

# Indicators and Tests of Fiscal Sustainability: An Integrated Approach

Giancarlo Marini<sup>a</sup> and Alessandro Piergallini<sup>b</sup>

JEL Classification: C12; C22; E60; H60

Keywords: Fiscal Sustainability; Fiscal Indicators; Tests of Government Solvency

May 2007

## Abstract

This paper shows that indicators and tests of government solvency should not be used alternatively. We present a simple and intuitive procedure to integrate simultaneously the results from the two approaches to fiscal sustainability. An application to U.S. post-World War II data demonstrates the empirical relevance of the proposed strategy. Our results suggest that U.S. fiscal policy is on a sustainable path, since the negative values of indicators merely reflect cyclical factors.

---

<sup>a</sup> Department of Economics, University of Rome “Tor Vergata”, Via Columbia 2, 00133, Roma, Italy. E-mail: [marini@uniroma2.it](mailto:marini@uniroma2.it).

<sup>b</sup> Department of Financial and Management Studies, SOAS, University of London, Thornhaugh Street, London WC1H 0XG, United Kingdom. E-mail: [ap50@soas.ac.uk](mailto:ap50@soas.ac.uk).

*Acknowledgments:* We thank Fabrizio Adriani, Barbara Annicchiarico, Bassam Fattouh, Norman Flynn, Pasquale Scaramozzino, Laixiang Sun and Damian Tobin for useful comments and discussions. Financial support from the FIRB project is gratefully acknowledged.

## 1. Introduction

The sustainability of fiscal policy arguably is one of the most debated issues in current macroeconomics. Following the high debt levels experienced by several developed economies since the early 1980s, concern over increasing government debt and deficits and their negative consequences for macroeconomic stability has attracted considerable interest both in the academic literature and in the policy debate. Two possible strategies for an empirical evaluation of fiscal sustainability, based on the government intertemporal budget constraint, appear to be worth pursuing. One strategy is to construct indicators of the sustainability of fiscal policy, along the lines suggested by Miller (1983), Buitier (1985, 1987), Blanchard (1990), and Buitier, Corsetti and Rubini (1993).<sup>1</sup> The other is to implement tests for fiscal solvency, following the seminal works by Hamilton and Flavin (1987), Trehan and Walsh (1988), and Bohn (1998). The indicators presented in the literature are either simple measures mainly based on current information, or summary values of model-based projections of future paths of fiscal policy. The testing techniques so far employed involve unit root tests (Hamilton and Flavin, 1986; Wilcox, 1989), cointegration analysis of fiscal data (Trehan and Walsh, 1988, 1991), or require the adoption of a model-based sustainability approach (Bohn, 1998, 2005).

The purpose of this paper is to demonstrate that indicators and tests for fiscal sustainability may usefully be integrated. We show that the simultaneous use of indicators and tests provides additional information on the issue of government solvency. Indicators

---

<sup>1</sup> Long-term sustainability is periodically evaluated through the use of fiscal indicators, for instance, in the European Union, where the Stability and Growth Pact prescribes specific constraints on the size of government deficits and debts of Member States. See European Commission (2006).

and tests simply reinforce each other when their predictions are not in conflict. However, in the event of conflicting results, indicators may signal the occurrence of a change in policy which may reverse the predictions of tests. This is because indicators respond to a set of current and expected future conditions. Tests, on the other hand, always rely on a sample of past data. For this reason, indicators are faster than tests in responding to changes in fiscal policy regimes. The key point is that a change in regime should cause a structural break in the data generating process of the primary surplus, which can be detected empirically.

We design a simple and intuitive strategy to combine the potentially different results of indicators and tests. To evaluate the empirical relevance of the proposed procedure, we provide an application to U.S. post-World War II data. Analyzing the case of U.S. fiscal policy proves to be particularly appropriate, for the deterioration in U.S. budget deficits since 2001 has given rise to renewed and widespread concerns over the question of sustainability.

The paper is structured as follows. Section 2 describes the most commonly used indicators of fiscal sustainability. Section 3 briefly reviews the main procedures to directly test for solvency. Section 4 presents the strategy we propose to combine the use of indicators and tests. Section 5 develops the empirical application to the U.S. fiscal position. Section 6 concludes.

## **2. Indicators of Fiscal Sustainability**

The “law of motion” of interest bearing public debt derived from the intertemporal budget constraint can be approximated as:

$$b_t = (1 + \rho - n) \cdot b_{t-1} - s_t, \quad (1)$$

where  $\rho$  is the real rate of return on public debt,  $n$  is the rate of growth of real output,  $b_t$  is the interest bearing public debt at the end of period  $t$ , and  $s_t$  is the primary surplus inclusive of seigniorage revenues. Both  $b_t$  and  $s_t$  are ratios to GDP. The assumption of a constant growth-adjusted real rate of interest is primarily made for expositional simplicity. It does not affect the essence of the present analysis.

Solving forward equation (1) for public debt as a function of future surpluses and the public debt at the terminal date  $N$ , one obtains:

$$b_t = \left[ \sum_{i=1}^N (1 + \rho - n)^{-i} \cdot s_{t+i} \right] + (1 + \rho - n)^{-N} \cdot b_{t+N}. \quad (2)$$

Equation (2) implies that when the rate of growth of the economy is greater than the real rate of interest, deficits can always be financed. Hence, “honest” Ponzi games are possible and no problem of solvency arises.

Both equation (1) and (2) have been used to construct indicators of sustainability. Buitier (1985) proposes measures based on the “net worth” approach. Blanchard (1990) suggests the use of two indicators designed to tackle the issue over different time horizons.<sup>2</sup>

The first indicator, called the *primary gap* ( $PG_t$ ), is given by:

$$PG_t = s_t - (\rho - n) \cdot b_t. \quad (3)$$

The primary gap is equal to the primary surplus plus the effect of the debt on the new debt creation.

A more complex set of indicators can be derived from equation (2), imposing a final condition on  $b_{t+N}$  over a predetermined time horizon  $N$  and solving for the primary

---

<sup>2</sup> See also Chouraqui, Hagemann and Sartor (1990).

surplus, or, alternatively, for the policy variable assumed to guarantee the required fiscal adjustment (tax, government expenditure or transfer rates). In particular, Blanchard (1990) proposes the so-called *tax gap*, measuring the change in the tax rate that would be necessary to bring the ratio of public debt to GDP at the current level after a given number of years. The tax gap at time  $t$  for a period of  $N$  years,  $TG_{t,N}$ , is given by:

$$TG_{t,N} = \tau_{t,N}^* - \tau_t, \quad (4)$$

where  $\tau_t$  is the observed tax rate and

$$\tau_{t,N}^* = (\rho - n) \cdot \left\{ \left[ 1 - (1 + \rho - n)^{-N} \right]^{-1} \cdot \sum_{i=1}^N (1 + \rho - n)^{-i} \cdot g_{t+i} + b_t \right\}, \quad (5)$$

where  $g_t$  is the ratio of public expenditure inclusive of transfers to GDP. Close inspection of equation (5) reveals that when  $\rho - n$  and  $N$  are not large, the tax rate  $\tau_{t,N}^*$  is approximately equal to the average value of public spending inclusive of transfers over the next  $N$  years, plus the interest rate net of economic growth multiplied by the outstanding public debt.<sup>3</sup> Blanchard (1990) suggests to construct this indicator for a small value of  $N$ , like three or five years, and for a high value of  $N$ , like twenty or fifty years.

In what follows, we are mainly concerned with indicators such as the primary surplus, the primary gap, and the three or five years tax gap. The reason is that these indicators are mainly based on current and expected future conditions according to the available forecasts and thus give simple “model-free”<sup>4</sup> measures of government solvency. Longer term measures of fiscal sustainability, like the twenty or fifty years tax gap, require

<sup>3</sup> As pointed out in Blanchard’s analysis, a similar indicator can also be constructed to evaluate the expenditure gap.

<sup>4</sup> Gramlich (1990) stresses the importance of model-free measures of sustainability in order to avoid “additional disagreement” among policy makers.

the full specification of the underlying model. The evaluation of this type of indicators is beyond the scope of the present paper.

