



HAL
open science

Public debt, the unit root hypothesis and structural breaks: a multi-country analysis

Merih Uctum, Thom Thurston, Remzi Uctum

► **To cite this version:**

Merih Uctum, Thom Thurston, Remzi Uctum. Public debt, the unit root hypothesis and structural breaks: a multi-country analysis. *Economica*, 2006, 73 (289), pp.129-156. halshs-00081527

HAL Id: halshs-00081527

<https://shs.hal.science/halshs-00081527v1>

Submitted on 23 Jun 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Public Debt, the Unit Root Hypothesis and Structural Breaks: A Multi-Country Analysis

by

Merih Uctum*, Thom Thurston, Remzi Uctum*****

June 2004

*Brooklyn College and the Graduate Center of the City University of New York.
**Queen's College and the Graduate Center of the City University of New York.
***Centre National de la Recherche Scientifique and Université de Paris X Nanterre.

Abstract

We assess fiscal performances in G7 and selected Latin American and Asian countries. We analyze two questions: (i) have public finances been sustainable? (ii) do countries follow more restrictive fiscal policies when debt starts to rise? We find that: (i) The traditional unit root tests often overlook the corrective actions taken by many governments. Controlling for structural breaks changes the nonstationarity results dramatically among the three groups; (ii) Estimation of a reaction function for governments, expanded by incorporating structural breaks, provides further evidence for significant active anti-debt policies among the G7 and to a lesser extent in the other regions.

Key Words: Fiscal policy, sustainability, government reaction function, structural breaks.
JEL Classification: E6, H87, H6, O23.

*Contact author: Merih Uctum, Economics Department, Brooklyn College of the CUNY, 2900 Bedford Avenue, Brooklyn, NY 11210. Phone: 732-549-5252. Email: muctum@brooklyn.cuny.edu, uctum@econ.rutgers.edu.

The authors would like to thank participants at the North American Summer Meetings of the Econometric Society, Washington, 2001, the Eastern Economics Association Meetings, New York, 2001, Applied Econometric Association Meetings, Toledo, 2003, Michael Binder for useful discussions and, in particular, Georgios Chortareas, Mike Wickens and Hiroki Tsurumi for many helpful suggestions and an anonymous referee whose comments improved the paper considerably. The first two authors gracefully acknowledge financial support from CUNY Research Foundation, Collaborative Grant #919210001.

INTRODUCTION

Notwithstanding a brief episode of surpluses during the late 1990s in most industrial countries, subsequent economic slowdown, aging of population and the burden that these impose on government budgets has kept alive the discussion of the sustainability of fiscal policy. Once investors begin to expect deterioration in the fiscal stance of the governments, currency crises may not be far behind. In assessing the soundness of fiscal policies, investors often project the long-horizon consequences of fiscal patterns that have occurred to date.

This paper evaluates the long-horizon implications of fiscal policies of countries, employing various statistical techniques. In particular, we examine two aspects of the same question: (i) have countries' public policies been sustainable in the presence of structural changes in the economy? (ii) do countries pursue active policies to avoid excessive debt accumulation?

The traditional approach in investigating the sustainability of countries' fiscal policy is based on the intertemporal budget constraint (IBC) of the government. Fiscal policy is considered sustainable if the IBC is expected to hold in present value terms. In other words, the no-Ponzi condition must be satisfied whereby debt-holders expect the current debt to be offset by the sum of expected future discounted primary budget surpluses.

The most widely adopted methodology consists of testing the null hypothesis of unit root in the statistical process governing an appropriately discounted government debt or deficit series or testing for no cointegration between government spending and revenue.ⁱ These tests frequently do not reject the null of a unit root in the series. The tests however are of notoriously low power in small sample and singularly suspect if there is reason to believe the country has experienced a shift or break in its fiscal management over the estimating sample. An alternative and complementary approach that bypasses some of the

technical difficulties associated with unit root tests is to test a hypothesis that governments systematically reduce deficits (increase surpluses) as the level of their debt rises. If such a mechanism is present, it provides a sufficient condition for stationarity of the debt or deficit variable and therefore sustainability of the budget process.ⁱⁱ

In the first part of this paper, we provide evidence as to how the traditional unit-root tests may be poorly designed to reject nonstationarity in the presence of breaks. Allowing for breaks changes the outcome dramatically for a majority of the countries in our sample. When countries with positive trends are excluded, mean-reversion appears in more than half of the countries. This compares with a single case of mean-reversion with the traditional tests and reverses the tendency not to reject unit roots when breaks are not allowed for.

In the second part of the paper, we address the second issue, whether accumulating debt tends to lead to more restrictive policies. For this, we develop a reaction function based on Barro's (1979) tax-smoothing model where a positive feedback from government debt to the primary surplus indicates stationarity in the debt variable. We estimate the reaction function for governments in a sample of eighteen countries to analyze the policy response of governments to debt accumulation, and check for potential breaks. We find substantial anti-debt policies in several countries. This result reinforces the conclusion of the unit root tests with endogenous breaks.

Taken together, the two approaches provide evidence that policies in several countries in our sample have been sustainable. This result is consistent with the widely discussed debt and deficit policies adopted by South American countries after the 1980 debt crisis, the fiscal restraint in many Asian countries prevailing before the Asian crisis and the fiscal rules adopted by the European countries to satisfy the conditions required for membership in the monetary union.

The paper is organized as follows. In the first section we overview the stylized facts. In sections II to IV, we explain the model, the methodology employed to test for sustainability and the data. We present the empirical results in the fifth section and conclude in the last section.

I. STYLIZED FACTS AND MOTIVATION

Analysts frequently note that since the 1970s, governments' debt gradually increased without rebate in most countries. On average, the debt/GDP ratios of the G7 countries have been increasing between the first oil-price shock and the 1990s, except that of the United Kingdom (Figure 1a). An initially declining trend in the United States, Canada, France and Germany reversed between the first and the second oil-price shocks and government debt has been on an increasing trend, at least into the mid-1990s. The Italian and the Japanese debt/GDP ratios show a positive trend since 1970. In the second half of the 1990s, however, debt ratios started to decline in Italy, Canada and the US, and earlier in the UK.

The picture is more complex for developing economies. Government debt ratios in Latin America (LA) were increasing even before the oil-price shocks (Figure 1b). Although these shocks adversely affected the debt/GDP ratio in several countries, government debt swelled further due to the spillover from the rapidly growing external debt. However, after the debt crisis of the 1980s, several countries managed to reverse debt accumulation and some have been since on a negative trend (e.g., Mexico, Uruguay, and Honduras).

The debt patterns for Asian countries are more heterogeneous (Figure 1c). Despite an early reversal, the trend in Malaysia and Thailand turned upward since the mid-1990s. In India, Indonesia and Turkey, the upward trend does not show a clear reversal, while in Korea the debt/GDP ratio has been declining after peaking in mid-1980s.

How in the face of such diverse patterns in debt can we test whether the fiscal policies are sustainable? As indicated, the standard test of sustainability is to examine whether the process generating the fiscal variables is stationary --i.e. does not contain a unit root within the existing time series. A unit root would imply that the current fiscal stance is unsustainable because, if the policy is extended into the indefinite future, it would violate the IBC.

However, the unit root tests are not insolvency tests of the government (Cuddington, 1997). If the tests find that the current fiscal situation is unsustainable, the government is expected to correct its policy to avoid insolvency and thus satisfy its IBC. Corrections to the existing policy process may include increased emphasis on primary surplus generation, inflation surprises that wipe out debt or even renegeing on debt. The required corrections can be smooth, traumatic if delayed, or may have already occurred in the historical time series data. In the latter two cases fiscal time series may reflect a structural change. Not accounting for these structural breaks would lead to nonrejection of the unit root hypothesis under the standard tests. We therefore use a methodology that allows testing the null of a unit root against the alternative of a structural break at unknown date. We show that this methodology dramatically increases the frequency of unit root rejection, thereby implying that the underlying fiscal processes in many countries have been stationary.

Another source of nonrejection of unit roots may simply be due to the low power that these tests have in short sample. To bypass this problem, we estimate a reaction function of government surplus in response to debt. A positive coefficient suggests mean reversion in debt and achieving efficiency in its estimate does not involve the long sample typically required in the unit root tests. As the dotted line in Figures 1a-1b shows, debt patterns are often correlated with primary balances generated by governments. Declining debt/GDP ratios are associated with prolonged surpluses (Costa Rica and Uruguay throughout the

1980s, Mexico, Malaysia and Thailand in late 1980s, and Canada, France, Germany, Italy and the US in the late 1990s). The second test checks whether the observed correlation between these two fiscal variables is due to a response by the government to debt accumulation.

In general, each test brings a different perspective to the question of sustainability of the fiscal stance of the government. The unit root tests describe the data generating process characterizing the debt series, while the estimates of the reaction function reflect the role of the government in shaping the fiscal policy.

II. MODEL

1. Government intertemporal budget constraint

The budget constraint of the government can be written as

$$(1) \quad \Delta B_t = -S_t + \rho_t B_{t-1},$$

where B is a measure of government debt, S is the seignorage inclusive primary surplus, and ρ is the *ex-post* interest rate on the outstanding stock of government debt. Earlier studies use various measures of government debt. Following the literature, we consider real debt, and debt normalized by GDP.ⁱⁱⁱ

Each debt measure has a corresponding discount rate, ρ . If B is real debt, then ρ is real interest rate, and if B is defined as debt/GDP ratio then ρ is the real interest rate adjusted for growth rate of real GDP.

Assuming perfect foresight, solving (1) forward and successively substituting out the future discounted debt measure, gives the n-period intertemporal budget equation:

$$(2) \quad B_t = \delta_{t,n} B_{t+n} + \sum_{i=1}^n \delta_{t,i} S_{t+i}$$

where $\delta_{t,n} = \prod_{s=1}^n (1 + \rho_{t+s})^{-1}$ is the time-varying discount factor n-periods ahead. A necessary and sufficient condition for sustainability is that as n goes to infinity, $\lim_{n \rightarrow \infty} \delta_{t,n} B_{t+n} = 0$. This is also known as the transversality condition and implies the debt market will not tolerate Ponzi games under which new debt is issued systematically to cover debt servicing. If this condition holds then $B_t = \sum_{i=1}^{\infty} \delta_{t,i} S_{t+i}$, and the IBC is satisfied.

