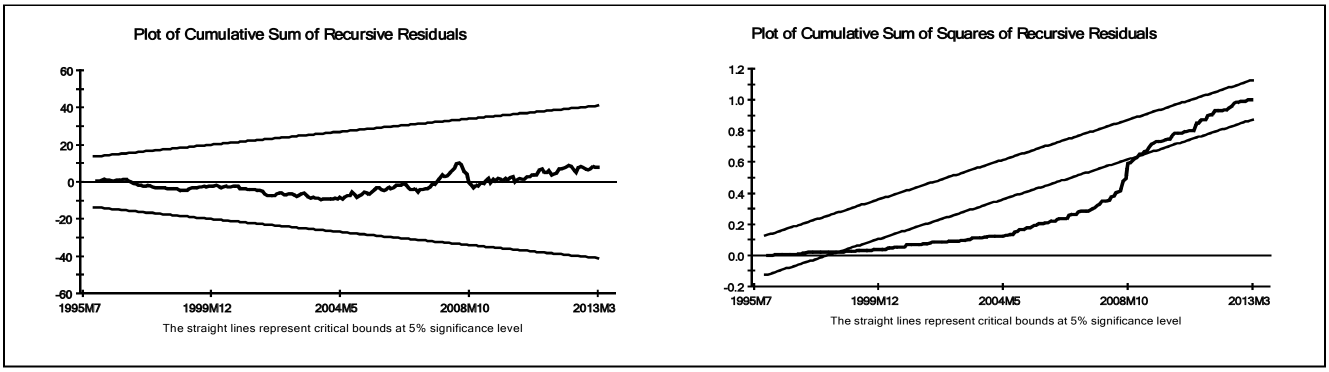

Fig. 2. CUSUM and CUSUMSQ tests on ARDL model (ROL / RCX)

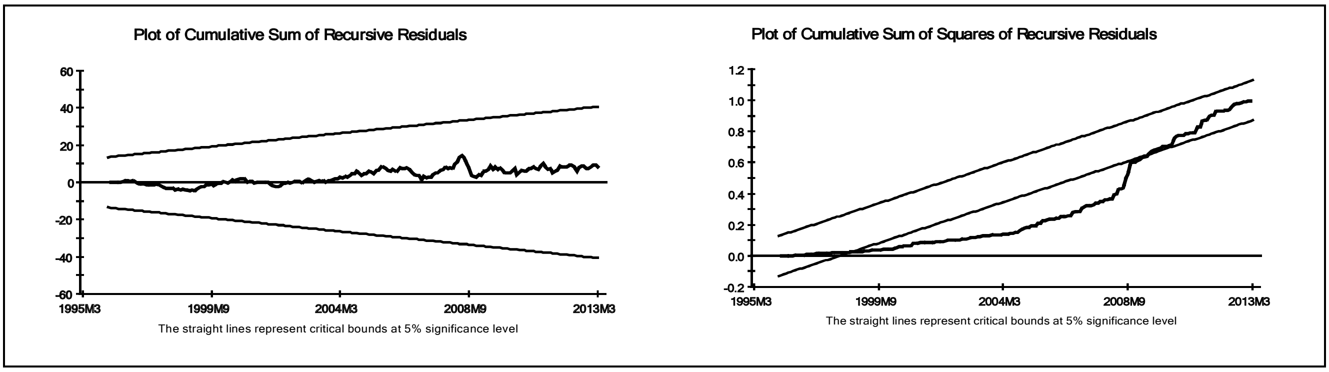

Fig. 3. CUSUM and CUSUMSQ tests on NARDL model (ROL / RIX^{+} RIX^{-})