RCEA THE RIMINI CENTRE FOR ECONOMIC ANALYSIS

WP 17-09

Mark J. Holmes
Waikato University, New Zealand

Jesús Otero
Universidad del Rosario, Colombia

Theodore Panagiotidis
University of Macedonia, Greece
and
The Rimini Centre for Economic Analysis, Italy

ARE EU BUDGET DEFICITS STATIONARY?

ARE EU BUDGET DEFICITS STATIONARY? [♦]

Mark J. Holmes
Department of Economics
Waikato University
New Zealand
holmesmj@waikato.ac.nz

Jesús Otero
Facultad de Economía
Universidad del Rosario
Colombia
jotero@urosario.edu.co

Theodore Panagiotidis
Department of Economics
University of Macedonia
Greece
tpanag@uom.gr

August 2008

Abstract

In this paper, we test for the stationarity of European Union budget deficits over the period 1971 to 2006, using a panel of thirteen member countries. Our testing strategy addresses two key concerns with regard to unit root panel data testing, namely (i) the presence of cross-sectional dependence among the countries in the panel and (ii) the identification of potential structural breaks that might have occurred at different points in time. To address these concerns, we employ an AR-based bootstrap approach that allows us to test the null hypothesis of joint stationarity with endogenously determined structural breaks. In contrast to the existing literature, we find that the EU countries considered are characterised by fiscal stationarity over the full sample period irrespective of us allowing for structural breaks. This conclusion also holds when analysing sub-periods based on before and after the Maastricht treaty.

---

1. Introduction

For the European Union (EU), the size of the government budget deficit has assumed a particular importance in recent years with the Maastricht Treaty and Stability and Growth Pact making fiscal discipline an explicit criterion for membership of European Monetary Union (EMU). The original Maastricht requirement set in 1992 that governments run a budget deficit of no more than 3% of GDP as a precondition to enter EMU has, for many EU countries, implied a shift towards a more restrictive fiscal stance and with this, the possibility of adverse consequences with respect to output and unemployment. However, if the budget is out of control, economic policies at both the macro- and microeconomic levels will quickly become unsupportable, and require changes to be made.

Empirical studies of budget deficit behavior have typically fallen into one of two categories. The first category has examined the possibility of non-stationary budget deficits by conducting tests of unit roots. Evidence against the existence of unit roots has been considered as support for the *strong* form of budget deficit sustainability insofar as satisfying the intertemporal budget constraint (IBC). The results of this line of research have been mixed. Studies such as Caporale (1995), Greiner and Semmler (1999) and Vanhorebeek and Rompuy (1995) paint a varied picture for the EU. In the case of the US, Hamilton and Flavin (1986) find that the budget deficit follows a stationary stochastic process and thus is regarded as sustainable. However, Wilcox (1989), Trehan and Walsh [(1988), (1991)], and Kremers (1989) find that the budget deficit is non-stationary implying an unsustainable budgetary process.

The second group of studies examines the long-run relationship between government revenues and expenditures using cointegration methodology. The existence of

a cointegrating relationship has been considered as evidence consistent with the IBC and can be regarded as the *weak* form of budget deficit sustainability. The results of this line of research also have also been mixed. With regard to EU countries, studies such as Bravo and Silvestre (2002) and Afonso (2005) find limited evidence in favor of cointegration. As to the US, Haug (1991) finds support for the existence of cointegration and the IBC, whereas Hakkio and Rush (1991) question the existence of cointegration when the sample period is extended towards the end of the 1980s, arguing that deficit sustainability may not hold in the later part of their sample period. The lack of consensus on both these approaches has motivated a further line of research that finds stronger evidence in favor of stationarity, cointegration and sustainability when allowance is made for the existence of structural breaks (see, *inter alia*, Tanner and Liu 1994, Quintos 1995, Martin 2000, Cunado *et al*. 2004).