### 3. Tests for Solvency

Tests for solvency based on the present value budget constraint have originally been proposed by Hamilton and Flavin (1986). The authors point out the similarity of equation (2) with the model of self-fulfilling hyperinflations first proposed by Flood and Garber (1983). Equation (2) is an identity and it is always satisfied ex-post. However, it is possible to test whether creditors perceive the government as solvent. This can be done by considering the ex-ante evaluation of (2), and testing the hypothesis that agents evaluate  $b_t$  as tending to:

$$b_t = \sum_{i=1}^{\infty} (1 + \rho - n)^{-i} \cdot E\{s_{t+i} / \Omega_t\}, \quad (2a)$$

where  $\Omega_t$  denotes the information set available at time  $t$ . Equation (2a) implies that agents expect the present value of public debt-GDP ratio to approach zero as  $N$  tends to infinity, that is:

$$\lim_{N \rightarrow \infty} (1 + \rho - n)^{-N} \cdot E\{b_{t+N} / \Omega_t\} = 0. \quad (2b)$$

Equation (2b) is formally equivalent to the condition for ruling out bubbles in Flood and Garber's model, and can be tested along the same lines. In particular, a direct test for debt sustainability which does not require a-priori assumptions on interest rates could usefully be derived employing the methodology proposed by West (1987) for testing the existence of speculative bubbles. Consider the following autoregressive representation of the surplus:

$$s_{t+1} = \alpha_0 + \alpha_1 \cdot s_t + \alpha_2 \cdot s_{t-1} + \dots + \alpha_q \cdot s_{t-q+1} + v_{t+1}, \quad (6)$$

where  $\alpha$ 's are parameters and  $v_{t+1}$  is a white noise process.<sup>5</sup> Substituting the forecasts based on (6) in equation (2a) and applying the formula derived by Hansen and Sargent (1981) to solve (2a) for the past and present values of the surplus, one obtains:<sup>6</sup>

$$b_{t+1} = \delta_0 + \delta_1 \cdot s_{t+1} + \delta_2 \cdot s_t + \dots + \delta_q \cdot s_{t-q} + w_{t+1}, \quad (7)$$

$$w_{t+1} = \left( \sum_{i=1}^{\infty} f^i \cdot e_{t+1+i} \right),$$

$$e_t = (E\{s_{t+1} / \Omega_t\} - E\{s_{t+1} / \Phi_t\}),$$

where  $\delta$ 's are parameters resulting from highly non-linear combinations of the  $\alpha$ 's and the discount factor  $f$ , and  $\Phi_t$  is the subset of  $\Omega_t$  including only present and lagged values of the surplus.<sup>7</sup> An unrestricted version of the model (6)-(7) can be estimated using OLS to yield, under the null hypothesis of sustainability, consistent and efficient estimates of the parameters. Using equation (2), rewritten in the form

$$b_t = f \cdot (b_{t+1} + s_{t+1}), \quad (8)$$

it is also possible to obtain an instrumental variables estimate of  $f$ . This estimate is consistent under both the null hypothesis of sustainability and the alternative, and can be combined with the OLS estimates of the parameters in (6)-(7) to perform a Wald test of the non-linear restriction implied by (2a-b).

---

<sup>5</sup> Equation (6) should be specified either in levels or in first differences, according to what is necessary to achieve stationarity.

<sup>6</sup> When the surplus equation is specified in differences, it is also convenient to specify the debt equation in the same way.

<sup>7</sup> The exact form of the non-linear constraints can be found in West (1987).

Tests for fiscal sustainability based on (2a-b) have also been derived analyzing the cointegration properties of fiscal variables. The seminal work by Trehan and Walsh (1988) demonstrates that if revenues, spending and debt have unit roots, a sufficient condition for sustainability is the stationarity of with-interest deficit. Equivalently, it is sufficient that the primary surplus and public debt are cointegrated with a cointegrating vector  $[1, -(\rho - n)]$ .

Trehan and Walsh (1991) generalize the cointegration analysis of fiscal data incorporating the possibility of a non-stationary with-interest deficit. Specifically, the authors show that the intertemporal budget constraint holds if a quasi-difference of debt  $b_t - \lambda \cdot b_{t-1}$  is stationary when

$$\lambda \in [0, 1 + \rho - n) \quad (9)$$

and primary surplus and debt are cointegrated.

As pointed out by Bohn (2005), Trehan and Walsh (1991)'s conditions imply that a strictly positive relationship between surplus and debt is sufficient for sustainability. Cointegration between surplus and debt implies that

$$s_t - \alpha \cdot b_{t-1} = u_t, \quad (10)$$

with  $\alpha \neq 0$ , is stationary. Substituting (10) into the flow budget constraint (1) yields:

$$b_t = \lambda \cdot b_{t-1} + u_t, \quad (11)$$

where  $\lambda = 1 + \rho - n - \alpha$ . Hence, Trehan and Walsh's condition (9) requires that  $\alpha > 0$ . This ensures sustainability because it guarantees that the growth rate of debt is strictly lower than the (growth-adjusted) real rate of interest (McCallum, 1984).

As shown by Bohn (1995, 1998, 2005), government solvency can also be tested using a model-based approach. According to this view, sustainability of fiscal policy is a

general equilibrium issue, based on optimizing individual behavior: the government ability to borrow is constrained by private agents' willingness to lend. Assuming complete financial markets and infinitely-lived optimizing individuals, sustainability must imply the respect of private agents' transversality condition:

$$\lim_{N \rightarrow \infty} E\{Q_{t,N} \cdot b_{t+N} / \Omega_t\} = 0, \quad (12)$$

where  $Q_{t,N}$  is the pricing kernel for contingency claims on period  $t+N$ . Combining the Euler equations, characterizing the agents' intertemporal optimality conditions on the consumption-saving decision, with the government's flow budget constraint (expressed with variable interest and growth rates) and applying the transversality condition yields the following intertemporal budget constraint (Bohn, 1995):

$$b_t^* = \sum_{N=0}^{\infty} E\{Q_{t,N} \cdot s_{t+N} / \Omega_t\}, \quad (13)$$

where  $b_t^* = (\rho_t - n_t) \cdot b_{t-1}$  denotes debt at the *beginning* of period  $t$ . Bohn (1998) demonstrates that sustainability can be tested by estimating the following class of policy rules:

$$s_t = \beta \cdot b_t^* + \mu_t, \quad (14)$$

where  $\beta > 0$  and  $\mu_t$  is a bounded set of other determinants of the primary surplus to GDP ratio.<sup>8</sup> A positive reaction coefficient  $\beta$ , in fact, implies:

$$E\{Q_{t,N} \cdot b_{t+N} / \Omega_t\} \approx (1 - \beta)^N \cdot b_{t+N}^* \rightarrow 0. \quad (15)$$

---

<sup>8</sup> The main advantages of this sustainability test are that it does not require any assumptions on interest rates and it maintains its validity under conditions of uncertainty.

Note that if public debt and primary surplus ratios to GDP both have unit roots and  $\mu_t$  is stationary,  $\beta > 0$  ensures cointegration between the two fiscal variables, thus satisfying the sustainability conditions obtained by Trehan and Walsh (1991). As a result, Trehan and Walsh's cointegration condition linking debt to primary surpluses implies an error-correction mechanism that can be interpreted as a fiscal reaction function (Bohn 2007).

In synthesis, sustainability surely holds if in the estimated fiscal policy rule one can detect a positive feedback reaction of the primary surplus to the debt-GDP ratio.

#### **4. Combining Indicators and Tests**

The simultaneous use of tests and indicators provides additional information on the issue of government solvency.<sup>9</sup> Indicators are *forward looking*, responding to a set of current and expected future fiscal conditions and hence not possessing long memory. Tests, on the other hand, although implicitly generating forecasts, are *backward looking*, always relying on historical data. For this reason, indicators may respond faster than tests to changes in the fiscal position and therefore may signal the possible occurrence of a *systematic* change in the policy regime.

