Review of Economics & Finance
Submitted on 13/11/2015
Article ID: 1923-7529-2016-03-65-09 Dimitris Hatzinikolaou

A "litmus test" of Deficit Sustainability: The Case of the Greek Budget Deficit

Dimitris Hatzinikolaou

University of Ioannina, Department of Economics, 451 10 Ioannina, GREECE
Tel: +30 26510 05925 Fax: +30 26510 05092 E-mail: dhatzini@uoi.gr

Abstract: The paper applies three sustainability tests to the Greek budget deficit, which is known to be unsustainable, and can therefore be used to check the reliability of the three tests. Using three deficit definitions, i.e., in levels, in per capita terms, and in percent of GDP, I find the following results: (1) the unit-root test rejects sustainability only for the last definition; (2) the Hakkio-Rush test does not reject sustainability for any definition; and (3) the Hatzinikolaou-Simos test, which is applicable only for the first two definitions, strongly rejects sustainability, and so it proves to be more reliable, as expected.

Keywords: Budget deficit, Sustainability tests, Greece

JEL Classifications: C15, E62, H62, H63

1. Introduction

Sustainability of budget and current-account deficits has been the most important financial issue in the Eurozone since late 2009. As of December 2012, five Eurozone countries (Cyprus, Greece, Ireland, Portugal, and Spain) requested emergency loans from the Troika (International Monetary Fund, European Central Bank, and European Commission) in order to service their debts, as it became increasingly difficult for them to borrow from private markets. These countries have signed "memoranda of understanding" with the Troika, which required, to various degrees, that austerity and reform measures be imposed on their domestic economies, especially in Greece.

In the spring of 2012, Greece partially defaulted on its debt, since domestic holders of government debt, e.g., private banks, social-security funds, hospitals, tertiary education institutions, and even small private savers suffered a 53.5-percent "haircut" of the value of their bonds under a program called Private Sector Involvement (PSI). Before the "haircut," the deposits of many of these organizations held at the Bank of Greece were obligatorily converted into government bonds. Thus, it is not surprising that a likely default of Greece on its debt and its exit from the Eurozone, dubbed "Grexit," has been fervently debated in the press, as such an outcome might cause a crisis of confidence and a financial contagion (e.g., bank runs) in the Eurozone, because many European banks held sovereign debt of other European countries.

Assuming that a budget deficit is sustainable only if it can be financed by borrowing from private markets (Hakkio and Rush, 1991, p. 433), the above facts suggest that Greece's budget deficit is not sustainable, and so a reliable test of the sustainability hypothesis should reject it. In other words, the Greek budget deficit can serve as a "litmus test" of the reliability of the existing deficit-sustainability tests.

Conventional tests of deficit sustainability exploit the conditions implied by the government ́s intertemporal budget constraint (IBC). The IBC is satisfied, and hence the sustainability hypothesis holds, if the present discounted value of the outstanding debt tends to zero. This definition gives rise to unit-root and cointegration tests of the sustainability hypothesis (see, e.g., Hamilton and Flavin, 1986; Hakkio and Rush, 1991; Liu and Tanner, 1996; Payne, 1997; Tanner and Liu, 1994; and Trachanas and Katrakilidis, 2014). According to Bohn (2007), however, these tests "are incapable of rejecting the consistency of data sets with the IBC" (p. 1838), because "the IBC *per se* imposes very weak econometric restrictions" (p. 1846). This means that the standard unit-root and cointegration tests for sustainability have low power.

Recently, Hatzinikolaou and Simos (2013), henceforth HS, developed a new test, where sustainability requires that both the present discounted value of the debt tend to zero (i.e., the IBC be satisfied) and the undiscounted debt be bounded. Thus, the HS test is more stringent than the standard ones, as it requires that an additional condition be satisfied. HS applied their test to United States budget and current-account deficit data and rejected sustainability, whereas three other papers that applied conventional tests to the same deficit measures and sample periods failed to reject it; see also Hatzinikolaou, Simos, and Tsoka (2013).

