Multi-step analysis of public finances sustainability☆

António Afonso a,b,c,⁎, Christophe Raulet d,e

a ISEG/Universidade de Lisboa, Department of Economics, Portugal
b ISEP/Universidade de Lisboa, Department of Economics, Portugal
c ISEP/Universidade de Lisboa, Department of Economics, Portugal
d The European Central Bank, Directorate General Economics, Kaiserstraße 29, D-60311 Frankfurt am Main, Germany
e University of Orléans (LEO, UMR CNRS 7322), France

A R T I C L E   I N F O

Article history:
Accepted 9 October 2014
Available online 29 October 2014

JEL classification:
C23
E62
H62

Keywords:
Fiscal sustainability
EU
Panel cointegration

A B S T R A C T

We use an original multi-step analysis to assess the sustainability of public finances. Firstly, we investigate the mean-reverting behaviour of government expenditures and revenues. Secondly, we apply bootstrap panel cointegration techniques to check for a long-run relationship between expenditures and revenues. Thirdly, we check the coefficient of expenditure in the cointegration relation. Fourthly, we estimate panel error correction models to identify short-run and long-run fiscal developments. While the results imply that public finances were not unsustainable for the EU panel, they highlight that fiscal sustainability is an issue for most countries, as evidenced by a below unit estimated coefficient of expenditure in the cointegration relation.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Studies on the sustainability of public finances regarding the European Union (EU) usually restrict themselves to the set of EU Member States from before the 1st of May, 2004 enlargement. To our knowledge, this is the first fully-fledged panel analysis of fiscal sustainability that encompasses the enlarged set of EU countries. The choice of the former group of countries is usually prompted by the lack of longer comparable time series data for the new EU Member States. In this paper we assess the sustainability of public finances, applying non-stationary panel data econometric techniques as well as the Pooled Mean Group (PMG) estimation method (see e.g. Pesaran and Smith, 1995; Pesaran et al., 1999; and the Common Correlated Effect (CCE) estimator (Pesaran, 2006, which allows for common factors in the cross equation covariances to be removed). We cover several sub-periods within the period of 1960–2012 and define different country groupings for the 27 members of the enlarged EU.

Even though the EU does not have a single fiscal policy, panel analysis of the sustainability of public finances is relevant in the context of the fact that the 27 EU countries are committed to complying with sound fiscal policies within the Stability and Growth Pact framework. Cross-country dependence is observed, either in the run-up to EMU or, for example, via integrated financial markets. Indeed, cross-country spillovers in government bond markets are to be expected, and interest rates co-movements inside the EU have also gradually become more noticeable. On the other hand, based on the fact that fiscal sustainability certainly needs to be tackled at country level, a country assessment is also necessary, and it is thus useful to obtain as many time series observations as possible. In this context, the use of the Seemingly Unrelated Regression D-F (SURADF) panel integration test provides additional country-specific results.

In the literature, fiscal sustainability analysis based on unit root or cointegration tests, has in the past been mostly performed for individual countries, giving rise to the problem of relatively short time series. Usually, such country analyses tend to conclude that only a few developed economies satisfy the sustainability hypothesis, those being notably Germany (allowing for the reunification break), the Netherlands, Finland, and Austria, and the U.K., even if the question of the time span under consideration can be an issue for some studies (see

http://dx.doi.org/10.1016/j.econmod.2014.10.011
0264-9993/© 2014 Elsevier B.V. All rights reserved.
Afonso, Blackley, and Fève and Hénin, 2000; Uctum and Wickens, 2000).

However, panel data methods have recently been used to assess fiscal sustainability, notably when testing for cointegration between general government expenditure and revenue, which is a relationship which is derived from the intertemporal budget constraint. In this context, panel unit root and panel cointegration analysis have been used, an example being Prohl and Schneider’s study (2006) for the EU, Westerlund and Prohl (2010) and Afonso and Rault’s (2007, 2010) for OECD countries, and Afonso and Jalles’ (2012) for the EU. Apart from the already-mentioned cross-country dependency issue, another argument for using this approach is the increased support that may be given to the cointegration hypothesis, through the increased number of observations that result from adding individual time series.

In this paper we contribute to the literature by proposing a multi-step empirical methodology to assess the sustainability of public finances in the EU. Firstly, the SURADF panel integration test from Breuer et al. (2002, 2006) is implemented for general government expenditure and revenue series, as a ratio of GDP. To the best of our knowledge, this is the first empirical application of this test for the context of fiscal sustainability. The test takes into account cross-sectional dependence among countries and permits the identification of how many, and which countries of the panel have a unit root, which has the advantage of enabling the consideration of the possibility of cross-section dependence, yet still provides country-specific sustainability assessments.

Secondly, for those countries for whom spending and revenue are found to be integrated of order one, we then carry out the panel bootstrap test of Westerlund and Edgerton (2007), which tests for the null hypothesis of cointegration between the two sides of the budget against the alternative of at least one country having two variables that are not cointegrated.

Thirdly, if cointegration is not evidenced, then long-run fiscal non-sustainability is consistent with the data, whereas if cointegration is in evidence, then the testing proceeds by checking with Pesaran’s CCE procedure (2006) (which allows for cross-section dependencies), for a unit slope on spending in a regression of revenues on spending. The latter condition is sufficient for the long-run sustainability of public finances to hold.

Fourthly, another important issue is how to model the reduced form relationship in the possible presence of non-stationarity in the panel. Indeed, a cursory reading of the formal literature on government spending and borrowing in stochastic general equilibrium, suggests that, given the panel data employed, there could also be short-run cyclical effects of relevance, which may vary across countries. Thus, in order to address this issue, we employ the Pooled Mean Group approach of Pesaran et al. (1999) to identify the long-run versus short-run effects for the respective fiscal policies of the EU member states. The advantage of such an approach is that not only does it inform about the issue of unit-roots in the country panel, but it also allows for short-run versus long-run analyses of fiscal sustainability for the same specification. Individual countries may well be on the same long-run path, albeit with different short-run developments.