The IBC thus imposes restrictions on the long-run relationship between expenditures and revenues of the government by requiring these two not to drift too far away from each other and the government to generate enough future net primary surpluses to pay back the outstanding stock of debt. If the IBC is satisfied, any short-run debt accumulation would, therefore, be mean reverting in the long run and the budget would be balanced in present value terms.

The usual procedure to investigate this condition is to test the hypothesis, using historical data, that the data generating process (DGP) for discounted debt is nonstationary. Rejection of the null of nonstationarity indicates that the sustainability condition holds and the no-Ponzi game condition is satisfied asymptotically. The IBC then suggests that government revenue and expenditure can continue to follow into future their past stochastic process without losing lenders' confidence. If the null hypothesis is not rejected, the present value of budget constraint is continually violated, the current policy is not sustainable and thus has to be changed (see Hamilton and Flavin, 1986, Wilcox, Trehan and Walsh, 1988, Wickens and Uctum, 1993 among others for an application of the transversality condition to government debt).

2. Government's reaction function

If there exists a positive relation between the primary surplus and the level of government debt – that is, governments react systematically to rising debts by increasing the current surplus-, this provides a sufficient condition that debt is mean reverting (see endnote 2). In Barro (1979) the government chooses a stream of future tax collections to minimize the present value of collection costs. Using the first-order conditions Barro derives an equation for the growth rate of nominal debt that varies positively with the deviation of government spending and negatively with income from their respective long-term (normal) trends. Based on this model, Bohn (1999) considers a reaction function of the form $S_t = \mu + \alpha B_t + \phi Z_t + \varepsilon_t$, where Z is a set of cyclical determinants of the primary surplus representing the Barro variables, such as temporary government spending and deviation of output from its trend.

Using a similar but simplified framework, we get a reaction function:

$$(3) \quad S_t = \mu + \alpha B_{t-1} + \beta \tilde{g}_t + \gamma \tilde{y}_t + v_t$$

where all variables except \tilde{y} are expressed as ratios to output, \tilde{g}_t and \tilde{y}_t are deviations of government spending and income from their respective normal levels. The surplus is expected to decrease with an increase in the transitory government spending ($\beta < 0$) and increase with deviations of output above its trend ($\gamma > 0$). The condition of sustainability is $\alpha > 0$, a positive feedback from government debt to the primary surplus. Finding such a coefficient with acceptable statistical confidence would be strong evidence that the fiscal policy has been on a sustainable path.

III. METHODOLOGY

1. The unit root test approach to testing for sustainability

The test based on equation (2) consists of conducting a stationarity test on the discounted value of the government debt. The traditional approach is to use the Augmented Dickey-Fuller (1979) stationarity tests of the type:

$$(4) \quad \Delta b_t = \mu_0 + \mu_1 Trend + \phi b_{t-1} + \sum_{i=1}^n \beta_i \Delta b_{t-i} + \varepsilon_t$$

To test for stationarity in each country's debt series, we estimate individually the equation (4), where we use as the dependent variable a compound discounted debt measure. The null hypothesis of a unit-root is $\phi = 0$ and $\mu_1 = 0$.

It is generally agreed that a structural break in a series biases results toward not rejecting the null hypothesis of a unit root. Many of the countries we consider have been subject to various economic shocks likely to affect their debt policies and, as discussed earlier, the behavior of their debt ratios hints of changes in debt policies. We therefore implement unit root tests with endogenous breaks to check if the series appear trend-stationary when the break is accounted for. For this, we conduct the analysis with procedures by Zivot and Andrews (1992), which test the null of a unit root with drift against the alternative of trend-stationarity with an unknown break in the trend function.

Zivot and Andrews (ZA) consider three models along the lines of Perron (1989). Model A admits an exogenous shift in the mean of the series, Model B in the trend and Model C allows for both changes (see appendix). Unlike Perron, the ZA methodology assumes that no prior information concerning the nature or the date of the break is available to the investigator. Treating the break endogenously is important because it prevents a data-dependent arbitrary choice of the break point. Following ZA, we estimate each model by shifting the break date successively in an interval excluding the extreme points of the data

sample, and compute each time the t-statistics associated to ϕ under the null of a unit root. For each model we retain the breakpoint leading to the most favorable results towards the rejection of the null hypothesis. We reject the null of a unit root if the minimum t statistic is less than the critical value from the asymptotic distribution derived by ZA for each model.^{iv}

The ZA test methodology determines whether the series are trend-stationary in the presence of a one-time structural break in the series. However, sustainability in a world with no structural break requires trendless stationarity. Once we introduce the breaks, the interpretation of sustainability and that of trends becomes more challenging. Although a positive trend is inconsistent with the idea of stationarity, in this paper we argue that a negative trend in a finite sample does not have the same interpretation as a positive trend for sustainability in an infinite horizon.

According to the theory with infinite horizon, a debt series with a negative trend means that the government would eventually become a creditor and continue accumulating assets, creating dynamic inefficiency. In a finite-horizon model, however, the transversality condition forces the government to generate sufficient surpluses to pay back its debt by the last period. In a finite, small sample as ours, a negative (broken) trend can be interpreted as government paying back its debt at least partially. For example, when the government reverts its policy and starts a fiscal austerity program, this would show up as a break in debt accumulation followed by a decline in debt. The breakpoint in this case would correspond to the specific date at which the government implements the change in policy. An alternative scenario is when the government starts with high debt but with a negative trend throughout the sample with or without a break. In both cases, the theory suggests that the out-of-sample debt can become asymptotically negative.

However, it would be an unwarranted inference that a negative trend would imply that the country would indefinitely decumulate debt. An in-sample negative trend simply

means that the government is following a policy that is reducing its debt for a finite period so that the fiscal stance is sustainable. A positive trend, on the other hand, is evidence of unsustainability and the government would not have a choice but to change its policy in the future, as discussed above.

In analyzing the results from our finite sample period, we will interpret fiscal policy as sustainable if the null of a unit root is rejected against the alternative of zero-trend stationarity as well as negative-trend stationarity. Thus our criteria for sustainability will be as follows: (i) if the DGP describing the discounted debt is not affected by a break, it should be mean-stationary or negative-trend stationary; (ii) in the presence of a break, it should be mean-stationary or negative trend-stationary after controlling for the break (see appendix A for details on coefficient restrictions).

There is considerable evidence supporting lag selection with data-dependent methods, which whiten the residuals, rather than fixing it exogenously. For each choice of the breakpoint, we follow the general-to-specific criterion advocated by Ng and Perron (1995). The procedure consists in starting with the most general lag specification and testing, by diminishing the order of lags one by one, the null that the coefficients of the last n lags are jointly significant. In each sequence a Wald test is performed, which is a χ^2 with n degrees of freedom under the null. The procedure is repeated sequentially until the null is rejected or until obtaining a model with zero lag. We set the upper bound on the lag length to equal 6 and choose $n=1$ (Perron, 1989).

2. Reaction function approach to testing for sustainability

Preliminary estimates based on Chow tests suggest instability in the feedback coefficient in the reaction function (equation 3). The instability is likely due to economic policy shocks that may alter the debt-surplus relation over the sample. To account for such major policy

adjustments, the feedback coefficient should be allowed to change at one or several date(s) corresponding to structural change(s). Since these dates are unknown, we will have to estimate them as together with the model parameters. Two common approaches to estimating breakpoints endogenously are (i) instability test-based approach (Andrews, 1992) and (ii) regression-based approach (Bai, 1997a, b and Bai and Perron, 1998). The first approach selects one breakpoint endogenously and is applicable to nonlinear models. However, a major drawback with this test procedure is that the asymptotic distributions of the F-type test statistics are constructed for non-trending regressors. It cannot be used to test parameter instability in equation (3), which contains debt series with deterministic trends.^v We, therefore, adopt the regression-based approach developed by Bai, Bai and Perron (BBP). This approach allows us to find multiple breaks, which can be estimated either with the global minimizers algorithm described in Perron (1997), or sequential methodology of BBP. We adopt the latter methodology because it is more robust to the wrong choice of the number of breaks and computationally less costly.

If the number of breaks is known, the sequential methodology estimates the first breakpoint \hat{t}_1 such that $\hat{t}_1 = \arg \min_{t_1} S_T(t_1)$, where $S_T(t_1)$ is the sum of squared residuals (SSR) resulting from estimating the model over the entire period. Then the sample is partitioned into two sub-periods around \hat{t}_1 and a one-break model is estimated over each subsample, which identifies two additional potential breakpoints. Among these points, the second breakpoint \hat{t}_2 is obtained by choosing the breakpoint that allows the largest reduction in the SSR computed over the whole sample. The same procedure is repeated sequentially until the predetermined number of breaks is reached.

If the number of breakpoints is unknown, the null hypothesis of m versus $m+1$ breaks ($m=0,1,2,\dots$) is tested sequentially and the breakpoint is estimated if the null is rejected. The

number of breakpoints is obtained at the first value of m for which the null is not rejected. An alternative to this test procedure is to combine the first sequential search method with the Bayesian Information Criteria $BIC(m)$ of Yao (1988) and $LWZ(m)$ of Liu, Wu and Zidek (1997). The optimal number of breakpoints m is then given when the minimum of these information criteria is reached. These criteria are appropriate with multiple break models because they introduce a penalty factor for additional breakpoints, which necessarily decrease the SSR.^{vi}

Bai (1997) shows that when all breakpoints are estimated, a reestimation (or refinement) of the first break dates over the last refined sample periods improves the estimation results. In the case of two breakpoints, if \hat{t}_2 is located to the right of \hat{t}_1 ($\hat{t}_2 > \hat{t}_1$), then \hat{t}_1 should be reestimated over the period $[1, \hat{t}_2]$.