The present paper applies three tests of deficit sustainability to the Greek budget deficit, which is known to be unsustainable, and so it can be used to check the reliability of the three tests, namely, the unit-root test, the Hakkio and Rush (1991) test, and the HS test. Following Hakkio and Rush (1991), Payne (1997), and others, I use three definitions of revenue and spending, namely, levels (in real terms), ratios to population, and ratios to real GDP. As Hakkio and Rush (1991, p. 430) explain, the last two definitions are more pertinent for a growing economy. Some authors (e.g., Tanner and Liu, 1994), however, use only the levels, which are more likely to be non-stationary and cointegrated than the ratios (see Cuddington, 1997, pp. 12-13). The contribution of the paper is that it confirms the expected result, namely, the rejection of the sustainability hypothesis by the HS test for a budget deficit that is already known to be unsustainable, whereas the standard tests fail to reject. Section 2 describes the data and implements the tests and section 3 concludes.

2. The Data and the Implementation of the Three Tests

2.1 The data

The sources of the data are the European Commission (AMECO) and the Eurostat, Government Finance Statistics. The two basic variables are total revenue (T_t) and total expenditure (G_t) of general government (billions of Euros in constant 2005 prices, using the GDP deflator, for the period 1970 to 2012), where T_t = taxes on production and imports + taxes on income and wealth + receivable property income + social-security contributions, and G_t = intermediate consumption + gross capital formation + compensation of employees + payable property income + social-security benefits + subsidies + social transfers. The real budget deficit is $DEF_t = G_t - T_t$.[1] As ratios to population and to real GDP (*RGDP*), these variables are defined as follows: $TPOP_t = T_t/POP_t$, $GPOP_t = G_t/POP_t$, $DEFPOP_t = DEF_t/POP_t$, $TGDP_t = T_t/RGDP_t$, $GGDP_t = G_t/RGDP_t$, and $DEFGDP_t = DEF_t/RGDP_t$, where POP_t is population (in thousands).

I begin the empirical analysis by testing the variables for unit roots, a necessary step for the implementation of all the three tests. I use the *KPSS* test of Kwiatkowski, et al. (1992), with and

[1] From the definition of G_t, it is clear that the deficit measure used here is interest inclusive, as the derivation of the cointegrating equation (Hakkio and Rush, 1991) depends on that assumption. This is why this choice is quite standard in the literature (see, e.g., Hakkio and Rush, 1991; Tanner and Liu, 1994; Martin, 2000; and Trachanas and Katrakilidis, 2014).

without trend (denoted as $KPSS_{\tau}$ and $KPSS_{\mu}$, respectively), and the Lee and Strazicich (2003, 2013) test, which allows for one or two structural breaks, namely, one or two level shifts ("Crash" Model A) and one or two changes in level and trend slope (Model C). Table 1 reports the results.

Table 1. Unit-root tests

Series \ Test	$KPSS_{\mu}$	$KPSS_{\tau}$	*LS* one crash (A)	*LS* two crashes (A)	*LS* one break (C)	*LS* two breaks (C)	Decision
T_t	0.92^{***}	0.16^{**}	-1.93	-2.30	-3.23	-4.58	I(1)
G_t	0.92^{***}	0.11	-2.57	-2.70	-5.63^{***} (2005)	-8.65^{***} (1990, 2005)	I(0)
DEF_t	0.60^{**}	0.10	-2.70	-2.77	-2.78	-5.22^{*} (1993, 2005)	I(0)
$TPOP_t$	0.91^{***}	0.14^{*}	-1.97	-2.38	-3.21	-4.43	I(1)
$GPOP_t$	0.92^{***}	0.08	-2.63	-2.91	-4.52^{**} (2004)	-7.90^{***} (1990, 2005)	I(0)
$DEFPOP_t$	0.56^{***}	0.11	-2.72	-2.81	-2.65	-4.98	I(0)
$TGDP_t$	0.91^{***}	0.09	-2.55	-3.36	-2.96	-4.32	I(0)
$GGDP_t$	0.88^{***}	0.20^{**}	-2.74	-3.13	-3.67	-4.50	I(1)
$DEFGDP_t$	0.39^{*}	0.15^{**}	-2.51	-2.91	-2.76	-4.32	I(1)

Notes:

(1) ***, **, * indicate significance at levels 1%, 5%, and 10%, respectively;

(2) subscripts μ and τ indicate "intercept-but-no-trend" and "intercept-plus-trend," respectively;

(3) *LS* denotes a Lee and Strazicich (2003, 2013) test, where the letters A and C in parentheses indicate the models of one or two level shifts ("Crash" Model A) and one or two changes in level and trend slope (Model C); the break dates are given in parentheses, except when the values of the test statistic or the dummy variables are insignificant;

(4) for the first differences of all the variables the tests strongly suggest that they are I(0), so, for space considerations, the test values are not reported;

(5) all data are expressed in real terms.

