Panel Cointegration Tests of the Sustainability Hypothesis in Rich OECD Countries

Joakim Westerlund, Silika Prohl

Panel Cointegration Tests of the Sustainability Hypothesis in Rich OECD Countries

Journal:	*Applied Economics*
Manuscript ID:	APE-06-0624.R1
Journal Selection:	Applied Economics
Date Submitted by the Author:	27-Jan-2007

Complete List of Authors:	Westerlund, Joakim; Lund University, Department of Economics Prohl, Silika; University of St. Gallen

JEL Code:	C22 - Time-Series Models < C2 - Econometric Methods: Single Equation Models < C - Mathematical and Quantitative Methods, C23 - Models with Panel Data < C2 - Econometric Methods: Single Equation Models < C - Mathematical and Quantitative Methods, H60 - General < H6 - National Budget, Deficit, and Debt < H - Public Economics
Keywords:	Fiscal Sustainability, Panel Cointegration, Structural Change

Panel Cointegration Tests of the Sustainability Hypothesis in Rich OECD Countries*

Joakim Westerlund† and Silika Prohl‡

January 27, 2007

Abstract

This study reexamines the sustainability hypothesis by testing whether government revenues and expenditures for eight rich OECD countries between 1977Q1 and 2005Q4 are cointegrated. For this purpose, a nonstationary panel data approach is adopted, which is general enough to permit for cross-country dependence as well as structural breaks representing major shifts in fiscal policy. In contrast to many earlier studies, the results reported in this study suggest that the sustainability hypothesis cannot be rejected.

JEL Classification: C22; C23; H60.
Keywords: Fiscal sustainability; Panel Cointegration; Structural Change.

1 Introduction

The increased budget deficit and public debt experienced by many OECD countries raises the issue of sustainability of the government finance in the long run, and its potential effects on the economy as a whole. Fiscal policy is constrained by the need to finance the government deficit, which over the longer term implies that the market value of public debt must be offset by the present value of all discounted future budget surpluses. In other words, the fiscal policy is sustainable if the discounted value of debt is zero in the limit. This is the essence of the sustainability hypothesis. If the hypothesis holds, then the ongoing fiscal policy can, at least in principle, be maintained indefinitely, whereas if it fails, then there is a need for discretionary policy actions. In particular, a failure

*Westerlund gratefully acknowledges financial support from the Jan Wallander and Tom Hedelius Foundation, research grant number W2006-0068:1.

†Corresponding author: Department of Economics, Lund University, P. O. Box 7082, S-220 07 Lund, Sweden. E-mail address: joakim.westerlund@nek.lu.se.

‡SIAW - University of St. Gallen, Bodanstraße 8, CH-9000 St. Gallen, Switzerland. E-mail address: silika.prohl@unisg.ch.

implies that the government will inevitably run into problems in managing its debt, thus requiring compensation in terms of higher interest rates leading to a slowdown in the economic growth. Thus, from a policy point of view, the issue of whether the sustainability hypothesis holds or not is indeed an important one. Because of this, and because of the growing concern regarding the debt situation in many OECD countries, sustainability analysis has received much interest in recent empirical research.

Most of the earlier work in this field focused on the stationarity of public deficit and debt as a way to empirically test the sustainability hypothesis. Prominent examples include, among others, Wilcox (1989), MacDonald (1992), Hakkio and Rush (1991) and Quintos (1995) for the United States, Smith and Zin (1991) for Canada, Baglioni and Cherubini (1993) for Italy, Olekalns (2000) for Australia, and perhaps most recently Kirchgaessner and Prohl (2006) for Switzerland. The overall picture being that the sustainability hypothesis does not hold.

As a response to this, more recent work has moved away from examining the stationarity of debt, and towards more flexible testing strategies based on cointegration. In particular, as shown by Quintos (1995), given that public revenues and expenditures are non-stationary, sustainability requires these variables to be cointegrated with a unit slope on expenditures. In the terminology of Quintos (1995), this is known as strong sustainability. It means that no problems, according to the ongoing fiscal policy, are likely to arise in the future. By contrast, weak sustainability refers to the case when the slope lies between zero and one, regardless of whether revenues and expenditures are cointegrated or not. In this case, although the deficit is sustainable, the government is expected to have difficulty in marketing its debt, thus requiring fiscal reforms. Finally, if the slope on expenditures is equal to zero, then the deficit is not sustainable, in which case the sustainability hypothesis is refuted.

Payne (1997) applies this approach to the G7 countries between 1950 and 1990, and rejects the sustainability hypothesis. Similarly, using the same approach to a sample of five countries within the European Union between 1961 and 1995, Papadopoulos and Sidiropoulos (1999) only find evidence of cointegration for Greece and Spain. However, the cointegrating slope on government expenditures is estimated significantly lower than one, suggesting that only the weak form of the sustainability hypothesis is supported. Apparently, despite the strong theoretical appeal of the sustainability hypothesis, most empirical studies tend to reject it.

Although there are many explanations for why the sustainability hypothesis does not seem to hold, this paper focuses on two that might potentially go a long way towards explaining the weak empirical evidence on the cointegration between revenues and expenditures. The first explanation is that conventional time series cointegration tests may have low power against persistent alternatives because of the short sample periods usually employed, see Afonso (2005).

The second explanation is that the empirical relationship between public revenues and expenditures that is being estimated need not be invariant to policy regime changes, which may lead to structural shifts in both revenues and expenditures, and hence in the relationship between them, see for example Quintos (1995) and Martin (2000).

However, while very reasonable and potentially appealing, when analyzed separately, these explanations have been found to be empirically inadequate and far from convincing. This paper therefore offers an alternative route. The idea is that to be able to provide any robust evidence on the cointegration between revenues and expenditures, one needs to consider not the low power and the presence of structural change separately but simultaneously. The intuition is simple. On the one hand, while increasing the length of the sample may be justified from a power point of view, it also increases the probability of a break. On the other hand, modifying existing tests so as to accommodate for structural change is typically very costly in terms of power. Therefore, it is the joint consideration of both these aspects that it is likely to be key when testing the sustainability hypothesis.

One way to accomplish this is to resort to panel data, and the recent econometrical advances within this field that make it possible to construct powerful cointegration tools while simultaneously entertaining the possibility of structural change. Applying this approach to a panel of eight rich OECD countries between 1977Q1 and 2005Q4, in contrast to much of the earlier evidence, we are unable to reject the strong sustainability hypothesis. We also provide evidence suggesting that all countries have been subject to several structural breaks representing major changes in fiscal policy.

The remainder of this paper is organized as follows. In Section 2, we give the brief description of the theoretical model of fiscal sustainability, and of the empirical panel data framework that will be used to test it. Section 3 then reports the empirical results, while Section 4 concludes.

2 Testing the sustainability hypothesis

In this section, we begin with a brief account of the economic theory underlying the sustainability hypothesis, and then we go on to discuss the empirical method that we will use to test it.

2.1 Theoretical underpinnings

The theoretical model of sustainability of the fiscal policy starts with the budget constraint of the government at time t, which is given in nominal terms by

$$g_t + (1 + i_t)b_{t-1} = r_t + b_t, \quad (1)$$

where b_t is the stock of public debt, i_t is the nominal interest rate payable on the public debt, r_t is the government revenue including revenue from seignorage, and g_t is the government expenditure excluding interest payments. Of course, these variables are relatively uninteresting as they do not take into account the size of the economy. We therefore follow the conventional practice of rewriting (1) in terms of GDP ratios as

$$\frac{g_t}{y_t} + \frac{(1+i_t)}{(1+\sigma_t)}\frac{b_{t-1}}{y_{t-1}} \;=\; \frac{r_t}{y_t} + \frac{b_t}{y_t},$$

where y_t is the nominal GDP, while σ_t represents the nominal GDP growth rate. Using capital letters to denote the ratios of the corresponding upper-case variables to nominal GDP, we get

$$C_t + (1+i_t^*)B_{t-1} \;=\; R_t + B_t, \tag{2}$$

where $C_t = G_t + (i_t^* - i^*)B_{t-1}$ and i^* is the mean of $i_t^* = (i_t - \sigma_t)/(1+\sigma_t)$, the growth adjusted interest rate. If we assume that the budget constraint in (2) holds continuously at each t, we can use forward substitution to obtain

$$B_{t-1} \;=\; \sum_{j=t+1}^{\infty}\left(\frac{1}{1+i^*}\right)^{j-t}(R_{j-1} - C_{j-1}) + \lim_{j\to\infty}\left(\frac{1}{1+i^*}\right)^{j-t}B_{j-1}. \tag{3}$$

Now, consider taking expectations conditional on all the information available at time t. Since B_{t-1} is known at t, the hypothesis that the government obeys its intertemporal budget constraint can be expressed as

$$\lim_{j\to\infty} E_t\left(\frac{1}{1+i^*}\right)^{j-t}B_{j-1} \;=\; 0. \tag{4}$$

This transversality condition, also known as the no Ponzi game rule, says that the public debt-GDP ratio should grow no faster on average than the mean rate of interest, meaning that the principal repayments and interests cannot be financed indefinitely by issuing new debt. It then follows that the current stock of public debt is offset by all current and expected discounted future primary surpluses.[1]

2.2 A time series cointegration test

The empirical literature has developed several tests to investigate whether the transversality condition (3) holds. The particular test opted for in this paper is that of Quintos (1995), who suggests testing the sustainability hypothesis by examining whether government revenues and expenditures are cointegrated or not. The intuition behind this test lies in first taking differences, and then

[1] Note that $1/(1+i^*) = (i+\sigma)/(1+i)$, where i and σ are the mean values of i_t and σ_t, respectively. Thus, if $i < \sigma$, then the transversality condition is satisfied automatically.