In this paper, we test for stationarity of the EU budget deficits using data for a panel comprising thirteen EU members. Since unit root tests applied to single series suffer from low power if cross dependence across the series exists, panel unit root techniques offer a way forward in terms of enhanced test power. In recent years a number of alternative procedures have been proposed to test for the presence of unit roots in panels that combine both time-series and cross-sectional information such that fewer time observations are required for these tests to have power. The most commonly used unit root test applied to panels include Maddala and Wu (MW) (1999), Levin, Lin and Chu (2002), Im, Pesaran and Shin (IPS) (2003), and Pesaran (2007) which test the joint null hypothesis of a unit root against the alternative of at least one stationary series in the panel. These tests are based on augmented Dickey–Fuller (ADF) (1979) statistics across the cross-sectional units of the

panel. A recent study of EU budget sustainability by Prohl and Schneider (2006) utilises a range of panel unit root tests and finds evidence in favour of sustainability.[1] However, IPS (2003, p.73) warn that due to the heterogeneous nature of the alternative hypothesis in their test, one needs to be careful when interpreting the results because the null hypothesis that there is a unit root in each cross section may be rejected when only a fraction of the series in the panel is stationary. A further issue here is that the presence of cross-sectional dependencies can undermine the asymptotic normality of the IPS test and lead to over-rejection of the null hypothesis of joint non-stationarity.

In sharp contrast to the literature on fiscal sustainability, this study examines the stationarity of EU budget deficits using the Hadri (2000) test of the null hypothesis that all individual series are stationary against the alternative of at least a single unit root in the panel. The Hadri tests offer the key advantage insofar as we may conclude that all the deficits in the panel are stationary if the joint null hypothesis is not rejected. However, in addition to this, an important feature of our analysis is that we allow for the presence of structural breaks, serial correlation, and cross-sectional dependency across the individuals in the panel. More specifically, we apply the Hadri and Rao (2008) panel stationarity test with structural breaks, which admits the possibility of different endogenously determined breaking dates across the individuals in the panel. Hadri and Rao (2008) take into account both serial correlation and cross-sectional dependency by means of the implementation of an AR-based bootstrap.

The outline of the paper is as follows. Section 2 briefly reviews the Hadri-based

---

[1] The range of panel tests are augmented by a procedure advocated by Banerjee and Carrion-i-Silvestre (2006) that tests the null of joint non-stationarity with an allowance for endogenously-determined structural breaks.

4

approaches for testing for stationarity of the budget deficit in heterogeneous panels of data allowing for the likely presence of cross section dependence. Section 3 describes the data and presents the results of the empirical analysis. We find that in contrast to earlier estimates, our bootstrap approach confirms sustainability across all members of the EU sample. Section 4 concludes.

2. Stationarity and sustainability of the budget deficit in heterogeneous panel data

Sustainability is the criterion which is usually used to evaluate whether or not fiscal policy is under control. In this context, *sustainable* fiscal policies have been judged by many in terms of whether or not the IBC holds in present value terms (see, for example, Hamilton and Flavin 1986). The IBC is based on the equality of current debt with the sum of expected future discounted primary surpluses. *Unsustainable* policies, on the other hand, are characterized by violation of the IBC implying that, at some time in the future, such policies will have to be changed, otherwise they will lead to the government becoming insolvent or to a collapse of the policy regime. This in turn would have serious implications for the credibility and functioning of EMU. It is important, therefore, to view fiscal sustainability as a long-run concept. The literature on budget deficit sustainability is primarily concerned with whether or not government's intertemporal solvency constraint is violated. This approach relies on the underlying stability of past data processes. In this paper, we focus on the time-series properties of the government budget deficit.[2]

Hadri (2000) proposes an LM procedure to test the null hypothesis that all the

---

[2] Bohn (2007) argues against the use of unit root testing as a means of validating whether or not countries are satisfying their IBCs. In our study, stationarity of the budget deficit is of interest in its own right. This is because both Maastricht and the Stability and Growth pact identify clear constraints that limit the acceptable degree of fluctuations in member countries' deficits.