The results of Table 1 suggest that the variables T_t, $TPOP_t$, $GGDP_t$, and $DEFGDP_t$ behave as I(1) and the other ones as I(0), according to at least one of the tests. Thus, I take DEF_t and $DEFPOP_t$ to be I(0), because, according to the $KPSS_{\tau}$ test, trend stationarity cannot be rejected even at the 10-percent level; and in the case of DEF_t, the Lee-Strazicich test also supports stationarity at the 10-percent level.

2.2 The unit-root test of deficit sustainability

Again, according to the IBC, stationarity of the deficit implies its sustainability. When testing stationarity, however, it is crucial that we allow for structural breaks, which, unless they are taken into account, bias the test in favor of non-stationarity and reduce their power when the null hypothesis is non-stationarity (Liu and Tanner, 1996; Uctum, Thurston, and Uctum, 2006). Thus, the Lee-Strazicich endogenous-break test, which allows for one or two breaks under *both* the unit-root null and the alternative hypotheses, becomes especially important, since rejecting the null *unambiguously* implies trend stationarity. Hence, when in fact there is a unit root with breaks, the Lee-Strazicich test is less likely to lead to the false conclusion of stationarity, an error that is likely to occur in previous endogenous-break unit-root tests, in which the unit-root null assumes no breaks, and so its rejection is interpreted as stationarity, although it could mean unit root with breaks. Here, as I already indicated, DEF_t ~ I(0), $DEFPOP_t$ ~ I(0), and $DEFGDP_t$ ~ I(1), so $DEFGDP_t$ is the only deficit measure for which the unit-root test rejects the sustainability hypothesis.

2.3 The Hakkio-Rush test

Next, I use the Hakkio and Rush (1991) test based on the following equation:

$$y_{1t} = a + by_{2t} + \varepsilon_t, \quad (1)$$

where the variables y_{1t} and y_{2t} are both defined in levels (e.g., $y_{1t} = T_t$ and $y_{2t} = G_t$), or as ratios to population or to *RGDP*. Because of the uncertainty regarding the order of integration of the variables (see the discussion of the results of Table 1), I use the Pesaran, Shin, and Smith (2001) "bounds test" (*BT*) of cointegration to find "levels relationships"; and when I find one, I estimate its parameters and test hypotheses of interest. In order to locate structural breaks, I apply Gregory and Hansen (1996a, 1996b) cointegration tests to Equation (1). Note that these tests are sensitive to the choice of maximum lag length (k) in the Augmented Dickey-Fuller test on the residuals from Gregory-Hansen's (*GH*) "level shift," "level shift with trend," and "full break" regressions. Assuming $k = 6$, I report weak evidence (at the 10-percent level) for two pairs of variables that each pair forms a cointegrating regression with a "level shift" in 1997 (see Table 2, regressions 1 and 2).

Table 2. Two cointegration tests

Regression \ Test	*GH* (*C*)	*GH* (*C*|*T*)	*GH* (Full break)	*BT*
1. T_t on G_t	-4.55^{*} (1997)	-4.28	-4.65	7.27^{***}
2. $TPOP_t$ on $GPOP_t$	-4.55^{*} (1997)	-4.22	-4.59	7.16^{**}
3. $TGDP_t$ on $GGDP_t$	-4.21	-3.94	-4.23	15.80^{***}

***Notes*:** (1) In both tests, the null hypothesis (H_0) is "no cointegration";

(2) ***, **, * indicate rejection of H_0 at significance levels 1%, 5%, and 10%, respectively;

(3) *GH*(*C*), *GH*(*C* |*T*), and *GH* (Full break) stand for Gregory-Hansen's "level shift," "level shift with trend," and "full break" models; maximum lag length was set equal to 6; the break point is given in parentheses;

(4) *BT* stands for the Pesaran, Shin, and Smith (2001) "bounds test" for a "levels relationship," where the maximum lag length was set equal to 4, and insignificant lags were dropped one at a time; standard errors are robust to heteroscedasticity and serial correlation; critical values are obtained from Table CI(iii) Case III of Pesaran, Shin, and Smith (2001, p. 300) when no trend is included in the *BT* regression, and from their Table CI(v) Case V (p. 301) when trend is included; each of these *BT* regressions includes a dummy variable (to allow for a level shift); it is assumed that the presence of these dummies does not affect the critical values of the test (Pesaran, Shin, and Smith, 2001, footnote 17).