The two approaches could usefully be integrated. A possible strategy is shown in Figure 1. The basic idea is that tests and indicators simply reinforce each other when their predictions are not in conflict. In other words, when both tests and indicators (do not) show the existence of a solvency problem, (no) some correction in the fiscal stance is required. In the case of conflicting results, their interpretation is unclear. Indicators may signal the occurrence of a change in policy which may reverse the predictions of tests. A change in

---

<sup>9</sup> An initial attempt to combine the two approaches is in de Luzenberger and Marini (1993).

the policy regime should, however, cause a structural break in the data generating process of the surplus, which can be detected using, e.g., a Chow-type test on the stability of the parameters of the surplus equation. Whether the results from indicators or from tests should be given priority in the assessment of the sustainability of public debt depend thus on the stability of the surplus equation.

In synthesis, in the case of conflicting results, the predictions of fiscal indicators about the issue of government solvency should be accepted if and only if one is able to detect the occurrence of a structural change in the fiscal policy regime. This is because the validity of *all* tests of sustainability is seriously questionable in the presence of a structural break. On the other hand, without a systematic break in policy, the predictions of tests are more reliable since the results of indicators are likely to reflect cyclical factors.

## **5. An Application to U.S. Data**

We apply the strategy suggested to the U.S. data over the post-World War II period. The annual data for the period 1948 to 2006 are obtained from *Historical Tables, Budget of the United States Government, Fiscal Year 2008* (see the Data Appendix for details).

Figures 2-4 show budget deficits, primary surpluses and public debt in the U.S. for 1948-2006, respectively. All variables are scaled by GDP. The U.S. post-World War II period has experienced systematic fiscal deficits. In particular, the deficit-GDP ratio has shown a sharp increase in the post-1970 and in the post-2000 periods. The primary surplus-GDP ratio presents a downward-trend in the 1948-1983 period, an upward-trend in the 1983-2000 period, and a strong deterioration in the post-2000 period. Increases in the public debt-GDP ratio are visible in the 1974-1993 period and in the post-2000 period.

The worsening of the U.S. fiscal balance in the post-2000 period has given rise to serious concerns over a potential sustainability problem. Such concerns are periodically expressed by the Congressional Budget Office (CBO). In a recent study by the CBO (2007) it is contended that “for the years beyond 2017... policies reflected in both CBO’s baseline and the President’s proposals are unsustainable in the long run”.

Figures 5-6 show the behavior of fiscal indicators.<sup>10</sup> Figure 4 displays the primary gap. Figure 5 displays the two, three and five years tax gaps. The computation of the tax gaps for recent years utilizes the forecasts of fiscal variables contained in *Historical Tables, Budget of the United States Government, Fiscal Year 2008*, that are available until 2012. Both the primary gap and the tax gaps show a high volatility within the sample and the possible existence of a sustainability problem starting from 2001. The recent behavior of tax gaps, whose positive values indicate a need for future fiscal corrections, appears to be consistent with the predictions presented by the CBO (2005), which points out that “even if taxation reached levels that were unprecedented in the United States, current spending policies could become financially unsustainable”.

We now test the sustainability of U.S. fiscal policy using historical data. In particular, we test whether the U.S. fiscal history displays a robust positive reaction of the primary surplus-GDP ratio to the public debt-GDP ratio. As shown in Section 3, such a positive response is sufficient for sustainability.

A preliminary analysis of data using unit root tests for the debt-GDP ratio  $b_t$ , the primary surplus-GDP ratio  $s_t$  and the deficit-GDP ratio  $d_t$  is illustrated in Tables 1-3,

---

<sup>10</sup> The fiscal indicators (3) and (4)-(5) are constructed using averages over the last 10 years to express the growth-adjusted interest rate, as suggested by Blanchard (1990).

respectively. From the unit root analysis, both  $b_t$  and  $s_t$  appear to be I(1) over the sample considered. This suggests specifying the possible relationship between the primary surplus-GDP ratio and the public debt-GDP ratio using a dynamic Error Correction Model (ECM) (Sargan, 1964; Engle and Granger, 1987). This empirical specification enables us to incorporate the presence of inertia, which represents a typical feature of policy reaction functions. As a result, we take into account the possibility of a gradual adjustment in the response of the primary surplus to debt. A gradual fiscal stabilization policy can take place due, for example, to the presence of distortionary taxation or political costs. At the same time, the ECM constitutes an adequate estimation procedure in the presence of cointegrated variables according to Granger's Representation Theorem.

The estimated equation is a second-order ECM, given by:

$$\Delta s_t = \beta_0 + \beta_1 \cdot \Delta b_t^* + \beta_2 \cdot \Delta s_{t-1} + \beta_3 \cdot \Delta b_{t-1}^* + \beta_4 \cdot s_{t-2} + \beta_5 \cdot b_{t-2}^* + \mu_t. \quad (16)$$

A non-linear long-run estimator of the reaction of the surplus-GDP ratio to the *beginning-of-period* debt-GDP ratio is given by:

$$\hat{\beta} = -\frac{\hat{\beta}_5}{\hat{\beta}_4}. \quad (17)$$

Table 4 displays the results of the Ordinary Least Squares estimation of the ECM equation (16). The estimated long-run response of the primary surplus to debt is  $(0.031649/0.47349) = 0.0667$ , which is positive and statistically significant. This means that an increase in the debt-GDP ratio, say, by 10% increases the primary surplus-GDP ratio by 0.667% eventually. In the estimated equation, there is neither residual serial

correlation nor heteroskedasticity. The residuals are  $I(0)$  and normally distributed. The hypothesis that the model is correctly specified is statistically supported.

Table 5 confirms the robustness of a positive relationship between the surplus-GDP ratio and the debt-GDP ratio. It shows that we can detect a long-run positive relationship analyzing sub-samples as well. The estimated surplus response to public debt is 0.097 for 1960-2006 (Regression (a)), 0.098 for 1970-2006 (Regression (b)), and 0.099 for 1980-2006 (Regression (c)). This indicates the occurrence of a systematic long-run feedback reaction of the fiscal authorities to the debt-GDP ratio, in favor of the sustainability hypothesis.

To sum up, the “warnings” of fiscal indicators, based on forecasts, are reversed by the predictions of the test, based on historical data, according to which the hypothesis of an unsustainable course of U.S. fiscal policy is strongly rejected. Hence, we test for a structural break to have occurred in 2001.<sup>11</sup> The Chow test is computed for both the primary surplus-GDP ratio when specified in first differences and the data generating process of the primary surplus-GDP ratio given by the fiscal policy reaction function (16).

Tables 6-7 give the results. From Table 6, the value of the Chow test for the change in the surplus-GDP ratio is 1.4524 using 2001-2006 as the forecast period, and 0.78057 using 2001-2012 as the forecast period.<sup>12</sup> Both values of the test statistics are less than the 95% critical values (2.2826 and 1.9475, respectively). From Table 7, the value of the Chow test on the equality of the parameters of the surplus equation (16) is 0.65552

---

<sup>11</sup> The issue of structural shifts in the U.S. fiscal policy history has been extensively discussed in the literature. See, e.g., Hakkio and Rush (1991), Tanner and Liu (1994), Haug (1995), and Quintos (1995). For the identification of “threshold effects” see Arestis, Cipollini and Fattouh (2004), and Cipollini, Fattouh and Mouratidis (2007).

<sup>12</sup> In the latter case, we use the forecasts contained in *Historical Tables, Budget of the United States Government, Fiscal Year 2008*.

using 2001-2006 as the forecast period, and 0.49755 using 2001-2012 as the forecast period. Both values of the test statistics are less than the 95% critical values (2.3083 and 1.9745, respectively). In all cases considered, the hypothesis of a structural break is rejected by the data.

In conclusion, the absence of a structural change in regime is a critical signal in favor of the sustainability hypothesis. It leads us to accept the sustainability predictions of the test and to conclude that the potentially unsustainable course of U.S. fiscal policy from 2001 onwards, suggested by the use of fiscal indicators, is likely to reflect cyclical factors. In other words, indicators are merely signalling the occurrence of a transitory period of high deficits. Such a short term phenomenon does not appear to have any substantial implication for the soundness and consistency of long-term fiscal policy.

