Does government size affect per-capita income growth?
A Hierarchical meta-regression analysis

Sefa Awaworyi Churchill
Department of Economics,
Monash University, VIC 3800, Australia
sefa.awaworyi@monash.edu

Mehmet Ugur
Business School,
University of Greenwich, London, UK
m.ugur@gre.ac.uk

Siew Ling Yew
Department of Economics,
Monash University, VIC 3800, Australia
Siew.ling.yew@monash.edu

Corresponding author:

Sefa Awaworyi Churchill

Email: sefa.awaworyi@monash.edu

Tel: +61470641883

Received Date : 30-May-2016

Revised Date : 06-Sep-2016

Accepted Date : 29-Sep-2016

Article type : Survey Article

Does government size affect per-capita income growth? A Hierarchical meta-regression analysis

Abstract

Since the late 1970s, the received wisdom has been that government size (measured as the ratio of total government expenditure to GDP or government consumption to GDP) is detrimental to economic growth. We conduct a hierarchical meta-regression analysis of 799 effect-size estimates reported in 87 primary studies to verify if such assertion is supported by existing evidence. Our findings indicate that the conventional prior belief is supported by evidence mainly from developed countries but not from less developed countries (LDCs). We argue that the negative relationship between government size and economic growth in developed countries may reflect endogeneity bias.

JEL code: O40, H50, C1

Keywords: Economic growth; Government size; Government expenditure; Government consumption; Meta-analysis; Evidence synthesis

1. Introduction

One of the most contentious issues in economics is whether 'big government' is good or bad for economic growth. In recent decades, the received wisdom has been that big government is detrimental to growth. This consensus has been tested by the onset of the recent financial crisis in Europe and the United States (US). Governments have been called upon to act not only as lenders of last resort but also as demand-managing and bank-nationalising fiscal heavyweights, the spending capacity of which was considered as a crucial ingredient for recovery. Nevertheless, the financial crisis has also demonstrated the persistence of the received wisdom. A large number of European countries including fiscally-comfortable countries such as Germany and the UK had adopted austerity programs with the aim of spurring growth by creating more room for private investment.

The continued appeal of the received wisdom may be due to ambiguity in economic theory, which suggests that government size may have both positive and negative effects on growth. Government can play a growth-enhancing role by providing public goods, minimising externalities and maintaining confidence in a reliable medium of exchange. Government can also contribute to growth by enhancing human capital through investments in health and education and by building and maintaining a sound infrastructure (Kormendi and Meguire, 1985; Ram, 1986; De Witte and Moesen, 2010). From a Keynesian perspective, increased government spending increases aggregate demand that in turn increases output.

On the other hand, big government may affect growth adversely because of crowding-out effects on private investment (Landau, 1983; Engen and Skinner, 1992). Big government also implies high taxes, most of which are distortionary and hence growth-reducing (De Gregorio, 1992). Increased government size can also be a source of inefficiency due to rent seeking and political corruption that harm economic growth (Gould and Amaro-Reyes, 1983; Mauro, 1995; Hamilton, 2013).

There are also some reasons to expect an inverted-U relationship between government size and economic growth (e.g., Barro, 1990; Armey, 1995). According to Barro (1990), when the ratio of government spending to output is low, the positive effect of government spending on the marginal product of capital tends to dominate the negative effect of taxes on the private return to capital. Thus, a rise in the ratio of government spending to output tends to increase economic growth. However, the opposite tends to occur when the ratio of government spending to output is high. Similarly, Armey (1995) also posits an inverted-U relationship between government size and growth by invoking the law of diminishing factor returns. When government is sufficiently small, an increase in government size may be associated with higher growth rates as the government ensures rule of law and property right protection. However, when the government size is beyond the optimal level, a further increase in government size is associated with lower growth rates. Since government size in developed countries is typically larger than that in less developed countries, the ‘Armey curve’ suggests that the relationship between government size and economic growth tends to be negative in developed countries.

Beyond the theoretical literature, a large empirical literature has explored the relationship between government size and economic growth but the empirical evidence on the relationship is inconclusive. On the one hand, a large number of studies report a negative relationship between government size and growth (see, e.g., Grier and Tullock, 1989; Barro, 1991; Ghura, 1995; Lee, 1995; Fölster and Henrekson, 2001; among others). On the other hand, a sizeable number report a positive relationship (see, e.g., Aschauer, 1989; Munnell, 1990; Evans and Karras, 1994).

Heterogeneous findings are to be expected because government size is measured differently,[1] countries may be at different stages of development and the optimal government size may differ between countries, depending on prevailing political/institutional structures (Bergh and Karlsson, 2010). In addition, model specification as well as estimation methods differ between studies. Finally, earlier studies on determinants of growth report that the effect of government size on growth is either not robust to model specification (Levine and Renelt, 1992) or does not remain a significant determinant of growth in a series of Bayesian averaging trials (Sala-i-Martin et al., 2004).[2]

[1] Government consumption is measured as all government expenditure for goods and services but excluding military expenditure, and in some cases education expenditure (see, e.g., Barro, 1991; Easterly and Rebelo, 1993). According to the OECD, total government expenditure, on the other hand, captures the total amount of expenditure by government that needs to be financed via government revenues.

[2] Sala-i-Martin et al. (2004) report that government consumption has a negative effect on growth (-0.034). Surprisingly, however, the adverse effect of government consumption is less severe than that of public investment (-0.062).

Given this landscape, there is evident need to verify where the balance of the evidence lies. The synthesis should address not only the question of whether government size is growth-enhancing or growth-retarding, but also the extent of publication selection bias and sources of heterogeneity in the evidence base.

A number of past reviews have already tried to synthesize the existing findings (e.g., Poot, 2000; Nijkamp and Poot, 2004; and Bergh and Henrekson, 2011). Some studies focus on the growth-effects of specific types of government expenditure and taxation but others (e.g. Bergh and Henrekson, 2011) focus on government size in general. These reviews provide useful narrative syntheses that reflect the state of the art in the research field but they fall short of allowing for robust inference about the 'effect size', the extent of selection bias and the sources of heterogeneity in the evidence base. One reason is sample selection bias which arises when reviewers rely on a subset of the effect-size estimates reported in primary studies. Secondly, even when the full set of effect-size estimates is utilized (e.g., Nijkamp and Poot, 2004), they do not have systematic tests for publication selection bias, which occurs when primary study authors and/or journal editors tend to report findings that reject a null hypothesis. Last but not least, discussions on the sources of heterogeneity in the existing reviews are largely descriptive as it relies mainly on "vote counting" or narrative summary.[3]

Hence, we aim to contribute to existing knowledge along four dimensions. First, we provide a quantitative synthesis of the evidence on the growth effects of government size (measured by the ratio of total government expenditure to GDP and the ratio of government consumption to GDP) by taking into account information provided in 87 empirical studies that report 799 effect-size estimates. Secondly, we address the issue of selection bias that arises when primary study authors search for samples, estimation methods or model specifications that yield statistically significant estimates; or when narrative reviews rely on 'representative' or 'preferred' estimates rather than all available information. Third, we address the issue of data dependence that arises when primary studies that draw on a particular dataset report multiple estimates or when different studies utilize overlapping segments of the existing country datasets. Finally, we account for sources of heterogeneity in the evidence base, including estimation methods, data

[3] Vote counting is a method of synthesizing evidence from multiple studies, which involves the comparisons of studies reporting the direction of effect. For instance, number of positive effects or negative effects. This approach does not take into account study specific characteristics such as the size of samples used, quality of methods or effect size.

periods, data types (cross-section versus panel data), publication types and the level of development.

With regards to heterogeneity, two of our findings are key. The first suggests that the effect of government size on per-capita GDP growth is negative in developed countries but insignificant in LDCs. This is the case irrespective of whether government size is measured as the share of total expenditure or consumption expenditure in GDP. In the full sample that consists of evidence on both developed and LDCs, the effect is insignificant when government size is measured as the ratio of total expenditure to GDP but is negative and smaller in magnitude when government size is measured as the share of government consumption in GDP. We argue that the negative relationship between government size and economic growth is more likely to hold in developed countries as opposed to LDCs. We also argue that the negative relationship observed in developed countries should be interpreted with caution for three reasons: (i) increasing government share in GDP and declining economic growth had been concurrent trends in developed countries since the mid-1970s; (ii) the negative effect in developed countries may be reflecting other structural factors that drive both lower growth rates and larger government sizes in these countries; (iii) in contrast to linear specifications in the literature, the relationship between government size and growth may be non-linear. Secondly, we find evidence that primary-studies that control for endogeneity through instrumental variable (IV) methods (e.g., 2SLS, 3SLS, etc.) report systematically less adverse effects when government size is measured as the share consumption in GDP.

2. Theoretical and Empirical Considerations

Several perspectives exist on what explains growth. According to the seminal work by Solow (1956) and Swan (1956), an economy grows over time due to exogenous technological progress. Although neoclassical growth models (Solow, 1956; Swan, 1956) have a successful track record in estimating capital and labour elasticities (i.e., capital and labour shares in output) at the macro, industry and firm levels, these models have some shortcomings. The main shortcoming comes from the assumption that the source of long-run growth (i.e., technology) is exogenous and so the resulting total factor productivity can be captured only through the residuals of the estimated model. Another important issue is that the original model has to be augmented with a measure of government size even though the latter is not theorised to have any effect on long-run growth.

Endogenous growth models address the latter issue by endogenizing the factors that determine long-run growth (e.g., Romer, 1986; Lucas, 1988; Aschauer, 1989; Barro, 1990; Romer, 1990; Rebelo, 1991; Aghion and Howitt, 1992). Romer (1986) leads the way to allow investment in knowledge to affect growth, while Lucas (1988) provides the first human capital approach to endogenize growth. Other endogenous growth models such as Romer (1990) and Aghion and Howitt (1992) endogenize technological progress. In Aschauer (1989) and Barro (1990), public investment enhances long-run growth. More recent work has focused on other determinants of growth such as geography, institutions and culture (e.g., Acemoglu et al., 2002; Tabellini, 2010; Dell et al., 2012). Overall, in endogenous growth models, sustained long-run growth arises from endogenous sources and cross-country differences in per-capita income can persist indefinitely. In these models, government policy can alter the level of endogenous variables such as human capital or investment rates and may have theory-driven implications for the country's long-run growth.

Hence, one factor that is likely to cause heterogeneity in the empirical evidence is the theoretical model utilized in the empirical studies. Some studies (e.g., Bajo-Rubio, 2000; Bodman et al., 2012) augment the neoclassical Solow-type growth model with government size (G), which yields a generic model of the type below:

$$Y_{it} = A_{it}e^{\lambda t}K_{it}^{\alpha}L_{it}^{\beta}G_{it}^{\gamma}e^{u_{it}} \qquad (1)$$

In (1), Y is income (usually measured as gross domestic product – GDP), which is a function of technology (A), capital stock (K) and labour force (L). Subscripts i and t represent country and time, respectively. Augmenting (1) with government size (G) has been proposed by Feder (1983) and Ram (1986).

In the standard model, technology is exogenous and treated as an unobservable country-specific fixed factor. However, technology can be decomposed into a country-specific fixed component ($\tilde{A}_i e^{\lambda t}$) and a component that depends on a country's observed characteristics (C_{it}^{θ}). Hence, we let $A_{it}e^{\lambda t} = \tilde{A}_i e^{\lambda t} * C_{it}^{\theta}$ where $\tilde{A}$ captures technology shocks unobservable to the researchers and C is a proxy for observable technological change due to country characteristics.

Dividing the output and the inputs by labour (L) and taking natural logarithms, we obtain:

$$lny_{it} = \eta_i + \lambda t + \theta lnC_{it} + \alpha lnk_{it} + \delta lnL_{it} + \gamma lng_{it} + u_{it} \qquad \text{(1a – levels equation)}$$

Here y is output per employee (or per-capita GDP as a proxy), k is capital per employee (or per-capita capital as a proxy) and g is the measure of government size (measured as the ratio of total government expenditure or government consumption to GDP). The coefficients on capital per employee (α) is the elasticity of output with respect to capital; whereas the coefficient on labour (δ) is a measure of returns to scale, that is $\delta = \alpha + \beta + \gamma - 1$.[4] The log of unobservable technical change ($\ln(\tilde{A}_i e^{\lambda t})$) yields a country-specific effect (η_i) and a time effect (λt). An observable technological effect (θ) depends on determinants of technological change at the country level which may include geography, culture, political and economic institutions. The error term u_{it} is a white-noise disturbance term with zero mean and a constant variance.

The country-specific fixed effect (η_i) can be eliminated by first-differencing (1a).

$$\Delta y_{it} = \lambda + \theta \Delta C_{it} + \alpha \Delta k_{it} + \delta \Delta L_{it} + \gamma \Delta g_{it} + v_{it} \qquad \text{(1b – first-differenced equation)}$$

where Δ is log difference between periods t and t-1. In most studies, the ratio of government expenditure or consumption to GDP (g_{it}) is used instead of its growth rate (Δg_{it}) (see, Rubinson, 1977; Landau, 1983). Hence, the estimated growth model usually takes the form:

$$\Delta y_{it} = \lambda + \theta \Delta C_{it} + \alpha \Delta k_{it} + \delta \Delta L_{it} + \gamma g_{it} + \varepsilon_{it} \qquad \text{(1c – empirical model in most studies)}$$

where ε_{it} is an idiosyncratic error term.[5] The coefficient of interest is γ, which is: (i) either the elasticity of per-capita GDP with respect to government size if primary studies use the logarithm of g; or (ii) a semi-elasticity of per-capita GDP with respect to government size if primary studies use g as a ratio only.

