Testing for Fiscal Sustainability within the European Union: New Evidence Based on Panel Data*

Joakim Westerlund† and Silika Prohl‡

March 12, 2007

Abstract

Most empirical evidence suggests that the sustainability hypothesis, stating that government revenues and expenditures should cointegrate with a unit slope on expenditures, does not hold within the European Union, a finding at odds with many theoretical models. This paper argues that these results can be attributed in part to the poor precision of univariate techniques, and that the use of panel data can generate more precise tests. By using newly devised panel unit root and cointegration tests it is shown that the sustainability hypothesis cannot be rejected when applied to a panel of annual data covering 15 European countries between 1970 and 2004.

JEL Classification: C22; C23; H60.
Keywords: Fiscal Sustainability; Panel Unit Root; Panel Cointegration.

1 Introduction

The large fiscal deficits experienced by many European countries during the last decade have caused raising concerns regarding the sustainability of fiscal policy within the European Union (EU). As a response to this, the Maastricht Treaty and the Stability and Growth Pact have been put in place with the hope of maintaining and enforcing the fiscal discipline in the EU area. The pact states that the annual budget deficit cannot be higher than 3% of GDP, and that 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.

national debt must be lower than 60% of GDP or approaching that value. To ensure that the pact is respected, offending members are fined.

A well-developed body of literature on the sustainability of fiscal policy already exists. One line of this literature states that, because of borrowing constraints, the government faces the problem of balancing its budget intertemporally, which means that the discounted value of public debt must go to zero in the long run. This is how the sustainability hypothesis is usually formulated in the empirical literature, and it will be employed in this paper as well. The intuition is that if the public debt goes to zero, the intertemporal budget constraint of the government ensures that the current market value of public debt is equal to the discounted sum of all future budget surpluses.

Earlier empirical studies that have attempted to test this hypothesis have done so by looking at the time series behavior of public deficit and debt, which should both be mean reverting, or stationary, for the sustainability hypothesis to hold. Strangely enough, however, the results obtained by using this method has not been very supportive of the sustainability hypothesis. In fact, most evidence indicate that the public deficit and debt of many European countries are non-stationary, and hence that the fiscal policy of these governments cannot be sustained in the long run. For example, Vanhorebeek and Rompuy (1995) applied this framework to a sample covering eight European countries between 1970 and 1994, and did not find any evidence of sustainability. Similarly, Caporale (1995) finds that in only six out of the 10 EU member countries in his sample have fiscal policy been sustainable.

Due to these disappointing results, more recent research has turned away from the stationarity approach and towards more flexible econometric tests based on cointegration. Within this framework, fiscal policy is said to be strongly sustainable if government revenues and expenditures are cointegrated with a unit slope coefficient on expenditures, and it is said to be weakly sustainable if the slope is less than unity, see Quintos (1995). Despite its greater flexibility, however, the results obtained by using this approach has been as disappointing as those obtained earlier, see for example Papadopoulus *et al.* (1999), Bravo and Silvestre (2002) and Afonso (2005).

In this paper, we argue that most of the previous studies in the literature are flawed in at least two respects, and that this may well at least partially explain why the sustainability hypothesis has been so hard to evidence.

One problem is that most studies employ methods that are designed to test the null hypothesis of no cointegration, or a unit root in case the stationarity of the public deficit or debt is tested, and these tests are known to suffer from low power when the equilibrium errors are highly persistent under the alternative of cointegration. Thus, low power in the tests could be one explanation for why cointegration has been difficult to find.

The second problem is that most studies employ a country-by-country approach, in which conventional cointegration tests are applied to each country separately. Although this makes the results comparable across countries, it does not really bring any more information into the analysis, since the information contained in the cross-sectional dimension is essentially disregarded.

These problems suggest that what is really needed here is a testing approach that takes full account of the panel structure of the data. Such a strategy is likely to be more powerful, and thus more successful in detecting deviations from the no cointegration null. The current paper makes an attempt in this direction by using a panel testing approach, drawing upon a panel of 15 EU member countries between 1970 and 2004. In so doing, we pay special attention to the many features that characterize this type of fiscal data. For example, given the high degree of fiscal syncretization that exist within the EU, dependence across countries is likely to be the rule rather than the exception. The results suggest that, once these features have been taken into account, the sustainability hypothesis cannot be rejected.

The rest of this paper is organized as follows. Section 2 describes the theoretical model, while Section 3 reports the empirical results. Section 4 concludes.

2 The sustainability hypothesis

This section gives a brief account of the sustainability hypothesis, and the panel cointegration approach that we will use to test it.

2.1 The theoretical model within the panel framework

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

$$G_{it} + (1 + i_{it})B_{it-1} \;=\; R_{it} + B_{it}, \tag{1}$$

where B_{it} is the current stock of public debt, i_{it} is the nominal interest rate payable on the public debt, R_{it} is government revenues including revenue from seignorage, and G_{it} is government expenditures excluding interest payments. What this equation says is that the public debt should either be paid off or refinanced by issuing new debt. In order to account for the size of the economy, we further rewrite (1) in terms of GDP ratios as

$$g_{it} + (1 + \rho_{it})b_{it-1} \;=\; r_{it} + b_{it}, \tag{2}$$

where we have used lower-case letters to denote the ratios of the corresponding upper-case variables to GDP, and where $\rho_{it} = (i_{it} - \sigma_{it})/(1 + \sigma_{it})$ is the interest rate i_{it} adjusted for the GDP growth rate, here denoted by σ_{it}.