## **5. Conclusions**

The evolution of public debt in OECD countries requires some form of monitoring. In the present paper, we have proposed a simple strategy that enables one to integrate the use of indicators and tests of fiscal sustainability. Indicators are *forward looking*, in the sense that they are based on published forecasts, thereby reacting to a set of current and expected future conditions in fiscal policy. Tests, on the other hand, are *backward looking*, in the sense that they are based on a sample of past data. Our integrated approach allows to test the consistency of fiscal plans over the relevant sample period. Fiscal indicators derived from the present value budget constraint should also be calculated. Their interpretation, however, must be made contingent on results by tests, that is, according to whether or not there is a genuine sustainability problem.

For instance, our results for the U.S. post-World War II data reinforce the earlier finding of sustainability obtained by Bohn (1998, 2005). According to the test performed, which is based on an Error Correction Model to allow for the possibility that the primary surplus adjusts gradually to achieve sustainability, the conduct of fiscal policy within sample is consistent with government solvency. On the other hand, indicators show the need for some fiscal adjustment starting from 2001 onwards.

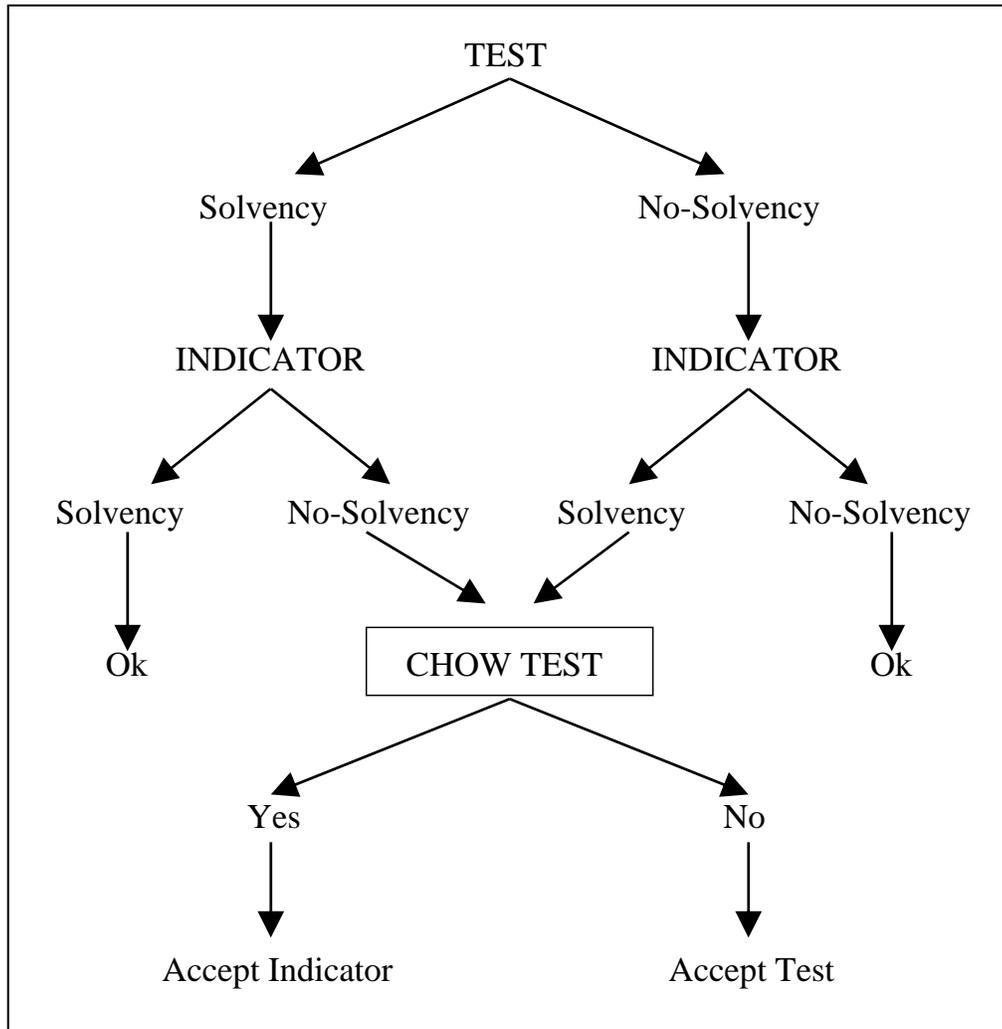
When indicators and tests give conflicting results, the stability of the surplus data generating process should be checked in order to establish whether indicators are merely responding to contingent factors or are detecting the existence of long-term sustainability problems. In the case considered here, the indicators fail to signal the existence of a structural break in the primary surplus to GDP ratio. Deficit increases in the U.S. were thus a transitory phenomenon to be eventually reversed.

In conclusion, an important insight of our paper is that empirical study of government solvency requires the simultaneous analysis of indicators and tests. Research aimed at further integrating the formulation of indicators of sustainability with testing techniques promises to be fruitful for policy evaluation and design.

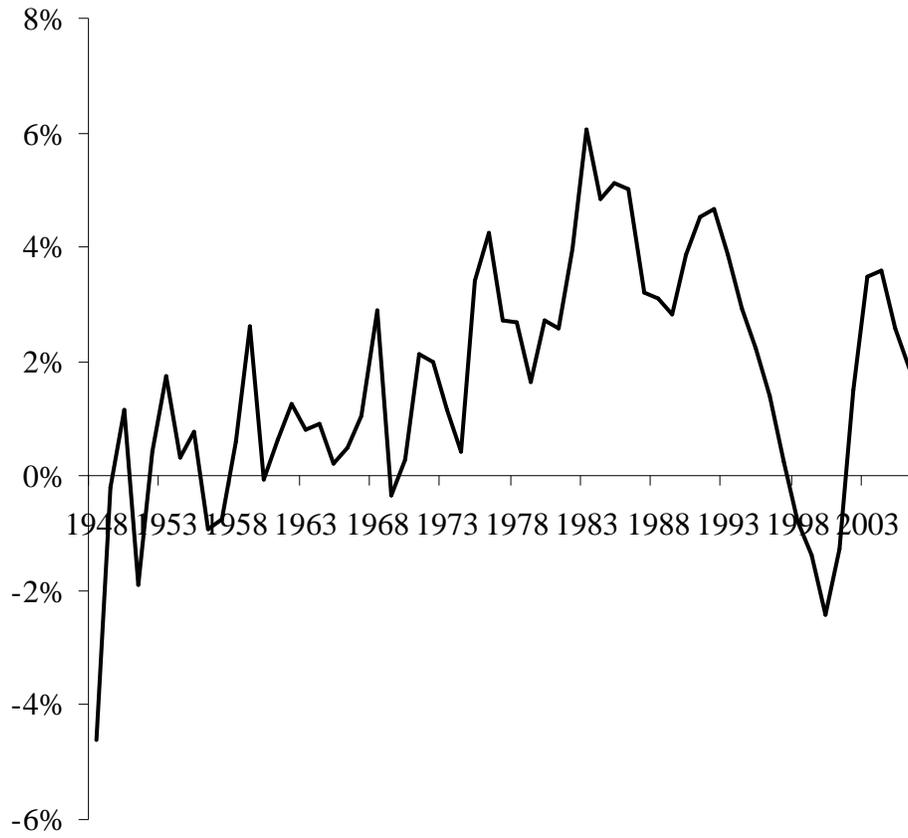
## Data Appendix

The data are taken from *Historical Tables, Budget of the United States Government, Fiscal Year 2008* (URL: <http://www.whitehouse.gov/omb/budget/fy2008/>). The debt series  $b_t$  is the Gross Federal Debt Held by the Public at the end of the fiscal year divided by the GDP of the same fiscal year. The primary surplus  $s_t$  is constructed by dividing the difference of Federal Receipts and Federal Non-Interest Outlays by GDP. Consistently with Bohn (2005), the interest rate on public debt is computed as the ratio of interest payments over the average of outstanding public debt at the start and at the end of each fiscal year. The main advantage of this procedure is that it takes into account the fact that government debt is composed of a portfolio of securities with different interest rates.