Note that an important benefit of our multi-step analysis (composed of the four steps mentioned above) is that it enables us to take into account the possible dependence that may exist across the different countries in the panel, thus yielding robust results. This question is crucial, and responds to the complex nature of the interactions and dependencies that generally exist over time and across the individual countries in the panel.

The rest of the paper is organised as follows. Section 2 briefly presents the analytical framework for fiscal sustainability. Section 3 explains our econometric strategy. Section 4 reports the empirical fiscal sustainability results for the EU, following our multi-step analysis, and Section 5 concludes.

2. Analytical framework for fiscal sustainability

The starting point for the analysis of the sustainability of public finances is the so-called present value borrowing constraint which is derived from the flow budget constraint (see Afonso, 2005, and Hakko and Rush, 1991), and can be written for a given country as:

$$B_{t-1} = \sum_{s=0}^{\infty} \frac{1}{(1+r)^{t-s}} (R_{t+s} - E_{t+s}) + \lim_{s \to \infty} \frac{B_{t+s}}{(1+r)^{t-s}}$$  

(1)

where $E_t = GP_t + (r_t - r)B_t - 1$, with $GP_t$ = primary government expenditure, $R_t$ = government revenue, $B_t$ = government debt at the end of $t$, and $r$ = real interest rate, assumed to be stationary with $r$ mean. A sustainable fiscal policy needs to ensure that the present value of the stock of government debt, the second term of the right-hand side of Eq. (1), reaches zero at infinity, constraining the debt to grow no faster than the real interest rate. The existence of this no-Ponzi game condition also implies that the outstanding stock of government debt equals the present value of future government surpluses. From Eq. (1), and defining $G_t = GP_t + rB_t - 1$, we have, after some manipulations:

$$G_t - R_t = \sum_{s=0}^{\infty} \frac{1}{(1+r)^{t+s}} (\Delta R_{t+s} - \Delta E_{t+s}) + \lim_{s \to \infty} \frac{B_{t+s}}{(1+r)^{t-s}}$$  

(2)

Using GDP ratios, with the GDP real growth rate ($y$) also assumed to be constant, we have, after some manipulations:

$$b_{t-1} = \sum_{s=0}^{\infty} \frac{1+y}{1+r} (\Delta E_{t+s} - \Delta E_{t+s}) + \lim_{s \to \infty} \frac{b_{t+s}}{(1+y)^{t-s}}$$  

(3)

where $b_t = B_t / Y_t$, $e_t = E_t / Y_t$ and $\rho_t = R_t / Y_t$. When $r > y$, the solvency condition of:

$$\lim_{s \to \infty} \frac{b_{t+s}}{(1+y)^{t-s}} = 0$$  

(4)

is needed to limit public debt growth (the inequality in this case would be a suboptimal result for the government).

With the no-Ponzi game condition, that is, with the second term of the right-hand side of Eqs. (1) and (2) converging to zero, $G$ and $R$ must be cointegrated of order one for their first differences to be stationary. If $R$ and $E$ are non-stationary, and the first differences are stationary, then $R$ and $E$ are I(1) in levels. Thus, for Eq. (3) to hold, it is sufficient that the left-hand side, i.e., the budget balance, is stationary, which is possible if $G$ and $R$ are integrated of order one, with a $(-1)$ cointegration vector. Therefore, assessing fiscal sustainability involves testing the cointegration regression, notably as GDP ratios:

$$R_t = a + \beta G_t + u_t.$$  

(5)

Naturally, and as mentioned, for instance, by Afonso (2005), if one of the two variables is I(0) and the other is I(1), then there may be a sustainability issue, which, to be precise, cannot be tested via cointegration. On the other hand, it may also be the case that even with different orders of integration, that there are no sustainability problems if revenue...
systematically remains greater than expenditure, and if the country consistently runs a budgetary surplus.4

The non-Ponzi condition has to be seen, in the panel context, as having relevance for the entire panel. Therefore, the overall fiscal position of the set of EU countries considered would be sustainable in spite of some problems that are evidenced in some countries.

In addition, a related strand of the literature estimates fiscal reaction functions, usually to assess the Ricardian behaviour of governments, which can qualify as an indirect way of studying the sustainability of public finances. This can be done in a simple setup, as performed, for instance, by Afonso (2008), or even with regime switches, as in Davig and Leeper (2005), and Afonso and Toffano (2013). However, this analysis is more suitable for characterising the fiscal regime itself, which is, the existence of “active” and “passive” fiscal regimes, as labelled by Leeper (1991), and it commonly mentions the use of the primary balance as the dependent variable. The analysis performed in our paper focuses directly on the intertemporal government budget constraint, which delivers a result of the degree of difficulty for a government to meet explicit liabilities in the future. Moreover, by checking the cointegration of the two sides of the budget, an assessment of the necessary co-movement is assured, which is crucial for sustainability.

3. Econometric strategy

3.1. Methodological issues

The literature on panel unit root and panel cointegration distinguishes between first generation tests developed on the assumption of the cross-sectional independence among panel units (except for common time effects), and second generation tests which allow for a variety of dependence across the different units.

Although panel data unit root tests are likely to have higher power than conventional time series unit root tests, as they include cross-section variations (which make them very popular in applied studies), their results must, however, be interpreted with some caution, especially when testing for fiscal policy sustainability. In particular, as noted by Taylor and Sarno (1998), and Taylor and Taylor (2004), the null and alternative hypotheses are less suitable when there is the possibility of using a mixed panel, as for example, when some members may be stationary, while others may be non-stationary. Indeed, rejection of the unit root null in the panel does not imply that stationarity holds, even for the majority of members in the panel. The most that can be inferred then, is that at least one country is mean reverting, or that stationarity holds only marginally for a few countries. In the context of fiscal sustainability, and in economic terms, this would imply that the stock of general government debt is stationary for at least one country, even though public finances may not be sustainable for the majority of the countries in the panel sample.