Figure 1. General government expenditure (G_t), revenue (T_t), and the budget deficit ($DEF = G_t - T_t$) of Greece, 1970-2012

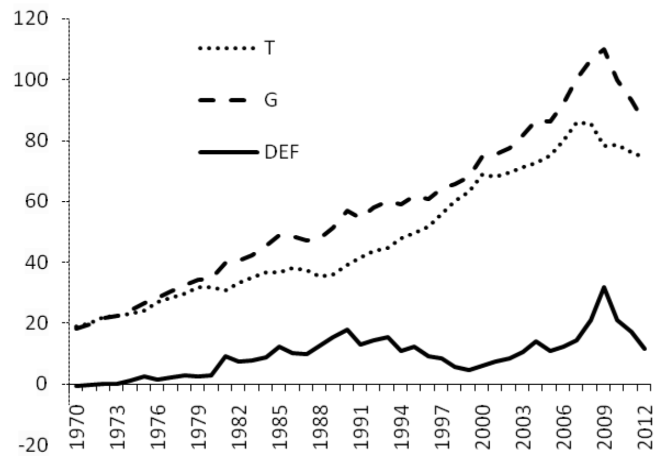

Figure 1 shows such shifts, where the dates may differ from those suggested by the *GH* test (i.e., 1997), apparently because the sample is not large enough to render the *GH* tests sufficiently powerful to locate these shifts with precision.

Since the *BT* rejects the non-cointegration hypothesis for all the pairs of variables (y_{1t}, y_{2t}) of Table 2, in each case I consider a levels relationship with a break in the level by including a dummy variable (D_t) whenever it is statistically significant. That is, I use the *BT* method to estimate the equation

$$y_{1t} = a + by_{2t} + cD_t + \varepsilon_t \qquad (2)$$

and test the following two hypotheses: (1) cointegration or, more generally, a "levels relationship"; and (2) H_0: $b \geq 1$ (H_1: $b < 1$). The deficit is "strongly sustainable" if and only if there is a levels relationship between y_{1t} and y_{2t} and $b = 1$; it is only "weakly sustainable" if there is a levels relationship and $0 < b < 1$; and it is unsustainable if $b \leq 0$; see Martin (2000, p. 86). Table 3 reports the results.

As an illustration of the *BT* method, consider regression 1 of Table 3, where $y_{1t} = T_t$, $y_{2t} = G_t$, and $D_t = D97_t$, where $D97_t = 1$ for $t \geq 1997$, and zero otherwise. This definition of D_t is not inconsistent with Figure 1, which shows the decrease in the budget deficit that started in 1991, thanks to the fiscal consolidation (mainly reforms to widen the tax base and limit tax evasion) pursued by the government, partly to meet the Maastricht Treaty targets (see, e.g., OECD, 1992, pp. 11-21; 1996, pp. 1, 24-40). After dropping the insignificant lags of ΔT_t and ΔG_t, the estimating *BT* regression is

$$\Delta T_t = \beta_0 + \beta_1 T_{t-1} + \beta_2 G_{t-1} + \beta_3 D97_t + \varphi_1 \Delta T_{t-1} + \psi_0 \Delta G_t + \varepsilon_t \qquad (3)$$

The parameters of interest, i.e., those of Equation (2), can be recovered from Equation (3) by setting $\Delta T_{t-i} = \Delta G_{t-i} = 0$, $i = 0, 1$, and then leading the equation by one period and solving for T_t. The result is $a = -\beta_0/\beta_1$, $b = -\beta_2/\beta_1$, and $c = -\beta_3/\beta_1$. Thus, in the context of Equation (3), testing the hypothesis H_0: $b = 1$ amounts to testing H_0: $\beta_1 + \beta_2 = 0$; testing H_0: $c = 0$ amounts to testing H_0: $\beta_3 = 0$; etc. The test for a levels relationship is a standard *F*-test of the hypothesis $\beta_1 = \beta_2 = 0$, with critical values obtained from Table CI(iii) Case III of Pesaran, Shin, and Smith (2001, p. 300) when no trend is included in (3), and from Table CI(v) Case V (p. 301) when trend is included. The hypothesis of no levels relationship is rejected at the 1-percent level (see Table 2).