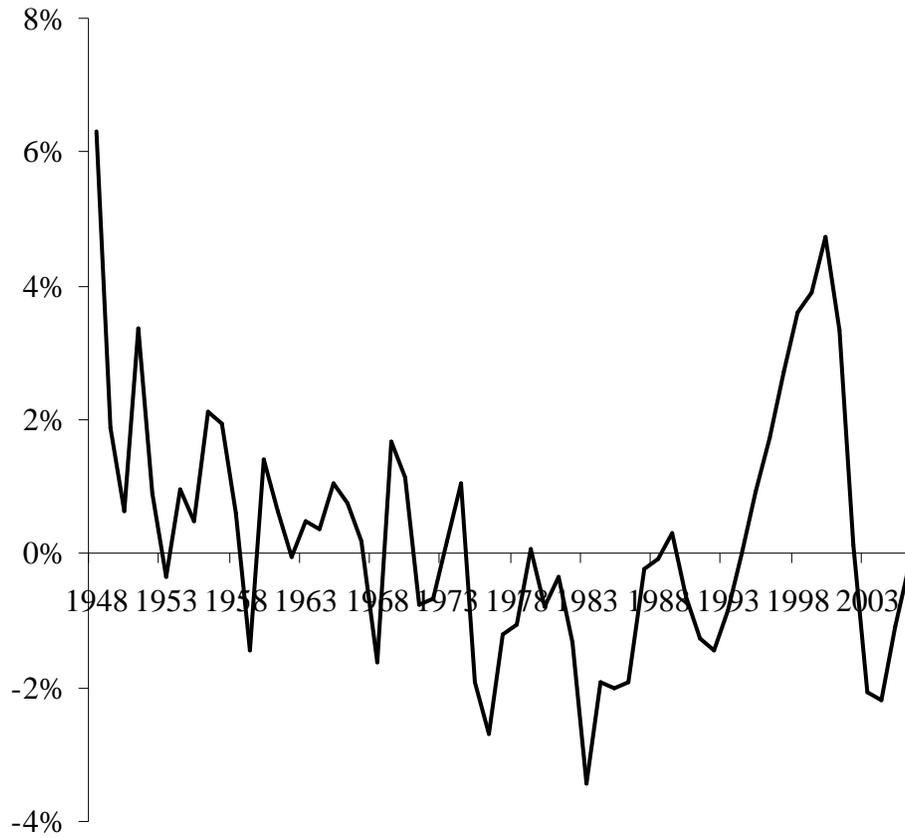
**Figure-1: A Procedure to Combine Indicators and Tests of Government Solvency**



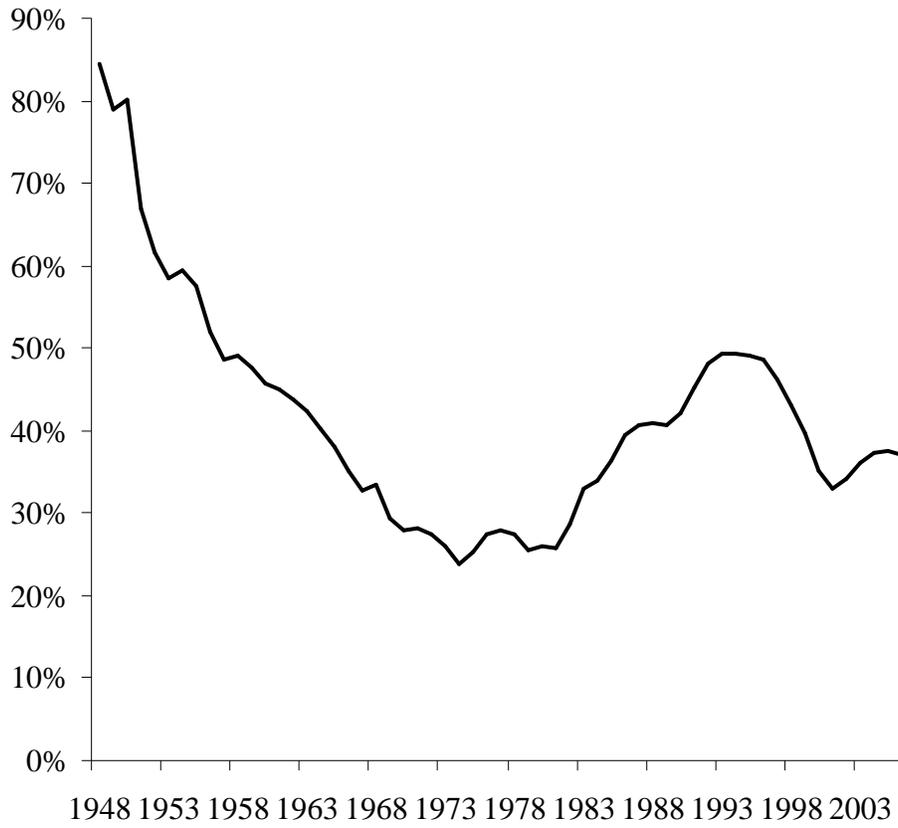
**Figure-2: The U.S. Deficit in Percent of GDP, 1948-2006**



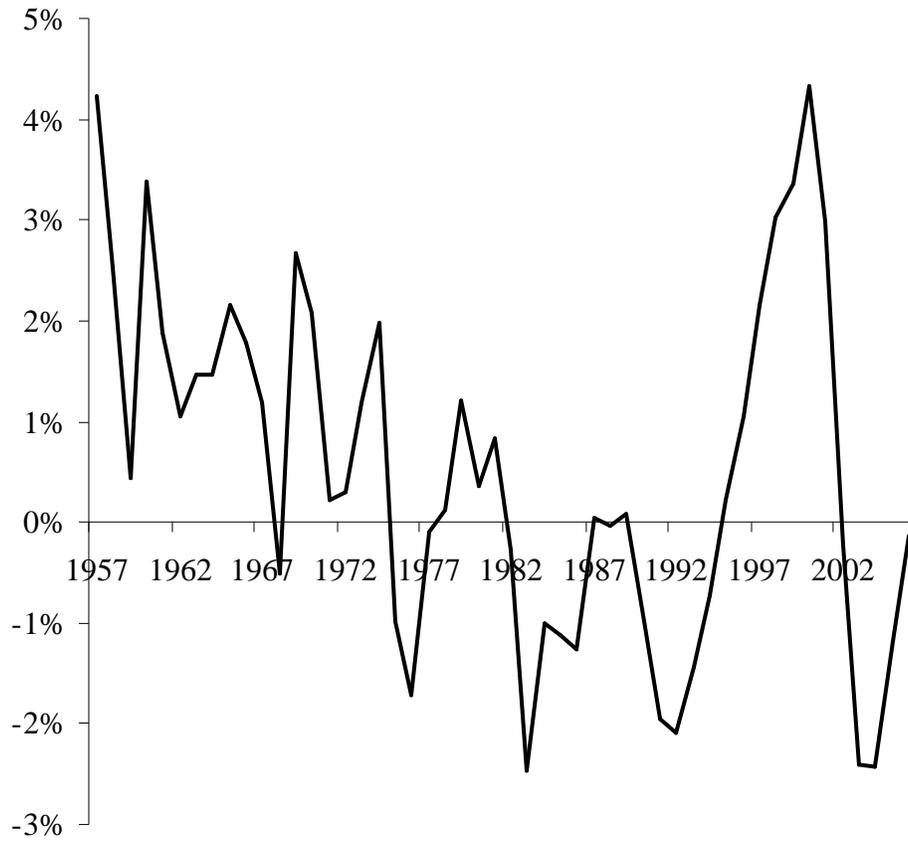
**Figure-3: The U.S. Primary Surplus in Percent of GDP, 1948-2006**



