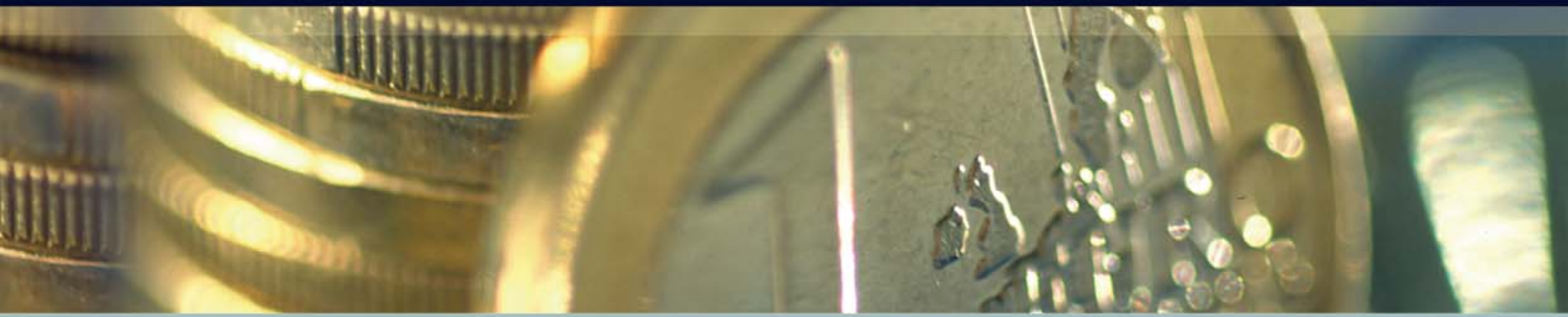
EUROPEAN ECONOMY

Economic Papers 300 | February 2008



Government expenditure and economic growth in the EU: long-run tendencies and short-term adjustment

Alfonso Arpaia, European Commission
Alessandro Turrini, European Commission and CEPR

Economic and Financial Affairs
DIRECTORATE-GENERAL

EUROPEAN COMMISSION

EMU@10 Research

In May 2008, it will be ten years since the final decision to move to the third and final stage of Economic and Monetary Union (EMU), and the decision on which countries would be the first to introduce the euro. To mark this anniversary, the Commission is undertaking a strategic review of EMU. This paper constitutes part of the research that was either conducted or financed by the Commission as source material for the review.

Economic Papers are written by the Staff of the Directorate-General for Economic and Financial Affairs, or by experts working in association with them. The Papers are intended to increase awareness of the technical work being done by staff and to seek comments and suggestions for further analysis. The views expressed are the author's alone and do not necessarily correspond to those of the European Commission. Comments and enquiries should be addressed to:

European Commission
Directorate-General for Economic and Financial Affairs
Publications
B-1049 Brussels
Belgium
E-mail: Ecfin-Info@ec.europa.eu

This paper exists in English only and can be downloaded from the website
http://ec.europa.eu/economy_finance/publications

A great deal of additional information is available on the Internet. It can be accessed through the Europa server (http://europa.eu)

ISBN 978-92-79-08225-2
doi: 10.2765/22776

Government expenditure and economic growth in the EU: long-run tendencies and short-term adjustment

Alfonso Arpaia* and Alessandro Turrini**

Abstract:

This paper analyses both the long and the short-run relation between government expenditure and potential output in EU countries by means of pooled mean group estimation (Pesaran, Shin, and Smith (1999)). Results show that, over a sample comprising EU-15 countries over the 1970-2003 period, it cannot be rejected the hypothesis of a common long-term elasticity between cyclically-adjusted primary expenditure and potential output close to unity. However, the long-run elasticity decreased considerably over the decades and is significantly higher than unity in catching-up countries, in fast-ageing countries, in low-debt countries, and in countries with weak numerical rules for the control of government spending. The average speed of adjustment of government expenditure to its long-tem relation is 3 years, but there are significant differences across countries. Anglo-Saxon and Nordic countries exhibit in general a faster adjustment process, while adjustment in Southern European countries appears somehow slower.

JEL Classification: E62, H50, C23

Key words: Government expenditure, Wagner's law, panel co integration

*European Commission

**European Commission and CEPR

The paper benefited from discussions with Martin Larch and Massimiliano Marcellino.

1. Introduction

This paper analyses the relation between government expenditures and economic growth in the EU. It focuses on three questions. By how much government expenditures change with GDP in the long-run and by how much in the short run? Is the relation between government expenditures and GDP robust over time? Is it significantly different across countries?

Better knowledge on the dynamic relation ship between government expenditure and GDP is relevant for policy in two major respects.

First, it improves the understanding of long-term, structural public finance issues. Is the size of government shrinking or expanding in the EU? Are long-term trends in the size of government similar across countries or there are relevant differences? Answering these questions is relevant for the debate on the sustainability of public finances in Europe. In particular, it could help to assess the impact on government expenditures and then on deficits arising from a structural deceleration in growth (e.g., associated with ageing populations or a decline in TFP growth) or, conversely, from an improvement in the growth potential (e.g., related to structural reforms).

Second, a better understanding of the dynamic relation between government expenditure and GDP helps the comprehension of policy-relevant issues over a short-to medium term horizon. Disposing of a reliable measure of the structural relation between the non-cyclical component of government expenditure and potential output is key to obtain a benchmark against which to evaluate the stance of expenditure policy and then of overall fiscal policy. Judging whether expenditure policy is expansionary or contractionary requires some idea about how a neutral expenditure policy would look like. However, while there is broad consensus that a neutral revenues policy is such that government revenues move together with output in a proportion depending on structural factors such as the degree of progression of the tax system and the responsiveness of the various tax bases with respect to output (the output elasticity of revenues), no clear a-priori exists for what concerns expenditure policy.[1] Estimating the long-

[1] In policy analysis, a constant primary cyclically adjusted budget balance is often taken as an indication of a neutral fiscal policy stance. This implies that expenditure policy is neutral as long as non cyclical primary expenditures grow in line with non-cyclical revenues. However, one may want to analyse separately the stance of revenue and expenditure policy, and this may require a different notion of neutral expenditure policy. Buti and Van den Noord (2003) adopt a definition of neutral expenditure policy according to which primary government expenditures grow in line with potential output plus expected inflation. Fatàs et al. (2003) and Hughes-Hallet et al. (2004) resort to three different definitions of 'neutral fiscal policy': government spending is held constant in volume terms; government expenditures grow in line with revenues; government expenditures grow in proportion with trend GDP. Moreover, Gali and Perotti (2003), among others, consider a broader concept of "non-discretionary" fiscal policy, obtained as the residual of an estimated fiscal reaction function where the primary cyclically-adjusted budget balance is regressed against its own lag, the lagged debt/GDP ratio and a measure of the output gap.

term relation between government expenditure and GDP permits to formulate a benchmark for neutral expenditure policy grounded on empirical evidence. Useful information for policy-making would also be provided by estimates of the speed at which government expenditure adjust to their long-term relation with GDP after a shock in economic activity. In the EU context, this information would be helpful, for instance, in formulating and assessing budgetary adjustment plans with a view to achieving medium term budgetary objectives or correcting deficits in excess of the 3% Maastricht reference value for the deficit.

This paper builds on the existing literature studying the long-term determinants of government expenditure and makes a step forward in two respects. First, there is an attempt to better disentangle cyclical from structural factors affecting the relation between government expenditure and GDP. Second, the panel dimension of the data set is exploited in such a way: (i) to improve the power of statistical tests for the analysis of the dynamic properties of macroeconomic series through panel unit root and cointegration tests; (ii) to obtain country-specific information on adjustment dynamics by means of pooled mean group estimation.

There is consensus that a relatively minor part of government spending, typically unemployment benefits, is a purely cyclical phenomenon, so that changes in the level of output matter only to the extent that the cyclical slack in the economy is affected. [2] A different and more complex issue is the one addressed in this paper, namely, how non-cyclical expenditures may be linked to non-cyclical movements in output over time. The empirical literature has tackled this issue from different corners. A branch of the literature investigates the determinants of the size of government across countries, focusing on alternative explanations such as per-capita income (e.g., Peltzman (1980), Borcherding (1985)), the relative price of government-provided goods and services (Baumol (1967)) demographic structures (Heller and Diamond, 1990), the size (Alesina and Wacziarg (1998)) or the degree of openness of the economy (Rodrik, 1998).[3] A growing strand of research aims at explaining cross-country structural differences in the size of government on the basis of political fundamentals that shape the extent of the deficit bias related with free-riding in government expenditure provision and governments' myopia. It has been shown that the size of government tends to be larger in parliamentary than in presidential regimes (Persson and Tabellini (2000)) and that countries with proportional electoral rules are characterized by

[2] See, e.g., van den Noord (2000) for an estimation of the elasticity of government expenditure to the cycle in OECD coountries. See also Bouthevillan et. al. (2001) for an empirical assessment of expenditure elasticities for the EU-15 countries based on an alternative methodology.
[3] Reviews of the findings in this strand of literature are provided, for instance, in Peltzman (1980), Borcherding (1985)), Tanzi and Schuknecht (2000).

higher government expenditure shares on GDP than countries with majoritarian election (Persson, Roland, and Tabellini (2006)) and by government expenditure tilted towards transfers rather than purchases of goods and services (Milesi-Ferretti, Perotti and Rostagno (2002)). It has also been shown that the fiscal performance of countries is affected by the way budgetary processes are structured (e.g., Von Hagen and Harden (1995), Hallerberg, Strauch and von Hagen (2001)).

A second strand of literature examines the link between expenditure and economic growth over time. Some work aims at describing long-term tendencies in history (Tanzi and Scuckencht (2000)). Other work is more specifically focused at the empirical estimation of elasticity of government expenditure with respect to output, often with the explicit aim of providing an empirical test of the so-called "Wagner law", i.e., the hypothesis that government expenditure increases more than proportionally with economic activity. The underlying idea is that goods and services generally provided by the government sector, including redistribution via transfers and the activities of public enterprises, have an income elastic greater then one, i.e., are superior goods. This last strand of studies includes the empirical analyses most closely related to that provided in this paper.

The Wagner law has been tested in different ways. In early time series analyses, government expenditure is regressed on GDP without taking into account the dynamic properties of the series (e.g., Ram, 1987).[4] More recently, new test specifications have been implemented taking into consideration non-stationarity and co-integration. This allows for a more structured modelling of expenditure dynamics introducing the distinction between a long-term relation ship and short term adjustment. Kolluri et al (2000), Akitoby et al. (2004) and Wahab (2004) are among the most recent cross-country analysis allowing for dynamic specifcations.

Based on country-specific single equation models, Kolluri et al. (2000) investigate G7 countries over the 1960-1993 period. They find that government expenditure is generally cointegrated with income, that the long-term income elasticities of government expenditure is slightly above unity in all countries both for government consumption and government transfers, and that short-term elasticities differ widely across countries and average around 0.5 (implying about 1/0.5=2 years for government spending to return to its long-term relation with GDP). Akitoby et al. (2004), focus on a set of developing countries between 1970 and 2002. Unit root and cointegration tests on individual country series reveal that government

[4] See, e.g., Bohl (1996), Payne and Ewing (1996), Chang (2002) for reviews on empirical studies on the Wagner's law.

expenditure is often cointegrated with income; country-level ECM estimation yields long-term income elasticities on average slightly above unity and short-term elasticities on average around 0.3. Wahab (2004) analyses a group of OECD countries over the 1950-2000 period. In this paper, individual country series are checked for unit roots and panel estimations for ECM specifications are performed for alternative country groupings. It is found that over the whole sample government expenditure increases less than proportionately with income (long-term income elasticity slightly below unity). The same result is obtained by limiting the sample only to EU countries. The response of government expenditure to GDP, however, is found to be asymmetric. While government expenditure increases less than proportionally when growth is below trend, it falls more than proportionally when growth is below trend.

The approach followed in this paper differs from previous work in two major respects.

First, by using cyclically-adjusted figures for both GDP and government expenditure, we manage to better disentangle short-term dynamics related to business cycle fluctuations and to concentrate the analysis on relations of structural nature. This also permits to contain the issue of reverse causation in interpreting results. Since the impact of government expenditure on GDP is mostly cyclical (the effect of government expenditure on potential output is associated with the composition rather than with the size of government expenditure. We interpret the relation between primary cyclically adjusted expenditure and potential output as reflecting the adaptation of expenditure to a changing size of the economy.[5] As the economy grows, governments need to face growing demand for public goods and services, adapt wages and salaries of government employees to meet higher remunerations in the private sector, revise the degree of generosity of transfers. This process of adaptation of government expenditure to changing potential output may take time. The possibility of distinguishing the long-run from the short-run impact of potential GDP on government expenditure in our empirical analysis permits to measure the speed at which this process of adaptation of government expenditure to a new value for potential output takes place.

Second, in our analysis we exploit as far as possible the variation both over time and across countries in our data set. This permit to improve the power of estimates when the number of observations over time is small, which is normally the case when analysing structural public finance issues with yearly data. Indeed, it has been shown that inference on the time series properties of the data can be improved upon when applying integration and cointegration tests

[5] See, for instance the results from the meta-analysis on empirical work aimed at testing the impact of government activity on growth by Nijkamp and Poot (2004).

to the whole panel rather than to each time series separately (see, e.g., Baltagi and Kao (2000), Phillips and Moon (2000), Smith (2000)). Furthermore, by resorting to Pooled Mean Group estimates (Pesaran, Shin, and Smith (1999)) we manage to extract country-specific information on adjustment coefficients while improving upon the precision of the estimates compared with the alternative of analysing each country separately.

Our analysis comprises EU-15 countries over the 1970-2003 period. Data on primary cyclically adjusted government expenditure and potential output are taken from the EU Commission AMECO database. The main results can be summarised as follows. The long term elasticity of government expenditure with respect to GDP across the countries in our panel is slightly below unity, meaning that expenditure is linked to potential output by roughly a one-to-one relationship. The average speed of adjustment of government expenditure to its long-tem relation is 3 years, but there are significant differences across countries. Anglo-Saxon and Nordic countries exhibit in general a faster adjustment process, while adjustment in Southern European countries appears somehow slower. Estimates of the long-term elasticity of expenditure are fairly robust over time and across countries, being the hypothesis of equal elasticity across countries accepted at standard confidence levels. However, there is evidence of a significantly higher elasticity in countries characterized by low initial per capita GDP, relatively fast ageing, low government debt/GDP ratios and weak numerical rules for the control of epxenditure.