However, researchers sometimes tend to draw a much stronger inference when in a given panel sample all government debt series are mean reverting, hence claiming to provide evidence that supports fiscal sustainability, which is not necessarily valid. Instead, for mixed panels, in the case of most interpretations, the preferred positioning of the null hypothesis would be “stationarity holds for all members of the panel”, against the alternative of “stationarity fails for at least some members of the panel”. This would allow for testing how pervasive the fiscal sustainability condition is for any given group of countries. One way to do this would be to use a panel test for the null of stationarity (see e.g. Hadri, 2000, whose null hypothesis is stationarity).

Another way to address this issue would be to use a procedure which allows the identification of how many, and which panel members are responsible for rejecting the joint null of non-stationarity. For example, Breuer et al. (2002, 2006) advocate a procedure whereby unit root testing is conducted within a Seemingly Unrelated Regression (SUR) framework, which exploits the information in the error covariance to produce efficient estimators and potentially more powerful test statistics. A further advantage of this procedure is that the SUR framework is another useful way of addressing cross-sectional dependency. We will pursue this second option in our paper in Step 1 of our empirical methodology.

Similar to most panel data unit root tests that are based on the null hypothesis of joint non-stationarity (against the alternative that at least one panel member is stationary), the well-known panel cointegration tests developed by Pedroni (1999, 2004), and generalised by Banerjee and Carrion-i-Silvestre (2006), are of the null of joint non-cointegration. The problem here is that a single series from the panel might be responsible for rejecting the joint null of non-stationary or non-cointegration, thus not necessarily implying that a cointegration relationship holds for the whole set of countries. In addition, such panel tests for the null hypothesis of no cointegration have been criticised in the literature, as it is usually the opposite null that is of primary interest in empirical research. A possible way to overcome this difficulty is to implement the bootstrap panel cointegration test, as proposed by Westerlund and Edgerton (2007), where the null hypothesis is cointegration, which implies, if not rejected, the existence of a long-run relationship for all panel members, and potential fiscal sustainability (the alternative hypothesis being that there is no cointegrating relationship for at least one country of the panel). The Westerlund and Edgerton (2007) test relies on the popular Lagrange multiplier test of Mccoskey and Kao (1998), and permits correlation to be accommodated both within, and between, individual cross-sectional units, and we use it in our Step 2.

Next we assess the magnitude of the β coefficient in the cointegration regression, via the computation of new common correlated CCE effect and CCE-MG (Mean Group) estimators (Pesaran, 2006), which constitutes our Step 3. Finally, we estimate panel ECM (PECM) models with the Pooled Mean Group approach of Pesaran et al. (1999), which allows us to assess the adjustment mechanism for a deviation from the long-run equilibrium relationship, along with the short-run dynamics, which is our Step 4.

3.2. Series specific panel unit root test: SURADF

The SURADF test developed by Breuer et al. (2002, 2006) is based on the following system of ADF equations:

\[ \Delta y_{1,t} = \alpha_1 + \phi_1 y_{1,t-1} + \sum_{i=1}^{p_1} \gamma_{1,i} \Delta y_{1,t-i} + \epsilon_{1,t} \quad t = 1, \ldots, T \]
\[ \Delta y_{2,t} = \alpha_2 + \phi_2 y_{2,t-1} + \sum_{i=1}^{p_2} \gamma_{2,i} \Delta y_{2,t-i} + \epsilon_{2,t} \quad t = 1, \ldots, T \]
\[ \Delta y_{N,t} = \alpha_N + \phi_N y_{N,t-1} + \sum_{i=1}^{p_N} \gamma_{N,i} \Delta y_{N,t-i} + \epsilon_{N,t} \quad t = 1, \ldots, T \]

where \( \phi = (\rho - 1) \) and \( \rho \) is the autoregressive coefficient for series \( j \) (\( j = 1, \ldots, N \)). This system is estimated by the SUR procedure and the null and the alternative hypotheses are tested individually as:

\[ H_0^\phi : \phi_j = 0; \quad H_1^\phi : \phi_j < 0 \]
\[ H_0^\rho : \rho_j = 0; \quad H_1^\rho : \rho_j > 0 \]

with the test statistics computed from SUR estimates of system (6), while the critical values are generated by Monte Carlo simulations. This procedure has three advantages. Firstly, by exploiting the information from the error covariances and allowing for autoregressive process,
it leads to efficient estimators over the single-equation methods. Secondly, the estimation also allows for heterogeneity in the lag structure across the panel members and, thirdly, the SURADF panel integration test accounts for possible cross-sectional dependence among countries, and allows for the identification of how many, and which countries of the panel have a unit root.

4. Investigating fiscal sustainability in the EU

All data for general government expenditure and revenue are taken from the European Commission Annual Macro-Economic Data (AMECO) database.5 The data cover the periods of 1960–2012 for the EU15 countries, 1998–2012 for the EU25 countries, and 2000–2012 for the EU26 countries, in order not to consider two countries with the smallest number of observations in the sample. The second sub-period covers the period after the setting up of the Economic and Monetary Union, and the last period concentrates on the years before, during, and after the 2008–2009 economic and financial crisis.6

4.1. Step 1: unit root analysis

There are good reasons to believe that there is considerable heterogeneity among the countries under investigation, and thus the typical panel unit root tests may lead to misleading inferences. Besides, the rejection of the null hypothesis that all series have a unit root does not imply that, under the alternative, all series are mean-reverting, as there may be a mixture of stationary and non-stationary processes in the panel under the alternative hypothesis. However, in case of the rejection of the null, for instance by the conventional Im et al. (2003), or by Breitung (2000) tests, this does not provide us with information about the exact mix of series in the panel, which is, for which series in the panel the unit root is rejected, and for which it is not. The SURADF test proposed by Breuer et al. (2002, 2006) addresses this issue.