rewriting (2) in terms of $\widetilde{G}_t = G_t + i_t^* B_{t-1}$, the total expenditure including interest payments, as

$$\begin{aligned}\widetilde{G}_{t-1} - R_{t-1} &= \sum_{j=t+1}^{\infty} \left(\frac{1}{1+i^*}\right)^{j-t} (\Delta R_{j-1} - \Delta C_{j-1}) \\ &+ \lim_{j\to\infty} \left(\frac{1}{1+i^*}\right)^{j-t} \Delta B_{j-1}. \qquad (5)\end{aligned}$$

As before, the sustainability hypothesis holds if the expected value of the second term on the right-hand side is zero. One way to test whether this is in fact the case, given that $\widetilde{G}_t$ and R_t possess unit roots, involves first estimating and then testing for cointegration in the following regression

$$R_t = \alpha + \beta \widetilde{G}_t + e_t, \qquad (6)$$

where e_t is a mean zero disturbance term. To see what this means for the test of the sustainability hypothesis, note that $\Delta B_t = \widetilde{G}_t - R_t$, which together with (5) implies that

$$\Delta B_t = (1-\beta)\widetilde{G}_t - \alpha - e_t. \qquad (7)$$

As shown by Quintos (1995), although the sustainability hypothesis holds regardless of the integratedness of ΔB_t, the speed with which this happens does not. If ΔB_t is stationary, then the second term in (5) vanishes at a much faster rate than if ΔB_t is non-stationary. In the terminology of Quintos (1995), the sustainability of the public deficit is said to be strong if ΔB_t is stationary, whereas it is said to be weak if ΔB_t is non-stationary.[2]

In terms of the regression in (6), by looking at (7), we see that cointegration with $\beta = 1$ is enough for strong sustainability, because only then will ΔB_t be stationary. But since $\widetilde{G}_t$ is non-stationary, the sustainability is weak if either $\beta = 1$ and there is no cointegration or $0 < \beta < 1$. It follows that $0 < \beta \leq 1$ is both necessary and sufficient for sustainability, while cointegration is only sufficient. The distinction between strong and weak sustainability is made because although $0 < \beta < 1$ is enough for the sustainability hypothesis to hold, the government is now spending more than it earns, thus making the public debt difficult to market in the long run. Finally, if $\beta \leq 0$ the deficit is no longer sustainable because this would make ΔB_t grow faster than the growth rate of the economy, as approximated by the mean interest rate.[3]

The above discussion suggests that testing for cointegration is key in this kind of sustainability analysis, not only because of its role in determining strong sustainability but also because we cannot do inference regarding β unless revenues and expenditures are cointegrated.

[2] Bergman (2001) extends the results of Quintos (1995) by showing that sufficient condition for government solvency is that government debt is integrated of any finite order.

[3] Note that the case with $\beta > 1$ is not consistent with a deficit, since revenues are growing faster than expenditures.

The conventional way in which earlier studies have been trying to test the sustainability hypothesis within this framework involves first estimating (6) by ordinary least squares (OLS) and then testing whether the residuals from that equation can be treated as stationary or not by using any conventional cointegration test. This test is then repeated for every country in the sample, each time using only the sample information for that particular country. Studies based on this approach are generally unable to reject the null hypothesis of no cointegration, which is seen to imply that the strong form of the sustainability hypothesis should be rejected. In this paper, we argue that this result should not be taken at face value as a failure to reject the null may well be due to low power.

In fact, it is well known that tests that take no cointegration as the null hypothesis can have very little power against nearly cointegrated alternatives, especially if there are structural breaks present. In a situation like this, it is therefore essential to device tests with increased power. A natural approach to do this would be to combine the sample information obtained from the time series dimension with that obtained from the cross-sectional. This not only increases the power by taking the total number of observations and their variation into account, but also increases the precision of the test by effectively reducing the noise coming from the individual time series regressions. Therefore, one way to augment the power of univariate tests would be to subject the residuals from (6) to a panel data cointegration test.

Unfortunately, most existing tests of this kind rely critically on assuming that the countries are independent of each other, which is unlikely to hold in the present application because of strong intra-economy linkages. Another limitation of most panel cointegration tests is that they ignore the possibility of structural change, which is likely to be the rule rather than the exception in this kind of fiscal data. Thus, a first crucial step in testing the sustainability hypothesis using panel data is to employ tests that allow for structural breaks, and that do not rely to such a large extent on the countries being independent.

2.3 A panel cointegration test with breaks

The previous section suggests that testing for cointegration in cross-country dependent data with structural change is key in inferring the sustainability hypothesis. In this section, we outline a test that fits this description, and that is general enough to allow for both cross-country dependence and an unknown number of breaks that may be located at different dates for different countries. In so doing, we will use the index $i = 1, ..., N$ to denote countries, while t again denotes time.

Consider the following panel version of the regression in (6)

$$R_{it} \;=\; \alpha_{ij} + \beta_i \widetilde{G}_{it} + e_{it}, \quad j \;=\; 1, ..., M_i + 1, \tag{8}$$

where β_i is a country specific slope that is assumed to be constant over time, while α_{ij} is a country specific intercept that is subject to M_i structural breaks. In other words, there are $M_i + 1$ regimes for each country i with the j^{th} regime running form T_{ij-1} to T_{ij} time series observations.

Sustainability requires (8) being a cointegrated relationship. If there is no structural change, then this hypothesis can be readily tested by using the existing tests for cointegration in panel data. If there are breaks, however, then this test procedure is no longer valid since the relationship in (8) is now non-linear. This poses a serious problem for inference since conventional tests cannot be used to discriminate between cointegration with structural change and the absence of cointegration. This issue was recently addressed by Westerlund (2006), who develops a panel Lagrange multiplier test for cointegration that allows for multiple structural breaks.

The null hypothesis is formulated as that all the countries in the panel are cointegrated, while the alternative is formulated as that there is at least some country for which cointegration does not hold. The test statistic for this particular hypothesis can be written as

$$Z(M) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M_i+1}\sum_{t=T_{ij-1}+1}^{T_{ij}} S_{it}^2/(T_{ij}-T_{ij-1})^2\widehat{\sigma}_i^2,$$

where $S_{it} = \sum_{s=T_{ij-1}+1}^{t} \widehat{e}_{is}$ and $\widehat{e}_{it}$ is the regression residual obtained by using any efficient estimator of the cointegration vector such as the conventional fully modified OLS (FMOLS) estimator. Given that the countries are independent, and some technical conditions regarding the persistency of the individual time series, Westerlund (2006) shows that $Z(M)$ reaches the following sequential limit as $T \to \infty$ and then $N \to \infty$ under the null hypothesis

$$\frac{\sqrt{N}(Z(M) - E(Z(M)))}{\sqrt{\text{var}(Z(M))}} \Rightarrow N(0,1),$$

where the mean and variance adjustment terms $E(Z(M))$ and $\text{var}(Z(M))$ are defined in Westerlund (2006). Thus, by standardizing $Z(M)$ by its mean and standard deviation, we obtain a new test statistic that has an asymptotic standard normal distribution under the null hypothesis. Under the alternative hypothesis, the statistic diverges to positive infinity suggesting that the right tail of the normal distribution should be used to reject the null.

For the estimation of the number of breaks and their locations, Westerlund (2006) suggests using the Bai and Perron (2003) procedure, which can be implemented in two steps. In the first, the breakpoints are estimated by minimizing the sum of squared residuals for all permissable values of M_i. In the second, the estimated breakpoints are used together with the associated sum of squares to estimate the number of breaks using an information criterion. These steps are then repeated N times to obtain the estimated number of breaks and their locations for each country.

The fact that both the number of breaks and their locations are treated as unknown here is a clear advantage in comparison to other approaches. For example, most studies that allow for structural change, see Wilcox (1989) and Hakkio and Rush (1991), often maintain that the breaks are known, which means that they are likely to suffer from pre-test bias. Other studies such as Haug (1995) and Quintos (1995) allow for unknown breakpoints but restrict the number of breaks to one, which is not likely to be the case.

To handle the impact of cross-country dependence Westerlund (2006) suggests using the bootstrap approach. The particular bootstrap scheme opted for in this paper uses the sieve approach of Psaradakis (2003), who proposes a bootstrapped stationarity test for pure time series. The advantage with this scheme is that it can be modified to preserve not only the cross-country correlations but also the serial correlations.

3 Empirical results

In this section, we first briefly describe the data, and then we present the empirical unit root and cointegration test results. Finally, we present some results of the estimated cointegration vectors.

3.1 Data

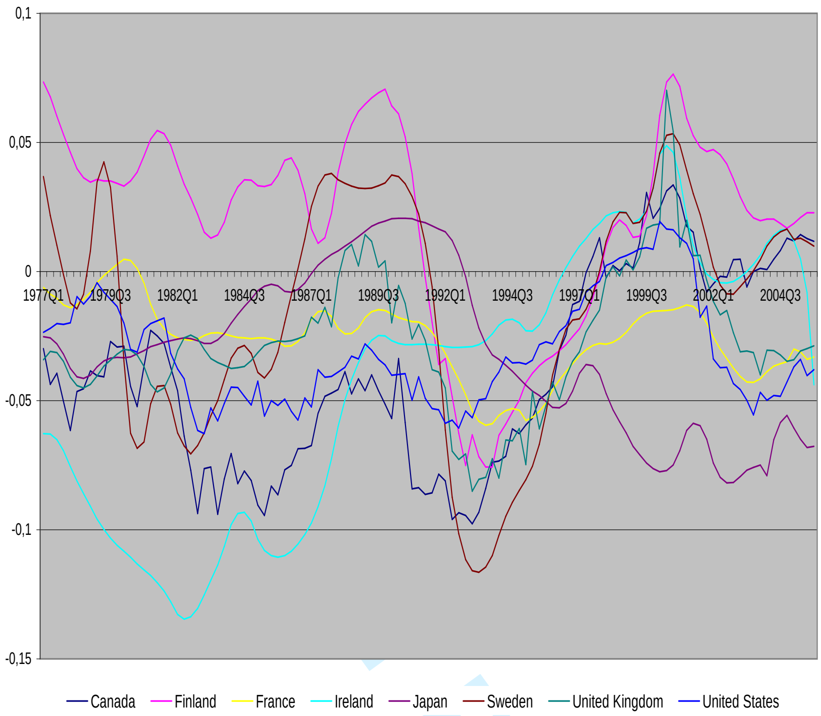
The sustainability hypothesis is tested by using quarterly data that covers eight rich OECD countries over the period 1977Q1 to 2005Q4. The countries included in the panel are Canada, Finland, France, Ireland, Japan, Sweden, the United Kingdom, and the United States. The data are taken from the OECD Economic Outlook Database Inventory, updated in February 2006, and include general government receipts and disbursement, and GDP at market prices. As in Section 2, the variables used in the analysis were converted into GDP ratios.