individual series, y_{it}, in the panel are stationary (either around a mean or around a trend) against the alternative of at least a single unit root. The two LM tests proposed by Hadri (2000) are based on the simple average of the individual univariate Kwiatkowski, Phillips, Schmidt and Shin (1992) stationarity test (denoted by KPSS), which after a suitable standardisation follows a standard normal distribution. In a recent paper, Hadri and Rao (2008) extend the Hadri stationarity tests to test the null hypothesis of stationarity allowing for the presence of a structural break. They analyse the following four different types of models of structural break under the null hypothesis:

$$\text{Model 0:} \quad y_{it} = \alpha_i + r_{it} + \delta_i D_{it} + \varepsilon_{it}, \tag{1}$$

$$\text{Model 1:} \quad y_{it} = \alpha_i + r_{it} + \delta_i D_{it} + \beta_i t + \varepsilon_{it}, \tag{2}$$

$$\text{Model 2:} \quad y_{it} = \alpha_i + r_{it} + \beta_i t + \gamma_i DT_{it} + \varepsilon_{it}, \tag{3}$$

$$\text{Model 3:} \quad y_{it} = \alpha_i + r_{it} + \delta_i D_{it} + \beta_i t + \gamma_i DT_{it} + \varepsilon_{it} \tag{4}$$

where r_{it} is a random walk, $r_{it} = r_{it-1} + u_{it}$, and ε_{it} and u_{it} are mutually independent normal distributions. Also, ε_{it} and u_{it} are *i.i.d* across i and over t, with $E[\varepsilon_{it}] = 0$, $E[\varepsilon_{it}^2] = \sigma_{\varepsilon,i}^2 > 0$, $E[u_{it}] = 0$, $E[u_{it}^2] = \sigma_{u,i}^2 \geq 0$, $t = 1,...,T$ and $i = 1,...,N$. The dummy variables that help characterise the type of structural break are defined as:

$$D_{it} = \begin{cases} 1, & \text{if} \quad t > T_{B,i}, \\ 0 & \text{otherwise} \end{cases}$$

and

$$DT_{it} = \begin{cases} t - T_{B,i}, & \text{if} \quad t > T_{B,i}, \\ 0, & \text{otherwise} \end{cases}$$

where $T_{B,i}$ denotes the occurrence of the break, and $T_{B,i} = \omega_i T$ with $\omega_i \in (0,1)$ indicating the fraction of the break point to the whole sample period for the individual i. It should be

noted that the parameters δ_i and γ_i measure the magnitude of the break and admit the possibility of different breaking dates across the individuals in the panel. Model 0 allows for a shift in the level of a series and there is no trend. Model 1 allows for a shift in the level of a series and there is a trend. Model 2 permits a change in the slope of the series. Lastly, Model 3 permits a change in both the level and the slope of the series.[3] The null hypothesis that all the series are stationary is given by $H_0 : \sigma_{u,i}^2 = 0$, $i = 1,...,N$, while the alternative that at least one of the series is nonstationary is $H_1 : \sigma_{u,i}^2 > 0$, $i = 1,...,N_1$ and $\sigma_{u,i}^2 = 0$, $i = N_1 + 1,...,N$.

The testing procedure put forward by Hadri and Rao (2008) starts off by determining an unknown break point endogenously. To do this, their suggested approach involves estimating for each individual in the panel and for each model, the break date $\hat{T}_{B,i,k}$. This is achieved by minimising the residual sum of squares (RSS) from the relevant regression under the null hypothesis, with $i = 1,...,N$ cross section units and $k = 0,1,2,3$ indicating the four models postulated in equations (1) to (4). Then, for each individual in the panel the break-type model is chosen by minimising the Schwarz Information Criterion.

Let $\hat{\varepsilon}_{it}$ be the residuals obtained from the estimation of the chosen break-type model. The individual univariate KPSS stationarity test where structural breaks are taken into account is given by:

---

[3] In their study of GDP per capita, Carrion-i-Silvestre et al. (2005) analyse two of the models considered by Hadri and Rao (2008), namely the model with breaks in the level and no time trend, and the model with breaks in the level and in the time trend.