In our framework we cannot perform the sequential estimation with a known number of breaks since inspection of data does not indicate an obvious number of breaks (Figure 1). The test procedure for unknown number of breaks is not appropriate either because the underlying parameter constancy test requires stationarity of the regressors, a condition that our trended debt series do not satisfy. We thus adopt the strategy of the sequential estimation with unknown number of breaks and information criteria (see Perron, 1997).

We also use the stationarity of the error term as an additional criterion for the number of breaks, which suggests that the model is well specified. To conduct the stationarity test, we perform a residual-based unit root test and use MacKinnon's (1991) all-sample estimated critical values. The stationarity of the residuals suggests that the number and location of the estimated breakpoints correspond to those of the « true » model.

Finally, to evaluate the degree of estimation accuracy of each estimated breakpoint we construct a confidence interval at the 5% level following the methodology in Bai, 1997 (see Appendix B for technical details).

3. Comparison of two approaches

Both the reaction function approach and the unit root tests assess the health of public finances, but each from a different perspective. The unit root test examines the sustainability of government finances by considering a single variable, the compound discounted real debt or debt/GDP ratio. The second test analyses whether the government tends to take offsetting adjustments when faced with rising debt by considering the relation between debt and the primary balances.

When the two tests are conducted simultaneously, there are four possible combinations for results: (i) unit root rejected and the feedback coefficient $\alpha > 0$ (or overall positive in the case of a structural break). Test results are consistent, fiscal policy is sustainable; (ii) unit root not rejected and $\alpha < 0$ (or overall negative in the case of a structural break). Test results are again consistent, fiscal policy is not sustainable; (iii) unit root not rejected and $\alpha > 0$. The primary surplus generated by the government has not been sufficient to revert the unsustainable path of fiscal policy and further efforts are required; (iv) unit root rejected and $\alpha < 0$. Although fiscal policy has been sustainable, the profligacy of government may put it at risk. The magnitude of the feedback before and after a break can give insight into the effects of reforms. For example, when there is a positive feedback, an increase (decrease) in the magnitude of the feedback coefficient over time suggests that the government increases (decreases) its efforts after the break(s) to reduce public debt.

IV. DATA

The data are annual and are from 1970 to 2002. Our choice of countries was dictated by the availability of data. The countries in the sample are the United States, Canada, Japan, Germany, France, the United Kingdom, Italy, Costa Rica, Honduras, Mexico, Panama, Uruguay, India, Indonesia, Korea, Malaysia, Thailand and Turkey.

In the first part of this analysis, we analyze the two most commonly used debt measures in the literature: real debt (br), and debt/GDP ratio (by). Each debt measure is compound-discounted with its corresponding discount rate detailed earlier.^{vii} We construct our debt measures based on the following series: nominal end-of-period government debt, the GDP deflator, nominal GDP (GNP when the GDP is not available), and interest rate. The G7 countries' data for nominal end-of-period gross debt, primary surplus, and the GDP deflator come from the *Organization for Economic Cultural and Development* (OECD). The interest rate series are government bond yields for maturities greater than 10 years and are from the *International Financial Statistics* (IFS) of the IMF. For LA and Asian countries, all data except the primary surplus data are from the IFS. For LA real debt is deflated by the country's CPI since the series for the GDP deflator are not available consistently across all countries.

Interest rate series for long-term government bonds either were unavailable or of widely-agreed notoriously poor quality, in particular in LA. To approximate an interest rate series that roughly reflects borrowing conditions in the economy and matches the sample length of government debt series, we used the commercial lending rate in most countries (Costa Rica, Honduras, Panama, Uruguay, India, Malaysia and Thailand), and the three-month time deposit rate (Turkey). The interest rate is computed as a geometric average of the two longest series of market rates in the other countries.^{viii} All interest rate series are from the IFS except when it is noted otherwise.

All variables in the government reaction function (equation 3) are proportional to GDP except \tilde{y}_t . We calculate \tilde{g}_t as the detrended series of nominal government spending with the Hodrick-Prescott filter with a smoothing parameter of 100 in all countries except Honduras, Panama and Turkey where the smoothing parameter is 25. All data for government spending are from the IFS except Turkey where the series are from *World Development Indicators* (WDI). The series \tilde{y}_t are also constructed as detrended series of the logarithm of the real GDP using the Hodrick Prescott filter, with a smoothing parameter of 100 for all countries except Malaysia, Mexico and Thailand where the parameter's value is 50. As expected, ADF test results (not reported) indicate that both variables are integrated of order 0.

The primary surplus for the G7 countries is provided on a fairly reliable basis from the OECD. For other countries, primary surplus figures are either not available or unreliable when they are. For those countries we generated S_t by adjusting total government deficit for total government interest payments, where the data on interest payments are from WDI. All primary balance series are inclusive of seignorage.

V. RESULTS

As Chart 1 shows, trends in the debt/GDP measure may exhibit several breaks, a condition which is likely to bias standard test results toward not rejecting the null of a unit root. We expect breaks to fall mainly in one of the four periods: the early 1970s (the first oil-price shock), late 1970s and early 1980s (the second oil-price shock, followed by a slowdown in the industrialized countries), mid-1980s (Latin American debt crisis), and the early 1990s (an episode that includes the German reunification, the European financial crisis, and the start of the protracted recession in Japan that culminated in the Asian crisis later in the decade).

1. Unit-root tests

We first test the simple unit root hypothesis in the ADF equation (Table 1). The first entry in each cell of the column t_{ADF} is the ADF statistic when trend is included in the test equation and the second one is the ADF statistic without the trend in the equation. We denote the compound discounted real debt and debt/GDP as br and by , respectively. If the null of a unit root is rejected when the trend is included and is significant, then the debt is trend-stationary. If the null is rejected when the trend is not present, then the debt is zero-trend stationary. Figures in the column t_{μ_t} are the t-statistics for the trend term from regressions pertaining to the first entry. The results are presented for 1 lag and they are robust to lag specification. Evidence based on standard ADF tests overwhelmingly fails to reject nonstationarity in the processes for both debt measures at the conventional significance levels in all but three countries.

< Table 1 here >

For both debt measures in France and Indonesia, and for by in Uruguay, the null of unit root without trend is significantly rejected (first entry, t_{ADF} column), suggesting that evidence supports trend stationarity. However, in the first two countries, the coefficient of the trend is positive. Discounted debt that grows at the trend rate will not exponentially decay towards zero and is considered to be unsustainable. Uruguay with a trend-stationary discounted debt/GDP and a negative trend is the only country whose policy is consistent with sustainability. We will return below to this concept of sustainability. In Korea results suggest zero-mean stationarity in discounted debt/GDP ratio at the 10 percent significance level.

The main conclusion to be drawn from this section is that the traditional unit-root tests suggest that countries in general do not satisfy their intertemporal budget constraint

during the time period considered, and to be solvent, governments have to change the course of their fiscal policy.

However, we know that structural breaks may bias ADF tests towards nonrejection of the null of a unit root when there is none. We now explore this possibility. Table 2 displays unit root test results with endogenous structural breaks for each discounted debt measure and for each model A, B and C of ZA (equations A1, A2, A3, appendix). The columns represent, in the order of appearance, the model, the estimated break date, the minimized Dickey-Fuller t-statistics, the trend coefficient, the slope-dummy coefficient, the shift-dummy coefficient and their respective t-statistics, and the optimal lag structure.

< Table 2 here >

In stark contrast to the simple ADF tests, the new results show that most countries in the sample have at least one stationary debt measure using one of the models. The ZA tests reject nonstationarity in at least one of the debt measures due to at least one type of break in 11 out of 18 countries at the 5 percent confidence level and in 13 countries at the 10 percent confidence level.^{ix} This compares with 2 and 3 cases, respectively, for the same confidence levels with the simple ADF tests.

More specifically, the evidence based on the ZA test rejects the null of a unit root in both debt measures in four G7 countries (France, Italy, Japan and the UK) at the 5 percent significance level with a breakpoint around 1979-81 and 1987-1990. The unit root hypothesis for at least one debt measure is also rejected in three LA countries (Costa Rica, Mexico and Uruguay) out of five at the 5 percent significance level, and in Honduras at the 10 percent level, albeit at different dates throughout the 1980s and the 1990s. In the Asian group, tests reject nonstationarity in four countries (Indonesia, Malaysia, Thailand and Turkey) out of six, at the 5 percent significance level, and in India at the 10 percent level,

with breakpoints mostly falling in the 1980s. The unit root null hypothesis cannot be rejected for US, Canada, Germany, Panama and Korea for either debt measure.

Closer inspection of Table 2 reveals that various models give different answers regarding the timing of the structural breaks. We can choose among the models as follows. For five countries the trend-stationary debt processes can be modelled with a single model (model A in Japan and Thailand, model C in Costa Rica and Mexico, and model C and A for br and by , respectively in India). In France, UK, and Indonesia model C includes an insignificant time-dummy with both debt measures, and thus fails to represent both a change in level and slope. In the latter two countries, model B is also inappropriate and the trend-stationarity is represented by model A. For the remaining six countries for which unit root is rejected, stationarity in debt processes is associated with different break dates. In these cases we focus on the model with the lowest Akaike and Schwarz Criteria (AIC and SC) or standard error of regression. This elimination process leaves us with the following specification: for br , the optimal model is B in France, and C in Italy, Honduras and Malaysia; for by , the optimal model is model A in France, B in Uruguay and Turkey, and C in Italy, and Malaysia.

What do these results imply for sustainability of fiscal policies? A close inspection of the data and the trend reveals that among the G7 countries for which ZA tests reject nonstationarity, in UK both debt measures, described by model A, have a downwards shift in level and a negative sloped trend, which is consistent with sustainability as defined above. In Italy, because the negative coefficient of the trend dummy is higher in absolute value than the positive coefficient of the trend, the sign of the overall trend becomes negative after 1989 in both debt measures (model C). In France, the policy is not sustainable. The negative trend dummy is not large enough to reverse the overall trend in br after the breakpoint (model B), and by is stationary around a positive trend (model A). Japan's policy appears to be

unsustainable because both debt measures exhibit a shift break in 1988 but with a significant and positive trend. Thus, we can argue that among the G7, only the UK and Italy follow sustainable policies.

