



GROWTH IN A TIME OF AUSTERITY: EVIDENCE FROM THE UK

Juergen Amann* and Paul Middleditch**

ABSTRACT

This paper uses an empirical approach to test the specific causal relationship between debt and growth in the UK, in the context of the debate surrounding the use of a policy known as austerity measures. This time series perspective makes use of more recent Granger causality and cointegration tests that allow for non-stationarity in macroeconomic time series data in the presence of structural breaks. Controlling for exogenous shocks associated with the period around the financial crisis, we find no evidence of a causal relationship between economic growth and public debt for the UK.

I INTRODUCTION

The European sovereign debt crisis of 2010 brought an increase in uncertainty for global financial markets already unnerved by the financial crisis that had begun a couple of years earlier. A concern with the ability of the governments across the Eurozone to service their public debt gave significant contribution to a second pan-European slump in economic activity. In July of the same year, the incoming UK coalition government came into power during an environment requiring immediate measures to calm financial markets. In order to achieve this, the freshly elected coalition announced a programme that has come to be known as *austerity*, a commitment to oversee the long-term reduction of public spending as a percentage of GDP.¹ Since this time and some years forward, we have seen growth return to more normal levels in the UK, raising the question of whether or not the recent recovery can be attributed in a direct way to the change in fiscal discipline.

This paper is motivated by the seminal contribution of Reinhart and Rogoff (2010) whose study has focused a debate around the issue of fiscal discipline and its relation to the rate of growth in income. This influential study found a direct link between public debt and economic growth, more specifically the

*The University of Nottingham

**University of Manchester

¹ See Konzelmann (2014) for an in-depth discussion on the concept of austerity as a measure of fiscal discipline and how this measure has transformed from being a means to achieve macroeconomic stabilisation to becoming an objective in its own right.

existence of a debt threshold (of 90%) at which economic growth is significantly impeded. In an environment of surging public debt and crumbling growth rates, international organisations and policy-makers have found their own interpretation of studies such as this to legitimise rigorous public spending cuts (Minea and Parent, 2012). Consequently, the effectiveness and legitimacy of the policy of austerity has been widely discussed in both the public, economic and political arenas. The findings in Reinhart and Rogoff (2010) have also provoked an extensive discussion in the field of applied economics and are not without reinforcement, in fact several other panel-type studies such as Cecchetti *et al.* (2011), Casni *et al.* (2014), Baum *et al.* (2013) and Woo and Kumar (2015) offer fair to mixed support of the debt-to-GDP threshold hypothesis.

Contrary to this, authors such as Minea and Parent (2012) have found that the coefficients between debt and growth predict a positive relation, Panizza and Presbitero (2014) found no statistical causal link and Kourtellos *et al.* (2013) a similar relation in terms of debt thresholds that only exists in low democracy countries. The Reinhart and Rogoff (2010) study has also been challenged on technical grounds through the work of Herndon *et al.* (2014) who replicated the study and discovered coding errors, selective exclusion of data and unconventional weighting methods. Controlling for these issues, their replication study finds that the effect of the 90% debt-to-GDP threshold on growth becomes neglectfully small. A recent study by Puente-Ajovín and Sanso-Navarro (2015) makes use of panel Granger causality tests across 16 OECD countries allowing for cross-country heterogeneity and cross-sectional dependence. Using annual data over a 30-year time span ending in 2009, they find little evidence of a causal relation between debt and growth for all the countries examined. In the UK, in particular and in line with Bell *et al.* (2015), they find no evidence of a causal relation between government debt and growth.

In this paper, we statistically test the proposition that there is a direct relation between the level of public debt and the rate of GDP growth, motivated by the natural experiment brought about by the UK government's policy of austerity announced in 2010. We do this by applying a time series causality testing framework to a bidirectional system of GDP growth and the same public debt measure used in the literature. Our approach deviates from the main literature in that it offers a time series perspective that focusses on the UK, uses high-frequency data, comfortably spans the financial crisis from 1995M01 to 2013M12 and puts special emphasis on the presence of structural breaks in our testing framework. For our empirical investigation, we first make robustness checks for structural interventions using endogenous break-point and unit root test methods to eliminate structural bias in our analysis. Secondly, employing an augmented Toda and Yamamoto (1995, TY procedure from now on) Granger causality testing procedure and the Johansen *et al.* (2000) cointegration framework that both allow for the presence of structural breaks, we draw a conclusion on the bidirectional relation between debt and growth.

Firstly, we find that there is consistency across tests predicting two important break dates for the debt-to-GDP ratio for the UK showing severe changes in the level in 2008 and 2010. Secondly, we find evidence of a bidirectional Granger causality relationship between the two variables by employing the TY procedure before controlling for these structural breaks. However, when accounting for the breaks determined in our earlier testing procedure, our results suggest no causal relationship in either direction for the two variables, a finding that confirms the importance of controlling for interventions in the testing procedure. These results are supported in a sub-sample analysis carried out for the purpose of robustness. Finally, we proceed with a break point-augmented cointegration test introduced by Johansen *et al.* (2000) and find further support for the denial of any causal relation when accounting for structural instabilities. Accordingly, we conclude that there is no evidence to support the use of changes in the debt-to-GDP ratio to influence economic activity and that policy actions solely concerned with restoring economic growth, through a reduction in the sovereign debt ratio, are an oversimplification of a more complex matter.

It is worth noting here what might be considered a design limitation of our approach; given the complexity of the interaction between common macroeconomic variables it is impossible to rule out a higher order/dimensional relation between the level of debt and growth, an issue discussed in Lütkepohl (1982). If growth is not caused by debt in a Granger sense then, in a purely bivariate system, past information on debt cannot be used to provide better forecasts of growth; but in multivariate terms we cannot make the same conclusion if there exist other variables that can influence the relationship between debt and growth in any way. Essentially, there exists possible bias from omitted variables in the standard Granger causality regression. In this study, we focus our efforts on the direct relation between debt and growth which allows us to contribute in the context of the common debate on austerity and to further the analysis beyond the measurement of correlation on to one of causality.

The rest of this paper is laid out as follows: In Section II, we consider the claims made in Reinhart and Rogoff (2010) by giving a short introductory and descriptive exploratory data analysis motivated by Amann and Middleditch (2015). Acknowledging the structural effects of the financial crisis reflected in recent UK time series data, we put special emphasis on the consideration of structural breaks in our testing framework in Section III. In doing so, we employ various unit root tests allowing for (multiple) endogenously determined structural breaks. In Section IV, we apply a modified flexible Granger causality testing framework allowing for different orders of integration among the variables in question introduced by Toda and Yamamoto (1995) which we augment to allow for structural breaks. Furthermore, we employ the Johansen *et al.* (2000) cointegration test, also allowing for structural breaks. Section 5 concludes with motivation for further research.

II PRELIMINARY DISCUSSION

What follows is a description and visual analysis of the data used in the seminal paper by Reinhart and Rogoff (2010) presented alongside our own monthly equivalent to provide some background on key objections in the literature surrounding this debate; namely the intuitive observation that, from a time series perspective, increases in debt appear to follow economic slumps, see Amann and Middleditch (2015) for a more detailed discussion of the evidence behind this reverse causality hypothesis, namely that a fall in economic activity will cause an increase in debt rather than vice versa.

The data

We use three different data sets in this paper. Our main data set is a specially derived, more recent monthly data set for the UK covering the period from 1995M01 to 2013M12. This monthly UK time series makes use of higher frequency data from the UK in order to evaluate the relationship between economic growth and public debt. For the UK monthly GDP growth rates we use an estimated monthly GDP index provided by the National Institute of Economic and Social Research (NIESR) that follows the methodology described in Mitchell *et al.* (2005). For monthly data on public debt, *Public Sector Net Debt* (PSND) published by the Office of National Statistics (ONS) is used. For the exploratory data analysis we compare this monthly data with the corrected data set of the Reinhart and Rogoff (2010) study provided by Herndon *et al.* (2014).² Furthermore, we demonstrate the robustness of our results by applying a selected set of tests to the quarterly OECD data series *general government total gross debt* as well as *quarterly growth rates of real GDP* from the OECDStats.StatExtracts data base.³

Exploratory data analysis

The scatter plot in Figure 1(a) depicts the debt-to-GDP ratio vs the GDP growth rate in percentage and considers the corrected Reinhart and Rogoff (2010) data set as well as quarterly OECD data and our monthly UK data set which we derived for this paper. For the purpose of this figure, we define two debt regimes: The first one is associated with the debt threshold of a gross government debt-to-GDP ratio (also DoG from now on) of above 90% as postulated by Reinhart and Rogoff (2010), and the second below 90%. In addition to that, we also define a positive growth and a negative growth regime associated with a GDP growth rate (also GR from now on) of $\pm 0\%$ to evaluate any changes in GDP growth rates with respect to the corresponding debt regime as well as any changes in the debt-to-GDP ratio considering the given growth regime.