A considerable number of studies also adopt a variant of the endogenous growth model. These studies follow Barro (1991) and Mankiw et al. (1992), where the determinants of growth include investment in physical capital (k) and human capital (h), augmented with other covariates such as government size (g) and other variables found to be related to growth in the empirical literature (e.g., initial level of per-capita GDP, openness, financial development, etc.). Endogenous growth models usually take the following form:

[4] In (1a), constant returns to scale are not imposed. Hence, the coefficient on labour lnL indicates increasing, decreasing or constant returns to scale – depending on whether δ is greater than, smaller than or equal to zero.
[5] Model (1c) has been used by a large number of primary studies included in this review, including Grossman (1990), Atesoglu and Mueller (1990) and Dar and AmirKhalkhali (2002), among others.

$$\Delta y_{it} = \alpha k_{it} + \beta h_{it} + \gamma g_{it} + \sum_j \theta_j Z_{j,it} + \varepsilon_{it} \quad \text{(2 – endogenous growth model)}$$

Here, Δy_{it} is the growth rate of per-capita GDP and g is the ratio of total government expenditure to GDP or government consumption to GDP; and Z is a vector of variables commonly used in the economic growth literature, including financial deepening or institutional quality, etc. This specification has been widely used in the empirical growth literature, including Barro and Sala-i-Martin (1995), Stroup and Heckelman (2001) and Bose et al. (2007).

There are also hybrid models - including simultaneous equation models where government size is modelled to have both direct and indirect effects on growth; and Keynesian models where government size affects growth from both the demand and supply side (see, e.g., Tanninen, 1999; Ghosh and Gregoriou, 2008).

To ensure comparability, we include studies and extract effect-size estimates for our meta-analysis if: (a) a study uses one of the three growth models summarized above; (b) the independent (intervention) variable in the model is measured as the ratio of total government expenditure to GDP or government consumption to GDP or their logarithms; and (c) the dependent (outcome) variable is measured as the growth rate of per-capita GDP.[6]

Given that the dependent variable is log-difference, the effect-size estimates (γ) can be either elasticities if the independent variable is in logs or semi-elasticities if it is a ratio. Therefore, the effect-size estimates are not comparable. To ensure comparability and allow pooling, we calculate partial correlation coefficients (PCCs) for each effect-size estimate, in accordance with the formula given in the Appendix.

The use of the aforementioned models to estimate the growth impact of government size may pose some estimation issues. For instance, there are reasons for expecting problems of endogeneity in the existing literature. The potential for omitted variables and reverse causality implied by Wagner's (1877) Law and the ratchet effect of Bird (1971, 1972), among others, lead us to suspect issues of endogeneity (Meltzer & Richard, 1981; Bellante & Porter, 1998; De Witte & Moesen, 2010). Wagner's Law suggests that government size increases as countries become

[6] We also conduct a sensitivity check to establish whether the exclusion of non-eligible studies based on our selection criteria affects the meta-regression results. The sensitivity check indicates that the results reported in our study are robust. The results of the sensitivity check are not presented here but are available from the authors upon request.

richer, while the ratchet effect suggests that government size increases significantly during periods of crisis and then grows to the new and higher level after the crisis.

Most studies use instrumental variable techniques (e.g., 2SLS and 3SLS) to properly identify the causal effect between government size and growth (e.g., Fölster & Henrekson, 2001; Afonso & Furceri, 2010). The unavailability of good instruments for government size has also led some studies to employ the Generalized Method of Moments (GMM) to deal with endogeneity (e.g., Romero-Avila & Strauch, 2008). Our meta-analysis controls for these differences in estimation methods together with other observed heterogeneity in the literature using multivariate meta-regressions.

The sources of heterogeneity include model choice (exogenous/neoclassical, endogenous, hybrid models), data type (panel and cross-section data), time period over which panel-data is averaged and whether there is control for business cycles. To account for the effects of these sources of heterogeneity, we code each estimate with respect to the growth model it is derived from, data type and the number of years over which the panel data is averaged.

Other sources of heterogeneity we control for include: (i) whether primary studies control for endogeneity through instrumental variable or dynamic panel-data techniques; (ii) whether control variables such as initial GDP, investment, population (growth or size), government tax revenues, etc. are included in estimated models; (iii) publication type (e.g., journal articles, working papers and book chapters); (iv) publication date; (v) journal quality ranking; (vi) country type (developed versus less developed); (vii) length of periods over which data is averaged in cross-section and panel-data studies; and (viii) the data period.

3. An overview of the evidence base

Our meta-analysis methodology draws on best practice for meta-analysis of research findings in economics and business research (Stanley et al., 2013). We searched five electronic databases - JSTOR, EconLit, Business Source Complete, Google Scholar and ProQuest - for journal articles, working papers and book chapters; using various keywords for government size and growth.[7] We also conduct a manual search which involves examining the references of key reviews and seminal studies that examine the relationship between government size and economic growth.

[7] The keywords for government size include government size, total government expenditure, government consumption, government spending, outlays, public spending, public expenditure and public consumption. Keywords for economic growth include economic growth, GDP, per capita income, growth, economic performance and economic activity.

We have used consistent criteria to include primary studies and effect-size estimates, as indicated above. The inclusion criteria have led to exclusion of primary studies that use other measures of government size such as expenditure or consumption levels or the growth rate of government expenditure/consumption. We also exclude studies that measure growth with GDP level.

Adhering to the above inclusion/exclusion criteria, we have constructed a sample of 799 effect-size estimates reported in 87 primary studies. Online Appendix Table A1 summarizes the characteristics of the primary studies with respect to publication date/type, data type, estimation method and model, data period, measure of government size and country composition.

Table A1 indicates that 59.77% of the studies rely on regression models with panel data, while 17.24% rely on cross-section data. The remaining 22.99% rely on regression models with either time-series data or a mixture of various datasets. With regards to country composition, 29 studies (33.33% of total number of studies) use data on developed countries, while 18 studies (20.69%) use data on less developed countries. The remaining 40 studies (45.98%) use data on a mixture of developed and less developed countries. Finally, 74 out of the 87 primary studies are journal articles, while the remaining 13 are working papers and book chapters.

3.1. Fixed-effect weighted means (FEWMs)

Fixed-effect weighted means (hereafter, FEWMs) are calculated for estimates reported in each study. We cluster the estimates in each study by expenditure type and take the average (or mean) of estimates in each cluster. Rather than take simple means, we take weighted averages given that they are more reliable than simple means. Specifically, FEWMs assign lower weights to less precise estimates (i.e., estimates with large standard errors) and they are also less biased than random effects weighted means when primary-study estimates are affected by publication selection bias (Stanley, 2008; Henmi and Copas, 2010; Stanley and Doucouliagos, 2014).

Tables 1A and 1B provide summary statistics, based on fixed-effect weighted means (FEWMs) and coefficients of variation per study.

Tables 1A and 1B Here

The FEWMs for the effect of total government expenditure on per-capita GDP growth (Table 1A) indicate that 17 studies (32.08% of the total) report 84 estimates (20.44% of the total) that are insignificant; 26 studies (49.06%) report 201 estimates (48.90%) that are negative and significant; and 10 studies (18.87%) report 126 estimates (30.66%) that are positive and significant. Overall, the FEWM for 411 estimates is negative (-0.0083) but statistically insignificant at 95% confidence.

With respect to the relationship between government consumption and economic growth (Table 1B), FEWMs indicate that the effect is insignificant in 14 primary studies (30.43% of total studies) that report 82 estimates (21.13% of the total). The effect is negative and significant in 31 studies (67.39%) reporting 290 estimates (74.75%). In the remaining one study with 16 estimates (4.12% of total estimates) the effect is positive and significant. The overall FEWM for all 388 estimates is negative (-0.1204) and significant with a 95% confidence interval. According to guidelines proposed by Cohen (1988) and Doucouliagos (2011), the FEWM indicates that the effect of government consumption on per-capita GDP is negative and small/medium.

The balance of the evidence from FEWMs indicates that government consumption is detrimental to per-capita GDP growth, whereas government expenditure has no effect. This inference, however, must be qualified on two grounds.

First, the evidence base is characterized by a high degree of heterogeneity. The within-study coefficients of variations are between 0.1 (Adam and Bevan, 2005 in Table 1A; Barro 1991 and 1996 in Table 1B) and 5.9 (Neycheva, 2010 in Table 1B) or as high as 11.7 (Mendoza et al, 1997 in Table 1A). Between-study variation is also high, ranging from -0.65 (Saunders, 1985) to +0.44 (Bukiewicz and Yanikkaya, 2011) in the case of total government expenditure in Table 1A; and from -0.65 (Barro, 2001) to +0.18 (Cronovich, 1998) in the case of government consumption in Table 1B. These variations reduce the reliability of the inference derived from summary measures even if the latter are free of publication selection bias. To address this issue, we model the sources of heterogeneity explicitly and estimate their effects on the variation in the evidence base in Sub-section 3.3.

Secondly, the inference above is valid only if the effect-size estimates reported by primary studies are not subject to selection bias. We adopt funnel plots as well as funnel-asymmetry and precision-effect tests (FAT/PET) in the next sub-section to examine whether selection bias exits and obtain average effect-size estimates corrected for selection bias.

3.2. Investigating publication selection bias

In the meta-analysis literature, a funnel plot is a useful way and a common trend to determine if publication bias exists. Thus, to visually inspect the possibility of publication bias, we first present funnel plots for the association between each government expenditure and growth. Funnel plots are scatter plots of effect sizes against their precision ($1/SE_{ri}$). Figures 1 to 6 present funnel plots for the associations between government size and economic growth. Figures 1 to 3, which show the association between total government expenditure and growth, illustrate less asymmetry considering our reference line, thus they suggest that there are no serious issues of publication selection bias. However, it might not be the case for Figures 4 to 6 that show the association between government consumption and growth.

[Insert Figures Here]

While funnel plots may be useful in determining the presence or absence of publication selection bias, a visual inspection alone does not guarantee the presence or absence of publication bias. In addition, funnel plots are not useful in determining the magnitude or direction of bias, if any exists. Therefore, to thoroughly investigate issues of publication selection bias, we adopt the precision effect test (PET) and the funnel asymmetry test (FAT).

The PET and FAT involve the estimation of a weighted least square bivariate model, in which the effect-size estimate is a linear function of its standard error (see Egger et al., 1997; Stanley, 2008). The theoretical rationale in Egger et al. (1997) is that researchers with small samples would search intensely across model specifications, econometric techniques and data measures to find sufficiently large (hence, statistically significant) effect-size estimates. However, Stanley and Doucouliagos (2007) and Moreno et al. (2009) indicate that a quadratic specification for the relationship between effect size and its standard error is more appropriate if the Egger regression results indicate the presence of significant effect after controlling for selection bias. This specification is referred to as precision-effect estimate with standard errors (PEESE) which applies if the PET rejects the null hypothesis of zero effect (see Appendix). Thus, we run the PEESE analysis only when the coefficient of the precision is significant in the PET-FAT analysis.

We estimate PET-FAT-PEESE models for two measures of government size: the ratio of total government expenditure to GDP and the ratio of government consumption to GDP. Our estimates are obtained using a hierarchical linear model (HLM) specification (Goldstein, 1995), whereby individual effect-size estimates are nested within studies reporting them. The choice is informed by likelihood ratio (LR) tests that compare the HLM with OLS; and the type of HLM is determined by additional LR tests that compare the random-intercepts specification with random-intercepts and random-slopes specification.[8] Estimation results are presented in Tables 2A and 2B, for two full samples and for two country types (developed and LDCs) within each sample.

Tables 2A and 2B Here

Regarding total government expenditure and growth, we find no evidence of genuine effect in the full sample or in LDCs as the coefficient of the precision is statistically insignificant (columns 1 and 3 of Table 2A). In the developed countries sample (column 2 of Table 2A), we find evidence of a negative effect (-0.13) without evidence of publication selection bias. This PET-FAT result is also supported by the PEESE result (column 4) with a slightly more adverse effect (-0.14). Thus, with respect to total government expenditure as a ratio of GDP, we report a negative partial correlation with growth in developed countries only.

With regards to government consumption (fraction of GDP) and growth (Table 2B), we find evidence of a negative effect together with significant negative publication selection bias for the entire sample (column 1) and for the developed countries sample (column 2), but no significant effect for the LDC sample (column 3). PEESE results that correct for non-linear relations between effect-size estimates and their standard errors (columns 4 and 5) confirm the existence of negative effects for the full sample and for developed-country sample (-0.10 and -0.14, respectively).

Statistical significance in the empirical literature has been clearly distinguished from economic (or practical) significance, especially when the size of a statistically significant coefficient is small (Ziliak and McCloskey, 2004). Cohen (1988) indicates that an estimate represents a small effect if its absolute value is around 0.10, a medium effect if it is 0.25 and over and a large effect if it is greater than 0.4. Doucouliagos (2011) argues that the guidelines presented by Cohen (1988) understate the economic significance of empirical effect when partial correlation coefficients (PCCs) are used. Thus, Doucouliagos (2011) suggests that PCCs larger (smaller)

[8] The HLM is employed to deal with data dependence by De Dominicis et al. (2008), Bateman and Jones (2003) and Alptekin and Levine (2012), among others. The likelihood ratio test results that compare HLM with OLS and the types of HLM structures are available on request. We run a series of robustness checks on our result using alternative estimation techniques such as the clustered data analysis (CDA) and these results are consistent with our main results. For brevity, these results are reported in the Online Appendix.

than 0.07 in absolute value can be considered as medium (small) effects whereas those with an absolute value of 0.33 or above can be considered as indicators of large effect.