3.2 Graphical analysis

In order to get a feeling of the behavior of the fiscal debt situation in the OECD area, we begin with a graphical inspection of the data. To foreshadow the more formal treatment in the next two sections, Figure 1 plots the deficit-GDP ratios of the eight countries in the sample. The figure clearly illustrates that while individually quite dispersed, the series tend to revert back towards a common mean value, which seem to have become more positive in recent years. Thus, if sustainability is to be interpreted as the mean reversion, or stationarity, of the public deficit, it appears that there is some indication of sustainability among the countries.

It is also interesting to consider Figure 1 in view of the major public reforms that has taken place during the length of the sample. The first thing to notice

Figure 1: Deficit-GDP ratios.



is increased deficit-GDP ratios in the first half of the sample, which originated with public reforms based on tight monetary and loose fiscal policy (OECD, 2002). These deficits made it impossible for governments to compensate the spiraling debt interest payments. Hence, in order to stabilize and restore public finances, and to enhance output growth in the long-run, new fiscal rules was put in place in the United States, and the European Union member countries introduced stricter budget deficit and fiscal debt criteria. These reforms have in recent years made it possible for governments to move away from a strategy of controlling unstable debt dynamics, and towards policy actions designed to strengthen the fiscal stance. This improvement in the overall debt-GDP stance in the latter half of the sample is clearly visible in the figure.

3.3 Stationarity tests

We begin by considering the integratedness of the government revenue and expenditure series using the panel stationarity test of Carrion-i-Silvestre *et al.*

(2005), which is very similar to the Westerlund (2006) cointegration test in the sense that it permits for both cross-country dependence and structural breaks. This is therefore an appropriate test for our purposes. Another similarity with the cointegration test is that the breaks are estimated using the Bai and Perron (2003) procedure and that the bootstrap is of sieve type.

In applying the Carrion-i-Silvestre *et al.* (2005) test, we follow the recommendation of Newey and West (1994) and use the Bartlett kernel, where the bandwidth parameter is permitted to grow with T at the rate $T^{1/3}$. In estimating the number of breaks, we allow for a maximum of five shifts for each country, which seem sufficient to capture most of the breaks in the data. The exact number of breaks for each country is estimated using the Schwartz information criterion. Also, to ensure that the break date estimator work properly, we set the minimum length of each regime equal to $0.1T$. The sieve bootstrap is implemented using a lag length of five and we make 1,000 replications.

As for the deterministic specification, since all variables are expressed in GDP ratios, no time trends are required. Two models are considered. The first assumes a fixed mean while the second allows for up to five breaks in the mean of each series. Among these models, since most of the literature indicates the presence of breaks, the latter stands out as the most natural one.

Table 1 shows the results of panel stationary test for each variable. For each model, the first row represents the test value, the second row contains the asymptotic p-values, and the third row contains the bootstrapped p-values. The results from the asymptotic p-values suggest that the null hypothesis of stationary can be safely rejected. However, these values assume that the countries are independent, which is unlikely to hold. In order to account for this dependence, we use the bootstrapped p-values instead. Except for revenues in the structural break model, the conclusions are not altered by taking the cross-country correlations into account. Thus, since the overall evidence in favor of a rejection is quite overwhelming, we choose to proceed as if the variables are in fact non-stationary.

3.4 Cointegration tests

The second step in our analysis is to test whether the variables are cointegrated using the $Z(M)$ test. The results are presented in the rightmost column of Table 1. It is seen that the p-values for the model with a fixed constant provide very little support of cointegration. However, these results do not take into account the possibility of structural breaks, and are therefore prone to erroneous conclusions. Indeed, if we allow for structural shifts and cross-country correlation, the null hypothesis of cointegration cannot be rejected at the 5 percent level. We therefore conclude that the variables appear to be cointegrated around a broken intercept.

Table 2 reports the estimated breakpoints obtained from the Bai and Perron (2003) procedure. It is seen that there are at least three breaks for each country, which is indicative of structural instability.[4] This evidence seems consistent with the gradual shift towards increased business cycles synchronization among the OECD countries, which has substantially affected the fiscal position of many governments, see Fatas and Mihov (2003). The first break period in the 1980's reflects the effect of the Great Depression, during which the synchronized recession coincided with an increase in real interest rate. This caused a substantial deterioration of the fiscal stance of of governments of rich OECD countries, as the deficits created by the spiralling interest payments could not be offset, see OECD (2002).

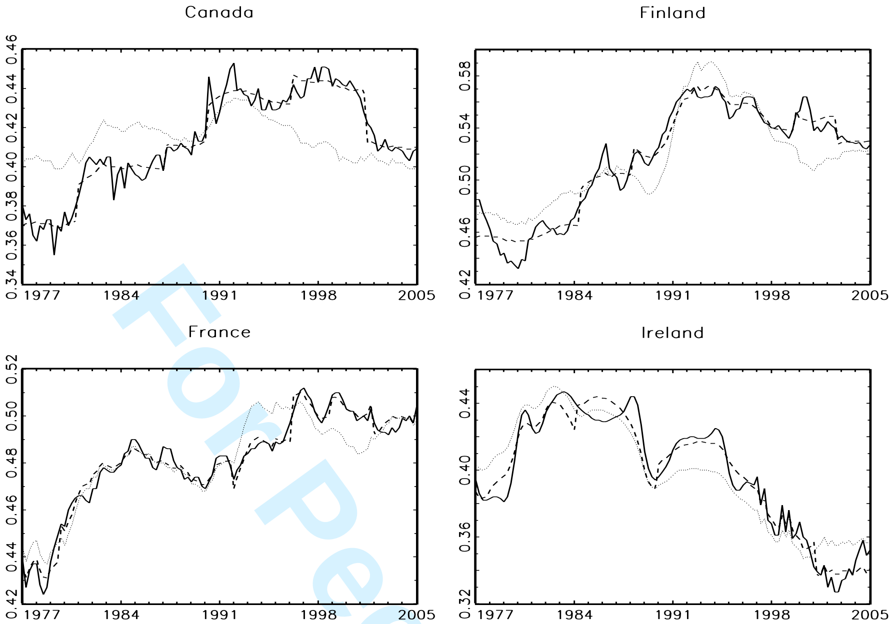
Later, following the Maastricht Treaty and the Stability and Growth Pact in the early 1990's, the fiscal stance of many OECD countries were driven primarily by an increased coordination of both fiscal and monetary policy within the European Monetary Union, see Darvas *et al.* (2005). This period is also clearly visible in the table, as indicated by the concentration of breaks in the early 1990's, and seems consistent with the results reported in many earlier studies, see for example Afonso (2005). The many breaks in the late 1990's are most likely due to the formation of the European Monetary Union in January 1999.

To further illustrate the importance of accounting for structural change, Figures 2 and 3 plot the revenues-GDP ratio, the dependent variable, together with the fitted trend functions for the models with and without breaks. In the figures, the solid line represents revenues measured as a fraction of GDP, the dashed line represents the model with breaks and the dotted line represents the model without any breaks. It can easily be seen that nearly all revenue series exhibit several level breaks, most of which seem to reflect the effect of economic fluctuations, and changes in the monetary and fiscal policy objectives in the individual countries. Interestingly, it is seen that the revenues-GDP ratio of most countries have been rising during the first half of the sample and falling during the second half, which reflects a more flexible response of the recent fiscal policy of most OECD governments to the economic downturn.

More importantly, we see that the trend lines for the model with breaks seem to provide a very good fit to the revenues for all eight countries. By contrast, the trend lines for the model without breaks are generally incapable of accounting for the all variation in revenues, thus leading to a poor fit. This effect is particularly striking for the United States.

[4]Note that for five of the eight countries we end up estimating the maximum number of breaks. This is in agreement with the results of Ahamada *et al.* (2004) and Perron (1997), who point out that the type of information criteria used here cannot directly take into account the effect of different distributions of the data and possible serial correlation in the regression errors. In fact, as Perron (1997) showed in simulations, although most criteria perform reasonably well when the errors are not correlated, they have a strong tendency to overestimate the number of breaks when serial correlation is present. Of course, the purpose in this paper is not the correct estimation of the number of breaks *per se* but rather the accounting of the effects of these breaks. Thus, overestimation is not a serious problem in the sense that the model is still free to estimate the associated break parameters to zero.

Figure 2: Revenues-GDP ratio with fitted trend functions.



