Social Welfare and Government Size*

J. Andrés[a], J. Bandrés[b], R. Doménech[a,c] and L. Gadea[b]

a University of Valencia
b University of Zaragoza
c BBVA Research

Preliminary draft, December, 2022

Abstract

In this paper, we analyze the effect of government size on social welfare and GDP per capita growth, using different econometric techniques and robustness analysis for the sample of 36 OECD countries in the last six decades. Our results show that the negative effects are smaller in absolute terms in the case of welfare than for GDP per capita. This result is robust to changes in the estimation method, to the use of smoothed variables, and to the inclusion of dummy variables that control for expansions and recessions over the business cycle. More interestingly, we find that the effect of government size follows an inverted U-shape: positive and greater on social welfare than for GDP per capita growth when government size is below 35% to 40%, and negative beyond that threshold. The range of values for which government size positively affects growth and welfare expands significantly with the quality of government.

Keywords: social welfare, GDP per capita, growth, government size.

JEL Classification: L11; L43; F11cm

1. Introduction

In the last decades, there has been a large body of literature evaluating the empirical effects of government size on GDP per capita growth (see, for example, the survey by Bergh and Henrekson, 2011). The evidence is far from conclusive and varies with model specifications, variable definitions, sample of selected countries, time periods, data quality, or estimation methods. For advanced economies, the empirical literature points to a negative effect of government size, although this result is not sufficiently robust either. Among other things, economic prosperity also depends on many public activities, such as the quality of public services and regulations or the configuration of the institutional framework. The literature on the effects of public spending on inequality is also abun-

* J. Andrés and R. Doménech thank the financial support of grant PID2020-116242RB-I00 funded by the Spanish Ministry of Science and Innovation (MCIN/AEI/ 10.13039/501100011033), grant GVPROMETEO2020-083 funded by the Generalitat Valenciana, and grant TED2021-132629B-I00 funded by MCIN/AEI/ 10.13039/501100011033 and the "European Union NextGenerationEU/PRTR". María Dolores Gadea acknowledges the financial support from the grants PID2020-114646RB-C44, and RED2018-102563-T funded by MCIN/AEI/.

dant. In general, larger and more effective governments tend to reduce inequality, mainly through spending on education and family and child support, and ex-post distributional policies, such as taxes and transfers.

Extending previous findings in the empirical literature, in this paper we systematically compare the impact of government size on GDP per capita growth with its effect on the rate of growth of a more comprehensive measure of social welfare like the one proposed by Jones and Klenow (2016) in a sample of 36 OECD countries in the last six decades. These authors show how their welfare index can be rigorously derived from the individual preferences used in the economic analysis of welfare and can be expressed as a function of private and consumption per capita, leisure (which depends on the number of hours worked), the distribution of income and life expectancy.

The preliminary evidence shows that there is a lot of heterogeneity across time and space on the correlation between welfare and public spending over *GDP*, which we take into account in our econometric analysis. Panel data estimates using annual data and fixed effects, which control for differences in GDP per capita and welfare steady states, show a negative and statistically significant impact of government size on both GDP per capita and welfare. In general, these effects are smaller in absolute terms in the case of welfare than for GDP per capita. This result is robust to changes in the estimation method, like the Mean Group (Pesaran and Smith, 1995) or the Pooled Mean Group (Pesaran et al., 1999) estimators, to the use of smoothed variables, either the Hodrick-Prescott filter or ten-year averages, and to the inclusion of dummy variables that control for expansions and recessions over the business cycle.

When we allow the effect of government size to be time-varying, we find that the estimated coefficients are negative, statistically significant and rather stable until the Great Recession. Since the beginning of the financial crisis to the sovereign debt crisis, the negative effect of government size on growth has almost doubled. There are two alternative explanations for this result. The first is that this was just a period of higher than normal government size and lower than normal growth rates. For this argument to be valid, it is necessary to assume that the time dummies included in our specification do not capture this singularity well. The second is that increases in government size during the Great Recession, though preventing bigger falls in economic activity at the time, dragged down economic growth in the following years. The aim of achieving the strongest multiplier effect in the sort run and of avoiding further deterioration in the distribution of income induced many governments to increase transfers at the cost of reducing public investment and other productive spending. Additionally, we find that the effect on welfare is smaller in absolute value than for GDP per capita, during most of the sample period, although the difference is smaller in the last six years.

Finally, we have tested the presence of non-linear effects of government size on growth, controlling for government quality. The effects on growth can be equally negative when the government is too small, with insufficient provision of critical public services to the economy, or when it is too large, so that high taxes are a drag on the efficiency of the private sector. At the same time, the optimal government size may be different for each country depending on the efficiency level of the public sector. Thus, we expect that countries with more efficient public sector will have larger governments. Our results show that the effect of government size on welfare and GDP per capita growth follows an inverted U-shape, being positive when this size is below 35% to 40% and turning negative beyond that level. We also find that the efficiency with which public resources are used matters, so that government quality improvements substantially expand the range of values for which government size has a positive effect on both indicators of well-being.

The structure of this paper is as follows. In Section 2, we survey the literature on social welfare and analyze the properties of the welfare measure used in the rest of the paper. The third section reviews the empirical evidence on the effects of government size on GDP per capita growth and inequality. Section 4 presents the main results of our panel data estimates, whereas section 5 analyses the robustness of these results to non-linear effects over the business cycle, to alternative smoothing approaches and to the non-linear effects of government size, taking into account government quality. Finally, Section 6 presents the main conclusions.

2. Social welfare and GDP per capita

GDP per capita is the most commonly used measure in comparative studies of economic performance, both across economies and time. It summarizes the value of market activities that capital and labor produce in an economy in a specific period. This makes it, in the best case, a partial measure of economic welfare (Aitken, 2019). For one thing, GDP misses nonmarket activities, the opportunity cost of leisure foregone in the production process, and the value of some goods and services associated, for example, with the digital economy. From a broader dynamic perspective, it has difficulties to accommodate the fact that a continuous increase in the amount of production also requires a growing environmental cost that will be unequally borne by current and future generations. But most importantly of all, being an average, GDP per capita fails to capture the important effect on personal well-being that unequal distribution of income has on aggregate welfare. Leaving aside preferences for more or less equity in the distribution, risk-averse individuals value more the welfare foregone for being in the lowest deciles of the income distribution than the increase in welfare they may achieve for being in the upper deciles.

In fact, most attempts to provide better indicators of social welfare include some measure of the inequality of income distribution among individuals in the economy. Nevertheless, it is difficult to integrate production or other similar macroeconomic variables along with inequality measures into a simple composite welfare indicator and many such proposals of broader welfare indicators have severe limitations of data gathering, homogeneity and relative weights given each to their components.

Berik (2020) distinguishes four different recent approaches to deal with the complexity of measuring welfare for comparison purposes: composite indexes (as the UN Human Development Index), subjective evaluations (as the UN World Happiness Report), dashboards (as the OECD Better Life Initiative) and monetary approaches (for example, the Measure of Economic Welfare developed by Nordhaus and Tobin, 1972, or the more recent variant proposed by Jones and Klenow, 2016). The monetary approach is perhaps less comprehensive than other alternatives, but, in contrast to composite indexes or the dashboard approaches, it provides a well-grounded theory-based aggregation procedure of different determinants of welfare. Furthermore, it allows cross-country and intertemporal comparability in a way that subjective evaluation approaches cannot.

In this paper we follow the monetary approach and study the effect of public spending on welfare, the latter being approximated by the Jones and Klenow's (2016) index . This welfare index can be rigorously derived from the individual preferences used in the economic analysis so that welfare and can be expressed as a function of private and consumption per capita (C), leisure (ℓ, which depends on the number of hours worked), the distribution of income and life expectancy (which depends on the probability of survival, S, of living beyond certain age, a):

$$U = E \sum_{a=1}^{100} \beta^a u(C_a, \ell_a) S(a) \tag{1}$$

The comparison of welfare across countries and time is made in terms of the equivalent annual consumption necessary for a person randomly chosen in any country to be indifferent to living in another (e.g., the United States). Although this form of comparison may seem very different from that carried using GDP, the underlying principle is similar, since with either indicator we are taking into account the capacity to enjoy the set of goods and services produced in each country. Furthermore, this comparison depends on how consumption is distributed among the individuals in an economy. If the average levels of consumption, leisure and life expectancy are equal, it is preferable to live in a country with less inequality, insofar as people are risk averse and/or they have personal preferences in favor of a more equitable distribution of income. This justifies that the comparison of equivalent annual consumption should take distributional issues into account.

If microeconomic data from representative surveys were available for all countries, a fairly robust approximation of equivalent annual consumption could be made for each country compared to the US. In practice, we do not have this information for all OECD countries, so some additional assumptions have to be made to use information from databases with macroeconomic information for large samples of countries. Jones and Klenow (2016) show that the approximation with macroeconomic data is quite good in the case of the 13 economies (advanced and developing) for which microeconomic data are available: the correlation of welfare levels measured with macroeconomic and microeconomic data is equal to 0.999.

In particular, the relative welfare measure (λ_i) of each OECD country in relation to the United States is calculated using the following expression:

$$\begin{aligned}\log\lambda_i = {} & \frac{e_i - e_{us}}{e_{us}}\left(\overline{u} + \log c_i + \nu\left(\ell_i\right) - \frac{1}{2}\sigma_i^2\right)\\ & + \log c_i - \log c_{us}\\ & + \nu\left(\ell_i\right) - \nu\left(\ell_{us}\right)\\ & - \frac{1}{2}\left(\sigma_i^2 - \sigma_{us}^2\right)\end{aligned} \tag{2}$$

where e, c, ν and σ are, respectively, life expectancy, per capita consumption, a function of leisure and the variance of consumption among individuals, for country i and the United States (us).

This index is particularly suitable for studying the economic and social impact of public spending. The role of the public sector in an economy is not only to foster productivity growth, through public investment, and the provision of public goods, such as property rights, defense, regulations or justice. Among its objectives is also to ensure that the benefits of the wealth created are spread among all members of society. Public spending affects market activities directly, whose effect on personal well-being is represented by consumption per capita and leisure (the latter reflecting the efficiency with which a given amount of goods and services are produced). It also influences the distribution of income and hence inequality in a country. Public spending is in particular a key determinant of pre-market inequality (schooling or health) and post-market inequality (taxes and transfers). Finally, health care, and other regulations, affect life expectancy.