In the light of the guidance proposed by Doucouliagos (2011), these findings indicate that: (i) total government expenditure (as a fraction of GDP) has a medium and adverse effect on per-capita income growth in developed countries only; (ii) government consumption (as a fraction of GDP) has a medium and adverse effect on per-capita income growth in developed countries and when all countries are pooled together; and (iii) neither total government expenditure nor government consumption has a significant effect on per-capita income growth in LDCs.

However, even PEESE results may have limited applicability when the underlying evidence base is highly heterogeneous. Indeed, the coefficients of variation for the full-sample PCCs in Table 1A and 1B are 9.11 and 1.28, respectively. In addition, the FEWMs and PEESE results are based on the assumption that, apart from the standard errors, all other moderating factors that affect the reported estimates are either zero (in the case of FEWMs) or at their sample means (in the case of PEESE). This assumption is too restrictive because the moderating factors that influence the effect-size estimates reported in primary studies differ between studies and between estimates reported by the same study. Therefore, it is necessary to identify the moderating factors (i.e., the sources of variation) in the evidence base and quantify their influence on the effect-size estimates reported in primary studies. This is done in the next sub-section, followed by a detailed discussion of the implications for the government size – growth relationship in the conclusions and discussion section.

3.3. Addressing Heterogeneity

To identify the sources of heterogeneity and quantify their influence on the reported effect-size estimates, we estimate a multivariate meta-regression model (MRM) for each sample (i.e., for total government expenditure and government consumption). As indicated in the Appendix, we estimate a general and a specific MRM for each sample. The general specification includes all moderating factors that can be measured on the basis of the information we obtain from the primary studies. However, the inclusion of all observable moderating factors poses issues of over-determination and multicollinearity. Therefore, we follow a general-to-specific model routine, which involves the exclusion of the moderating variables with high p-values (highly insignificant variables) one at a time until all remaining variables are statistically significant.

We utilize three sets of moderator variables, which are informed by the theoretical, empirical and methodological dimensions of the research field. The first set captures the variations in econometric specifications and theoretical models adopted by the primary studies. The second captures data characteristics in primary studies and the third reflects the publication characteristics of the primary studies. Summary statistics for moderator variables and their description are presented in Tables 3A and 3B.

Tables 3A and 3B Here

Results from the general and specific MRMs are presented in Table 4A (for total government expenditure) and Table 4B (for government consumption) below. The paragraphs below summarize the findings and interpret their implications for the relationship between government size and per-capita income growth.

Moderator Set 1: Theoretical Models and Econometric Specifications

Differences between theoretical models are captured through a dummy variable that takes the value of 1 if the reported effect-size estimates are obtained from an endogenous growth model, with Solow-type growth model used as base. Results in Table 4A indicate that the underlying theoretical model does not have a significant effect on reported effect-size estimates when the latter are about the effects of total government expenditure (fraction of GDP) on per-capita income growth. However, we note from Table 4B that studies that utilize an endogenous growth model tend to report more adverse effect-size estimates for the relationship between government consumption (fraction of GDP) and per-capita income growth.

The endogenous growth theory offers two reasons as to why government consumption may not be conducive to higher levels of long-run growth (see, e.g., Romer, 1986; Barro, 1990; King and Rebelo, 1990; Lucas, 1990): (i) most government consumption is 'unproductive' (Parente and Prescott, 1991; Barro and Sala-i-Martin, 1995); and (ii) taxes used to finance government consumption are distortionary (Barro, 1990; Parente and Prescott, 1991). Thus, endogenous growth theory predicts higher levels of government consumption lead to lower levels of long-run growth because higher distortionary taxes lead to lower levels of investment in productive activities. Specifically, higher capital income taxes that are used to finance government consumption can lead to lower levels of investment in technology adoption and the differences in these institutional taxes across countries can explain the diversity in long-run growth rate (e.g., Parente and Prescott, 1991). Although the theoretical construction is consistent, the problem arises from the linear specification of the relationship between the level of government consumption and the level of technology. Given that countries are heterogeneous in terms of development levels, institutional structures and the composition of government consumption, a non-linear specification may be more appropriate. However, primary studies included in this meta-analysis do not control for non-linear relationship between government size and growth.

Tables 4A and 4B Here

With respect to econometric dimension, we first examine the difference between estimates based on cross-section data as opposed to panel data. This control is relevant because cross-section estimations overlook fixed effects that may reflect country-specific differences in preferences and technology. In the presence of fixed-effects, estimates based on cross-section data may yield biased results. For instance, Islam (1995) argues that country-specific effects that are ignored in cross-section regressions could be correlated with included explanatory variables and this leads to omitted variable bias. Panel-data estimations can address this source of bias by purging the country-specific fixed effects and focusing on temporal variations in the data. However, the direction of bias (whether upward or downward) in the effect of government size on growth is not certain. Controlling for this econometric dimension in our meta-regressions allows us to draw a conclusion, based on the existing literature, on the direction of bias.

In Table 4A where the focus is on total government expenditure, we find that the use of cross-section data (as opposed to panel data) is associated with more adverse effects on growth, but the effect is insignificant. However, when government size is proxied by government consumption (Table 4B), the use of cross-section data is associated with more adverse effects; and the coefficient is significant. Given the potential bias associated with cross-section data, we interpret this finding as follows: inference about a negative relationship between government size and growth is likely to be biased when it is based on cross-section data only. Therefore, the received wisdom about a negative association between government consumption and growth should be qualified to the extent that it is informed by studies based on cross-section data – even though the data may be averaged over a long time period.

The second dimension of the econometric specification we consider is model specification. In the empirical growth literature, it is well known that the inclusion (or exclusion) of certain regressors in growth regressions can affect the reported effect-size estimate.[9] We include dummies for studies that control for initial GDP per capita, investment share of GDP and population (growth or size). We also include a dummy for studies that control for government tax revenues in their growth regressions, given that the distortionary effects of taxation are a major factor in the debate on government size and growth. Levine and Renelt (1992) indicate that initial GDP per capita, investment share of GDP and population growth are important growth determinants, so parameter estimates may be biased if primary studies do not control for these factors (Easterly and Rebelo, 1993; Agell et al., 1997). Other factors such as geography, institution and culture may also affect economic growth. These factors have been regarded as fundamental sources of growth by some recent studies (e.g., Acemoglu et al., 2002; Tabellini, 2010; Dell et al., 2012). However, very few studies reviewed in this meta-analysis control for these fundamental determinants in their growth regressions. The growth impacts of these fundamental determinants however may work through proximate determinants of growth such as investment (e.g., Rodrik et al., 2004). Thus, we focus our attention to the common set of growth determinants such as initial GDP per capita, investment share of GDP and population (growth or size) in our meta-regression analysis.

MRM results in Tables 4A and 4B confirm that the inclusion of these variables in growth regressions tends to affect the estimates reported in primary studies. For instance, results for total government expenditure and government consumption show that studies that control for population (growth or size) or initial GDP (compared to those that do not) tend to report more adverse effects. We therefore conclude that it would be good practice for researchers to include the key regressors in their regressions with a view to minimize the risk of model specification bias and the additional heterogeneity that would result from such biases.

Another dimension of the econometric specification that may affect the reported estimates concerns the length of time over which both regressors and regressands are averaged. Two arguments can be put forward in favour of averaging. First, averaging over a period equal to the business cycle (usually five years) eliminates the effect of business cycle and this is particularly important if measures of business cycle (e.g., output gap) are not included in the model. Secondly, estimates based on data averaged over 5 years or more can be interpreted as medium- to long-run effects as opposed to short-run effects. Thus, to verify if estimates reported in

[9] For reviews of the literature on the importance of various variables, see Levine and Renelt (1992), Durlauf et al. (2005) and Glewwe et al. (2014).

primary studies are affected by the period of data averaging, we control for studies where data is averaged over five years or more, with others where annual data is used or the data is averaged over periods of less than 5 years as base. In both government consumption and total government expenditure samples, we find that the data averaging period has no statistically significant effect on estimates reported in primary studies.

We further examine the nature of reported estimates for studies that use panel data and adopt data averaging of 5 years or more and also those that use cross-section data with data averaging of 5 years and above (as opposed to those that do not). In the total government expenditure specification, the coefficient for studies that use panel data with data averaging of 5 years or more is statistically insignificant. However, we find that studies that use cross-section data with data averaging of 5 years and above tend to report less adverse effects of government consumption on growth. This is an interesting finding because it indicates the effect of government consumption tends to be less adverse in the long run.

This finding also indicates that the bias that results from failure to account for country fixed-effects in cross-section data is larger when the data averaging period is short. This is to be expected because country fixed effects are more likely to remain fixed over shorter time horizons. Another implication of this finding is that the relatively larger adverse effects reported by studies using cross-section data are likely to be driven by the dominance of the effect-size estimates based on short time horizons.

The last dimension relating to econometric and theoretical specification concerns the econometric methodology used by primary studies. In the empirical growth literature, various econometric methods have been used and these methodologies aim at addressing specific issues. For instance, OLS estimates have been found to be inconsistent and biased in the presence of endogeneity. In the government size-growth literature, reverse causality is a potential problem and a source of endogeneity given that higher income countries have been identified to choose larger governments. This is consistent with Wagner's Law which postulates that government size tends to grow with economic development. The ratchet effect, i.e., government size rises significantly during periods of crises and grows further after the crisis, is also often not accounted for in primary studies.

Omitted variables bias could be another source of endogeneity. This bias emerges as a result of unobservable factors in the literature that relates government size to several political and economic variables (e.g., Karras, 1993; Gali, 1994; Fatás & Mihov, 2001). The growth literature refers to these unobservable factors as country-specific effects and they are usually ignored in cross-section regressions. Several studies also fail to control for relevant observable variables in growth regressions. For instance, a growing body of literature demonstrates the role of institutional quality, culture and geography in growth (e.g., Acemoglu et al., 2002; Tabellini, 2010; Dell et al., 2012) but very few studies reviewed in this meta-analysis control for these factors in their growth regressions.

To address endogeneity, some primary studies tend to use instrumental variable (IV) techniques such as 2SLS, 3SLS and GMM. Therefore, we control for studies that control for endogeneity as opposed to those that do not. The coefficient on the dummy for studies that control for endogeneity is positive and significant in the total government expenditure sample but insignificant in the government consumption sample. This finding suggests that the received wisdom about the adverse effect of government expenditure on growth may be informed by biased estimates from primary studies that do not control for heterogeneity.

Moderator Set 2: Data Characteristics

With regards to data characteristics, we first examine if the government size-growth relationship is time variant. Studies often re-examine the government size-growth relationship using different datasets when newer datasets become available. Since new datasets are likely to reveal different trends in government spending, we include dummy variables to capture the 'recentness' of data and how data time periods affect reported estimates. We include dummy variables to capture the decade in which the beginning year of the data period falls. For instance, "Data Period (1980+)" captures studies with data beginning in the 1980s. The excluded category is "Data period (1950+)".

MRA results for the total government expenditure sample mainly show statistically insignificant coefficients for data period dummies. However, from Table 4B coefficients for data period dummies are negative and statistically significant. Particularly, we note that the magnitude of the coefficient increases as the decades increase. Thus, the most adverse effect is observed for "Data Period 1990+". This suggests that studies that use newer datasets tend to report more adverse effects of government consumption on growth. This is also the case for total government expenditure as the only included dummy in the specific model (Data Period 2000+), is negative and statistically significant as well.

Our results therefore show that the use of data from more recent time periods is associated with increased adverse effect of government size on growth. This could be a result of the increased globalization and financial integration in recent times. Data from the World Trade Organization shows that the rate of globalization and economic integration proxied by trade levels has more than doubled since the beginning of the 1970s. A large body of literature has demonstrated how increased globalization affects the size of government and various government spending policies (see, e.g., Rodrik, 1998; Garrett, 2001; Garrett & Mitchell, 2001; Dreher, 2006; Dreher et al., 2007; Gastaldi & Liberati, 2011). For example, it has been argued that government spending and taxes tend to increase as governments try to ameliorate the adverse consequences of globalization for income volatility and inequalities (Rodrik, 1998; Gastaldi & Liberati, 2011). Therefore, the likely adverse effects of government size on growth should be evaluated in the light of where and why governments are spending limited resources and raising taxes to finance them.

We also examine the effect of country type. Although the PET-FAT results reveal a negative effect of big government on growth in developed countries, it is worthwhile to control for this in our MRA as well. This ensures the inclusion of all relevant moderator variables that capture the necessary dimensions. We therefore control for studies that report estimates using data on developed countries as opposed to those that use data from LDCs and a mixture of both developed and LDCs. Results from Table 4A confirm what the PET-FAT results suggest. The developed country dummy is significant and negative, suggesting that studies that use data on developed countries tend to report more adverse effects of total government expenditure on growth compared to those that use LDC samples and mixed samples. From Table 4B, the coefficient of studies that use developed countries data is also negative but insignificant.

Moderator Set 3: Publication Characteristics

Under the publication characteristics dimension, we first control for publication type. Here, we examine if journal articles tend to systematically report different effect sizes in comparison to book chapters and working papers. This allows us to determine whether researchers, authors and editors are predisposed to publishing and/or accepting studies with statistically significant results that are consistent with theory to justify model selection. Using book chapters and working papers as base, we include a dummy for journal articles in our MRA specification. Results reveal that studies published in journals tend to report less adverse effects of government size on growth. This is consistent across both measures of government size and specification type (i.e., both general and general-to-specific).