² Code and data of the Herndon *et al.* (2014) study is provided at <https://www.peri.umass.edu/publication/item/526-does-high-public-debt-consistently-stifle-economic-growth-a-critique-of-reinhart-and-rogooff> (last visit March 2017).

³ We provide more information on our data sets in Appendix A.

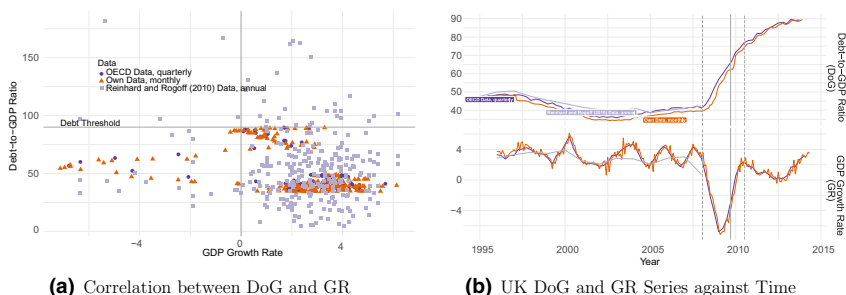


Figure 1. Exploratory data series analysis. (a) Correlation between DoG and GR
 (b) UK DoG and GR Series against Time. [Colour figure can be viewed at wileyonlinelibrary.com]

The horizontal line on the scatter plot reflects the 90% debt threshold level and the vertical line corresponds to a 'recession line' at a GDP growth level of 0%. Following the line of argument in Reinhart and Rogoff (2010) with public debt over 90% impeding growth, we would expect to find a higher density of data points in the upper left and lower right quadrant of the correlogram in Figure 1(a). Instead, the plot reveals anything but a negative relation between the two variables of interest for either series. Figure 1(b) plots the corrected annual Reinhart and Rogoff (2010) data for the UK as well as the quarterly OECD data and the monthly UK data against time. In all cases the impact of the financial crisis 2007/08 can be seen very clearly. Although the GDP growth rate experiences a sharp setback with quick recovery afterwards, the debt-to-GDP ratio was subject to a severe shift in levels due to the crisis. Initially we can see that for the earlier periods the time series plot presents little evidence of a clear correlation, in fact there seems to exist a time-dependent rather than debt-level dependent relation for the UK:

From 2008 – the first dashed, vertical line in Figure 1(b) corresponds to March 2008 – we see a sudden and severe increase in the debt-to-GDP ratio coinciding with a sharp fall in the rate of growth. However, this is just as severely reversed in 2009; the data reflecting a continuation of the explosion in debt coupled with a sharp increase in the UK growth rate. A breakdown of the corrected Reinhart and Rogoff (2010) data, shown in Amann and Middleditch (2015), reveals a similar pattern for many other countries: Relatively low debt levels and moderate growth rates characterising the late 1990s to mid-2000s were followed by declines in GDP growth rates and, subsequently, a rise in the gross debt levels. The financial crisis is to be marked as an exceptional event, rather than following a path predicted as a consequence of higher debt levels.

As the above explanatory analysis suggests, it was not the high level of public debt that brought GDP growth to a decline, it was the financial crisis that triggered changes in both macro variables and led to a significant increase in DoG. If indeed the postulated negative threshold link between gross debt and GDP growth existed, then for the case of the UK, which experienced a level shift in gross debt to slightly above 90% due to the crisis, the present high debt level should have had a more negative effect on GDP growth. That is, one would not expect GDP growth to return to a similar rate to that of before the crisis.

In light of the above statement, we now investigate the link between DoG and GR by means of an initial data analysis making use of our monthly UK data series spanning from 1995M01 to 2013M12 to closely analyse the possibility of any causal link between the variables in question, beginning with an investigation into the structural stability of the relationship between the two variables of interest.

III STRUCTURAL BREAKS

Testing for structural breaks in both series

Given that our sample encompasses the period around the financial crisis of 2008, we will naturally want to make some preliminary investigation of the data, specifically to test for structural change. For this purpose, this study uses a global maximiser test following a technique introduced by Bai and Perron (1998, 2003b) to determine the dates of any structural breaks in the DoG and GR series. The authors use the following multiple linear regression model in T periods with m unknown structural breaks, and thus $m + 1$ regimes:

$$y_t = X_t' \beta + Z_t' \delta_j + \varepsilon_t. \quad (1)$$

Note that X_t represents the set of regressors whose parameters are time invariant and Z_t the j regime-specific variables. This study is only concerned with pure structural change and consequently we only consider the restricted model with no X_t variables. For a specific set of m breakpoints, the test minimises the sum of squared residuals in the above linear regression model. A null hypothesis of no breaks, equivalent to $\delta_0 = \delta_1 = \dots = \delta_{i+1}$ where $1 \leq i \leq m$ against an alternative of l breaks, where l is pre-specified as a maximum number of all possible breaks, which in our test is set at 2. The general form of the Bai and Perron (1998, 2003b) test statistic is given as:

$$F(\hat{\delta}) = \left(\frac{T - (i+1)q - p}{Tk_q} \right) (R\hat{\delta})' (R\hat{V}(\hat{\delta})R')^{-1} R\hat{\delta} \quad (2)$$

where $\hat{V}(\hat{\delta})$ is the estimated covariance matrix of the optimal l break estimate and R a matrix such that $(R\hat{\delta}) = (\delta'_1 - \delta'_2 \dots \delta'_k - \delta'_{k+1})$. The distribution of the test statistics is non-standard and so they are compared to the critical values provided by Bai and Perron (2003a).

Table 1 reports the statistics for the sequentially determined structural breaks for the variable DoG and the globally maximised Bai and Perron (1998, 2003b) test for the variable GR. The reason for the global test on GR was that the sequential test failed to achieve one break and thus the possibility of two breaks or more required utilisation of the global version. When comparing Table 1 with the earlier time series plot in Figure 1 we see that our quantitative analysis is largely consistent with our observations from the data.

The tests predict two breaks for each variable and thus three regimes partitioned in 1991M01 and 2009M06 for DoG and 2008M03 and 2011M01 for GR. The null hypothesis for the sequential test is rejected for zero vs. one

Table 1
Bai and Perron (1998, 2003b) multiple break point test

Breakp	Breaks	F-Statistic	Crit. val.	Estimated break dates
DoG	0 vs 1†	92.110	8.58	2009M07
	1 vs 2†	19.153	10.13	1999M01, 2009M06
	2 vs 3	4.458	11.14	–
GR	1	1.742	8.58	2008M05
	2†	26.04	7.22	2008M03, 2011M01

Notes: For DoG we use L-1 vs. L sequentially determined breaks. For GR we test for global L breaks vs none with 0.15 trimming and a maximum of five breaks. Significant results on the 5% level of significance designated †. Values are rounded.

break or one break vs. two but cannot be rejected for two vs. three breaks. For the global test, the null hypothesis cannot be rejected for just one break but is rejected in favour of two breaks; The number of breaks seems reasonable given the sample size (two breaks in two decades) and considering that the sample encompasses the financial crisis.

The break date, 2009M07, is identified as the most likely date for the variable DoG when allowing for only one break. This is of no surprise, as this date reflects the ‘mid-point’ in the considerable increase in government gross debt materialising through the financial crisis. For variable GR and allowing for two break points, the dates 2008M03 and 2011M01 are selected. The predicted breaks make economic sense as they flag the beginning and the end of the financial crisis with growth rates recovering after a severe slump in 2009.

Unit root tests

In a next step, we test the stationarity properties of our series. To begin with, both series are analysed neglecting the possibility of any structural break using traditional Augmented Dickey Fuller (ADF), Phillips–Perron (PP) and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests. This is done in order to point out the differences in results when (not) taking structural breaks into account. In doing so we also conduct the same analysis of two different sub-samples as defined by the Bai and Perron (1998, 2003b) test results: In addition to the full sample analysis we also conduct our empirical investigation on two sub-samples: Sub-samples pre- and post-crisis I splits the sample in 2009M07 and pre- and post-crisis II generates two sub-samples spanning the periods from 1995M01 to 2008M03 and from 2010M03 to 2013M12, respectively, which are also presented in Figure 1(b) for visual evidence.⁴ Unit root model specifications as well as the autocorrelation of the residuals were tested and the lag length automatically selected using the Schwarz Information

⁴ As we will demonstrate in the remainder of this section, the definition of these sub-samples is very robust. We draw this conclusion from the fact that the break point test results for the monthly data series are confirmed when applying the same set of tests to quarterly OECD data. This observation is backed by the set of unit root tests allowing for structural break(s) which we introduce in the next section and all consistently confirm the breaks suggested by the Bai and Perron (1998, 2003b) test for both data sets. This is shown in Table B13.