The remainder of the paper is structured as follows. Section 2 illustrated the empirical strategy followed, describes the data set and provides prima-facie descriptive statistics. Unit root and cointegration tests are performed in section 3. Section 4 presents the estimation of the dynamic relationship between government expenditure and potential output and discusses the results concerning long and short-term elasticities. Section 5 deals with robustness issues, both with respect to different sub-periods and different grouping of countries. Section 6 discusses the policy implications of results. The concluding remarks follow.

2. Empirical strategy and data

2.1. Empirical strategy

Our aim is that of establishing the dynamic properties of the relationship between government expenditures and GDP in the EU-15 countries over the 1970-2003 period. In particular, we are interested in the following questions. Are government expenditures and potential output linked by a stable long-run relationship? Is the long-term elasticity between government expenditure and potential GDP greater than one, as predicted by the Wagner law? Do countries share the same long-run elasticity or are there considerable differences? Is the relation robust over time? What is the value of the speed at which expenditure adjusts to the level predicted by the long-run relationship with potential output? Are there relevant cross-country differences in the speed of adjustment?

Rather than exploring the relation between economic activity and various definitions of subcategories of government expenditure as in other papers, we focus on overall primary expenditure. Although the dynamics of different categories of government expenditure are undoubtedly explained by different determinants, we concentrate our attention on a broad expenditure aggregate because of two main reasons. First, what matters for the determination of government deficit and debt, and ultimately for the overall sustainability of public finances is overall government expenditure. Second, existing work analysing separately different government expenditure categories via the estimation of dynamic equations does not find evidence of a strongly different relation with economic activity across types of expenditure (e.g., Kolluri et al. (2004), Akitoby et al. (2004)).

To overcome the issue of spurious regression that characterized earlier studies on the relation between government expenditure and GDP due to the neglect of the time series properties, we follow the now standard three-step approach consisting of (i) assessing the stationarity of the time series, (ii) in case the variables are not stationary, checking whether they are characterized by a cointegration relationship, (iii) in case cointegration holds, estimating error correction mechanism (ECM), which permits to analyse the long-run relationship between the variables jointly with the short-term adjustment towards the long-run equilibrium.

We abstract from cyclical considerations, by using cyclically-adjusted figures for both government expenditure and GDP. More precisely, the series used in our analysis are, respectively, primary cyclically-adjusted government expenditure and potential output. This has two major advantages. First, it permits to disentangle from the analysis business cycle gyrations and to concentrate the analysis on relations of structural nature, unrelated to interest rate shocks or to the "automatic" response of government expenditure to cyclical conditions (associated with the working of automatic stabilizers like unemployment benefits and subsidies). Second, abstracting form cyclical dynamics helps to reduce the issue of reverse causation in interpreting results. Government expenditure reacts to changing potential output as a result of the adaptation of the public sector to a modified size of the economy. However, it also true that shocks to government expenditure translate into aggregate demand and then changed GDP levels, i.e., it is difficult to disentangle a priori whether the relation between government expenditure and GDP goes from the latter to the former or vice-versa. To the extent that the impact of government expenditure on GDP is mostly in terms of aggregate demand impulse rather than changed output potential, focusing the analysis on figures adjusted for the cycle contributes to contain the issue of reverse causality.[6] Hence, in our context, the use of cyclically-adjusted variables implies that the temporary deviations from the long-run relationships do not reflect the evolution of the business cycles, but rather temporary deviations due to a lagged response of fiscal authorities in adjusting expenditure to changes in potential output

We aim at exploiting as far as possible both the time series and the cross-section (i.e., across countries) properties of the data. This has the major advantage of improving the statistical properties of estimates when the number of observations over time is limited, which is typically the case when analysing structural public finance issues with yearly data. Indeed, in estimating and testing the stochastic properties of time series with "small" sample sizes one has to face the well-known risks of low power of stationarity and cointegration tests. To circumvent these problems, the recent literature on non-stationary panel data has concluded that inference on the time series properties of the data can be improved upon when applying integration and cointegration tests to the whole panel rather than to each unit separately.[7] Opting for panel estimates as opposed to repeated time-series estimates for each of the country in the sample has the advantage of improving the power of the statistical tests but at the cost of losing country-specific information. We use the recently introduced Pooled Mean

[6] An impact of government expenditure on potential output cannot be excluded. However, the effect can be opposite depending on which type of expenditures are considered. While government investment or public education expenditures are likely to improve the growth potential, other types of expenditure may reduce growth by crowding out resources to private investment (Kneller, Bleaney, and Gemmell, (1999)). Moreover, Levine and Renelt (1992) show that fiscal variables are generally non robust when included in cross-country growth regressions. See also the main conclusions of the meta-analysis on cross-country and panel empirical studies on fiscal policy on growth by Nijkamp and Poot (2004).

[7] When there are similarities between the data generation processes of cross sectional units, some form of averaging may improve the power of unit roots tests and the consistency of cointegrating relationships - i.e. the potential risks of spurious regression are largely reduced (the cross-section dimension can in fact be considered as repeated draws from the same distribution). Hence, independent cross sections in the data add more information and lead to a strong signal than that of the pure time series. Averaging over units (individuals, regions or countries) attenuates the noise of the least square estimates due to the covariance of two independent random variables. For a survey of the literature on non-stationary panel data see for instance Banerjee (1999), Baltagi and Kao (2000), Phillips and Moon (2000), and Smith (2000).

Group estimators (PMG) that allows for country-specific adjustment coefficients in panel estimation but pool countries over the long-run (Pesaran, Shin, and Smith (1999)).

The empirical analysis in the remainder of the paper proceeds as follows. First, we give a description of the data set employed and inspect by means of graphical analysis the dynamic behaviour of government expenditure and potential output.[8] Second, panel unit root tests are performed to assess whether the variables we use in the analysis are stationary. Third, the existence of a long-run relationship between cyclically adjusted primary expenditure and potential output is verified by means of the residual-based Pedroni (1999) panel cointegration tests. Fourth, the dynamic relation between government expenditure and GDP is analysed empirically by means of testing an error correction mechanism (ECM) with the PMG estimator. Fifth, the robustness of results is discussed with respect to different time sub-periods in the sample and different groupings of countries.

2.2. Data

To investigate the relationship between public expenditure and GDP, we use yearly observations from 1970 to 2003 for the EU-15 countries. Data are taken from the Annual Macroeconomic (AMECO) database of the European Commission. All data are expressed at constant 1995 prices and denominated in common currency (ECU). Expenditure data are net of interest expenditure and are adjusted for the cycle. GDP data refer to potential GDP. Potential GDP series are obtained by means of the production function approach, i.e., potential output is estimated starting from an assumed aggregate production function for the economy and estimates of the capital stock, labour inputs and total factor productivity (see Denis et al. (2002)). The government expenditure/GDP ratio is adjusted for the cycle following the approach used by the European Commission, i.e., by deducting a measure of "cyclical" government expenditure consisting of a country-specific expenditure "sensitivity" parameter multiplied by the output gap.[9] The sensitivity of expenditure to the cycle captures the monetary change in expenditure associated with a unit monetary change in the difference between actual and potential output as a result of the operation of existing legislation (automatic stabilizers). Sensitivity parameters are constructed on the basis of budgetary elasticities estimated in Van den Noord (2001). Unemployment subsidies is the only

[8] Henceforth, in the paper we will use, for brevity, the terms "government expenditure" to refer to primary cyclically-adjusted primary government expenditure.
[9] For the years 1970-1978, cyclically adjusted expenditure/GDP ratio based on potential output is not available for Luxembourg; for these years the cyclical adjustment of primary expenditure is based on trend GDP.

government expenditure component assumed to react “automatically” to the cycle. [10] To get the cyclically adjusted expenditure net of interest spending, the expenditure/GDP ratio is multiplied by the GDP at current prices, and finally deflated.

Figure 1 plots the series for each country in log scale. Despite the apparent common positive relation between expenditures and potential output in all countries over the long-run, there are notable differences over time periods and across countries. While in most countries during the 1970s the growth rate of government expenditures outpaced that in potential output, starting from the 1980s it is observed a generalised deceleration in expenditure. In some countries, government expenditure decelerated already during the 1980s as a result of a general restructuring of the government sector (UK) or as a consequence of expenditure-based consolidations carried out to stabilize debt-GDP ratios (Belgium, Denmark, Ireland). In other countries, the downward adjustment in the growth rate of expenditure was enacted in the 1990s, in some cases with the express objective to achieve the respect of the deficit Maastricht criterion in the run-up to EMU (Spain, Italy). Overall, in the 1980s government expenditure kept growing at lower pace compared with the 1970s (see Table 1). However, while during the 1980s the growth rate of government expenditure was less than that of potential output in Belgium, Ireland, Germany, the Netherlands and the UK, the opposite holds for Spain, Italy, Finland, France, and Portugal. In this second group of countries (and in the Netherlands), government expenditure growth decelerated in the 1990s and declined relative to that of potential output (except in Portugal). Expenditure grew less than potential output also in the 1990s in Ireland, Luxembourg and the UK. In recent years, several countries are witnessing a change in the behaviour of government expenditure. Starting from 2000, government expenditures relative to potential output picked up in UK, Luxembourg, Ireland, Belgium, Sweden and Italy.

Figure 2 reports on the horizontal and on the vertical axis, respectively, the average growth rate over the different decades of potential output and of government expenditure across countries. The figure shows that while in the 1970s there was a clear and almost linear cross-country positive relation between the two variables such that at higher rates of potential growth it was associated a more than proportionally higher growth in government expenditure, this cross country relation changes in the 1980s and 1990s. The relation is still positive, but when potential output grows faster, government expenditure tends to grow faster

[10] See European Commission (2002 and 2004) for an explanation of the European Commission methodology for the cyclical adjustment of public finance variables.

but less than proportionally. Additionally, this cross-country relation weakens: there is a greater dispersion in data points indicating that cases where potential growth is high and expenditure growth is low, or vice-versa, become more frequent. Finally, it is to notice that on average across countries, while in the 1970s government expenditure grew faster than potential output, starting from the 1980s the growth rate in the two variables is roughly equal. Overall, this indicates that the relation between government expenditure changed over time to some extent. Part of this change reflected a general tendency observed in all countries, part was related to specific country cases.

3. Panel unit root tests and cointegration analysis

3.1. Panel unit root tests

A first specification assumes that all units are stationary with the same autoregressive coefficient across units (the *homogeneous* alternative hypothesis). This implies that the relevant variable in all countries converge towards their average at the same speed. The statistics developed by Levin, Lin and Chu (2002) - LLC hereafter - and Breitung (2000), both test the null of unit root against this homogeneous alternative of stationarity. These tests allow for heterogeneous serially correlated errors, country-specific fixed effects and country-specific deterministic trends, and are based on an Augmented Dickey-Fuller (ADF) regression of the following type:

(1) $$\Delta y_{it} = \delta_i \tau + \phi_i y_{it-1} + \sum_{j}^{p_i} \beta_{ij} \Delta y_{it-j} + \xi_{it}$$

where y_{it} is a given variable (expenditure or GDP in our case), i denotes panel units (countries in our case), t is time, τ is a common trend across countries, p_i is the country-specific lag order, and ξ_{it} are stochastic errors which could be serial correlated.

The formulation of the null (H_0) and alternative (H_1) hypothese in this set up is as follows: H_0: ϕ_i=0; H_1: $\phi_i=\phi<0$. This dynamic structure is likely to be restrictive for variables with a time path strongly influenced by country-specific factors such as public finance variables.[11] Neglecting this source of heterogeneity makes the use of the pooled estimators such as those proposed by LLC and Breitung inappropriate and the estimates of the parameters inconsistent

[11] Certainly it is too restrictive in the case of public finance variables whose dynamic properties are influenced by the characteristics of national institutions.

even when the time and the cross-section dimension of the sample are large (Pesaran and Smith (1995)).[12]

Second-generation unit root tests allowing *heterogenous short-run dynamics* help to overcome the above limitations. The test devised by Im Pesaran and Shin (2003) - IPS hereafter - allows for some (but not all) of the individual series to have a unit root under the alternative hypothesis, implying that the degree of persistence of the variable of interest is not forced to be the same. The heterogeneous alternative hypothesis is that al least some of the units have stationary processes. [13] Maddala and Wu (1999) - MW hereafter - suggest instead a test of unit root against the heterogeneous alternative that combines the p-values from unit root statistics in each cross-sectional unit.[14]

To smooth series and permit an interpretation of regression coefficients in terms of elasticities, all the regression analysis is performed on the natural logarithms of expenditure and GDP series described in section 2.2. With trending variables, the testing equation should have intercepts when variables are expressed in first differences.

Moreover, since panel unit root tests require cross-sectional independence, the tests are also applied to de-meaned data. If countries are equally affected by common factors (i.e. aggregate disturbances common to all), then demeaning the data permits to eliminate cross-sectional dependence. In the presence of country-specific deterministic trends, Phillips and Moon (2000) suggest to test the unit root hypothesis on OLS de-trended. Tests are therefore also performed on demeaned and OLS de-trended data.

Tables 2a-2c present the results from LLC, Breitung, IPS and MW unit-root tests. For each variable, the table displays the p-value associated with the testing equations including,

---

[12] Moreover, when the independence across units is violated, unit root tests tend to over-reject the null hypothesis (Banerjee et al (2004)).

[13] Formally, the null hypothesis is H_0: ϕ_i=0 against the alternative H_1: ϕ_i<0 for i=1,2,…,N_1 and H_1: ϕ_i=0 for N_1+1,…,N, where N is the total number of cross-section units and N_1 is the number of cross-section units having a stationary process.