When the information needed is available, the welfare measure for OECD countries has been calculated since 1960 or the first available year. Life expectancy at birth (e) is obtained from the Gapminder database (2020). Consumption per capita (c), GDP per capita (gdp), and the number of hours worked over the working age population are from PWT 10 (Feenstra, Inklaar and Timmer, 2015). For inequality of disposable income after taxes and transfers, we use the Gini coefficient of Eurostat (2020) and the OECD (2020a). Data from SWIID 8.3 (Solt, 2020), Atkinson et al. (2017) and Prados de la Escosura (2008), in the case of Spain, are used to extrapolate backward.

Data availability allows us to construct an unbalanced or incomplete panel for 35 OECD countries, with observations since 1960 for Australia, Austria, Belgium, Canada, Switzerland, Germany, Denmark, Spain, Finland, France, United Kingdom, Greece, Ireland, Italy, Japan, Korea, Netherlands, Norway, New Zealand, Portugal, Sweden and United States, since 1970 for Colombia, Mexico and Iceland, since 1980 for Hungary, since 1990 for Estonia, Israel, Lithuania, Slovakia, Slovenia and Turkey, and since 2000 for Luxembourg, Poland and Czech Republic.

Evidence in Figure 1 for OECD countries using averages from 2010 to 2019 shows that the measure of social welfare is closely related to GDP per capita, which is able to explain 83 percent of the welfare differences between most countries. However, behind this high correlation there are interesting differences between these two indicators. For example, in Ireland, GDP per capita is more than 22 percent higher than in the US, as a result of many multinationals having moved their tax bases to Ireland, but the welfare index corresponding to Ireland is 21 percent lower, because GDP per capita does not translate into a similar relative level of consumption per capita (see Byrne, Conefrey and O'Grady, 2021). Something similar happens in the case of Luxembourg: with a GDP per capita 47 per cent higher than in the US, its welfare level is only 13 per cent higher. In this case, this disparity is mainly explained by the weight of financial services, which represent around a third of Luxembourg's GDP. With the exception of these particular cases, and to a lesser extent Switzerland, if we compare the most advanced countries with the most backward OECD ones, the differences in welfare are greater than those observed in terms of GDP per capita, since in countries with lower per capita income, inequality and the number of hours worked tend to be greater, and life expectancy lower.

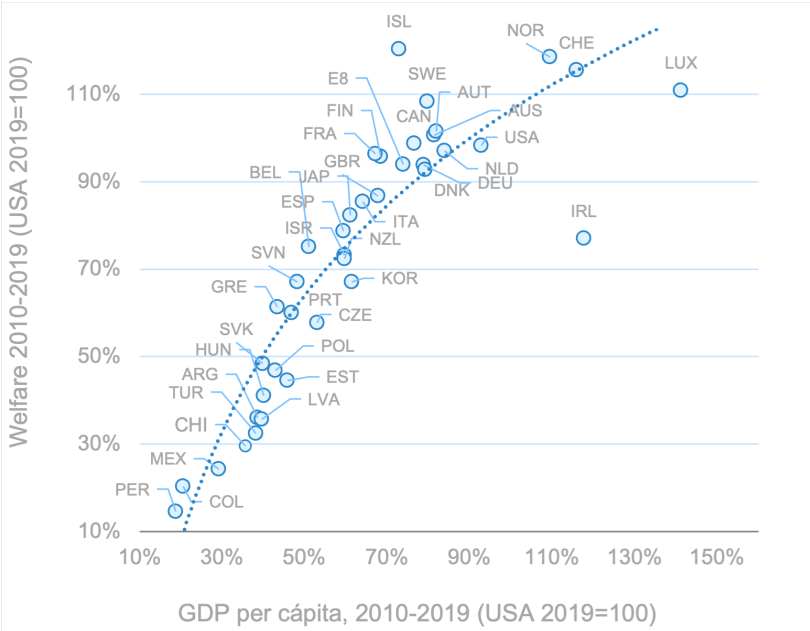


Figure 1: *GDP per capita and welfare in OECD and selected Latam countries, 2010-2019. Source: own elaboration based on Jones and Klenow (2016), PWT10, SWIID, Eurostat, OECD and Gapminder.*

In contrast to less developed countries, in the case of the advanced economies their differences in well-being with respect to the United States are significantly smaller than the distance measured in terms of GDP per capita. Longer life expectancy, a better distribution of income, and fewer hours worked partially compensate for the advantage of the United States relative to the European economies in GDP and per capita consumption. For example, the average gap of the eight most advanced European economies (E8) with respect to the United States was approximately 23 percentage points in 2018 when measured in income per capita terms, while it was just over 8 points in terms of the welfare index.

3. Government size, growth and inequality: Review of the evidence

Given the novelty of the more recent contributions to the monetary approach to social welfare, there is no empirical evidence on the effects of size and composition of public expenditure on welfare, in contrast to the large existing literature on their effects on economic growth and income inequality (see, for example, Awaworyi Churchill et al, 2017).

The complexity of the relationship between the size of public spending and growth is reflected in a vast empirical literature that, using different methodologies, time periods, and samples of countries, has tried to assess the extent to which public revenues and expenditures are a driving force for growth. The relationship between public spending and economic growth responds to two opposing forces. On the one hand, the positive effects of public spending on productivity, mainly through the accumulation of physical and human capital, the correction of market failures, and the improvement of social stability by reducing inequality. On the other hand, the distortionary effects of taxes on capital productivity and growth, as well as inefficiencies in public management and the extension of rent-seeking activities. The final outcome will depend on the relative importance of both effects, which, under certain assumptions about the shape of the production function, is mainly related to the size and composition of public spending and revenues. The combined impact of these two types of effects also suggests that the relationship between the size of public spending and growth could be non-monotonic (Tanzi and Zee, 1997).

Early empirical work mostly supported the hypothesis of a statistically significant and robust negative relationship between public spending and economic growth. However, more recent empirical findings have shown that this result is not always robust to the introduction of changes in the control variables and other issues associated with the specification of the models (Slemrod, 1995). Economic prosperity also depends, among other things, on other public activities that are not necessarily reflected in public finances, such as the quality of public services and regulations, or the configuration of the institutional framework.

An examination of the abundant empirical studies on the subject shows that one factor that has a very significant influence on the sign and value of the coefficients that relate the size of public expenditure to growth is the level of development of the countries in the sample. For example, Landau (1983) detected that the negative effect of public consumption on growth was mainly a characteristic of advanced economies, which disappeared for countries whose income was below the median of the more than 100 countries analyzed.

Grier and Tullock (1989) also obtained significant results on the negative influence of the increase in the public expenditure/GDP ratio on growth. However, in the sample of less developed countries, the estimated value of the coefficient was much lower and less statistically significant. When the analysis was confined to advanced countries, different econometric methodologies show very consistent and robust results, pointing to the negative effects of public spending on growth (Saunders, 1985, Hansson and Henrekson, 1994, or Pevcin, 2004). The surveys by Bergh and Henrekson (2011 and 2015) for advanced countries confirm the negative influence of spending on growth, even after correcting for possible reverse causality problems.

As regards the composition of public expenditures, Gemmell, Kneller and Sanz (2016) evaluate the long-run effects on GDP per capita of changes in total government expenditure and in the shares of different spending categories for a sample of OECD countries since the 1970s. The authors take into account the structure of taxes used to finance these expenditures and potential endogeneity problems of the variables. They find strong evidence indicating that reallocating total spending toward productive expenditures, such as infrastructure and education, has positive effects on long-run output levels. On the contrary, the reallocation of public spending towards social expenditures is associated with a modest negative effect on growth.

The literature on the effects of public spending on inequality is also abundant. Fournier and Johansson (2016) estimate the effect of public spending on inequality looking not only at the size but also at the composition of public spending in OECD countries over the 1987-2014 period. The more relevant findings are the following. First, larger and more effective governments tend to reduce inequality. Second, the size and effectiveness of the government have little effect on the distribution of market income. Thus, ex-post distributional policies are the main instruments used by OECD governments to reduce inequality: about two-thirds of the reduction in inequality between market and disposable income is due to transfers, and one-third due to taxes. Third, regarding composition, spending on education and family and child support unambiguously reduce inequality. Other subsidies (for example, to firms) and generous pension systems (those cases in which pensioner income is already above the country's average) increase inequality.

Johansson (2016) explores in more detail the empirical evidence on the impact of different components of public spending on inequality. In general the distributional effects of tax and transfers are significant across OECD countries, leading to a reduction of up to 40 per cent of the market income Gini coefficient, mostly among European countries. Almost 75 percent of this reduction is accounted for by transfers in most countries (with the exception of Israel and the United States), whereas the progressivity of the tax system explains the remaining 25 percent.

The literature also shows the positive distributional effect of education, public health, social spending and labor market policies. Nevertheless, this positive effect does not always respond to the amount of public funds invested but to the quality of the programs involved. In particular, public spending in education and generalized access to schooling is unequivocally equity enhancing, as both of them facilitate people from poorer backgrounds to move up in the social and economic ladder. However, the impact on the quality of student achievements (PISA scores, for example) seems to be explained by the quality of teaching and the use of other resources at the school rather than by the size of public spending on education.

In this vein, Afonso, Schuknecht and Tanzi (2010) confirm that social spending and the performance of the education system are critical to redistribute income, but these factors do not work separately. Large social spending helps to reduce inequality only in those countries where the education system is efficient. The authors find that the social spending framework displays a great deal of variance in terms of efficiency in OECD countries, in particular comparing Southern European with Nordic countries. This variance suggests that better (or similar) results can be achieved with the same (or lower) level of public spending. In addition, the degree of efficiency in social spending is significantly correlated with the level of education achievement and the quality of the institutional framework (civil servants, the judiciary, etc.).

Unemployment benefits and other social protection programs designed to reduce inequality may end up having unintended consequences, unless they are accompanied by efficient active labor market policies and other mechanisms to incentivize job search and labor market participation by their recipients. Otherwise, although they may reduce inequality temporarily (for instance in the aftermath of recessions and severe crises), they may end up making poverty and unemployment more persistent.