Equation (2) is a non-linear differential equation in b_{it}, which is assumed to be stable.[1] If we further assume that $E_t(\rho_{it})$, the expected value of ρ_{it}

[1]Formally, (2) is stable if $\rho_{it} < 0$ for all i and t.

conditional on the information available at time t, is constant through time, then (2) can be solved forwards to obtain

$$b_{it} = \sum_{j=0}^{\infty} E_t \left(\frac{1}{1+\rho_i}\right)^{j+1} (r_{it+j} - g_{it+j}) + E_t \left(\frac{1}{1+\rho_i}\right)^{j+1} b_{it+j}. \quad (3)$$

This condition is crucial for testing the sustainability of fiscal policy. The first part on the right-hand side embodies the monetary and fiscal policy objectives of the government, which do not impair the ability of the government to conduct its policies in a sustainable way. Hence, this part is relatively unimportant for our purposes. The second part, on the other hand, is absolutely crucial, and determines whether sustainability hypothesis holds or not. Sustainability requires that the following transversality condition holds

$$\lim_{j\to\infty} E_t \left(\frac{1}{1+\rho_i}\right)^{j+1} b_{it+j} = 0. \quad (4)$$

This condition says that the public debt can grow no faster than the economy itself. If (4) holds, then it is clear that (3) reduces to

$$b_{it} = \sum_{j=0}^{\infty} E_t \left(\frac{1}{1+\rho_i}\right)^{j+1} (r_{it+j} - g_{it+j}).$$

In other words, the fiscal policy is sustainable if the sum of all discounted future primary surpluses is enough to offset the market value of public debt.

2.2 An empirical panel cointegration test

It is useful to rewrite (3) in its first differenced form as

$$\begin{aligned} g^r_{it} - r_{it} &= \sum_{j=0}^{\infty} E_t \left(\frac{1}{1+\rho_i}\right)^{j-1} (\Delta r_{it+j} - \Delta c_{it+j}) \\ &+ \lim_{j\to\infty} E_t \left(\frac{1}{1+\rho_i}\right)^{j+1} \Delta b_{it+j}, \end{aligned} \quad (5)$$

where $g^r_{it} = g_{it} + \rho_{it} b_{it-1}$ are the total expenditures of the government including interest payments, while $c_{it} = g_{it} + (\rho_{it} - \rho_i) b_{it-1}$ is the same variable with the discount factor taken around a nonzero mean. In this formulation, the transversality condition in (4) can be written as

$$\lim_{j\to\infty} E_t \left(\frac{1}{1+\rho_i}\right)^{j+1} \Delta b_{it+j} = 0. \quad (6)$$

Our objective is to test if this condition holds, which can be done using the following panel regression

$$r_{it} = \alpha_i + \beta_i g^r_{it} + e_{it}, \quad (7)$$

where α_i and β_i are country specific intercept and slope parameters, respectively, while e_{it} is a mean zero disturbance term. Because all variables appearing on the right-hand side of (5) are stationary by assumption, e_{it} should be stationary. In the terminology of Quintos (1995), the fiscal policy is said to be strongly sustainable if $\beta_i = 1$, whereas it is said to be weakly sustainable if $0 < \beta_i < 1$. The intuition follows by rewriting (7) as

$$g_{it}^r - r_{it} \;=\; (1 - \beta_i)g_{it}^r - \alpha_i - e_{it}. \tag{8}$$

This shows that strong sustainability implies, and is implied by, the transversality condition in (6), as it ensures that the right-hand side of (8) is stationary. By contrast, the sustainability is weak if $0 < \beta_i < 1$, regardless of whether there is cointegration or not. The relevance of this latter form is that although the right-hand side of (8) is now non-stationary, as shown by Quintos (1995), it is still enough to ensure that (6) holds. However, the speed with which this happens is reduced so that the government is forced to continuously issuing new debt in order to balance its budget intertemporally. It follows that unless the government is willing to take steps to improve the fiscal stance, it will eventually run into problems with marketing its debt.[2]

The conventional way in which researchers have been implementing this test involves first estimating (7) for each country in the sample, and then testing whether the residuals from these regressions can be characterized as stationary or not. The unit slope hypothesis is usually tested using a simple t-test. This exercise is then repeated N times, which yields one test for each country in the sample.

However, studies based on this approach are generally unable to find any evidence of sustainability, which has been taken to imply that the fiscal policy of most EU countries is unsustainable, see for example Papadopoulus *et al.* (1999), Bravo and Silvestre (2002) and Afonso (2005). But this does not seem very realistic, and this study therefore offers an alternative interpretation of these results. In particular, we argue that the weak evidence in favor of the sustainability hypothesis may not reflect the actual data generating process but rather the poor performance of the tests that have been employed to test it. The intuition is simple. Since the country-by-country approach only uses information for one country at a time, the information contained in the rest of cross-sectional dimension is essentially disregarded.

Moreover, this waste of information is likely to have become more severe in recent years, as the degree of intra-EU collaboration and synchronization has increased. There is also the issue of overlapping policy shocks that affect the fiscal positions of all member countries, thus adding to the dependence across the EU. Yet another reason for why accounting for the cross-sectional

[2]Note that the deficit is no longer sustainable if $\beta_i \leq 0$, as this would make the deficit grow faster than the economy as a whole. Similarly, the deficit is automatically sustainable if $\beta_i > 1$, as revenues are now growing faster than expenditures.

dimension is likely to be important is that the time series dimension is usually relatively short. In other word, the cross-section can be regarded as a source of complementary information.