The effect of breaks on debt stationarity in the LA group is even more striking. Debt in three out of five countries appears to be negative trend-stationary at the 5% significance level in at least one debt measure. In Costa Rica and Mexico, the positive trend of br in the first period is reversed in the second period by a negative trend-dummy that dominates the debt pattern (model C). In Mexico a similar pattern exists for by . In both countries the break affects the mean and the trend of the series. The positively sloped trend in Uruguay is not strong enough to reverse the negative trend-stationarity in by (model B). Although weaker (10% significance level), evidence also suggests a negative trend-stationarity in Honduras in both debt measures after the breaks for br and for by .

Evidence supports zero-trend stationarity or negative trend-stationarity in fewer Asian countries, namely, in Malaysia and Thailand, and weaker in India. Malaysia satisfies the sustainability condition with both debt measures. The null of unit root is rejected at the 5 (10) percent significance level in br (by) with model C where the positive trend is reversed after the break date. Since the trend is not significantly different from zero, the Thai by is constant-mean stationary before and after a shift-in-mean in 1989. In India test results suggest negative trend-stationarity in br (model C) and constant-mean trend-stationarity in by (model A) at the 10 percent level. Indonesia and Turkey have unsustainable policies because both exhibit positive trend-stationarity for at least one debt measure.

2. Governments' reaction functions and parameter instability

In this section we examine whether governments respond to accumulating debt by generating a primary surplus in the manner suggested by equation (3) and where the feedback

coefficient α captures this effect. To take into account possible breaks we perform a sequential least squares method as explained above.^x The estimates, α_1 , α_2 (and α_3) in Table 3 are the feedback response over the first, the second (and the third) sub-period or regime following the BBP procedure. For one break we verified that the residual stationarity criterion and the information criteria lead to the same estimated breakpoint. Because of our limited sample size, we restricted the maximum number of breaks to two and therefore could not verify that information criteria were minimized. However, estimating two breaks was sufficient to make the error terms stationary, and reduced the BIC and LWZ values. In rare cases where the error term is not stationary despite incorporating two breaks, we follow the specification of Wickens and Uctum (1993) and add the lagged dependent variable to the right-hand-side of the equation.

We should note that, for a given country the break dates estimated in the previous section for all debt measures are not expected to be necessarily the same as the break dates estimated in the present section for two reasons: (i) the variables involved in this section are in levels and as a ratio to GDP, not discounted; (ii) here the breaks characterize an economic relationship between two variables (debt and surplus), whereas in the previous section the break was characterizing the deterministic trend of a variable (debt).

< Table 3 here >

Several interesting results emerge. First, all countries except UK, has experienced at least one structural change. The confidence intervals are often the closest possible to the estimated breakpoints (one year on either side), reflecting the reliability of the estimates. Second, the significant coefficients keep their sign throughout the sub-periods in all of the G7, LA and in half of the Asian countries. Third, frequent occurrence of positive coefficients suggests more frequent mean reversion than in the ZA unit-root tests. In countries with a

positive and significant feedback coefficient, the coefficient estimates vary from 0.03 to 0.65 suggesting that given the level of GDP, for every \$100 increase in the public debt, governments generate a surplus of \$3 to \$65 the following year.

More specifically, except in Japan where the fiscal stance deteriorated since the collapse of the bubble, in all other G7 countries the feedback coefficient is positive. The magnitude of the positive coefficient declines in two G7 countries (France, Germany), reflecting loosening of fiscal discipline. It increases in two G7 countries (US, Italy) and is strong and positive in the UK. In the two other areas, the coefficient becomes positive from being negative or insignificant in several countries (Costa Rica, Mexico, Indonesia, Malaysia, Thailand), suggesting an overturn of the previous policies. The magnitude of the negative sign decreases in three countries (Honduras, Panama, India), showing some improvement in the degree of unsustainability. In four other countries (Canada, Costa Rica, Uruguay, Korea) it fluctuates with a positive sign over three subperiods, reflecting continuous fiscal effort with positive but erratic results on surplus.

Comparison with the unit root results

It is possible to reconcile the two approaches to sustainability. In four countries (UK, Italy, Uruguay and Thailand) both tests are consistent. Unit root is rejected for at least one debt measure (with or without breaks) and the feedback coefficient is positive in general, suggesting that these countries fall in the first category discussed in section III. Nonstationarity of debt measures in Canada, France, Germany, Japan, the US, together with positive and declining feedback coefficients indicate that the surplus generated by the governments are not sufficient and further efforts are required (third category). In Costa Rica, Mexico, and Malaysia unit root in debt is rejected while the surplus/debt relation

becomes positive after the break. This suggests that the change in the government stance towards debt accumulation put the public finances to a sustainable path (first category).

In India and Honduras with a weak rejection in unit root in one of the debt measures, the negative feedback coefficient, although declining in absolute value, indicates that governments need to generate a stream of surpluses for fiscal policy to become clearly sustainable. In the remaining two countries (Indonesia and Turkey), policy is unsustainable and the brief period of primary surplus has visibly not been sufficient to improve the long-term outlook.

VI. CONCLUSION

The traditional ADF tests of sustainability overwhelmingly fail to reject nonstationarity of public debt in the troubled areas of the world. However, these tests suffer a major weakness: they are sensitive to structural breaks, which bias results toward not rejecting unit roots. In this paper, we provide evidence that once possible breaks are taken into account, the strong nonsustainability results weaken and may disappear altogether. We examine the zero-trend sustainability as well as the negative-trend stationarity characteristic of the data, and we discuss how this is consistent with sustainability of fiscal policy. We find that controlling for structural breaks changes the outcome drastically for a majority of countries. Results show that in half of the eighteen countries, fiscal policy is sustainable, invalidating the nonstationarity results obtained from the standard econometric approaches.

We also explore an alternative indicator of sustainability, which consists in examining whether a government reacts to debt accumulation by generating primary surpluses. This alternative measure of mean reversion in debt reveals substantial active anti-debt policies in several countries, reinforcing the conclusion of the unit root test with endogenous

breakpoints. Frequent occurrence of a positive coefficient suggests that mean reversion is likely to be more common than any standard unit-root test would suggest.

Despite some evidence indicating a structural change in the debt/surplus relation, the sign of the government reaction parameter remains mostly stable through our sample.

Among the three groups, G7 countries fare best in feedback to debt accumulation. In the two other regions, in a majority of the cases the surplus/debt relation either becomes positive after the break or, if negative, declines in magnitude after the break.

Our results are consistent with the ongoing debate about debt and deficit policies adopted by Latin American countries after the 1980 debt crisis, the fiscal restraint in many Asian countries prevailing before the Asian crisis and the fiscal rules abided by the European countries to satisfy the conditions required for membership in the monetary union.

References

- Ahmed, S. and Rogers J.H. (1996). Government budget deficits and trade deficits: are present value constraints satisfied in long-term data? *Journal of Monetary Economics*, **36**, 351-74.
- Andrews, D. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, **61** (4), 821-56
- Bai J. (1997a). Estimating multiple breaks one at a time. *Econometric Theory*, **13**, 315-52.
- Bai J. (1997b). Estimation of a change point in multiple regression models. *Review of Economics and Statistics*. **79** (4), 551-63.
- Bai J. and Perron P. (1998). Estimating and testing linear models with multiple structural changes", *Econometrica*. **66** (1), 47-78.
- Banerjee, A., Lumsdaine R.L., and Stock J.H. (1992). Recursive and sequential tests of the unit root and trend break hypotheses: theory and international evidence. *Journal of Business and Economic Statistics*, **10**, 271-288.
- Ben-David, D., Lumsdaine R.L., Papell, D.H. (1998). Unit roots, post-war slowdowns and long-run growth: evidence from two structural breaks. NBER WP #6397.
- Bohn, H. (1999). The behavior of US public debt and deficits. *The Quarterly Journal of Economics*, 949-963.
- Buiter, W., Corsetti, G., Roubini, N. (1993). Excessive deficits: sense and nonsense in the Treaty of Maastricht. *Economic Policy*, spring.
- Brown, R.L., Durbin J. and Evans J.M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society*, B 37, 149-92
- Cuddington, J.T. (1997). Analyzing the sustainability of fiscal deficits in developing countries. Georgetown University.
- Dickey, D.A and Fuller, W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, **74**, pp. 427-431.
- Hakkio, C.S. and Rush, M. (1991). Is the 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, B.E. (1992). Testing for parameter instability in regressions with I(1) processes. *Journal of Business and Economic Statistics*, **10**, 321-336.
- Kremers, J.J.M. (1989). US federal indebtedness and the conduct of fiscal policy. *Journal of Monetary Economics*, **23**, 219-238.
- Liu, J., Wu, S. Zidek, V. (1997). On segmented multivariate regressions. *Statistica Sinica*, **7**, 497-527.
- Lumsdaine, R.L. and Papell, D.H. (1997). Multiple trend breaks and the unit-root hypothesis. *The Review of Economics and Statistics*, **79** (2). p 212-18.
- MacKinnon, J.G. (1991). Critical values for cointegration tests. In R.F. Engle and C.W.J. Granger (eds.), *Long-run Economic Relationships: Readings in Cointegration*. Oxford University Press.
- Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, **57** (6), 1361-1401.
- Perron P. (1997). L'estimation de modèles avec changements structurels multiples. *L'Actualité Economique*, **73** (1-2-3), 457-505.
- Perron, P. and Ng S. (1995). Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association*, **90**, Issue 429, 268-281.
- Ploberger W, Krämer W. and Kontrus, K. (1989). A new test for structural stability in the linear regression model. *Journal of Econometrics*, **40**, 307-18.