Furthermore, we examine if perceived quality of the publication outlet is associated with variations in reported estimates. We measure perceived publication outlet quality using two sources of journal ranking data.[10] From Table 4A, the coefficient for studies published in high-ranked journals is statistically insignificant. However, studies published in high-ranked journals tend to report more adverse effects of government consumption on growth (Table 4B).

Next, we control for publication year. Examining publication year enables us to identify whether more recent studies, as opposed to older studies, tend to report different estimates. Thus, we include dummy variables similar to those constructed for data period. For instance, studies published between 1998 and 2013 fall under "Publication Year (1990+)" and those published between 2001 and 2013 fall under "Publication Year (2000+)". Leaving 1980+ as base, we control for studies published in the decades starting 1990, 2000 and 2010. In both government consumption and total government expenditure specifications, publication year dummies are significant; but they do not reflect a consistent pattern. Hence, we conclude that it is not possible to infer whether newer studies tend to report more or less adverse effects compared to studies published before them.

3.3.1. Robustness check

For robustness checks, we run additional regressions using different variables to capture the dimensions of the research field pertaining to data period, publication dates, journal quality and data type. With regards to data period, we include a dummy variable which captures studies that include data from anywhere before the 1980s (inclusive) and a second dummy variable which captures studies that include data for the 1990s and 2000s. We exclude studies that have a mixture of these two groups as base. MRA results indicate that the government size-growth effect changes over time and hence is time variant. Specifically, consistent with our main results, newer datasets tend to report more adverse effects of government size on growth.

For publication date, we include a dummy for studies that were published in the 2000s, leaving out other publication years as base. We find that in the total expenditure-growth literature, the

[10] The Australian Business Dean's Council (ABDC) and the Australian Research Council (ARC) present classifications for journal quality. Journals are ranked in descending order of quality as A*, A, B and C. Thus, we introduce a dummy for A* and A ranked journals (high quality) in our MRA and use other ranks as base.

year of study publication does not affect the nature of reported estimates. However, in the government consumption-growth literature, we find that more recent publications tend to report more negatively on the effects of government consumption on growth and this is consistent with our main results which use a different classification for publication dates.

With regards to journal rank, our main results classify A* and A ranked journals into one category. We examine if results related to journal rankings are robust to a further distinction between A* and A. Thus, we include two dummy variables capturing A* and A ranked journals independently. We find that the dummy for A* ranked studies is significant, while that of A ranked studies is not.

Lastly, we examine if there are any systematic differences in estimates reported by primary studies using time-series data and those using cross-section. Thus, we include a dummy for cross-section studies (as opposed to those using time-series data). We find that cross-section studies tend to report more negatively on the impact of government size on growth. Overall, results from these regressions are consistent with our main results. Given space constraints, these results are reported in the Online Appendix.

4. Summary and conclusions

This paper reviews the empirical literature on the association between government size and economic growth. We focus on total government expenditure and government consumption expenditure (as a share of GDP) as measures of government size. Results are based on a synthesis of 87 studies solely examining the effect of the government size on per-capita GDP growth. We control for publication selection bias and address issues of heterogeneity in the existing literature.

Bivariate meta-regression results reported above indicate that the average effect of government size on growth, using both proxies of government size, is medium and negative in developed countries. The average effect of total government expenditure is insignificant in both LDCs and mixed-country samples (i.e., when developed and LDCs are pooled together). On the other hand, the average effect of government consumption is insignificant in LDCs, but it is medium and negative in both developed countries and mixed-country samples.

These findings suggest that the existing evidence does not support an overall inference that establishes a negative relationship between government size and per-capita income growth for several reasons including: (i) potential biases induced by reverse causality between government size and per-capita income; (ii) lack of control for country-specific effects in cross-section studies; and (iii) absence of control for non-linear relationships between government size and per-capita GDP growth.

Furthermore, the effect is specific to the level of development: a larger government size tends to have a negative effect on per-capita income growth as the level of income increases. This finding ties in with the Armey curve hypothesis (Armey, 1995) which posits an inverted-U relationship between government size and economic growth. A small government can enhance economic growth by providing a conducive institutional environment characterized by rule of law and protection of property rights, which are often regarded as important factors for economic growth. However, when an economy becomes richer, the size of the government tends to grow beyond its efficient level, so a further rise in government size would hamper economic growth. There are several possible reasons for this argument. First, government size may be characterized by decreasing returns when government size is sufficiently large. The second reason is related to the distortionary nature of taxes, which is minimal for low levels of taxation, but beyond a certain threshold, they grow rapidly and become extremely large (e.g., Barro, 1990; Agell, 1996) to have adverse effects on saving, investment and other forms of productive behaviour. Third, rent-seeking activities tend to increase in countries with larger governments (Buchanan 1980). Hence, our findings suggest that estimates of the relationship between government size and growth obtained from linear estimations may be biased (see also, Barro, 1990).

In addition, developed countries tend to have well-developed systems of automatic stabilisers such as social security expenditure and progressive taxation. According to the World Social Security Report 2010/11, Europe spends between 20 and 30 per cent of GDP on social security, while in most African countries social security spending accounts only for 4–6 per cent of GDP.

According to Devarajan et al. (1996), social security expenditure is unproductive and as such they may be driving the negative relationship between government size and per-capita income growth in developed countries. However, social security expenditure and other forms of automatic stabilisers may be conducive to lower growth rates because of the reverse causality they inject into the government size-growth relationship. As indicated by Bergh and Henrekson (2011), automatic stabilisers on the expenditure sides would increase as GDP falls. This well-known feature of the automatic stabilisers introduces a negative bias in the estimates for the effect of government size on growth. The risk of such bias is higher in developed countries with higher incidence of automatic stabilisers. Indeed, this risk of bias is confirmed by our findings that the effect of government size is less adverse: (i) in developing as opposed to developed countries; and (ii) in studies that control for endogeneity as opposed to those that do not.

Furthermore, the more pronounced negative effects for developed countries may be related to Wagner's Law, which indicates that government size increases with the level of income. There is evidence indicating that the long-run elasticity of government size with respect to growth in developed countries is large (Lamartina and Zaghini, 2011). In this case, the government size-GDP ratio for developed countries will grow faster than LDCs for a given increase in GDP. This additional endogeneity problem leads to what Roodman (2008) describes as 'the looking glass problem: if the government size-GDP ratio increases with GDP (i.e., if Wagner's Law holds), then the stronger negative effects reported on developed countries may be due to either lack of control for endogeneity in the growth regressions or absence of adequate instruments or both.

With regards to other sources of heterogeneity, we find that model specification, study design and sample used by primary studies do affect effect-size estimates. Our findings complement those of Nijkamp and Poot (2004), who report that cross-section studies are more likely to report detrimental effects of big government on economic growth.

We also find that studies published in journals tend to report less adverse effects compared to working papers and book chapters. This is consistent across both measures of government size and thus raises the question as to whether the negative association between government size and per-capita income growth may be driven by less rigorous external reviewing processes in the case of book chapters and working papers. However, we do not wish to overemphasize this because in the government consumption sample, we find that studies published in higher-ranked journals tend to report more adverse effects of government size on growth. This may be an indication of the 'Winner's curse' - whereby journals with good reputation capitalize on their reputation and publish 'more selected' findings (see Costa-Font et al., 2013; Ugur, 2014).

In conclusion, our findings show that where an evidence base is too diverse, meta-analysis can be highly effective in synthesizing the evidence base and accounting for the sources of heterogeneity among reported findings. Our findings in this study indicate that government size is more likely to be associated with negative effects on per-capita income growth in developed countries. They also indicate that the medium-sized adverse effects in developed countries may be biased due to endogeneity and reverse causality problems, which are either unaddressed in a large segment of the evidence base or the instruments used to address these problems are weak or both. Therefore, we call for caution in establishing casual links between government size and per-capita income growth. We also call for use of non-linear models in the estimation of the government size – growth relationship. As indicated by Agell (1996), non-linear models may provide richer evidence on the optimal government size, particularly when the latter is measured in terms of tax revenues. Finally, as indicated by Kneller et al. (1999), Poot (2000) and Bergh and Henrekson (2011), we call for further research on the relationship between particular components of the government size and growth as such studies are more likely to produce policy-relevant findings compared to studies that focus on total measures of government size.

5. References

Abreu, M., de Groot, H., & Florax, R. (2005). A meta-analysis of beta convergence. Tinbergen Institute Discussion Papers, No. TI 2005-001/3.

Acemoglu, D., Johnson, S., & Robinson, J. A. (2002). Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution. Quarterly Journal of Economics, 117, 1231-1294.

Adam, C., & Bevan, D. (2005). Fiscal deficits and growth in developing countries. Journal of Public Economics 89 (89), 571– 597.

Afonso, A., & Furceri, D. (2010). Government Size, Composition, Volatility and Economic Growth. European Journal of Political Economy, 26(4), 517-532.

Afonso, A., & Jalles, J. T. (2013). Fiscal composition and long-term growth. Applied Economics, 46(3), 349-358. doi: 10.1080/00036846.2013.848030

Afonso, A., & Jalles, J. T. (2013). Do Fiscal Rules Matter for Growth? Applied Economics Letters, 20(1-3), 34-40.

Afonso, A. A., Gruner, H. P., Kolerus, C. E., & Grüner, H. P. (2010). Fiscal Policy and Growth: Do Financial Crises Make a Difference? European Central Bank, Working Paper Series: 1217. University of Mannheim - Department of Economics; University of Mannheim - Center for Doctoral Studies in Economics and Management (CDSEM).

Agell, J. (1996). Why Sweden's welfare state needed reform. Economic Journal, 106(439), 1760–1771.

Agell, J., Lindh, T., & Ohlsson, H. (1997). Growth and the public sector: A critical review essay. European Journal of Political Economy, 13(1), 33-52.

Aghion, P., & Howitt, P. (1992). A model of growth through creative destruction. Econometrica, 60, 323-351.

Alptekin, A., & Levine, P. (2012). Military expenditure and economic growth: A meta-analysis. European Journal of Political Economy, 28(4), 636-650.

Andrés, J., Doménech, R., & Molinas, C. (1996). Macroeconomic performance and convergence in OECD countries. European Economic Review, 40(9), 1683-1704.

Angelopoulos, K., & Philippopoulos, A. (2007). The Growth Effects of Fiscal Policy in Greece 1960-2000. Public Choice, 131(1-2), 157-175.

Angelopoulos, K., Economides, G., & Kammas, P. (2007). Tax-spending policies and economic growth: theoretical predictions and evidence from the OECD. European Journal of Political Economy, 23(4), 885-902.

Angelopoulos, K., Philippopoulos, A., & Tsionas, E. (2008). Does Public Sector Efficiency Matter? Revisiting the Relation between Fiscal Size and Economic Growth in a World Sample. Public Choice, 137(1-2), 245-278.

Arin, K. P. (2004). Fiscal Policy, Private Investment and Economic Growth: Evidence from G-7 Countries. Massey University - Department of Commerce.

Armey, D. (1995). The Freedom Revolution. Washington: Regnery Publishing.

Aschauer, D. (1989). Is government spending productive? Journal of Monetary Economics, 23, 177-200.

Atesoglu, H. S., & Mueller, M. J. (1990). Defence spending and economic growth. Defence Economics, 2, 19 - 27.

Bajo-Rubio, O. (2000). A further generalization of the Solow growth model: the role of the public sector. Economics Letters, 68, 79–84.

Barro, R. J. (1989). A Cross-Country Study of Growth, Saving, and Government. NBER Working Paper No. 2855.

Barro, R. J. (1990). Government spending in a simple model of endogenous growth. Journal of Political Economy, 98(5), S103-S125.

Barro, R. J. (1991). Economic Growth in a Cross Section of Countries. The Quarterly Journal of Economics, 106(2), 407-443.

Barro, R. J. (1996). Determinants of Economic Growth: A Cross-Country Empirical Study. NBER Working Paper No. 5698.

Barro, R. J. (2001). Human Capital and Growth. American Economic Review, 91(2), 12-17.

Barro, R. J., & Sala-i-Martin, X. (1995). Economic Growth Cambridge, MA: MIT Press.

Barro, R. J., & Sala-i-Martin, X. (2004). Economic Growth. Cambridge, MA: MIT Press.

Bateman, I., & Jones, A. (2003). Contrasting conventional with multi-level modelling approaches to meta-analysis: expectation consistency in U.K. woodland recreation values. . Land Economics 79(2), 235–258.

Bellante, D., & Porter, P. (1998). Public and private employment over the business cycle: A ratchet theory of government growth. Journal of labor research, 19(4), 613-628. doi: 10.1007/s12122-998-1052-9

Bellettini, G., & Ceroni, C. B. (2000). Social Security Expenditure and Economic Growth: An Empirical Assessment. Research in Economics, 54(3), 249-275.

Benos, N., & Zotou, S. (2014). Education and Economic Growth: A Meta-Regression Analysis. World Development, 64(0), 669-689.

Bergh, A., & Henrekson, M. (2011). Government Size and Growth: A Survey and Interpretation of the Evidence. Journal of Economic Surveys, 25(5), 872-897.

Bergh, A., & Karlsson, M. (2010). Government size and growth: Accounting for economic freedom and globalization. Public Choice, 142(1-2), 195-213.

Bergh, A., & Öhrn, N. (2011). Growth Effects of Fiscal Policies: A Critical Appraisal of Colombier's (2009) Study. IFN Working Paper No. 865.

Bernhard, H. (2001). The Scope of Government and its Impact on Economic Growth in OECD Countries. Kieler Arbeitspapiere, No. 1034.

Bird, R. M. (1971). Wagner‘s Law of Expanding State Activity. Public Finance, 26(1), 1–26.