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, which are then used for forecasting government revenues and expenditures.

3.1 Data

The data is annual and covers 15 EU member countries over the period 1970 to 2004. The countries included are Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlandxs, Portugal, Spain, Sweden and the United Kingdom. The data are taken from the European Commission annual macroeconomic database AMECO, last updated in December 2005, and include for each country general government revenues and expenditures, and GDP.

3.2 Preliminary data analysis

In order to get a feeling of the behavior of the fiscal debt situation among the 15 EU countries in our sample, we begin with a graphical inspection of the calculated deficit-GDP ratios, which can be expressed in terms of our panel regression framework as

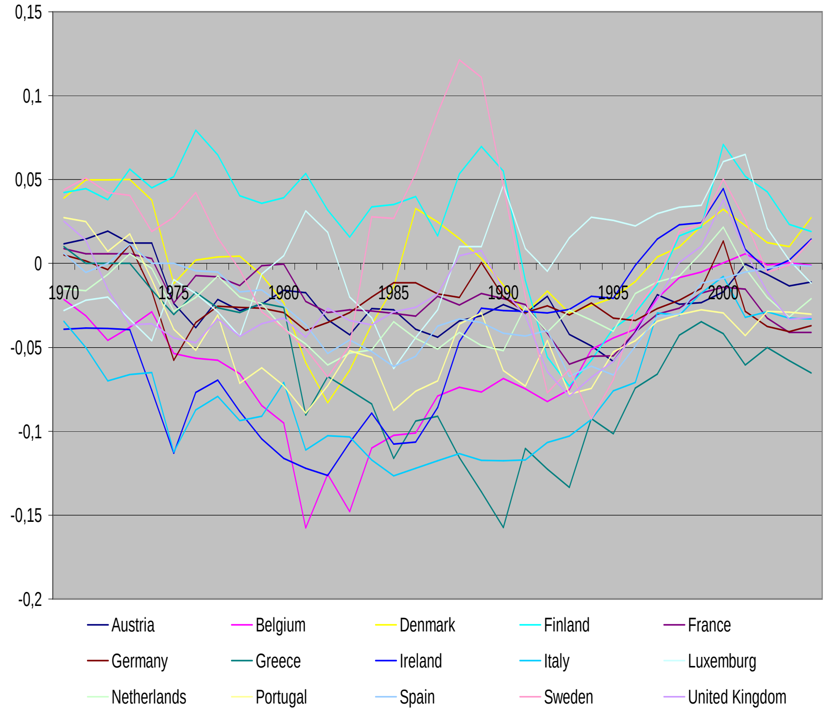
$$r_{it} - g_{it}^{r} \;=\; \alpha_i + e_{it},$$

where we have imposed the unit restriction on β_i. Thus, if the sustainability hypothesis holds, each of the deficit-GDP ratios should be stationary around some mean value, which is here denoted by α_i. Interestingly, this is exactly what we find if we look at the data in Figure 1. In fact, all series seem to be highly mean reverting, which is a strong indication not only of cointegration but also of a unit slope, as otherwise the deficit-GDP ratios should look non-stationary.

In other words, if sustainability is to be interpreted as the mean reversion of the public deficit, then Figure 1 gives a quite strong indication of sustainability within the EU. The fact that the mean of the individual series seem to differ suggests that our regression specification with an individual specific intercept is appropriate.

Another interesting observation is that, except possibly for Sweden which has a bump in the middle of the sample, there are no apparent breaks in the series. Note also how the degree of mean reversion has increased markedly during the

Figure 1: Deficit-GDP ratios.



second half of the sample. This is presumably a reflection of the Maastricht Treaty and the Stability and Growth Pact. The fact that this increased mean reversion also seem to be very similar across countries supports our claim that the assumption of cross-sectional independence is likely to be violated in the data. In order to get a measure of the size of this problem, we computed the cross-sectional correlation matrix of the deficit-GDP ratios. The results are reported in Table 1. It is seen that all correlations lie between zero and 0.83, with an overall average of 0.45, which indicates that pervasive cross-correlation is present amongst the EU countries.

3.3 Panel unit root tests

The preliminary results of the previous section suggest that government revenues and expenditures are in fact cointegrated with a unit cointegrating slope. Of course, since the integratedness of the variables is not yet known, these results should be interpreted with caution as the possibility remains that the variables are stationary. In order to shed some light on this issue, we now proceed by testing the variables for unit roots.

However, because the data appear to be cross-sectionally correlated, we cannot use the conventional approach of just combining individual unit root tests as if they were independent. For this purpose, we employ the G_{ols}^{++}, P_m and Z tests of Phillip and Sul (2003), the t_a and t_b tests of Moon and Perron (2004), and the Bai and Ng (2004) P_e^c test, which all permit for cross-sectional dependence by assuming that the variables admit to a common factor representation. For example, revenues may decomposed as

$$r_{it} \; = \; \lambda_i' f_t + u_{it}, \tag{9}$$

where f_t is a vector of unobserved common factors, which could represent fiscal shocks or any other feature affecting revenues that is common for all counties. These factors are introduced to model the cross-sectional dependence in r_{it}, whose magnitude is determined by λ_i. The remaining part u_{it} is assumed to be completely idiosyncratic.