Another advantage is that the SUR framework is a useful way of addressing cross-sectional dependency. To test for the presence of such cross-sectional dependence we have implemented the simple Pesaran test (2004). This test is based on the average of pair-wise correlation coefficients of the OLS residuals, obtained from standard augmented Dickey–Fuller regressions for each individual. Its null hypothesis is cross-sectional independence and it is asymptotically distributed as a two-tailed standard normal distribution. Results (which are available upon request) indicate that the null hypothesis is always rejected, regardless of the number of lags included in the augmented DF auxiliary regression (up to five lags) at the 5% level of significance.7

In the context of our paper, cross-dependence can mirror possible changes in the behaviour of fiscal authorities as a result of the signing of the EU Treaty in Maastricht on the 7th of February, 1992. The setting up of the fiscal convergence criteria urged the EU countries to consolidate their public finances from the mid-1990s onwards, during the run-up to the EMU on the 1st of January, 1999, which was when most EU legacy currencies were replaced by the Euro, and more recently after the adoption of the EU fiscal framework by the New Member States.

As the SURADF test has non-standard distributions, the critical values need to be obtained via simulations. In the case of the Monte Carlo simulations, the intercepts and the coefficients on the lagged values for each series were set equal to zero in each of the three EU15, EU25 and EU26 panel sets (see Breuer et al., 2002, 2006). Following on from this, the lagged differences and the covariance matrix were obtained from the SUR estimation on the general government expenditure and revenue ratio series. The SURADF test statistic for each series was computed as the t-statistic, calculated individually for the coefficient on the lagged level. To obtain the critical values, the experiments were replicated 10,000 times, and the critical values of 5% were tailored to the number of countries considered in the three panel sets.8

As it is now well known, in the presence of cross-section, the Ordinary Least Squares (OLS) estimator is not efficient, which makes it a poor option for inference. A common approach to alleviate this problem is to use the Seemingly Unrelated Regressions (SUR) estimator, which is more efficient than the OLS estimator in this framework. However, as noted by Westerlund (2007), this approach is not feasible when the number of countries considered in the three panel sets is too large. Indeed, the SURADF test statistic is actually performed on the (unbalanced) panel under the alternative hypothesis. However, in case of the rejection of the null hypothesis that all series have a unit root does not imply that, under the alternative, all series have a unit root. Consequently, it is beneficial to use SURADF tests which permit new insights.

11 The simulations provided by Breuer et al. (2002) demonstrate that the SURADF tests provide increased power over the single equation Dickey–Fuller tests when residual cross-correlations are high, as is the case in our country-sustainability analysis. Indeed, the SUR estimators exploit the information in the error covariances to produce efficient estimators. This is not the case for the conventional country-specific ADF tests, which we have implemented for comparison purpose, as these were unable to reject the individual unit root (under the null) for each country of our three panel sets (results are available on request). Consequently, it is beneficial to use SURADF tests which permit new insights.
The SURADF column refers to the estimated Augmented Dickey-Fuller statistics. Each of the estimated equation includes a constant, but not a time trend. The time-series suggest that general government revenue and expenditure ratios are non-stationary.

4.2. Step 2: panel cointegration

Given the results of the SURADF tests, we define three new panel sets: EU14, which includes all countries of the EU15 panel except for Germany; EU21 and EU22, which correspond to the previous EU25 and EU26 panel sets, without Estonia, Hungary, Poland and Slovakia. In-deed, in these countries, at least one of the two series of general government revenue and expenditure is integrated of order zero, thus preventing us from using cointegration techniques. We then perform 2nd generation panel data cointegration tests between government expenditure and revenue, which allow for cross-sectional dependence among countries.13

More specifically, we use the bootstrap panel cointegration test of Westerlund and Edgerton (2007). Unlike the panel data cointegration tests of Pedroni (1999, 2004), which were generalised by Banerjee and Carrion-i-Silvestre (2006), this test has the advantage that the joint null hypothesis is cointegration. Therefore, in the case of null non-rejection, we know that a cointegration relationship exists for the joint null hypothesis is cointegration. Therefore, in the case of null

---

12 As pointed out by a referee, possible breaks can affect estimated short-run and long-run parameters, if found to be present in the data. For this reason, it seems important to test for the presence of breaks in the data, and also to estimate their dates, in order to assess whether different fiscal realities and behaviours can be detected. Therefore we employed the panel data unit root test based on the Lagrange multiplier (LM) principle developed by Im et al. (2005). This approach is very flexible, as it can be applied not only when a structural break occurs at a different time period in each time series, but also when the structural break occurs in only some of the time series. It is also robust to the presence of structural breaks and more powerful than the popular Im et al. (2003) test in the basic scenario, where no structural break is involved. Furthermore, as reported by Im et al. (2005), as the LM test loses little power when controlling for spurious structural breaks when they do not exist, this represents a reasonable strategy for controlling for breaks, even when they are only at a suspicious level. We have implemented such tests on all series taken in level (general government expenditure-to-GDP ratios, and general government revenue-to-GDP ratios) for the three panel sets (EU14, EU21, EU22) and there is no clear support of the existence of statistically significant breaks in the data (even if results are borderline for the 1992 year associated with the Maastricht event). Similar results are obtained with Lee and Strazicich (2003) approach, employing minimum Lagrange Multiplier (LM) tests.

13 We previously found significant evidence of cross-sectional dependence in the data.

14 This bootstrap test is based on the sieve-sampling scheme, with the advantage of significantly reducing the distortions of the asymptotic test. We are grateful to Joakim Westerlund for sending us his Gauss codes.
sample. This null hypothesis is only not rejected in the case of the EU22 panel.