[14] Their test statistics is $-2\log\prod_i^N p_i$, where N is the number of cross-section units and p_i is the p value associated to unit i. The test is distributed as χ^2 with 2N degrees of freedom. The p-value is the smallest significance level at which H_0 can be rejected (not the probability of H_0 itself). If the significance level is less than the p-value it is not possible to reject the H_0. If H_0 were to be rejected at significance level α, this would be the case for $p<\alpha$. For example, if the p-value is 0.027, the results are significant (i.e. it is not possible not to reject the null) for significance levels greater than 0.027 (such as 0.05) and not significant for all significance levels less than 0.027 (such as 0.01). Somebody using the 5% level would reject the null hypothesis while a person who uses the 1% level would fail to reject it. The inferential step to conclude that the null hypothesis is false goes as follows: the data (or data more extreme) are very unlikely given that the null hypothesis is true. This means that: (1) a very unlikely event occurred or (2) the null hypothesis is false.

alternatively, country fixed effects only or also country-specific trends. The lags included in the ADF regressions are selected on the basis of the Akaike Information Criterion (AIC).

Table 2a shows that in almost all cases, government expenditure in levels has a unit root, while the series appear stationary once taken in first differences. In the case of potential output, the tests give instead conflicting result. When country-specific trends and intercepts are included in the testing regression, the null hypothesis of unit root is accepted by LLC, while it is rejected by the IPS and the MW statistics.[15]

Table 2b presents tests on cross-sectional de-meaned data. It turns out that the null hypothesis of expenditure being integrated of order 1 (I(1)) when a trend is included in the testing regression cannot be rejected. However, the results are still uncertain for the potential output. The results do not change significantly when the unit root tests are run on de-meaned and OLS-de-trended data (Table 2c)

When the effect of the common component differ across countries, de-meaning is not sufficient to eliminate cross-sectional dependence. Pesaran (2005) suggests a unit root test which controls for the common factor proxied by the cross section-averages of lags and differences of the individual series (named cross-section IPS or CIPS). Similarly to the IPS test, panel unit root tests are based on the averages of individual Cross-sectional Augmented Dickey-Fuller t-statistics (CADF).[16] Table 2d shows that, based on the CIPS test, the hypothesis of unit root both for the expenditure and the potential output cannot be rejected at the 5 significance level (Table 2d).

Overall, there is evidence that primary cyclically-adjusted government expenditure and potential output are non-stationary and therefore candidate for being cointegrated - i.e. there is a potential long-run relationship tying cyclical adjusted primary expenditure and potential output.

4.2. Panel cointegration tests

[15] According to the Breitung test it is not possible to reject the assumption that potential output is I(2). This outcome is clearly inconsistent with the hypothesis of balanced growth as it implies that temporary shocks to the growth rate turn out to be permanent. One problem with panel unit root tests is that they tend to over-reject the null hypothesis of non-stationarity when there are errors with a large negative root and the lag selected by the traditional information criteria is small. Ng and Perron (2001) propose a Modified Akaike Information Criteria (MAIC) that is data-dependent. The MAIC takes into account the nature of the deterministic components and the de-trending procedure, which allows for a better measurement of the cost of each lag choice. When lag length is determined with the MAIC, in all cases it is not possible to reject the null of unit root (results are available by the authors upon request).

[16] The cross-sectional IPS (CIPS) test is defined as the average of the individual CADF. The CIPS test has a non standard distribution with critical values tabulated in Pesaran (2005).

Difficulties analogous to those encountered with unit root tests are found when testing for cointegrating relationships in panel data. Firstly, it is necessary that the idiosyncratic error terms are independent across units in the panel. This implies that disturbances to one unit are not diffused to other units.[17] Secondly, Banerjee et al. (2001) warn against the existence of cointegration *between* some units in the panel.[18] Thirdly, there is the issue of possible multiple cointegration vectors. Available residual based panel cointegration tests make the assumption of a single cointegrating vector. In our particular application this is not an issue since panel cointegration is tested between two variables only: government expenditure and potential output.

This paper uses residual-based tests of the null hypothesis of no cointegration developed by Pedroni (1995, 1997, 1999). The tests is performed on the residuals of a static regression and allow for country-specific short-term dynamics and long-run relationships.[19] In symbols, the tests are based on the following regression:

$$e_{it} = \alpha_i + \theta_i\, y_{it} + u_{it} \qquad (2)$$

Where e_{it} and y_{it} are, respectively, the log of primary cyclically adjusted government expenditure and of potential GDP in country i and year t, u_{it} is a stochastic residual and α_i the country specific intercept. The elasticity of expenditure to output θ_i is allowed to vary across individual countries. Cointegration occurs when the linear combination of I(1) variables is stationary, implying that deviations of one variable from the path prescribed by the cointegrating relationship are transitory (i.e. without memory). In such a case, there is a long-run relationship between the variables and temporary deviations can be modelled with an error correction mechanism (ECM).

Starting from equation (2), Pedroni proposes seven tests for the null hypothesis of no cointegration using the residuals estimated from panel regressions, in analogy with the Engle and Granger method. These tests differ according to the way in which information is combined. Four tests are based on pooling information along the within dimension and three

[17] Asymptotic distributions of the tests are derived under the hypothesis of cross-sectional independence.

[18] In the case of testing PPP, they show that the hypothesis of a unit root tends to be rejected too often in the presence of cross-unit cointegrating relationships.

[19] No hypothesis of exogeneity is imposed on the regressors of the cointegrating equation. The test control for endogeneity/reverse causality. In contrast, the test is based on the assumption of a single cointegrating vector, although this does not need to be the same across countries. As in the case of panel unit root tests, the individual processes are assumed to be independent cross-sectionally.

tests on pooling along the between-dimension.[20] For the within-dimension statistics, the test for the null hypothesis of no cointegration is a residual-based test of the hypothesis that the residuals are non-stationary (i.e. no cointegration between the variables) against the alternative of stationary residuals (i.e. cointegration) with exactly the same autocorrelation coefficients of residuals across countries. Regarding the tests performing pooling along the between dimension, the null hypothesis does not change (i.e. no cointegration) while the alternative presumes country-specific autocorrelation coefficients of residuals.

Pedroni (1997) performs Monte Carlo simulations to study the small sample properties of the tests. He shows that in terms of power, panel ADF tests (obtained pooling along the within dimension) followed by group ADF tests (constructed pooling along the between dimension) perform better than the other. Hence, we restrict our analysis to panel ADF and group ADF Pedroni cointegration tests.[21]

As already mentioned, the cointegration tests used in this paper are valid only under the assumption of cross-sectional independence, i.e, disturbances to one unit are not diffused to other units. A general form of cross-sectional dependence can be modelled as follows:

(3) $$e_{it} = \alpha_i + \theta_i y_{it} + \delta_i t + \delta_t + \varepsilon_{it}$$

where δ_t is a common residual component which impacts all countries in the same way, $\delta_i t$ is a common trend which may have a different impact depending on the country, and ε_{it} is serially uncorrelated disturbance. The main idea to achieve cross-sectional independence is to eliminate the common factor before applying cointegration tests on filtered data. The structure assumed for the common component of equation (3) is quite flexible to model alternative forms of cross-sectional dependence. When $\delta_i = 0$, the common component has the same effect on expenditure for all countries and cross-section independence is achieved by simply de-meaning the data. When $\delta_i \neq 0$ and $\delta_t = 0$, the effect of the common component differs across countries and independence can be achieved de-trending the original data.[22] Finally,

[20] The within dimension statistics are based on estimators that pool the autoregressive coefficient across different members for the unit root tests on the estimated residuals. The between dimension statistics are based on estimators that average the individually estimated coefficients for each member of the panel (see Pedroni (1999)).
[21] These tests, after appropriate normalisation, converge to a standard normal under the null of no cointegration. These statistics are normally distributed and diverge to negative infinite under the alternative of cointegration. Hence, the null of no cointegration is rejected for large and negative values of the test statistics.
[22] When there is a considerable heterogeneity in the deterministic trends, Phillips and Moon (1999) suggest OLS de-trending. They argue that a consistent estimate of the cointegrating vectors can be obtained when data are OLS de-trended. They also show that OLS de-trending is more efficient than GLS-de-trending.

when $\delta_i \neq 0$ and $\delta_t \neq 0$, there is a common component of trend expenditure which impacts all countries in the same way and one whose effect is country specific: both de-trending and de-meaning are required to get cross-sectional independence. Since there is no clear a-priori on the form in which cross-sectional dependence could manifest, we perform Pedroni tests alternatively on original data, de-meaned data, OLS de-trended data and data that are both de-meaned and de-trended.

Table 3 reports the results of the cointegration tests.[23] In interpreting results, it is important to bear in mind that different transformations of the original data reflect different assumptions on the common component. The tests are performed both including and non including a trend in the cointegration regression. Trends are dropped from the cointegration regression when tests are performed on de-trended data. Results show that when variables are not de-trended, the null hypothesis of no cointegration is rejected by the group ADF test on original data if the cointegrating equation does not include a trend and on de-meaned data if there is no trend in the cointegration regression. In the case of de-trended data, Pedroni tests always reject the null hypothesis of no cointegration. Overall, on the basis of this evidence, and given that the group ADF, which allows for a more general structure of the residual correlation under the null hypothesis is also the most powerful test in small samples (Pedroni (1997)), we conclude that the primary expenditure and potential output are cointegrated.

Having established that government expenditure is cointegrated with potential output, we proceed modelling the error correction mechanism allowing for country specific short-run coefficients. The approach is based on the pooled mean group estimator (PMG, see Pesaran et al (1999)) which allows testing the hypothesis that the cointegration relation across the cross section units is the same, in our case, that the long-run elasticity between government expenditure and potential output is the same for all countries.

4. Heterogeneous panel ECM estimation

4.1. The approach

Building on the existence of a long-term relation between government expenditure and potential output in our panel of EU countries, the aim of this section is to estimate this long-

[23] When data are de-trended and then cross-sectionally de-meaned the cointegration test exclude a trend from the cointegrating regression.

run relationship jointly with the short-term dynamics. A fairly general dynamic specification is represented by an auto-regressive distributed lag model of order p_i and q_i, ARDL(p_i,q_i):

(4) $$e_{it} = \sum_{j=1}^{p_i} \lambda_{ij} e_{it.j} + \sum_{j=0}^{q_i} \delta_{ij} y_{it.j} + \mu_i + u_{it},$$

where μ_i is an unobserved country-specific effect and u_{it} is the error term. The ARDL(p_i,q_i) can be rewritten in the following error correction model form (Pesaran et al.(1999)):

(5) $$\Delta e_{it} = \phi_i \left(e_{it-1} + \frac{\beta_i}{\phi_i} y_{it} \right) + \sum_{j=1}^{p_i-1} \lambda_{ij}^* \Delta e_{it.j} + \sum_{j=0}^{q_i} \delta_{ij}^* \Delta y_{it.j} + \mu_i + u_{it}$$

where $\phi_i = -\left[1 - \sum_{j=1}^{p_i} \lambda_{ij}\right]$; $\beta_i = \sum_{j=0}^{q_i} \delta_{ij}$; $\lambda_{ij}^* = -\sum_{k=j+1}^{p_i} \lambda_{ik}$; $\delta_{ij}^* = -\sum_{k=j+1}^{p} \delta_{ik}$

When the ARDL(p_i,q_i) is stable (i.e., error correcting), the adjustment coefficient ϕ_i is negative and less than 1 in absolute value. In this case, the long-run relationship is defined by:

(5) $$e_{it} = -\frac{\beta_i'}{\phi_i} y_{it} + \eta_{it} \quad ,$$

where η_{it} is a stationary process. In steady-state, trend expenditure and potential output are tied one to the other, with a *long-term elasticity* of by $\theta_i = -\frac{\beta_i}{\phi_i}$. Under the Wagner law, the long-term elasticity is expected to be positive and larger than 1. Conversely, the assumption underlying widely used methods to adjust government budgets for the effect of the cycle is that the long-term elasticity between government expenditure and potential output is unitary.

Temporary deviations from this relationship are possible and may be driven by common and/or country specific shocks. The parameter ϕ_i measures the adjustment coefficient of the error correction term. It says how much of a temporary deviation of trend government expenditure from potential output is eliminated in one year.

The ECM in equation (5) can be estimated in different ways. Traditional time series models do not take into account the information on the cross-country correlation in the data. Dynamic fixed effect models control for country fixed effects but impose the same coefficients for all countries.[24] Pesaran and Smith (1995) show that pooling produces inconsistent estimates of the parameters value unless the slope coefficients are identical.[25] To tackle this issue, Pesaran and Smith (1995) propose a mean group estimator (MG) consisting of estimating the coefficient of each cross section and then taking an average of them. Although consistent, the MG estimator does not take into account that some of the parameters may be the same across countries, implying that its estimates, especially in small samples, are likely to be inefficient and strongly affected by the presence of outliers.

An intermediate choice between imposing slope homogeneity and no restrictions is the pooled mean group estimator (PMG) proposed in Pesaran, Shin and Smith (1999), which combines the characteristics of the pooled estimators (namely the fixed effect) with those of the mean group estimator.[26] The PMG estimator treats differently the short- and the long–run dynamics. [27] The short-run dynamics are allowed to differ across countries but the long-run effects are constrained to be the same. Formally, the PMG estimator imposes the restriction that the long-run-coefficients are the same across units: $\theta_i = \frac{\beta_i}{\varphi_i} = \theta$ [28]

The PMG estimator is appropriate when data have complex country-specific short-term dynamics which cannot be captured imposing the same lag structure on all countries. This estimator combines the properties of efficiency of the pooled dynamic estimators while avoiding the inconsistency problem deriving from slope heterogeneity. [29] The restriction of

[24] It is well known that with a small time dimension, dynamic fixed effects estimators give biased and inconsistent estimates of the parameters. However, when the number of observations over time is large enough, the asymptotic bias of the estimator is likely to be rather small (Baltagi, 2005).

[25] The inconsistency does not disappear even when the size of the cross-section and of that of the time periods is large.

[26] There is an increasing use of PMG estimates in applied econometric work. PMG estimates have been recently used in the analysis of the effects of institutions on innovation and growth (OECD (2001)), for modelling the Euro area demand of money (Golinelli and Pastorello (2002)), to analyse the wealth effects in the consumption function (Barrel and Davis (2004)), to explore the impact of policies on fertility rates (D'Addio and Mira D'Ercole (2005)), to identify the determinants of the sovereign risks in the gold standard (Cameron and Tan (2006)), to the analysis of the link between fiscal policies and the trade balance (Funke and Nickel, 2006), to investigate the effects of financial intermediation on economic activity (Loyaza, 2006).