Criterion (SBC) for the ADF test. For the KPSS and PP test, the spectral estimation was conducted using a Bartlett kernel estimation with the automatic bandwidth selection through the Newey–West estimator.

The results of these tests are presented in Table B8 for the full sample and those for the truncated samples in Table B9 to Table B12: Whereas DoG is found to be I(2) employing standard unit root tests for the full sample and sub-sample I, it is stationary in first differences for sub-sample II. Similarly, GR is found to be I(1) for the full sample analysis and possibly I(2) for sub-sample I. We will show in the next paragraph that these results are highly questionable and hence provide empirical evidence of the crucial point made in Perron (1989) that structural changes and unit root processes are closely related. For this purpose, we employ unit root tests allowing for endogenously determined structural breaks and show that our previously derived conventional unit root tests are biased towards not rejecting H_0 , as we find strong statistical evidence of breaks in both our variables.

Unit root tests with endogenous break dates

We start our analysis with tests proposed by Perron (1997) as well as Zivot and Andrews (1992) allowing for one endogenous break. The motivation behind this is twofold. Firstly, as mentioned before, these tests are not biased towards the non-rejection of the null hypothesis as stressed in Perron (1997). Secondly, since these procedures can identify the possible presence of structural breaks, this group of tests provides useful information on parameter stability; for instance analysing whether a structural break in a certain variable is associated with a particular global event such as a war, an economic or political crisis or new regulations etc. This method is convenient in that it allows us to cross-validate the empirical findings of the Bai and Perron (1998, 2003b) tests. Additionally, as Perron (1997) noted, we avoid the shortcomings of only capturing the most significant break in each variable and answer the question of what would happen if one allowed for multiple break points: For this reason we also incorporate the unit root test proposed by Lee and Strazicich (2002) to allow for two endogenous structural breaks.

For the methodological discussion of the endogenous break point testing framework, we follow Patterson (2012) with a slightly different notation. We incorporate Perron (1989)'s 'combined model' which assumes a gradual break and construct a Dickey–Fuller test equation of the form

$$y_t = \mu + \beta t + \theta DI_t + \gamma DT_t + \vartheta D_t + \rho y_{t-1} + \sum_{i=1}^p \phi_i \Delta y_{t-i} + \varepsilon_t, \quad (3)$$

where, for a specific break data t^* , DI_t is defined as an intercept break $DI_t = 1(t > t^*)$, the trend break variable $DT_t = t - t^*$ and the one-time break dummy $D_t = 1(t = t^* + 1)$. Using the general form of equation (3), we now specify different models that, given the restrictions we impose on the trend and break parameters β, θ, γ and ϑ , which correspond to different assumptions regarding the nature of trend(s) and break(s) of our data and correspond to

Perron (1989)'s 'crash' and 'changing growth' model respectively. More specifically, we define

$$\text{Model A } y_t = \mu + \beta t + \theta DI_t + \vartheta D_t + \rho y_{t-1} + \sum_{i=1}^p \phi_i \Delta y_{t-i} + \varepsilon_t \quad (4)$$

$$\text{Model C } y_t = \mu + \beta t + \gamma DT_t + \rho y_{t-1} + \sum_{i=1}^p \phi_i \Delta y_{t-i} + \varepsilon_t,$$

where, in Model A, as we set the break trend coefficient of $\gamma = 0$, we test for a random walk with drift vs a trend stationary process with structural break. In Model C, we allow for a trend, with a break in the data by setting both the intercept break as well as the break dummy coefficients zero, that is $\theta = \vartheta = 0$, and test H_0 of a random walk with drift against H_a of a trend stationary process with a break in the trend. Finally, we also test for a Model B, representing a random walk with drift vs. a trend stationary process. This model features a break in the intercept and the trend which, by construction, is equivalent to the general formulation in equation (3) and is therefore not written out explicitly again.

The methodology described in Zivot and Andrews (1992) does not allow for a structural break under the null hypothesis when endogenously determining t^* based on the minimum t-statistic of the ADF test of a unit root. We see this as a shortcoming as failing to reject H_0 in this context does not imply the existence of a unit root but rather of a unit root without break. Given this limitation, we also incorporate the test procedure described in Perron (1997) which allows for slope changes under both the null and the alternative hypotheses and allows us to conveniently compare the previously obtained breakpoint test results for consistency.

Table 2 summarises the results of both these tests. As can be seen by comparing the results, both tests strongly indicate the presence of unit roots with endogenously determined structural breaks around the time of the financial crisis. The null hypothesis of non-stationarity with a break in intercept and trend cannot be rejected for the Perron (1997) test on DoG levels. Yet, in first differences, H_0 is rejected at the 1% level of significance for a break in the intercept.⁵ This confirms the previously formulated suspicion that the non-rejection of H_0 for DoG in first differences might be due to a structural break.

For the case of GR, which is considered to be most accurately described by a model only allowing for a break in the intercept, results indicate that this series may be stationary with one structural break in levels already. However, by the nature of the series and the results of the standard unit root tests without breaks, there is a strong presumption that this series may in fact be subject to a second break. This will be investigated in the next section. With respect to the timing of the structural break, both tests consistently identify

⁵ As stressed in Perron (1997) and implemented in studies such as Stern and Enflo (2013); if the test in levels follows the model allowing for an intercept and trend, then for conducting the subsequent test in first difference, any test should only include a break in the intercept. The reason for this is that by differentiating a dummy variable marking a permanent effect, an impulse dummy that corresponds to a regime change is created. This holds true for the Perron (1997), Zivot and Andrews (1992) as well as the Lee and Strazicich (2002) test.

Table 2

Results of Zivot and Andrews (1992) and Perron (1997) unit root tests

		Break in intercept		Break in intercept and trend	
		t-Statistic	Break Point	t-Statistic	Break Point
Zivot and Andrews (1992) Test					
DoG	Levels	-5.972	2008M05	-3.318	2005M01
	1st diff.	-6.214*	2008M04	-	-
GR	Levels	-5.457*	2007M09	-6.988*	2008M03
	1st diff.	-5.02†	2009M03	-	-
Perron (1997) Test					
DoG	Levels	-5.958*	2008M04	-3.353	2004M12
	1st diff.	-6.738*	2010M01	-	-
GR	Levels	-6.054*	2007M08	-7.832*	2008M02
	1st diff.	-6.122*	2009M02	-	-

Notes: Table contains t-statistic of both tests and the endogenously chosen break point. Significant results on the 1%, 5% and 10% level of significance designated *, † and ‡ respectively. Values are rounded.

the time of the financial crisis as the most promising candidate for a structural break, thereby confirming the results of the break point analysis presented in Table 1.

In order to verify the results of the previous unit root as well as break point tests, we employ the minimum Lagrange multiplier unit root test as proposed in Lee and Strazicich (2002) (also referred to as LM test from now on). The LM test is capable of endogenously determining two structural breaks in either the intercept, this is also referred to as the 'crash model', or the intercept and trend, also known as the 'break model' under H_0 and H_a . The implementation of the LM test follows Stern and Enflo (2013) who apply the LM test in its 'break model' form to levels of their variables of interest and the 'crash model' to the first differences of the same series for the same reason as laid out above.⁶

As can be seen in Table 3, both series are found to be integrated of order one with two structural breaks. Although the case is clear for DoG, for GR a borderline rejection of H_0 in levels at the 5% level is reported. Given these results it is safe to assume that GR is either (i) $I(0)$ with two breaks or (ii) $I(1)$ with two breaks due to the borderline rejection for the test in levels. Since no clear-cut results can be reported for this series, both variants are being investigated in the next step. More precisely, when assuming case (i), the TY procedure allowing one and two structural breaks is the appropriate choice. In case (ii), a cointegration test allowing for one and two structural breaks can be conducted. All these variants will be considered in the next step.

With respect to the timing of the structural breaks, the LM test results are consistent with previous findings: When testing in levels for variable DoG,

⁶ For a more detailed discussion on this test see either the original paper Lee and Strazicich (2002) or applications of the technique as in *exempli gratia* Stern and Enflo (2013) or Ferreira *et al.* (2014).

Table 3
Results of Lee and Strazicich (2002) Unit Root Tests

	Levels			1st Differences		
H_0 :	Unit root with two breaks			Unit root with two breaks		
H_1 : for DoG	Break model: trend-stationary with two breaks			Crash model: level-stationary with two breaks		
H_1 : for GR	Crash model: level-stationary with two breaks			Crash model: level-stationary with two breaks		
Variable	Test statistic	Break point 1	Break point 2	Test statistic	Break point 1	Break point 2
DoG	-4.013	2004M11	2009M12	-6.2629*	2002M08	2007M12
GR	-4.577†	2006M03	2010M04	-4.9113*	2002M02	2010M06

Notes: Significant values at the 1% and 5% level are designated * and † respectively.