- Sen, P.K. (1980). Asymptotic theory of some tests for a possible change in the regression slope occurring at an unknown time point. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **52**, 203-18.
- Trehan, B. and Walsh C.E. (1988). Common trends, the government budget constraint and revenue smoothing. *Journal of Economics Dynamics and Control*, **17**, 423-441.
- Uctum, M. and Wickens M.R. (2000). Debt and deficit ceilings and sustainability of fiscal policy. *Oxford Bulletin of Economics and Statistics*, **62** (2).
- Wickens, M.R. and Uctum M. (1993). The Sustainability of Current Account Deficits: a Test of the US Intertemporal Budget Constraint. *Journal of Economic Dynamics and Control*, **17** (3), 423-441.
- Wilcox, D.W. (1989). The sustainability of government deficits: implications of the present-value constraint. *Journal of Money Credit and Banking*, **21**, 291-306
- Yao Y.C. (1988). Estimating the number of change-points via Schwarz' criterion. *Statistics and Probability Letters*, **6**, 181-9.
- Zivot, E. and Andrews, D. (1992). Further evidence on the great crash, the oil price shock, and the unit-root hypothesis. *Journal of Business and Economic Statistics*, **10** (3), 25-44.

Table 1: ADF unit root tests‡

Countries	br		by	
	t_{ADF}	t_{μ_1}	t_{ADF}	t_{μ_1}
G7				
US	-0.27 -1.23	-1.3	0.37 -1.28	1.3
Canada	0.79 -1.33	-2.0	0.91 -1.45	-2.2
France	-3.82 -1.74	3.2	-3.49 -1.66	2.9
Germany	-1.96 -1.31	1.6	-1.98 -1.27	1.7
Italy	0.61 -2.36	-2.4	0.61 -2.39	2.2
Japan	-2.22 -0.53	2.2	-2.22 -0.53	2.2
UK	-2.64 -1.01	-2.5	-2.61 -1.04	-2.4
Latin America				
Costa Rica	-1.91 -1.94	0.9	-2.29 -2.10	1.4
Honduras	-0.35 -1.62	-0.6	-0.33 -1.63	0.6
Mexico	-2.15 -1.88	1.2	-1.92 -1.78	1.0
Panama	-2.01 -0.49	-2.0	-2.13 -0.49	-2.1
Uruguay	-3.05 -0.92	-2.9	-3.61 -0.79	-3.5
Asia				
India	-2.17 -2.33	0.0	-2.53 -1.71	2.1
Indonesia	-3.81 -1.45	3.4	-4.08 -1.44	3.7
Korea	-2.39 -2.65	-0.7	-2.23 -2.60	-0.0
Malaysia	-1.47 -1.92	0.1	-1.43 -1.97	-0.1
Thailand	-2.65 -2.24	-1.2	-2.54 -2.28	-0.9
Turkey	-1.75 -0.80	1.1	-0.79 1.60	1.9

‡The ADF test regression is $\Delta b_t = \mu_0 + \mu_1 Trend + \phi b_{t-1} + \sum_{i=1}^n \beta_i \Delta b_{t-i} + \varepsilon_t$ for $n=1$ and where b_t is measured as br and by, which stand for real debt, and debt/GDP, all compound discounted. t_{ADF} is the ADF t-statistics for testing the null hypothesis $\phi = 0$, and t_{μ_1} is the t-statistics for $\hat{\mu}_1$. The first entry in each cell in the t_{ADF} column is the ADF-t statistics for $\mu_1 \neq 0$ and the second entry is for $\mu_1 = 0$; MacKinnon's 5% and 10% critical values are -3.56 and -3.21 for the first entry and -2.93 and -2.60 for the second entry.

Table 2. Unit Root Tests With Endogenous Structural Break‡

G7	Model	br						by					
		date	t _{ZA}	T	DT	DU	lag	date	t _{ZA}	T	DT	DU	Lag
US	A	1996	-1.51	32.60 (1.0)		-11.74 (-2.4)	1	1996	-1.39	0.00 (0.9)		-0.03 (-2.3)	1
	B	1994	-3.86	141.00 (3.5)	-590.54 (-4.6)		1	1994	-3.77	0.00 (3.4)	-0.02 (-4.4)		1
	C	1994	-3.75	148.73 (3.5)	-657.51 (-3.9)	263.9 (0.6)	1	1994	-3.63	0.00 (3.4)	-0.02 (-3.8)	0.01 (0.5)	1
Canada	A	1995	-0.60	-0.06 (-0.0)		-172.9 (-2.6)	0	1995	-0.53	-0.00 (-0.2)		-0.04 (-2.5)	0
	B	1993	-2.18	12.2 (1.8)	-61.6 (-3.3)		0	1993	-1.94	0.00 (1.5)	-0.01 (-3.0)		0
	C	1992	-1.94	12.1 (1.7)	-62.1 (-3.2)	66.7 (1.1)	0	1992	-1.91	0.00 (1.4)	-0.01 (-2.9)	0.01 (0.9)	0
France	A	1988	-4.76*	239.44 (3.9)		-1556.8 (-2.3)	1	1981	-5.89***	0.00 (4.3)		0.05 (4.7)	5
	B	1979	-4.63**	383.9 (3.5)	-288.72 (-2.5)		1	1979	-4.57**	0.01 (3.7)	-0.01 (-2.8)		1
	C	1987	-4.92*	275.97 (4.1)	-88.63 (-1.3)	-1444.7 (-2.4)	1	1981	-5.68***	0.00 (0.6)	0.00 (0.4)	0.051 (4.0)	5
Germany	A	1980	-4.22	126.35 (4.0)		495.6 (2.3)	5	1980	-4.25	0.01 (4.0)		0.02 (2.1)	5
	B	1977	-2.15	110.66 (2.0)	-80.34 (-1.6)		0	1977	-2.07	0.01 (1.9)	-0.00 (-1.5)		0
	C	1980	-4.30	43.83 (0.5)	87.73 (0.9)	684.6 (2.3)	5	1980	-4.33	0.00 (0.5)	0.00 (1.0)	0.03 (2.2)	5
Italy	A	1981	-1.70	-37 10 ³ (-1.4)		723 10 ³ (2.5)	2	1981	-1.77	-0.00 (-1.2)		0.08 (2.4)	2
	B	1990	-6.63***	403 10 ³ (6.1)	-670 10 ³ (-6.7)		2	1990	-6.87***	0.05 (6.3)	-0.08 (-6.9)		2
	C	1989	-6.96***	450 10 ³ (6.6)	-756 10 ³ (-7.1)	110 10 ³ (4.2)	2	1989	-7.15***	0.05 (6.8)	-0.09 (-7.2)	0.12 (4.3)	2
Japan	A	1988	-4.95**	33 10 ³ (4.5)		-150 10 ³ (-2.7)	3	1988	-4.94**	0.02 (4.5)		-0.08 (-2.7)	3
	B	1977	-3.41	33 10 ³ (2.0)	-20 10 ³ (-1.4)		2	1977	-3.85	0.02 (1.9)	-0.01 (-1.3)		3
	C	1999	-3.22	12 10 ³ (3.1)	87 10 ³ (1.0)	-105 10 ³ (-0.7)	2	1988	-3.24	0.02 (2.5)	0.00 (0.1)	-0.08 (-2.6)	3
UK	A	1987	-5.66***	-18.92 (-2.9)		-462.4 (-4.0)	4	1987	-5.51***	-0.00 (-2.5)		-0.10 (-3.8)	4
	B	1977	-3.13	24.71 (1.0)	-44.76 (-1.6)		1	1977	-3.15	0.01 (1.1)	-0.01 (-1.6)		1
	C	1987	-5.21**	-15.06 (-1.8)	-9.91 (-0.7)	-485 (-4.0)	4	1987	-5.15**	-0.00 (-1.2)	-0.00 (-1.0)	-0.10 (-3.9)	4

‡ Variables br and by refer to real debt, and debt/GDP, both compound discounted with the appropriate discount rates. Figures in the t_{ZA} -column are the values of the t-statistics from models with shift-level dummy (model A), slope dummy (model B) and the combination dummies (model C) following Zivot and Andrews (1992), where serial correlation is corrected using Perron and Ng (1995) optimal lag selection procedure. Significance levels of 1%, 5% and 10% are represented by ***, ** and * and the corresponding asymptotic critical values are -5.34, -4.8 and -4.58 for model A, -4.93, -4.42 and -4.11 for model B and -5.57, -5.08 and -4.82 for model C. Figures in the t_{DU}-columns are the t-statistics of the shift-dummy. T and DT columns are, respectively, the coefficient and the associated value of the t-statistics of the trend variable and the slope dummy.