$$\eta_{i,T,k}\left(\hat{\omega}_i\right) = \frac{\sum_{t=1}^{T} S_{it}^2}{T^2 \hat{\sigma}_{\varepsilon_i}^2},$$

where S_{it} denotes the partial sum process of the residuals given by $S_{it} = \sum_{j=1}^{t} \hat{\varepsilon}_{ij}$, and $\hat{\sigma}_{\varepsilon_i}^2$ is a consistent estimator of the long-run variance of $\hat{\varepsilon}_{it}$ from the appropriate regression. In their original paper, KPSS propose a nonparametric estimator of $\hat{\sigma}_{\varepsilon_i}^2$ based on a Bartlett window having a truncation lag parameter of $l_q = \text{integer}\left[q\left(T/100\right)^{1/4}\right]$, with $q = 4,12$ (the value of the test statistics appears sensitive to the choice of q). However, in a subsequent paper Caner and Kilian (2001) pointed out that stationarity tests, like the KPSS tests, exhibit very low power after correcting for size distortions. Thus, in our paper we follow recent work by Sul, Phillips and Choi (2005), who propose a new boundary condition rule to obtain a consistent estimate of the long-run variance $\hat{\sigma}_{\varepsilon_i}^2$, that improves the size and power properties of the KPSS stationarity tests. In particular, Sul et al. suggest the following procedure. First, an AR model for the residuals is estimated, that is:

$$\hat{\varepsilon}_{it} = \rho_{i,1}\hat{\varepsilon}_{i,t-1} + ... + \rho_{i,p_i}\hat{\varepsilon}_{i,t-p_i} + \upsilon_{it} \tag{5}$$

where the lag length of the autoregression can be determined for example using the Schwarz Information Criterion (SIC), or applying the GEneral-To-Specific (GETS) algorithm proposed by Hall (1994) and Campbell and Perron (1991). Second, the long-run variance estimate of $\hat{\sigma}_{\varepsilon_i}^2$ is obtained with the boundary condition rule:

$$\hat{\sigma}_{\varepsilon_i}^2 = \min\left\{T\hat{\sigma}_{\upsilon_i}^2, \frac{\hat{\sigma}_{\upsilon_i}^2}{\left(1 - \hat{\rho}_i\left(1\right)\right)^2}\right\},$$

where $\hat{\rho}_i(1) = \hat{\rho}_{i,1}(1) + ... + \hat{\rho}_{i,p_i}(1)$ denotes the autoregressive polynomial evaluated at $L = 1$. In turn, $\hat{\sigma}_{\upsilon_i}^2$ is the long-run variance estimate of the residuals in equation (5) that is obtained using a quadratic spectral window Heteroscedastic and Autocorrelation Consistent (HAC) estimator.[4]

The Hadri and Rao (2008) panel stationarity test statistic takes structural breaks into account through a simple average of individual univariate KPSS stationarity tests:

$$\widehat{LM}_{T,N,k}(\hat{\omega}_i) = \frac{1}{N}\sum_{i=1}^{N}\eta_{i,T,k}(\hat{\omega}_i),$$

After a suitable standardisation, using appropriate moments that are functions of the break fraction parameter $\hat{\omega}_i$ (see Theorem 3, in Hadri and Rao (2008)), this follows a standard normal limiting distribution, that is:

$$Z_k(\hat{\omega}_i) = \frac{\sqrt{N}\left(\widehat{LM}_{T,N,k}(\hat{\omega}_i) - \overline{\xi}_k\right)}{\overline{\zeta}_k} \Rightarrow N(0,1) \qquad (6)$$

where $\overline{\xi}_k = \frac{1}{N}\sum_{i=1}^{N}\xi_{i,k}$ and $\overline{\zeta}_k^2 = \frac{1}{N}\sum_{i=1}^{N}\zeta_{i,k}^2$.

An important assumption underlying the Hadri and Rao (2008) test is that of cross section independence among the individual time series in the panel.[5] To allow for the presence of cross-sectional dependency, these authors recommend implementing the following AR bootstrap method. To begin with, we correct for serial correlation using

---

[4] Additional Monte Carlo evidence reported by Carrion-i-Silvestre and Sansó (2006) also indicates that the proposal in Sul et al. (2005) is to be preferred since the KPSS statistics exhibit less size distortion and reasonable power.