Break Point 2 corresponds very closely to the previous break point analysis results. This is indeed very interesting as all tests seem to recognise the ‘pivot point’ of the crisis which is also reflected by the choice of the second break point for the variable GR tested in levels: Here, again, early 2010 is found to be a structural break. This date, in turn, also marks the turning point of the GDP growth rate (see Figure 1).

The important conclusion that we can draw from the above detailed unit root test exercise is that, firstly, structural breaks indeed play a crucial role in analysing the data set at hand. Not accounting for them can lead to wrong inference and add serious bias to the results of any further analysis. Coincidentally, this exercise provides further evidence for the arguments stressed in Perron (1989). Furthermore, using sophisticated unit root test procedures, we were able to show that the structural breaks in both series are an immediate consequence of the financial crisis and that, given the results of all three unit root tests with structural breaks together, both series are most likely to be integrated of order one with (a) structural break (s). Additionally, concerning the timing of these breaks, it was shown that the endogenously determined break points are selected quite consistently around the time of the financial crisis and also reflect the nature of both series.

This observation is furthermore confirmed when employing the same set of tests on quarterly OECD data. Table B13 shows that our observations regarding the break points hold irrespectively of the data set we used. There is an almost perfect overlap of suggested break points for the full range of tests employed on either the monthly or quarterly data considered in this paper. Consequently, given the fact that (i) the Bai and Perron (1998, 2003b) break point test results confirm the break points suggested by the endogenous break unit root tests, and vice versa, and (ii) that this holds true for our monthly UK data set as well as for a second, quarterly

OECD data series, we conclude that, for any further analysis, special emphasis has to be put on the correct incorporation of these structural breaks. We account for this observation in the subsequent sections by providing results of two causality testing frameworks which we apply to the full sample by accounting for the suggested break dates as well as by purposely omitting them. Additionally, we also implement the same set of tests for both sub-samples on which we elaborated in this section to demonstrate robustness of our results. We provide a detailed summary of our testing procedure alongside the suggested order of integration of each series in question in Table B7.

IV TESTING FOR STATISTICAL CAUSALITY

As widely discussed in the empirical literature, the standard Granger causality test has probable shortcomings of specification bias and spurious regression: If the variables under consideration are non-stationary and cointegrated, the standard Granger causal inference will be invalid. Granger causality tests can also lead to misleading results in the presence of structural breaks, see Bianchi (1998) also Lütkepohl (1982) or Rossi (2013). Bianchi (1998) discusses the following possibilities for Granger causality tests in the presence of regime shifts: One approach is for the original series to be ‘purified’ of the structural break, that is made stationary or ‘detrended’ by subtracting local mean values computed within the appropriate sub-samples/regimes before Granger causality tests can then be performed on the transformed series. Alternatively, these causality tests can be performed on structurally stable sub-samples, or exogenous dummy variables can be introduced in the VAR estimation to reflect the regime shifts included. Rossi (2006) shows that parameter instability in the standard Granger causality testing framework may lead to a failure to reject the null hypothesis of non-Granger causality even if the regressor does Granger-cause by construction. As we acknowledge that parameter stability is of primary concern, we begin the next section with a test for the structural stability of our system before discussing how structural breaks may affect inference for Granger causality as well as cointegration tests in the subsequent sections.

Furthermore, we acknowledge the fact that standard Granger causality tests have shortcomings for non-stationary and cointegrated series, a fact that motivates our choice of the Toda and Yamamoto (1995) procedure. This augmented VAR model is superior to the ordinary Granger causality test in that it does not require pre-testing for cointegrating properties of the system and thus avoids the potential bias associated with unit roots and cointegration tests. Hence, it can be applied to any system regardless of the order of possible integration or cointegration of the series. We see this as an advantage of our method as standard unit root tests can also be affected by structural breaks.

Testing for structural stability

For the sake of simplicity, we follow closely the notation used in the original test as discussed in Chu *et al.* (1995). We consider a simple linear model where $\{x_t, y_t\} = \{DoG_t, GR_t\}$ with time varying coefficient

$$y_t = x_t' \beta_t + \epsilon_t$$

and are interested in testing for parameter stability of the coefficient, that is, if $\beta_t = \beta_0 \forall t = 1, \dots, T$. Hence, following simplistic assumptions that $\{\epsilon_t\}$ is a normally distributed, *iid* error term with zero mean and known variance and independent of x_t , we can describe a break at an unknown point t^* as

$$y_t = \begin{cases} x_t' \beta_1 + \epsilon_t & t = 1, \dots, t^* \\ x_t' \beta_2 + \epsilon_t & t = t^* + 1, \dots, T \end{cases} \quad (5)$$

Following the rational in Chu *et al.* (1995), the OLS estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ of equation (5) should be reasonably 'close' to each other under the null hypothesis of structural stability. Consequently, when estimating the full sample as $\hat{\beta}_T = (\sum_{t=1}^T x_t x_t')^{-1} (\sum_{t=1}^T x_t y_t)$, we substitute for y_t as defined in equation (5), which allows us to perform a test of parameter stability of the form $\hat{\beta}_1(t^*) - \hat{\beta}_T$ where $\hat{\beta}_1(t^*) = (\sum_{t=1}^{t^*} x_t x_t')^{-1} (\sum_{t=1}^{t^*} x_t y_t)$. We evaluate the possibility of structural changes in the above test by analysing the moving sums of residuals over the full sample period in a data window chosen by a bandwidth parameter that we set to $h = 0.15$. In the scenario depicted above with a break in t^* in the coefficients, one would expect a deviation away from the zero mean of the residual function around the break point. Table 4 presents strong evidence of structural breaks in 2008M07 and 2010M03 which strongly supports our univariate break point test results.

Testing for Granger causality in the presence of unit root processes and structural breaks

In the presence of non-stationarity and a possible cointegrating relationship among variables, the causality test proposed in Granger (1969) is inadequate. Given this and because our analysis shows that structural breaks are a crucial feature of our data, we put special emphasis on the inclusion of these breaks. We employ a modified flexible Granger causality testing framework allowing for structural breaks and different orders of integration among the variables

Table 4
Recursive moving sums of residual test

Structural instabilities	<i>p</i> -value
2008M07, 2010M03	0.01*

Notes: Recursive moving sums of residual test with chosen bandwidth parameter of $h = 0.15$. Results are robust with regards to reasonable changes in the bandwidth. Significant results on the 1% level of significance designated * respectively. Values are rounded.

in question. This methodology was first introduced by Toda and Yamamoto (1995), and we augment this procedure by allowing for (multiple) exogenous structural breaks following the approach presented in Stern and Enflo (2013). To retain consistency, we apply the standard Toda and Yamamoto (1995) procedure to all sample periods (both sub-samples as well as the full sample). Furthermore, we employ the augmented model with break points in 2009M07 as well as 2008M03 and 2010M03 to analyse the entire sample more thoroughly.⁷

More precisely, we define our bivariate system as

$$\begin{aligned}
 DoG_t &= \sum_{k=0}^n (\zeta_k \Delta t_{k,t} + \delta_k t_{k,t}) \\
 &+ \sum_{i=1}^p \alpha_{1,i} DoG_{t-i} + \sum_{j=p+1}^{p+m} \alpha_{2,j} DoG_{t-j} \\
 &+ \sum_{i=1}^p \gamma_{1,i} GR_{t-i} + \sum_{j=p+1}^{p+m} \gamma_{2,j} GR_{t-j} + \varepsilon_{1,t} \\
 GR_t &= \sum_{k=0}^n (\zeta_k \Delta t_{k,t} + \delta_k t_{k,t}) \\
 &+ \sum_{i=1}^p \beta_{1,i} GR_{t-i} + \sum_{j=p+1}^{p+m} \beta_{2,j} GR_{t-j} \\
 &+ \sum_{i=1}^p \delta_{1,i} DoG_{t-i} + \sum_{j=p+1}^{p+m} \delta_{2,j} DoG_{t-j} + \varepsilon_{2,t},
 \end{aligned} \tag{6}$$

where $\alpha_{h,l}, \beta_{h,l}, \gamma_{h,l}, \delta_{h,l}$, $h = \{1,2\}$ and $l = 1, \dots, p, \dots, p+m$, are the coefficients of the system to be estimated for the variables DoG and GR respectively. Furthermore, p denotes the lag length selected through the optimal lag length criteria test plus additional lags added afterwards to eliminate any autocorrelation in the residuals which we determined by means of the Schwarz Information Criterion (SBC). Additionally, m corresponds to the maximum order of integration of the processes for each individual series. For example, if DoG and GR are both found to be I(1) then we set $m = 1$.