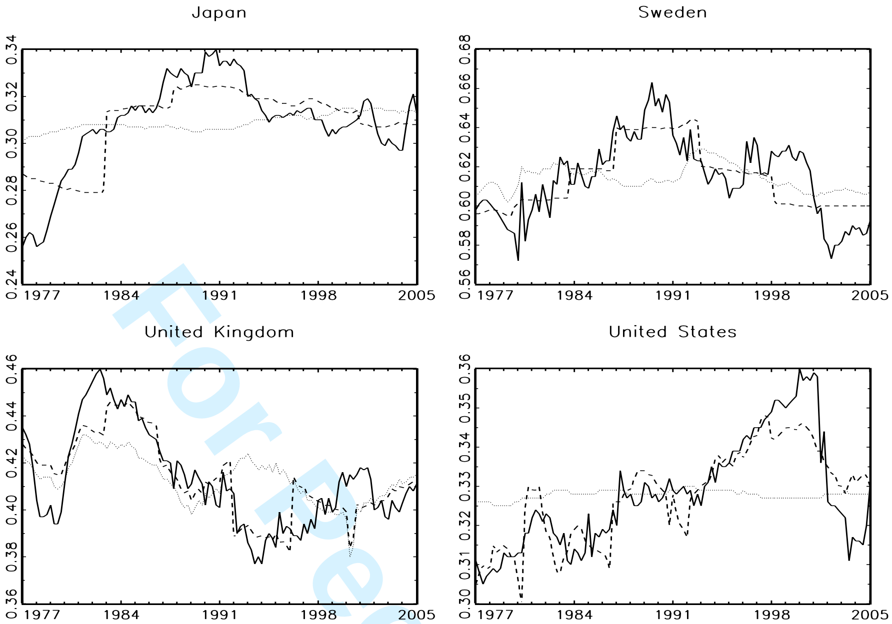
3.5 Cointegration estimation

Since the variables appear to be cointegrated, it is possible to estimate and test whether the cointegrating slope is indeed equal to one as required by sustainability hypothesis.

It is well known that the OLS estimator is consistent under fairly general conditions when applied to the cointegrated regression in (6). Unfortunately, if the regressor is endogenous, then this estimator suffers from nuisance parameter dependencies even asymptotically, which makes it a poor candidate for inference. To account for this, in addition to using OLS, we employ both FMOLS and dynamic OLS (DOLS) panel estimation techniques, see Kao and Chiang (2000). All three estimators are based on pooling along the cross-sectional dimension, and are appropriate for testing the null hypothesis that $\beta_i = 1$ for all i against the alternative hypothesis that β_i is equal to some common slope value, which is different from unity.

Moreover, as when testing for stationarity and cointegration in panel data, the presence of cross-country dependence makes inference based on the asymptotic normal distribution inappropriate, in which case bootstrap inference might be better. The particular bootstrap opted for this section is taken from Chang *et al.* (2006), who propose a sieve resampling scheme that preserves the serial correlation properties of the errors, and that can be generalized along the lines

Figure 3: Revenues-GDP ratio with fitted trend functions.



of Westerlund (2006) to also accommodate cross-country dependence. As an indication of the severity of this problem, we applied the Pesaran (2004) test of cross-section correlation to the OLS residuals of the estimated cointegrated regression.[5] The computed test value was 4.98, which, when compared to the right tail of the asymptotic normal distribution, lead to a clear rejection of the no correlation null at all conventional significance levels. Thus, cross-country correlation is indeed a problem that needs to be addressed.

The results from the OLS, FMOLS and DOLS estimators are reported in Table 3 together with the double-sided p-values for the null hypothesis of a unit slope. The first thing to notice is that the estimated slopes lie very close to their hypothesized values of one. Indeed, the range of the estimated slopes is 0.909 to 0.934. The closeness of these estimates to their expected value based on the sustainability hypothesis is supported by the p-values.

Indeed, by looking at the p-values based on the asymptotic normal distribution, we see that the null hypothesis of a unit slope cannot be rejected even at the most generous 10 percent level. The results from the bootstrapped p-values, which allow for cross-country dependence, are even more supportive of the unit slope null, and are all well above 0.6. Thus, based on this evidence we cannot

[5]The Pesaran (2004) test is basically the average pair-wise correlation coefficient of the regression residuals, which has an asymptotic normal distribution under the null hypothesis of no cross-correlation.

reject the null hypothesis of a unit slope in the panel. Thus, in contrast to much of the earlier evidence regarding the sustainability of fiscal policy, we are unable to reject the sustainability hypothesis for the panel as a whole.

4 Concluding remarks

Earlier empirical research based on testing for stationarity of public debt and budget deficit has tended to reject the hypothesis of fiscal sustainability. As a response to this, recent work within this field has turned towards more flexible testing approaches based on cointegration. The idea is that if fiscal policy is sustainable, then government revenues and expenditures should be cointegrated with a unit slope on expenditures. Unfortunately, the results of the earlier studies have not been very convincing.

In this paper, we provide some insights suggesting that the weak empirical support of the sustainability hypothesis may be due to inadequate econometric methods. In particular, we argue that a failure of accounting for structural breaks and the poor precision of commonly applied time series tests may well result in a rejection of the sustainability hypothesis. To circumvent these problems, we resort to some recent advances in the area of non-stationary panel data. These methods are more powerful than conventional time series methods, and are general enough to allow for cross-country dependence and an unknown number of structural breaks in the cointegrating relationship.

Based on data covering eight rich OECD countries over the period from 1977Q1 to 2005Q4, we show that revenues and expenditures are non-stationary and cointegrated. We also provide evidence suggesting that the null hypothesis of a common unit slope on expenditures cannot be rejected, suggesting that the sustainability hypothesis cannot be rejected. The results on the estimated structural breaks are suggestive of strong structural instability, and are compatible with the findings of many earlier studies such as Quintos (1995), Martin (2000) and Afonso (2005). Some of the estimated breaks are related to the increased business cycle synchronization among the OECD countries, while others are related to the policy interventions resulting from the increased coordination of the monetary and fiscal policy within the European community.

Thus, although there is a growing concern about increased budget deficits within the OECD area, our results indicate that after taking into account structural change and economic growth, the budget of the eight rich countries considered in this study has in fact been balanced, at least intertemporally over the sample considered. In other words, for these countries the sequence of primary surpluses has been sufficient to cover the marked value of public debt. Thus, from an historical point of view, there is no reason to be concerned about the public debt situation.

This finding is in agreement with the results reported by Quintos (1995) for the United States, and is particularly interesting in view of the recent discussion regarding the effect of the projected increase in future public liabilities to age-related public spending and health care on public finances during the demographic transition of OECD countries to older societies. Our results imply that if governments are able to raise revenues at the same rate as the expected increase in spending, so that revenues and expenditures are one-to-one in the long run, then one need not worry about the future public debt situation.

Table 1: Panel stationarity and cointegration tests.

Specification	Test	Stationarity tests R_{it}	$\widetilde{G}_{it}$	$Z(M)$
Fixed constant	Value	20.147	18.684	16.661
	p-value^a	0.000	0.000	0.000
	p-value^b	0.000	0.002	0.028
Constant with breaks	Value	2.192	5.114	1.416
	p-value^a	0.014	0.000	0.078
	p-value^b	0.242	0.008	0.678

Notes: All test results allow for up to five structural breaks for each state. To handle the serial correlation, the Bartlett kernel is used with a bandwidth parameter of $T^{1/3}$.

^{a}The p-value is based on the asymptotic normal distribution.

^{b}The p-value is based on the bootstrapped distribution. The number of lags in the sieve approximation is five and we make 1,000 bootstrap replications.

Table 2: Estimated breaks.

Country	No.	Breakpoint 1	2	3	4	5
Canada	5	1981Q1	1987Q2	1990Q2	1996Q2	2001Q3
Finland	4	1984Q3	1988Q2	1999Q4	2002Q4	–
France	5	1979Q4	1992Q1	1996Q2	1999Q1	2002Q1
Ireland	3	1983Q1	1987Q4	2001Q1	–	–
Japan	5	1979Q4	1983Q4	1987Q1	1993Q1	1998Q2
Sweden	5	1983Q1	1986Q4	1992Q1	1996Q2	2001Q2
United Kingdom	4	1980Q3	1986Q4	1992Q2	1998Q1	–
United States	5	1980Q3	1987Q2	1994Q2	1997Q3	2002Q1

Notes: The breaks are estimated using the Bai and Perron (2003) procedure with a maximum number of five breaks for each state. The minimum length of each break regime is set to $0.1T$.

Table 3: Panel cointegration slope estimates.

Estimator	Value	p-value^a	p-value^b
OLS	0.925	0.594	0.708
DOLS	0.934	0.689	0.718
FMOLS	0.909	0.444	0.602

Notes: All results allow for up to five structural breaks for each state. The p-values for the between and within estimates are for the null of a unit slope against the alternative that the slope is different from unity.
[a]The p-value is based on the asymptotic normal distribution.
[b]The p-value is based on the bootstrapped distribution. The number of lags in the sieve approximation is five and we make 1,000 bootstrap replications.

References

Afonso, A. (2005). Fiscal Sustainability: the unpleasant European case. *FinanzArchive* 61, 19-44.

Ahamada, I., Jouini, J., and Boutahar, M. (2004). Detecting multiple breaks in time series covariance structure: a non-parametric approach based on the evolutionary spectral density. *Applied Economics* 36, 1095-1101.

Ahmed, S., and Rogers, J. (1995). Government budget deficits and trade deficits. Are present value constraint satisfied in long-term data? *Journal of Monetary Economics* 36, 351-374.

Baglioni, A., and Cherubini, U. (1993). Intertemporal budget constraint and public debt sustainability: the case of Italy. *Applied Economics* 25, 275-283.

Bai, J., and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics* 18, 1-22.

Bergman, M. (2001). Testing government solvency and the No Ponzi Game condition. *Applied Economics Letters* 8, 27-29.

Bohn, H. (1991). The sustainability of budget deficit with lump-sum and with income-based taxation. *Journal of Money, Credit and Banking* 23, 581-604.

Bordo, M.D., and Heibling, T. (2003). Have national business cycles become more synchronized? NBER Working Paper No. 10130.

Breitung, J., and Pesaran, M. H. (2005). Unit roots and cointegration in panels. In Matyas, L., and Sevestre, P. (Eds.), *The Econometrics of Panel Data*. Klüver Academic Publishers.

Carrion-i-Silvestre, J. L., Del Barrio-Castro, T., and López-Bazo, E. (2005). Breaking the panels: an application to the GDP per capita. *Econometrics Journal* 8, 159-175.

Chang, Y., Park, J. Y., and Song, K. (2006). Bootstrapping cointegrating regressions. *Journal of Econometrics* 133, 703-739.

Darvas, Z., Rose, A. K., and Szapary, G. (2005). Fiscal divergence and business cycle synchronization: irresponsibility is idiosyncratic. NBER Working Paper No. 11580.

