Does the Composition of Government Expenditure Matter for Long-run GDP Levels?

Norman Gemmell, Richard Kneller and Ismael Sanz

WORKING PAPER 10/2014
August 2014

Working Papers in Public Finance

VICTORIA UNIVERSITY OF WELLINGTON
TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI

Chair in Public Finance
Victoria Business School

The Working Papers in Public Finance series is published by the Victoria Business School to disseminate initial research on public finance topics, from economists, accountants, finance, law and tax specialists, to a wider audience. Any opinions and views expressed in these papers are those of the author(s). They should not be attributed to Victoria University of Wellington or the sponsors of the Chair in Public Finance.

Further enquiries to:
The Administrator
Chair in Public Finance
Victoria University of Wellington
PO Box 600
Wellington 6041
New Zealand

Phone: +64-4-463-9656
Email: cpf-info@vuw.ac.nz

Papers in the series can be downloaded from the following website:
http://www.victoria.ac.nz/cpf/working-papers

Does the composition of government expenditure matter for long-run GDP levels?

NORMAN GEMMELL[‡], RICHARD KNELLER[†] AND ISMAEL SANZ[**]

[‡]*Victoria Business School, Victoria University of Wellington, New Zealand*

(Corresponding author: e-mail: norman.gemmell@vuw.ac.nz)

[†] *School of Economics, University of Nottingham, UK*

(e-mail: richard.kneller@nottingham.ac.uk)

*** Universidad Rey Juan Carlos. Madrid, Spain.*

(e-mail: ismael.sanz@urjc.es)

Abstract

We examine the long-run GDP impacts of changes in *total* government expenditure and in the *shares* of different spending categories for a sample of OECD countries since the 1970s, taking account of methods of financing expenditure changes and possible endogenous relationships. We provide more systematic empirical evidence than available hitherto for OECD countries. Our results provide strong evidence that reallocating total spending towards infrastructure and education would be positive for long-run income levels. Increasing the share of social welfare spending (and away from all others pro-rata) may be associated with, at most, modestly lower long-run GDP levels.

JEL codes: H50; O40

Keywords: government expenditure composition, fiscal policy, GDP

This version: March 2015

I. Introduction

The fiscal stimulus packages contemplated or enacted in many countries from 2009 onwards to combat the global economic crisis typically included significant expansions in various public expenditure programs. Much debate surrounded the merits of these fiscal packages, not least with respect to the wisdom of attempts, and governments' ability, to stimulate the economy via expenditure injections. The specific public spending choices in these short-term fiscal packages were motivated in part by a desire to be consistent with these governments' long-term growth objectives, such as expanding infrastructure spending which was perceived as benefiting long-run GDP levels or growth rates. This brings to the fore, the twin questions of how strong is the evidence-base that long-run income levels or growth rates respond to public expenditure changes, and which expenditure categories have greatest impact?

This paper focuses on these two questions; it does not deal with whether there are short-run benefits from such stimulus packages, which requires quite different analytics and empirical methodologies. We first briefly review the relevant theory. This builds on Barro (1990) and Devarajan *et al.* (1996) who proposed that so-called 'productive' public expenditures can influence long-run growth rates via impacts on private sector production functions. Following a brief discussion of the existing evidence on the long-run public expenditure-output relationships, we provide more systematic empirical evidence than available hitherto for OECD countries.

There are a number of reasons to re-examine this issue. Firstly, most empirical studies focus on a subset of individual expenditure categories, such as defence, education or transport and communication expenditures. These have produced a mixed set of findings and do not typically allow the trade-offs between different forms of public expenditure to be identified explicitly. Our approach provides evidence on these trade-offs.

Secondly, a number of recent papers have used alternative methodologies to examine long-run impacts of *tax* policy on GDP levels or growth rates (see Adam and Bevan, 2005; Lee and Gordon, 2005; Romero-Avila and Strauch, 2008; Arnold *et al*, 2011; Gemmell *et al.*, 2011, 2014). However, *public expenditure* is rarely the primary focus of attention in these studies, or it is tested as an aggregate 'productive' spending category. In this paper we limit attention to potential effects of public expenditure on long-run GDP levels, but we focus both on total public expenditure (suitably financed via the government budget constraint) and on how long-run GDP levels may be differentially affected by specific functional spending categories.

Thirdly, difficulties dealing with endogeneity associated with estimates of fiscal impacts of on GDP, have afflicted most previous studies. While we do not claim to have resolved these concerns in this paper, we do carefully address potential endogeneity problems. Our estimated fiscal-GDP parameters would appear to be at least ‘weakly exogenous’ (based on standard econometric definitions, see Johansen, 1992, and Boswijk, 1995).

Fourthly, increased use of panel methods and datasets has increased the reliability of recent studies. However, the generally short time-series dimensions of these panels have led to fixed effects estimators, which impose parameter homogeneity, typically being used. Furthermore, as Haque’s (2004) re-examination of the Devarajan *et al* (1996) results shows, how the time-series properties of the data are handled can be crucial for estimated output effects of fiscal variables.