Table 2. Continued‡

LA	Model	br						by					
		date	t _{ZA}	T	DT	DU	lag	Date	t _{ZA}	T	DT	DU	lag
Costa Rica	A	1981	-3.47	6.22 (0.2)		2235.2 (2.7)	0	1977	-2.91	0.01 (0.8)		0.70 (1.9)	0
	B	1985	-3.18	134.41 (2.9)	-168.71 (-2.5)		0	1982	-3.31	0.14 (2.7)	-0.13 (-2.4)		0
	C	1981	-6.40 ^{***}	155.33 (4.0)	-237.60 (-4.6)	3781.8 (5.3)	1	1981	-3.58	0.10 (1.8)	-0.11 (-1.9)	0.60 (1.6)	0
Honduras	A	1996	-3.32	2.49 (2.9)		-53.02 (-3.9)	0	1996	-3.31	0.07 (2.9)		-0.35 (-4.0)	0
	B	1995	-4.35 [*]	4.03 (3.9)	-20.29 (-4.7)		0	1995	-4.40 [*]	0.03 (4.0)	-0.13 (-4.8)		0
	C	1989	-5.05 [*]	2.28 (3.3)	-9.84 (-6.3)	53.3 (5.2)	0	1989	-4.75	0.01 (3.1)	-0.06 (-5.9)	0.34 (4.9)	0
Mexico	A	1995	-4.00	1651.00 (3.5)		-20 10 ³ (-3.3)	3	1995	-3.62	0.35 (3.1)		-4.06 (-3.0)	3
	B	1988	-3.60	1473.60 (3.1)	-2081.46 (-2.8)		1	1990	-3.45	0.32 (2.9)	-0.53 (-2.7)		1
	C	1986	-7.35 ^{***}	1977.29 (5.0)	-3857.17 (-6.3)	37 10 ³ (5.8)	3	1986	-10.74 ^{***}	0.42 (6.8)	-0.86 (-9.1)	3.37 (9.0)	3
Panama	A	1982	-3.48	-0.05 (-1.9)		-1.12 (-3.0)	0	1983	-4.09	-0.00 (-2.1)		-0.05 (-3.4)	0
	B	1989	-3.24	-0.15 (-3.7)	0.14 (3.0)		0	1990	-3.17	-0.00 (-3.5)	0.00 (2.6)		0
	C	1992	-3.41	-0.25 (-3.4)	0.14 (1.7)	1.11 (3.2)	3	1983	-3.18	-0.00 (-1.7)	0.00 (0.8)	-0.04 (-2.3)	0
Uruguay	A	1984	-0.77	2.41 (1.0)		-25 (-1.5)	5	1996	-4.60 [*]	-0.02 (-4.1)		0.09 (2.8)	3
	B	1991	-3.16	-18.28 (-2.8)	16.80 (3.6)		5	1993	-6.35 ^{***}	-0.03 (-5.8)	0.03 (4.5)		3
	C	1990	-3.05	-18.89 (-2.7)	17.44 (3.4)	-21.4 (-1.8)	5	1992	-6.21 ^{***}	-0.03 (-5.7)	0.03 (4.0)	-0.02 (-0.6)	3

‡See footnote above.

Table 2. Continued‡

Asia	Model	br						by					
		date	t _{ZA}	T	DT	DU	lag	date	t _{ZA}	T	DT	DU	lag
India	A	1985	-4.57	-75.06 (-3.1)		877.3 (2.6)	5	1987	-4.59*	-0.00 (-0.6)		0.03 (3.0)	5
	B	1981	-3.15	-387.10 (-0.8)	373.59 (0.8)		5	1989	-3.18	0.00 (2.4)	-0.00 (-1.5)		0
	C	1984	-4.89*	-342.63 (-2.8)	302.02 (2.8)	1028.5 (3.1)	5	1987	-4.47	0.00 (0.4)	-0.00 (-0.6)	0.03 (2.9)	5
Indonesia	A	1990	-4.76*	155 10 ³ (3.9)		-745.6 10 ³ (-1.9)	0	1990	-5.04**	0.36 (4.1)		-1.47 (-1.7)	0
	B	1988	-4.35*	133 10 ³ (3.4)	-68 10 ³ (-1.2)		0	1988	-4.71**	0.32 (3.8)	-0.15 (-1.2)		0
	C	1990	-4.65	154 10 ³ (3.4)	12.5 10 ³ (0.2)	-792 10 ³ (-1.6)	0	1985	-4.99**	0.25 (2.6)	-0.11 (-0.9)	1.53 (1.6)	0
Korea	A	1986	-3.09	479.34 (1.6)		-10 10 ³ (-2.2)	0	1986	-2.88	0.00 (1.8)		-0.06 (-2.1)	0
	B	1981	-2.92	1528.85 (1.9)	-2268.92 (-2.1)		0	1981	-3.08	0.01 (2.3)	-0.02 (-2.4)		0
	C	1979	-2.95	1095.82 (1.2)	-1814.11 (-1.7)	6085.5 (1.5)	0	1979	-3.05	0.01 (1.8)	-0.01 (-2.1)	0.04 (1.4)	0
Malaysia	A	1980	-3.30	-1.68 (-2.4)		40.18 (3.0)	1	1980	-3.14	-0.01 (-2.1)		0.16 (2.9)	1
	B	1986	-4.41**	6.80 (3.8)	-11.96 (-4.0)		1	1986	-4.28*	0.03 (3.8)	-0.05 (-3.9)		1
	C	1980	-5.09**	2.76 (1.9)	-5.85 (-3.4)	43.9 (3.9)	1	1980	-4.84*	0.01 (2.1)	-0.02 (-3.3)	0.18 (3.8)	1
Thailand	A	1983	-4.58*	-38.53 (-3.3)		349.18 (2.5)	5	1989	-4.94**	0.00 (0.8)		-0.06 (-2.6)	5
	B	1985	-3.33	31.11 (1.2)	-81.88 (-1.7)		5	1986	-3.46	0.01 (1.4)	-0.01 (-1.8)		5
	C	1983	-4.21	-22.57 (-0.6)	-20.61 (-0.5)	343.45 (2.6)	5	1983	-4.31	-0.00 (-0.6)	-0.00 (-0.4)	0.06 (2.5)	5
Turkey	A	1993	-1.47	56 10 ⁶ (1.7)		940 10 ⁶ (1.8)	0	1993	-2.99	3.14 (2.9)		72.66 (3.2)	0
	B	1991	-3.06	91 10 ⁶ (2.9)	299 10 ⁶ (3.5)		0	1990	-4.99***	7.34 (5.1)	33.19 (4.8)		3
	C	1991	-2.97	92 10 ⁶ (2.8)	304 10 ⁶ (3.3)	-77 10 ⁶ (-0.2)	0	1989	-4.88*	7.49 (4.6)	33.36 (4.7)	-36.56 (-2.0)	3

‡See footnote above

Table 3. Government Reaction Function and Feedback Coefficients ‡

G7 countries	t_1	t_2	μ	α_1	α_2	α_3	γ	β	BIC	LWZ	\bar{R}^2	ADF
US 1972-2001	1996 [95,97]		-0.02 (-2.38)	0.03 (1.58)	0.08 (6.82)		0.18 (1.60)	-1.73 (-2.73)	-9.01	-8.69	0.82	-3.97*
Canada 1971-2001	1974 [73,75]	1995 [94,96]	-0.03 (-3.25)	0.10 (5.03)	0.03 (2.24)	0.09 (9.01)	0.54 (3.96)	-0.09 (-0.25)	-8.55	-8.11	0.91	-4.21*
France 1972-2001	1988 [86,90]		-0.05 (-5.81)	0.13 (5.06)	0.08 (6.00)		0.19 (2.42)	-0.91 (-3.25)	-9.62	-9.29	0.70	-4.57*
Germany 1971-2001	1974 [73,75]	1989 [88,90]	-0.07 (-13.62)	0.32 (10.04)	0.20 (12.15)	0.14 (14.11)	0.15 (2.95)	-1.11 (-5.98)	-9.65	-9.21	0.89	-4.55**
Italy 1971-2001	1991 [89,93]	1996 [94,98]	-0.10 (-10.22)	0.08 (6.39)	0.11 (12.07)	0.13 (14.88)	0.21 (1.33)	0.37 (1.36)	-8.36	-7.92	0.94	-6.13***
	t_1		μ	θ	α_1	α_2	γ	β	BIC	LWZ	\bar{R}^2	ADF
Japan 1971-2001	1991 [89,93]		-0.01 (-4.07)	0.74 (12.61)	0.03 (5.40)	-0.00 (-0.04)	-0.04 (-0.76)	-2.59 (-5.85)	-9.39	-9.06	0.93	-6.12***
UK 1979-2001			-0.09 (-5.06)	0.86 (16.00)	0.18 (4.98)		0.08 (0.49)	-0.43 (-0.58)	-6.46		0.91	-4.75***

‡ Regression equation is $S_t = \mu + \sum_i \alpha_i B_{t-1} 1_{t \in I_i} + \beta \tilde{g}_t + \gamma \tilde{y}_t + v_t$ (see Appendix B), plus the lagged dependent variable in the case of Japan

and UK with the parameter θ . Columns t_1 and t_2 are the estimated breakpoints and the brackets are 95% confidence intervals computed following the procedure in Bai (1997). Figures in parentheses in the estimates columns are the t-statistics. Columns BIC and LWZ are the minimum values of Yao's (1988) and Liu, Wu and Zidek's (1997) information criteria depending on the number of breaks. The last column is the residual-based ADF test statistic. The 1%, 5% and 10% significance levels from MacKinnon (1991) are represented by ***, ** and *. See Appendix B for details. The end-of-period debt is dated at t-2 for US, France, Costa-Rica, Indonesia, Malaysia, Thailand and at t-1 for the others.

Table 3. Continued ‡

L.A. countries	t_1	t_2	μ	α_1	α_2	α_3	γ	β	BIC	LWZ	\bar{R}^2	ADF
Costa-Rica 1972-1999	1976 [74,78]	1992 [84,00]	-0.01 (-0.38)	0.00 (0.02)	0.29 (2.26)	0.11 (0.97)	-0.34 (-5.20)	-1.77 (-4.05)	-6.21	-5.76	0.64	-6.10***
Honduras 1971-2001	1987 [86,88]		0.02 (1.56)	-0.17 (-5.43)	-0.06 (-4.53)		0.08 (1.22)	0.03 (0.08)	-7.86	-7.54	0.65	-4.42**
Mexico 1971-1998	1987 [86,88]		0.03 (1.73)	-0.20 (-3.79)	0.06 (1.39)		0.15 (0.46)	1.47 (0.74)	-5.44	-5.12	0.43	-3.97*
Panama 1972-1998	1976 [74,78]	1989 [83,95]	0.16 (2.28)	-0.56 (-2.55)	-0.21 (-1.70)	-0.12 (-1.06)	-0.24 (-1.66)	-1.07 (-2.07)	-5.46	-5.13	0.40	-5.71***
Uruguay 1973-1999	1988 [86,90]	1990 [89,91]	-0.02 (-1.26)	0.30 (5.94)	0.54 (8.79)	0.22 (2.85)	-0.00 (-0.05)	-0.48 (-0.83)	-6.20	-5.74	0.59	-5.73***