European Commission. (2002). Public finances in EMU. *European Economy* 3.

Fatas, A., and Mihov, I. (2003). The case of restricting fiscal policy discretion. *Quarterly Journal of Economics* 118-4, 1419-1447.

Hakkio, G., and Rush, M. (1991). Is the budget deficit too large? *Economic Inquiry* 29, 429-445.

Haug, A. (1995). Has federal budget deficit policy changed in recent years? *Economic Inquiry* 33, 104-118.

IMF (2001). Business cycle linkages among major advanced economics. *World Economic Outlook*, October 2001.

Kao, C., and Chiang, M. H. (2000). On the estimation and inference of a cointegrated regression in panel data. *Advances in Econometrics* 15, 179-222.

Kirchgaessner, G., and Prohl, S. (2006). Sustainability of Swiss fiscal policy. Forthcoming in *Swiss Journal of Economics and Statistics*.

MacDonald, R. (1992). Some tests of the government's intertemporal budget constraint using U.S. data. *Applied Economics* 24, 1287-1292.

Martin, G. (2000). U.S. deficit sustainability: a new approach based on multiple endogenous breaks. *Journal of Applied Econometrics* 15, 83-105.

Newey, W., and West, K. (1994). Autocovariance lag selection in covariance matrix estimation. *Review of Economic Studies* 61, 631-653.

OECD (2002). Fiscal sustainability: the contribution of fiscal rules. *OECD Economic Outlook* 72.

Olekalns, N. (2000). Sustainability and stability? Australian fiscal policy in the 20th century. *Australian Economic Papers* 39, 138-151.

Papadopoulos, A., and Sidiropoulos, M. (1999). The sustainability of fiscal policies in the European Union. *International Advances in Economic Research* 5, 289-307.

Payne, J. (1997). International evidence on the sustainability of budget deficits. *Applied Economic Letters* 12, 775-779.

Perron, P. (1997). L'estimation de modèles avec changements structurels multiples. *Actualit conomique* 73, 457-505.

Pesaran, H. (2004). General diagnostic tests for cross section dependence in panels. Cambridge Working Paper in Economics 435.

Psaradakis, Z. (2003). A sieve bootstrap test for stationary. *Statistics and Probability Letters* 62, 263-274.

Quintos, C. (1995). Sustainability of the deficit process with structural shifts. *Journal of Business and Economic Statistics* 13, 409-417.

Smith, G., and Zin, S. (1991). Persistent deficits and the market value of government debt. *Journal of Applied Econometrics* 6, 31-44.

Westerlund, J. (2006). Testing for panel cointegration with multiple structural breaks. *Oxford Bulletin of Economics and Statistics* 68, 101-132.

Wilcox, D. (1989). The sustainability of government deficits: implications of the present-value borrowing constraint. *Journal of Money, Credit, and Banking* 21, 291-306.

Panel Cointegration Tests of the Sustainability Hypothesis in Rich OECD Countries*

Joakim Westerlund† and Silika Prohl‡

October 7, 2006

Abstract

This study reexamines the sustainability hypothesis by testing whether government revenues and expenditures for eight rich OECD countries between 1977Q1 and 2005Q4 are cointegrated. For this purpose, a nonstationary panel data approach is adopted, which is general enough to permit for cross-country dependence as well as structural breaks representing major shifts in fiscal policy. In contrast to many earlier studies, the results reported in this study suggest that the sustainability hypothesis cannot be rejected.

JEL Classification: C22; C23; H60.

Keywords: Fiscal Sustainability; Panel Cointegration; Structural Change.

1 Introduction

The increased budget deficit and public debt experienced by many OECD countries raises the issue of sustainability of the government finance in the

*Westerlund gratefully acknowledges financial support from the Jan Wallander and Tom Hedelius Foundation, research grant number W2006-0068:1.

†Corresponding author: Department of Economics, Lund University, P. O. Box 7082, S-220 07 Lund, Sweden. E-mail address: joakim.westerlund@nek.lu.se.

‡SIAW - University of St. Gallen, Bodanstraße 8, CH-9000 St. Gallen, Switzerland. E-mail address: silika.prohl@unisg.ch.

long run, and its potential effects on the economy as a whole. Fiscal policy is constrained by the need to finance the government deficit, which over the longer term implies that the market value of public debt must be offset by the present value of all discounted future budget surpluses. In other words, the fiscal policy is sustainable if the discounted value of debt is zero in the limit. This is the essence of the sustainability hypothesis. If the hypothesis holds, then the ongoing fiscal policy can, at least in principle, be maintained indefinitely, whereas if it fails, then there is a need for discretionary policy actions. In particular, a failure implies that the government will inevitably run into problems in managing its debt, thus requiring compensation in terms of higher interest rates leading to a slowdown in the economic growth.

Thus, from a policy point of view, the issue of whether the sustainability hypothesis holds or not is indeed an important one. Because of this, and because of the growing concern regarding the debt situation in many OECD countries, sustainability analysis has received much interest in recent empirical research. Most of the earlier work in this field focused on the stationarity of public deficit and debt as a way to empirically test the sustainability hypothesis. Prominent examples include, among others, Wilcox (1989), MacDonald (1992), Hakkio and Rush (1991) and Quintos (1995) for the United States, Smith and Zin (1991) for Canada, Baglioni and Cherubini (1993) for Italy, Olekalns (2000) for Australia, and perhaps most recently Kirchgaessner and Prohl (2006) for Switzerland. The overall picture being that the sustainability hypothesis does not hold.

As a response to this, more recent work has moved away from examining the stationarity of debt, and towards more flexible testing strategies based on cointegration. In particular, as shown by Quintos (1995), given that public revenues and expenditures are non-stationary, sustainability requires these variables to be cointegrated with a unit slope on expenditures. In the terminology of Quintos (1995), this is known as strong sustainability. It means that no problems, according to the ongoing fiscal policy, are likely to arise in the future. By contrast, weak sustainability refers to the case when the slope lies between zero and one, regardless of whether revenues and expenditures are cointegrated or not. In this case, although the deficit is sustainable, the government is expected to have difficulty in marketing its debt, thus requiring fiscal reforms. Finally, if the slope on expenditures is equal to zero, then the deficit is not sustainable, in which case the sustainability hypothesis is refuted.

Payne (1997) applies this approach to the G7 countries between 1950 and 1990, and rejects the sustainability hypothesis. Similarly, using the same approach to a sample of five EU countries between 1961 and 1995, Papadopoulos and Sidiropoulos (1999) only find evidence of cointegration for Greece and Spain. However, the cointegrating slope on government expenditures is estimated significantly lower than one, suggesting that only the weak form of the sustainability hypothesis is supported. Apparently, despite the strong theoretical appeal of the sustainability hypothesis, most empirical studies tend to reject it.

Although there are many explanations for why the sustainability hypothesis does not hold, this paper focuses on two that might potentially go a long way towards explaining the weak empirical evidence on the cointegration between revenues and expenditures. The first explanation is that conventional time series cointegration tests may have low power against persistent alternatives because of the short sample periods usually employed, see Afonso (2005). The second explanation is that the empirical relationship between public revenues and expenditures that is being estimated need not be invariant to policy regime changes, which may lead to structural shifts in both revenues and expenditures, and hence in the relationship between them, see for example Quintos (1995) and Martin (2000).

However, while very reasonable and potentially appealing, when analyzed separately, these explanations have been found to be empirically inadequate and far from convincing. This paper therefore offers an alternative route. The idea is that, to be able to provide any robust evidence on the cointegration between revenues and expenditures, one needs to consider not the low power and the presence of structural change separately but simultaneously. The intuition is simple. First, while increasing the length of the sample may be justified from a power point of view, it also increases the probability of a break. On the other hand, modifying existing tests so as to accommodate for structural change is typically very costly in terms of power. Therefore, it is the joint consideration of both these aspects that it is likely to be key approach when testing the sustainability hypothesis.

One way to accomplish this is to resort to panel data, and the recent econometrical advances within this field that make it possible to construct powerful cointegration methods while simultaneously entertaining the possibility of structural change. Applying this approach to panel of eight rich OECD countries between 1977Q1 and 2005Q4, in contrast to much of the earlier evidence, we are unable to reject the strong sustainability hypothesis. We also provide evidence suggesting that all countries have been subject to several structural breaks representing major changes in fiscal policy.

The remainder of this paper is organized as follows. In Section 2, we give the brief description of the theoretical model of fiscal sustainability and the empirical panel data framework that will be used to test it. Section 3 then reports the empirical results, while Section 4 concludes.

2 Testing the sustainability hypothesis

In this section, we begin with a brief account of the economic theory underlying the sustainability hypothesis, and then we go on to discuss the empirical method that we will use to test it.