But the impact of the size of the public sector on inequality does not only depend on the productive/unproductive nature of public spending. The recent increase in inequality in some countries is mainly due to the increase in the wage gap between high and low-performance firms, even after controlling for workers' skills. In these circumstances, technology adoption by the former may worsen the Gini coefficient, whereas public incentives and programs aimed at making technological progress available for small and medium firms might not only promote GDP growth, but also improve the distribution of income.

Agenello and Souisa (2014) look at the effect of fiscal consolidation processes, their structure and timing on income inequality in a panel of 18 industrialized countries from 1978 to 2009. These events are informative about the connection between fiscal variables and inequality, although the channels through which this connection works may not be the same as those operating between the size and composition of the public sector and the distribution of income. In general, public spending-driven consolidations, even those successful in reducing the debt to GDP ratio, are detrimental to the distribution of income, particularly in the aftermath of banking crises. Low-income earners are particularly affected by consolidations in terms of low growth and high inflation. High inflation generally worsens the distribution of income and accentuates the negative effect of consolidations in public spending. The negative effects of fiscal retrenchments after a crisis are increasing on the size of public spending.

Doumbia and Kinda (2019) analyze the effect on income inequality of reallocating different components of total public spending, in a panel of 82 advanced and developing countries. The authors control for an extensive list of variables that have been found to affect income inequality, and control for the level of public spending as a share of GDP. The results can be interpreted as the impact on inequality of an increase of a given public spending component (i.e., health or social spending) financed by the reduction in other spending components. As suggested by previous results in the literature, controlling for total public expenditure, an increase in social spending (e. g., family support, poverty programs, or unemployment benefits) and infrastructure reduce inequality, and more so if financed by a reduction in defense or education spending. Interestingly, the reduction in defense spending is only positive in reducing inequality in countries with a high level of internal security and stability. Otherwise, less defense spending is indeed detrimental to income equity, implying that social unrest and political instability undermine mostly the income of poor people. These authors find that increases in spending on education and health produce mixed results. An increase in these components, even financed by (cutting) defense spending, does not help to reduce inequality in the short term. However, when looking at a 10-year horizon, the effect of these expenses is inequality reducing (although the effect of health spending is barely significant). This implies that the positive impact of these components of public spending accrues first to the middle (mostly urban) class and spreads to the rest of the population as time goes by.

In a meta-regression analysis applied to more than 900 regression results in 84 different studies, including low, middle and high income countries, Anderson et al. (2017) also confirm a moderate negative effect of public spending on inequality. This effect is stronger for social spending, which helps to reduce the gap between high-income and middle-income households, but less so between the middle income and the poorest 20 percent of the population.

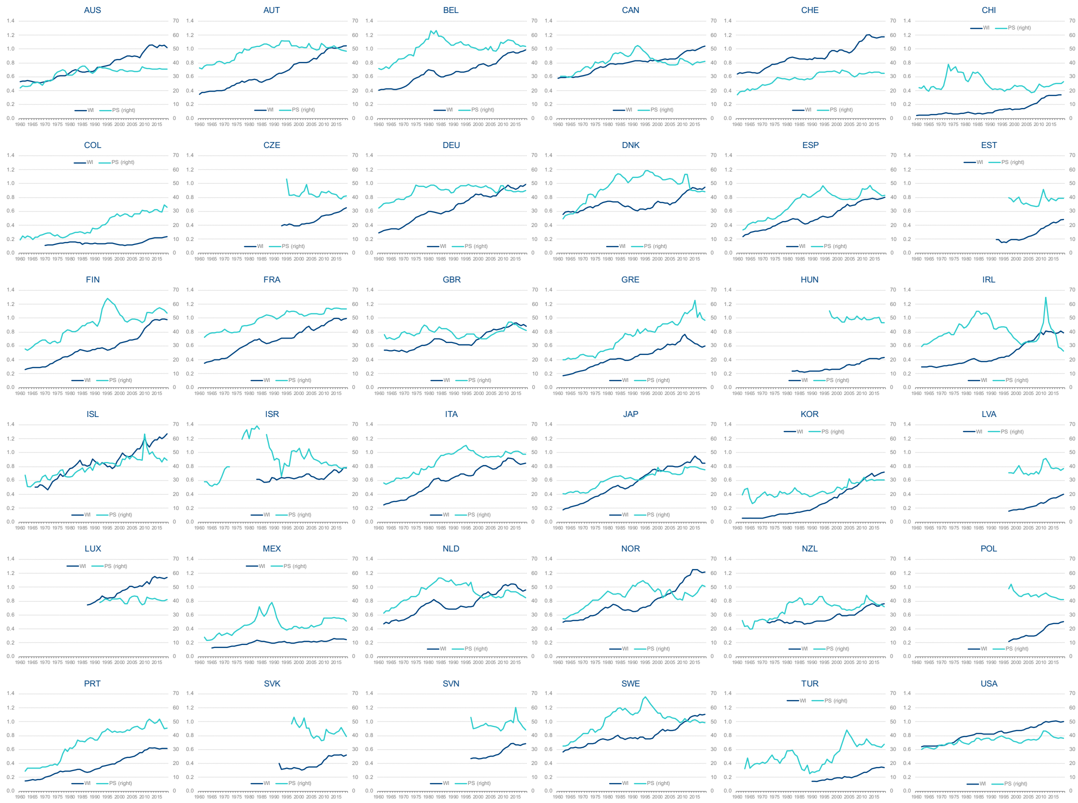


Figure 2: *relative welfare to the* $US_{2019} = 1.0$ *and public spending over GDP in percentage (right scale), OECD, 1960-2019. Source: see main text.*

4. Main results

4.1. Data

Extending previous findings in the empirical literature, we systematically compare the effects of government size on welfare growth with respect to GDP per capita growth. Our results are based on data of the welfare index (*WI*), described in section 2, *GDP* per capita and total public spending (*PS*) in relation to GDP. The last two variables have been taken from Penn World Table 10.0 and the OECD Economic Outlook for a sample of 36 countries. *PS/GDP* has been extrapolated backward, until 1960, whenever possible, using the database of Mauro et al. (2015). The lower case variables $gdp_{i,t}$, $wi_{i,t}$, and $ps_{i,t}$ represent logarithms, *i=1;...,N*, where *N* is the number of countries and *t=1,...,T* the time period from 1960 to 2019, although the sample is not complete for all countries. The original data for *WI* and *PS* are shown in Figure 2.

A preliminary analysis of Figure 2 shows some interesting results. First, there is a great deal of heterogeneity in the sample of OECD countries. Total public spending over *GDP* ranges from 10 per cent in Colombia during the 1960s to close to 70 per cent of Israel around 1980. Second, on average, total public expenditures over *GDP* increased in the 1960s and 1970s, but has remained relative constant since then and with a countercyclical pattern[2]. Nevertheless, there is again heterogeneity across countries with respect this general trend. For example, in Sweden there was an upward trend until the beginning of the 1990s, approaching 70 per cent, and since then *PS* has converged gradually to 50 per cent. Third, in general, relative welfare to the US in 2019 shows a positive trend. Again, there are some exceptions. In Colombia relative welfare was almost constant from 1970 to 2005, whereas in the US increased 24 percentage points. Something similar happened in Mexico from 1981 to 2010, Israel from 1981 to 2006 or Denmark from 1977 to 2005. Fourth, there are countries and decades in which the correlation between WI and PS is positive, for example in most European countries in the 1960s and 1970s. However, it is easy to find examples of a negative correlation, as in Canada, Denmark or the Netherlands from 1985 to the present, Poland from 1995 onward, or Sweden since 1993.

The correlation analysis of the cross-section variables (Table 1) also offers some interesting preliminary results. $\Delta gdp_{i,\bar{t}}$, which represents the average growth rate of income per capita for each country for the available period is positively and strongly correlated with $\Delta c_{i,\bar{t}}$ and $\Delta wi_{i,\bar{t}}$, which represent the average growth rate of consumption per capita and the welfare index. It also shows an appreciable negative correlation with the initial values of gdp, c and wi. However, the correlation with $ps_{i,\bar{t}}$, which represents the average size of the public sector, is also negative but very small. The results are very similar for $c_{i,\bar{t}}$ and $wi_{i,\bar{t}}$, although in the latter case the correlation increases in absolute terms.

4.2. Panel data analysis with annual data

We begin our econometric analysis performing a panel data estimation using annual data and fixed effects, according to the following equation:

$$\Delta y_{it} = \beta ps_{it} + \rho y_{it-1} + \delta_t + \alpha_i + \epsilon_{it} \quad (3)$$

where $y_{i,t} = ln(Y_{i,t})$, Y_{it} is *GDP* per capita or *WI* (the welfare index), $ps_{i,t}$ is the log of the

[2] It should be noted that, by construction, the variable *PS* is countercyclical as long as public expenditures are at least procyclical but with less volatility than *GDP*. As noted before, given the disconnection between *GDP* and domestic economic activity in Ireland from 2015, the fall of *PS* since then is artificially affected by the large inflows of intellectual property assets

Table 1: *Correlations between cross country variables*

	$\Delta gdp_{i,\bar{t}}$	$\Delta c_{i,\bar{t}}$	$\Delta wi_{i,\bar{t}}$	$gdp_{i,0}$	$c_{i,0}$	$wi_{i,0}$	$ps_{i,\bar{t}}$
$\Delta gdp_{i,\bar{t}}$	1.000	0.937	0.845	-0.537	-0.459	-0.501	-0.099
$\Delta c_{i,\bar{t}}$	-	1.000	0.913	-0.606	-0.553	-0.575	-0.144
$\Delta wi_{i,\bar{t}}$	-	-	1.000	-0.601	-0.567	-0.736	-0.223
$gdp_{i,0}$	-	-	-	1.000	0.971	0.871	0.566
$c_{i,0}$	-	-	-	-	1.000	0.868	0.611
$wi_{i,0}$	-	-	-	-	-	1.000	0.649
$ps_{i,\bar{t}}$	-	-	-	-	-	-	1.000

ratio of public expenditure to GDP of the country i in year t, δ_t is the time fixed effect for year t, which allows correction of common economic cycles between countries, and α_i is the fixed effect for country i that controls differences in GDP per capita and welfare steady states.

Two types of panels have been estimated. The first includes all 36 countries considered in our study, although they have different sample sizes. It is, therefore, an unbalanced panel and we denote it as the *full panel*. The second includes only those countries that have data for the full period 1960-2019, in general, more advanced economies. In this case, the panel is balanced. We denote it as the *restricted panel* and the number of countries is reduced to 22. The choice of a fixed effects specification instead of random effects is supported by the Hausman test in all estimates. In any case, the results obtained with the random effects model are very similar. In addition, the standard errors of the fixed effects model have also been robustly estimated, adjusting the heteroskedasticity for N clusters.