Table 3. Estimation of Eq. (3) by the method of Pesaran, Shin, and Smith (2001)

Regression of y_{1t} on y_{2t}	$\hat{a}$	$\hat{b}$	$\hat{c}$	***p*-values for H$_0$: *a*=0, *b*=1 and H$_0$: *b*≥1**	***p*-value for LM_H**	***p*-values for LM_1 & LM_2**	***p*-value for *RESET***	***p*-value for *BJ***
1. T_t on G_t	12.70** (0.010)	0.539*** (0.000)	14.60** (0.010)	0.029** 0.004***	0.283	0.848 0.904	0.426 0.293	0.000***
2. $TPOP_t$ on $GPOP_t$	0.001*** (0.007)	0.495*** (0.000)	0.001*** (0.008)	0.025** 0.003***	0.279	0.771 0.847	0.763 0.320	0.000***
3. $TGDP_t$ on $GGDP_t$	0.133*** (0.004)	0.482*** (0.000)	0.050** (0.021)	0.013** 0.002***	0.248	0.381 0.551	0.740 0.541	0.208

***Notes*:** (1) ***, **, * indicate significance at levels 1%, 5%, and 10%, respectively;

(2) the numbers in parentheses underneath coefficient estimates are the *p*-values of the χ^2 statistic for the hypothesis that the true coefficient is zero;

(3) LM_H is a test for heteroscedasticity, namely, the *t*-test on the slope coefficient in the regression of the squared residuals on the squared fitted values;

(4) LM_1 and LM_2 are the standard Breusch-Godfrey *LM* tests for autocorrelation of orders 1 and 2;

(5) the estimation method is robust to heteroscedasticity and to autocorrelation;

(6) *RESET* tests the significance of the squared fitted value of the dependent variable as an additional regressor, as well as the joint significance of the squared and the cubed fitted values;

(7) *BJ* is the Bera-Jarque χ^2 test for normality; when *BJ* rejects, the results are still considered reliable by invoking the central-limit theorem.

The diagnostic tests for regression 1 of Table 3 suggest that the results are reliable, as: (1) there is no evidence of heteroscedasticity, since the *p*-value of an *LM* test, denoted as LM_H (a *t*-test on the slope coefficient in the regression of the squared residuals on the squared fitted values), is 0.283; (2) there is no evidence of autocorrelation in the residuals, since the *p*-values of the standard Breusch-Godfrey *LM* test for autocorrelation of order 1 and 2 are 0.848 and 0.904; (3) there is no evidence of misspecification, since the *p*-values of the *RESET*, when the squared as well as the squared and the cubed fitted values of the dependent variable are used as additional regressors in (3), are 0.426 and 0.293, respectively; (4) normality can be rejected at any level, since the *p*-value of the Bera-Jarque test is 0.000, but the results are still considered reliable by invoking the central-limit theorem.

Now, the hypothesis H$_0$: $b \geq 1$ can be rejected at the 1-percent level (*p*-value = 0.004) in favor of the alternative, H$_1$: $b < 1$. Thus, according to the Hakkio-Rush test, since (1) there is strong evidence for a levels relationship between T_t and G_t (see Table 2); and (2) $0 < b < 1$; it follows that the deficit can be considered weakly sustainable.

Note that regressions 2 and 3 of Table 3 generate similar results, which is evidence of robustness to the choice of definition of the deficit (as level or as ratio to population or to real GDP). Note also that regression 3 is more reliable than the first two, as it passes the normality test as well (*p*-value of the Bera-Jarque test = 0.208).