2.1 Theoretical underpinnings

The theoretical model of sustainability of the fiscal policy starts with the budget constraint of the government at time t, which is given in nominal terms by

$$g_t + (1 + i_t)b_{t-1} = r_t + b_t, \tag{1}$$

where b_t is the stock of public debt, i_t is the nominal interest rate payable on the public debt, r_t is the government revenue including revenue from seignorage, and g_t is the government expenditure excluding interest payments. Of course, these variables are relatively uninteresting as they do not take into account the size of the economy. We therefore follow the conventional practice of rewriting (1) in terms of GDP ratios as

$$\frac{g_t}{y_t} + \frac{(1+i_t)}{(1+\sigma_t)}\frac{b_{t-1}}{y_{t-1}} \;=\; \frac{r_t}{y_t} + \frac{b_t}{y_t},$$

where y_t is the nominal GDP, while σ_t represents the nominal GDP growth rate. Using capital letters to denote the ratios of the corresponding upper-case variables to nominal GDP, we get

$$C_t + (1+i_t^*)B_{t-1} \;=\; R_t + B_t, \tag{2}$$

where $C_t = G_t + (i_t^* - i^*)B_{t-1}$ and i^* is the mean of $i_t^* = (i_t - \sigma_t)/(1+\sigma_t)$, the growth adjusted interest rate. If we assume that the budget constraint in (2) holds continuously at each t, we can use forward substitution to obtain

$$B_{t-1} \;=\; \sum_{j=t+1}^{\infty}\left(\frac{1}{1+i^*}\right)^{j-t}(R_{j-1} - C_{j-1}) + \lim_{j\to\infty}\left(\frac{1}{1+i^*}\right)^{j-t} B_{j-1}. \tag{3}$$

Now, consider taking expectations conditional on all the information available at time t. Since B_{t-1} is known at t, the hypothesis that the government obeys its intertemporal budget constraint can be expressed as

$$\lim_{j\to\infty} E_t\left(\frac{1}{1+i^*}\right)^{j-t} B_{j-1} \;=\; 0. \tag{4}$$

This transversality condition, also known as the no Ponzi game rule, says that the public debt-GDP ratio should grow no faster on average than the mean rate of interest, meaning that the principal repayments and interests cannot be financed indefinitely by issuing new debt. It then follows that the current stock of public debt is offset by all current and expected discounted future primary surpluses.[1]

[1] Note that $1/(1+i^*) = (i+\sigma)/(1+i)$, where i and σ are the mean values of i_t and σ_t, respectively. Thus, if $i < \sigma$, then the transversality condition is satisfied automatically.

2.2 A time series cointegration test

The empirical literature has developed several tests to investigate whether the transversality condition (3) holds. The particular test opted for in this paper is that of Quintos (1995), who suggests testing the sustainability hypothesis by examining whether government revenues and expenditures are cointegrated or not. The intuition behind this test lies in first taking differences, and then rewriting (2) in terms of $\widetilde{G}_t = G_t + i_t^* B_{t-1}$, the total expenditure including interest payments, as

$$\begin{aligned}\widetilde{G}_{t-1} - R_{t-1} &= \sum_{j=t+1}^{\infty} \left(\frac{1}{1+i^*}\right)^{j-t} (\Delta R_{j-1} - \Delta C_{j-1}) \\ &+ \lim_{j\to\infty} \left(\frac{1}{1+i^*}\right)^{j-t} \Delta B_{j-1}. \qquad (5)\end{aligned}$$

As before, the sustainability hypothesis holds if the expected value of the second term on the right-hand side is zero. One way to test whether this is in fact the case, given that $\widetilde{G}_t$ and R_t possess unit roots, involves first estimating and then testing for cointegration in the following regression

$$R_t = \alpha + \beta \widetilde{G}_t + e_t, \qquad (6)$$

where e_t is a mean zero disturbance term. To see what this means for the test of the sustainability hypothesis, note that $\Delta B_t = \widetilde{G}_t - R_t$, which together with (5) implies that

$$\Delta B_t = (1-\beta)\widetilde{G}_t - \alpha - e_t. \qquad (7)$$

As shown by Quintos (1995), although the sustainability hypothesis holds regardless of the integratedness of ΔB_t, the speed with which this happens does not. If ΔB_t is stationary, then the second term in (5) vanishes at a much faster rate than if ΔB_t is non-stationary. In the terminology of Quintos (1995), the sustainability of the public deficit is said to be strong if ΔB_t is stationary, whereas it is said to be weak if ΔB_t is non-stationary.

In terms of the regression in (6), by looking at (7), we see that cointegration with $\beta = 1$ is enough for strong sustainability, because only then will ΔB_t be stationary. But since $\widetilde{G}_t$ is non-stationary, the sustainability is weak if either $\beta = 1$ and there is no cointegration or $0 < \beta < 1$. It follows that $0 < \beta \leq 1$ is both necessary and sufficient for sustainability, while cointegration is only sufficient. The distinction between strong and weak sustainability is made because although $0 < \beta < 1$ is enough for the sustainability hypothesis to hold, the government is now spending more than it earns, thus making the public debt difficult to market in the long run. Finally, if $\beta \leq 0$, then the deficit is no longer sustainable, because this would make ΔB_t grow faster than the growth rate of the economy, as approximated by the mean interest rate.[2]

The above discussion suggests that testing for cointegration is key in this kind of sustainability analysis, not only because of its role in determining strong sustainability but also because we cannot do inference regarding β unless revenues and expenditures are cointegrated.

The conventional way in which earlier studies have been trying to test the sustainability hypothesis within this framework involves first estimating (6) by ordinary least squares (OLS) and then testing whether the residuals from that equation can be treated as stationary or not by using any conventional cointegration test. This test is then repeated for every country in the sample, each time using only the sample information for that particular country. Studies based on this approach are generally unable to reject the null hypothesis of no cointegration, which is seen to imply that the strong form of the sustainability hypothesis should be rejected. In this paper, we argue that this result should not be taken at face value as a failure to reject the null may well be due to low power.

In fact, it is well known that tests that take no cointegration as the null hypothesis can have very little power against nearly cointegrated alternatives, especially if there are structural breaks present. In a situation like this, it is therefore essential to device tests with increased power. A natural approach

[2]Note that the case with $\beta > 0$ is not consistent with a deficit, since revenues are growing faster than expenditures.

to do this would be to combine the sample information obtained from the time series dimension with that obtained from the cross-sectional. This not only increases the power by taking the total number of observations and their variation into account, but also increases the precision of the test by effectively reducing the noise coming from the individual time series regressions. Therefore, one way to augment the power of univariate tests would be to subject the residuals from (6) to a panel data cointegration test.

Unfortunately, most existing tests of this kind rely critically on assuming that the countries are independent of each other, which is unlikely to hold in the present application because of strong intra-economy linkages. Another limitation of most panel cointegration tests is that they ignore the possibility of structural change, which is likely to be the rule rather than the exception in this kind of fiscal data. Thus, a first crucial step in testing the sustainability hypothesis using panel data is to employ tests that allow for structural breaks, and that do not rely to such a large extent on the countries being independent.

2.3 A panel cointegration test with breaks

The previous section suggests that testing for cointegration in cross-country dependent data with structural change is key in inferring the sustainability hypothesis. In this section, we outline a test that fits this description, and that is general enough to allow for both cross-country dependence and an unknown number of breaks that may be located at different dates for different countries. In so doing, we will use the index $i = 1, ..., N$ to denote countries, while t again denotes time.

Consider the following panel version of the regression in (6)

$$R_{it} = \alpha_{ij} + \beta_i \widetilde{G}_{it} + e_{it}, \quad j = 1, ..., M_i + 1, \tag{8}$$

where β_i is a country specific slope that is assumed to be constant over time, while α_{ij} is a country specific intercept that is subject to M_i structural breaks. In other words, there are $M_i + 1$ regimes for each country i with the j^{th} regime running form T_{ij-1} to T_{ij} time series observations.

Sustainability requires (8) being a cointegrated relationship. If there is no structural change, then this hypothesis can be readily tested by using the existing tests for cointegration in panel data. If there are breaks, however, then this test procedure is no longer valid since the relationship in (8) is now non-linear. This poses a serious problem for inference since conventional tests cannot be used to discriminate between cointegration with structural change and the absence of cointegration. This issue was recently addressed by Westerlund (2005), who develops a panel Lagrange multiplier test for cointegration that allows for multiple structural breaks.

The null hypothesis is formulated as that all the countries in the panel are cointegrated, while the alternative is formulated as that there is at least some country for which cointegration does not hold. The test statistic for this particular hypothesis can be written as

$$Z(M) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M_i+1}\sum_{t=T_{ij-1}+1}^{T_{ij}} S_{it}^2/(T_{ij}-T_{ij-1})^2\widehat{\sigma}_i^2,$$

where $S_{it} = \sum_{s=T_{ij-1}+1}^{t}\widehat{e}_{is}$ and $\widehat{e}_{it}$ is the regression residual obtained by using any efficient estimator of the cointegration vector such as the conventional fully modified OLS (FMOLS) estimator. Given that the countries are independent, and some technical conditions regarding the persistency of the individual time series, Westerlund (2005) shows that $Z(M)$ reaches the following sequential limit as $T \to \infty$ and then $N \to \infty$ under the null hypothesis

$$\frac{\sqrt{N}(Z(M) - E(Z(M)))}{\sqrt{\text{var}(Z(M))}} \Rightarrow N(0,1),$$

where the mean and variance adjustment terms $E(Z(M))$ and $\text{var}(Z(M))$ are defined in Westerlund (2005). Thus, by standardizing $Z(M)$ by its mean and standard deviation, we obtain a new test statistic that has an asymptotic standard normal distribution under the null hypothesis. Under the alternative hypothesis, the statistic diverges to positive infinity suggesting that the right tail of the normal distribution should be used to reject the null.

For the estimation of the number of breaks and their locations, Westerlund (2005) suggest using the Bai and Perron (2003) procedure, which can be implemented in two steps. In the first, the breakpoints are estimated by minimizing the sum of squared residuals for all permissable values of M_i. In the second, the estimated breakpoints are used together with the associated sum of squares to estimate the number of breaks using an information criterion. These steps are then repeated N times to obtain the estimated number of breaks and their locations for each country.

The fact that both, the number of breaks and their locations are treated as unknown here is a clear advantage in comparison to other approaches. For example, most studies that allow for structural change, for example Wilcox (1989) and Hakkio and Rush (1991), often maintain that the breaks are known, which means that they are likely to suffer from pre-test bias. Other studies such as Haug (1995) and Quintos (1995) allow for unknown breakpoints but restrict the number of breaks to one, which is not likely to be the case.

To handle the impact of cross-country dependence Westerlund (2005) suggests using the bootstrap approach. The particular bootstrap scheme opted for in this paper uses the sieve approach of Psaradakis (2003), who proposes a bootstrapped stationarity test for pure time series. The advantage with this scheme is that it can be modified to preserve not only the cross-country correlations but also the serial correlations.

3 Empirical results

In this section, we first briefly describe the data, and then we present the empirical unit root and cointegration test results. Finally, we present some results on the estimated cointegration vectors.