The first column in Table 2 shows the result of the panel regression of equation 3 for GDP per capita, with the unbalanced (top) and balanced (bottom) samples, imposing the same β for all countries and omitting country fixed effects. The effect of government size is negative and statistically significant in the balanced panel, whereas the rate of convergence (the coefficient of $gdp_{i,t-1}$) is very similar. In column (2) we confirm the negative effect of PS, now using the growth of welfare as dependent variable, although it is smaller than in column (1). In both cases we observe that the negative effects of the government size are greater and the convergence rates smaller in the case of the balanced sample. A possible explanation is that, in general, the countries excluded in the balanced sample are those with lower initial levels of GDP per capita and rapid convergence.

Table 2: *Results of panel estimates*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	Panel A: full sample															
	Original variables								Smoothed variables							
	Non Fixed-effects		Fixed-effects		Mean group		Pool mean group		Non Fixed-effects		Fixed-effects		Mean group		Pool mean group	
	$\Delta gdp_{i,t}$	$\Delta wi_{i,t}$	$\Delta gdp_{i,t}$	$\Delta wi_{i,t}$	$\Delta gdp_{i,t}$	$\Delta wi_{i,t}$	$\Delta gdp_{i,\bar{t}}$	$\Delta wi_{i,t}$	$\Delta\widetilde{gdp}_{i,\bar{t}}$	$\Delta\widetilde{wi}_{i,t}$	$\Delta\widetilde{gdp}_{i,\bar{t}}$	$\Delta\widetilde{wi}_{i,\bar{t}}$	$\Delta\widetilde{gdp}_{i,t}$	$\Delta\widetilde{wi}_{i,t}$	$\Delta\widetilde{gdp}_{i,\bar{t}}$	$\Delta\widetilde{wi}_{i,t}$
$ps_{i,t}$	−0.011 (−1.41)	−0.006 (−0.72)	−0.049 (3.80)	−0.033 (−2.91)	−0.096 (−5.65)	−0.052 (−4.43)	−0.028 (−2.50)	−0.036 (−3.33)	−0.005 (−1.22)	−0.003 (−0.37)	−0.034 (−2.51)	−0.026 (−1.99)	−0.087 (−4.13)	−0.042 (−3.46)	−0.040 (−2.53)	−0.043 (−2.16)
$gdp_{i,t-1}$	−0.009 (−2.16)		−0.015 (−1.67)		−0.026 (−3.23)		−0.040 (−5.91)									
$wi_{i,t-1}$		−0.010 (−3.84)		−0.016 (−1.76)		−0.015 (−1.86)		−0.033 (−4.13)								
$\widetilde{gdp}_{i,t-1}$									−0.010 (−3.38)		−0.008 (−1.67)		−0.010 (−1.57)		−0.028 (−3.88)	
$\widetilde{wi}_{i,t-1}$										−0.010 (−4.05)		−0.005 (−0.72)		0.015 (3.08)		−0.020 (−1.99)
Error Correction term							−0.875 (−17.39)	−0.891 (−17.04)							−0.007 (−10.32)	−0.007 (−10.32)
R^2	0.33	0.29	0.42	0.32	0.52	0.39	0.59	0.56	0.40	0.37	0.70	0.58	0.82	0.75	0.99	0.99
Countries	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36	36
Observations	1775	1775	1775	1775	1775	1775	1596	1596	1775	1775	1775	1775	1775	1775	1596	1596
	Panel B: restricted sample															
	Original variables								Smoothed variables							
	Non Fixed-effects		Fixed-effects		Mean group		Pool mean group		Nonfixed effects		Fixed-effects		Mean group		Pool mean group	
	$\Delta gdp_{i,\bar{t}}$	$\Delta wi_{i,t}$	$\Delta gdp_{i,\bar{t}}$	$\Delta wi_{i,t}$	$\Delta gdp_{i,\bar{t}}$	$\Delta wi_{i,t}$	$\Delta gdp_{i,\bar{t}}$	$\Delta wi_{i,t}$	$\Delta\widetilde{gdp}_{i,\bar{t}}$	$\Delta\widetilde{wi}_{i,\bar{t}}$	$\Delta\widetilde{gdp}_{i,\bar{t}}$	$\Delta\widetilde{wi}_{i,\bar{t}}$	$\Delta\widetilde{gdp}_{i,\bar{t}}$	$\Delta\widetilde{wi}_{i,\bar{t}}$	$\Delta\widetilde{gdp}_{i,\bar{t}}$	$\Delta\widetilde{wi}_{i,\bar{t}}$
$ps_{i,t}$	−0.025 (−2.76)	−0.019 (−2.20)	−0.068 (−6.11)	−0.051 (−4.66)	−0.078 (−11.32)	−0.053 (−8.22)	−0.028 (−3.77)	−0.040 (−5.67)	−0.021 (−2.38)	−0.018 (−2.00)	−0.057 (−6.01)	−0.047 (−4.73)	−0.062 (−13.35)	−0.046 (−6.44)	−0.060 (−9.88)	−0.064 (−5.51)
$gdp_{i,t-1}$	−0.007 (−1.05)		−0.011 (−1.14)		−0.029 (−2.92)		−0.031 (−4.26)									
$wi_{i,t-1}$		−0.010 (−2.02)		−0.009 (−1.07)		−0.013 (−1.50)		−0.024 (−3.17)								
$\widetilde{gdp}_{i,t-1}$									−0.007 (−1.16)		−0.006 (−1.06)		−0.011 (−1.60)		−0.025 (−3.22)	
$\widetilde{wi}_{i,t-1}$										−0.010 (−2.03)		−0.001 (−0.21)		0.011 (2.28)		−0.012 (−1.33)
Error Correction term							−0.924 (−15.28)	−0.950 (−15.64)							−0.008 (−8.80)	−0.007 (−7.02)
R^2	0.42	0.37	0.52	0.42	0.55	0.44	0.60	0.55	0.58	0.55	0.79	0.69	0.85	0.80	0.99	0.99
Countries	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22
Observations	1298	1298	1298	1298	1298	1298	1188	1188	1298	1298	1298	1298	1298	1298	1188	1188

Cluster robust t-ratios in brackets; all estimates include time effects; $\Delta gdp_{i,t}$ and $\Delta wi_{i,t}$ denote the annual rate of growth of *GDP per capita* and *WI* growth; $ps_{i,t}$ denotes the public spending GDP ratio in logs; $\widetilde{gdp}$, $\widetilde{gdp}$ and $\widetilde{gdp}$ represent the log of GDP, WI and PS/GDP smoothed with the Hodrick-Prescott filter ($\lambda = 100$); The model with fixed-effects estimates equation 3 which includes time effects; in columns (1) and (2) cross-country effects are removed; mean-group models are estimated according to equation 4 and pooled mean group according to equation 7 .

In columns (3) and (4) we now include fixed effects to control for omitted variables that determine the steady state of the growth equation. As we observe, the effect of the government size is larger in absolute terms and statistically significant in both the balanced and unbalanced panels. As in columns (1) and (2), the effects of PS are smaller in absolute terms in the case of welfare than for GDP per capita.

As noted above, in columns (1) to (4) in Table 2 we have assumed that β is the same between countries. In columns (5) to (8) we relax this restrictions using two different approaches. In the first one (columns (5) and (6)), we estimate a specification similar to the Mean Group (MG) estimator (Pesaran and Smith, 1995). In particular, we allow now in equation (4) for country specific effects of government size on GDP per capita and welfare (β_i), that is

$$\Delta y_{it} = \beta_i ps_{it} + \rho y_{it-1} + \delta_t + \alpha_i + \epsilon_{it} \tag{4}$$

Then we define the Mean Group estimate of β as:

$$\widehat{\beta}_{MG} = N^{-1} \sum_{i=1}^{N} \widehat{\beta}_i \tag{5}$$

and its standard error as

$$se(\widehat{\beta}_{MG}) = \left\{ [N(N-1)]^{-1} \sum_{i=1}^{N} (\widehat{\beta}_i - \widehat{\beta}_{MG})^2 \right\}^{-1/2} \tag{6}$$

In the second approach (columns (7) and (8)), the Pooled Mean Group (PGM) (Pesaran et al., 1999), we start from an ARDL specification that estimates the long-run relationship, including the short-run dynamics. The PMG can be considered as an intermediate estimator, which involves both pooling and averaging approaches. In our specification we allow the intercepts (α_i) and the long-run coefficient of the government size (β_i) to differ freely across countries, but we constrain the error correction and the short-run dynamics coefficients and the time dummies to be the same between countries:

$$\Delta\Delta y_{i,t} = \theta(\Delta y_{i,t-1} - \alpha_i - \rho y_{i,t-2} - \beta_i ps_{i,t-1} - \delta_t) + \sum_{p=1}^{P} \lambda_p \Delta\Delta y_{i,t-p} + \sum_{q=1}^{Q} \gamma_q \Delta ps_{i,t-q} + \epsilon_{it} \tag{7}$$

Columns (5) to (8) in Table 2 summarize the results obtained with these two models. The main conclusion is that, regardless of the estimation method, we find a negative and significant effect of the size of the public sector on the growth of GDP per capita and of the welfare index, once country fixed effects are taken into account, although this negative effect is usually smaller in the latter case. Additionally, the negative sign and the strong significance of the error correction term confirm the hypothesis that we are working with stationary variables or that there is a long-run cointegration relationship between them.[3]

The negative sign of the government size coefficient may be a consequence of the business cycle. For example, as growth is negative in recessions, GDP falls, and government expenditure over GDP increases, even in the case of a neutral fiscal policy. As a result, *PS* is countercyclical. As a first attempt to correct for this potential bias, in columns (9) to (16) we replicate the previous estimations of columns (1) to (8) but now using the trend component obtained after smoothing the logs of GDP per capita, welfare and government expenditures over GDP with the Hodrick and Prescott filter. The dependent variables corresponds to the annual rate of growth of the trend components of GDP per capita and welfare. As expected, the coefficients of *PS* are smaller in absolute value (except for the Pool Mean Group estimator), but the effect is still negative and statistically significant in all specifications in which country-fixed effects are included.