[5] Giulietti et al. (2008) examine the effect of cross sectional dependency in the Hadri (2000) panel stationarity tests in the absence of structural breaks and with no serial correlation. They find that even for relatively large T and N the Hadri (2000) tests suffer from severe size distortions, the magnitude of which increases as the strength of the cross-sectional dependence increases. To correct the size distortion caused by cross-sectional dependence, Giulietti et al. (2008) apply the bootstrap method and find that the bootstrap Hadri tests are approximately correctly sized.

equation (5) and obtain $\hat{\upsilon}_{it}$, which are centred around zero. Next, as in Maddala and Wu (1999), the residuals $\hat{\upsilon}_{it}$ are resampled with replacement with the cross-section index fixed, so that their cross-correlation structure is preserved; the resulting bootstrap innovation $\hat{\upsilon}_{it}$ is denoted $\hat{\upsilon}_{it}^*$. Then, $\hat{\varepsilon}_{it}^*$ is generated recursively as:

$$\hat{\varepsilon}_{it}^* = \hat{\rho}_{i,1}\hat{\varepsilon}_{i,t-1}^* + ... + \hat{\rho}_{i,p_i}\hat{\varepsilon}_{i,t-p_i}^* + \upsilon_{it}^*,$$

where, in order to ensure that initialisation of $\hat{\varepsilon}_{it}^*$ becomes unimportant, a large number of $\hat{\varepsilon}_{it}^*$ are generated, let us say $T+Q$ values and then the first Q values are discarded (see Chang, 2004). For our purposes we choose $Q = 30$. Lastly, the bootstrap samples of y_{it}^* are calculated by adding $\hat{\varepsilon}_{it}^*$ to the deterministic component of the corresponding model, and the Hadri LM statistic is calculated for each y_{it}^*. The results shown in Table 1 are based on 5,000 bootstrap replications used to derive the empirical distribution of the LM statistic.

3. Data and empirical analysis

We examine the sustainability of the budget deficit for a panel of thirteen EU countries over the study period 1971-2006. These are Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Spain, Sweden and the United Kingdom[6]. Annual data for each of these countries are taken from the *European Commission AMECO* (Annual Macroeconomic Data) database[7]. In all cases, fiscal deficits are expressed as a

---

[6] From the sample of countries we are able to consider, Belgium France, Germany, Italy and the Netherlands are the founding states, Denmark, Ireland and the UK joined in 1973, Greece in 1981, Spain in 1986 and Austria, Finland and Sweden in 1995.

[7] This range of countries is dictated by the availability of consistent data with respect to the study period. For this reason, other long-standing members such as Luxembourg and Portugal are excluded from the sample.

proportion of GDP. The data exclude seigniorage revenues whereby countries use money finance to fund a budget deficit. This exclusion is justified on the grounds that ten countries from the sample have proceeded towards adopting the Euro as their currency and no longer have the ability to pursue an independent monetary policy.[8] Since money finance is no longer an option for these countries, it therefore seems appropriate to judge sustainability using measures of the budget deficit that exclude money financing. A second issue concerns data availability and German unification in July 1990. Data for West Germany only can be obtained for 1971-1990, while data for Germany only can be obtained for 1991-2006. A final issue concerns the Maastricht treaty which was signed in February 1992, but negotiations were completed in 1991.

Our empirical analysis of fiscal stationarity begins by considering two sub-samples, namely from 1971 to 1990 and from 1991 to 2006. This allows us to examine the possibility of differences in pre- and post-Maastricht behaviour as well as the effect of the German series due to unification. Table 1 presents the results of applying the KPSS stationarity test to the budget deficits of the countries listed above, based on the model with intercept only.[9] As indicated in the previous section, the long-run variance required to calculate the KPSS statistic is consistently estimated using the new boundary condition rule put forward by Sul, Phillips and Choi (2005). Furthermore, to correct for possible serial correlation the autoregressive processes in (5) are estimated for up to $p = 4$ lags, and the optimal number of lags is chosen based on the SIC and the GETS algorithm. This algorithm involves testing whether the last autoregressive coefficient is statistically different from zero (say, at the 10% significance level); if it is not statistically significant,

---

[8] Denmark, Sweden and the UK are currently not members of the single currency.
[9] Qualitatively similar findings are obtained when using the model with intercept and trend. In the interests of brevity, these results are not reported here.