[27] If a long-run relationship between y_{it} and x_{it} with coefficients identical across groups exist and assuming that disturbances uit are normally and independently distributed across countries, the equation (5) is estimated with Maximum Likelihood by means of the Newton-Raphson algorithm.

[28] Long-run homogeneity can also be imposed on a subset of variable and/or countries.

[29] The test of homogeneity of the long-run coefficients consists of an Hausman test that compares the MG and the PMG estimators (Pesaran et al, 1999; Pesaran et al, 1996). The PMG estimator is consistent and efficient under the null hypothesis of long-run slope homogeneity and inconsistent under the alternative of long-run slope heterogeneity. The MG estimator provides a consistent estimate of the mean of the long-run parameters although this is inefficient under null of homogeneity.

homogenous long-run coefficients can be tested by means of a Hausman test.[30] Moreover, since the PMG estimator does not impose any restriction on short–term coefficients, it provides important information on country specific values of the speed of convergence towards the long-run relationship linking government expenditure and potential output.

4.2. Pooled mean group ECM estimation

PMG estimates are valid under the assumption that disturbances are independently distributed across units and over time with zero mean and constant variances. The independence of the disturbances across countries is needed for the consistent estimation of the short-term coefficients. Following Pesaran et al. (1999), we model cross-sectional dependence assuming the existence of observable common components in the residual, captured by the EU15 aggregate potential output, which is assumed to have an impact on government expenditure that differs across countries. In formal terms, the error component of the ARDL is defined as follows:

(6) $$u_{it} = \boldsymbol{\psi}_i^{'} \lambda_t + \varepsilon_{it} \ ,$$

where λ_t is a common factor and ε_{it} are stochastic disturbances assumed to be with zero mean and constant variance and independently distributed across *i* and *t.* We make the further assumption that

(7) $$\boldsymbol{\psi}_i^{'} \lambda_t = \psi_i \, y_{EU,t} \, ,$$

i.e., that the EU aggregate potential output, $y_{EU,t}$, affects government expenditure in each country with an intensity measured by parameter ψ_i .[31]

Table 4 reports PMG estimates of the ECM. Lags are chosen on the basis of AIC and are allowed to vary across countries. Table 4 shows that the long-run elasticity of expenditure to output is not significantly different from 1. On the basis of the Hausman test it is not possible to reject the hypothesis of poolability of the long-run elasticity of public expenditure (p-value

[30] Also when the restriction of long-run homogeneity is rejected, pooling may still be preferable to averaging across country specific parameters as it reduces the effects of outliers, especially in small samples.

[31] Bai and NG (2002) propose to model cross sectional dependence of the error terms constructing the common factor λ_t from the error term using principal component analysis. However, Pesaran (2006) shows that the principal components approach can still yield inconsistent estimates. Pesaran (2006) shows that linear combinations of unobserved factors can be approximated by cross-section averages of the dependent variable and the observed regressors.

0.34). The error correction coefficient is negative and statistically different from zero, implying that any deviation of government expenditure from the value predicted by the long-run relationship with the potential output triggers a change in the opposite direction in government expenditure. The average value of the error correction coefficient of government expenditure is -0.35, implying a speed of adjustment of about 3 years.

Specification tests indicate that in most countries there is no evidence of misspecification (Table 5). In all countries but Belgium, there is no first order serial autocorrelation. The RESET test rejects the functional form of the ECM only for Belgium, Germany and France. The heterosckedasticity test rejects the hypothesis of constant variances for Belgium, Germany, Ireland, Italy, Netherlands and Finland. Finally, the Jarque-Bera test suggests non-normal errors for Austria, Finland and Germany.

Table 5 shows that countries' individual estimates of the error correction coefficient are all negative, implying convergence of expenditure towards its long-run equilibrium.

The adjustment coefficient for Belgium is equal to 1 as the AIC criteria selects an ARDL(0,0). However, tests of functional form suggest possible problems with this specification (Table 5). Hence, for this country, an ARDL has been estimated,imposing the long-run elasticity given by the PMG estimator and selecting the lags on the basis of their statistical significance and of the usual diagnostic tests. The final model has normal, serially, uncorrelated and homoskedastic residuals and a speed of adjustment of 0.11.[32]

Similarly, the equation for Germany suffers from non-normal and heteroskedastic errors and the test of functional form is rejected by the data. The equation for Germany has therefore been re-estimated imposing the long-run elasticity estimated by the PMG. A model with 2 lags of the EU-15 potential output, a dummy variable for 1991 (a unification dummy), and the ECM lagged by one year yields a satisfactory representation of the data. The short-run elasticity is in this case 0.073, implying a very persistent out-of-steady-state dynamics.

With the revised coefficients for Belgium and Germany, the average speed of adjustment is about 0.29, implying that on average it takes about 3 years for public expenditure to close a temporary deviation from the level predicted by the long-run relationship with potential output.

[32] An ARDL with 3 lags for both government expenditure and potential output, no lags for EU potential output and a shift dummy for 1981 yield well behaved disturbances in the case of Belgium.

The speed of adjustment is relatively fast in the UK, Ireland, Sweden, Greece and Finland, while it is relatively slow in Spain, Italy and Portugal. Overall, there is some evidence that Anglo-Saxon and Nordic countries exhibit a faster adjustment of government expenditure to its long-term equilibrium, while adjustment is slower in Southern European countries. There are two exceptions to this pattern. First, Greece appears to be characterized by a very fast adjustment process. Second, Germany in our alternative specification exhibits a considerably low speed of adjustment.

4.3. Alternative modelling of the common component and estimation

To what extent are the estimates robust to the chosen approach? To provide an answer to such question we compute first the same estimates as in Table 4 with a different modelling of the common component affecting government expenditure across the EU. Instead of relying on an observed common factor (EU-15 aggregate potential output) as in our previous estimates, we assume now that the common factor cannot be easily identifiable with an observable variable. We simply assume instead the presence of a common EU wide deterministic trend that affects the relation between government expenditure and potential output. Since there is not strong a priori on the functional form for this trend, we allow both a linear and a quadratic trend component.[33] By estimating by PMG the model specified in this way with an unobserved common component, our previous results are to a large extent confirmed (see Table A1). The common long-term elasticity of government expenditure is above but close to unity (1.29) and the average short-term elasticity is -0.46. The short-run elasticity in all countries is negative, implying convergence of government expenditure towards its long-term relation with potential output.[34] The Hausman test accepts the poolability of countries. [35]

Second, we estimate the model via mean group estimation (MG), i.e., as suggested by Pesaran and Smith (1995), we estimate the model for each country and take an average of the value of the coefficients to infer the behaviour of government expenditure across the whole panel. As discussed, such method has a cost in terms of estimates' precision (efficiency) and, especially

[33] In terms of equation (7), $\boldsymbol{\psi}_i^{'}\lambda_t = \psi'_i\ \tau + \psi''_i\ \tau^2$.

[34] The country-specific values of the short-term elasticity broadly reflect the ranking in Table 5, with the exception of Spain, Italy and Austria (higher elasticities compared with baseline) and Ireland (lower elasticity).

[35] Although, in light of the wide array of interdependencies among European countries we consider the lack of any form of cross-section dependence unrealistic, to check robustness we also estimated the model excluding a common component altogether. Without controlling for common factors the long-term elasticity of government expenditure is the same as that of our baseline estimates in Table 4 (0.93), while the estimated short-term elasticity is somehow smaller (-0.19), implying an adjustment of government expenditure taking on average about 5 years.

in small samples, the presence of outliers can greatly affect the estimated coefficients. there is no cost PMG as . Nonetheless, to check robustness, in Table A2 we repeat the estimation of the model via MG estimation. Results show that, while the long-term government expenditure elasticity is higher than that obtained with PMG (2.16), the average short-term elasticity is close (-0.44).

Overall, our baseline results in Table 4 seem relatively robust with respect to alternative specifications of the common component affecting government expenditure across the panel. Conversely, the estimation of the long-term elasticity is clearly affected by the estimation method. By choosing the MG rather than the PMG method a much higher value is obtained.

5. Robustness analysis

In this section, the robustness of the relation between government expenditure and potential output is checked against alternative definitions of the sample. We address the following questions. Is the relation significantly stable over time? Are there countries with a significantly different behaviour? Which country characteristics appear to be related with the values of the elasticity of government expenditure with respect to potential output?

5.1 Stability over time

We first check the stability over time of our results via a recursive PMG estimation of the empirical model illustrated in section 5. The model is estimated initially over the 1970-1979 sub-period and repeatedly adding 5 additional years until the entire sample period (1970-2003) is covered. Figure 3 displays the results. Figure 3a plots the values of the long-run elasticity of government expenditure over the various sub-samples considered; Figure 3b does the same for short-run elasticities.

Results show that the long-term elasticity changed substantially over the period considered. In the 1970s the value of the elasticity was around 2. The 1980s were marked by a substantial decline in the long-term elasticity, whose value appears to have stabilised only at mid nineties. Overall, recursive estimations suggest a significantly different and higher response of expenditure to output for the earlier decades. Indeed, for the 1970s and the 1980s the confidence bands of the coefficient estimated recursively (the dotted lines in Figure 3a) do not include any of the values included in the corresponding bands estimated including also the 1990s.

In contrast, Figure 3b shows that the speed of adjustment appears rather stable over time. In almost all sub-periods, the estimates of the adjustment coefficients are significantly negative and less than 1 in absolute value, implying that following a shock, expenditure converges back towards its long-run relationship with the potential output. The set of values falling into confidence intervals overlaps with that obtained for the whole period already in the 1980s.

In light of the evidence of substantial changes in the long-term relation between government expenditure and potential output during the 1980s, PMG estimations and Hausman tests are performed over two different sub-periods: 1970-1989 and 1990-2003 (Table 6). The Hausman test suggests that hypothesis of long-run homogeneity (i.e. equality across countries of the long-run elasticity) is supported by the data over all the sub-periods. In line with the findings from recursive estimations, the hypothesis of unit elasticity of expenditure to output is rejected for the earlier period, while it is accepted for the more recent sub-period. Regarding the short-term elasticity, it appears relatively stable across-sub-periods.

5.2. Stability across countries

The findings from our ECM estimation via PMG might be affected also by the relative small number of countries in the sample. As a further robustness check we have re-estimated the model excluding from the sample one country at a time. This permits to understand whether the results are strongly driven the behaviour of a single country. Figures 5a and 5b plot, respectively, the value of the long-run and of the short-run elasticity of government expenditure on the country excluded from the sample.

The only country that appears to influence significantly the estimation of the long-run elasticity is Ireland. When this country is excluded from the sample, the estimated elasticity is significantly higher: there are values of the long-run elasticity falling inside the 95% confidence band (i.e., +/- 2 times the standard deviation of the estimated elasticity) that would be too high to fall inside the corresponding confidence bands when the estimates concern any sample with Ireland included. This indicates that the presence of Ireland contributes to keep low the value of the long-run elasticity estimated on the whole sample. The impact, though significant, appears not to be strong enough to alter qualitative results: the long-run elasticity of government expenditure once Ireland is excluded is still close to unity.

Regarding the short-run elasticity, Spain appears to reduce (in absolute value) significantly the value estimated across the whole panel: its exclusion leads to an estimated elasticity of about -0.4, with values falling within the confidence band that are too negative to be included also in the corresponding confidence bands obtained with any sample including Spain.

As a further check of stability we have computed the Hausman test of poolability of long-run coefficients 15 times excluding each time one country from the estimation. The test is always accepted, implying that the long-run countries' coefficients are indeed poolable in all cases.

5.3. Checking rebustness across country groupings: development stage, demography, public finances, fiscal governance

The relationship between government expenditure and potential output is affected by a series of factors (economic, demographic, institutional,..). For instance, there are reasons to think that catching-up countries are likely to exhibit a higher long-run elasticity of government expenditure compared with countries at a later stage of development. Catching up countries are in general characterized by a less developed social welfare system, which tends to grow in size as income per-capita rises. The demand for government investment is also likely to grow faster during the catching up process, since public infrastructure needs to adapt to the requests of an expanding private sector. This means that, by grouping countries according to their initial per-capita GDP, one should expect different long-run elasticities for government expenditure for different country groupings: relatively high for countries starting with low income-per capita, relatively low for those countries where initial income per-capita was high.

This section aims at testing robustness of results by splitting countries according to particular characteristics that are likely to affect the relation between government expenditure and economic activity. A systematic analysis of the influence of all possible factors that could play a role is beyond the scope of this paper. We focus instead on a limited set of factors that appear obvious candidates for such an exercise: the stage of economic development, demography, the state of public finances. Results are reported in Table 8.

Development stage

The first robustness check consists of splitting the sample according to the per-capita GDP (measured in PPP) at the beginning of the sample period. The expectation is that the long-run elasticity in countries with initial low per-capita income should be higher, being those the countries likely to have experienced a catching up process during the period considered.

In order to obtaining country groups of about equal size, the median of the initial per-capita GDP has been used as a cut-off value to split countries. According to the most recent update of the Penn World Tables (Heston, Summers, and Aten (2002)), the median annual real GDP measured in PPP across EU-15 countries in 1970 was 3640.3 US dollars (base year 1996). Countries with income per-capita below this value median in 1970 were Portugal, Ireland, Greece, Spain, Italy, Austria, and Finland (see Table 9).

Table 8 shows that our expectation is fully met: while the long-run elasticity of government expenditure is close to unity for high per-capita GDP countries, the elasticity for the countries with low initial per-capita GDP is about 3. This result seems to suggest that the Wagner law is a phenomenon that mostly pertains to catching-up countries. Regarding the short-run elasticity, the difference between the two country groups is rather limited and values are close in both cases to those estimated for the whole sample. The Hausman test accepts the poolability of countries for both country groups.