Bird, R. M. (1972). The Displacement Effect: A Critical Note. Finanzarchiv, 30(3), 454–463.

Bodman, P., Campbell, H., & Le, T. (2012). Public investment, taxation, and long-run output in economies with multi-level governments. Economic Modelling, 29, 1603–1611.

Bojanic, A. N. (2013). The Composition of Government Expenditures and Economic Growth in Bolivia. Latin American Journal of Economics, 50(1), 83-105.

Bose, N., Haque, M. E., & Osborn, D. R. (2007). Public Expenditure and Economic Growth: A Disaggregated Analysis for Developing Countries. Manchester School, 75(5), 533-556.

Brumm, H. J. (1997). Military Spending, Government Disarray, and Economic Growth: A Cross-Country Empirical Analysis. Journal of Macroeconomics, 19(4), 827-838.

Buchanan, J. (1980). Rent-seeking and Profit-seeking. In: Buchanan, J.M., Tullock, G. (eds) Towards a theory of the rent-seeking society. TX, College Station.

Butkiewicz, J. L., & Yanikkaya, H. (2011). Institutions and the Impact of Government Spending on Growth. Journal of Applied Economics, 14(2), 319-341.

Castro, V. (2011). The Impact of the European Union Fiscal Rules on Economic Growth. Journal of Macroeconomics, 33(2), 313-326.

Chen, S.-T., & Lee, C.-C. (2005). Government size and economic growth in Taiwan: A threshold regression approach. Journal of Policy Modeling, 27(9), 1051-1066.

Cohen, J. (1988). Statistical Power Analysis for the Behavioural Sciences Hillsdale, NJ.

Colombier, C. (2009). Growth effects of fiscal policies: an application of robust modified M-estimator. Applied Economics, 41(7), 899-899.

Commander, S., Davoodi, H., & Lee, U. (1999). The Causes and Consequences of Government for Growth and Well-being. World Bank Policy Research Working Paper.

Cooray, A. (2009). Government Expenditure, Governance and Economic Growth. Comparative Economic Studies, 51(3), 401-418.

Costa-Font, J., McGuire, A., & Stanley, T. (2013). Publication selection in health policy research: The winner's curse hypothesis. Health policy, 109(1), 78-87.

Cronovich, R. (1998). Measuring the Human Capital Intensity of Government Spending and Its Impact on Economic Growth in a Cross Section of Countries. Scottish Journal of Political Economy, 45(1), 48-77.

d'Agostino, G., Dunne, J. P., & Pieroni, L. (2010). Assessing the Effects of Military Expenditure on Growth Oxford Handbook of the Economics of Peace and Conflict. Oxford University Press.

d'Agostino, G., Dunne, J. P., & Pieroni, L. (2012). Corruption, Military Spending and Growth. Defence and Peace Economics, 23(6), 591-591.

Dalic, M. (2013). Fiscal policy and growth in new member states of the EU: a panel data analysis. Financial Theory and Practice, 37(4), 335-360.

Dar, A. A., & AmirKhalkhali, S. (2002). Government size, factor accumulation, and economic growth: evidence from OECD countries. Journal of Policy Modeling, 24(7–8), 679-692.

De Dominicis, L., Florax , R., & Groot, H. (2008). A meta-analysis on the relationship between income inequality and economic growth. Scottish Journal of Political Economy, 55(5), 654-682.

De Gregorio, J. (1992). Economic growth in Latin America. Journal of Development Economics, 39(1), 59-84.

Demidenko, E. (2004). Mixed Models: Theory and Applications. Hoboken, NJ: Wiley.

De Witte, K., & Moesen, W. (2010). Sizing the government. Public Choice, 145, 39-55.

Dell, M., Jones, B. F., & Olken, B. A. (2012). Temperature Shocks and Economic Growth: Evidence from the Last Half Century. American Economic Journal: Macroeconomics, 4(3), 66–95.

Devarajan, S., Swaroop, V., & Zou, H.-f. (1996). The Composition of Public Expenditure and Economic Growth. Journal of Monetary Economics, 37(2), 313-344.

Diamond, J. (1998). Fiscal indicators for economic growth: The Government own saving concept re-examined. Journal of Public Budgeting, Accounting & Financial Management, 9(4), 627-651.

Doucouliagos, H. (2011). How Large is Large? Preliminary and relative guidelines for interpreting partial correlations in economics. Deakin University, School of Accounting, Economics and Finance Working Paper Series (5).

Doucouliagos, H., & P. Laroche (2009). Unions and profits: A meta-regression analysis. Industrial Relations: A Journal of Economy and Society, 48(1), 146-184.

Dowrick, S. (1996). Estimating the Impact of Government Consumption on Growth: Growth Accounting and Endogenous Growth Models. Empirical Economics, 21(1), 163-186.

Dreher, A. (2006). The influence of globalization on taxes and social policy: An empirical analysis for OECD countries. European Journal of Political Economy, 22(1), 179-201. doi: http://dx.doi.org/10.1016/j.ejpoleco.2005.04.006

Dreher, A., Sturm, J.-E., & Ursprung, H. W. (2007). The impact of globalization on the composition of government expenditures: Evidence from panel data. Public Choice, 134(3), 263-292. doi: 10.1007/s11127-007-9223-4

Durlauf, S. N., Johnson, P. A., & Temple, J. R. W. (2005). Growth Econometrics. In A. Philippe & N. D. Steven (Eds.), Handbook of Economic Growth (Vol. Volume 1, Part A, pp. 555-677): Elsevier.

Easterly, W., & Rebelo, S. (1993). Fiscal policy and economic growth: An empirical investigation. Journal of Monetary Economics, 32(3), 417-458.

Egger, M., Smith, D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. BMJ, 315, 629 - 634.

Engen, E. M., & Skinner, J. S. (1992). Fiscal Policy and Economic Growth. NBER Working Paper No. w4223.

Evans, P., & Karras, G. (1994). Is government capital productive? Evidence from a panel of seven countries. Journal of Macroeconomics, 16(2), 271-279.

Fatás, A., & Mihov, I. (2001). Government size and automatic stabilizers: international and intranational evidence. Journal of International Economics, 55(1), 3-28.

Feder, G. (1983). On exports and economic growth. Journal of Development Economics, 12(1–2), 59-73.

Fölster, S., & Henrekson, M. (1999). Growth and the public sector: a critique of the critics. European Journal of Political Economy, 15(2), 337-358.

Fölster, S., & Henrekson, M. (2001). Growth effects of government expenditure and taxation in rich countries. European Economic Review, 45(8), 1501-1520.

Gali, J. (1994). Government size and macroeconomic stability. European economic review, 38(1), 117-132.

Garrett, G. (2001). Globalization and government spending around the world. Studies in comparative international development, 35(4), 3-29.

Garrett, G., & Mitchell, D. (2001). Globalization, government spending and taxation in the OECD. European Journal of Political Research, 39(2), 145-177.

Garrison, C. B., & Lee, F.-Y. (1995). The effect of macroeconomic variables on economic growth rates: A cross-country study. Journal of Macroeconomics, 17(2), 303-317.

Gastaldi, F., & Liberati, P. (2011). Economic integration and government size: a review of the empirical literature. Financial Theory and Practice, 35(3), 327-384.

Ghali, K. (2003). Government Spending, Budget Financing, and Economic Growth: The Tunisian Experience. The Journal of Developing Areas, 36(2), 19-37.

Ghosh, S., & Gregoriou, A. (2008). The Composition of Government Spending and Growth: Is Current or Capital Spending Better? Oxford Economic Papers, 60(3), 484-516.

Ghura, D. (1995). Macro Policies, External Forces, and Economic Growth in Sub-Saharan Africa. Economic Development and Cultural Change, 43(4), 759-778.

Glewwe, P., Maïga, E., & Zheng, H. (2014). The Contribution of Education to Economic Growth: A Review of the Evidence, with Special Attention and an Application to Sub-Saharan Africa. World Development, 59, 379-393.

Goldstein, H. (1995). Multilevel statistical models 2nd ed. . London: Edward Arnold.

Gould, D. J., & Amaro-Reyes, J. A. (1983). The Effects of Corruption on Administrative Performance. World Bank Staff Working Paper No. 580.

Grier, K. B., & Tullock, G. (1989). An empirical analysis of cross-national economic growth, 1951-1980. Journal of Monetary Economics, 24(2), 259-276.

Grimes, A. (2003). Economic growth and the size & structure of government: Implications for New Zealand. New Zealand Economic Papers, 37(1), 151-174.

Grossman, P. J. (1990). Government and Growth: Cross-Sectional Evidence. Public Choice, 65(3), 217-227.

Guseh, J. S. (1997). Government Size and Economic Growth in Developing Countries: A Political-Economy Framework. Journal of Macroeconomics, 19(1), 175-192.

Hamdi, H., & Sbia, R. (2013). Dynamic Relationships between Oil Revenues, Government Spending and Economic Growth in an Oil-Dependent Economy. Economic Modelling, 35, 118-125.

Hamilton, A. (2013). Small is Beautiful, at Least in High-Income Democracies: The Distribution of Policy-Making Responsibility, Electoral Accountability, and Incentives for Rent Extraction. The World Bank Institute. Policy Research Working Paper 6305.

Hansen, P. (1994). The government, exporters and economic growth in New Zealand. New Zealand Economic Papers, 28(2), 133-142.

Hansson, P., & Henrekson, M. (1994). A new framework for testing the effect of government spending on growth and productivity. Public Choice, 81(3-4), 381-401.

Henmi, M., & Copas, J. B. (2010). Confidence intervals for random effects meta-analysis and robustness to publication bias. Stat Med, 29(29), 2969-2983.

Husnain, M. I. u., & Ghani, E. (2010). Expenditure-Growth Nexus: Does the Source of Finance Matter? Empirical Evidence from Selected South Asian Countries [with Comments]. The Pakistan Development Review, 49(4), 631-640.

Islam, N. (1995). Growth Empirics: A Panel Data Approach. The Quarterly Journal of Economics, 110(4), 1127-1170.

Kalaitzidakis, P., & Tzouvelekas, V. (2011). Military spending and the growth-maximizing allocation of public capital: a cross-country empirical analysis. Economic Inquiry, 49(4), 1029-1041.

Karras, G. (1993). Employment and output effects of government spending: is government size important? Economic Inquiry, 31(3), 354.

Kelly, T. (1997). Public Expenditures and Growth. Journal of Development Studies, 34(1), 60-84.

King, R., Rebelo, S., 1990. Public policy and economic growth: Developing neoclassical implications. Journal of Political Economy 98 (1), s126–151.

Kneller, R., Bleaney, M. F., & Gemmell, N. (1999). Fiscal Policy and Growth: Evidence from OECD Countries. Journal of Public Economics, 74(2), 171-190.

Kormendi, R. C., & Meguire, P. G. (1985). Macroeconomic determinants of growth: Cross-country evidence. Journal of Monetary Economics, 16(2), 141-163.

Lamartina, S., & Zaghini, A. (2011). Increasing Public Expenditure: Wagner's Law in OECD Countries. German Economic Review, 12(2), 149-164.

Landau, D. (1983). Government Expenditure and Economic Growth: A Cross-Country Study. Southern Economic Journal, 49(3), 783-792.

Landau, D. (1986). Government and Economic Growth in the Less Developed Countries: An Empirical Study for 1960-1980. Economic Development and Cultural Change, 35(1), 35-75.

Landau, D. (1997). Government expenditure, human capital creation and economic growth. Journal of Public Budgeting, Accounting & Financial Management, 9(3), 467-467.

Lee, B. S., & Lin, S. (1994). Government Size, Demographic Changes, and Economic Growth. International Economic Journal, 8(1), 91-108.

Lee, J.-W. (1995). Capital goods imports and long-run growth. Journal of Development Economics, 48(1), 91-110.

Levine, R., & Renelt, D. (1992). A Sensitivity Analysis of Cross-Country Growth Regressions. The American Economic Review, 82(4), 942-963.

Lucas, R., (1988). On the mechanics of economic development. Journal of Monetary Economics, 22, 3-42.

Lucas, R., (1990). Supply-side economics: an analytical review. Oxford Economic Papers 42 (2), 293–316.

Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A Contribution to the Empirics of Economic Growth. The Quarterly Journal of Economics, 107(2), 407-437.

Marlow, M. (1986). Private sector shrinkage and the growth of industrialized economies. Public Choice, 49(2), 143-154.

Martin, R., & Fardmanesh, M. (1990). Fiscal variables and growth: A cross-sectional analysis. Public Choice, 64(3), 239-251.

Mauro, P. (1995). Corruption and Growth. Quarterly Journal of Economics, 110(3), 681–712.

McCulloch, C. E., S. R. Searle, and J. M. Neuhaus (2008). Generalized, Linear, and Mixed Models (2nd ed). Hoboken, NJ: Wiley.

Meltzer, A. H., & Richard, S. F. (1981). A rational theory of the size of government. The Journal of Political Economy, 914-927.

Mendoza, E. G., Milesi-Ferretti, G. M., & Asea, P. (1997). On the ineffectiveness of tax policy in altering long-run growth: Harberger's superneutrality conjecture. Journal of Public Economics, 66(1), 99-126.

Miller, S. M., & Russek, F. S. (1997). Fiscal Structures and Economic Growth: International Evidence. Economic Inquiry, 35(3), 603-613.

Mo, P. H. (2007). Government Expenditures and Economic Growth: The Supply and Demand Sides. Fiscal Studies, 28(4), 497-522.