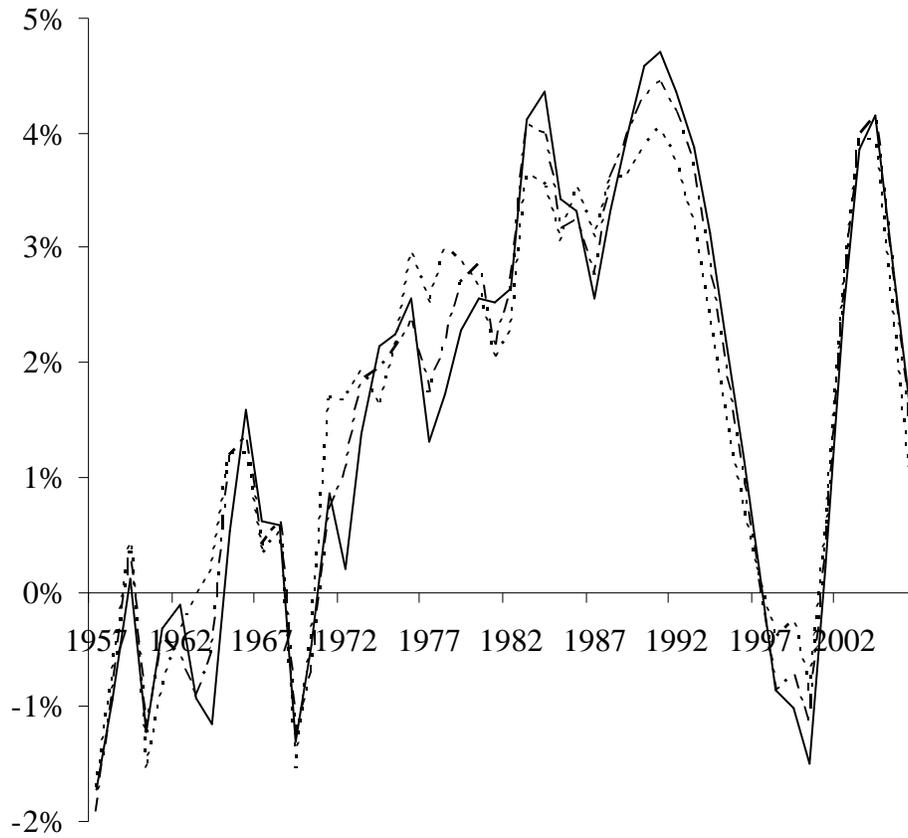
**Figure-4: The U.S. Public Debt in Percent of GDP, 1948-2006**



**Figure-5: The U.S. Primary Gap, 1957-2006**



**Figure-6: The U.S. Tax Gap, 1957-2006**



Legenda:

- 2 Years Tax Gap
- 3 Years Tax Gap
- - - - 5 Years Tax Gap

**Table-1: Unit Root Tests: The U.S. Public Debt in Percent of GDP, 1948-2006****(a) Unit root tests for the variable  $b_t$** 

	Test statistic	LL	AIC	SBC	HQC
<b>DF</b>	-2.3140	136.9898	133.9898	131.0063	132.8392
<b>ADF(1)</b>	-1.8758	146.2019	142.2019	138.2239	140.6677
<b>ADF(2)</b>	-2.0586	147.3098	142.3098	137.3374	140.3922
<b>ADF(3)</b>	-2.0365	147.3103	141.3103	135.3433	139.0090
<b>ADF(4)</b>	-1.9523	149.3421	142.3421	135.3806	139.6573

**(b) Unit root tests for the variable  $\Delta b_t$** 

	Test statistic	LL	AIC	SBC	HQC
<b>DF</b>	-3.7755	140.7638	138.7638	136.7935	138.0061
<b>ADF(1)</b>	-3.8565	141.2893	138.2893	135.3339	137.1528
<b>ADF(2)</b>	-3.5650	141.5923	137.5923	133.6517	136.0770
<b>ADF(3)</b>	-2.6190	143.8900	138.8900	133.9643	136.9958
<b>ADF(4)</b>	-2.4512	144.0220	138.0220	132.1111	135.7489

Notes:

The Dickey-Fuller regressions in Table 1a include an intercept and a linear trend.

The Dickey-Fuller regressions in Table 1b include an intercept but not a trend.

95% critical value for the augmented Dickey-Fuller statistic with an intercept and a linear trend = -3.4935.

95% critical value for the augmented Dickey-Fuller statistic with an intercept but not a trend = -2.9167.

LL = Maximized log-likelihood

AIC = Akaike Information Criterion

SBC = Schwarz Bayesian Criterion

HQC = Hannan-Quinn Criterion

From Table 1a, we cannot reject the hypothesis that the debt-GDP ratio  $b_t$  is I(1). The SBC and the HQC for the Dickey-Fuller regressions with intercept and a linear trend support an ADF(1) specification. The test statistic for the ADF(1) is -1.8758, which is greater than the 95% critical value for the augmented-Dickey-Fuller statistic (-3.4935). The AIC would tend to support the choice of an ADF(4) specification, for which again we cannot reject that  $b_t$  is I(1). The unit root tests reported in Table 1b are consistent with the hypothesis that  $\Delta b_t$  is I(0). The SBC and the HQC for the Dickey-Fuller regressions with intercept and without trend lead to the choice of a DF specification. The value of the test statistic for the DF specification is -3.7755, which is less than the 95% critical value for the Dickey-Fuller statistic (-2.9167). The AIC would instead select an ADF(3) specification, for which the hypothesis of a unit root cannot be rejected even for  $\Delta b_t$ . However, the AIC tends to overestimate the number of parameters whereas the SBC and the HQC are consistent estimators.

**Table-2: Unit Root Tests: The U.S. Primary Surplus in Percent of GDP, 1948-2006****(a) Unit root tests for the variable  $s_t$** 

	Test statistic	LL	AIC	SBC	HQC
<b>DF</b>	-3.0358	161.9850	158.9850	156.0015	157.8344
<b>ADF(1)</b>	-3.3351	162.9315	158.9315	154.9535	157.3973
<b>ADF(2)</b>	-2.7661	163.0613	158.0613	153.0889	156.1437
<b>ADF(3)</b>	-2.7799	163.2719	157.2719	151.3050	154.9707
<b>ADF(4)</b>	-2.5984	163.2803	156.2803	149.3189	153.5955

**(b) Unit root tests for the variable  $\Delta s_t$** 

	Test statistic	LL	AIC	SBC	HQC
<b>DF</b>	-7.2688	154.5477	152.5477	150.5774	151.7900
<b>ADF(1)</b>	-6.3864	155.8301	152.8301	149.8747	151.6936
<b>ADF(2)</b>	-5.0076	155.9676	151.9676	148.0271	150.4523
<b>ADF(3)</b>	-4.7651	156.6700	151.6700	146.7442	149.7758
<b>ADF(4)</b>	-4.1960	156.7360	150.7360	144.8252	148.4630

Notes:

The Dickey-Fuller regressions in Table 2a include an intercept and a linear trend.

The Dickey-Fuller regressions in Table 2b include an intercept but not a trend.

95% critical value for the augmented Dickey-Fuller statistic with an intercept and a linear trend = -3.4935.

95% critical value for the augmented Dickey-Fuller statistic with an intercept but not a trend = -2.9167.

LL = Maximized log-likelihood

AIC = Akaike Information Criterion

SBC = Schwarz Bayesian Criterion

HQC = Hannan-Quinn Criterion

From Table 2a, we cannot reject the hypothesis that the primary surplus-GDP ratio  $s_t$  is I(1). The AIC, the SBC and the HQC all select a DF specification. The test statistic for the DF regression is -3.0358, which is greater than the 95% critical value for the Dickey-Fuller statistic (-3.4935). Table 2b provides empirical support to the hypothesis that  $\Delta s_t$  is I(0). The SBC and the HQC conduct to a DF specification. The value of the test statistic for the DF specification is -7.2688, which is less than the 95% critical value for the Dickey-Fuller statistic (-2.9167). The AIC would tend to support an ADF(1) specification, which again allows us to reject the hypothesis of a unit root in  $\Delta s_t$ .

**Table-3: Unit Root Tests: The U.S. Deficit in Percent of GDP, 1948-2006****(a) Unit root tests for the variable  $d_t$** 

	Test statistic	LL	AIC	SBC	HQC
<b>DF</b>	-2.6884	162.2635	159.2635	156.2801	158.1129
<b>ADF(1)</b>	-2.9097	162.9188	158.9188	154.9408	157.3846
<b>ADF(2)</b>	-2.3767	163.2697	158.2697	153.2972	156.3520
<b>ADF(3)</b>	-2.3401	163.3320	157.3320	151.3651	155.0308
<b>ADF(4)</b>	-2.1496	163.3761	156.3761	149.4147	153.6913

**(b) Unit root tests for the variable  $\Delta d_t$** 

	Test statistic	LL	AIC	SBC	HQC
<b>DF</b>	-7.2656	155.7211	153.7211	151.7508	152.9634
<b>ADF(1)</b>	-6.3709	156.9546	153.9546	150.9992	152.8181
<b>ADF(2)</b>	-4.9955	157.0822	153.0822	149.1416	151.5668
<b>ADF(3)</b>	-4.7317	157.7469	152.7469	147.8212	150.8527
<b>ADF(4)</b>	-4.1413	157.7962	151.7962	145.8853	149.5232

Notes:

The Dickey-Fuller regressions in Table 3a include an intercept and a linear trend.