In Figure 3 we have represented the estimated coefficients of government size for each country in the welfare regression of column (14) in Table 2, based on data filtered by business cycle. The vertical lines around the point estimate represent the 95 per cent confidence intervals. Although the average of coefficients is negative (-0.042), there is a lot of heterogeneity across countries. In five countries the coefficients are smaller than -0.10 and in seven are positive. The estimated coefficients of *PS* range from -0.317 for the Czech Republic to 0.098 for Slovenia. In most of the countries, the coefficients are estimated with a small standard error, so the confidence intervals are quite small.

In Figure 4 we present the scatter of the government size coefficients for each country in the welfare regression of column (14) in Table 2 versus those estimated for GDP per capita in column (13). The average value of the government coefficient for welfare is half the value (-0.042) of the average for GDP per capita (-0.087). Although the correlation is positive, it is quite low (0.28). In the case of GDP per capita, the coefficient of *PS* is positive only for two countries (Colombia and Turkey). At the other extreme, the negative coefficient for Luxembourg (-0.72) is very large and atypical with respect to the estimated coefficients for other countries.

All the empirical work above can be summarized in two conclusions. First, when we replace *GDP* with a more inclusive measure such as the welfare index (*WI*), the nega-

[3] Unit root tests to check if ΔGDP and ΔWI they are integrated of order 1 would not be very reliable in this case due to their known problem of lack of power in finite samples and the more than likely presence of structural breaks.

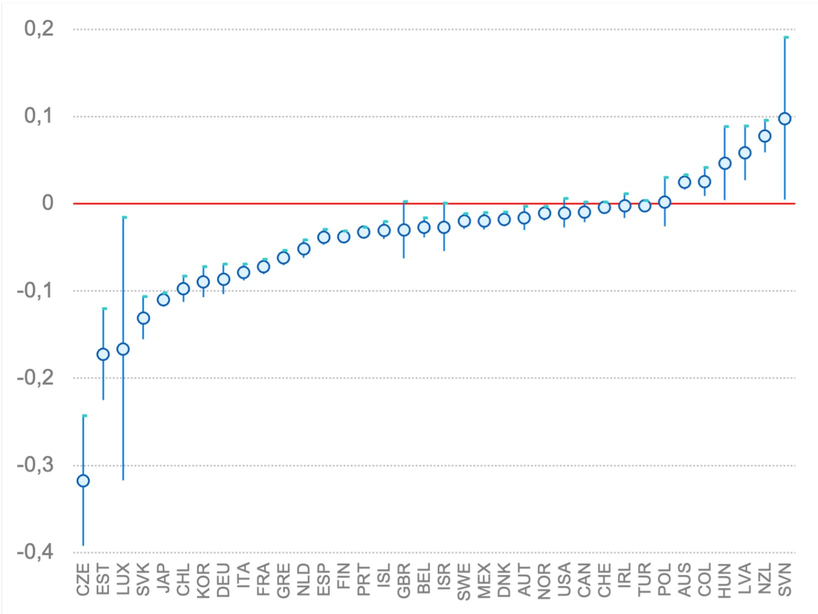


Figure 3: *Coefficients of government size for each country in welfare regression. Source: own elaboration based on column (14), Table 2.*

Figure 4: *Coefficients of government size for each country in welfare and GDP per capita regressions. Source: own elaboration based on columns (13) and (14) of Table 2.*

tive effect is reduced in most specifications. Second, there is strong cross-country heterogeneity in the results, and taking into account unobserved differences across countries is very relevant.

5. Robustness analysis and the relevance of government quality

5.1. Filtering the business cycle effect

As mentioned above, the business cycle may affect the relationship between the size of public spending and economic growth and the welfare index due to the action of economic stabilizers. Because of this potential problem, in the previous section we have compared the results obtained using the raw annual data with those of the variables filtered with the Hodrick-Prescott filter, in both cases controlling for annual dummies that estimate a common economic cycle. In this section, the effect of the business cycle is filtered out in two alternative ways. First, by establishing a business cycle chronology and inputting it into the model. Second, by taking averages every 10 years.

5.1..1 Business cycle chronology

In the spirit of Burns and Mitchell (1946), we consider phases of expansion and recession. According to this framework, *GDP* generally grows until it reaches its (local) maximum and a contraction phase begins as *GDP* starts to decline. After some time of negative growth, *GDP* reaches its (local) minimum, and an expansion begins as *GDP* growth rates become positive. We estimate the specific business cycle turning point chronologies by applying a yearly adaptation of the nonparametric dating procedure that was introduced by Bry and Boschan (1971) at the monthly frequency.[4] This algorithm consists of a set of filters and rules, based on moving averages of the data with different windows, that isolates the local minima and maxima in the log levels of the national series of seasonally adjusted *GDP*, subject to constraints on both the length and amplitude of expansion and contraction periods. Figure 5 shows the result of the business cycle dating for different countries.

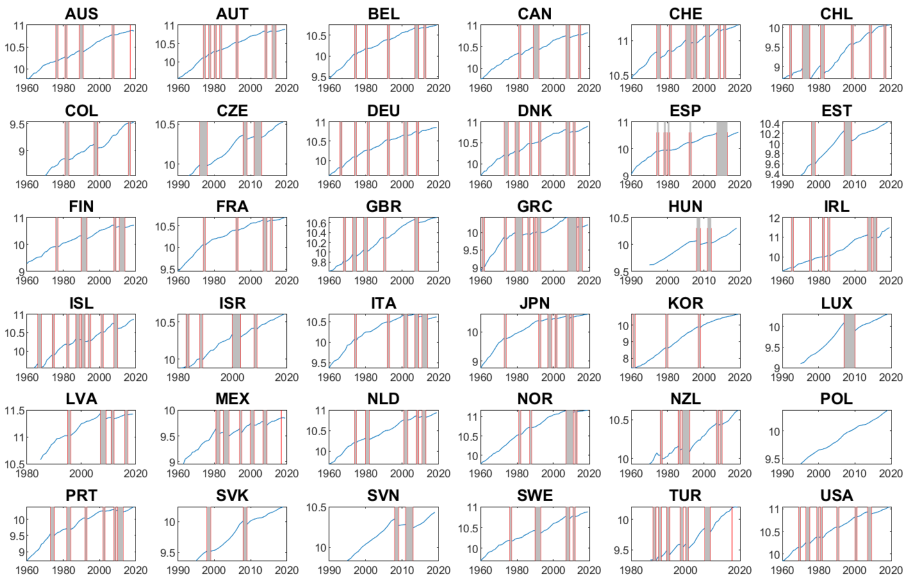
The panel data model is now re-estimated with an additional variable *bc* that takes the value 1 in times of recession and 0 in times of expansion.

$$\Delta y_{i,t} = \beta ps_{i,t} + \rho y_{i,t-1} + \alpha_i + \theta bc_{i,t} + \delta_t + \epsilon_{i,t} \quad (8)$$

In all the specifications in Table 3 the business cycle is, as expected, clearly signifi-

[4] See Harding and Pagan (2002) for an extension to a quarterly frequency.

Figure 5: *Business cycle index*



cant. Nevertheless, the effects of ps are statistically significant and only slightly smaller, in absolute value, than in fixed-effect models in Table 2, both in the full and balanced samples. Therefore, after including bc in the full sample, β is now -0.036 for GDP per capita growth (compared to -0.049 in column (3), Table 2), and -0.024 for welfare growth (-0.033 in column (4), Table 2).

5.1..2 Averaging the sample every 10 years

An alternative way to eliminate the influence of the economic cycle is to average the sample every 10 years. In Table 4 we present the results of the reestimation of columns (1) to (8) in Table 2, using rolling 10-year averages. In particular, in the first fourth columns we estimate the following equation

$$\frac{y_{it} - y_{it-10}}{10} = \beta \overline{ps}_{it}^{10} + \rho y_{it-10} + \delta_t + \alpha_i + \epsilon_{it} \qquad (9)$$

where the dependent variable is the average rate of growth for a 10-year interval, and $\beta \overline{ps}_{it}^{10}$ in the average level of public spending for each 10-year interval.

Table 3: *Panel estimation with a business cycle dummy*

	Full sample		Balanced sample	
	$\Delta gdp_{i,t}$	$\Delta wi_{i,t}$		
$ps_{i,t}$	−0.036 (−3.335)	−0.024 (−2.505)	−0.055 (−6.372)	−0.043 (−5.060)
$gdp_{i,0}$	−0.016 (−2.204)		−0.013 (−1.579)	
$wi_{i,0}$		−0.015 (−2.137)		−0.010 (−1.347)
$bc_{i,t}$	−0.042 (−12.641)	−0.028 (−7.770)	−0.037 (−12.954)	−0.023 (−6.137)
$\overline{R}^2$				
Countries	36	36	22	22
Observations	1775	1775	1298	1298

Columns (1) to (2) do not include fixed effects. The government size coefficients in the equations for GDP per capita and welfare are close to zero or even positive in the unbalanced sample. However, they are not statistically significant. As we control for fixed effects, in columns (3) to (4), the coefficients of *ps* become negative and statistically significant. These results suggests that the omission of country specific effects may bias the estimated coefficients of government size towards zero, in contrast to the negative coefficients estimated in Table 2.

In columns (5) to (6) in Table 4 we re-estimate the Mean Group (MG) estimator, using rolling 10-year averages, that is

$$\frac{y_{it} - y_{it-10}}{10} = \beta_i \overline{ps}_{it}^{10} + \rho y_{it-10} + \delta_t + \alpha_i + \epsilon_{it} \tag{10}$$

The estimated values of β are again negative and statistically significant. Again, the coefficient of government size is smaller in absolute terms for the welfare equation (column (6)) than for GDP capita (column (5)). The last two columns of Table 4 show the results of the Pool Mean Group and, as in columns (3) to (6), the coefficients of *ps* are negative in both specifications. Nevertheless, we observe that, in general, this coefficient in smaller in absolute value for welfare than for GDP per capita.