Demography

The second assumption that we are interested in testing is whether countries characterized by a population that is ageing faster are also distinguished by a high elasticity of government expenditure over the long-run, and whether this has any bearing on the short-run elasticity as well. The idea is that an ageing population entails expenditure dynamics that are independent of those relating to potential output growth. Government expenditure grows not only to satisfy rising demands for public goods and services stemming from rising incomes, but also to accommodate a changing composition of the population (a rising fraction of old people with higher social welfare claims). Since old dependency ratios tend to raise over time as well as potential output, the time series relation between government expenditure and potential output is inevitably affected by ageing.

Countries have been split in two groups on the basis of the median change in the old dependency ratio over the period 1970-2003. The median change in the dependency ratio is 4.5.[36] Countries classifiable as slow-ageing according to this criterion are Ireland, Austria, Luxembourg, Denmark, UK, the Netherlands, while Germany and France are very close to the median.

[36] This indicator is the ratio between the total number of elderly persons of an age when they are generally economically inactive (aged 65 and over) and the number of persons of working age (from 15 to 64). Source: Eurostat

Results in Table 8 show that, in line with our expectations, the long-run elasticity of slow-ageing countries is lower than that of fast-ageing countries. [37] Slow-ageing countries are also characterized by a somehow faster adjustment process of government expenditure to its long-term relation. Hausman tests accept the hypothesis of countries poolability for both country groups.

Public finances

A further hypothesis we are interested in is whether countries with high debt/GDP ratios are characterised by a lower long-run government elasticity. Since Bohn (1991) an expanding literature has analysed the issue of public finances sustainability by looking at the relation between flow and stock public finance variables via the econometric estimation of fiscal reaction functions. The aim is establishing whether any increase in government debt induces a rise or a fall in primary government surpluses, the former implying debt sustainability. Fiscal reaction functions generally analyse the behaviour of the share of primary government budget balance over GDP, but estimates have been carried out separately for government primary expenditure as a share on GDP. Results show in general that government expenditure tend to fall in relation to GDP as debt/GDP ratios fall, a result consistent with the hypothesis that fiscal authorities set expenditure motivated also by the purpose of stabilising debt (see, e.g.,).

To shed light to the above assumption, the sample has been slit between high-debt and low-debt countries on the basis of the median value of the debt/GDP ratio observed on average across the period (see Table 9).[38]

Table 8 shows that, in accordance with the hypothesis outlined above, the long-run elasticity estimated for the group of low-debt countries is considerably higher than that for high-debt countries. The latter group of countries also appears to adjust government expenditure at a speed that is about twice that high-debt countries. In the case of high-debt countries, however, Hausman tests reject the hypothesis of poolability, a possible indication of heterogeneity in the cross section's long-run coefficients and panel mis-specification.

Fiscal governance

[37] This finding is robust with respect to the exclusion of France and Germany from the group of slow-ageing countries. In this case, the elasticity is 0.84 (t-Student of 7.46), the speed of adjustment -0.48 (t-Student -4.22) and the Hausman test 0.26 (t-student 0.61).

[38] This criterion to split the sample depends to some extent on the expenditure dynamics itself. An alternative would be to use the median debt/GDP ratio observed at a given point in time. This is however problematic, since the rank of countries according to their debt/GDP ratio changes over time.

Finally, we check robustness of our baseline results concerning the long run relation between expenditure and potential output and the error correction coefficient splitting country groups according to their fiscal governance, namely the set of rules and institutions that contribute to the government control of fiscal variables, notably public expenditure. To that purpose, the sample was split on the basis of the average value of the EU Commission indicators of national-level expenditure rule across years for which information are available. The indicators vary across countries and over time and capture both the degree of coverage of numerical rules to keep expenditures under control (i.e., which share of general government expenditures are subject to the rule) and a series of qualitative features of the rule: their statutory basis, their monitoring and enforcement procedures, and their visibility in the media (for details on the construction of the indexes see European Commission (2006) and Ayuso-i-Casals et al. (2006)). The information for the construction of the indicators was collected via questionnaires targeted to experts on finance ministries and covers 22 EU countries over the 1990-2005 period. The countries with expenditure rules receiving a raking higher than that of the median country are Denmark, Germany, France, Ireland, Luxemburg, the Netherlands and Sweden.

The expectation is that countries with "stronger" numerical rules to control expenditure should exhibit a lower long-run coefficient linking government expenditure to potential output compared with the group of "weak-rule" countries: for the former it would be easier to contrast the tendency for government expenditures to grow over time as a result of ageing or increased pressures for spending. The same group of countries is also expected to be able to correct faster any divergence between current developments in expenditure and the long-term trend (i.e., to exhibit a higher error correction coefficient). Table 8 shows that our results are in line with expectations. Hausman tests reject the hypothesis of poolability only for countries with low numerical expenditure rules. This finding suggests that the presence of strong fiscal rules is sufficient to identify an homogenous group of countries with an elasticity of expenditure to output below one, while the group of countries with weak expenditure rules exhibits a higher average long-run elasticity but is not homogenous.

6. Implications for policy

Overall, the analysis shows that on average, across our sample of EU countries, government expenditure and potential output are linked by a long-run relation such that government spending grows roughly in proportion with potential output.

This finding has a clear implication for the EU debate on public finances sustainability. It is often claimed in the EU policy debate that rising potential growth would be key to ensure the compatibility of relatively generous welfare systems with the sustainability of public finances over the long-run.[39] Satisfactory rates potential growth are a necessary condition for satisfactory growth rates of government revenues and would ensure a rapid reduction of the existing stock of government debt as a share of GDP. However, the net impact of potential growth on the future stream of government budget balances ultimately depends also on its impact on government expenditure. Our results suggest that, on average, increased rates of potential growth would leave the share of government expenditure on potential output roughly unaffected, but the impact would differ quite considerably across countries.

Evidence of a roughly proportional relation between cyclically adjusted primary government expenditure and potential GDP also sheds some light on the empirical validity of the alternative approaches followed to construct measures of cyclically-adjusted budget balances and to assess the stance of fiscal policy. Our findings yield empirical support to the models for the cyclical adjustment of budget balances based on the assumption that the share of government expenditure on potential output is constant in the long-run and that possible deviations have a cyclical nature. Our results also lend support to the analysis of the stance of expenditure policy based on the share of cyclically adjusted primary expenditure over potential output: an increase in such a ratio would be an indication of an expansionary stance of government expenditure.

The evidence on the speed of adjustment of government expenditure to potential output has implications for budgetary surveillance. This is particularly relevant in the EU context, where national budgetary policies are subject to a common framework for fiscal policy enshrined in the EU Treaty and in the Stability and Growth Pact (SGP).

Some implications concern the so-called preventive arm of the Stability and Growth Pact. In order to prevent the risk of breaching the 3% of GDP reference value for deficits, EU countries aim at medium term budgetary objectives defined in structural terms well below this threshold. The respect of such medium term objectives implies that the growth of government expenditure adjusts to changes in the growth rate of potential output. Our estimates indicate that such adjustment could take few years and be largely country-specific.

[39] See, e.g., Sapir et al. (2004).

Regarding the so-called corrective arm of the SGP, budgetary deteriorations ensuing from sluggish economic growth could lead to the breach of the 3 per cent of GDP reference value for deficits and the opening of an Excessive Deficit Procedure (EDP), in which countries are subject to enhanced surveillance by the Commission and the Council with a view to correct budgetary imbalances within deadlines defined in the SGP. The estimated short-run elasticities of government expenditure with respect to potential output provide information on the feasibility of the budgetary effort of EDP countries.

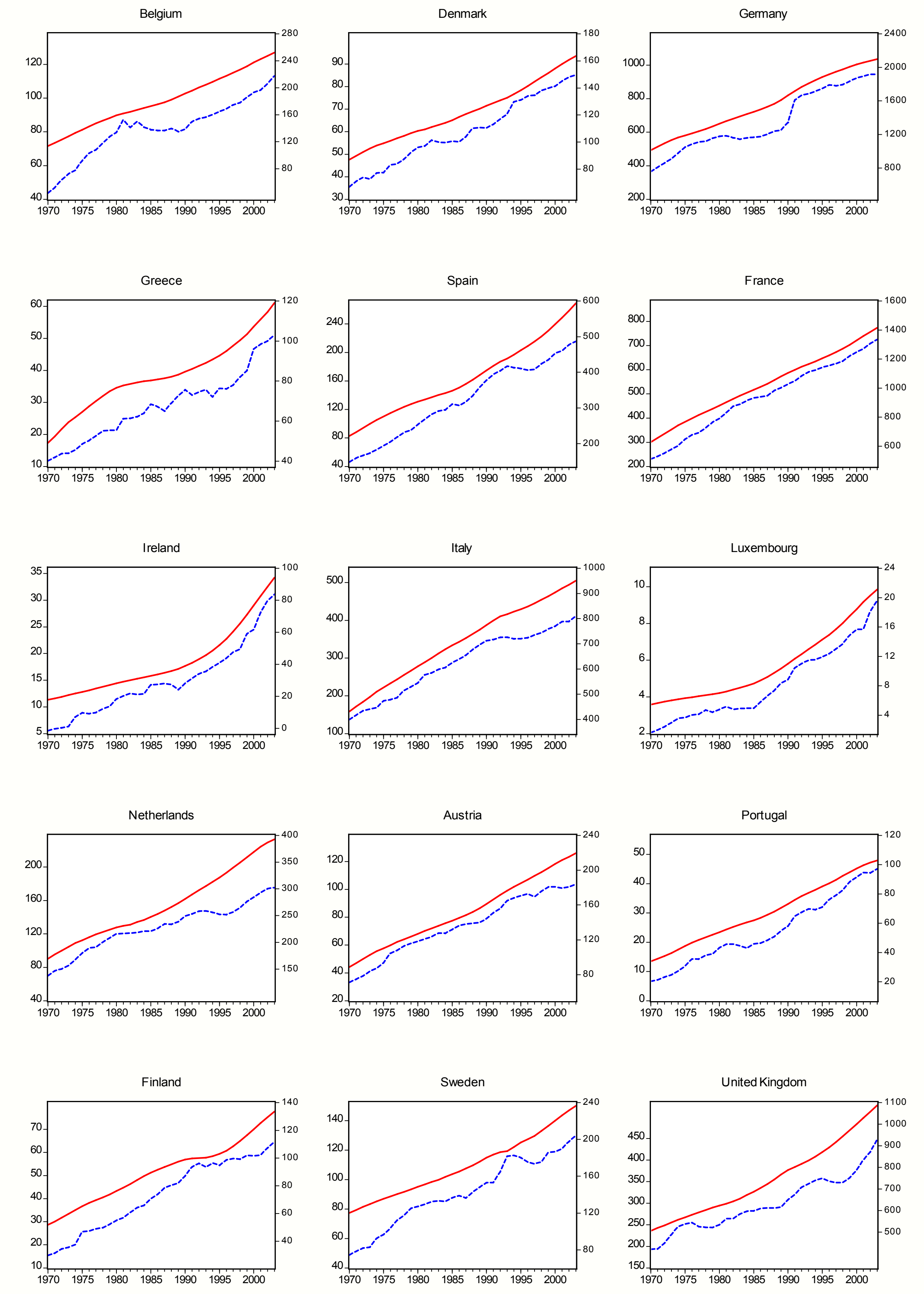
7. Concluding remarks

This paper has provided an estimation of the long and short-run relation between government expenditure and potential output across EU countries. Panel cointegration tests reveal that government expenditure and potential output in the EU are linked by a stable long-run relation. The estimation of the dynamic relation between the two variables by means of the Pooled Mean Group (PMG) estimator (Pesaran, Shin, and Smith (1999)) permits to combine the precision of the estimates allowed by pooling the data across the cross-country dimension while limiting the risk of inconsistency of the estimates associated with the possible heterogeneity of regression coefficients across countries. The PMG imposes a common long-term elasticity for all countries while allowing country-specific short-term elasticities.

Results show that the assumption of a common long-run elasticity is accepted by the data and that such elasticity is slightly below unity. The long-run elasticity is however not stable over time (it decreased considerably over the decades) and is significantly higher than unity in catching-up countries, in fast-ageing countries, in low-debt countries, and in countries with weak numerical rules for expenditure control. Country-specific short-term elasticities imply on average a speed of adjustment of government expenditure to potential output of about 3 years, even though coefficients vary quite widely across countries, with Anglo-Saxon and Nordic countries exhibiting in general higher speed of adjustment than Southern European countries. Such findings have implications for policy, notably for the EU, where countries are subject to a common framework for budgetary surveillance.

Overall, the paper shows that the estimation method matters substantially for the measurement of the relation between government expenditure and potential output. Relying and the average of individual country-level estimates would have yielded a long-run elasticity of government expenditure well above unity. However, such estimate would be less precise than one exploiting the panel dimension of the data.

Figure 1: Government expenditure and potential output (log scale).