3.1 Data

We test the sustainability hypothesis by using the quarterly data that covers eight rich OECD countries over the period 1977Q1 to 2005Q4. The countries included in the panel are Canada, Finland, France, Ireland, Japan, Sweden, the United Kingdom, and the United States. The data are taken from the

OECD Economic Outlook Database Inventory, updated in February 2006, and include general government receipts and disbursement, and GDP at market prices. As in Section 2, the variables used in the analysis are defined in terms of GDP ratios.

3.2 Stationarity tests

We begin by considering the integratedness of the government revenue and expenditure series using the panel stationarity test of Carrion-i-Silvestre *et al.* (2005), which is very similar to the Westerlund (2005) cointegration test in the sense that it permits for both cross-country dependence and structural breaks. This is therefore an appropriate test for our purposes. Another similarity with the cointegration test is that the breaks are estimated using the Bai and Perron (2003) procedure and that the bootstrap of the sieve type.

In applying the Carrion-i-Silvestre *et al.* (2005) test, we follow the recommendation of Newey and West (1994) and use the Bartlett kernel, where the bandwidth parameter is permitted to grow with T at the rate $T^{1/3}$. In estimating the number of breaks, we allow for a maximum of five shifts for each country, which seem sufficient to capture most of the breaks in the data. The exact number of breaks for each country is estimated using the Schwartz information criterion. Also, to ensure that the break date estimator work properly, we set the minimum length of each regime equal to $0.1T$. The sieve bootstrap is implemented using a lag length of five and we make 1,000 replications.

As for the deterministic specification, since all variables are expressed in GDP ratios, no time trends are required. Two models are considered. The first assumes a fixed mean while the second allows for up to five breaks in the mean of each series. Among these models, since most of the literature indicates the presence of breaks, the latter stands out as the most natural one.

Table 1 shows the results of panel stationary test for each variable. For each model, the first row represents the test value, the second row contains the asymptotic p-values, and the third row contains the bootstrapped p-values.

The results from the asymptotic p-values suggest that the null hypothesis of stationary can be safely rejected. However, these values assume that the countries are independent, which is unlikely to hold. In order to account for this dependence, we use the bootstrapped p-values instead. Except for revenues in the structural break model, the conclusions are not altered by taking the cross-country correlations into account. Thus, since the overall evidence in favor of a rejection is quite overwhelming, we choose to proceed as if the variables are in fact non-stationary.

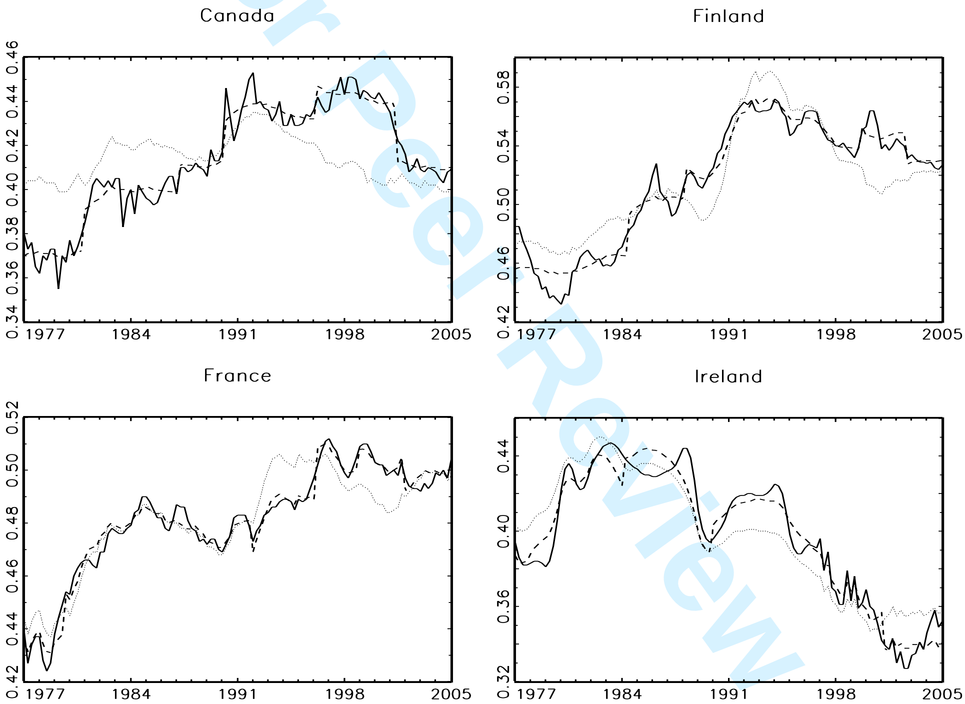
3.3 Cointegration tests

The second step in our analysis is to test whether the variables are cointegrated using the $Z(M)$ test. The results are presented in the rightmost column of Table 1. It is seen that the p-values for the model with a fixed constant provide very little support of cointegration. However, these results do not take into account the possibility of structural breaks, and are therefore prone to erroneous conclusions. Indeed, if we allow for structural shifts and cross-country correlation, the null hypothesis of cointegration cannot be rejected at the 5 percent level. We therefore conclude that the variables appear to be cointegrated around a broken intercept.

Table 2 reports the estimated breakpoints obtained from the Bai and Perron (2003) procedure. It is seen that there are at least three breaks for each country, which is indicative of structural instability. This evidence seems consistent with the gradual shift towards increased business cycles synchronization among the OECD countries, which has substantially affected the fiscal position of many governments, see Fatas and Mihov (2003). The first break period in the 1980's reflects the effect of the Great Depression, during which the synchronized recession coincided with an increase in real interest rate. This caused a substantial deterioration of the fiscal stance of of governments of rich OECD countries, as the deficits created by the spiralling interest payments could not be offset, see OECD (2002).

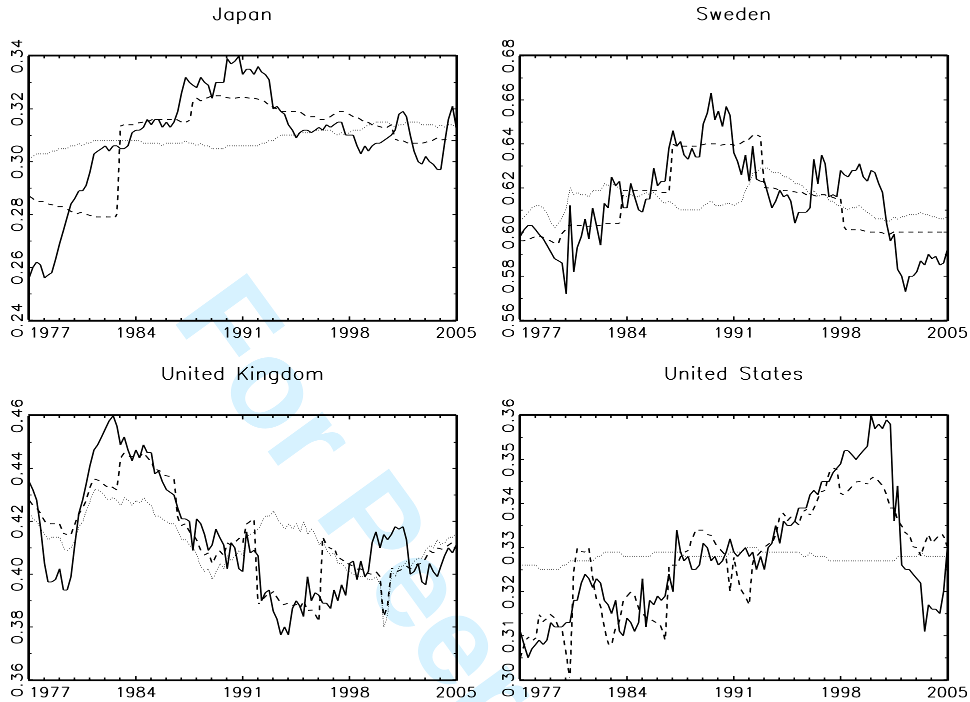
Later, following the Maastricht Treaty and the Stability and Growth Pact in the early 1990's, the fiscal stance of many OECD countries were driven primarily by an increased coordination of both fiscal and monetary policy within the European Monetary Union, see Darvas *et al.* (2005). This period is also clearly visible in the table, as indicated by the concentration of breaks in the early 1990's, and seems consistent with the results reported in many earlier studies, see for example Afonso (2005). The many breaks in the late 1990's are most likely due to the formation of the European Monetary Union in January 1999.

Figure 1: Revenues-GDP ratio with fitted trend functions.



To further illustrate the importance of accounting for structural change, Figures 1 and 2 plot revenues-GDP ratio, the dependent variable, together with the fitted trend functions for the models with and without breaks. In the figures, the solid line represents revenues measured as GDP ratio, the dashed line represents the model with breaks and the dotted line represents the

Figure 2: Revenues-GDP ratio with fitted trend functions.



model without any breaks. It can easily be seen that nearly all revenue series exhibit several level breaks, most of which seem to reflect the effect of economic fluctuations, and changes in the monetary and fiscal policy objectives in the individual countries. Interestingly, it is seen that the revenues measured as GDP ratio of most countries have been rising during the first half of the sample and falling during the second half, which reflects a more flexible response of the recent fiscal policy of most OECD governments to the economic downturn.

More importantly, we see that the trend lines for the model with breaks seem to provide a very good fit to the revenues for all eight countries. By contrast, the trend lines for the model without breaks are generally incapable of accounting for the all variation in revenues, thus leading to a poor fit. This effect is particularly striking for the United States.

3.4 Cointegration estimation

Since the variables appear to be cointegrated, it is possible to estimate and test whether the cointegrating slope is indeed equal to one as required by sustainability hypothesis.

It is well known that the OLS estimator is consistent under fairly general conditions when applied to the cointegrated regression in (7). Unfortunately, if the regressor is endogenous, then this estimator suffers from nuisance parameter dependencies even asymptotically, which makes it a poor candidate for inference. To account for this, in addition to using OLS, we employ both FMOLS and dynamic OLS (DOLS) panel estimation techniques, see Kao and Chiang (2000). All three estimators are based on pooling along the cross-sectional dimension, and are appropriate for testing the null hypothesis that $\beta_i = 1$ for all i against the alternative hypothesis that β_i is equal to some common slope value, which is different from unity.