The first three tests, G_{ols}^{++}, P_m and Z, assume that there is a single factor, while the remaining three, t_a, t_b and P_e^c, allow for an unknown number of common factors that are estimated using the method of principal components. The essential idea is to first estimate and subtract the common component from the data, and then to test for a unit root in the remaining estimated idiosyncratic component.

All tests take non-stationarity as the null hypothesis, and all tests except t_a and t_b permit the individual autoregressive roots to differ across countries, which is likely to be important in this kind of heterogenous data. Each statistic is normally distributed under the null hypothesis. Moreover, while G_{ols}^{++}, P_m, t_a and t_b are left-tailed, Z and P_e^c are right-tailed.

For the implementation of the tests, we follow the usual practice and estimate all long-run variances using the Newey and West (1994) estimator. The lag lengths are determined using the Akaike information criterion. Several different information criteria were applied to determine the appropriate number of common factors. However, this always resulted in the maximum number being chosen, which in not totally unexpected given that these criteria are known to suffer from an upwards bias in the presence of serial correlation. On the other hand, the exact number of common factors to use is relatively unimportant as long as it is larger than the true value. So the problem is how to determine how many factors to use in order to get rid of the effects of the cross-sectional dependence. In this paper, we use the results from both the estimated principal components and cross-correlations, which suggest that three factors should be enough to capture most of the cross-sectional dependence in the data.[3]

[3]The share of the total variation explained by the third principal component is 3.6% for expenditures and 2.6% for revenues.

Table 2 reports the results from the panel unit root tests for each variable. It is seen that the unit root null is accepted for all tests at the 5% level, and for all but one test at the more liberal 10% level. In view of this, we conclude that both variables appear to be non-stationary. In fact, since the entire panel is non-stationary under the null hypothesis, we can even say that we have evidence of non-stationarity for all 15 countries.

3.4 Panel cointegration tests

We know from earlier that the variables are correlated cross-sectionally, which suggests that existing tests for cointegration in independent panels cannot be used.

In this paper, we handle this issue by applying the DH_g and DH_p tests recently proposed by Westerlund (2007a), which are general enough to permit for an unknown number of common factors in the regression errors, just as in (9). These tests are therefore appropriate for our purposes. Similar to the unit root tests of Moon and Perron (2004) and Bai and Ng (2004), DH_g and DH_p are based on first estimating and subtracting the common component from the regression errors, and then testing for a unit root in the estimated idiosyncratic error component.

Note that because the tests are constructed under the maintained assumption of a unit root, the null hypothesis is formulated as that the cointegration is absent, while the alternative hypothesis is formulated as that there is at least some of the countries for which cointegration holds. Of course, with this formulation of the alternative, we cannot really say which countries that are cointegrated in case of a rejection, and we will therefore also consider a test that takes cointegration for all 15 countries as the null hypothesis. Such a test was recently developed by Westerlund and Edgerton (2007), who propose using the bootstrap approach to handle the problem with cross-sectional dependence.

The idea here is to first apply the DH_g and DH_p tests. If the no cointegration null is accepted, we conclude that the variables are not cointegrated, and proceed no further. If the null is rejected, however, the testing proceeds by applying the Westerlund and Edgerton (2007) test, denoted LM_N^+. If the null of cointegration is accepted, we conclude that the whole panel is cointegrated, while, if it is rejected, we conclude that there is cointegration but only for a subset of the countries. Only if the whole panel is cointegrated do we conclude that the hypothesis of strong sustainability cannot be rejected for the panel as a whole.

One of the main advantages of using the DH_g and DH_p tests, in spite of the unnatural formulation of the alternative hypothesis, is that they can be used to test for cointegration with a prespecified cointegrating slope, which is of course highly relevant in situations such as this where economic theory implies a particular slope value.

As before, all long-run variances are estimated using the Newey and West (1994) estimator, and the DH_g and DH_p tests are computed using three common factors, which again seem sufficient to capture the cross-sectional dependence in the data. The LM_N^+ test is based on a sieve bootstrap, where the order of the sieve approximation is permitted to increase with T. The number of bootstrap replications is set to $1,000$.

The results from the DH_g and DH_p tests are reported in Table 3. It is seen the no cointegration null is strongly rejected at all conventional significance levels, which means that there is at least some countries that are cointegrated. This conclusion is further strengthened by the LM_N^+ test, which result in an acceptance of the null of cointegration. In other words, there is no evidence against cointegration for the full panel. We also see that the no cointegration null is rejected if the cointegrating slope is fixed at unity as postulated by theory. Thus, we also have some evidence of strong sustainability.

3.5 Cointegration estimates

It is well known that the least squares estimator is consistent under fairly general conditions when applied to cointegrated regressions such as ours. Unfortunately, the presence of endogeneity and cross-sectional dependence means that the least squares estimator will generally be inefficient and biased, which makes it a poor candidate for inference.

A common approach to alleviate this problem is to use seemingly unrelated regressions techniques. However, this approach is not feasible in our case where T is only moderately larger than N. To circumvent this problem, we apply the newly developed panel estimators of Bai and Kao (2005) and Westerlund (2007b), which are similar to the tests employed earlier in the sense that the cross-sectional dependence is modelled using a small number of common factors. This makes them applicable even in situations when N is larger than T. The Bai and Kao (2005) estimator can be seen as a factor augmented version of the more conventional panel fully modified estimator of Kao and Chiang (2000), while the Westerlund (2007b) estimator is basically a factor augmented version of their bias-adjusted estimator.