Total expenditure excluding interest cyclically adjusted at 1995 market prices — **Potential gross domestic product at 1995 market prices**

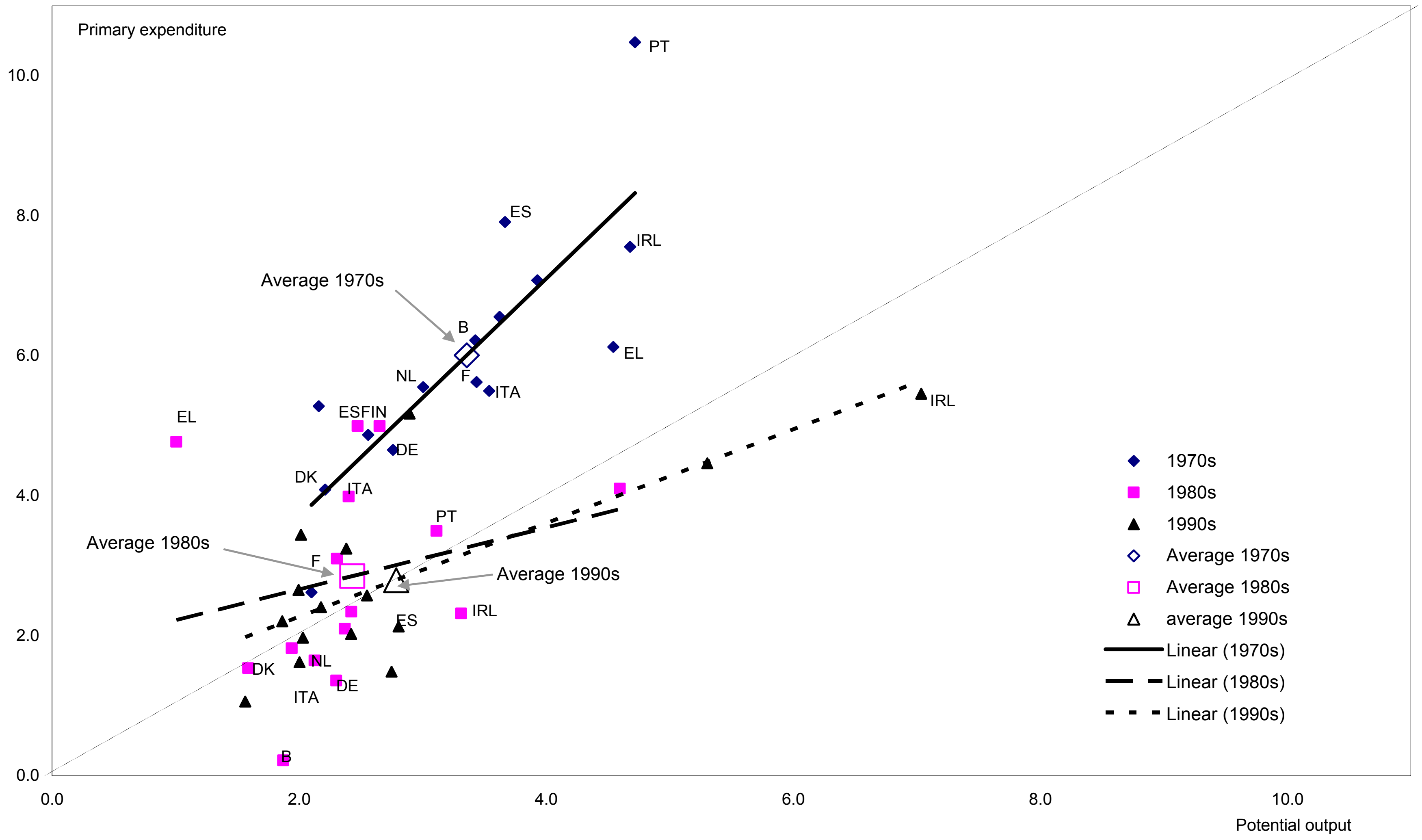
Source: European Commission AMECO database

Table 1 Growth rates of government expenditure and potential output (average annual growth rates)

	Cyclically-adjusted government expenditure				Potential output			
	1970-1980	1980-1990	1990-2000	2000-2003	1970-1980	1980-1990	1990-2000	2000-2003
Belgium	6.2	0.2	2.4	3.0	3.4	1.9	2.2	2.0
Denmark	4.1	1.5	2.7	2.0	2.2	1.6	2.0	2.0
Germany	4.7	1.4	3.4	0.8	2.8	2.3	2.0	1.0
Greece	6.1	4.8	3.2	3.0	4.5	1.0	2.4	3.6
Spain	7.9	5.0	2.1	2.8	3.7	2.5	2.8	3.6
France	5.6	3.1	2.2	2.5	3.4	2.3	1.9	2.1
Ireland	7.6	2.3	5.5	8.3	4.7	3.3	7.0	7.0
Italy	5.5	4.0	1.1	2.3	3.5	2.4	1.6	1.7
Luxemburg	4.9	4.1	4.5	6.4	2.6	4.6	5.3	4.6
The Netherlands	5.6	1.7	1.5	2.3	3.0	2.1	2.7	2.1
Austria	6.6	2.3	2.6	0.5	3.6	2.4	2.6	2.0
Portugal	10.5	3.5	5.2	2.1	4.7	3.1	2.9	1.9
Finland	7.1	5.0	1.6	3.4	3.9	2.7	2.0	3.4
Sweden	5.3	1.8	2.0	3.0	2.2	1.9	2.0	2.4
UK	2.6	2.1	2.0	5.9	2.1	2.4	2.4	2.8
Simple average	6	2.9	2.8	3.2	3.4	2.4	2.8	2.8
Coefficient of variation	0.30	0.52	0.48	0.65	0.26	0.34	0.52	0.53

Source: European Commission AMECO database.

Figure 2: Growth in government expenditure relative to potential output



Source: European Commission AMECO database

Table 2a Panel unit root test (p-values), EU-15, 1970-2003

Variable Method	Cyclically-adjusted primary government expenditure		Potential output	
	Without trend	With trend	Without trend	With trend
H_0: Unit root (common unit root process)				
LLC t				
Level	0.00 (473)	**0.05** (456)	**0.94** (456)	**0.25** (463)
First Difference	0.00 (468)	0.00 (454)	0.00 (460)	0.00 (459)
Breitung t-stat				
Level	**0.35** (458)	**0.12** (441)	**0.21** (441)	**0.62** (448)
First Difference	0.00 (453)	0.00 (439)	**0.09** (445)	**0.29** (444)
H_0: Unit root (individual unit root process)				
IPS t-stat				
Level	0.004 (473)	**0.05** (456)	**1.00** (456)	0.00 (463)
First Difference	0.00 (468)	0.00 (454)	0.00 (460)	0.00 (459)
ADF-MW χ^2				
Level	0.00 (473)	**0.06** (456)	**0.99** (456)	0.001 (463)
First Difference	0.00 (468)	0.00 (454)	0.05 (460)	0.00 (459)
PP - MW χ^2				
Level	0.00 (495)	**0.64** (495)	0.00 (510)	0.00 (495)
First Difference	0.00 (480)	0.00 (480)	0.002 (495)	**0.90** (495)

All data are expressed as natural logarithms of differences with respect to the cross-country averages. Country-specific intercepts are included in the testing equation. The p value of the test when the null hypothesis of unit root is *not* rejected is in bold. The null of unit root is accepted at significance level α when the p-values are bigger than $\alpha/100$. The number of observations is reported in parentheses. Automatic selection of lags based on the Akaike Information Criterion.
ADF and PP are two tests that uses Fisher's (1931) result to derive test that combine the p-values from individual unit roots tests. The tests are distributed as a χ^2 with 2*N degrees of freedom where N is the number of cross-sections.

Table 2b Panel unit root test (p-values): cross-sectionally de-meaned data, EU-15, 1970-2003

Variable Method	Cyclically-adjusted primary government expenditure		Potential output	
	Without trend	With trend	Without trend	With trend
H_0: Unit root (common unit root process)				
LLC t				
Level	0.00 (482)	**0.91** (473)	**0.81** (451)	**0.37** (452)
First Difference	0.00 (473)	0.00 (471)	**0.11** (459)	**0.86** (440)
Breitung t-stat				
Level	**0.60** (467)	**0.43** (458)	0.05 (436)	**0.56** (437)
First Difference	0.00 (458)	0.00 (456)	**0.31** (444)	**0.21** (425)
H_0: Unit root (individual unit root process)				
IPS t-stat				
Level	**0.61** (482)	**0.72** (473)	**0.94** (451)	**0.48** (452)
First Difference	0.00 (473)	0.00 (471)	**0.07** (459)	0.04 (440)
ADF-MW Chi-square				
Level	**0.48** (482)	**0.48** (473)	**0.29** (451)	**0.16** (452)
First Difference	0.00 (473)	0.00 (471)	**0.08** (459)	0.01 (440)
PP - MW Chi-square				
Level	**0.85** (495)	**0.98** (495)	0.04 (495)	**0.77** (495)
First Difference	0.00 (480)	0.00 (480)	**0.64** (480)	**0.94** (480)

All data are expressed as natural logarithms of differences with respect to the cross-country averages. Country-specific intercepts are included in the testing equation. The p value of the test when the null hypothesis of unit root is *not* rejected is in bold. The null of unit root is accepted at significance level α when the p-values are bigger than $\alpha/100$. The number of observations is reported in parentheses. Automatic selection of lags based on the Akaike Information Criterion.
ADF and PP are two tests that uses Fisher's (1931) result to derive test that combine the p-values from individual unit roots tests. The tests are distributed as a χ^2 with 2*N degrees of freedom where N is the number of cross-sections.

Table 2c Panel unit root test (p-values): De-trended and cross-sectionally de-meaned data, EU-15, 1970-2003

Method / Variable	Cyclically-adjusted primary government expenditure	Potential output
H_0: Unit root (common unit root process)		
LLC t		
Level	0.00 (482)	**0.82** (451)
First Difference	0.00 (473)	0.11 (459)
Breitung t-stat		
Level	**0.60** (467)	**0.07** (459)
First Difference	0.00 (458)	**0.31** (444)
H_0: Unit root (individual unit root process)		
IPS t-stat		
Level	**0.61** (482)	**0.94** (451)
First Difference	0.00 (473)	**0.07** (459)
ADF-MW Chi-square		
Level	**0.48** (482)	**0.28** (451)
First Difference	0.00 (473)	**0.08** (459)
PP - MW Chi-square		
Level	**0.85** (495)	0.03 (495)
First Difference	0.00 (480)	**0.63** (480)

All data are expressed natural logarithms of differences with respect to the cross-country averages. Country specific intercepts are included in the testing equation. The p value of the test when the null hypothesis of unit root is *not* rejected is in bold. The null of unit root is accepted at significance level α when the p-values are bigger than $\alpha/100$. The number of observations is reported in parentheses. Automatic selection of lags based on Akaike Information Criterion. ADF and PP are two tests that uses Fisher's (1931) result to derive test that combine the p-values from individual unit roots tests. The tests are distributed as a χ^2 with 2N degrees of freedom, where N is the number of panels.

Table 2d Cross sectional augmented IPS test for panel unit root (CIPS), EU-15, 1970-2003

Variable	Cyclically-adjusted primary government expenditure		Potential output	
	Without trend	With trend	Without trend	With trend
p=1	-0.22	-0.30	-0.074	-0.05
p=2	-0.25	-0.34	-0.069	-0.04
p=3	-0.29	-0.40	-0.072	-0.05

H_0: unit root. The critical values for the CIPS are tabulated in Pesaran (2005). For T=30 and N=15 the 5% critical value of the test in the case of models with an intercept is -2.25; for models with an intercept and a linear trend the critical value is -2.76. The 1% critical values are, respectively, -2.45 and -2.96. p is the number of lags in the cross-sectionally augmented Dickey-Fuller test.

Table 3 Panel cointegration test between cyclically adjusted primary government expenditure and potential output, EU-15, 1970-2003

	Original data	Cross-sectionally de-meaned data	De-trended	De-trended and cross-sectionally de-meaned data
With trend				
Panel ADF	0.2	-0.43		
Group ADF	-1.00	**-1.91**		
Without trend				
Panel ADF	-1.61	-0.28	**-2.10**	**-2.61**
Group ADF	**-1.91**	-0.94	**-3.31**	**-4.14**

H_0: no cointegration. The critical level of the test at 5% is -1.65. The calculated statistics must be in absolute value larger than this value to reject the null hypothesis of absence of cointegration for all units in the panel. In bold are reported the values for which it is not possible to reject the null at the 5 % level.

Table 4: Pooled Mean Group ECM estimates: common parameters, EU-15, 1970-2003

Long and short term elasticities	
Long-run elasticity $(-\frac{\beta_i}{\phi_i})$	0.93*** (6.83)
Error correction coefficient (cross country average of error correction coefficients ϕ_i)	-0.35*** (-4.83)
Short run coefficients	
GDP	0.33*** (4.83)
Δ government expenditure(-1)	0.07 (1.86)
Δ government expenditure(-2)	0.034 (1.00)
Δ potential output	-1.28 (-1.5)
Δ potential output (-1)	0.40 (0.84)
Δ potential output (-2)	-0.27 (-1.00)
Intercept	-0.69** (-2.08)
EU-15 potential output	0.15* (1.7)
Hausman test for poolability of countries	0.89 (0.34)

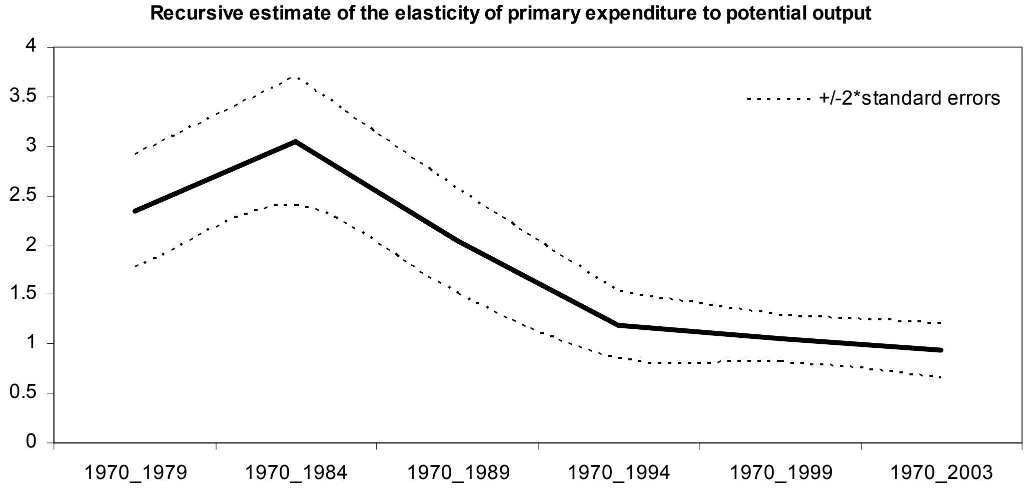
Lags are selected by the Akaike Information Criterion. t-statistic are reported in parentheses. The error correction coefficient measures the speed of adjustment and is computed as the average of each country speed of adjustment. The Hausman test is a test of poolability of the long-run coefficient (i.e. of the restriction that all countries have the same long-run elasticity). t-statistic in parentheses excepted for the Hausman where p-values are reported in parenthesis. The null of homogenous long-run coefficient is accepted at 5% when the p-values are bigger than 0.05. *** Significant at 1%; ** Significant at 5%; * Significant at 10%.