2.4 The Hatzinikolaou-Simos test

Implementing the HS test, first note that for each of the three pairs of variables, there is evidence for a levels relationship (see Table 2), and the hypothesis H_0: $a = 0$, $b = 1$ [or, in the context of Equation (3), H_0: $\beta_0 = 0$, $\beta_1 + \beta_2 = 0$] can be rejected at the 5-percent level (see Table 3). Thus, in accordance with Case 3 of HS (p. 66), further testing is required. In particular, the following two left-sided hypotheses must be tested separately: (i) H_0: $b \geq 1$ against H_1: $b < 1$; and (ii) H_0: $a \geq 0$ against H_1: $a < 0$. As shown in Table 3, for each of the three pairs of variables, the hypothesis H_0: $b \geq 1$ is rejected at the 1-percent level, whereas the hypothesis H_0: $a \geq 0$ cannot be rejected, since the estimate of a is positive. Thus, the three regressions of Table 3 clearly fall into Case 3c of HS (p. 66), and so I need to test their Condition (9). As I have already indicated, however, this condition cannot be tested for the series *DEFGDP*, because the latter behaves as I(1). Thus, in what follows, I test this condition for the definitions *DEF* and *DEFPOP*, which are taken to be I(0) (see Table 1). For both of these two definitions, the critical values for the HS test reported in the two columns of Table 4 have been produced by an AR(1) model, which includes the dummy variable $D08_t$ as an additional regressor, where $D08_t = 1$ for $t \geq 2008$, and $D08_t = 0$ otherwise.

Note that in all of the six cases considered in Table 4, the symmetry of the sampling distributions produced by the Monte Carlo simulations (each with 50,000 replications) is easily accepted, even when the estimator of the spectrum is given by Equation (14) of HS, since the *p*-values of the test for symmetry range from 0.30 to 0.96. These distributions are characterized by kurtosis, however, in the form of somewhat fatter tails than those of the standard normal distribution, so the Jarque-Bera test strongly rejects normality. Table 4 also reports the values of the HS test statistic (*TS*) calculated from the actual data, along with its *p*-value. All of these values have been calculated in the same way as in HS, so, to save space, I will not explain their methodology here. According to the results of Table 4, sustainability of the deficit can be rejected even at the 1-percent level.

Table 4. Values of the HS test statistic (10) calculated from the actual data for two definitions of the deficit, *DEF* and *DEFPOP*, and 1%, 5%, and 10% critical values (*CVs*) derived from simulations with 50,000 replications

***DEF*, AR(1) including $D08_t$**	***DEFPOP*, AR(1) including $D08_t$**
Equation (12) of HS:	**Equation (12) of HS:**
$q = 2$, $TS = 5.78^{***}$ [0.009], *CVs*: 5.68, 3.73, 2.83	$q = 2$, $TS = 6.04^{***}$ [0.007], *CVs*: 5.68, 3.73, 2.83
$q = 4$, $TS = 4.75^{**}$ [0.014], *CVs*: 5.14, 3.30, 2.48	$q = 4$, $TS = 4.96^{**}$ [0.011], *CVs*: 5.14, 3.30, 2.48
Equation (14) of HS:	**Equation (14) of HS:**
$TS = 3.90^{**}$ [0.015], *CVs*: 4.34, 2.57, 1.82	$TS = 3.98^{**}$ [0.014], *CVs*: 4.35, 2.57, 1.82

***Notes*:** (1) ***, **, and * denote significance at levels 1%, 5%, and 10%, respectively;

(2) *DEFPOP* is the real per capita budget deficit;

(3) Equations (12) and (14) of HS give two alternative estimators of the spectrum;

(4) q is the number of autocovariances of the deficit series taken into account in the first estimator of the spectrum, Equation (12) of HS, which employs the Bartlett kernel; and

(5) the number in square brackets following the value of *TS* is its *p*-value.

3. Summary and Policy Conclusions

This paper applies three tests of deficit sustainability to the Greek budget deficit, which is known to be unsustainable, since (1) there has already been a partial default (see section 1), and (2) Greece is still under a "memorandum of understanding" with the Troika, as it is unable to borrow in international markets. This application can therefore serve as a "litmus test" of the reliability of the three tests used, namely, the unit-root test, the Hakkio-Rush test, and the HS test. Using three definitions of the deficit, i.e., in levels, in per capita terms, and in percent of GDP, I find that the unit-root test rejects sustainability only for the last definition, the Hakkio-Rush test does not reject for any of the three definitions, and the HS test, which is applicable only for the first two definitions (where the deficit series is stationary) rejects at the 1-percent level in both cases. Thus, as was expected, the HS test proves to be more powerful than the traditional tests, which, according to Bohn (2007), are "incapable" of rejecting sustainability. Note, however, that the traditional tests gain power when structural breaks are treated properly; see Trachanas and Katrakilidis (2014).