Both estimators can be implemented either in two steps, by first estimating the factors and then the cointegration vector, or iteratively until convergence. The idea behind the latter is that estimation accuracy may be gained by iterating between the first and second step. The estimation is carried out as before, using three common factors and the Newey and West (1994) long-run variance estimator. In addition to the asymptotic p-values based on the normal distribution, we also compute bootstrapped p-values, which are based on making $1,000$ replications of the same bootstrap scheme used by Westerlund (2007b). The reason for having both types of p-values is that the asymptotic normal distribution can sometimes provide a poor approximation to the empirical distribution of the t-statistics.

The estimation results are reported in Table 4. It is seen that neither the asymptotic nor the bootstrapped p-values lend any evidence against the null of a unit slope at the 10% level for the fully modified estimator, which implies that the hypothesis of strong sustainability cannot be rejected. The fact that the bias-adjusted estimator leads to a different conclusion is somewhat puzzling. What we do know, however, is that this difference is due to the construction of the estimated slopes, as their asymptotic variance is the same. Nevertheless, since both estimators lie reasonably close to the hypothesized value of one, we chose to interpret Table 4 as giving evidence in favor of strong sustainability.

3.6 Out-of-sample forecasts

In this section, we make use of our cointegration test results in forecasting government revenues and expenditures. This undertaking is motivated in part by studies such as Artis and Marcellino (2001), who find that the fiscal forecasts of major international organizations rarely outperforms the naive random walk model, in part by the operating procedure of the Stability and Growth Pact, which involves reference to the forecast values of the fiscal stance. Intuitively, the fact that strong sustainability cannot be rejected should be beneficial for forecasting purposes.

The econometrical approach used here is taken from Mark and Sul (2001), who tested the predictive ability of the monetary model of exchange rate determination over the random walk. It proceeds as follows. We begin by using the fact that r_{it} and g_{it} are cointegrated with $\beta_i = 1$, which suggests that the deficit $e_{it} = r_{it} - g_{it}^r$ should be stationary. A k period ahead forecast for revenues, say, can therefore be obtained as

$$r_{it+k} - r_{it} = \eta_{ik} + \gamma_{ik} e_{it} + u_{it+k}, \tag{10}$$

where η_{ik} is an individual specific constant and u_{it+k} is a stationary forecast error. Note that (10) is basically an error correction model for $r_{it+k} - r_{it}$, which makes γ_{ik} a key parameter in the sense that it governs the error correction towards the deficit, e_{it}. A negative γ_{ik} implies that present day deviations from the deficit will be reversed in the future.

The idea is now to compare the forecast from (10) with the random walk forecast obtained by setting γ_{ik} to zero. If the sustainability of the public deficit has any bearing on the ability to forecast government revenues and expenditures, then we expect the forecast obtained from (10) to outperform the naive random walk.

The mean squared forecast error of the two models can be evaluated as in Mark and Sul (2001) by using the Theil U statistic, or by using the S statistic of Diebold and Mariano (1995). This paper uses both tests, which compare the null hypothesis of equal forecast accuracy against the one-sided alternative that the forecast obtained from (10) is more accurate than that obtained when setting γ_{ik} to zero. Unfortunately, asymptotic critical values for this type of tests can be severely biased in small samples because of the overlapping observations when $k > 1$.

In order to mitigate this bias, Mark and Sul (2001) propose using the bootstrap approach to calculate critical values based on the empirical distribution of the tests under the null of equal predictability. Unlike asymptotic critical values, bootstrapped critical values adjust for the serial correlation induced by the presence of overlapping observations and should thus enable valid inference even in the case when $k > 1$. The bootstrap algorithm used for this purpose is taken directly from Mark and Sul (2001), and is very general in the sense that it accounts for both the time series and cross-sectional dependence of the data.

Moreover, since the number of time series observations at hand is relatively small, it seems natural to make use also of the cross-sectional information when estimating (10). Two versions of (10) will therefore be considered, one heterogeneous where γ_{ik} is permitted to vary across i and one pooled where γ_{ik} is assumed to be equal across i.

The forecasting results are reported in Table 5 for revenues and in Table 6 for expenditures. Two horizons are considered, one year and four years. In both cases, we begin by estimating (10) using least squares on the observations available through 1985. This estimate is then used to generate the out-of-sample forecasts.

We begin by looking at Table 5, which shows substantial evidence of predictability, both at the one and four year horizons. As expected, we see that the random walk model is more easily beaten at shorter horizons when the forecasting equation is pooled. We also see that the Theil U statistic generally results in more rejections of the no predictability null. Thus, the sustainability of the public deficit is indeed a valuable piece of information when it comes to forecasting revenues. The results for expenditures, however, are less encouraging. In fact, based on the results reported in Table 6, there seem to be little or no evidence of predicability beyond that of a random walk. Apparently, predicting expenditures is much more difficult than predicting revenues when using the considered model.

One explanation for this difference in predictability is that a relatively large component of expenditures is determined outside the control of the government, see Alho and Vanne (2006). This means that expenditures will tend to be only slowly error correcting, which in turns implies that the resulting forecast will not be very different from that of a random walk.

4 Concluding remarks

The sustainability of the fiscal policy within the EU area has been, and continues to be, one of the most widely investigated empirical issues. The most popular approach used for testing this so-called sustainability hypothesis is based on the intertemporal budget constraint. The idea is that for the sustainability hypothesis to hold, the discounted public debt must be zero over the long term, which means that government revenues and expenditures should cointegrate with a unit slope on expenditures.