then the order of the autoregression is reduced by one until the last coefficient is statistically significant. Both criteria tend to pick up the same optimal lag length, although when they do not coincide the SIC favours a more parsimonious specification than GETS. Focusing first on the results of the pre-Maastricht period when GETS is used to select the lag length, the null hypothesis of stationarity is rejected at the 10% significance level for two countries, and for one more country rejection is at the 5% level. Turning to the post-Maastricht period, the null hypothesis of stationarity is rejected for seven countries: for two countries rejection is at the 10%, for three countries at the 5%, and for two more countries at the 1% level. Lastly, when one considers the full sample period, stationarity is rejected for three countries at the 10% level of significance. A similar picture emerges when inspecting the results based on SIC lag length selection, thereby not providing a clear indication of sustainability.

We now consider the Hadri panel stationarity test to the budget deficit series. The main motivation for testing stationarity in a panel of data instead of individual time series is that it has been noted that the power of the tests increases with the number of cross-sections in the panel. However, failure to account for potential cross section dependence will result in severe size distortion of the Hadri test statistics. We therefore apply an AR-based bootstrap to the Hadri tests as outlined in the previous section. The resulting Hadri test statistics are reported in the bottom row of Table 1, along with their corresponding bootstrap p-values in brackets, which in turn are based on 5,000 replications used to derive the empirical distributions of the test statistics. The results indicate that the null hypothesis of panel stationarity is not rejected (at the 10% significance level) independently of the sample period considered.

Thus far, the analysis has implicitly assumed the presence of a known structural break at the time of the Maastricht treaty and the German unification. However, it may be easily argued that there is no guarantee that a structural break, if present, occurred for all countries at precisely the same time (for example, the process towards the single currency might have been more influential for some countries for instance). Hence, in what follows we apply the Hadri and Rao (2008) procedure outlined in the previous sections. This allows us to test for panel stationarity in the presence of an unknown break point that is endogenously determined for each individual country in the panel, and for the four models postulated in equations (1) to (4).[10] The results reported in the third column on Table 2 indicate that other breaks may in fact be more important than those associated with the Maastricht treaty. Indeed, while Finland, France, Greece, Italy and Spain feature breaks dated in the 1990s, we observe that half of the Euro members included in our sample are characterised by structural breaks that pre-date the 1992 Maastricht Treaty.

The residuals from the chosen break-type model for each country are then used to construct the KPSS statistics and these, in turn, are used as in equation (6) to compute the Hadri and Rao (2008) panel stationarity test, in the presence of an unknown structural break and allowing for cross section dependence. As can be seen from the bottom row in Table 2, we are unable to reject the joint null hypothesis of panel stationarity, independently of the method used to select the optimal lag length of the autoregressive processes in (5). It is worth noting that if we wrongly assume cross-sectional independence among the countries in the panel, and use the standard normal distribution for the purposes of inference, then the null hypothesis of panel stationarity is clearly rejected when using the

---

[10] The empirical results reported in the paper were implemented using the computer software RATS and are based on a GAUSS code which was kindly provided to the authors by Yao Rao.

GETS procedure. This finding highlights the importance of allowing for the possibility of potential cross-sectional dependencies among the individual countries in the panel.

4. Concluding remarks

This paper applies the Hadri and Rao (2008) test for panel stationarity to examine evidence on budget deficits stationarity and sustainability for thirteen EU countries. In contrast to standard panel unit root tests, the Hadri test employs the null hypothesis of joint stationarity. The standard tests are of a joint non-stationary null, the rejection of which may be attributable to the stationary behaviour of as little as one panel member. This study also addresses problems associated with structural breaks in the data as well as cross-sectional dependence among panel members through pursuing an AR bootstrap approach to the Hadri and Rao tests. Using these tests that are more powerful than previous tests of the null of stationarity, we have failed to find evidence that any of the budget deficits of the EU countries are nostationary. This applies to the full sample of EU countries irrespective of whether or not membership of the single currency is present. An implication of this is that long-run fiscal discipline may not necessarily be restricted to Euro membership. This finding is unaffected if we estimate for sub-periods based on a known structural break or allow for endogenously-determined breaks across the sample.

References

Afonso A (2005) Fiscal sustainability: The unpleasant European case. Finanz Archiv 61: 19-44.