As a policy conclusion, if the Eurozone countries are to help Greece recover from its depression and its debt crisis, and to avoid financial instability, which would follow Greece's likely default on its debt and its exit from the Eurozone, they should consider adopting the following policies: (1) provide Greece with a substantial debt relief; (2) lessen the austerity measures that have been in effect since May 2010; (3) lift the capital controls that have been in place since June 2015; and (4) request the European Central Bank to play its role as a lender of last resort by providing Greek banks with the necessary liquidity.

Acknowledgment: I am grateful to two anonymous referees of this Journal for their useful comments and suggestions, which improved my paper significantly. Of course, the usual disclaimer applies.

References

[1] Bohn, H. (2007). "Are stationarity and cointegration restrictions really necessary for intertemporal budget constraint?", *Journal of Monetary Economics*, 54(7): 1837-1847.

[2] Cuddington, J.T. (1997). "Analyzing the sustainability of fiscal deficits in developing countries", World Bank, *Policy Research Working Paper* No.1784.

[3] Gregory, A.W., and B.E. Hansen (1996a). "Residual-based tests for cointegration in models with regime shifts", *Journal of Econometrics*, 70(1): 99-126.

[4] Gregory, A.W., and B.E. Hansen (1996b). "Tests for cointegration in models with regime and trend shifts", *Oxford Bulletin of Economics and Statistics*, 58(3): 555-560.

[5] Hakkio, C.S., and M. Rush (1991). "Is the budget deficit 'too large'?", *Economic Inquiry*, 29(3): 429-445.

[6] Hamilton, J.D., and M.A. Flavin (1986). "On the limitations of government borrowing: a framework for empirical testing", *American Economic Review*, 76(4): 808-819.

[7] Hatzinikolaou, D., and T. Simos (2013). "A new test for deficit sustainability and its application to US data", *Empirical Economics*, 45(1): 61-79.

[8] Hatzinikolaou, D., T. Simos, and A. Tsoka (2013). "Is the US current-account deficit sustainable? The importance of structural breaks in testing sustainability", *Economics Bulletin*, 33(4): 2817-2827.

[9] Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin (1992). "Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root?", *Journal of Econometrics*, 54(1-3): 159-178.

[10] Lee, J., and M.C. Strazicich (2003). "Minimum Lagrange multiplier unit root test with two structural breaks", *Review of Economics and Statistics*, 85(4): 1082-1089.

[11] Lee, J., and M.C. Strazicich (2013). "Minimum LM unit root test with one structural break", *Economics Bulletin*, 33(4): 2483-2492.

[12] Liu, P.C., and E. Tanner (1996). "International intertemporal solvency in industrialized countries: evidence and implications", *Southern Economic Journal*, 62(3): 739-749.

[13] Martin, G.M. (2000). "US deficit sustainability: a new approach based on multiple endogenous breaks", *Journal of Applied Econometrics*, 15(1): 83-105.

[14] Organization for Economic Cooperation and Development (OECD) (1992, 1996). *Economic Surveys*: *Greece*, Paris: OECD Publishing.

[15] Payne, J.E. (1997). "International evidence on the sustainability of budget deficits", *Applied Economics Letters*, 4(12): 775-779.

[16] Pesaran, M.H, Y. Shin, and R.J. Smith (2001). "Bounds testing approaches to the analysis of level relationships", *Journal of Applied Econometrics*, 16(3): 289-326.

[17] Tanner, E., and P. Liu (1994). "Is the budget deficit 'too large'? Some further evidence", *Economic Inquiry*, 32(3): 511-518.

[18] Trachanas, E., and C. Katrakilidis (2014). "Is the Greek budget deficit sustainable after all? Empirical evidence accounting for regime shifts," *Applied Economics*, 46(20): 2387-2397.

[19] Uctum, M., T. Thurston, and R. Uctum (2006). "Public debt, the unit root hypothesis and structural breaks: a multi-country analysis", *Economica*, 73(289): 129-156.