Yet for a theory so widely held among economic researchers, the postulated long-run one-for-one relationship between revenues and expenditures has proven very difficult to verify empirically, in spite of the recent advances in econometric methodology for testing long-run relationships using time series cointegration techniques. In fact, most studies of the EU tend to reject the sustainability hypothesis, which is especially alarming as the Maastricht Treaty and the Stability and Growth Pact should stabilize the public debt of these countries in the long run.

This paper can be seen as a response to these puzzling findings. The aim is to show that a failure to reject the absence of cointegration for an individual country need not be taken as evidence against the sustainability hypothesis, and that the use of panel data can lead to more accurate tests. The reason is that the conventional country-by-country approach can have low power because of the short time period usually applied. By contrast, the panel data approach used here enables us to selectively pool the information contained in both time series and cross-section dimensions, and should therefore yield more powerful tests. This power advantage is expected to be particularly pronounced in the light of the recent EU-wide legislative harmonization.

The data we use covers 15 EU member countries between 1970 and 2004. We begin by applying a battery of panel unit root and cointegration tests, which shows that revenues and expenditures are indeed non-stationary and cointegrated, as predicted by the sustainability hypothesis. In order to examine whether the sustainability is of the strong form, we further test whether the slope on expenditures is equal to one. The results suggest that the unit slope restriction cannot be rejected, which means that there is evidence of strong sustainability.

Having shown that the fiscal policy within the EU seem to be strongly sustainable, we further show how this information can be used to forecast government revenues and expenditures. This is particularly interesting in the light of the fact that the Stability and Growth Pact is operated in part by forecasting the fiscal stance. Our results show that the sustainability of the public debt is a useful piece of information in predicting revenues, and to a lesser extent also expenditures.

Table 1: Estimated cross-correlations.

No	Country	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Austria	1.00	0.42	0.60	0.42	0.69	0.41	0.63	0.16	0.50	0.02	0.54	0.75	0.77	0.28	0.43
2	Belgium	0.42	1.00	0.70	0.02	0.08	0.16	0.40	0.81	0.83	0.42	0.75	0.61	0.58	0.24	0.39
3	Denmark	0.60	0.70	1.00	0.33	0.49	0.57	0.49	0.43	0.51	0.23	0.66	0.71	0.66	0.64	0.51
4	Finland	0.42	0.02	0.33	1.00	0.72	0.27	0.41	0.26	0.08	0.14	0.17	0.28	0.59	0.67	0.57
5	France	0.69	0.08	0.49	0.72	1.00	0.50	0.70	0.23	0.18	0.16	0.40	0.54	0.68	0.49	0.51
6	Germany	0.41	0.16	0.57	0.27	0.50	1.00	0.19	0.22	0.17	0.00	0.37	0.51	0.17	0.50	0.58
7	Greece	0.63	0.40	0.49	0.41	0.70	0.19	1.00	0.03	0.57	0.17	0.63	0.62	0.78	0.11	0.27
8	Ireland	0.16	0.81	0.43	0.26	0.23	0.22	0.03	1.00	0.70	0.62	0.57	0.41	0.20	0.09	0.35
9	Italy	0.50	0.83	0.51	0.08	0.18	0.17	0.57	0.70	1.00	0.34	0.75	0.59	0.59	0.03	0.47
10	Luxembourg	0.02	0.42	0.23	0.14	0.16	0.00	0.17	0.62	0.34	1.00	0.34	0.06	0.08	0.07	0.16
11	Netherlands	0.54	0.75	0.66	0.17	0.40	0.37	0.63	0.57	0.75	0.34	1.00	0.58	0.65	0.18	0.37
12	Portugal	0.75	0.61	0.71	0.28	0.54	0.51	0.62	0.41	0.59	0.06	0.58	1.00	0.72	0.41	0.51
13	Spain	0.77	0.58	0.66	0.59	0.68	0.17	0.78	0.20	0.59	0.08	0.65	0.72	1.00	0.40	0.42
14	Sweden	0.28	0.24	0.64	0.67	0.49	0.50	0.11	0.09	0.03	0.07	0.18	0.41	0.40	1.00	0.67
15	United Kingdom	0.43	0.39	0.51	0.57	0.51	0.58	0.27	0.35	0.47	0.16	0.37	0.51	0.42	0.67	1.00

Notes: The table reports the absolute value of the estimated cross-correlations.

Table 2: Panel unit root tests.

		Revenue		Expenditure	
Study	Test	Value	p-value	Value	p-value
Bai and Ng (2004)	P_e^c	−0.487	0.687	0.279	0.390
Phillips and Sul (2003)	G_{ols}^{++}	4.450	1.000	−0.012	0.495
	Z	2.957	0.998	−0.345	0.365
	P_m	−2.471	0.993	1.317	0.094
Moon and Perron (2004)	t_a	0.397	0.654	−0.511	0.305
	t_b	2.157	0.984	−1.174	0.120

Notes: The tests are based on an intercept and the Newey and West (1994) variance estimator. All lag lengths are chosen using the Akaike information criterion and the number of common factors is set to three.

Table 3: Panel cointegration tests.