However, if we refer to the bootstrap critical values (which are valid here, given the high degree of cross-section dependence found in the data with the Pesaran CD statistic of 2004), we can conclude that a cointegration relationship exists between government revenue and expenditure ratios for the three panel data sets. This result is robust to the significance levels used (1%, 5%, or 10%) and clearly underlines the importance of explicitly taking cross-section dependence into account through using appropriate (bootstrap) critical values during the research.

Interestingly, and as the two last panel sets essentially start at the end of the 1990s, the evidence regarding the existence of a long-run relationship between government revenue and expenditure is pretty much in line with the results obtained by Afonso and Rault (2010) for the EU15, for the sub-period of 1992–2006 (even if it was for a smaller set of countries). Such results can be evidence of an increased awareness by policymakers after that period, of the need to redress public finances, in order to prepare for Stage Three of the Economic and Monetary Union, which started on the 1st of January, 1999.

Moreover, we welcome the fact that the hypothesis of fiscal sustainability is not rejected in the EU, as this puts into perspective the eventual individual difficulties of carrying out decentralised fiscal policies. Indeed, this is notably important for two reasons. Firstly, the degree of government indebtedness in the EU (or in the Euro area) is no bigger than, say, the case of the US or that of Japan. Secondly, in the context of the 2010–2011 European sovereign debt crisis, fiscal policy started being individually less decentralised, with European institutions taking more ownership of country-level fiscal policies. Therefore, the path to a more centralised fiscal policy implementation in the Euro area is reinforced with these results, which shows the absence of fiscal unsustainability in the overall EU panels.

4.3. Step 3: the magnitude of the cointegration relationship

We estimate for each country the cross-section augmented cointegrating regression:

$$R_{it} = \alpha_i + \beta_iG_{it} + \mu_{it}R_{it} + \mu_{2t}C_{it} + u_{it}, \quad i = 1, ..., N; \quad t = 1, ..., T \quad (8)$$

for the three panel sets, to assess the magnitude of the individual $\beta_i$ coefficient in the cointegrating relationship with the CCE estimation procedure, which was developed by Pesaran (2006), which addresses cross-sectional dependency. The regression includes the cross-section averages of the dependent variable and observed regressors as proxies for the unobserved factors. Accordingly, $\bar{R}_t$ and $\bar{G}_t$ respectively denote the cross-section averages of $R_{it}$ and $G_{it}$ in year $t$. The estimation results are reported in Table 4a, 4b, and 4c.

According to our estimation results, although the $\beta$ coefficient is always with the right sign, and almost always statistically significant, its magnitude is also below unity. Nevertheless, it seems fair to point out that the size of the $\beta$ coefficient is quite high in some cases, and consistently close or above 0.9, which is notable in the case of: Denmark, Germany, Portugal, the Netherlands, and the U.K. These results, which mostly hold for all three country panels, can be interpreted as an indicator that public finances may have been less unsustainable in the past for the above-mentioned countries.

<table>
<thead>
<tr>
<th>Table 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel cointegration test results between government revenue and expenditure (Westerlund and Edgerston, 2007)*.</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>EU14 (1960–2012)</strong></td>
</tr>
<tr>
<td><strong>Model with a constant term</strong></td>
</tr>
<tr>
<td><strong>EU22 (2000–2012)</strong></td>
</tr>
</tbody>
</table>

Note: the bootstrap is based on 2000 replications.

* The null hypothesis of the test is cointegration between government revenue and expenditure. EU14 excludes Germany, vis-à-vis the initial EU15 set; EU21 excludes Bulgaria, Estonia, Hungary, Poland, Slovakia, and Slovenia; EU22 excludes Bulgaria, Estonia, Hungary, Poland, and Slovakia.

On the other hand, it is also possible to observe the lower magnitude of the $\beta$ estimated coefficient for several countries, such as: Belgium, Greece, Ireland, Italy, Portugal and France, which reflects a bigger departure from a one-to-one linkage between expenditure and revenue in the cointegration relationship. Interestingly, and as a result of running significant budget deficits, most of these countries also experienced a divergent behaviour of their respective debt-to-GDP ratio during continued phases over the sample period, which would theoretically increase in infinite horizon, if the $\beta$ magnitude was to remain too distant and be below unity. Indeed, for most of the period, expenditure ratios were systematically above, and growing faster in some cases, than revenue ratios, in Belgium, Greece, Ireland, and Italy. Some of these countries also experienced important fiscal deteriorations in the aftermath of the 2007–2008 economic and financial crisis, and following on from the 2010 sovereign debt crisis.

On the other hand, in Finland and in Sweden, this difference was particularly acute during the first half of the 1990s (when the two countries faced important fiscal deteriorations).

The common correlated effects mean group (CCE-MG) method yields the following results (t-statistics are in parentheses):

**EU14 panel (1960–2012)**

$$R_{it} = 1.085 + 0.73G_{it} + 0.76R_{it} - 0.46C_{it}, \quad (7.85) \quad (3.98) \quad (6.45) \quad (5.53) \quad (8a)$$

**EU21 panel (1960–2012)**

$$R_{it} = 2.78 + 0.85G_{it} + 0.36R_{it} - 0.23C_{it}, \quad (10.27) \quad (4.98) \quad (7.47) \quad (4.75) \quad (8b)$$

**EU22 panel (1960–2012)**

$$R_{it} = 2.95 + 0.88G_{it} + 0.47R_{it} - 0.29C_{it}, \quad (12.01) \quad (6.02) \quad (8.32) \quad (3.38) \quad (8c)$$