The first sum of both equations in (6) corresponds to the break point augmentation of the bivariate system, where n denotes the number of structural breaks allowed for in the data and Δ is the first-difference operator. Expanding the sum, t_0 is a simple linear trend and Δt_0 equals a constant term corresponding to simple intercept terms in both equations, α_0 and β_0 say respectively. This is equivalent to the set-up when not allowing for any structural breaks in the system and consequently setting $n = 0$. For $n \geq k > 0$, Δt_k

⁷ The specific break date of 2010M03 has been chosen for consistency with the other dates derived from the various break point tests. It fits with the earlier data partitioning and also with the time series plot in Figure 1. For robustness we have re-tested the closest derived date of the Lee and Strazich (2002) test of 2010M04 with no effect on the results of this analysis.

is equal to a vector containing zeros up to and including the break point period t^* , and unity afterwards. Consequently, t_k equals zero up to and including the break point period and is increasing in unit steps afterwards.

For each test, the assumed value of m can be verified in the last column of Table 5 or Table B7; please see table note for more details. For cases where no definite conclusion on the order of integration of the variables was reached, different values for the maximum order of integration were used to cross-validate the results. These scenarios, however, change neither the direction nor the statistical evidence presented in Table 5.

To test H_0 of non-Granger causality of GR_t on DoG_t we set

$$H_0 : \gamma_{1,i} = 0 \forall i = 1, \dots, p$$

where $\gamma_{1,i}$ are the p lagged coefficients of GR_{t-i} , $i = 1, \dots, p$, in the first equation in (6). The exogenous coefficients $\gamma_{2,j}$ for $j = p + 1, \dots, p+m$ are not considered for this purpose. The test for non-causality of DoG_t on GR_t is calculated in a similar fashion.

With the model outlined above it is now worth emphasising that we are explicitly imposing the assumption of linearity in our empirical model. Even though it may be appealing for future research to elaborate on alternative, non-linear derivations from the set-up outlined above; in line with other research in this field such as Chen *et al.* (2010), the motivation of this piece is not to evaluate and compare different model alternatives but rather to empirically test the time-dependent linkage of public debt and economic growth in a linear setting. One way to control for the issue of linearity might be to use a change-point model as in Koop and Potter (2007) or an infinite hidden Markov model as in Song (2014) which would allow the inclusion of both regime switching and structural break dynamics to exist in one test together, in other words, a more unified approach.

Table 5

Results of granger-causality tests following the TY procedure

	DoG \Rightarrow GR	GR \Rightarrow DoG	Proposed value for m
Full sample			
No breaks in either series	0.0218†	0.0019*	2
One break in each series	0.0261†	0.0077*	1
Two breaks in each series	0.3415	0.1613	1
Sub-sample			
Split in 2009M07			
Pre-crisis I	0.016†	0.082‡	2
Post-crisis I	0.587	0.581	2•
Split in 2008M03 and 2010M03			
Pre-crisis II	0.914	0.884	1
Post-crisis II	0.544	0.872	1•

Notes: Table 5 reports p -values of the Wald tests. Arrows denote the direction of Granger causality. Significant results on the 1%, 5% and 10% level of significance designated *, † and ‡ respectively. Values are rounded. Results where no definite conclusion on the order of integration was derived are designated • or • for $m = \{0,1\}$ or $m = \{1,2\}$ respectively.

Results of the Toda and Yamamoto (1995) procedure

Table 5 presents the results of the TY procedure. For the full sample not allowing for any breaks in either series, the null hypothesis that DoG does not Granger-cause changes in GR is rejected at the 1% level of significance. The same is true for the hypothesis that GR does not Granger-cause DoG, thereby confirming a bidirectional relation. This also holds true if one allows for a structural break in 2009M07. Again, both null hypotheses of non-Granger causality are rejected at the 5% as well as 1% level respectively. For the case of two breaks, however, a rather different picture can be drawn which makes it worth considering the sub-sample analysis first.

Concerning the question whether the relationship between both variables of interest has changed due to the financial crisis, the previously reported bidirectional causation remains statistically significant for the pre-crisis I period. Here, again H_0 of DoG not-Granger-causing GR is rejected at the 5% level and the hypothesis GR not-Granger-causing DoG at the 10% level. Nonetheless, when looking at the second sub-sample, an interesting pattern emerges. The statistically strong evidence of bidirectional causation in the full sample cannot be confirmed for sub-sample II. In other words, when accounting for the financial crisis and its immediate aftermath, there is no statistical evidence of any Granger causality in either direction. This evidence is strongly supported by what is found by including two breaks for 2008M03 and 2010M03 and investigating the full sample from 1995M01 to 2013M12. For this test set-up, no Granger causality link in either direction can be reported.

We argue that the direction of Granger causality strongly depends on the time period(s) looked at and the way the financial crisis is dealt with. Not accounting for the fact that the crisis struck both series differently, an incorrect pattern of causality may be invoked. This becomes quite clear when comparing both sub-samples. In sub-sample I, the immediate effects of the crisis were simply 'split' and inherited by the corresponding sub-samples. It is interesting to see, that for the pre-crisis I but not the post-crisis I period, the same causality links as for the full sample with none/one breaks were reported. Yet, sub-sample II which neglects the slump-and-recovery path of GR as well as the level-shift in DoG, fails to provide similar causality links. Essentially this means that, excluding 24 observations which correspond to the financial crisis and its immediate aftermath, the strong case of a bidirectional Granger causality link becomes statistically insignificant.⁸

Cointegration in the presence of structural breaks

As for the Granger causality testing framework, we implement the $HI(r)$ test of cointegration allowing for structural breaks in Johansen *et al.* (2000) and

⁸ In order to check for robustness of our results, we employed the same TY procedure on the quarterly series of DoG and GR. As reported in Table B14 these results are in line with our findings on the monthly series and stress the importance of adjusting for structural break (s).

follow the discussion in Joyeux (2007). Similarly to the TY procedure allowing for structural breaks, we define the same VAR(P) system as before and, in addition, the following variables:

$$i_{2,t} = \begin{cases} 1 & \text{if } t = t^* + 1 \\ 0 & \text{else} \end{cases} \quad \text{and} \quad d_{2,t} = \begin{cases} 1 & \text{if } t \geq t^* \\ 0 & \text{if } t < t^*. \end{cases} \quad (7)$$

Again, t^* denotes the break point period for any $t \in \{1, \dots, T\}$. For the cointegration test allowing for *one* structural break, we include (i) a linear trend c , (ii) $d_{2,t-k}$ where k is designated the maximum lag length, (iii) an interaction term $c \times d_{2,t-k}$ as well as (iv) $i_{2,t-l}$, where $l = 0, 1, \dots, k-1$, as exogenous variables to the system. For the cointegration test allowing for two structural breaks, the same dummy variables as above were defined, however, instead of only one break point t^* , *two* break points, t_1^* and t_2^* are used to appropriately define the newly generated set of dummy variables. For hypothesis testing, the critical values are derived as discussed by Giles (2011) which depend on the position of the previously defined exogenous break point(s).

Even though testing cointegrating relationships is a very powerful tool in assessing causality links between variables, stringent assumptions have to be imposed for this method to work such as an equal order of integration of one and a reasonable sample size. Hargreaves (1994) suggests that Johansen's cointegration test performs best for fairly large samples of 100 observations and more. Our study makes use of 228 observations, boosted by the higher frequency monthly data set. However, when splitting the sample due to regime changes the shorter span prevents us from making a robustness check for cointegrating relationships in the sub-samples with meaningful results.

We compare the findings of the Johansen's cointegration test when allowing for no, one and two structural breaks. The motivation for this stems from the borderline rejection of non-stationarity for GR when testing for unit roots with one and two structural breaks in level in order to see if the results of the TY procedure can be confirmed when assuming that GR is I(1) with breaks. The cointegration test without structural breaks is added to this set of tests to compare and cross-validate all results and confirm the point made by Giles (2011) who argues that erroneously neglecting actual structural breaks in a cointegration analysis may produce misleading results.

Therefore, the cointegrating analysis presented in the forthcoming paragraphs is undertaken for the full data sample spanning 1995M01 to 2013M12 with, firstly, no structural break, secondly, one structural break in 2009M07 as well as, thirdly, two structural breaks in 2008M03 and 2010M03 which mark the corner stones of the financial crisis. It has to be pointed out that in order to correctly undergo a cointegration analysis when not allowing for any structural breaks, a different approach would have to be used. This is because ADF, PP as well as KPSS tests consistently find DoG to be I(2) and GR to be I(1). Therefore, if these results were assumed to be correct, one would have to implement a cointegration analysis allowing for these data features.⁹ Yet,

⁹ See Kurita (2013) or Nielsen (2002) for further discussion on testing for cointegration between macroeconomic variables of differing orders of integration.

as further unit root tests have shown that both series can be assumed to be I (1) with (a) break(s), this approach is dismissed. Rather, the forthcoming analysis investigates what happens if one fails to consider the presence of structural breaks altogether. Both cointegration tests allowing for structural breaks follow Johansen *et al.* (2000). With respect to the deterministic component of the system, we allow for a trend in both the VAR as well as the cointegrating equation as we are explicitly interested in any trend associated with the relationship between both variables.