Banerjee A, Carrion-i-Silvestre JL (2006) Cointegration in panel data with breaks and cross-section dependence. Economics Working Papers ECO2006/5, European University Institute.

Bohn H (2007) Are stationary and cointegration restrictions really necessary for the temporal budget constraint?. Journal of Monetary Economics, 54: 1837-47.

Bravo ABS, Silvestre AL (2002) Intertemporal sustainability of fiscal policies: some tests for European countries. European Journal of Political Economy 18: 517-528.

Campbell J, Perron P (1991) Pitfalls and opportunities: what macroeconomists should know about unit roots. In: Blanchard O, Fischer S (eds) NBER Economics Annual 6. MIT Press, Cambridge MA, pp. 141-201.

Caner M, Kilian L (2001) Size distortions of tests of the null hypothesis of stationarity: Evidence and implications for the PPP debate. Journal of International Money and Finance 20: 639-657.

Carrion-i-Silvestre JL, Sansó A (2006) A guide to the computation of stationarity tests. Empirical Economics 31: 433-448.

Carrion-i-Silvestre JL, Del Barrio T, López-Bazo E (2005) Breaking the panels: An application to the GDP per capita. Econometrics Journal 8: 159-175.

Caporale GM (1995) Bubble finance and debt sustainability: A test of the government's intertemporal budget constraint. Applied Economics 27: 1135-1143.

Chang Y (2004). Bootstrap unit root tests in panels with cross-sectional dependency.

Journal of Econometrics 120: 263-293.

Cunado J, Gil-Alana LA, Gracia FP (2004) Is the US fiscal deficit sustainable? A fractionally integrated approach. Journal of Economics and Business 56: 501-526.

Dickey DA, Fuller WA (1979) Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association 74: 427-431.

Giulietti M, Otero J, Smith J (2008) Testing for stationarity in heterogeneous panel data in the presence of cross section dependence. Journal of Statistical Computation and Simulation, forthcoming.

Greiner A, Semmler W (1999). An inquiry into the sustainability of German fiscal policy: Some Time-Series Tests. Public Finance Review 27: 220-236.

Hadri K (2000) Testing for stationarity in heterogeneous panels. The Econometrics Journal 3: 148-161.

Hadri K, Rao Y (2008) Panel stationarity test with structural break. Oxford Bulleting of Economics and Statistics 70: 245-269.

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

Hall A (1994) Testing for a unit root in time series with pretest data-based model selection. Journal of Business and Economic Statistics 12: 461-470.

Hamilton JD, Flavin MA (1986). On the limitations of government borrowing: A framework for testing. American Economic Review 76: 808-819.

Haug AA (1991) Cointegration and government borrowing constraints: Evidence for the United States. Journal of Business and Economic Statistics 9: 97-101.

Im K, Pesaran MH, Shin Y (2003). Testing for unit roots in heterogeneous panels. Journal of Econometrics 115: 53-74.

16

Kremers JJM (1989) US federal indebtedness and the conduct of fiscal policy. Journal of Monetary Economics 23: 219-238.

Kwiatkowski D, Phillips PCB, Schmidt P, Shin Y (1992) Testing the null hypothesis of stationarity against the alternative of a unit root. Journal of Econometrics 54: 159-178.

Levin A, Lin C-F, Chu C-S (2002) Unit root tests in panel data:asymptotic and Finite-sample properties. Journal of Econometrics 108: 1-24.

Maddala GS, Wu S (1999) A comparative study of unit root tests with panel data and a new simple test. Oxford Bulletin of Economics and Statistics 61: 631-652.

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

Pesaran MH (2007) A simple panel unit root test in the presence of cross section dependence. Journal of Applied Econometrics 22: 265-312.

Prohl S, Schneider F (2006) Sustainability of public debt and budget deficit: Panel cointegration analysis for the European Union member countries. Department of Economics, Johannes Kepler University of Linz, Working Paper No. 0610.

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

Sephton PS (1995) Response surface estimates of the KPSS stationarity test. Economics Letters 47: 255-261.

Sul D, Phillips PCB, Choi CY (2005) Prewhitening bias in HAC estimation. Oxford Bulleting of Economics and Statistics 67: 517-546.