Moreover, as when testing for stationarity and cointegration in panel data, the presence of cross-country dependence makes inference based on the asymptotic normal distribution inappropriate, in which case bootstrap inference might be better. The particular bootstrap opted for this section is taken from Chang *et al.* (2006), who propose a sieve resampling scheme that preserves the serial correlation properties of the errors, and that can be generalized along the lines of Westerlund (2006) to also accommodate cross-country dependence. As an indication of the severity of this problem, we applied the Pesaran (2004) test of cross-section correlation to the OLS residuals of the estimated cointegrated regression. The computed test value was 4.98, which, when compared to the right tail of the asymptotic normal distribution, lead to a clear rejection of the no correlation null at all conventional significance levels. Thus, cross-country correlation is indeed a problem that needs to be addressed.

The results from the OLS, FMOLS and DOLS estimators are reported in Table 3 together with the double-sided p-values for the null hypothesis of a unit slope. The first thing to notice is that the estimated slopes lie very close to their hypothesized values of one. Indeed, the range of the estimated slopes is 0.909 to 0.934. The closeness of these estimates to their expected value based on the sustainability hypothesis is supported by the p-values.

Indeed, by looking at the p-values based on the asymptotic normal distribution, we see that the null hypothesis of a unit slope cannot be rejected even at the most generous 10 percent level. The results from the bootstrapped p-values, which allow for cross-country dependence, are even more supportive of the unit slope null, and are all well above 0.6. Thus, based on this evidence we cannot reject the null hypothesis of a unit slope in the panel. Thus, in contrast to much of the earlier evidence regarding the sustainability of fiscal policy, we are unable to reject the sustainability hypothesis for the panel as a whole.

4 Concluding remarks

Earlier empirical research based on testing for stationarity of public debt and budget deficit has tended to reject the hypothesis of fiscal sustainability. As a response to this, recent work within this field has turned towards more flexible testing approaches based on cointegration. The idea is that if fiscal policy is sustainable, then government revenues and expenditures should be cointegrated with a unit slope on expenditures. Unfortunately, the results of the earlier studies have not been very convincing.

In this paper, we provide some insights suggesting that the weak empirical support of the sustainability hypothesis may be due to inadequate econometric methods. In particular, we argue that a failure of accounting for structural breaks and the poor precision of commonly applied time series tests may well result in a rejection of the sustainability hypothesis. To circumvent these problems, we resort to some recent advances in the area of non-stationary panel data. These methods are more powerful than conventional time series methods, and are general enough to allow for cross-country dependence and an unknown number of structural breaks in the cointegrating relationship.

Based on data covering eight rich OECD countries over the period from 1977Q1 to 2005Q4, we show that revenues and expenditures are non-stationary and cointegrated. We also provide evidence suggesting that the null hypothesis of a common unit slope on expenditures cannot be rejected, suggesting that the sustainability hypothesis cannot be rejected. The results on the estimated structural breaks are suggestive of strong structural instability, and are compatible with the findings of many earlier studies such as Quintos (1995), Martin (2000) and Afonso (2005). Some of the estimated breaks are related to the increased business cycle synchronization among the OECD countries, and the policy interventions resulting from the increased coordination of the monetary and fiscal policy within the European community.

Table 1: Panel stationarity and cointegration tests.

		Stationarity tests		
Specification	Test	R_{it}	$\widetilde{G}_{it}$	$Z(M)$
Fixed constant	Value	20.147	18.684	16.661
	p-value^a	0.000	0.000	0.000
	p-value^b	0.000	0.002	0.028
Constant with breaks	Value	2.192	5.114	1.416
	p-value^a	0.014	0.000	0.078
	p-value^b	0.242	0.008	0.678

Notes: All test results allow for up to five structural breaks for each state. To handle the serial correlation, the Bartlett kernel is used with a bandwidth parameter of $T^{1/3}$.

^{a}The p-value is based on the asymptotic normal distribution.

^{b}The p-value is based on the bootstrapped distribution. The number of lags in the sieve approximation is five and we make 1,000 bootstrap replications.

Table 2: Estimated breaks.

Country	No.	Breakpoint 1	2	3	4	5
Canada	5	1981Q1	1987Q2	1990Q2	1996Q2	2001Q3
Finland	4	1984Q3	1988Q2	1999Q4	2002Q4	–
France	5	1979Q4	1992Q1	1996Q2	1999Q1	2002Q1
Ireland	3	1983Q1	1987Q4	2001Q1	–	–
Japan	5	1979Q4	1983Q4	1987Q1	1993Q1	1998Q2
Sweden	5	1983Q1	1986Q4	1992Q1	1996Q2	2001Q2
United Kingdom	4	1980Q3	1986Q4	1992Q2	1998Q1	–
United States	5	1980Q3	1987Q2	1994Q2	1997Q3	2002Q1

Notes: The breaks are estimated using the Bai and Perron (2003) procedure with a maximum number of five breaks for each state. The minimum length of each break regime is set to $0.1T$.

Table 3: Panel cointegration slope estimates.

Estimator	Value	p-value^a	p-value^b
OLS	0.925	0.594	0.708
DOLS	0.934	0.689	0.718
FMOLS	0.909	0.444	0.602

Notes: All results allow for up to five structural breaks for each state. The p-values for the between and within estimates are for the null of a unit slope against the alternative that the slope is different from unity.

[a]The p-value is based on the asymptotic normal distribution.

[b]The p-value is based on the bootstrapped distribution. The number of lags in the sieve approximation is five and we make 1,000 bootstrap replications.

References

Afonso, A. (2005). Fiscal Sustainability: the unpleasant European case. *FinanzArchive* 61, 19-44.

Ahmed, S., and Rogers, J. (1995). Government budget deficits and trade deficits. Are present value constraint satisfied in long-term data? *Journal of Monetary Economics* 36, 351-374.

Baglioni, A., and Cherubini, U. (1993). Intertemporal budget constraint and public debt sustainability: the case of Italy. *Applied Economics* 25, 275-283.

Bai, J., and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics* 18, 1-22.

Bohn, H. (1991). The sustainability of budget deficit with lump-sum and with income-based taxation. *Journal of Money, Credit and Banking* 23, 581-604.

Bordo, M.D., and Heibling, T. (2003). Have national business cycles become more synchronized? NBER Working Paper No. 10130.

Breitung, J., and Pesaran, M. H. (2005). Unit roots and cointegration in panels. Forthcoming in Matyas, L., Sevestre, P. (Eds.), *The Econometrics of Panel Data*, Klüver Academic Press.

Carrion-i-Silvestre, J. L., Del Barrio-Castro, T., and López-Bazo, E. (2005). Breaking the panels: an application to the GDP per capita. *Econometrics Journal* 8, 159-175.

Chang, Y., Park, J. Y., and Song, K. (2006). Bootstrapping cointegrating regressions. *Journal of Econometrics*, forthcoming.

Darvas, Z., Rose, A.K., and Szapary, G. (2005). Fiscal divergence and business cycle synchronization: irresponsibility is idiosyncratic. NBER Working Paper No. 11580.

European Commission. (2002). Public finances in EMU. *European Economy* 3.

Fatas, A., and Mihov, I. (2003). The case of restricting fiscal policy discretion. *Quarterly Journal of Economics* 118-4, 1419-1447.

Hakkio, G., and Rush, M. (1991). Is the budget deficit too large? *Economic Inquiry* 29, 429-445.

Haug, A. (1995). Has federal budget deficit policy changed in recent years? *Economic Inquiry* 33, 104-118.

IMF (2001). Business cycle linkages among major advanced economices. *World Economic Outlook*, October 2001.

Kao, C., and Chiang, M. H. (2000). On the estimation and inference of a cointegrated regression in panel data. *Advances in Econometrics* 15, 179-222.

Kirchgaessner, G., and Prohl, S. (2006). Sustainability of Swiss fiscal policy. forthcoming in *Swiss Journal of Economics and Statistics*.

MacDonald, R. (1992). Some tests of the government's intertemporal budget constraint using U.S. data. *Applied Economics* 24, 1287-1292.

Martin, G. (2000). U.S. deficit sustainability: a new approach based on multiple endogenous breaks. *Journal of Applied Econometrics* 15, 83-105.

Newey, W., and West, K. (1994). Autocovariance lag selection in covariance matrix estimation. *Review of Economic Studies* 61, 631-653.

OECD (2002). Fiscal sustainability: the contribution of fiscal rules. *OECD Economic Outlook 72*.

Olekalns, N. (2000). Sustainability and stability? Australian fiscal policy in the 20th century. *Australian Economic Papers* 39, 138-151.

Papadopoulos, A., and Sidiropoulos, M. (1999). The sustainability of fiscal policies in the European Union. *International Advances in Economic Research* 5, 289-307.

Payne, J. (1997). International evidence on the sustainability of budget deficits. *Applied Economic Letters* 12, 775-779.

Pesaran, H. (2004). General diagnostic tests for cross section dependence in panels. Cambridge Working Paper in Economics 435.

Psaradakis, Z. (2003). A sieve bootstrap test for stationary. *Statistics and Probability Letters* 62, 263-274.

item Quintos, C. (1995). Sustainability of the deficit process with structural shifts. *Journal of Business and Economic Statistics* 13, 409-417.

Smith, G., and Zin, S. (1991). Persistent deficits and the market value of government debt. *Journal of Applied Econometrics* 6, 31-44.

Westerlund, J. (2005). Testing for panel cointegration with multiple structural breaks. *Oxford Bulletin of Economics and Statistics* 68, 101-132.

Wilcox, D. (1989). The sustainability of government deficits: implications of the present-value borrowing constraint. *Journal of Money, Credit, and Banking* 21, 291-306.