Moreno, S.G., A. J. Sutton, A. Ades, T. D. Stanley, K. R. Abrams, J. L. Peters, and N. J. Cooper (2009). Assessment of regression-based methods to adjust for publication bias through a comprehensive simulation study, BMC Medical Research Methodology, 9(2), 1–17.

Munnell, A. (1990). Why Has Productivity Growth Declined? Productivity and Public Investment. . New England Economic Review, Federal Reserve Bank of Boston Jan, 1990 Issue, 3–22.

Murphy, K. M., Shleifer, A., & Vishny, R. W. (1991). The Allocation of Talent: Implications for Growth. The Quarterly Journal of Economics, 106(2), 503-530.

Neycheva, M. (2010). Does public expenditure on education matter for growth in Europe? A comparison between old EU member states and post-communist economies. Post - Communist Economies, 22(2), 141-141.

Nijkamp, P., & Poot, J. (2004). Meta-Analysis of the Effect of Fiscal Policies on Long-Run Growth. European Journal of Political Economy, 20(1), 91-124.

Nketiah-Amponsah, E. (2009). Public Spending and Economic Growth: Evidence from Ghana (1970-2004). Development Southern Africa, 26(3), 477-497.

Odedokun, M. O. (1997). Relative effects of public versus private investment spending on economic efficiency and growth in developing countries. Applied Economics, 29(10), 1325-1336.

Parente, S., Prescott, E., 1991. Technology adoption and growth. NBER Working Paper No. 3733.

Plümper, T., & Martin, C. W. (2003). Democracy, Government Spending, and Economic Growth: A Political-Economic Explanation of the Barro-Effect. Public Choice, 117(1/2), 27-50.

Poot, J. (2000). A Synthesis of Empirical Research on the Impact of Government on Long-Run Growth. Growth and Change, 31(4), 516-546.

Ram, R. (1986). Government Size and Economic Growth: A New Framework and Some Evidence from Cross-Section and Time-Series Data. The American Economic Review, 76(1), 191-203.

Rebelo, S. T. (1991). Long-Run Policy Analysis and Long-Run Growth. Journal of Political Economy, 99(3), 500–521.

Rodrik, D. (1998). Why do More Open Economies Have Bigger Governments? Journal of Political Economy, 106(5), 997-1032. doi: 10.1086/250038

Rodrik, D., Subramanian, A., & Trebbi, F. (2004). Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development. Journal of Economic Growth, 9(2), 131-165.

Romer, P. M. (1986). Increasing Returns and Long-Run Growth. Journal of Political Economy, 94(5), 1002-1037. doi: 10.2307/1833190

Romer, P. M. (1989). Human Capital and Growth: Theory and Evidence. NBER Working Paper No. w3173.

Romer, P. M. (1990). Endogenous Technological Change. Journal of Political Economy, 98, S71-S102.

Romero-Avila, D., & Strauch, R. (2008). Public Finances and Long-Term Growth in Europe: Evidence from a Panel Data Analysis. European Journal of Political Economy, 24(1), 172-191.

Roodman, D. (2008). Through the Looking Glass, and What OLS Found There: On Growth, Foreign Aid, and Reverse Causality. Center for Global Development Working Paper No. 137.

Roubini, N., & Sala-i-Martin, X. (1992). Financial repression and economic growth. Journal of Development Economics, 39(1), 5-30.

Rubinson, R. (1977). Dependency, Government Revenue, and Economic Growth 1955-70. Studies in Comparative International Development, 12, 3-28.

Sala-I-Martin, X. (1995). Transfers, Social Safety Nets, and Economic Growth. Universitat Pompeu Fabra Economics Working Paper 139.

Sala-i-Martin, X., Doppelhofer, G. and Miller, R. I. (2004). Determinants of Long-term Growth: A Bayesian Averaging of Classical Estimates (BACE) approach, American Economic Review, 94(4), 813-835.

Sattar, Z. (1993). Public Expenditure and Economic Performance: A Comparison of Developed and Low-Income Developing Economies. Journal of International Development, 5(1), 27-49.

Saunders, P. (1985). Public Expenditure and Economic Performance in OECD Countries. Journal of Public Policy, 5(1), 1-21.

Saunders, P. (1986). What can we learn from international comparisons of public sector size and economic performance? European Sociological Review, 2(1), 52-60.

Saunders, P. (1988). Private Sector Shrinkage and the Growth of Industrialized Economies: Comment. Public Choice, 58(3), 277-284.

Scully, G. W. (1989). The Size of the State, Economic Growth and the Efficient Utilization of National Resources. Public Choice, 63(2), 149-164.

Sheehey, E. (1993). The Effect of Government Size on Economic Growth. Eastern Economic Journal, 19(3), 321-328

Solow, R. (1956). A contribution to the theory of economic growth. Quarterly Journal of Economics, 70(1), 65–94.

Stanley, T. (2008). Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection. Oxford Bulletin of Economics and Statistics, 70(2), 103-127.

Stanley, T., & Doucouliagos, H. (2007). Identifying and correcting publication selection bias in the efficiency-wage literature: Heckman meta-regression Economics Series 2007, 11. Deakin University.

Stanley, T., & Doucouliagos, H. (2012). Meta-Regression Analysis in Economics and Business. New York: Routledge.

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. Research Synthesis Methods, 5(1), 60-78.

Stanley, T. D., Doucouliagos, H., Giles, M., Heckemeyer, J. H., Johnston, R. J., Laroche, P., . . . Rost, K. (2013). Meta-Analysis of Economics Research Reporting Guidelines. Journal of Economic Surveys, 27(2), 390-394.

Stroup, M., & Heckelman, J. (2001). Size Of The Military Sector And Economic Growth: A Panel Data Analysis Of Africa And Latin America. Journal of Applied Economics, 4(2), 329-360.

Swan, T. W. (1956). Economic Growth and Capital Accumulation. Economic Record, 32(2), 334-361.

Tabellini, G. (2010). Culture and Institutions: Economic Development in the Regions of Europe. Journal of the European Economic Association, 8(4), 677-716.

Tanninen, H. (1999). Income inequality, government expenditures and growth. Applied Economics, 31(9), 1109-1117.

Ugur, M. (2014). Corruption's Direct Effects on Per-Capita Income Growth: A Meta-Analysis. Journal of Economic Surveys, 28(3), 472-490.

Wagner, A. (1877). Finanzwissenschaft. Leipzig: C.F. Winter.

Yan, C., & Gong, L. (2009). Government expenditure, taxation and long-run growth. Frontiers of Economics in China, 4(4), 505-525.

Zhang, J., & Casagrande, R. (1998). Fertility, growth, and flat-rate taxation for education subsidies. Economics Letters, 60(2), 209-216.

Ziliak, S. T., & McCloskey, D. N. (2004). Size matters: the standard error of regressions in the American Economic Review. The Journal of Socio-Economics, 33(5), 527-546.

Table 1A: Fixed-effect weighted means (FEWMs) - Total government expenditure and growth

Study	Number of estimates	FEWM	*Coeff. of variation	Significant	Conf. interval
Adam and Bevan (2005)	9	0.2256	0.1128	Yes	(0.2060, 0.2452)
Afonso and Furceri (2010)	10	-0.3092	0.2295	Yes	(-0.3600, -0.2585)
Afonso and Jalles (2014)	12	-0.1651	0.5900	Yes	(-0.2270, -0.1032)
Afonso and Jalles (2013)	13	-0.1292	0.5945	Yes	(-0.1756, -0.0828)
Afonso et al. (2010)	32	0.0411	0.3841	Yes	(0.0354, 0.0468)
Agell et al. (1997)	3	-0.0828	1.7525	No	(-0.4430, 0.2775)
Angelopoulos et al. (2007)	2	-0.2819	0.3554	No	(-1.1819, 0.6181)
Angelopoulos et al. (2008)	18	-0.2245	1.0092	Yes	(-0.3372, -0.1119)
Arin (2004)	20	-0.2822	0.3048	Yes	(-0.3224, -0.2419)
Bergh and Karlsson (2010)	9	-0.2652	0.3208	Yes	(-0.3306, -0.1998)
Bergh and Öhrn (2011)	9	0.0082	2.6514	No	(-0.1338, 0.1501)
Bernhard (2001)	2	-0.3843	0.1833	No	(-1.0173, 0.2486)
Bojanic (2013)	14	0.3319	0.6608	Yes	(0.2053, 0.4586)
Bose et al. (2007)	2	0.4339	0.0182	Yes	(0.3630, 0.5047)
Butkiewicz and Yanikkaya (2011)	30	-0.1341	0.8184	Yes	(-0.1750, -0.0931)
Chen and Lee (2005)	9	-0.0975	3.7851	No	(-0.3812, 0.1862)
Colombier (2009)	4	0.1499	1.7667	No	(-0.2715, 0.5712)
Cooray (2009)	10	0.1095	0.2122	Yes	(0.0929, 0.1262)
Dalic (2013)	4	-0.2711	0.0932	Yes	(-0.3113, -0.2309)
Dar and AmirKhalkhali (2002)	3	-0.1519	0.7874	No	(-0.4490, 0.1452)
Devarajan et al. (1996)	16	0.0447	1.4274	Yes	(0.0107, 0.0786)
Diamond (1998)	2	0.0394	1.0397	No	(-0.3286, 0.4074)
Engen and Skinner (1992)	6	-0.3843	0.8564	Yes	(-0.7297, -0.0389)
Fölster and Henrekson (1999)	7	-0.4693	0.2372	Yes	(-0.5722, -0.3663)
Fölster and Henrekson (2001)	8	-0.3579	0.4022	Yes	(-0.4783, -0.2376)
Ghali (2003)	2	0.4332	0.0947	Yes	(0.0645, 0.8020)
Ghosh and Gregoriou (2008)	36	0.1896	0.2403	Yes	(0.1742, 0.2050)
Grimes (2003)	5	-0.4706	0.4347	Yes	(-0.7247, -0.2165)
Hamdi and Sbia (2013)	3	0.1629	2.7945	No	(-0.9682, 1.2941)
Hansen (1994)	1	-0.2133	n.a.	n.a.	n.a.
Husnain and Ghani (2010)	6	-0.1872	0.3127	Yes	(-0.2486, -0.1258)
Kalaitzidakis and Tzouvelekas (2011)	1	0.1004	n.a.	n.a.	n.a.
Kelly (1997)	4	-0.1877	0.5738	Yes	(-0.3592, -0.0163)
Lee and Lin (1994)	8	-0.2569	0.2145	Yes	(-0.3030, -0.2108)
Levine and Renelt (1992)	3	-0.1931	0.5896	No	(-0.4758, 0.0897)
Marlow (1986)	6	-0.5519	0.4461	Yes	(-0.8102, -0.2935)
Martin and Fardmanesh (1990)	12	0.0361	1.9956	No	(-0.0097, 0.0820)
Mendoza et al. (1997)	3	-0.0059	11.7364	No	(-0.1789, 0.1670)
Miller and Russek (1997)	6	-0.1767	0.5151	Yes	(-0.2721, -0.0812)
Nketiah-Amponsah (2009)	1	-0.3985	n.a.	n.a.	n.a.
Odedokun (1997)	1	-0.0267	n.a.	n.a.	n.a.
Plümper and Martin (2003)	2	-0.1319	0.3782	No	(-1.1055, 0.8417)
Ram (1986)	8	-0.2074	0.6273	Yes	(-0.3243, -0.0905)
Romer (1989)	3	-0.3026	0.0105	Yes	(-0.5869, -0.0183)

Romero-Avila and Strauch (2008)	3	-0.1534	3.7126	No	(-0.3934, 0.0858)
Sala-I-Martin (1995)	2	-0.3420	0.3523	Yes	(-0.3743, -0.3096)
Sattar (1993)	9	0.0047	3.7126	No	(-0.0087, 0.0181)
Saunders (1985)	2	-0.6847	0.3523	No	(-2.8519, 1.4825)
Saunders (1988)	12	-0.5150	0.5613	Yes	(-0.6987, -0.3313)
Scully (1989)	4	0.2639	0.0969	Yes	(0.2232, 0.3046)
Stroup and Heckelman (2001)	5	-0.1561	0.9438	No	(-0.3391, 0.0268)
Tanninen (1999)	1	-0.0360	n.a.	n.a.	n.a.
Yan and Gong (2009)	8	0.0594	2.7707	No	(-0.0782, 0.1970)
Government expenditure overall	**411**	**-0.0083**	**9.1092**	**No**	**(-0.0238, 0.0071)**

*Absolute values reported

Table 1B: Fixed-effect weighted means (FEWMs) - Government consumption and growth