The Dickey-Fuller regressions in Table 3b include an intercept but not a trend.

95% critical value for the augmented Dickey-Fuller statistic with an intercept and a linear trend = -3.4935.

95% critical value for the augmented Dickey-Fuller statistic with an intercept but not a trend = -2.9167.

LL = Maximized log-likelihood

AIC = Akaike Information Criterion

SBC = Schwarz Bayesian Criterion

HQC = Hannan-Quinn Criterion

From Table 3a, we cannot reject the hypothesis that the deficit-GDP ratio  $d_t$  is I(1). The AIC, the SBC and the HQC all support a DF specification. The value of the test statistic for the DF specification is -2.6884, which is greater than the 95% critical value for the Dickey-Fuller statistic (-3.4935). The results illustrated in Table 3b are consistent with the hypothesis that  $\Delta d_t$  is I(0). The SBC and the HQC lead to select a DF specification. The value of the test statistic for the DF specification is -7.2656, which is less than the 95% critical value for the Dickey-Fuller statistic (-2.9167). The AIC would support an ADF(1) specification, for which again we reject the hypothesis of a unit root in  $\Delta d_t$ .

**Table-4: The Relationship between the Primary Surplus-GDP Ratio and the Public Debt-GDP Ratio, 1948-2006**

Dependent variable: $\Delta s_t$	Regression coefficient
constant	-0.012348** (0.0057002)
$\Delta b_t^*$	-0.23182** (0.10156)
$\Delta s_{t-1}$	-0.44642*** (0.14457)
$\Delta b_{t-1}^*$	0.24227** (0.10213)
$s_{t-2}$	-0.47349** (0.15512)
$b_{t-2}^*$	0.031649** (0.014172)
<i>F</i> test	$F(5,51) = 6.0516***$
Serial correlation <sup>1</sup>	$\chi^2(1) = 0.90644$
Heteroskedasticity <sup>2</sup>	$F(1,55) = 0.091382$
Residuals unit root test	$DF = -7.8465** [-5.0141^3]$
Normality <sup>4</sup>	$\chi^2(2) = 1.5700$
Functional form <sup>5</sup>	$F(1,50) = 0.073850$

Notes:

Standard errors in brackets.

1. Lagrange Multiplier (LM) test of residual serial correlation (Godfrey, 1978a, 1978b).
2. Test based on the regression of squared residuals on squared fitted values (Koenker, 1981).
3. 95% critical value of the Dickey-Fuller statistic (MacKinnon, 1991).
4. Jarque-Bera test of skewness and kurtosis of residuals (Jarque and Bera, 1987).
5. Ramsey's RESET test using the square of the fitted values (Ramsey, 1969).

\* Statistically significant at 10%.

\*\* Statistically significant at 5%.

\*\*\* Statistically significant at 1%.

In the estimated equation (16), we cannot reject the null hypothesis of no residual serial correlation. The value of the LM statistic for the test of AR(1) disturbances is 0.90644, which is less than the 95% critical value (3.8415). Also, we cannot reject the null hypothesis of homoskedasticity. The test for heteroskedasticity based on the regression of squared residuals on squared fitted values (Koenker, 1981) yields a test statistic (0.091382) which is less than the 95% critical value (4.0162). The value of the test statistic for a DF specification of residuals is -7.8465, which is less than the 95% critical value for the Dickey-Fuller statistic (-5.0141), in favor of the hypothesis of I(0) residuals. Moreover, we cannot reject the null hypothesis that the residuals are normally distributed. The value of the Jarque-Bera test statistic is 1.5700, which is less than the 95% critical value (5.9915). Finally, the hypothesis that the model is correctly specified is supported by the Ramsey's RESET test. The test statistic has a value of 0.073850, which is less than the 95% critical value of 4.0343, thereby allowing us to accept the null hypothesis of a correct specification of the surplus equation.

**Table-5: The Relationship between the Primary Surplus-GDP Ratio and the Public Debt-GDP Ratio, 1960-2006, 1970-2006, 1980-2006**

Dependent variable: $\Delta s_t$	Regression (a) 1960-2006	Regression (b) 1970-2006	Regression (c) 1980-2006
constant	-0.032262*** (0.0086217)	-0.042799*** (0.010504)	-0.036184*** (0.011352)
$\Delta b_t^*$	-0.30951 (0.20739)	-0.51451** (0.24007)	-0.50100* (0.27771)
$\Delta s_{t-1}$	-0.63980*** (0.20522)	-0.78154*** (0.25992)	-0.59149* (0.30600)
$\Delta b_{t-1}^*$	-0.053280 (0.21558)	-0.18874 (0.28135)	-0.11619 (0.31165)
$s_{t-2}$	-0.88299*** (0.20700)	-1.2039*** (0.29283)	-1.0222*** (0.33321)
$b_{t-2}^*$	0.085220*** (0.022587)	0.11848*** (0.029312)	0.10079*** (0.030636)
<i>F</i> test	$F(5,41) = 5.9024***$	$F(5,31) = 7.1933***$	$F(5,21) = 6.1092***$
Serial correlation <sup>1</sup>	$\chi^2(1) = 1.8774$	$\chi^2(1) = 1.6255$	$\chi^2(1) = 0.079803$
Heteroskedasticity <sup>2</sup>	$F(1,45) = 0.21830$	$F(1,35) = 0.086148$	$F(1,25) = 0.18042$
Residuals unit root test	$DF = -6.2060**$ [-5.0823 <sup>3</sup> ]	$DF = -5.4768**$ [-5.1890 <sup>3</sup> ]	$DF = -5.1777*$ [-5.3798 <sup>3</sup> ]
Normality <sup>4</sup>	$\chi^2(2) = 1.0257$	$\chi^2(2) = 0.63101$	$\chi^2(2) = 3.4543$
Functional Form <sup>5</sup>	$F(1,40) = 0.017533$	$F(1,30) = 3.8746*$	$F(1,20) = 6.6691**$

Notes:

Standard errors in brackets.

1. Lagrange Multiplier test of residual serial correlation (Godfrey, 1978a, 1978b).
2. Test based on the regression of squared residuals on squared fitted values (Koenker, 1981).
3. 95% critical value for the Dickey-Fuller statistic (MacKinnon, 1991).
4. Jarque-Bera test of skewness and kurtosis of residuals (Jarque and Bera, 1987).
5. Ramsey's RESET test using the square of the fitted values (Ramsey, 1969).

\* Statistically significant at 10%.

\*\* Statistically significant at 5%.

\*\*\* Statistically significant at 1%.

**Table-6: Chow Tests for the Variable  $\Delta s_t$ , 2000 Split**

Chow tests <sup>1</sup>	$F(6,51) = 1.4524^2$ $F(12,51) = 0.78057^3$
-------------------------	--

Notes:

1. Chow's (1960) second test.
2. Sample period for structural stability: 2001-2006.
3. Sample period for structural stability: 2001-2012. For the 2007-2012 period, the forecasts obtained from *Historical Tables, Budget of the United States Government, Fiscal Year 2008* are used.