Results of the cointegration test

The results of the different cointegration tests are presented in Table 6. As can be seen there, if one does not allow for a structural break when testing for cointegrating relationships, Johansen's test finds very strong evidence of a relationship between both variables: The null hypothesis of no cointegrating relationship between both variables is rejected at a 1% level of significance. This result confirms the findings of the TY procedure which concluded that a bivariate causality between both variables at the 5% level is given. However, when allowing for one structural break in 2009M07, it can be seen that, even though the presence of a cointegrating relationship between both variables is confirmed, the trace statistic is very close to the 5% critical value. Once more, these results are in line with the findings in Table 5. More interestingly, however, and again in line with previous findings, when allowing for two structural breaks in 2008M03 and 2010M03, the Johansen cointegration test fails to reject H_0 of no cointegrating relationships between both series at the 5% interval. Again, this is confirmed through the results of the TY procedure with two structural breaks.

These results are encouraging in that they reinforce the proposition that not accounting for structural breaks in the cointegration testing framework may result in useless and incorrect inference. We believe that this is the correct point to stop the analysis as for both cases where a cointegrating relationship was reported, there is strong empirical evidence for the claim that this relationship is solely due to not having accounted for the financial crisis. We conclude that when properly controlling for the financial crisis through

Table 6
Results cointegration test

No. of breaks	Hypothesized no. of CE(s)	Trace statistic	5% Critical value
None	None	52.359	25.872†
	At most 1	5.729	12.518
One	None	38.070	35.022†
	At most 1	9.163	17.684
Two	None	39.020	41.757
	At most 1	5.667	21.512

Notes: The table reports the trace statistic of the Johansen (1991, 1995) as well as the Johansen *et al.* (2000) cointegration tests. Significant results on the 5% level of significance designated † respectively. Values are rounded.

endogenously determined break points from various analysis tools, no empirical evidence of a causality relation between the gross debt level and the GDP growth rate can be reported through either a Granger causality test following Toda and Yamamoto (1995) or Johansen *et al.* (2000) cointegration tests allowing for structural breaks.

V CONCLUSION

The European sovereign debt crisis of 2010 acted as a catalyst for an extensive debate centred around the seminal contribution of Reinhart and Rogoff (2010) on the interaction between government debt and economic growth. In the context of this debate, we have tested for a statistical causal link between public debt and growth for the UK between 1995 and 2013 using higher frequency data, an approach that allows us to utilise the intra-quarterly fluctuations displayed by the widely available monthly time series for UK government debt. In doing so, we have carried out Granger causality and cointegration tests capable of controlling for structural breaks caused by the extraordinary shocks that appear as the result of the recent financial crisis in macroeconomic time series data.

Using monthly time series data for growth in GDP and the sovereign debt-to-GDP ratio for the UK, we test for and suggest two structural breaks that mark the beginning and end of the financial crisis for both series. Our extensive unit root tests confirm that these two structural breaks play a crucial role in our analysis and that omitting them could have led to misleading findings. Furthermore, we test both series using an augmented Toda and Yamamoto (1995) test of Granger causality and a Johansen *et al.* (2000) cointegration test that are both modified to account for the presence of structural breaks and find that for the case of the UK there is no evidence of a direct causal relation of the two variables in either direction when one correctly accounts for the different regimes found in the data.

In our linear analysis, we find no evidence of a time-dependent relationship between economic growth and debt for the UK when accounting for structural breaks around the financial crisis. Our results are robust over different testing frameworks and sub-sample analyses providing consistent results that even hold true when applied to a second data set. Given our analysis presents clear empirical evidence of a lack of a causal relationship between these two variables we conclude that any policy measure directly targeting public expenditures in the sole hope of bolstering economic growth could be an overly simple approach to a more complex economic problem. This result is consistent with other findings in the literature such as in Bell *et al.* (2015) or Puente-Ajovín and Sanso-Navarro (2015) who also suggest that negative correlations between the two variables might be due to the increase in debt levels witnessed after economic downturns and in particular find no causal relationship from high debt to low growth for the case of the UK.

One way to extend this analysis might be to use a framework that does not rely on the assumption of linearity in the model used to test for statistical causality or one that includes some control for possible higher dimensional

relationships between debt and growth and, while firmly on our agenda, we leave these interesting issues for future research.

APPENDIX A: DATA MANIPULATION

Below we provide a discussion of the data derivation process. Code to reproduce all figures as well as the data presented in this paper are available online.¹⁰

Derivation of monthly debt-to-GDP ratio (DoG)

Quarterly data on gross government debt are readily available, for example, quarterly government debt from Eurostat and general government total gross debt from the OECDStats.StatExtracts.¹¹ However, only the United Kingdom Office of National Statistics (ONS) offers Public Sector Net Debt (PSND from now on) on a monthly basis.¹² As quarterly gross debt and quarterly PSND data are freely available and in order to generate a monthly gross debt measure, we first extract the difference between the quarterly OECD gross debt measure and the quarterly PSND measure by ONS. This difference is then disaggregated from quarterly to monthly data using the Denton–Cholette (Denton, 1971) method through the *R* (R Core Team, 2013) package *tempdisagg* (Sax and Steiner, 2013). The newly derived monthly difference is then added to the monthly PSND series to generate a close approximation of a monthly gross debt measure. In a last step, we use Census X-13, the Seasonal Adjustment Program developed by the United States Census Bureau to seasonally adjust the series using the *R* package *seasonal* (Sax, 2017). Figure A. 2a plots the monthly debt-to-GDP ratio against time and compares it with the same measure from other publicly available data sources.

Derivation of GDP growth rate (GR)

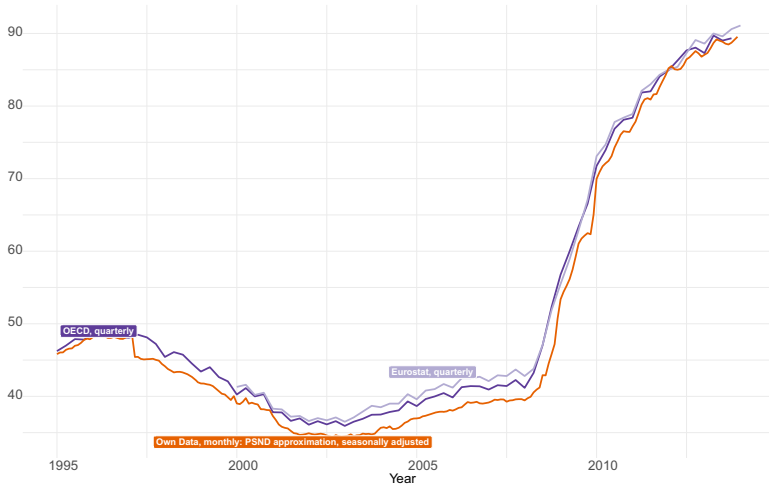
For monthly GDP growth rates, we make use of monthly estimates from the National Institute of Economic and Social Research (NIESR). This series was estimated following the methodology described in Mitchell *et al.* (2005), expressed as an index where 2008 = 100. Using these indexed GDP estimates we calculate the percent growth rate in change over the same month of the previous year. Figure A.2b compares this calculated series with other GDP measures commonly used by researchers.¹³

¹⁰ See <https://amannj.github.io/publications>.

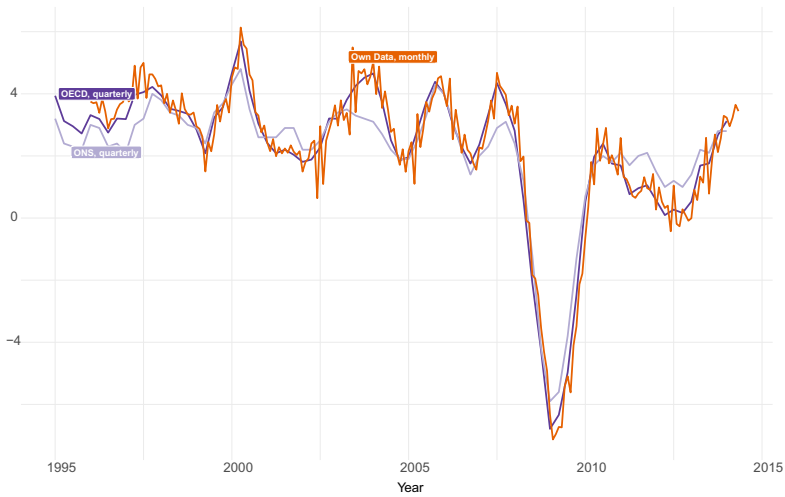
¹¹ Eurostat data: The full data set including meta data can be downloaded from <http://ec.europa.eu/eurostat/data/database> (last visit March 2017). OECD.StatExtract quarterly National Accounts (Quarterly Public Sectors Debt, percent of GDP; data can be found at http://stats.oecd.org/Index.aspx?DataSetCode=QASA_TABLE7PSD (last visit March 2017).