Following previous contributions to the empirical growth literature, such as Acemoglu and Molina (2022) on the effects of democracy upon economic growth, we have reestimated equation (9) also allowing the effect of government size to vary over time, that is:

Table 4: *Results of panel estimates with 10-year averages*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Full sample							
	Non-fixed effects		Fixed Effects		Mean Group		Pool Mean Group	
	Δgdp	Δwi	Δgdp	Δwi	Δgdp	Δwi	Δgdp	Δwi
$\overline{ps}_{it}^{10}$	-0.001	0.001	-0.033	-0.024	-0.085	-0.047	-0.010	-0.111
	(-0.09)	(0.16)	(-2.48)	(-1.65)	(-5.43)	(-2.77)	(-0.18)	(-3.16)
$gdp_{i,t-11}$	-0.013		-0.026		-0.04		-0.01	
	(-3.57)		(-3.18)		(-5.57)		(-1.67)	
$wi_{i,t-11}$		-0.013		-0.026		-0.039		0.02
		(-4.69)		(-2.57)		(-3.95)		(1.98)
ECM coefficient							-0.152	-0.145
							(-10.62)	(-10.06)
$\overline{R}^2$	0.43	0.43	0.74	0.68	0.84	0.76	0.52	0.41
Countries	36	36	36	36	36	36	36	36
Observations	1451	1451	1451	1451	1451	1451	1272	1272
	Balanced sample							
	Non-fixed effects		Fixed Effects		Mean Group		Pool Mean Group	
	Δgdp	Δwi	Δgdp	Δwi	Δgdp	Δwi	Δgdp	Δwi
$\overline{ps}_{it}^{10}$	-0.015	-0.014	-0.051	-0.043	-0.054	-0.042	-0.028	-0.064
	(-2.64)	(-1.18)	(-5.10)	(-3.31)	(-7.50)	(-6.24)	(-4.24)	(-5.28)
$gdp_{i,t-11}$	-0.013		-0.018		-0.037		-0.006	
	(-2.34)		(-2.70)		(-4.99)		(-1.07)	
$wi_{i,t-11}$		-0.012		-0.015		-0.025		0.022
		(-2.33)		(-2.05)		(-2.66)		(2.63)
ECM coefficient							-0.155	-0.169
							(-9.15)	(-9.98)
$\overline{R}^2$	0.59	0.58	0.81	0.72	0.87	0.79	0.52	0.46
Countries	22	22	22	22	22	22	22	22
Observations	1100	1100	1100	1100	1100	1100	990	990

$$\frac{y_{it} - y_{it-10}}{10} = \beta_t \overline{ps}_{it}^{10} + \rho y_{it-10} + \delta_t + \alpha_i + \epsilon_{it} \tag{11}$$

In Figure 6 we represent the estimated values of β_t for the equations with GDP per capita and welfare. The estimated government size coefficients are negative, statistically significant, and fairly stable until the Great Recession. Since the beginning of the financial crisis to the sovereign debt crisis, the negative effect of $\overline{ps}_{it}^{10}$ almost doubled. However, in principle, we cannot reject the hypothesis of reverse causation. That is, if the financial crisis reduced potential growth of GDP and welfare but countries maintained the rate of growth of public spending during the next decade, then $\overline{ps}_{it}^{10}$ increased, reinforcing the effect estimated from government size to growth and welfare estimated in previous

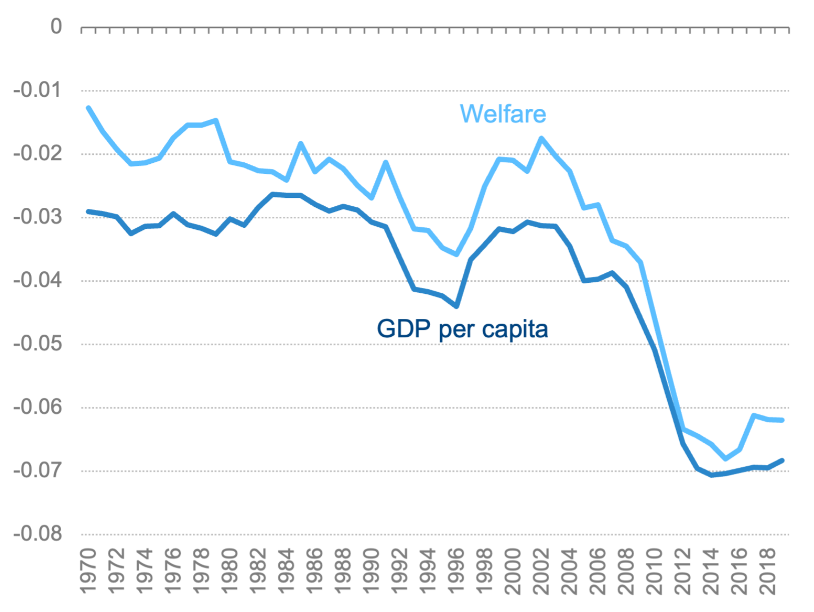


Figure 6: *Coefficients of government size over time in welfare and GDP per capita regressions. Source: own elaboration.*

decades. There is some evidence that this could have been the case in some countries in the sample.

Figure 6 also shows that the negative effect of government size on welfare is smaller in absolute value than for GDP per capita, during most of the sample period, although the difference almost disappeared in the final part of the sample.

5.2. Cross-country estimates

In the previous subsection we have taken 10-year averages to reduce the potential bias of business expansions and recessions on the effects of government spending on GDP per capita and welfare growth. The traditional approach of growth equations has used averages for the whole sample period and has estimated cross-country specification. In particular, in columns (1) and (2) of Table 5 we present the estimation results of the following equation:

$$\Delta y_{i,\bar{t}} = \theta y_{i,0} + \beta ps_{i,\bar{t}} + \alpha + \epsilon_i \tag{12}$$

where $y = gdp, wi$, therefore, $\Delta ln(Y_{i,t}) = \Delta y_{i,t}$ represents the average growth rate of GDP per capita or WI over the whole sample period; $y_{i,0}$ and $wi_{i,0}$ the initial values of the GDP per capita and welfare log for the first year available for each country, and $ps_{i,\bar{t}}$ the

Table 5: *Results of the cross-section estimation*

	(1)	(2)	(3)	(4)
	$\Delta gdp_{i,\bar{t}}$	$\Delta wi_{i,\bar{t}}$	$\Delta gdp_{i,\bar{t}}$	$\Delta wi_{i,\bar{t}}$
Constant	0.094	0.004	0.258	0.099
	(3.34)	(0.20)	(11.7)	(6.69)
$ps_{i,\bar{t}}$	0.012	0.019	-0.043	-0.011
	(2.20)	(3.10)	(5.99)	(2.30)
$gdp_{i,0}$	-0.012		-0.008	
	(4.67)		(5.89)	
$wi_{i,0}$		-0.015		-0.012
		(6.31)		(9.21)
Fixed effects			0.800	0.509
			(8.07)	(6.39)
Countries	36	36	36	36
R^2 *adj.*	0.350	0.660	0.816	0.790

average of the government size log, again for the entire sample period.

Contrary to the robust negative effects of *ps* estimated in all previous sections, the results in columns (1) and (2) of Table 5 show now a surprising positive and significant effect for both *gdp* and welfare. However, in these specifications, we are controlling only for the initial levels of GDP per capita and welfare. We have seen that in the panel estimates the inclusion of country fixed effects contributed to uncover the negative effects of government size and their significance in all cases. Therefore, it could be the case that the omission of these country-time invariant effects could bias the coefficient of *ps* toward positive values.

To test this hypothesis, we have recovered the estimated country fixed effects (α_i) in columns (3) and (4) of Table 3, and we have included them as an additional control in the following cross-section equation:

$$\Delta y_{i,\bar{t}} = \theta y_{i,0} + \beta ps_{i,\bar{t}} + \gamma\alpha_i + \alpha + \epsilon_i \tag{13}$$

The results in columns (3) and (4) in Table 5 show that the inclusion of α_i does not affect to much the $\beta-$convergence coefficient of the initial GDP per capita and welfare, but makes again the coefficient of government size negative and statistically significant. As in the panel estimates, the negative effect of *ps* on welfare was smaller (-0.011) in absolute values than that estimated for GDP per capita (-0.043).

6. The relevance of government quality

The heterogeneity across countries of the effects of government size estimated in previous sections leads us to think about the importance of the institutional framework, the efficiency of the public sector and the presence of potential non-linear effects when estimating the effects of *ps* on GDP per capita and welfare.

There are good reasons to think that the effects of government size on welfare and growth may be nonlinear. As pointed out by theoretical contributions such as Barro (1990), there could be an optimal level of government size that maximizes GDP per capita or welfare so that too small or too large government sizes would be less than optimal, or outright detrimental, in promoting GDP and welfare growth. In the former case, due to an insufficient provision of critical public services to the economy, and in the latter case, due to the waste of resources that could be more efficiently used by the private sector. However, this optimal level may be different for each country depending on the level of efficiency of the public sector. We expect that countries with a more efficient public sector will have larger governments.

To test to what extent the effects of government size are affected by the quality of the public sector, we have used the Government Quality Indicator from The Quality of Government Institute of the University of Goteborg. This quality indicator is the result of adding three components: the corruption of the political system, the impartiality of the judicial system and compliance with the law, and the quality of public administrations. Scandinavian countries are at the top in this ranking, followed by others from central Europe, New Zealand, Canada, and Australia. We have data for all countries in our sample with the exception of Colombia. For most of them the initial year is 1984, although for some countries the available data begin in 1993 or even later.