**Table-7: Chow Tests for the Primary Surplus Equation, 2000 Split**

<b>Dependent variable:</b> $\Delta s_t$	<b>Regression coefficient</b> <b>1948-2000</b>
constant	0.0092531 (0.0063331)
$\Delta b_t^*$	-0.21958** (0.10416)
$\Delta s_{t-1}$	-0.46849** (0.10416)
$\Delta b_{t-1}^*$	0.25559*** (0.10552)
$s_{t-2}$	-0.36667** (0.19652)
$b_{t-2}^*$	0.025352* (0.015342)
<i>F</i> test	$F(5,45) = 3.8079***$
Chow tests <sup>1</sup>	$F(6,45) = 0.65552^2$ $F(12,45) = 0.49755^3$
Serial correlation <sup>4</sup>	$F(1,44) = 2.5140$
Heteroskedasticity <sup>5</sup>	$F(1,49) = 0.36184$
Residuals unit root test	$DF = -7.5523** [-5.0517^6]$
Normality <sup>7</sup>	$\chi^2(2) = 0.34829$
Functional Form <sup>8</sup>	$F(1,44) = 0.0097583$

Notes:

Standard errors in brackets.

1. Chow's (1960) second test of equality between sets of coefficients.
2. Sample period for structural stability: 2001-2006.
3. Sample period for structural stability: 2001-2012. For the 2007-2012 period, the forecasts obtained from *Historical Tables, Budget of the United States Government, Fiscal Year 2008* are used.
4. Lagrange Multiplier test of residual serial correlation (Godfrey, 1978a, 1978b).
5. Test based on the regression of squared residuals on squared fitted values (Koenker, 1981).
6. 95% critical value for the Dickey-Fuller statistic (MacKinnon, 1991).
7. Jarque-Bera test of skewness and kurtosis of residuals (Jarque and Bera, 1987).
8. Ramsey's RESET test using the square of the fitted values (Ramsey, 1969).

\* Statistically significant at 10%.

\*\* Statistically significant at 5%.

\*\*\* Statistically significant at 1%.

## References

- Arestis, P., Cipollini, A., and Fattouh, B. (2004), “Threshold Effects in the U.S. Budget Deficit”, *Economic Inquiry* 42, 214-222.
- Blanchard, O.J. (1990), “Suggestions for a New Set of Fiscal Indicators”, *OECD Economics Department Working Paper* N. 79.
- Bohn, H. (1995), “The Sustainability of Budget Deficits in a Stochastic Economy”, *Journal of Money, Credit and Banking* 27, 257-271.
- Bohn, H. (1998), “The Behavior of U.S. Public Debt and Deficits”, *Quarterly Journal of Economics* 113, 949-963.
- Bohn, H. (2005), “The Sustainability of Fiscal Policy in the United States”, *CESifo Working Paper* N. 1446.
- Bohn, H. (2007), “Are Stationarity and Cointegration Restrictions Really Necessary for the Intertemporal Budget Constraint?” *Journal of Monetary Economics*, forthcoming.
- Buiter, W.H. (1985), “A Guide to Public Sector Debt and Deficits”, *Economic Policy* 1, 14-79.
- Buiter, W.H. (1987), “The Current Global Economic Situation, Outlook and Policy Options, with Special Emphasis on Fiscal Policy Issues”, *CEPR Discussion Paper* N. 210.
- Buiter, W.H., Corsetti, G., and Rubini, N. (1993), “Excessive Deficits: Sense and Nonsense in the Treaty of Maastricht”, *Economic Policy* 8, 57-100.
- Chouraqi, J-C., Hagemann, R.P., and Sartor, N. (1990), “Indicators of Fiscal Policy: A Reassessment”, *OECD Economics Department Working Paper* N. 78.
- Chow, G.C. (1960), “Tests of Equality between Sets of Coefficients in Two Linear Regressions”, *Econometrica* 28, 591–605.
- Cipollini, A., Fattouh, B., and Mouratidis K. (2007), “Fiscal Readjustments in the U.S.: A Non-linear Time Series Analysis”, *CeFiMS Discussion Paper* N. 68.
- Congressional Budget Office (2005), *Long-Term Budget Outlook*, December.
- Congressional Budget Office (2007), *An Analysis of the President’s Budgetary Proposals for Fiscal Year 2008*, March.

- De Luzenberger, R., and Marini, G. (1993), “The Sustainability of Public Debt: Tests and Indicators”, in *Politica Fiscale e Debito Pubblico*, Ministero del Tesoro.
- Engle, R.F., and Granger, C.W.J. (1987), “Cointegration and Error–Correction: Representation, Estimation, and Testing”, *Econometrica* 55, 391–407.
- European Commission (2006), “The Long-term Sustainability of Public Finances in the European Union”, *European Economy*, N. 4.
- Flood, R.P., and Garber, P.M. (1980), “Market Fundamentals Versus Price Level Bubbles: The First Tests”, *Journal of Political Economy* 88, 745-770.
- Godfrey, L.G. (1978a), “Testing against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables”, *Econometrica* 46, 1293–1301.
- Godfrey, L.G. (1978b), “Testing for Higher Order Serial Correlation in Regression Equations when the Regressors Include Lagged Dependent Variables”, *Econometrica* 46, 1303–1310.
- Gramlich, E.M. (1990), “Fiscal Indicators”, *OECD Economics Department Working Paper* N. 80.
- Hakkio, C.S., and Rush, M. (1991), “Is the Budget Deficit “Too Large?”” *Economic Inquiry* 29, 429-445.
- Hamilton, J.D., and Flavin, M.A. (1986), “On the Limitations of Government Borrowing: A Framework for Empirical Testing”, *American Economic Review* 76, 808-819.
- Hansen, L.P., and Sargent, T.J. (1981), “Formulating And Estimating Dynamic Linear Rational Expectations Models”, in Lucas R.E., and Sargent, T.J. (ed.), *Rational Expectations and Econometric Practice*, University of Minneapolis Press, Minneapolis.
- Haug, A. (1995), “Has Federal Budget Deficit Policy Changed in Recent Years?”, *Economic Inquiry* 33, 104-118.
- Jarque, C.M., and Bera, A.K. (1987), “A Test for Normality of Observations and Regression Residuals”, *International Economic Review* 55, 163-172.
- Koenker, R. (1981), “A Note on Studentizing a Test for Heteroskedasticity”, *Journal of Econometrics* 17, 107–112.

- MacKinnon, J. (1991), "Critical Values for Cointegration Tests", in Engle R.F. and Granger, C.W.J. (eds.), *Long-run Economic Relationships*, Oxford University Press, Oxford, 267-276.
- McCallum, B. (1984), "Are Bond-financed Deficits Inflationary? A Ricardian Analysis", *Journal of Political Economy* 92, 125-135.
- Miller, M. (1983), "Inflation Adjusting the Public Sector Financial Deficit", in Kay, J. (ed.), *The 1982 Budget*, Basil Blackwell, London.
- Quintos, C.E. (1995), "Sustainability of the Deficit Process with Structural Shifts", *Journal of Business and Economic Statistics* 13, 409-417.
- Ramsey, J.B. (1969), "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis", *Journal of the Royal Statistical Society* 31, 350-371.
- Sargan, D. (1964), "Wages and Prices in the United Kingdom: A Study in Econometric Methodology", in P.E. Hart, G. Mills and J.K. Whitaker (eds.), *Econometric Analysis of National Economic Planning*, London, Butterworth, 25-63.
- Schwert, W.G. (1987), "Effects of Model Specification on Tests for Unit Roots in Macroeconomic Data", *Journal of Monetary Economics* 20, 73-103.
- Schwert, W.G. (1989), "Tests for Unit Roots: A Monte Carlo Investigation", *Journal of Business and Economic Statistics* 7, 149-159.
- Tanner, E., and Liu, P. (1994), "Is the Budget Deficit 'Too Large': Some Further Evidence", *Economic Inquiry* 32, 511-518.
- Trehan, B., and Walsh, C. (1988), "Common Trends, The Government Budget Constraint, and Revenue Smoothing", *Journal of Economics Dynamics and Control* 12, 425-444.
- Trehan, B., and Walsh, C. (1991), Testing Intertemporal Budget Constraints: Theory and Applications to U.S. Federal Budget and Current Account Deficits", *Journal of Money, Credit and Banking* 23, 210-223.
- West, K.D. (1987), "A Specification Test for Speculative Bubbles", *Quarterly Journal of Economics* 102, 553-580.
- Wilcox, D.W. (1989), "The Sustainability of Government Deficits: Implications of the Present-Value Borrowing Constraint", *Journal of Money, Credit and Banking* 21, 291-306.