Study	Number of estimates	FEWM	*Coeff. of variation	Significant	Conf. interval
Afonso and Furceri (2010)	4	-0.3023	0.3793	Yes	(-0.4847, -0.1199)
Afonso and Jalles (2014)	18	-0.0742	2.4684	No	(-0.1652, 0.0169)
Afonso and Jalles (2013)	8	-0.1326	0.5046	Yes	(-0.1886, -0.0767)
Andrés et al. (1996)	2	-0.0388	0.3888	No	(-0.1745, 0.0968)
Angelopoulos and Philippopoulos (2007)	6	-0.3752	0.5452	Yes	(-0.5899, -0.1605)
Angelopoulos et al. (2008)	18	-0.1868	0.3702	Yes	(-0.2211, -0.1524)
Barro and Sala-i-Martin (1995)	24	-0.3670	0.2985	Yes	(-0.4133, -0.3208)
Barro (1989)	5	-0.4445	0.1340	Yes	(-0.5185, -0.3705)
Barro (1991)	20	-0.4226	0.1346	Yes	(-0.4492, -0.3960)
Barro (1996)	8	-0.2810	0.0827	Yes	(-0.3004, -0.2615)
Barro (2001)	1	-0.6490	n.a.	n.a.	n.a.
Bellettini and Ceroni (2000)	24	-0.2127	0.6031	Yes	(-0.2669, -0.1585)
Bernhard (2001)	1	-0.2551	n.a.	n.a.	n.a.
Brumm (1997)	1	-0.1385	n.a.	n.a.	n.a.
Butkiewicz and Yanikkaya (2011)	29	-0.1069	0.7060	Yes	(-0.1356, -0.0782)
Castro (2011)	12	-0.3450	0.2771	Yes	(-0.4058, -0.2843)
Commander et al. (1999)	9	-0.2173	0.3288	Yes	(-0.2722, -0.1624)
Cooray (2009)	5	0.0166	1.3136	No	(-0.0105, 0.0436)
Cronovich (1998)	4	0.1820	0.8977	No	(-0.0780, 0.4420)
De Gregorio (1992)	5	-0.1562	0.8494	No	(-0.3209, 0.0085)
Dowrick (1996)	11	-0.0782	0.7209	Yes	(-0.1160, -0.0403)
Easterly and Rebelo (1993)	3	-0.0429	0.3829	Yes	(-0.0837, -0.0021)
Fölster and Henrekson (2001)	2	-0.3816	0.2618	No	(-1.2790, 0.5159)
Garrison and Lee (1995)	4	0.0129	1.6873	No	(-0.0217, 0.0475)
Ghura (1995)	6	-0.1737	0.0863	Yes	(-0.1894, -0.1580)
Grier and Tullock (1989)	10	-0.2261	0.9932	Yes	(-0.3867, -0.0655)

Grossman (1990)	16	0.0583	1.2122	Yes	(0.0207, 0.0960)
Guseh (1997)	8	-0.0692	1.5077	No	(-0.1565, 0.0180)
Hansson and Henrekson (1994)	6	-0.1967	0.6395	Yes	(-0.3288, -0.0647)
Landau (1983)	14	-0.2222	0.6145	Yes	(-0.3010, -0.1433)
Landau (1986)	12	-0.1025	0.6618	Yes	(-0.1456, -0.0594)
Landau (1997)	8	-0.0311	1.3723	No	(-0.0668, 0.0046)
Lee (1995)	4	-0.3022	0.1994	Yes	(-0.3981, -0.2063)
Levine and Renelt (1992)	10	-0.2199	0.6508	Yes	(-0.3224, -0.1176)
Mo (2007)	10	-0.4806	0.1761	Yes	(-0.5411, -0.4200)
Murphy et al. (1991)	2	-0.3039	0.4310	No	(-1.4809, 0.8730)
Neycheva (2010)	13	-0.0206	5.9556	No	(-0.0947, 0.0535)
Romero-Avila and Strauch (2008)	5	-0.1177	0.6042	Yes	(-0.2060, -0.0294)
Roubini and Sala-i-Martin (1992)	20	-0.4523	0.1173	Yes	(-0.4771, -0.4274)
Sala-i-Martin (1995)	1	-0.3117	n.a.	n.a.	n.a.
Saunders (1986)	3	-0.6488	0.2878	Yes	(-1.1127, -0.1849)
Sheehey (1993)	6	0.1093	2.5298	No	(-0.1809, 0.3994)
Tanninen (1999)	3	-0.1855	2.8551	No	(-1.5008, 1.1299)
Zhang and Casagrande (1998)	2	-0.4291	0.0306	Yes	(-0.5470, -0.3111)
d’Agostino et al. (2010)	2	-0.1173	0.6149	No	(-0.7651, 0.5306)
d'Agostino et al. (2012)	3	-0.1833	0.1279	Yes	(-0.2416, -0.1251)
Government consumption overall	**388**	**-0.1204**	**1.2846**	**Yes**	**(-0.1359, -0.1049)**

*Absolute values reported

Table 2A Total government expenditure and growth

PET-FAT and PEESE Results

	A. PET-FAT			**B. PEESE**
	(1)	(2)	(3)	(4)
VARIABLES	Entire Dataset	Developed	LDCs	Developed
Precision (β_0)	-0.0317	-0.1311***	-0.0700	-0.1397***
	(0.0193)	(0.0459)	(0.0467)	(0.0316)
Bias (α_0)	-0.5963	0.0275	1.0715	
	(0.4042)	(0.7804)	(0.7519)	
Std. Error				4.9584
				(3.3918)
Observations	411	165	139	165
No of studies in cluster	53	28	22	28

Robust standard errors (in brackets) are clustered at the study level. *, **, *** indicate significance at 10%, 5% and 1%, respectively.

Panel A reports PET/FAT results; and Panel B reports results that take account of quadratic relationship between effect size and its standard error (PEESE).

PEESE results are reported only when PET/FAT results indicate significant effect after controlling for selection bias.

Table 2B Government consumption and growth

PET-FAT and PEESE Results

	A. PET-FAT			**B. PESEE**	
	(1)	(2)	(3)	(4)	(5)
VARIABLES	Entire Dataset	Developed	LDCs	Entire Dataset	Developed
Precision (β_0)	-0.0474***	-0.0862**	-0.0091	-0.0996***	-0.1397***
	(0.0182)	(0.0403)	(0.0320)	(0.0141)	(0.0260)
Bias (α_0)	-1.5525***	-1.1544*	-1.4529**		
	(0.3595)	(0.6206)	(0.7231)		
Std. error				-2.7107	-2.3687
				(2.0915)	(3.0106)
Observations	388	105	70	388	105
No of studies in cluster	46	19	14	46	19

Robust standard errors (in brackets) are clustered at the study level. *, **, *** indicate significance at 10%, 5% and 1%, respectively.

Panel A reports PET/FAT results; and Panel B reports results that take account of quadratic relationship between effect size and its standard error (PEESE).

PEESE results are reported only when PET/FAT results indicate significant effect after controlling for selection bias.

Table 3A Summary Statistics – Total Expenditure

Variables	Definition	N	Mean	S.D.	Min	Max
t-value	t-statistics reported in primary studies	411	-0.66	2.79	-12.17	6.33
Precision	Inverse of standard error of the partial correlation coefficient	411	15.07	9.99	3.16	51.03
SE_{ri}	Standard errors of the partial correlation coefficients	411	0.09	0.05	0.02	0.32
Developed	Takes value 1 if the primary study data is from developed countries, otherwise 0	411	5.09	7.39	0	29.36
LDCs	Takes value 1 if the primary study data is from LDCs, otherwise 0	411	4.26	6.71	0	33.33
Time Series	Takes value 1 if Time Series is used by primary study, otherwise 0	411	0.55	2.01	0	10.84
Cross-section	Takes value 1 if cross-section data is used by primary study, 0 if panel is used	411	0.64	2.13	0	11.04
Panel Data	Takes value 1 if panel data is used by primary study, otherwise 0	411	13.88	11.17	0	51.03
Control for Endogeneity	Takes value 1 if primary study controls for endogeneity, otherwise 0	411	2.22	5.35	0	22.41
Endogenous Growth Model	Takes value 1 if the model is based on endogenous growth model, otherwise 0.	411	0.06	0.75	0	10.11
Data Average (=>5)	Takes value 1 if data averaging period is =>5 years otherwise 0	411	6.05	7.03	0	23.36
Data Average*Panel Data	Takes value 1 if study used panel data and averaging period is =>5 years otherwise 0	411	6.94	7.23	0	27.50
Data Average*Cross Section	Takes value 1 if study used cross section and averaging period is =>5 years otherwise 0	411	0.64	2.13	0	11.04
Data Period (1960+)	Takes value 1 if data year>= 1960, otherwise 0	411	12.97	11.51	0	51.03
Data Period (1970+)	Takes value 1 if data year>= 1970, otherwise 0	411	8.38	12.48	0	51.03
Data Period (1980+)	Takes value 1 if data year>= 1980, otherwise 0	411	11.19	7.43	0	44.44
Data Period (1990+)	Takes value 1 if data year>= 1990, otherwise 0	411	13.29	10.98	0	51.03
Data Period (2000+)	Takes value 1 if data year>= 2000, otherwise 0	411	14.19	10.45	0	51.03
Initial GDP	Takes value 1 if the primary study control for initial per capita GDP, otherwise 0	411	5.49	7.02	0	26.26
Population	Takes value 1 if the primary study control for population, otherwise 0	411	1.84	4.10	0	18.71
Investment	Takes value 1 if the primary study control for investment, otherwise 0	411	5.98	8.04	0	44.44
Tax	Takes value 1 if the primary study control for taxes, otherwise 0	411	3.99	7.01	0	27.50
Journal Rank	Takes value 1 if the primary study is published in high-ranked journal, otherwise 0	411	6.79	7.32	0	29.69
Journal	Takes value 1 if the primary study is published in a journal, otherwise 0	411	10.48	7.71	0	44.44
Publication Year (1990+)	Takes value 1 if publication year>=1990, otherwise 0	411	3.05	7.29	0	44.44
Publication Year (2000+)	Takes value 1 if publication year>=2000, otherwise 0	411	4.67	6.79	0	29.69

Publication Year (2010+)	Takes value 1 if publication year>=2010, otherwise 0	411	6.79	12.09	0	51.03

Notes: weighted variables are divided by SE_{ri}

Table 3B Summary Statistics – Government Consumption

Variables	Definition	N	Mean	S.D.	Min	Max
t-value	t-statistics reported in primary studies	388	-2.24	2.13	-10.32	3.53
Precision	Inverse of standard error of the partial correlation coefficient	388	13.49	8.17	3.25	43.49
SE_{ri}	Standard errors of the partial correlation coefficients	388	0.10	0.05	0.02	0.31
Developed	Takes value 1 if the primary study data is from developed countries, otherwise 0	388	3.33	6.34	0	26.94
LDCs	Takes value 1 if the primary study data is from LDCs, otherwise 0	388	3.54	8.68	0	39.73
Time Series	Takes value 1 if Time Series is used by primary study, otherwise 0	388	0.08	0.67	0	6.48
Cross-section	Takes value 1 if cross-section data is used by primary study, 0 if panel is used	388	2.91	4.87	0	24.89
Panel Data	Takes value 1 if panel data is used by primary study, otherwise 0	388	10.49	10.30	0	43.49
Control for Endogeneity	Takes value 1 if primary study controls for endogeneity, otherwise 0	388	1.99	5.03	0	26.45
Endogenous Growth Model	Takes value 1 if the model is based on endogenous growth model, otherwise 0.	388	0.26	1.59	0	11.01
Data Average (=>5)	Takes value 1 if data averaging period is =>5 years otherwise 0	388	7.28	7.47	0	27.16
Data Average*Panel Data	Takes value 1 if study used panel data and averaging period is =>5 years otherwise 0	388	5.19	7.57	0	27.16
Data Average*Cross Section	Takes value 1 if study used cross section and averaging period is =>5 years otherwise 0	388	2.09	4.49	0	24.89
Data Period (1960+)	Takes value 1 if data year>= 1960, otherwise 0	388	7.49	9.15	0	43.49
Data Period (1970+)	Takes value 1 if data year>= 1970, otherwise 0	388	9.12	8.91	0	42.60
Data Period (1980+)	Takes value 1 if data year>= 1980, otherwise 0	388	12.80	8.52	0	12.51
Data Period (1990+)	Takes value 1 if data year>= 1990, otherwise 0	388	12.11	9.02	0	25.26
Data Period (2000+)	Takes value 1 if data year>= 2000, otherwise 0	388	0.14	1.23	0	11.01
Initial GDP	Takes value 1 if the primary study control for initial per capita GDP, otherwise 0	388	8.48	8.32	0	43.49
Population	Takes value 1 if the primary study control for population, otherwise 0	388	5.09	7.59	0	26.94

Investment	Takes value 1 if the primary study control for investment, otherwise 0	388	7.25	7.95	0	42.60
Tax	Takes value 1 if the primary study control for taxes, otherwise 0	388	2.49	7.62	0	43.49
Journal Rank	Takes value 1 if the primary study is published in high-ranked journal, otherwise 0	388	8.68	9.88	0	43.49
Journal	Takes value 1 if the primary study is published in a journal, otherwise 0	388	11.96	9.27	0	43.49
Publication Year (1990+)	Takes value 1 if publication year>=1990, otherwise 0	388	6.89	8.38	0	34.36
Publication Year (2000+)	Takes value 1 if publication year>=2000, otherwise 0	388	11.79	9.36	0	43.49
Publication Year (2010+)	Takes value 1 if publication year>=2010, otherwise 0	388	3.16	6.56	0	27.16

Notes: weighted variables are divided by SE_{ri}

Table 4A – MRA (Total Government Expenditure and Growth)

	(1)	(2)
VARIABLES	General Model	General-to-Specific Model
Precision	-0.0049	0.0397
	(0.1450)	(0.0880)
Theoretical and econometric dimensions		
Control for Endogeneity	0.0671***	0.0649***
	(0.0205)	(0.0203)
Cross Section	-0.0828	
	(0.1182)	
Endogenous Growth Model	0.3139	
	(0.2669)	
Data Average (=>5 years)	0.0005	
	(0.0812)	
Data Average*Panel Data	-0.0192	
	(0.0778)	
Population	-0.1733**	-0.1739***
	(0.0683)	(0.0605)
Initial GDP	-0.1684***	-0.1733***
	(0.0405)	(0.0375)
Investment	-0.0640*	-0.0574*
	(0.0366)	(0.0347)
Data Characteristics		
Data Period (1960+)	0.0257	
	(0.0319)	
Data Period (1970+)	0.0215	
	(0.0501)	
Data Period (1980+)	-0.0327	
	(0.0673)	
Data Period (1990+)	0.0055	
	(0.0613)	
Data Period (2000+)	-0.1271	-0.1232*
	(0.0915)	(0.0698)
Developed	-0.0280*	-0.0312**
	(0.0170)	(0.0159)
Publication Characteristics		
Journal Rank	0.0319	
	(0.0471)	