‡See footnote above

Table 3. Continued ‡

Asian countries	t_1	t_2	μ	α_1	α_2	α_3	γ	β	BIC	LWZ	\bar{R}^2	ADF
India 1975-2000	1986 [85,87]	1990 [89,91]	0.06 (6.22)	-0.27 (-10.95)	-0.24 (-12.17)	-0.18 (-9.71)	0.11 (2.80)	-0.59 (-3.29)	-9.88	-9.42	0.90	-6.24***
Indonesia 1973-1999	1983 [81,85]		0.03 (4.15)	-0.11 (-4.65)	0.06 (3.29)		0.09 (1.39)	-0.36 (-1.02)	-8.32	-7.99	0.84	-5.04***
Korea 1971-1997	1985 [84,86]	1989 [87,91]	-0.03 (-1.43)	0.35 (3.01)	0.64 (4.63)	0.38 (2.20)	-0.06 (-1.03)	-0.91 (-3.45)	-8.14	-7.68	0.72	-7.32***
Malaysia 1972-1997	1984 [83,85]		-0.00 (-0.08)	-0.11 (-2.28)	0.08 (2.55)		0.16 (1.12)	-2.51 (-3.79)	-6.73	-6.40	0.85	-5.05***
Thailand 1972-2000	1982 [80,84]	1989 [88,90]	-0.03 (-2.65)	0.01 (0.09)	0.21 (5.21)	0.30 (4.39)	0.40 (4.29)	-1.50 (-4.25)	-7.80	-7.35	0.81	-4.49*
Turkey 1971-2000	1995 [94,96]		-0.03 (-3.45)	0.09 (3.08)	-0.06 (-3.06)		0.02 (1.02)	-1.25 (-3.73)	-7.68	-7.36	0.64	-4.60**

‡See footnote above

Appendix A: Zivot and Andrews (1992) unit root test with an unknown breakpoint in the deterministic trend function

The deterministic trend of the debt series is characterized by a change in mean (A1), a change in slope (A2) or by both changes (A3) :

$$(A1) \quad \text{Model A: } b_t = \mu_{A,\pi} + \theta_{A,\pi} DU_t(\pi) + \beta_{A,\pi} t + \phi_{A,\pi} b_{t-1} + \sum_{j=1}^k \delta_{A,\pi,j} \Delta b_{t-j} + e_t$$

$$(A2) \quad \text{Model B: } b_t = \mu_{B,\pi} + \gamma_{B,\pi} DT_t(\pi) + \beta_{B,\pi} t + \phi_{B,\pi} b_{t-1} + \sum_{j=1}^k \delta_{B,\pi,j} \Delta b_{t-j} + e_t$$

(A3) Model C:

$$b_t = \mu_{C,\pi} + \theta_{C,\pi} DU_t(\pi) + \gamma_{C,\pi} DT_t(\pi) + \beta_{C,\pi} t + \phi_{C,\pi} b_{t-1} + \sum_{j=1}^k \delta_{C,\pi,j} \Delta b_{t-j} + e_t$$

where $\pi = d/T$ is the unknown sample ratio to be estimated ($\pi \in \Pi = [2/T, (T-1)/T]$), T is the number of observations and d is the unknown breakpoint. The level dummy $DU_t(\pi)$ equals 1 if $t > d$ and 0 otherwise. The slope dummy $DT_t(\pi)$ equals $t-d$ if $t > d$ and 0 otherwise. The null hypothesis of a unit root is that $\phi_{i,\pi} = 1, i = A, B, C$. For each value of π one computes the ADF t -statistics, $t(\hat{\phi}_{i,\pi})$, and the $\inf_{\pi \in \Pi} [t(\hat{\phi}_{i,\pi})]$ statistics are used for testing the null against the trend-stationary alternative. One rejects the null if these statistics are less than their asymptotic critical values provided by the authors.

In a finite sample, we consider three cases as compatible with concept of sustainability:

$$\beta_{i,\pi} < 0 \text{ and } \gamma_{i,\pi} = 0, \quad i = A, B, C \quad (\text{decreasing linear trend})$$

$$\beta_{i,\pi} < 0 \text{ and } \gamma_{i,\pi} + \beta_{i,\pi} < 0 \quad (\gamma_i \neq 0), \quad i = A, B, C \quad (\text{decreasing breaking trend})$$

$$\beta_{i,\pi} > 0 \text{ and } \gamma_{i,\pi} + \beta_{i,\pi} < 0, \quad i = A, B, C \quad (\text{increasing followed by decreasing trend})$$

Appendix B: Government reaction function with unknown multiple breaks (Bai (1997), Bai and Perron (1998))

Equation (3) with m ($m=0,1,2,\dots$) unknown breakpoints can be written as:

$$(B1) \quad S_t = \mu + \sum_{i=1}^{m+1} \alpha_i B_{t-1} 1_{t \in I_i} + \beta \tilde{g}_t + \gamma \tilde{y}_t + v_t$$

where I_i is the sub-period between break dates t_{i-1} and t_i , and $1_{t \in I_i}$ an indicator function such that $1_{t \in I_i} = 1$ for $t_{i-1} < t \leq t_i$ and 0 elsewhere (we set $t_0 = 1$ and $t_{m+1} = T$). The parameters α_i ($i=1,\dots, m+1$) are estimated over the sub-period I_i , while β and γ are defined over the full sample. When $m=0$, (B1) reduces to model (3).

The method of sequential least squares estimation consists in first, estimating (B1) for $m=1$, over the entire period and identifying the first breakpoint $\hat{t}_1 = \arg \min_{t_1} S_T(t_1)$, where $S_T(t_1)$ is the sum of squared residuals from the one-break model with the candidate break date t_1 . The sample is then split into two and a one-break model is estimated over each sub-sample $[1, \hat{t}_1]$ and $[\hat{t}_1, T]$, yielding two potential break dates, $\hat{\tau}_1$ and $\hat{\tau}_2$, respectively. The second estimate $\hat{t}_2 = \hat{\tau}_1$ if $S_T(\hat{t}_1, \hat{\tau}_1) < S_T(\hat{t}_1, \hat{\tau}_2)$ and $\hat{t}_2 = \hat{\tau}_2$ otherwise, where for example $S_T(\hat{t}_1, \hat{\tau}_1)$ is the SSR from model (B1) for $m=2$ breakpoints \hat{t}_1 and $\hat{\tau}_1$. Bai and Perron (1998) show that if t_1^* and t_2^* are the true breakpoints, then (\hat{t}_1, \hat{t}_2) is consistent for (t_1^*, t_2^*) . The sample is then partitioned into three and a one-break model is estimated over each sub-sample $[1, \hat{t}_1]$, $[\hat{t}_1, \hat{t}_2]$ and $[\hat{t}_2, T]$, and so forth.

We estimate the additional break dates until the BIC and LWZ information criteria are minimized and the residuals \hat{v}_t from (B1) are $I(0)$. For m breakpoints, these criteria are

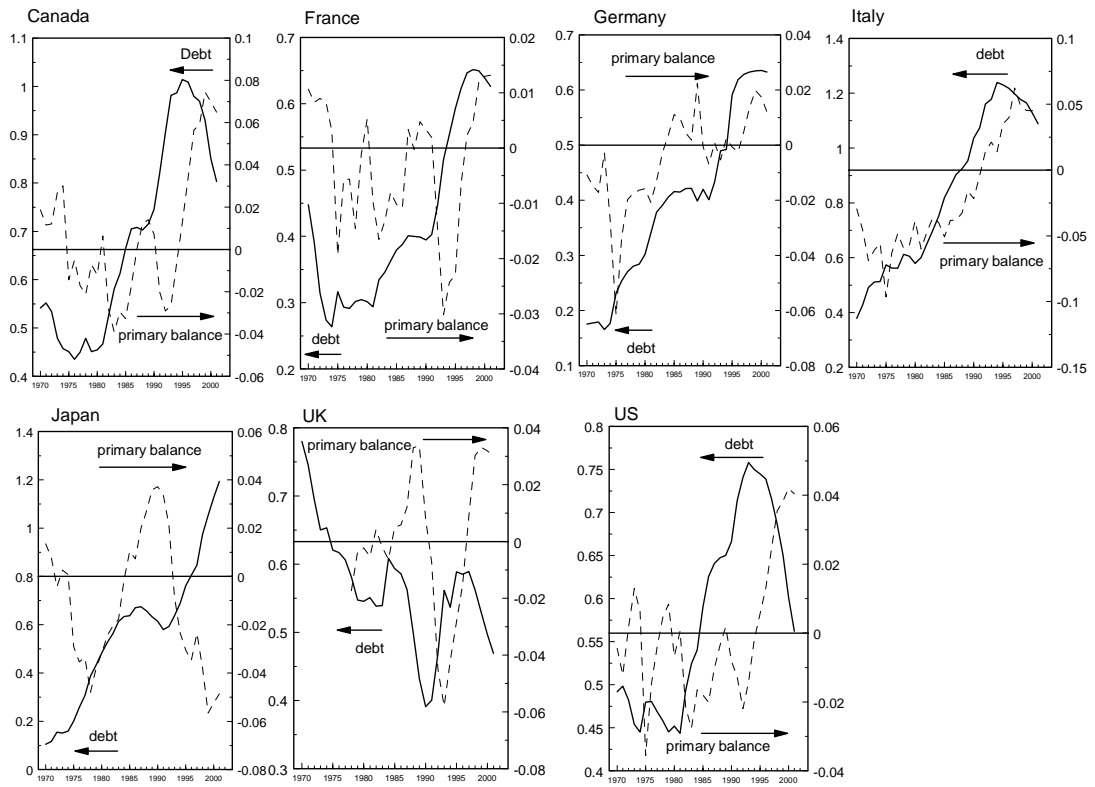
$$BIC(m) = \log S_T(\hat{t}_1, \dots, \hat{t}_m) - \log T + (p^*/T) \log T \text{ and}$$

$$LWZ(m) = \log S_T(\hat{t}_1, \dots, \hat{t}_m) - \log(T - p^*) + c_0(p^*/T)(\log T)^{c_1},$$

where $c_0 = 0.299$, $c_1 = 2.1$, and p^* , the penalty factor compensating the decrease in the SSR with each additional new break is $(m+1)q + m + p$, with q , p , defined as the number of changing, and unchanging coefficients, respectively (see Perron 1998).