We further investigate whether public finances were sustainable for the above models, using a Student test statistic, to test whether the panel cointegration coefficient of the general government expenditure-to-GDP ratio is equal to one in the cointegrating regression, where revenue is the dependent variable. Over the 1960–2012 period, and for the EU14 panel data set, the calculated Student test statistic is 1.47, with an associated p-value of 14.16%, which provides evidence in favour of the null of a common unit slope equal to one, at the 10% level of significance. Evidence of the sustainability of public finances is obtained for the EU21 and EU22 panel data set over the 1998–2012 and 2000–2012 periods, as the calculated Student test statistics for the above hypothesis are respectively 0.87 and 0.82, with their associated p-values being respectively 38.43%, and 41.22%.
4.4. Step 4: estimation of a panel ECM (PECM) representation

Having established the long-run structure of the underlying data in the previous step, and given that there exists a long-run relationship for all countries in our three panel sets, we now estimate the complete panel error-correction model (PECM) which is described by:

\[ \Delta R_{it} = \sum_{j=1}^{p} \Gamma_{ij} \Delta R_{i,t-j} + \sum_{j=0}^{p} \mathbf{G}_{j} \Delta y_{i,t-j} + \lambda_{i} R_{i,t-j-1} - \alpha - \beta \Delta y_{i,t-j} + \epsilon_{it}. \]  

(9)

\( i = 1, \ldots, N; t = 1, \ldots, T. \)

We use the Pooled Mean Group (PMG) approach of Pesaran et al. (1999), with long-run parameters obtained with CCE techniques, in order to obtain the estimates of the \( \lambda_{i} \) loading factors (weights or error correction parameters, or speed of adjustment to the equilibrium values), as well as those of the \( \Gamma_{ij} \) and \( \mathbf{G}_{j} \) short-run parameters for each country in our panel. Consequently, the loading factors and short-run coefficients are allowed to differ across countries.\[^{16}\]

The \( p \) lag length structure is chosen using the Schwarz (SC) and Hannan–Quinn (HQ) selection criteria, and by carrying out a standard likelihood ratio testing-down type procedure to examine the lag significance from a long-lag structure (started with \( p = 5 \)), to a more parsimonious one. Afterwards, in order to improve the statistical specification of the model, we systematically implement Wald tests of exclusion of lagged variables from the short-run dynamic (which are not reported here), as a means of eliminating insignificant short-run estimates at the 5% level. We tested the residuals from each PECM model for the absence of heteroscedasticity, autocorrelation and ARCH effect, and consequently can report that they are not subject to misspecification. The results of the PECM estimations based on Eq. (9) for the three panel sets are reported in Tables 5a,5b, and 5c, but only for significant short-run estimates at the 5% level.

Tables 5a,5b, and 5c show that there is mostly a short-run positive response of government revenue to government spending in the EU14 panel set, with changes in the revenue ratio, following on from contemporaneous changes in the spending ratio. One exception is Finland, where only the lagged response is not statistically positive.

For the more country comprehensive panel sets, the results are broadly similar, with a stronger short-run contemporaneous responsiveness of government revenues to government spending in the cases of Denmark, France, Germany, Ireland, the Netherlands, Spain and the

\[^{16}\] Note that before considering Eq. (9), we first used a Wald statistic to test for common parameters across countries (i.e., \( \lambda_{i} = \lambda \), and \( \beta_{i} = \beta \), for \( i = 1, \ldots, N \)) with the Pesaran (2006) CCE techniques which allow common factors in the cross-equation covariances to be removed. We also found that only the \( \beta_{i} = \beta \), for \( i = 1, \ldots, N \) was not rejected by the data, whereas speeds of \( \lambda_{i} \) adjustment vary considerably across countries (results are available upon request).
of Belgium, Greece, Italy, Portugal, and Sweden, bringing attention again to some unwelcome fiscal dynamics for these countries. These results are similar to the conclusions of the previous section, where we estimated long-run parameters and cointegration relationships.\(^{17}\) Therefore, in this final step of our sustainability analysis, we can gain insight into both the existence of a long-run relationship between the two sides of the general government budget, and also, of short-run dynamics ruling fiscal behaviour across these countries, which is in addition to the existing results.\(^{18}\)

\(^{17}\) We also performed a robustness check for the three panel sets and estimated panel error-correction models for \(R_t\) in line of Eq. (9), including time-specific intercepts \(\alpha_t\), in order to control for possible common macroeconomic shocks affecting all countries at period \(t\). Our short and long-run estimated parameters, which are available on request, are unchanged qualitatively.

\(^{18}\) Note that on account of data availability, the samples considered for new entrants are quite short (post-1998 or post-2000). Therefore, as underlined by a referee, these samples may not temporally be long enough to identify clear short-run or long-run behaviours for these countries.
5. Conclusion

Even in the absence of a single fiscal policy for the EU, panel analysis of fiscal sustainability is certainly relevant in the context of 27 EU countries that seek to follow sound fiscal policies within the framework of the Stability and Growth Pact. In this paper, starting from the intertemporal government budget constraint, and taking advantage of non-stationary panel data econometric techniques, we assessed the sustainability of public finances covering several sub-periods within the period of 1960–2012 and also defined different country groupings for the 27 members of the EU.

More specifically, we proposed and implemented an original multi-step analysis comprising: (i) SURADF panel integration analysis (which tests for individual unit root within the panel member, and handles heterogeneous serial correlation across panel members), which we believe is the first empirical application in the context of the sustainability of public finances; (ii) panel bootstrap, to test the null hypothesis of cointegration between expenditure and revenue ratios; (iii) testing for cross-section dependencies (potentially arising from multiple unobserved common factors); and, (iv) panel error correction models of the revenue ratios, where short-run and long-run effects are estimated jointly from a general autoregressive distributed-lag model, and where the short-run effects vary across countries. This multi-step approach takes advantage of the increased power of panel techniques, and also provides specific information regarding how far a given country has been from fiscal sustainability in the past.