		Estimated β_i^{a}		Prespecified β_i^{b}	
Study	Test	Value	p-value	Value	p-value
Westerlund (2007a)	DH_g	8.129	0.000	6.671	0.000
	DH_p	12.464	0.000	11.467	0.000
Westerlund and Edgerton (2007)	LM_N^+	4.317	0.999	–	–

Notes: All tests are based on an intercept and the Newey and West (1994) variance estimator. The p-values for the DH_g and DH_p tests are based on the asymptotic normal distribution while the p-value for the LM_N^+ test is based on the bootstrapped distribution using $1,000$ replications.

[a]The tests are based on estimating β_i.

[b]The tests are based on a prespecified β_i that is set equal to one.

Table 4: Panel estimation results.

Value	BA[a]	FM[a]	BA[b]	FM[b]
β	0.647	0.951	0.649	0.950
p-value[c]	0.000	0.125	0.000	0.116
p-value[d]	0.000	0.554	0.000	0.544

Notes: The values in the top row refer to the estimated cointegrating slope, while BA and FM refer to the bias-adjusted and fully modified estimators of Westerlund and (2007b) and Bai and Kao (2005), respectively. The results are based on an intercept and the Newey and West (1994) variance estimator.

[a]The estimation is carried out in two steps.

[b]The estimation is carried out iteratively until convergence.

[c]The p-values are for a double-sided unit slope test based on the bootstrapped distribution using $1,000$ replications.

[d]The p-values are for a double-sided unit slope test based on the normal distribution.

Table 5: Forecasts of government revenues.

	Pooled estimation								Heterogeneous estimation							
	$k = 1$				$k = 4$				$k = 1$				$k = 4$			
Country	D	p-val	U	p-val	D	p-val	U	p-val	D	p-val	U	p-val	D	p-val	U	p-val
Austria	1.60	0.97	1.06	1.00	2.94	0.97	1.16	0.98	1.02	1.00	1.03	0.97	−2.48	0.07	0.84	0.07
Belgium	0.44	0.58	1.03	1.00	1.27	0.78	1.21	0.99	0.96	0.05	1.02	0.00	0.49	0.21	1.08	0.24
Denmark	−0.76	0.13	0.98	0.06	−1.17	0.02	0.92	0.03	−0.77	0.00	0.98	0.00	−0.69	0.00	0.91	0.00
Finland	1.93	0.99	1.16	1.00	1.62	0.89	1.31	0.84	1.02	0.00	1.08	0.00	1.66	0.14	1.81	0.00
France	2.16	1.00	1.09	1.00	2.65	0.95	1.27	1.00	1.68	0.98	1.08	0.99	2.06	0.89	1.35	0.98
Germany	−1.20	0.11	0.96	0.00	−1.31	0.28	0.93	0.13	1.06	0.32	1.03	0.04	1.25	0.52	1.14	0.48
Greece	−1.30	0.09	0.93	0.00	−2.01	0.22	0.82	0.02	−0.27	0.70	0.98	0.65	−1.60	0.22	0.83	0.12
Ireland	−1.85	0.01	0.91	0.00	−1.75	0.17	0.87	0.01	−1.09	0.28	0.91	0.33	−1.58	0.09	0.88	0.04
Italy	−1.53	0.04	0.91	0.00	−1.45	0.25	0.85	0.02	−0.87	0.00	0.93	0.00	−0.53	0.01	0.89	0.01
Luxembourg	−1.04	0.24	0.93	0.00	−0.98	0.28	0.90	0.05	0.61	0.60	1.16	0.98	0.96	0.58	1.46	0.97
Netherlands	−0.14	0.44	1.00	0.00	0.53	0.63	1.03	0.70	0.67	0.09	1.03	0.03	0.28	0.14	1.04	0.17
Portugal	−0.27	0.46	0.99	0.19	1.21	0.52	1.11	0.34	−0.05	0.00	1.00	0.00	0.44	0.00	1.02	0.00
Spain	1.12	0.86	1.09	1.00	1.88	0.91	1.21	1.00	1.01	0.42	1.06	0.20	3.50	0.97	1.15	0.36
Sweden	−1.20	0.00	0.93	0.00	−1.01	0.00	0.93	0.00	−0.47	0.00	0.97	0.00	0.16	0.00	1.03	0.00
United Kingdom	1.03	0.78	1.07	1.00	0.32	0.59	1.07	0.86	2.03	0.96	1.36	1.00	0.55	0.60	1.11	0.90

Notes: The value k refers to the forecast horizon, while D and U refer to the Diebold and Mariano (1995) and Theil statistics, respectively. The reported p-values are one-sided, based on $1,000$ bootstrap replications. The pooled results are based on assuming that the slopes of the forecasting equation are equal across the cross-section, while the heterogeneous results are not.

Table 6: Forecasts of government expenditures.