Table 5. Pooled Mean Group ECM estimates: country-specific parameters and specification tests EU-15, 1970-2003

	Number of lags selected by AIC[1]	Error correction coefficients ϕ_i	**Auto-correlation test**	**Functional form test**	**Norma-lity test**	**Heterosch-edasticity test**	**Adjusted R square**
Belgium	0; 1	-1 (NA)	43.45	31.9	0.54	8.75	-0.71
Denmark	1; 0	-0.318 (2.7)	0.14	0.66	0.17	0.43	0.15
Germany	1; 2	-0.150 (1.7)	4.66	11.13	27.0	8.33	0.38
Greece	1; 2	-0.570 (3.9)	1.33	0.12	0.02	0.08	0.26
Spain	2; 0	-0.064 (1.6)	0.05	7.39	1.69	0.54	0.37
France	1; 0	-0.126 (2.8)	0.46	12.98	1.61	3.53	0.52
Ireland	1; 2	-0.558 (3.5)	0.26	0.62	0.53	13.88	0.36
Italy	2; 2	-0.087 (1.06)	0.27	1.54	0.25	11.00	0.21
Luxembourg	1; 1	-0.287 (2.8)	0.12	0.48	0.40	0.34	0.20
Netherlands	2; 0	-0.133 (2.8)	1.58	7.35	0.69	7.54	0.39
Austria	1; 0	-0.139 (2.6)	0.54	5.36	15.41	0.77	0.45
Portugal	1; 1	-0.093 (1.01)	2.06	0.39	2.93	0.06	0.48
Finland	1; 1	-0.446 (3.2)	0.35	2.05	70.11	7.45	0.33
Sweden	2; 3	-0.572 (5.7)	1.21	0.03	0.33	0.34	0.50
United Kingdom	3; 1	-0.70 (7.2)	4.12	0.26	0.95	0.65	0.72

In the first column, the first figure indicates the lag selected by the Akaike Information Criterion for the dependent variable; the second number refers to the lags selected by the Akaike Information Criterion for the explanatory variable. t-statistics of the error correction coefficients in parentheses. Diagnostic checks refer to the equations that pool the long-run coefficients but leaving unconstrained the short-run dynamics. Specification tests are as follows: Godfrey's test of residual serial correlation distributed as $\chi^2(1)$ under the null of no autocorrelation; Ramsey's RESET test of functional form distributed as $\chi^2(1)$ under the null of no autocorrelation; Jarque-Bera's test of normality distributed as $\chi^2(2)$ under the null of no autocorrelation; Test of Heterosckedasticity distributed as $\chi^2(1)$ under the null of no autocorrelation

Figure 3: Results from recursive estimates EU-15, different sub-periods



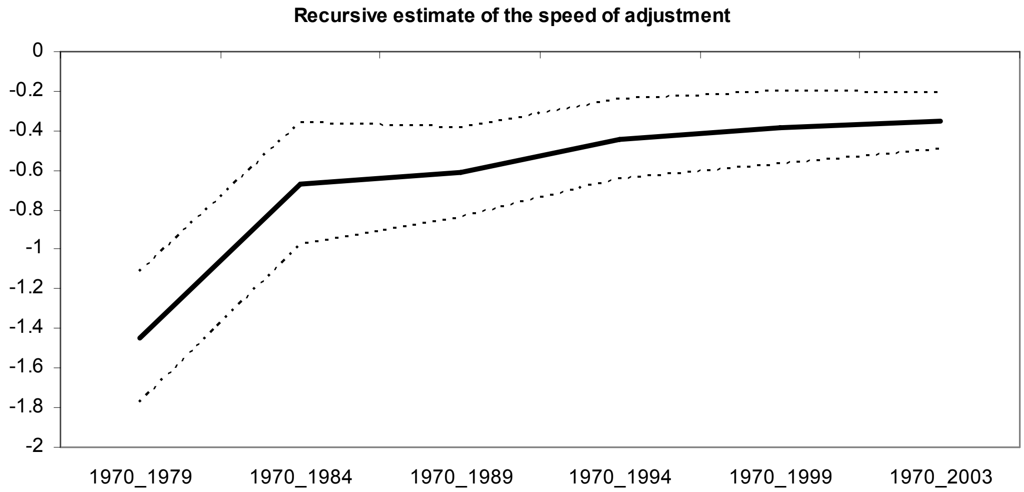


Table 6: Pooled Mean Group ECM estimates, EU-15 over different sub-periods

	Coefficient values	Hausman test for poolability of countries
1970-2003		
Long-run Coefficients	0.93*** (6.8)	3.15 (0.08)
Error Correction Coefficient	-0.35*** (-4.8)	
1970-1989		
Long-run Coefficients	2.03*** (7.8)	0.13 (0.72)
Error Correction Coefficient	-0.61*** (-5.4)	
1990-2003		
Long-run Coefficients	1.18*** (4.35)	1.25 (0.26)
Error Correction Coefficient	-0.69*** (-4.86)	

Lags are selected by the Akaike Information Criterion. t-statistic are reported in parentheses. The Hausman test is a test of poolability of the long-run coefficient (i.e. of the restriction that all countries have the same long-run elasticity). The p-value is reported in parenthesis. ***, **, and * denote, statistical significance at, respectively, 1%, 5%, and 10% level.

Figure 4 Cross-sectional stability of the long-run elasticity of government expenditure with respect to potential output

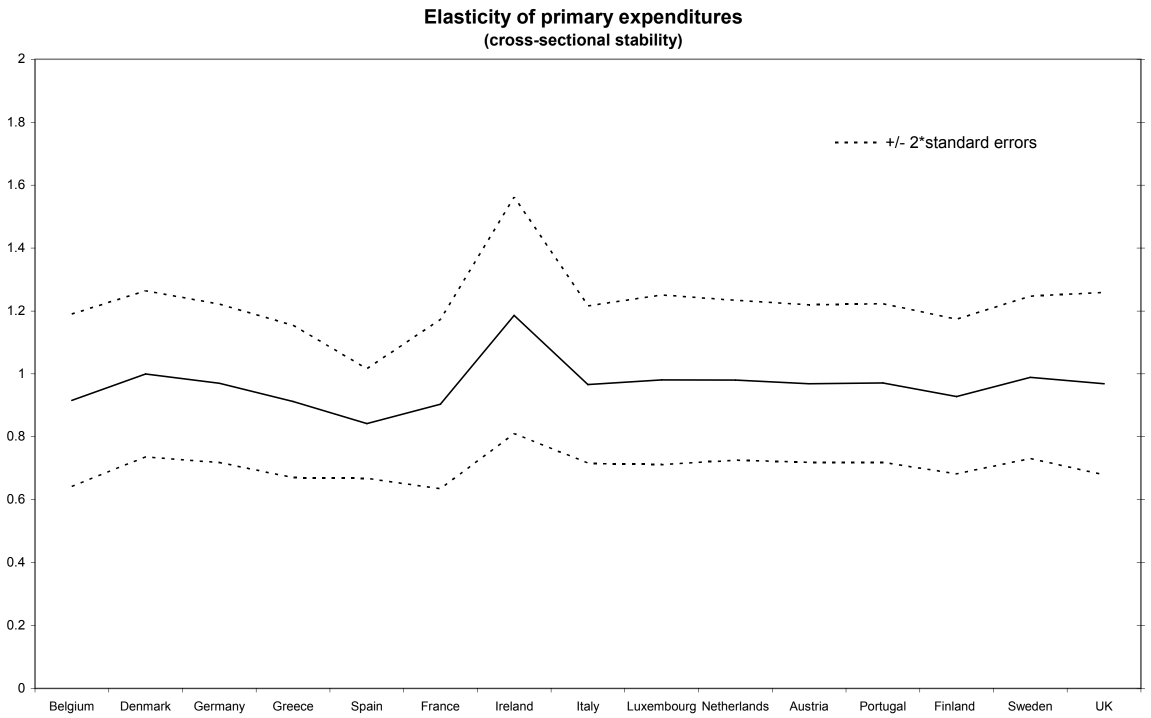


Chart 5b Cross-sectional stability of the speed of adjustment.

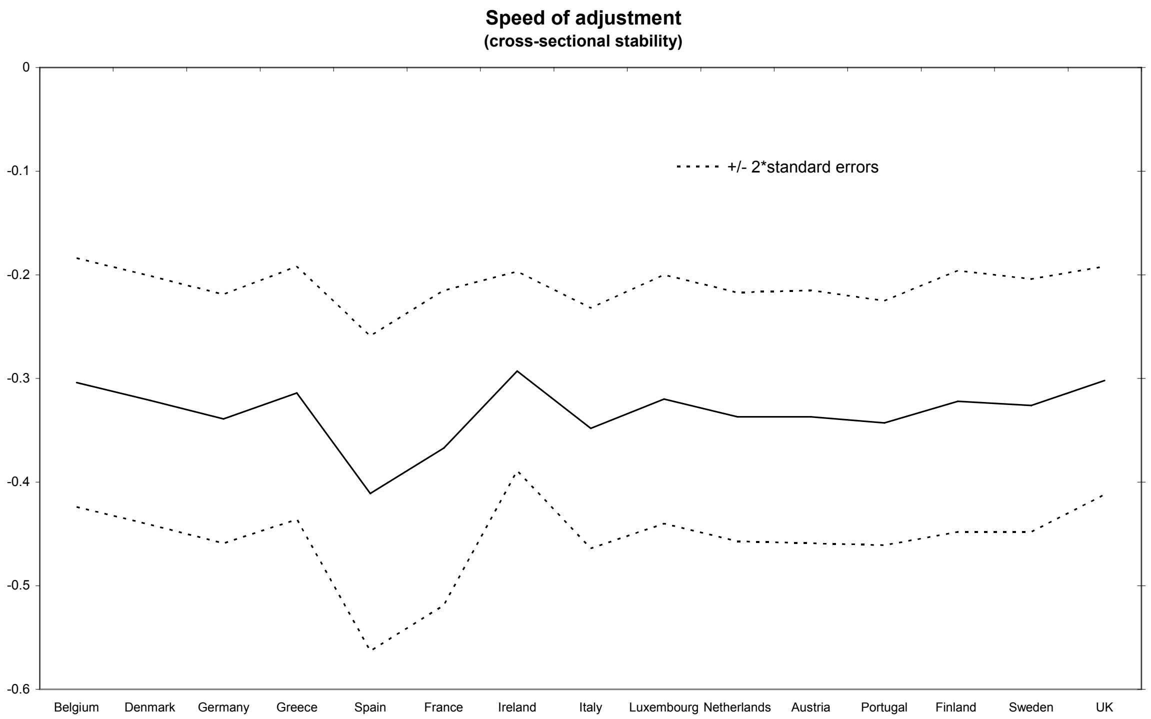


Table 8: Pooled Mean Group ECM estimates, different EU-15 sub-samples, 1970-2003

	Coefficient values	Hausman test for poolability of countries
Full Sample		
Long-run coefficient	0.93	3.15
	(6.83)	(0.08)
Error correction coefficient	-0.35	
	(-4.83)	
Development stage: split based on the median level of GDP per capita in 1970		
Low GDP per capita		
Long-run coefficient	3.1***	0.05***
	(19.6)	(0.83)
Error correction coefficient	-0.25***	
	(-2.1)	
High GDP per capita		
Long-run coefficient	1.04***	0.01***
	(5.3)	(0.91)
Error correction coefficient	-0.40***	
	(-3.6)	
Low GDP per capita		
Long-run coefficient	3.1***	0.05***
	(19.6)	(0.83)
Error correction coefficient	-0.25***	
	(-2.1)	
Demography : split based on the median yearly change in the dependency ratio over the sample		
Slow ageing countries		
Long-run coefficient	0.61***	0.20***
	(4.6)	(0.65)
Error correction coefficient	-0.35***	
	(-3.5)	
Fast ageing countries		
Long-run coefficient	1.5***	0.42***
	(5.1)	(0.51)
Error correction coefficient	-0.27***	
	(-3.3)	
Slow ageing countries		
Long-run coefficient	0.61***	0.20***
	(4.6)	(0.65)
Error correction coefficient	-0.35***	
	(-3.5)	
Public finances: split based on the median level of the average debt/GDP ratio		
Low average debt		
Long-run coefficient	1.5	0.89
	(5.7)	(0.34)
Error correction coefficient	-0.26	
	(-3.96)	
High average debt		
Long-run coefficient	0.6	3.73
	(4.1)	(0.05)
Error correction coefficient	-0.52	
	(-3.53)	
.Fiscal governance: split based on the median level of expenditure rule indexes		
Above Median countries		
Long-run coefficient	0.28	0.05
	(1.7)	(0.83)
Error correction coefficient	-0.32	
	(-3.3)	
Below Median countries		
Long-run coefficient	3.2	0.89
	(20.9)	(0.35)
Error correction coefficient	-0.39	
	(-2.7)	

Lags are selected by the Akaike Information Criterion. t-statistic are reported in parentheses. The Hausman test is a test of poolability of the long-run coefficient (i.e. of the restriction that all countries have the same long-run elasticity). The p-value is reported in parenthesis. ***, **, and * denote, statistical significance at, respectively, 1%, 5%, and 10% level.

Table 9: Variables used to split the sample of countries (see Table 8)

	Debt/GDP ratios		GDP per capita		Yearly change in dependency ratio		Expenditure rule index	
	Avg. level	Rank	Level in 1970	Rank	Avg. level	Rank	Avg. level	Rank
Belgium	98.4	16	3697.2	9	20.4	8	0.49	9
Denmark	49.5	9	4783.4	14	16.5	5	0.79	11
Germany	40.4	5	3748.8	10	19.1	6	0.90	13
Greece	63.9	13	2473.6	3	39.8	11		
Spain	36.9	3	2729.5	4	50.5	14	0.25	7
France	39.8	4	3764.1	11	19.8	7	-0.33	5
Ireland	70.9	14	2219.8	2	-16.3	1	-0.67	2
Italy	85.0	15	3417.2	5	48.4	12	-0.66	3
Luxembourg	10.5	1	5064.2	15	9.0	3	1.27	14
Netherlands	59.8	12	4051.3	12	22.6	9	0.84	12
Austria	46.7	7	3434.3	6	0.4	2	-0.54	4
Portugal	46.1	6	1849.3	1	49.8	13	-1.03	1
Finland	25.4	2	3453.6	7	52.1	15	0.26	8
Sweden	49.4	8	4604.9	13	24.7	10	0.14	6
United Kingdom	51.5	11	3640.3	8	14.0	4	0.68	10

Per-capita GDP data are based on Penn World Tables mark 6.1. Expenditure indexes are described in Ayuso et al. (2006). The average reported in the table covers the period 1990-2003, i.e., the whole years for which index data are available. The source of debt/GDP ratios and dependency ratios are AMECO database and Eurostat, respectively.