Journal	0.1963***	0.1894***
	(0.0635)	(0.0483)
Publication Year (1990+)	-0.0749	-0.0921*
	(0.0804)	(0.0540)
Publication Year (2000+)	-0.0604	-0.0839*
	(0.0704)	(0.0430)
Publication Year (2010+)	0.1525**	0.1682***
	(0.0752)	(0.0451)
Constant	0.8082*	0.7136
	(0.4646)	(0.4440)
Observations	411	411
Number of studies	53	53

Standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table 4B – MRA (Government Consumption and Growth)

	(1)	(2)
VARIABLES	General Model	General-to-Specific Model
Precision	0.0730	0.1776
	(0.1510)	(0.1256)
Theoretical and econometric dimensions		
Control for Endogeneity	0.0258	
	(0.0243)	
Cross Section	-0.2738***	-0.2328***
	(0.0628)	(0.0579)
Endogenous Growth Model	-0.1378**	-0.1226**
	(0.0609)	(0.0603)
Data Average (=>5)	-0.0234	
	(0.0304)	
Data Average*Cross-section	0.1320*	0.1029*
	(0.0733)	(0.0634)
Population	-0.0766***	-0.0713***
	(0.0269)	(0.0243)
Initial GDP	0.0255	
	(0.0234)	

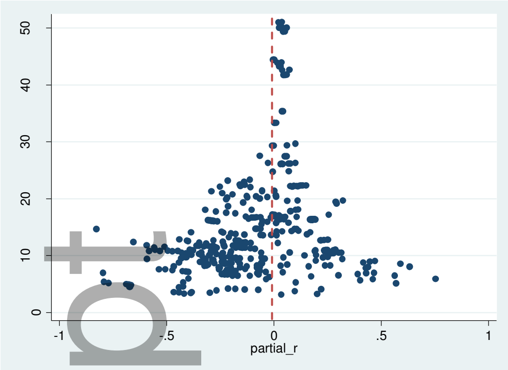
Tax	-0.0372	
	(0.0227)	
Investment	0.0935***	0.0908***
	(0.0199)	(0.0194)
Data Characteristics		
Data Period (1960+)	-0.0685	-0.0837**
	(0.0469)	(0.0425)
Data Period (1970+)	-0.0982**	-0.1216***
	(0.0462)	(0.0402)
Data Period (1980+)	-0.1212**	-0.1487***
	(0.0530)	(0.0475)
Data Period (1990+)	-0.2304***	-0.2760***
	(0.0683)	(0.0501)
Developed	-0.0137	
	(0.0180)	
Publication Characteristics		
Journal Rank	-0.0761***	-0.0866***
	(0.0288)	(0.0265)
Journal	0.2114***	0.2040***
	(0.0441)	(0.0424)
Publication Year (1990+)	0.0530**	0.0408*
	(0.0241)	(0.0231)
Publication Year (2000+)	0.1911***	0.1921***
	(0.0386)	(0.0360)
Publication Year (2010+)	-0.1667***	-0.1822***
	(0.0359)	(0.0312)
Constant	-1.0983***	-1.0963***
	(0.2867)	(0.2747)
Observations	388	388
Number of studies	46	46

Standard errors in parentheses.

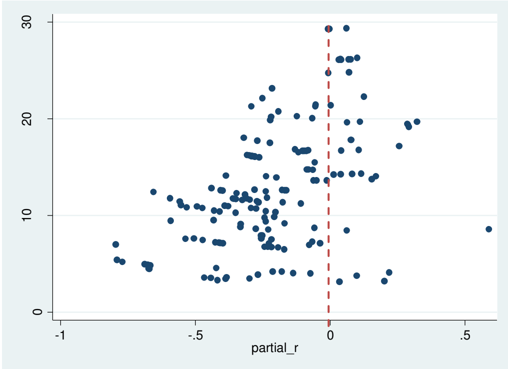
*** p<0.01, ** p<0.05, * p<0.1

Figures 1 – 3 (Total Government Expenditure and Growth)

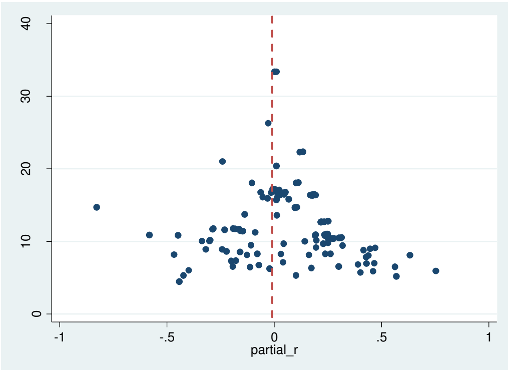
Entire Dataset (1)



Developed Countries Sample (2)

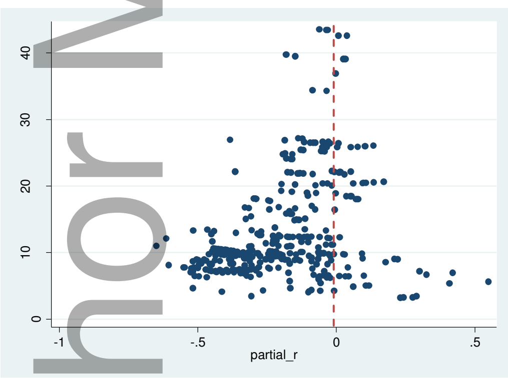


LDCs Sample (3)

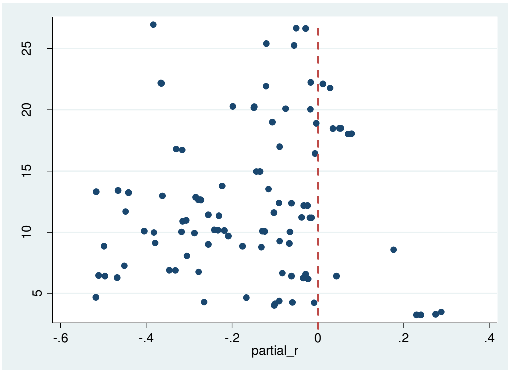


Figures 4 – 6 (Government Consumption and Growth)

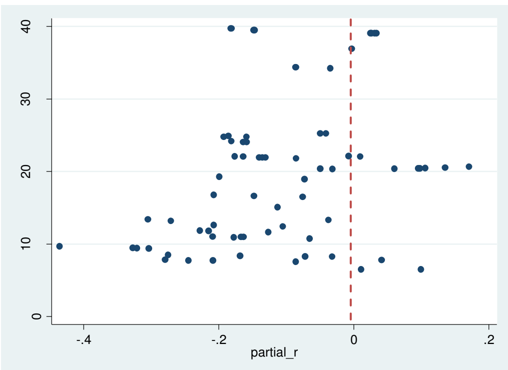
Entire Dataset (4)



Developed Countries Sample (5)



LDCs Sample (6)



Appendix – Overview of Methods

1. Partial Correlation Coefficients (PCCs)

PCCs measure the association between government expenditure and per-capita GDP growth. Given that they are independent of the metrics used in measuring both independent and dependent variables, they allow for the comparability of studies and reported effect-size estimates. They are mostly used in meta-analysis (see e.g. Alptekin and Levine, 2012; Ugur; 2014; Benos and Zotou, 2014).

We use equations (A1) and (A2) to calculate a PCC and standard error, respectively, for each relevant effect-size estimate reported by primary studies.

$$r_i = \frac{t_i}{\sqrt{t_i^2 + df_i}} \tag{A1}$$

and

$$SE_{ri} = \sqrt{\frac{1 - r_i^2}{df_i}} \tag{A2}$$

r_i and SE_{ri} represent PCC and its associated standard errors, respectively. SE_{ri} represents variations due to sampling error and its inverse is used as weight in the calculation of study-by-study fixed-effect weighted averages. t_i and df_i represent t-value and degrees of freedom, respectively, associated with estimates reported in primary studies.

2. Fixed Effect Weighted Means

We calculate FEEs using (A3) below.

$$\bar{X}_{FEE} = \frac{\sum r_i \left(\frac{1}{SE_{ri}^2}\right)}{\sum \frac{1}{SE_{ri}^2}} \tag{A3}$$

$\bar{X}_{FEE}$ is the fixed effect weighted average and all other variables remain as explained before. FEEs account for within-study variations by assigning higher weights to more precise estimates and lower weights to less precise estimates.

3. Bivariate meta-regressions

To estimate 'genuine effect' beyond publication selection bias, we draw on meta-regression analysis (MRA) models proposed and developed by Stanley (2008) and Stanley and Doucouliagos (2012, 2014). The underpinning theoretical framework is that of Egger et al. (1997), who postulate that researchers with small samples and large standard errors would search intensely across model specifications, econometric techniques and data measures to find sufficiently large (hence statistically significant) effect-size estimates. Hence:

$$e_i = \beta + \alpha SE_i + u_i \tag{A4}$$

Here, e_i is the effect-size reported in primary studies and SE_i is the associated standard error. Rejecting the null hypothesis of $\alpha = 0$ indicates the presence of publication bias. This is also known as the funnel-asymmetry test (FAT), which evaluates the asymmetry of the funnel graphs that chart the effect-size estimates against their precisions. Testing for $\beta = 0$ is known as precision-effect test (PET), and allows for establishing whether genuine effect exists beyond selection bias.

However, estimating (A4) poses several issues. First, the model is heteroskedastic: effect-size estimates have widely different standard errors (hence variances), violating the assumption of independently and identically distributed (i.i.d.) error term (u_i). To address this issue, Stanley (2008) and Stanley and Doucouliagos (2012) propose a weighted least squares (WLS) version, obtained by dividing both sides of (A4) with precision ($1/SE_i$).

Secondly, primary-study estimates may be affected by data dependence, which arises when primary studies using a particular dataset report multiple estimates or when different studies use overlapping segments of the country data compiled by national statistical agencies (Doucouliagos and Laroche, 2009). Clustered data analysis (CDA), an approach often used in the meta-analysis literature, only corrects the standard errors for within-study dependence. However, hierarchical linear models (HLMs) allow for robust standard errors clustered on studies and take account of both within-study and between-study dependence explicitly. We model data dependence by allowing for random variation between study-specific estimates, which may be due to study-specific intercepts and/or study-specific slopes (Demidenko, 2004; McCulloch et al., 2008). Stated differently, estimates reported by primary studies are nested within each study; and the estimates are modelled to differ between studies either because they share a common intercept (a fixed component) and/or a common slope within each primary study. The HLM can be stated as follows:

$$t_{ij} = \alpha_0 + \beta_0 \left(\frac{1}{SE_{rij}}\right) + v_j + \varepsilon_{ij} \tag{A5}$$

where t_{ij} is the t-value associated with effect-size estimate i (i.e., the partial correlation coefficient calculated using A1) of study j; SE_{rij} is the corresponding standard calculated in accordance with (A2); v_j is the study-level random effect; and ε_{ij} is the multivariate-normal error term with mean zero. The random effects (v_j) are not estimated directly, but their variance (or standard error) is. We conclude in favour of publication selection bias if α_0 is statistically significant at conventional levels. In the presence of publication bias, α_0 determines the magnitude and the direction of bias. Similarly, we conclude in favour of genuine effect beyond selection bias if β_0 is statistically significant at conventional levels.

The third issue is that Egger et al. (1997) assume a linear relationship between primary-study estimates and their standard errors. However, Moreno et al. (2009) and Stanley and Doucouliagos (2014) provide simulation evidence indicating that a quadratic specification is superior if 'genuine effect' exists beyond selection bias – i.e., if the PET in (A5) rejects the null hypothesis of zero effect. Then, the correct specification is referred to as precision-effect test corrected for standard errors (PEESE) and can be stated as follows:

$$t_{ij} = \beta_0 \left(\frac{1}{SE_{rij}}\right) + \alpha_0 (SE_{rij}) + v_j + \varepsilon_{ij} \qquad \text{(A6)}$$

Given that study-level random effects may be observed at the intercept or slope levels or both, we establish which type of HML is appropriate using LR tests, where the null hypothesis is that the preferred specification is nested within the comparator specification. Therefore, we estimate HLMs with random-intercepts only and HLMs with random intercepts and random slopes; and test whether the latter are nested within the former.

4. Multivariate Meta-Regression Model (MRM)

To address the issues of heterogeneity, we estimate a multivariate hierarchical meta-regression model specified in (A7) below.

$$t_{ji} = \alpha_0 + \beta_0 \left(\frac{1}{SE_{jri}}\right) + \sum \beta_k \frac{(Z_{ki})}{SE_{jri}} + v_j + \varepsilon_{ij} \qquad \text{(A7)}$$

Here, t_{ji} is the ith t-value from the jth study, while Z_{ki} is a $k \times 1$ vector of moderator variables that capture the observable sources of heterogeneity in the government size-growth evidence base.

To minimise the risk of multicollinearity and over-fitting, we estimate (A7) through a general-to-specific estimation routine, whereby we omit the most insignificant variables (variables associated with the largest p-values) one at a time until all remaining covariates are statistically significant. We present the findings from the specific and general models side by side to: (a) establish the extent of congruence between the significant moderating factors; and (ii) identify the range of moderating variables that do not affect the variation in the evidence base.