As expected, the correlation between the trend component of the government size log (*ps*), obtained with the Hodrick and Prescott filter, and the log of the Government Quality Indicator (*gq*) is positive and statistically significant. The coefficient of *gq* in a regression equation for *ps* is equal to 0.508 with a $t-$ratio of 16.7. This potential endogeneity problem of government size to government quality causes multicollinearity. Additionally, *gq* is rather stable in some countries, making it difficult to differentiate its effects from the country dummies included in our specifications. Both problems could make it difficult to estimate significant coefficients for *gq* and its interaction with *ps*. To circumvent these problems, we follow an alternative two-stage approach. First, we regress *ps* on *gq*. Second, we include the square of the residuals of this regression $(ps_{i,t} - \widehat{ps}_{i,t})^2$ in our panel equation, as well as the level of government quality ($gq_{i,t}$) and its interaction

Table 6: *Panel results with the government efficiency variable*

	(1)	(2)	(3)	(4)	(5)	(6)
	Δgdp	Δwi	Δgdp	Δwi	Δgdp	Δwi
$ps_{i,t}$	-0.022	-0.032	-0.030	-0.039	-0.039	-0.046
	(-4.26)	(-7.31)	(-7.46)	(-9.64)	(-8.07)	(-10.2)
$(ps_{i,t} - \widehat{ps}_{i,t})^2$			-0.030	-0.024	-0.033	-0.026
			(-4.04)	(-2.93)	(-4.64)	(-3.38)
$gq_{i,t}$					0.072	0.061
					(1.92)	(1.94)
$ps_{i,t} * gq_{i,t}$					-0.020	-0.017
					(-1.95)	(-2.02)
$gdp_{i,t-1}$	-0.019		-0.020		-0.022	
	(-6.66)		(-6.91)		(-6.81)	
$wi_{i,t-1}$		-0.003		-0.003		-0.004
		(-1.16)		(-0.96)		(-1.32)
R^2	0.742	0.719	0.751	0.725	0.752	0.726
Countries	36	36	36	36	36	36
Observationss	1157	1157	1157	1157	1157	1157

with government size ($ps_{i,t} * gq_{i,t}$), that is,

$$\Delta y_{it} = \beta_1 ps_{it} + \beta_2 (ps_{i,t} - \widehat{ps}_{i,t})^2 + \beta_3 gq_{i,t} + \beta_4 ps_{i,t} * gq_{i,t} + \rho y_{it-1} + \delta_t + \alpha_i + \epsilon_{it} \quad (14)$$

Table 6 depicts the estimated parameters of equation (14) for the full sample and the trend components of GDP per capita, welfare and government size. As government quality evolves smoothly and does not depend on the business cycle, we do not filter this variable. For comparability, in columns (1) and (2) we estimate the same equations as in columns (11) and (12) of Table 2, but for the sample with data for government quality. Again, the coefficients of *ps* are statistically significant. When we add the squared residuals $(ps_{i,t} - \widehat{ps}_{i,t})^2$, the statistical significance of the size of the government increases. The squared residuals enter with the expected negative sing. The interpretation of $\beta_2 < 0$ is that the deviation of government size with respect to the level predicted by the quality of government reduces the growth of GDP per capita and welfare.

Finally, in columns (5) and (6) we add the quality of government and its interaction with government size. As expected, we estimate a positive coefficient for government quality. Note that the inclusion of this additional variable introduces a quadratic effect of *ps* for a given level of government quality. To further analyze the effects of government

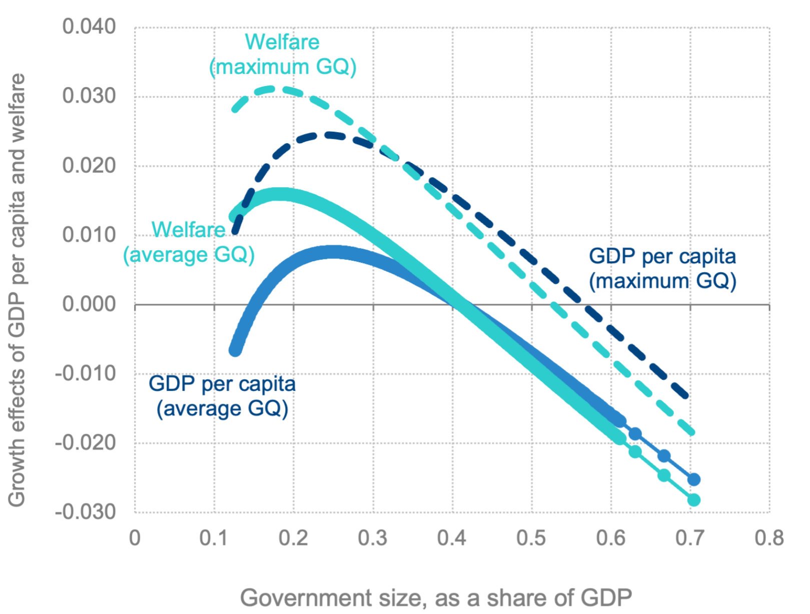


Figure 7: *Effects of government size on GDP per capita and welfare growth rates for two different values of government quality.*

size, we have simulated its effects on the growth rates of GDP per capita and welfare, holding constant government quality at two different values: its average and the maximum level of quality.[5]

Figure 7 represents the results of this simulation. As we observe, when government quality is at its average value, the effects of government size are positive on growth rates for values of *ps* below 41.1 per cent of GDP. According to our estimates, the maximum positive effects occur when the government size is below 30 per cent of GDP. When government quality is at its maximum level, the curves shift upward and to the right. The positive effects on growth are greater since β_3 is positive; also, as the quality of government improves, the range of government sizes that have positive effects on growth and welfare expands.

7. Conclusions

In this paper, we have systematically compared the effects of government size on social welfare growth with respect to GDP per capita growth, using different econometric techniques and robustness analysis for the sample of 36 OECD countries in the last six

[5] In this exercise the government size has been expressed in deviation with respect to its sample mean.

decades. For this purpose, we rely on the welfare measure proposed by Jones and Klenow (2016).

Panel data estimates using annual data and fixed effects show negative and statistically significant effects of government size on both GDP per capita and welfare. These effects are smaller in absolute terms in the case of welfare than in the case of GDP per capita. This result is robust to changes in the estimation method, to the use of smoothed variables, and to the inclusion of dummy variables that control for expansions and recessions over the business cycle.

When we allow the effect of government size to be time-varying, we find that the estimated coefficients are negative, statistically significant and rather stable until the Great Recession. Since the beginning of the financial crisis to the sovereign debt crisis, the negative effect of government size on growth has almost doubled. Additionally, we find that the effect on welfare is smaller in absolute value than for GDP per capita, during most of the sample period.

Finally, we have tested the presence of non-linear effects of government size on growth, controlling for government quality. We have found that the effects of government size follow an inverted U-shape. The effects on social welfare are positive and greater than for GDP per capita growth when the government size is below 40%, and improvements in government quality can substantially increase the government size before having negative effects on growth.

References

Afonso, A. y Furceri, D. (2010). Government size, composition, volatility and economic growth. *European Journal of Political Economy*, 26, 517-532.

Afonso, A., L. Schuknecht, and V. Tanzi (2010): "Income distribution determinants and public spending efficiency", *Journal of Economic Inequality*, 8, 367–389. https://rb.gy/fyznb6

Agnello, L. and R. Sousa (2014): "How does Fiscal Consolidation Impact on Income Inequality", *Review of Income and Wealth*. Series 60, Number 4, December. https://rb.gy/mnmevg

Aitken, A. (2019): "Measuring Welfare Beyond GDP", National Institute Economic Review No. 249 August. https://rb.gy/dpnvjd

Anderson, E., M. Jalles D'Orey, M. Duvendack, and L. Esposito (2017): "Does Government Spending affect Income Inequality? A Meta-Regression Analysis", *Journal of Economic Surveys*, 31(4), 961–987.

Awaworyi Churchill, S., Ugur, M., and Yew, S. L. (2017). Does Government Size Affect Per-capita Income Growth? A Hierarchical Meta-regression Analysis. *Economic Record*, 93(300), 142-171.

Bandrés, E. and Gadea, M. D. (2019). Investigating casual relations between public spending and economic growth in Europe. *Revista de Economía Mundial*, 51, 51-78.

Barro, R. J. (1990). Government spending in a simple model of endogenous growth. *Journal of Political Economy*, 98, S103-S125.

Barro, R. J. y Sala-i-Martin, X. (1992). Public finance in models of economic growth. *Review of Economic Studies*, 59, 645-661.

Bergh, A. and Henrekson, M. (2011): "Government size and growth: a survey and interpretation of the evidence." *Journal of Economic Surveys*, 25 (5), 872-897.

Bergh, A., and Henrekson, M. (2016): "Government size and growth: A rejoinder." *Journal of Economic Surveys*, 30(2), 393-396. http://bit.ly/3s1433x

Berik, G. (2020): "Measuring what matters and guiding policy: An evaluation of the Genuine Progress Indicator." *International Labour Review*, Vol. 159, No. 1.

Bhardwaj, R. (2018): "Explaining Income Inequalities: A Cross Sectional-Study of OECD Countries". Mimeo. http://bit.ly/3IhGzgh

Byrne, S.; T. Conefrey and M. O'Grady (2021): "The Disconnection of GDP from Economic Activity Carried out in Ireland," Quarterly Bulletin – Q4 2021, Central Bank of Ireland. http://bit.ly/3LE3vXA

Doumbia, D. and T. Kinda (2019): "Reallocating Public Spending to Reduce Income Inequality: Can It Work?", IMF Working Paper, WP/19/88. https://rb.gy/ei67eo

Fournier, J.M. y Johansson, A (2016). The effect of the size and the mix of public spending on growth and inequality. OECD Economics Department Working Papers, 1344. https://rb.gy/izj2ls

Gemmell, N., R. Kneller and I. Sanz (2016): "Does the composition of government expenditure matter for long-run GDP levels?" *Oxford Bulletin of Economics and Statistics*, 78(4), 522-547. http://bit.ly/3LFOSEy

Grier, K.B. and Tullock, G. (1989): "An empirical analysis of cross-national economic growth, 1951-80." *Journal of Monetary Economics*, 24, 259-276.

Hansson, P. and Henrekson, M. (1994): "A new framework for testing the effect of government spending on growth and productivity." *Public Choice*, 81, 381-404.

Johansson, A. (2016): "Public Finance, Economic Growth and Inequality: A Survey of the Evidence", OECD Economics Department Working Papers, 1346. https://rb.gy/mehawm

Jones, C.I. and P. Klenow (2016): "Beyond GDP? Welfare across Countries and Time." *American Economic Review*, vol. 106(9), 2426-2457. http://bit.ly/3H2O09k

Landau, D. L. (1983): "Government expenditure and economic growth: a cross-country study." *Southern Economic Journal*, 49, 783-792.

Pevcin, P. (2004): "Economic output and the optimal size of government." *Economic and Business Review*, 6(3),

213-227.
Saunders, P. (1985): "Public expenditure and economic performance in OECD countries." *Journal of Public Policy*, 5, 1-21.
Slemrod, J. (1995): "What do cross-country studies teach about government involvement, prosperity, and economic growth." *Brookings Papers on Economic Activity*, 2, 373-431.
Tanzi, V. and Zee, H. H. (1997). "Fiscal policy and long-run growth." *IMF Staff Papers*, 44 (2), 179-2019.