Appendix

Table A1: Pooled Mean Group ECM estimates: common parameters. Unobserved common component, EU-15, 1970-2003

Long and short term elasticities	
Long-run elasticity $(-\frac{\beta_i}{\phi_i})$	1.29^{***} (8.14)
Error correction coefficient (cross country average of error correction coefficients ϕ_i)	-0.46^{***} (-6.58)
Short run coefficients	
GDP	0.59^{***} (6.58)
Δ government expenditure(-1)	0.12^{**} (2.02)
Δ government expenditure(-2)	0.045^{*} (1.73)
Δ potential output	-0.39 (-0.52)
Δ potential output (-1)	0.61 (0.83)
Δ potential output (-2)	-0.02 (0.13)
Intercept	0.04 (0.76)
Trend	-0.001 (-0.65)
Trend square	0.00 (0.36)
Hausman test for poolability of countries	0.89^{***} (0.34)

Lags are selected by the Akaike Information Criterion. t-statistic are reported in parentheses. The Hausman test is a test of poolability of the long-run coefficient (i.e. of the restriction that all countries have the same long-run elasticity). The p-value is reported in parenthesis. ***, **, and * denote, statistical significance at, respectively, 1%, 5%, and 10% level.

Table A2: Mean Group ECM estimates: cross-country average of estimated parameters, EU-15, 1970-2003

Long and short term elasticities	
Long-run elasticity $(-\frac{\beta_i}{\phi_i})$	2.16*** (3.1)
Error correction coefficient (error correction coefficients ϕ_i)	-0.44*** (-5.66)
Short run coefficients	
GDP	0.95*** (2.70)
Δ government expenditure(-1)	0.095** (2.50)
Δ government expenditure(-2)	0.034 (1.0)
Δ potential output	-1.36 (-1.22)
Δ potential output (-1)	1.08 (1.08)
Δ potential output (-2)	-0.29 (-1.00)
I ntercept	0.21 (0.17.1)
EU-15 potential output	-0.92 (1.30)

Lags are selected by the Akaike Information Criterion. t-statistic are reported in parentheses. ***, **, and * denote, statistical significance at, respectively, 1%, 5%, and 10% level.

References

Abizadeh, S., and M. Yousefi, (1988), An empirical examination of Wagner’s law, Economics Letters, 26, 169-73.

Alesina, A., and R. Wacziarg, (1998), Openness, country size, and government, Journal of Public Economics, 49, 305-21.

Alesina, A., Perotti, R. (1995) Fiscal expansions and Adjustments in the OECD Countries, Economic Policy 21, 207-248.

Alesina, A., Perotti, R. (1996) Budget beficits and budget institutions, NBER Working Paper , 5556.

Axfentiou, P.C., (1992), The importance of transfers in Canada: an analysis within the framework of Wagner’s law, Rivista Internazionale di Scienze Economiche e Commerciali, 39, 709-25.

Ayuso-i-Casals. J, D. Gonzalez Hernandez, L. Moulin, and A. Turrini, (2006), "Beyond the SGP: Features and effects of EU national-level fiscal rules", paper presented at the Workshop organised by the European Commission on "The role of national fiscal rules and institutions in shaping budgetary outcomes", Brussels 24 November 2006.

Baltagi, B. H. (2000) “Non stationary panels, panel cointegration, and dynamic panels”, Amsterdam: JAI Press.

Baltagi, B. H. (2005) “Econometric Analysis of Panel Data”, John Wiley & Sons.

Baltagi, B. H. and Kao, C. (2000), “Non stationary panels, cointegration in panels, and dynamic panels: a survey” in B. Baltagi (ed.), *Advances in Econometrics, Vol. 15: Non-stationary Panels, Panel Cointegration, and Dynamic Panels*, Amsterdam: JAI Press

Banerjee, A (1999) “Panel unit roots and cointegration:an overview” Oxford Bullettin of Economics and Statistics, Special Issue.

Banerjee, A., Marcellino, M. and C. Osbat (2004), “Some cautions on the use of panel methods for integrated series of macroeconomic data” *Econometrics Journal*, 7, 322-340.

Banerjee, A., Marcellino, M. and C. Osbat (2005), “Testing for PPP should we use panel methods”, *Empirical Economics*, 30 pp 77-91.

Blanchard, O., Perotti, R. (2001) An empirical characterisation of the dynamic effects of changes in government spending and taxes on output, mimeo.

Bogetic, Z. Randolph, S. and D. Heffley (1996) “Determinants of public expenditure on infrastructure”, *World Bank Policy Research Working Paper* 1661.

Bohl, M. T. (1996), Some international evidence on Wagner's Law, *Public Finance*, 51, 185-200.

Bouthevillan, C., Cour-Thimann P., Van den Dool, G., Hernandez de Cos, P., Langenus, G., Mohr, M., Momigliano, S., Tujula, M. (2001), Cyclically adjusted budget balances: an alternative approach, ECB Working Paper , 77.

Breitung, J. (2000). “The local power of some unit root tests for panel data,” in B. Baltagi (ed.), *Advances in Econometrics, Vol. 15: Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, Amsterdam: JAI Press, p. 161–178.

Breitung, J. and Pesaran, M. H. (2005), "Unit roots and cointegration in panels". *Institute for Economic Policy Research* (IEPR) Research Papers Series , 32, forthcoming in L. Matyas, and P. Sevestre, The Econometrics of Panel Data (Third Edition), Kluwer Academic Publishers.

Buti, M. P. van den Noord (2003) Discretionary fiscal policy and elections: the experience of the early years of EMU, OECD Working Paper , 351.

Chang, T., (2002), An econometric test for Wagner's law for six countries based on cointegration and error correction modeling techniques, Applied Economics, 34, 1157-1169.

Choi, I. (2005), “Nonstationarity panels” in Palgrave Handbook of Econometrics vol. 1 Econometric theory Patterson, K. and Mills, T.C. ed.

D’Addio, A., and Mira D’Ercole, M. (2005), "Policies, institutions and fertility rates: A panel data analysis in OECD countries," *OECD Economic Studies* , 41.

Denis, C., K. McMorrow, and W. Roeger, (2002), Production function approach to calculating potential output growth and output gaps- Estimates for the EU Member States and the US, European Economy, Economic Paper , 176.

Engle, R. F., and C. W. J. Granger (1987), Cointegration and error correction: representation, estimation, and testing, *Econometrica*, 55, 251-76.

European Commission (2002) Public Finances in EMU 2002, European Economy , 3/2002, Bruxelles.

European Commission (2006), Public Finances in EMU 2002, European Economy , 3/2002, Bruxelles.

Fatás, A., Mihov, I. (2002) "The case for restricting fiscal policy discretion", CEPR Discussion Paper , 3277.

Fatàs, A., J. von Hagen, A. Hughes Hallet, A. Siebert, R. R. Strauch (2003) *Stability and Growth in Europe: Towards a Better Pact*, Centre for Economic Policy Research.

Funke, K. and Nickel, C., (2006), "Does fiscal policy matter for the trade account? A panel cointegration study", ECB Working Paper, 620.

Golinelli, R. and S. Pastorello, (2002) "Modelling the demand for M3 in the Euro area", *The European Journal of Finance*, 8, 371-401

Halleberg, M., Strauch, R., von Hagen, J. (2001) "The use and effectiveness of budgetary rules and norms in EU Member States," Report prepared for the dutch Ministry of Finance by the Institute of European Integration Studies.

Heller, P. S., and J. Diamond, (1990), "International Comparison of Government Expenditure Revisited: the Developing Countries 1975-1986", IMF Occasional Paper, 69.

Henrekson, M., (1993), "Wagner's law – a spurious relationship?" *Public Finance*, 48, 406-15.

Heston, A., R. Summers and B. Aten (2002), Penn World Table Version 6.1, Center for International Comparisons at the University of Pennsylvania (CICUP).

Hughes Hallet, A., J. Lewis, and J. Von Hagen, (2004), *Fiscal Policy in Europe, 1991-2003. An Evidence-based Analysis*, CEPR, London.

Im, K., Pesaran, M.H., Shin, Y (2003), "Testing Unit Roots in Heterogeneous Panels", *Journal of Econometrics*, 115, 53-74.

Kneller, R., M. F. Bleaney, and n. Gemmell, (2004), "Fiscal policy and growth: Evidence from OECD countries", *European Journal of Political Economy*, 74, 171-190.

Kolluri, B.R., M. J. Panik, M.S. Wahab (2000), "Government expenditure and economic growth: evidence from G7 countries", *Applied Economics*, 32, 1059-1068.

Cameron, G and Tan, K.Y., (2006), "Sovereign risk in the classical gold standard era", University of Oxford Discussion Paper, 258.

Larch, M., and M. Salto, (2003) "Fiscal rules, inertia and discretionary fiscal policy", Economic Paper, 194, European Economy.

Levin A., Lin, C.-F. and C.-S.J. Chu (2002), "Unit roots tests in panel Data: Asymptotics and finite sample properties", *Journal of Econometrics* 108, 1-24.

Levine, R., and D. Renelt, (1992), "A sensitivity analysis of cross-country growth regressions", *American Economic Review*, 82, 942-963.

Loayza, N., Ranciere, R. (2006), "Financial development, financial fragility, and growth" *Journal of Money, Credit, and Banking*, 38,. 1051-1076.

Mackenzie, G. A. (1989) "Are all summary indicators of the stance of fiscal policy misleading?" *IMF Staff Papers* 36, 743-70.

Martinez-Mongay, C. (2002) "Fiscal policy and the size of governments", in Buti, von Hagen, Martinez-Mongay (eds), *The Behaviour of Fiscal Authorities*, Palgrave.

Milesi Ferretti, G.M., R. Perotti, and M. Rostagno, (2002), "Electoral systems and public spending", *Quaterly Journal of Economics*, 117, 609-656.

Nijkamp, P., and J. Poot (2004), "Meta-analysis of the effect of fiscal policy on long-run growth", *European Journal of European Economy*, 20, 91, 124.

Oxley, L., (1994), "Cointegration, causality, and Wagner's law: a test for Britain 1870-1913", *Scottish Journal of Political Economy*, 41, 286-98.

Payne, J., E., and B. T. Ewing, (1996), "International evidence on Wagner's hypothesis: A cointegration analysis", *Public Finance*, 51, 258-74.

Peacock, A., and A. Scott, (2000), "The curious attraction of Wagner's Law", *Public Choice*, 102, 1-17.

Pedroni, P. (2001) “Purchasing power parities tests in cointegrated panels”, *Review of Economic and Statistics*, 83, 727-731.

Pedroni, P. (2000) “Critical values for cointegration tests in heterogeneous panels with multiple regressors," *Oxford Bulletin of Economics and Statistics*, 61, 653-670.

Pedroni, P. (2000) "Fully-modified OLS for heterogeneous cointegrated panels," in B. Baltagi (ed.), *Advances in Econometrics, Vol. 15: Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, Amsterdam: JAI Press, 93-130.

Persson, T. , and G. Tabellini, (1999), "The size and scope of government: comparative politcs with rational politicians", *European Economic Review*, 63, 699-735.

Perotti, R. (2002) "Estimating the effects of fiscal policy in OECD countries", ECB Working Paper,168.

Persson, T., and G. Tabellini, (1990), "Comparative politics and public finance", *Journal of Political Economy*, 108, 1121–1161.

Pesaran, M. H. and Smith, R. (1995), “Estimating long-run relationship from dynamic heterogeneous panel”, *Journal of Econometrics,* 68, 79-113.

Pesaran, M. H., Shin, Y. and Smith, R.(1999) “Pooled mean group estimator of dynamic heterogeneous panels” *Journal of the American Statistical Association,* Vol 94 pp.621-634.

Pesaran M. H., (2003), "A simple panel unit root test in the presence of cross section dependence," Cambridge Working Papers in Economics, 346.

Pesaran, M. H. (2005), "A simple panel unit root test in the presence of cross section dependence", Cambridge Working Papers in Economics 346, revised version.

Phillips, P.C.B. and Moon, H. R. (1999). "Nonstationary panel data analysis: An overview of some recent developments", Cowles Foundation Discussion Papers 1221, Cowles Foundation, Yale University.

Phillips, P.C.B. and Moon, H. R. (1999), “Linear regression limit theory for nonstationary panel data”, *Econometrica*, 67, 1057-1112.

Ram, R. (1987) "Wagner’s hypothesis in time-series and cross-section perspectives: Evidence from ‘real’ data for 115 countries," *Review of Economics and Statistics*, 69, 194-204.

Rodrik, D., (1998), "Why do more open economies have bigger governments?", *Journal of Political Economy*, 151, 997-1032.

Sapir, A., P. Aghion, G. Bertola, M. Hellwig, J. Pisani-Ferry, D. Rosati, J. Vinals and H. Wallace with M. Buti, M. Nava and P. M. Smith, (2004), *An Agenda for a Growing Europe – The Sapir Report*, Oxford University Press, Oxford.

Smith (2000), R.P. “Estimation and inference with non-stationary panel time-series data” mimeo, Department of Economics, Birbeck College, London.

Tanzi, V. and L. Schuknecht, (2000), *Public Spending in the 20th Century: A Global Perspective*, Cambridge University Press, Cambridge.

Van den Noord, P. (2000) "The Size and Role of Automatic Stabilizers in the 1990s and Beyond", OECD Economics Department Working Paper , 230.

Wahab, M. (2004) "Economic growth and government expenditure: evidence from a new test specification", *Applied Economics*, 36, 2125-2135.