According to the results, notably with regard to the SURADF test, general government revenue and expenditure-to-GDP ratios are not stationary for the overwhelming majority of the EU27 countries. Additionally, at the conventional 5% level of significance, we can also conclude that there is a cointegrating relationship between government revenue and expenditure ratios for the EU14 panel data set over the period of 1960–2012. A similar conclusion is found for the existence of a cointegration relation for the country panel sets, which includes the more recent members of the EU: the EU21 for the period of 1998–2012 and the EU22 for the period of 2000–2012.

Moreover, even though we identified a cointegration vector for all countries, the estimated coefficient for expenditures in the cointegration equations is always less than one. In other words, for the overwhelming majority of the EU27 countries. Additionally, at the conventional 5% level of significance, we can also conclude that there is a cointegrating relationship between government revenue and expenditure ratios for the EU14 panel data set over the period of 1960–2012. A similar conclusion is found for the existence of a cointegration relation for the country panel sets, which includes the more recent members of the EU: the EU21 for the period of 1998–2012 and the EU22 for the period of 2000–2012.

Table 5b
Panel error-correction estimations for Rit for the EU21 panel (1960–2012).

<table>
<thead>
<tr>
<th>Country</th>
<th>Δ R_{i,t-1}</th>
<th>Δ R_{i,t-2}</th>
<th>Δ C_{i}</th>
<th>Δ C_{i,t-1}</th>
<th>Δ C_{i,t-2}</th>
<th>Loading factor λ</th>
<th>Intercept</th>
<th>C_{i,t-1}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>0.24</td>
<td>−0.23</td>
<td>0.11</td>
<td>−0.26</td>
<td>−0.49</td>
<td>2.47</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>Czech</td>
<td>0.14</td>
<td>0.05</td>
<td>0.04</td>
<td>−0.09</td>
<td>0.17</td>
<td>2.42</td>
<td>0.71</td>
<td></td>
</tr>
<tr>
<td>Cyprus</td>
<td>2.17</td>
<td>1.46</td>
<td>−2.87</td>
<td>−2.22</td>
<td>−1.58</td>
<td>2.43</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td>0.29</td>
<td>0.42</td>
<td>0.26</td>
<td>0.08</td>
<td>0.22</td>
<td>2.45</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>Finland</td>
<td>0.27</td>
<td>(3.10)</td>
<td>0.05</td>
<td>−0.04</td>
<td>0.07</td>
<td>2.45</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>−0.21</td>
<td>−0.24</td>
<td>0.39</td>
<td>−0.04</td>
<td>0.27</td>
<td>2.45</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>Ireland</td>
<td>(−2.24)</td>
<td>(−2.87)</td>
<td>(3.76)</td>
<td>(−2.98)</td>
<td>(2.21)</td>
<td>(3.34)</td>
<td>(9.76)</td>
<td></td>
</tr>
<tr>
<td>Italy</td>
<td>0.16</td>
<td>0.17</td>
<td>0.29</td>
<td>0.17</td>
<td>0.19</td>
<td>2.45</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>Latvia</td>
<td>0.29</td>
<td>0.29</td>
<td>0.17</td>
<td>0.19</td>
<td>0.31</td>
<td>2.45</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>Luxembourg</td>
<td>0.29</td>
<td>0.29</td>
<td>0.18</td>
<td>0.22</td>
<td>0.07</td>
<td>2.45</td>
<td>0.77</td>
<td></td>
</tr>
<tr>
<td>Malta</td>
<td>0.37</td>
<td>0.36</td>
<td>0.08</td>
<td>0.22</td>
<td>0.10</td>
<td>2.44</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.37</td>
<td>0.37</td>
<td>0.23</td>
<td>0.23</td>
<td>0.23</td>
<td>2.46</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>Portugal</td>
<td>−0.08</td>
<td>−0.08</td>
<td>0.01</td>
<td>0.22</td>
<td>0.02</td>
<td>0.10</td>
<td>2.44</td>
<td></td>
</tr>
<tr>
<td>Romania</td>
<td>0.11</td>
<td>0.05</td>
<td>0.56</td>
<td>−0.09</td>
<td>0.25</td>
<td>0.09</td>
<td>2.44</td>
<td></td>
</tr>
<tr>
<td>Spain</td>
<td>0.35</td>
<td>0.35</td>
<td>0.23</td>
<td>−0.04</td>
<td>0.09</td>
<td>2.44</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td>0.28</td>
<td>0.10</td>
<td>0.17</td>
<td>−0.25</td>
<td>0.35</td>
<td>2.44</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>UK</td>
<td>0.43</td>
<td>0.45</td>
<td>0.29</td>
<td>0.29</td>
<td>0.09</td>
<td>2.42</td>
<td>0.73</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The estimations are obtained from the Pooled Mean Group approach with long-run parameters estimated with CCE techniques. The coefficients of the variables \( \Delta R_i \) and \( \Delta C_{i,t} \) of Eq. (8b) have not been reported in the table. t-statistics are in brackets.
in order to avoid encountering unpleasant fiscal sustainability pitfalls in the future.

Overall, while the results imply that public finances were not unsustainable for the EU panels, some country diversity emerged. Indeed, while a higher degree of fiscal sustainability was found in the cases of Denmark, Germany, the Netherlands, and Spain, somewhat opposite conclusions were discovered for Belgium, Greece, Italy, Portugal, and Sweden. Such behaviour of the respective budgetary components naturally helped bring about increasing fiscal sustainability for these countries in the past.

Acknowledgements

We are grateful to two anonymous referees for their helpful comments and suggestions on earlier drafts of this paper, and to participants in the 68th European meeting of the Econometric Society, at the CESifo Area Conference on Macro, Money and International Finance for their helpful comments, and also to those from the 18th ICMAIF Conference.

The opinions expressed herein are those of the authors, and do not necessarily reflect those of the ECB or Eurosystem.