Tanner E, Liu P (1994) Is the budget deficit too large? Some further evidence. Economic Inquiry 32: 511-518.

Trehan B, Walsh C (1988) Common trends, the government budget constraint and revenue smoothing. Journal of Economic Dynamics and Control 12: 425-444.

Trehan B, Walsh C (1991) Testing intertemporal budget constraints: Theory and applications to US federal budget and current account deficits. Journal of Money, Credit and Banking 23: 206–223.

Vanhorebeek F, Rompuy PV (1995) Solvency and sustainability of fiscal policies in the EU. De Economist 143: 457-473.

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

Table 1. Panel stationarity test with a known structural break (model with constant)

Lag length based on SIC

Countries	1971 – 1990		1991 – 2006		1971 – 2006	
	Lag	Statistic	Lag	Statistic	Lag	Statistic
Austria	1	0.381*	1	0.132	1	0.185
Belgium	1	0.152	4	22.61***	1	0.236
Denmark	2	0.260	2	0.565**	2	0.176
Finland	1	0.114	1	0.213	2	0.153
France	1	0.358	2	0.622**	1	0.185
Germany	1	0.139	1	0.146	2	0.124
Greece	1	0.251	1	0.180	1	0.150
Ireland	2	0.165	1	0.167	1	0.459*
Italy	1	0.464*	4	2.248***	4	0.336
Netherlands	1	0.176	2	0.615**	1	0.099
Spain	1	0.283	1	0.354	1	0.066
Sweden	1	0.073	2	0.458*	1	0.035
United Kingdom	1	0.114	2	0.530**	3	0.208
Hadri test	1.301 [0.575]		53.618 [0.138]		0.357 [0.748]	

Lag length based on GETS

Countries	1971 – 1990		1991 – 2006		1971 – 2006	
	Lag	Statistic	Lag	Statistic	Lag	Statistic
Austria	1	0.381*	1	0.132	1	0.185
Belgium	3	0.281	4	22.61***	3	0.367*
Denmark	2	0.260	1	0.269	2	0.176
Finland	1	0.114	3	0.377*	2	0.153
France	1	0.358	2	0.622**	1	0.185
Germany	1	0.139	1	0.146	2	0.124
Greece	1	0.251	1	0.180	1	0.150
Ireland	1	0.048	1	0.167	1	0.459*
Italy	1	0.464*	4	2.248***	4	0.336
Netherlands	1	0.176	2	0.615**	1	0.099
Spain	1	0.283	1	0.354	1	0.066
Sweden	4	0.552**	2	0.458*	1	0.035
United Kingdom	4	0.239	2	0.530**	4	0.380*
Hadri test	2.526 [0.712]		53.351 [0.140]		0.940 [0.749]	

*, ** and *** indicate 10, 5 and 1% levels of significance, respectively, based on finite sample critical values calculated from the response surfaces in Sephton (1995). The long-run variance required to calculate the KPSS statistic is consistently estimated using the new boundary condition rule put forward by Sul, Phillips and Choi (2005). The bootstrap p–values of the Hadri test appear in brackets, and are based on 5,000 replications.

Table 2. Panel stationarity test with endogenously determined structural break

Countries	Model	Break date	Lag length based on:			
			SIC		GETS	
			p	KPSS statistic	p	KPSS statistic
Austria	1	1975	1	0.076	1	0.076
Belgium	3	1981	1	0.032	4	0.153
Denmark	3	1984	2	0.067	2	0.067
Finland	3	1992	1	0.028	1	0.028
France	1	1997	1	0.116	2	0.056
Germany	0	1974	1	0.083	1	0.083
Greece	3	1994	1	0.117	2	0.091
Ireland	3	1988	1	0.042	1	0.042
Italy	1	1997	1	0.096	1	0.096
Netherlands	1	1980	1	0.048	1	0.048
Spain	2	1994	1	0.029	1	0.029
Sweden	3	1992	1	0.021	4	0.210
United Kingdom	3	1998	1	0.030	4	0.138
Hadri statistic				1.043 [0.741]		3.385 [0.632]

The bootstrap p-values reported are reported in brackets, and are calculated using 5,000 replications.