¹² These data are available at http://www.quandl.com/UKONS/PUSF_HF6X_M-PSND-excluding-NR-as-a-of-GDP-Monthly (last visit March 2017) .

¹³ The OECD data series, 'quarterly growth rates of real GDP, changes over same quarter, previous year', is taken from <http://stats.oecd.org/index.aspx?queryid=26674> (last visit March 2017). Data from ONS, 'gross domestic product: quarter on quarter previous year: CVM SA (Quarterly)', is taken from <https://www.ons.gov.uk/economy/grossdomesticproductgdp/timeseries/ihyq> (last visit March 2017).



(a) Gross Government Debt-to-GDP Ratios against Time



(b) GDP Growth Rates against Time

Figure A.2. Comparison of various data sources for DoG and GR. [Colour figure can be viewed at wileyonlinelibrary.com]

APPENDIX B: TABLES

Table B7
Summary of testing procedures

Full sample	Unit root tests	Results	Further analysis
No breaks in either series	standard ADF, PP and KPSS tests	DoG is I(2) GR is (1)	TY p.
One break in each series	Perron (1997) and Zivot and Andrews (1992) tests	DoG is I(1) GR is I(1) [•]	Cointegration analysis with structural breaks and TY p.
Two breaks in each series	Lee and Strazicich (2002) test	DoG is I(1) GR is I(1) [•]	Cointegration analysis with structural breaks and TY p
Sub-samples			
Split in 2009M07:			
Pre-crisis I	standard ADF, PP and KPSS tests	DoG is I(2) GR is I(1)	TYp.
Post-crisis I		DoG is I(1) GR is I(2) [•]	TYp.
... split in 2008M03 and 2010M03			
Pre-crisis II	standard ADF, PP and KPSS tests	DoG is I(1) GR is I(0)	TYp.
Post-crisis II		DoG is I(1) [•] GR is I(1) [•]	TY

Notes: Results where no definite conclusion on the order of integration was derived are designated • or • in the above table if $m=\{0,1\}$ or $m=\{1,2\}$ respectively.

Table B8
Results of traditional unit root tests, full sample

	DoG			GR		
	Levels	1st diff	2nd diff	Levels	1st diff	2nd diff
ADF Test						
Intercept & trend	-1.70	-2.74	-13.87*	-3.07	-6.54*	-
Intercept	-0.51	-2.25	-13.91*	-1.93	-6.54*	-
None	0.58	-2.11†	-13.94*	-1.51	-6.54*	-
PP Test						
Intercept & trend	-0.40	-11.33*	-	-3.13	-21.26*	-
Intercept	1.54	-10.30*	-	-2.79‡	-21.27*	-
None	2.50	-10.01*	-	-2.03†	-21.30*	-
KPSS Test						
Intercept & trend	0.47*	0.16†	0.06	0.08	-	-
Intercept	1.14*	0.93*	0.07	0.78*	0.05	-

Notes: Table contains t-statistics (ADF and PP Test) and LM-statistics (KPSS Test). Significant results on the 1%, 5% and 10% level of significance designated *, † and ‡ respectively. Values are rounded.

Table B9
Results of traditional unit root tests, pre-crisis I period

	DoG			GR		
	Levels	1st diff	2nd diff	Levels	1st diff	2nd diff
ADF Test						
Intercept & trend	3.98	-2.01	-10.17*	-2.17	-5.18*	-
Intercept	-0.64	-0.933	-9.98*	-1.30	-5.06*	-
None	0.70	-0.82	-9.99*	-1.22	-5.00*	-
PP Test						
Intercept & trend	3.79	-8.14*	-	-1.65	-19.22*	-
Intercept	0.89	-7.13*	-	-0.81	-19.03*	-
None	1.06	-6.82*	-	-1.23	-18.93*	-
KPSS Test						
Intercept & trend	0.36*	0.23*	0.07	0.16†	0.08	-
Intercept	0.43†	0.72*	0.23	0.58†	0.20	-

Notes: Table contains t-statistics (ADF and PP Test) and LM-statistics (KPSS Test). Significant results on the 1%, 5% and 10% level of significance designated *, † and ‡ respectively. Values are rounded.

Table B10
Results of traditional unit root tests, post-crisis I period

	DoG			GR		
	Levels	1st diff	2nd diff	Levels	1st diff	2nd diff
ADF Test						
Intercept & trend	-1.92	-7.26*	-	-3.01	-1.73	-9.37*
Intercept	-3.85*	-	-	-2.84‡	-1.91	-9.40*
None	4.31	-1.66‡	-7.67*	-2.33†	-1.83‡	-9.48*
PP Test						
Intercept & trend	-2.14	-7.96*	-	-2.86	-10.22*	-
Intercept	-6.95*	-	-	-3.10†	-10.06*	-
None	3.50	-4.76*	-	-2.40†	-9.76*	-
KPSS Test						
Intercept & trend	0.24*	0.11	-	0.12‡	0.19‡	0.07
Intercept	0.94*	0.71†	0.23	0.24	0.26	-

Notes: Table contains t-statistics (ADF and PP Test) and LM-statistics (KPSS Test). Significant results on the 1%, 5% and 10% level of significance designated *, † and ‡ respectively. Values are rounded.

Table B11

Results of traditional unit root tests, pre-crisis II period

	DoG			GR		
	Levels	1st diff	2nd diff	Levels	1st diff	2nd diff
ADF Test						
Intercept & trend	0.11	-11.26*	-	-4.65*	-	-
Intercept	-0.98	-11.11*	-	-4.38*	-	-
None	-1.19	-11.06*	-	-1.20	-5.61*	-
PP Test						
Intercept & trend	-0.26	-11.54*	-	-5.36*	-	-
Intercept	-1.04	-11.48*	-	-21.30*	-	-
None	-0.91	-11.46*	-	-21.41*	-	-
KPSS Test						
Intercept & trend	0.33*	0.28*	0.07	0.05	0.03	-
Intercept	0.98*	0.44‡	0.07	0.19	0.03	-

Notes: Table contains t-statistics (ADF and PP Test) and LM-statistics (KPSS Test). Significant results on the 1%, 5% and 10% level of significance designated *, † and ‡ respectively. Values are rounded.

Table B12

Results of traditional unit root tests, post-crisis II period

	DoG			GR		
	Levels	1st diff	2nd diff	Levels	1st diff	2nd diff
ADF Test						
Intercept & trend	-1.20	-3.80†	-8.16*	-2.20	-2.61	-8.98*
Intercept	-2.78‡	-5.49*	-	-2.38	-2.87†	-8.70*
None	4.67	-2.65*	-	-1.15	-2.30*	-
PP Test						
Intercept & trend	-1.08	-8.18*	-	-3.18	-11.92*	-
Intercept	-4.48*	-	-	-3.21†	-11.94*	-
None	4.01	-4.03*	-	-0.60	-11.87*	-
KPSS Test						
Intercept & trend	0.23*	0.15*	0.14*	0.19†	0.14‡	0.07
Intercept	0.85*	0.52†	0.15	0.22	0.18	0.01

Notes: Table contains t-statistics (ADF and PP Test) and LM-statistics (KPSS Test). Significant results on the 1%, 5% and 10% level of significance designated *, † and ‡ respectively. Values are rounded.

Table B13

Summary table break point test results

	DoG		GR	
	1 Break	2 Breaks	1 Break	2 Breaks
Monthly data				
Bai and Perron (1998, 2003b)	2009M07	1999M01 2009M06	2008M05	2008M03 2011M01
Zivot and Andrews (1992)	2005M01	–	2008M03	–
Perron (1997)	2004M12	–	2008M02	–
Lee and Strazicich (2002)	–	2004M11 2009M12	–	2006M03 2010M02
Bivariate sums of residuals test		2008M07 and 2010M03		
Quarterly data				
Bai and Perron (1998, 2003b)	2009Q3	1999Q1 2009Q3	2008Q2	2008Q2 2011Q1
Zivot and Andrews (1992)	2005Q4	–	2008Q2	–
Perron (1997)	2004Q4	–	2008Q1	–
Lee and Strazicich (2002)	–	2000Q4 2008Q1	–	2006Q2 2010Q1
Bivariate sums of residuals test		2008Q3 and 2010Q1		

Notes: Summary table of previously conducted break point tests as well as unit root tests allowing for up to two endogenously defined breaks. All tests were conducted for the constructed monthly data set as well as quarterly OECD data.