References


<table>
<thead>
<tr>
<th>Country</th>
<th>$\Delta R_{it-1}$</th>
<th>$\Delta R_{it-2}$</th>
<th>$\Delta G_{it}$</th>
<th>$\Delta G_{it-1}$</th>
<th>$\Delta G_{it-2}$</th>
<th>Loading factor $\lambda_1$</th>
<th>Intercept</th>
<th>$c_{it-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>0.17 (2.45)</td>
<td>−0.22 (−2.23)</td>
<td>0.19 (3.33)</td>
<td>−0.176 (−2.78)</td>
<td>(−4.17) (11.68)</td>
<td>2.44 (11.68)</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>Belgium</td>
<td>0.07 (2.56)</td>
<td>0.27 (3.92)</td>
<td>0.07 (3.92)</td>
<td>−0.09 (−2.27)</td>
<td>(11.01) (4.46)</td>
<td>2.40 (4.46)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyprus</td>
<td>2.08 (12.54)</td>
<td>1.34 (5.45)</td>
<td>−2.72 (−11.22)</td>
<td>−2.17 (−6.13)</td>
<td>(−3.48) (10.87)</td>
<td>2.40 (4.08)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Czech Republic</td>
<td>0.17 (2.67)</td>
<td>0.06 (−1.32)</td>
<td>−0.13 (3.14)</td>
<td>(−3.96) (11.66)</td>
<td>(4.78)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denmark</td>
<td>0.27 (2.12)</td>
<td>0.33 (−0.53)</td>
<td>0.26 (2.75)</td>
<td>−0.26 (−3.65)</td>
<td>(10.88) (4.45)</td>
<td>2.41 (4.76)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finland</td>
<td>0.26 (2.32)</td>
<td>0.07 (−0.21)</td>
<td>−1.04 (−3.12)</td>
<td>(7.21) (11.00)</td>
<td>(4.08)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>France</td>
<td>0.23 (4.43)</td>
<td>0.36 (−0.13)</td>
<td>0.18 (3.77)</td>
<td>(11.76) (4.26)</td>
<td>2.45 (4.76)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>0.19 (2.66)</td>
<td>0.28 (3.97)</td>
<td>0.36 (−0.06)</td>
<td>(10.96) (4.13)</td>
<td>(4.76)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greece</td>
<td>0.12 (3.98)</td>
<td>−0.06 (−3.13)</td>
<td>0.98 (−0.27)</td>
<td>(11.78) (4.12)</td>
<td>(4.76)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ireland</td>
<td>−0.19 (−2.67)</td>
<td>−0.25 (−3.07)</td>
<td>0.37 (3.56)</td>
<td>−0.09 (−4.08)</td>
<td>(2.94) (−3.54)</td>
<td>(11.78) (4.12)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Italy</td>
<td>0.16 (4.43)</td>
<td>−0.09 (−3.48)</td>
<td>0.16 (3.45)</td>
<td>(11.160) (4.78)</td>
<td>2.43 (4.76)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latvia</td>
<td>0.19 (3.45)</td>
<td>−0.13 (−4.08)</td>
<td>0.19 (3.45)</td>
<td>(11.167) (4.45)</td>
<td>(4.76)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lithuania</td>
<td>−0.20 (7.78)</td>
<td>−0.46 (−20.22)</td>
<td>0.28 (9.17)</td>
<td>0.19 (6.88)</td>
<td>(9.43) (−2.98)</td>
<td>(11.56) (4.89)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luxembourg</td>
<td>0.34 (2.45)</td>
<td>−0.18 (−3.23)</td>
<td>0.34 (2.45)</td>
<td>(−2.92) (10.98)</td>
<td>(4.09)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malta</td>
<td>0.32 (3.76)</td>
<td>0.09 (−0.24)</td>
<td>0.32 (3.76)</td>
<td>(11.88) (4.56)</td>
<td>(4.76)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.32 (3.55)</td>
<td>2.43 (−4.92)</td>
<td>0.32 (3.55)</td>
<td>(11.36) (4.44)</td>
<td>(4.76)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Portugal</td>
<td>−0.08 (−2.67)</td>
<td>−0.52 (−3.76)</td>
<td>0.03 (2.09)</td>
<td>0.23 (2.65)</td>
<td>(−2.98) (10.88)</td>
<td>(4.27)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Romania</td>
<td>0.11 (2.77)</td>
<td>−0.32 (−9.04)</td>
<td>0.53 (14.34)</td>
<td>−0.08 (−2.65)</td>
<td>(7.88) (6.87)</td>
<td>(11.54) (4.55)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slovenia</td>
<td>−1.00 (−21.34)</td>
<td>−0.57 (−17.47)</td>
<td>0.9 (8.87)</td>
<td>0.90 (17.14)</td>
<td>(14.87) (12.01)</td>
<td>(11.54) (4.45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spain</td>
<td>0.32 (3.19)</td>
<td>−0.04 (−4.68)</td>
<td>0.32 (3.19)</td>
<td>(12.09) (4.89)</td>
<td>(4.89)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweden</td>
<td>0.35 (2.54)</td>
<td>0.09 (−0.16)</td>
<td>0.08 (2.54)</td>
<td>−0.16 (−2.65)</td>
<td>(−3.12) (−4.54)</td>
<td>(11.22) (4.12)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UK</td>
<td>0.48 (4.09)</td>
<td>0.44 (−0.42)</td>
<td>0.48 (4.09)</td>
<td>−0.29 (−4.15)</td>
<td>(11.01) (4.89)</td>
<td>(4.89)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: The estimations are obtained from the Pooled Mean Group approach with long-run parameters estimated with CCE techniques. The coefficients of the variables $\Gamma_t$ and $\lambda_0 t$ of Eq. (8c) have not been reported in the table. t-statistics are in brackets.

Table Sc
Panel error-correction estimations for $R_{it}$ for the EU22 panel (1960–2012).