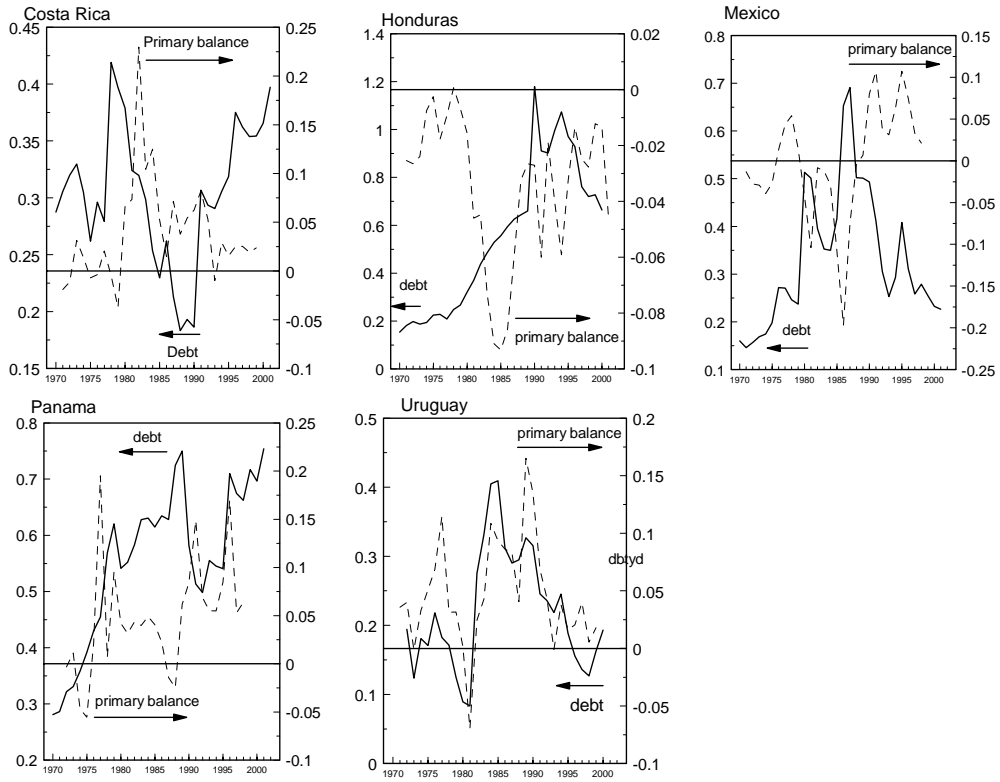
Following Bai (1997b), a 95% confidence interval for the estimated break date \hat{t}_i in the case of trending regressors can be computed as $[\hat{t}_i - [c/\hat{L}_i] - 1, \hat{t}_i + [c/\hat{L}_i] + 1]$, where $[c/\hat{L}_i]$ is the integer part of c/\hat{L}_i , c the 97.5th quintile (computed from the symmetric case cdf formula as $c=11$), and $\hat{L}_i = (\hat{\alpha}_{i-1} - \hat{\alpha}_i)^2 B_{\hat{t}_i}^2 / \hat{\sigma}_v^2$, a scale factor with $\hat{\alpha}_{i-1}$, $\hat{\alpha}_i$ and $\hat{\sigma}_v^2$ defined as the feedback estimates before and after the breakpoint \hat{t}_i , and the estimated variance of \hat{v}_t in (B1), respectively.

Figure 1a. Debt/GDP and primary balance/GDP: G7



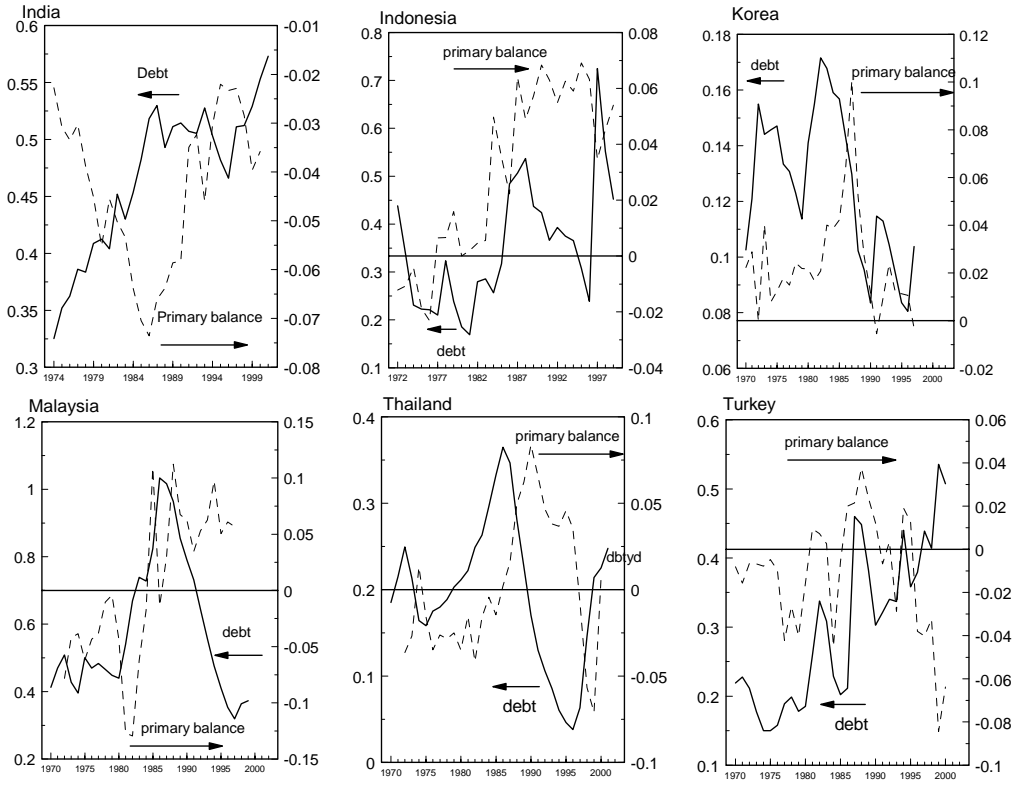
*Debt/GDP is the bold line (left axis), primary balance/GDP is the dashed line (right axis).

Figure 1b. Debt/GDP and Primary Surplus/GDP: LA*



*TDebt/GDP is the bold line (left axis), primary balance/GDP is the dashed line (right axis).

Figure 1c. Debt/GDP and Primary Surplus/GDP: Asia*



*Debt/GDP is the bold line (left axis), primary balance/GDP is the dashed line (right axis).

ENDNOTES

ⁱ Most of the sustainability analysis is conducted for industrialized nations (Buiter et al., 1993, Uctum and Wickens, 2000) and in particular the United States (Hamilton and Flavin, 1986, Wilcox, 1989, Trehan and Walsh, 1988, Kremers, 1989, Hakkio and Rush, 1991, Wickens and Uctum, 1993) and the United Kingdom (Ahmed and Rogers, 1996). Results vary depending on (i) the specification of the budget constraint, (ii) the length of the data set (the longer the data the more likely to reject nonstationarity), and (iii) whether data after the 1980s are included (financial crises or slowdown in activity have an adverse effect on government finances).

ⁱⁱ The mean reversion approach for testing sustainability has been used by Wickens and Uctum (1993) in the context of the US current account balance and by Bohn (1999) for the US public debt. Wickens and Uctum show that a sufficient condition for satisfying a national intertemporal budget constraint is that there is a negative feedback from debt to the trade account.

ⁱⁱⁱ See Hakkio and Rush (1991) and Cuddington (1997) for a discussion of various debt measures. We also conducted the analysis with per capita real debt. Since the results do not alter the conclusion reached in this study we are not reporting them, but they are available from the authors upon request.

^{iv} We also conducted the tests with the methodology by Banerjee et al. (1992). However, these tests require truncating a large fraction of data at both ends of the sample (1/4 to 1/3 of the data) and therefore cannot capture breaks occurring at the beginning or at the end of the sample. This compares with only two extreme observations excluded by ZA, which is more appropriate for small samples like ours. By applying Banerjee et al., we often observed that breakpoints were located before or after the trimming parameters. For this reason, we focus our discussion on ZA test results.

^v Other methodologies also test parameter instability but these tests do not estimate a breakpoint and are not appropriate to examine the structural change in the debt-surplus relation (examples of such methodologies are Cusum or Cusum Square tests by Brown, Durbin and Evans (1975), the fluctuation test by Sen (1980) or Ploberger et al. (1989).

^{vi} See Perron (1997) for a comparison of the two information criteria.

^{vii} See Uctum and Wickens (2000) for the derivation and a discussion for using a compound discount rate.

^{viii} These are: the average cost of funds and the three-month deposit rate from *Global Finance* (Mexico), commercial discount rate and the lending rate (Costa Rica), call money rate and term deposits in Indonesia, and the yield on housing bonds and the discount rate (Korea).

^{ix} It is possible that most countries experienced multiple breaks in the processes governing their fiscal variables. Studies involving multiple breaks show that allowing for more than one endogenous break points, in general, reinforces the evidence against the unit-root hypothesis (Lumsdaine and Papell, 1997, Ben-David, Lumsdaine and Papell, 1998).

^x Since the primary deficit is inversely related to government spending, S_t and \tilde{g}_t share the same information set, which introduces the problem of simultaneity bias. A potential bias might also be caused by the relation between S_t and \tilde{y}_t . Then OLS would be inappropriate and TSLS or IV method should be used. To see if there is a significant bias, we checked the correlation between \tilde{g}_t and \hat{v}_t , \tilde{y}_t and \hat{v}_t from equation (3) for each country. We found that it is systematically insignificant, suggesting that OLS is appropriate.