8. Appendix

8.1. Additional country-analysis: DOLS estimates

As an additional analysis of country heterogeneity, we have performed an individualized analysis country by country, using the Dynamic Ordinary Least Squares (DOLS) estimator, which allows for up to a maximum of four lags and four leading values of the independent variable, the public spending to GDP ratio. The number of lags and leads -between one and four- is selected using the Schwartz information criterion, so that a different number is allowed for each country. The estimated equation for $i = 1...N$ is as follows:

$$\Delta y_{i,t} = \alpha_i + \beta_i ps_{i,t} + \rho_i y_{i,t-1} + \sum_{j=-q}^{r} \delta_{ij} \Delta ps_{i,t+j} + \epsilon_{i,t} \qquad (15)$$

where $y_{i,t} = ln(Y_{i,t})$, Y_{it} is *GDP* per capita or *WI* (the welfare index), $PS_{i,t}$ is the public spending to GDP ratio of country i in year t, q is the number of delays and r is the number of leads. Compared with the Mean Group estimation in Table 2, DOLS estimation allows for country heterogeneity also in the convergence coefficient (ρ) and for additional dynamics of *PS*, but at the cost of omitting the common time effects (δ_t) estimated in equations (4) and (5).

The DOLS technique has the advantage of correcting for serial autocorrelation, considering the possible endogeneity of the relationship and thus the causality of Δy_{it} to ps_{it}, and it is an alternative to other methods of estimating the cointegrating relationship, which prevents the non-stationarity of the series (Phillips and Hansen, 1990 and Stock and Watson, 1993). In this way, we take into account the dynamic effect of government spending on growth and welfare, which may be extended over time.

What we are looking for is not the Keynesian-type impact of spending on the rate of change of *GDP* per capita or the Welfare index over a short period of time, but a more long-lasting type of effect, in line with models that try to explain the determinants of long-run growth. It is quite possible that the effects of an increase in spending, even if larger in the early periods of its implementation, will end up maintaining a dynamic of persistence in the medium and long term. The research by Gemmell, Kneller and Sanz (2011) and Bandrés and Gadea (2019) using impulse-response functions provides evidence in this sense.

The results obtained with the DOLS estimation (see Table 6 in the Appendix) show much more heterogeneity in the estimates of the coefficients (β_i) of the government size than the panel specifications in Table 2. In the case of the filtered data in the equation for

GDP per capita , the correlation of β_i between the DOLS and the Mean Group estimates of column (13) in Table 2 is negative (-0.35), with some outliers as Poland and Czech Republic. The average value of the DOLS estimates of β_i is -0.048, below the Mean Group estimate (-0.087). In the case of welfare, this correlation is positive but low (0.21), with the outliers of Latvia and Hungary. However, the average value of the DOLS estimates of β_i is positive (0.002), whereas in column (14) of Table 2 the Mean Group estimate is negative and equal to -0.042.

Table 7: *Estimation of DOLS regressions with business cycle*

countries/variables	constant	β	θ	constant	β	θ
		GDP growth			Welfare index	
AUS	10.0212	-0.1902	-3.6688	-1.6748	0.0856	1.0852
	(6.0720)	(-3.7487)	(-9.1016)	(-0.5923)	(0.9877)	(0.7004)
AUT	7.1528	-0.0920	-3.4493	5.6527	-0.0742	-0.2510
	(0.9538)	(-0.6398)	(-4.6650)	(1.2768)	(-0.8166)	(-0.2074)
BEL	11.0786	-0.1631	-3.1478	-1.0807	0.0478	-0.4417
	(6.1128)	(-4.7327)	(-3.8703)	(-0.1817)	(0.4293)	(-0.5886)
CAN	9.2511	-0.1437	-5.2815	1.8736	-0.0239	0.9522
	(4.0641)	(-2.6210)	(-7.0844)	(1.0038)	(-0.5536)	(1.0078)
CHL	34.5260	-1.3145	-5.0171	21.2820	-0.6934	-6.3656
	(5.9282)	(-5.2400)	(-5.3949)	(1.4242)	(-1.0497)	(-1.3566)
COL	3.1852	0.0221	-6.4721	-2.7878	0.1824	-3.1375
	(3.9091)	(0.5430)	(-6.5776)	(-0.7846)	(1.1434)	(-1.1658)
CZE	16.7007	-0.3087	-5.0789	-14.3839	0.3856	0.1677
	(1.9149)	(-1.5494)	(-4.5540)	(-1.5635)	(1.8363)	(0.1130)
DNK	2.1039	0.0047	-3.5551	5.1031	-0.0847	0.5121
	(1.0591)	(0.1129)	(-7.0659)	(1.7182)	(-1.3040)	(0.4897)
EST	46.4924	-1.1157	-4.8021	25.6591	-0.5681	0.7911
	(4.7719)	(-4.3302)	(-1.8594)	(2.7713)	(-2.1874)	(0.3884)
FIN	6.7332	-0.0713	-4.2917	5.6958	-0.0695	-0.3965
	(3.8024)	(-1.8892)	(-2.6715)	(2.1941)	(-1.2221)	(-0.3877)
FRA	15.9865	-0.2597	-3.8882	10.3942	-0.1711	0.9415
	(24.8009)	(-19.7724)	(-5.9693)	(3.3435)	(-2.8154)	(1.0369)
DEU	13.4032	-0.2389	-3.8749	0.2081	0.0329	-0.1373
	(4.5043)	(-3.7181)	(-7.0108)	(0.0327)	(0.2379)	(-0.1533)
GRC	9.3633	-0.1441	-5.6832	7.5256	-0.1384	-1.1005
	(6.2148)	(-3.8067)	(-5.1336)	(5.8698)	(-4.1763)	(-0.7920)
HUN	9.6169	-0.1309	-7.2043	24.5038	-0.4477	-0.1980
	(1.0833)	(-0.7062)	(-4.1803)	(1.2703)	(-1.1182)	(-0.1592)
ISL	8.6843	-0.1029	-6.3276	3.9744	-0.0311	-5.0380
	(4.1940)	(-2.0934)	(-7.7145)	(0.8736)	(-0.2885)	(-2.9454)
IRL	13.7161	-0.2033	-3.0387	4.8049	-0.0597	-2.2798
	(3.5697)	(-2.2159)	(-1.9667)	(1.4903)	(-0.7658)	(-1.5532)
ISR	3.9574	-0.0052	-1.2468	6.4603	-0.1198	1.6166
	(0.6455)	(-0.0370)	(-0.7412)	(0.7741)	(-0.6259)	(0.9713)
ITA	9.8494	-0.1644	-3.9035	9.5759	-0.1714	-0.6536
	(5.6508)	(-4.2633)	(-5.8024)	(3.4120)	(-2.8200)	(-0.5522)
JPN	19.9528	-0.4987	-3.2458	11.5903	-0.2881	-0.7337
	(10.2807)	(-8.3152)	(-3.7666)	(7.1461)	(-5.3172)	(-0.6763)
KOR	23.7337	-0.6817	-11.0410	11.9799	-0.2884	-6.8266
	(20.4721)	(-16.5904)	(-8.8177)	(3.1077)	(-1.7522)	(-3.3652)
LVA	50.0588	-1.1908	-5.9987	-7.1443	0.2207	-1.1520
	(3.5185)	(-3.2563)	(-3.7925)	(-0.5852)	(0.6593)	(-0.3621)
LUX	63.1779	-1.4453	-4.8473	17.5732	-0.3295	0.4088
	(2.8135)	(-2.6418)	(-4.5209)	(0.4054)	(-0.3125)	(0.1474)
MEX	6.8393	-0.1082	-6.7827	4.0139	-0.0910	-4.5505
	(3.7308)	(-1.4145)	(-4.9132)	(1.2209)	(-0.6775)	(-3.1633)
NLD	8.6302	-0.1173	-3.8379	8.0061	-0.1444	-0.5317
	(3.7002)	(-2.5515)	(-6.6733)	(1.9247)	(-1.6479)	(-0.4666)
NZL	11.8076	-0.2339	-5.0438	-3.9343	0.1345	-1.5759
	(3.9821)	(-3.0236)	(-7.2716)	(-0.9210)	(1.1725)	(-1.7721)
NOR	7.5479	-0.0954	-3.8895	2.4154	-0.0153	-0.6181
	(6.0379)	(-3.2369)	(-4.3022)	(1.0769)	(-0.3157)	(-0.2185)
POL	-	-	-	-	-	-
	-	-	-	-	-	-
PRT	8.3671	-0.1295	-3.7842	4.2334	-0.0240	-2.7185
	(6.1506)	(-3.6700)	(-4.8359)	(1.4409)	(-0.3714)	(-1.8338)
SVK	16.5756	-0.2854	-6.1224	23.8583	-0.5054	-0.6409
	(1.7022)	(-1.2678)	(-2.4545)	(1.9711)	(-1.7766)	(-0.1770)
SVN	11.9174	-0.1763	-6.3431	-6.6298	0.1920	-1.8381
	(1.4666)	(-1.0362)	(-2.1950)	(-0.8347)	(1.1017)	(-1.5898)
ESP	10.1103	-0.1741	-2.7264	4.6808	-0.0736	-0.1184
	(8.6014)	(-5.3634)	(-3.8622)	(2.4564)	(-1.2630)	(-0.0707)
SWE	6.3576	-0.0667	-4.0367	4.9371	-0.0729	0.6773
	(4.1025)	(-2.2083)	(-3.8658)	(2.7286)	(-2.0067)	(0.8821)
CHE	1.1874	0.0304	-3.4228	2.0758	-0.0330	0.0009
	(0.4690)	(0.3802)	(-3.8692)	(0.5960)	(-0.2869)	(0.0010)
TUR	4.2921	0.0559	-9.3043	4.0032	0.0146	-7.5096
	(3.6232)	(1.2975)	(-7.3561)	(1.2865)	(0.1390)	(-4.4183)
GBR	10.7368	-0.1990	-2.9961	6.3882	-0.1443	2.0754
	(4.8692)	(-3.5781)	(-5.4975)	(0.8079)	(-0.7351)	(1.4974)
USA	15.1317	-0.3178	-3.6478	1.8882	-0.0365	1.3994
	(7.6101)	(-5.8418)	(-7.8910)	(0.8160)	(-0.5864)	(1.9753)