Table B14

Results of granger-causality tests following the TY procedure, quarterly data

Full sample	DoG \Rightarrow GR	GR \Rightarrow DoG
No breaks in either series	0.0138†	0.0090*
One break in each series	0.0173†	0.0119†
Two breaks in each series	0.7119	0.7945
Sub-sample		
Split in 2009Q03		
Pre-crisis I	0.1430	0.0643‡
Post-crisis I	0.0937‡	0.4981
Split in 2008Q03 and 2010Q01		
Pre-crisis II	0.9512	0.1746
Post-crisis II	0.2282	0.6180

Notes: The table reports p-values of the Wald test. Arrows denote the direction of Granger causality. The levels of significance are denoted ‡, † and * for the 10%, 5% and 1% level respectively. Proposed value for reported above for maximum order of integration remain as in the initial analysis.

ACKNOWLEDGEMENTS

We gratefully acknowledge the constructive comments by two anonymous reviewers on earlier versions of the manuscript.

REFERENCES

AMANN, J. and MIDDLEDITCH, P. (2015). Revisiting Reinhart & Rogoff after the crisis: a time series perspective. Centre for Growth and Business Cycle Research Discussion Paper Series 198.

- BAI, J. and PERRON, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, **66**, pp. 47–78. <http://www.jstor.org/stable/2998540>
- BAI, J. and PERRON, P. (2003a). Critical values for multiple structural change tests. *Econometrics Journal*, **6**, pp. 72–8.
- BAI, J. S. and PERRON, P. (2003b). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, **18**, pp. 1–22.
- BAUM, A., CHECHERITA-WESTPHAL, C. and ROTHER, P. (2013). Debt and growth: New evidence for the euro area. *Journal of International Money and Finance*, **32**, pp. 809–21.
- BELL, A., JOHNSTON, R. and JONES, K. (2015). Stylised fact or situated messiness? The diverse effects of increasing debt on national economic growth. *Journal of Economic Geography*, **15**, pp. 449–72.
- BIANCHI, M. (1998). Granger causality tests in the presence of structural changes. Bank of England Working Paper 33.
- CASNI, A. C., BADURINA, A. A. and SERTIC, M. B. (2014). Public debt and growth: evidence from Central, Eastern and Southeastern European countries. *Proceedings of Rijeka Faculty of Economics*, **32**, pp. 35–51.
- CECCHETTI, S. G., MOHANTY, M. S. and ZAMPOLLI, F. (2011). Achieving growth amid fiscal imbalances: the real effects of debt. *Economic Symposium Conference Proceedings* **352**, 145–96. <http://www.bis.org/publ/work352.htm>
- CHEN, Y.-C., ROGOFF, K. S. and ROSSI, B. (2010). Can exchange rates forecast commodity prices? *The Quarterly Journal of Economics*, **125**, pp. 1145–94.
- CHU, C.-S. J., HORNİK, K. and KUAN, C.-M. (1995). The moving-estimates test for parameter stability. *Econometric Theory*, **11**, pp. 699–720.
- DENTON, F. T. (1971). Adjustment of monthly of quarterly series to annual totals: an approach based on quadratic minimizations.
- FERREIRA, N. B., MENEZES, R. and BENTES, S. (2014). Cointegration and structural breaks in the eu sovereign debt crisis. *International Journal of Latest Trends in Finance and Economic Sciences*, **4**, pp. 680–90.
- GILES, D. (2011). Cointegrated at the Hips. Tech. rep., Econometrics Beat: Dave Giles' Blog. <http://davegiles.blogspot.co.at/2011/05/cointegrated-at-hips.html>
- GRANGER, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, **37**, pp. 424–38. <http://www.jstor.org/stable/1912791>
- HARGREAVES, C. P. (1994). *Review of Methods of Estimating Cointegrating Relationships*. New York: Oxford University Press Inc., 87 pp.
- HERNDON, T., ASH, M. and POLLIN, R. (2014). Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff. *Cambridge Journal of Economics*, **38**, pp. 257–79.
- JOHANSEN, S. (1991). Estimation and hypothesis-testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, **59**, pp. 1551–80.
- JOHANSEN, S. (1995). Likelihood-based inference in cointegrated vector autoregressive models. No. 9780198774501 in OUP Catalogue. Oxford University Press.
- JOHANSEN, S., MOSCONI, R. and NIELSEN, B. (2000). Cointegration analysis in the presence of structural breaks in the deterministic trend. *Econometrics Journal*, **3**, pp. 216–49. <http://ideas.repec.org/a/ect/emjrn/v3y2000i2p216-249.html>
- JOYEUX, R. (2007). How to deal with structural breaks in practical cointegration analysis. *Cointegration for the Applied Economist*, **5**, pp. 195–221.
- KONZELMANN, S. J. (2014). The political economics of austerity. *Cambridge Journal of Economics*, **38**, pp. 701–41.
- KOOP, G. and POTTER, S. M. (2007). Estimation and forecasting in models with multiple breaks. *Review of Economic Studies*, **74**, pp. 763–89.
- KOURTELLOS, A., STENGOS, T. and TAN, C. M. (2013). The effect of public debt on growth in multiple regimes. *Journal of Macroeconomics*, **38**, pp. 35–43.
- KURITA, T. (2013). Modelling time series data of monetary aggregates using I(2) and I(1) cointegration analysis. *Bulletin of Economic Research*, **65**, pp. 372–88. <http://doi.wiley.com/10.1111/j.1467-8586.2011.00400.x>

- LEE, J. and STRAZIČIČ, M. C. (2002). Minimum lagrange multiplier unit root test with two structural breaks. *Review of Economics and Statistics*, **85**, pp. 1082–9.
- LÜTKEPOHL, H. (1982). Non-causality due to omitted variables. *Journal of Econometrics*, **19**, pp. 367–78.
- MINEA, A. and PARENT, A. (2012). Is high public debt always harmful to economic growth? Reinhart and Rogoff and some complex nonlinearities. Working Papers 201218, CERDI. <http://ideas.repec.org/p/cdi/wpaper/1355.html>
- MITCHELL, J., SMITH, R. J., WEALE, M. R., WRIGHT, S. and SALAZAR, E. L. (2005). An indicator of monthly GDP and an early estimate of quarterly GDP growth. *Economic Journal*, **115**, pp. F108–29.
- NIELSEN, H. B. (2002). An I(2) cointegration analysis of price and quantity formation in Danish manufactured exports. *Oxford Bulletin of Economics and Statistics*, **64**, pp. 449–72.
- PANIZZA, U. and PRESBITERO, A. F. (2014). Public debt and economic growth: is there a causal effect? *Journal of Macroeconomics*, **41**, pp. 21–41.
- PATTERSON, K. (2012). Unit root tests in time series. Volume 2, Extensions and developments. Palgrave Texts in Econometrics.
- PERRON, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, **57**, pp. 1361–401.
- PERRON, P. (1997). Further evidence on breaking trend functions in macroeconomic variables. *Journal of Econometrics*, **80**, pp. 355–85.
- PUNTE-AJOVÍN, M. and SANJO-NAVARRO, M. (2015). Granger causality between debt and growth: evidence from OECD countries. *International Review of Economics & Finance*, **35**, pp. 66–77. <https://ideas.repec.org/a/eee/reveco/v35y2015icp66-77.html>
- R Core Team (2013). R: a language and environment for statistical computing. <http://www.r-project.org/>
- REINHART, C. M. and ROGOFF, K. S. (2010). Growth in a time of debt. *American Economic Review*, **100**, pp. 573–8.
- ROSSI, B. (2006). Are exchange rates really random walks? Some evidence robust to parameter instability. *Macroeconomic Dynamics* **10**, pp. 1–25.
- ROSSI, B. (2013). Advances in forecasting under instability. *Handbook of Economic Forecasting*, **2**, pp. 1203–324.
- SAX, C. (2017). seasonal: R Interface to X-13-ARIMA-SEATS. <https://cran.r-project.org/web/packages/seasonal/index.html>
- SAX, C. and STEINER, P. (2013). Temporal disaggregation of time series. *The R Journal*, **5**, pp. 80–7. <https://journal.r-project.org/archive/2013-2/sax-steiner.pdf>
- SONG, Y. (2014). Modelling regime switching and structural breaks with an infinite hidden markov model. *Journal of Applied Econometrics*, **29**, pp. 825–42.
- STERN, D. I. and ENFLO, K. (2013). Causality between energy and output in the long-run. *Energy Economics*, **39**, pp. 135–46.
- TODA, H. Y. and YAMAMOTO, T. (1995). Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics*, **66**, pp. 225–50.
- WOO, J. and KUMAR, M. S. (2015). Public debt and growth. *Economica*, **82**, pp. 705–39.
- ZIVOT, E. and ANDREWS, D. W. (1992). Further evidence on the great crash, the oil price shock, and the unit root hypothesis. *Journal of Business & Economic Statistics*, **10**, pp. 251–70. <http://ideas.repec.org/p/cwl/cwldpp/944.html>

Date of receipt of final manuscript: March 14 2017