	Pooled estimation								Heterogeneous estimation							
	$k = 1$				$k = 4$				$k = 1$				$k = 4$			
Country	D	p-val	U	p-val	D	p-val	U	p-val	D	p-val	U	p-val	D	p-val	U	p-val
Austria	−2.58	0.05	0.93	0.40	−1.83	0.16	0.89	0.09	−1.70	0.24	0.90	0.19	−2.20	0.08	0.71	0.01
Belgium	0.84	0.93	1.03	0.94	1.40	0.84	1.13	0.92	1.33	0.91	1.10	0.97	1.59	0.84	1.30	0.97
Denmark	−1.07	0.75	0.98	0.78	0.10	0.55	1.01	0.60	−0.43	0.74	0.99	0.78	0.82	0.64	1.11	0.82
Finland	−1.71	0.59	0.95	0.69	−2.13	0.11	0.77	0.00	−0.22	0.86	0.98	0.87	−1.40	0.19	0.81	0.07
France	−2.21	0.09	0.92	0.40	−1.80	0.17	0.86	0.08	−0.60	0.64	0.94	0.48	−1.02	0.23	0.87	0.13
Germany	−0.64	0.57	0.99	0.67	−0.64	0.36	0.95	0.25	−0.13	0.54	0.99	0.46	0.36	0.51	1.08	0.75
Greece	0.25	0.74	1.02	0.77	0.90	0.72	1.22	0.91	1.24	0.91	1.08	0.96	3.06	0.98	1.23	0.93
Ireland	2.48	0.99	1.09	0.94	2.99	0.96	1.29	0.99	0.29	0.69	1.01	0.69	2.14	0.93	1.21	0.91
Italy	1.63	0.99	1.08	0.98	1.07	0.75	1.17	0.93	1.69	0.96	1.10	0.99	1.09	0.73	1.11	0.82
Luxembourg	0.19	0.97	1.01	0.98	1.46	0.88	1.14	0.97	1.31	0.99	1.07	0.98	1.38	0.86	1.19	0.95
Netherlands	−1.26	0.67	0.97	0.70	−0.69	0.39	0.98	0.40	−0.28	0.72	0.98	0.64	−0.09	0.41	0.99	0.38
Portugal	−1.61	0.64	0.94	0.59	−1.68	0.22	0.86	0.08	−1.36	0.63	0.95	0.75	−1.64	0.16	0.84	0.12
Spain	−1.21	0.51	0.96	0.46	−1.24	0.21	0.87	0.11	0.24	0.73	1.01	0.78	0.88	0.59	1.12	0.80
Sweden	1.02	0.88	1.05	0.93	−0.63	0.30	0.90	0.08	1.54	0.90	1.06	0.92	−0.15	0.36	0.97	0.29
United Kingdom	−1.15	0.44	0.97	0.57	−2.46	0.06	0.79	0.01	−0.28	0.51	0.99	0.52	−3.13	0.03	0.71	0.01

Notes: See Table 5 for an explanation of the various features.

References

Alho, J., and R. Vanne (2006). On Predictive Distributions of Public Net Liabilities. *International Journal of Forecasting*, Vol. 22, pp. 725-733.

Afonso, A. (2005). Fiscal Sustainability: The Unpleasant European Case. *FinanzArchiv*, Vol. 61, pp. 19-44.

Artis, M., and M. Marcellino (2001). Fiscal Forecasting: The Track Record of the IMF, OECD and EC. *Econometrics Journal*, Vol. 4, pp. S20-S36.

Bai, J., and C. Kao (2005). On the Estimation and Inference of a Panel Cointegration Model with Cross-Sectional Dependence. In Baltagi, B. (Ed.), *Contributions to Economic Analysis*, Elsevier, Amsterdam.

Bai, J., and S. Ng (2004). A Panic Attack on Unit Roots and Cointegration. *Econometrica*, Vol. 72, pp. 1127-1177.

Bravo, A., and A. Silvestre (2002). Intertemporal Sustainability of Fiscal Policies: Some Tests for European countries. *European Journal of Political Economy*, Vol. 18, pp. 517-528.

Caporale, G. (1995). Bubble Finance and Debt Sustainability: A Test of the Government's Intertemporal Budget Constraint. *Applied Economics*, Vol. 27, pp. 1135-1143.

Diebold, F. X., and R. S. Mariano (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, Vol. 13, pp. 253-263.

Kao, C., and M. H. Chiang (2000). On the Estimation and Inference of a Cointegrated Regression in Panel Data. *Advances in Econometrics*, Vol. 15, pp. 179-222.

Mark, N. C., and D. Sul (2001). Nominal Exchange Rates and Monetary Fundamentals: Evidence from a Small Post-Bretton Woods Panel. *Journal of International Economics*, Vol. 53, pp. 29-52.

Moon, H. R., and B. Perron (2004). Testing for a Unit Root in Panels with Dynamic Factors. *Journal of Econometrics*, Vol. 122, pp. 81-126.

Newey, W., and K. West (1994). Autocovariance Lag Selection in Covariance Matrix Estimation. *Review of Economic Studies*, Vol. 61, pp. 631-653.

Papadopoulos, A., and M. Sidiropoulos (1999). The Sustainability of Fiscal Policies in the European Union. *International Advances in Economic Research*, Vol. 5, pp. 289-307.

Phillips, P. C. B., and D. Sul (2003). Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence. *Econometrics Journal*, Vol. 6, pp. 217-259.

Quintos, C. (1995). Sustainability of the Deficit Process with Structural Shifts. *Journal of Business and Economic Statistics*, Vol. 13, pp. 409-417.

Vanhorebeek, F., and P. V. Rompuy (1995). Solvency and Sustainability of Fiscal Policies in the EU. *De Economist*, Vol. 143, pp. 457-473.

Westerlund, J. (2007a). Panel Cointegration Tests of the Fisher Hypothesis. Forthcoming in *Journal of Applied Econometrics*.

Westerlund, J. (2007b). Unbiased Estimation of Cointegrated Panel Regressions with Cross Section Dependence. Forthcoming in *Journal of Financial Econometrics*.

Westerlund, J., and D. L. Edgerton (2007). A Panel Bootstrap Cointegration Test. Forthcoming in *Economics Letters*.