In this paper, access to an extended panel dataset with a longer time dimension than previous studies, permits application of the more flexible Pooled Mean Group (PMG) estimator proposed by Pesaran *et al.* (1999). This enables us to explore both short-run dynamics and long-run equilibrium relationships among the variables of interest, account for heterogeneity across countries in their short-run dynamic relationships. We compare these results with those obtained using more restrictive dynamic fixed effects (DFE) methods, and the more flexible, but information-intensive, Mean Group (MG) approach.

These improvements provide more robust evidence on the potential long-run association and causation between public spending and GDP. This supports some traditional views, such as that a spending reallocation towards infrastructure and education spending (and to a lesser extent, health) can raise GDP over the long-run. Our results also suggest that, relative to public spending ‘on average’, social welfare spending may have moderate GDP-reducing effects. That is, there is likely to be a trade-off between public spending aimed at income redistribution (via social welfare spending) and that aimed at raising overall income levels.

The remainder of this paper is structured as follows. Section II briefly describes the links between public expenditure composition and GDP growth hypothesised by recent theory, and summarises current evidence for OECD countries. Section III then discusses our testing methodologies and dataset; while section IV reports results for a sample of 17 OECD countries over 1972-2008. Some conclusions are drawn in section V.

II. Public spending, taxes and growth

Theory

As is well known, in the neoclassical growth model, if the incentives to save or to invest in new capital are affected by fiscal policy, this alters the equilibrium capital-output ratio, and therefore the level of the output path, but not its slope. There are effects on growth only for a transitional period as the economy moves onto its new output path, though the length of this transition remains subject to debate.[1]There is, however, a resulting impact on the long-run *level* of output (GDP). The 1990s saw the development of a number of growth models with a permanent, or at least persistent, role for fiscal policy such as those of Barro (1990), Futagami *et al.* (1993) and Deverajan *et al.* (1996).

A novel feature of these models was that fiscal policy can determine both the level of the output path and the steady-state (long-run) growth rate.[2] Key fiscal policy aspects include that some ‘productive’ public expenditures affect the productivity of the private sector while other ‘unproductive’ expenditures only impact on citizens’ welfare (including the possibility of zero welfare effects), and that some taxes levied to fund public expenditures distort investment decisions. In a model with multiple productive expenditures Devarajan *et al.* (1996) show that the long-run growth effects depend upon a combination of the *relative* productivities of these expenditures and their *relative* budget shares.

As Kneller *et al.* (1999) show, the categorising of public expenditures into ‘productive’ and ‘unproductive’, and taxes into ‘distortionary’ and ‘non-distortionary’ (with respect to investment decisions) yields predicted long-run growth impacts from fiscal policy that depend on these decompositions. Positive, negative or zero effects are each possible depending on the tax/expenditure combinations chosen. When the above models are extended to allow for the growth effects of deficits/surpluses, outcomes are again positive, negative or ambiguous depending on what the deficit is financing; see Adam and Bevan (2005), Gemmell *et al.* (2011).

A number of recent papers have modelled the relationship between particular public spending categories and growth. In a series of papers, Agénor and Neanidis (2006) and Agénor (2008) have examined various extensions of the Barro/Devarajan framework which explicitly model (i) infrastructure, education and/or health spending as inputs into private production; and (ii) interactions between these spending types, for example, by allowing the supply of health services or infrastructure spending to enter into the production function for education.

In a similar vein, Semmler *et al.* (2007) develop an endogenous growth model to consider the output effects of these three types of spending, in which the tax rate is assumed to be chosen optimally. They solve the model numerically and calibrate it to explore the impact of shifts between public investment in infrastructure assets that directly influence market production and public investment devoted to the production of human capital accumulation (education and

[1]On fiscal response dynamics see, for example, Turnovsky (2004) who estimates ‘transitional’ output adjustments to fiscal policy changes in terms of decades. Lee *et al.*’s (1997) analysis of convergence, on the other hand, finds convergence speeds to equilibrium could be as rapid as 2-3 years.
[2] Not all endogenous growth models predict long-run growth effects from fiscal policy. These have sometimes been labelled ‘semi-endogenous’ or ‘non-scale’ growth models; see Eicher and Turnovsky (1999).

health). Blankenau and Simpson (2004) propose a different model of the relationship between public education expenditures and growth. They show that the relationship need not be monotonic when account is taken of tax-financing methods (analogously to Barro) and the specification of the technology of human capital production (e.g. how public and private inputs into human capital formation, and the input of human capital from preceding generations, are combined)[3].

Each of the above models hypothesises mechanisms by which permanent impacts on GDP growth rates can occur in association with changes in particular public expenditure categories, especially those related to infrastructure and human capital production. Of course as is well known, such endogenous growth models depend for their 'permanent' growth effects on the so-called 'knife-edge' properties whereby these permanent growth effects depend on exactly constant returns to scale (Solow, 1994; Dalgaard and Kreiner, 2003). While the conditions for such permanent effects may be restrictive, and hard to verify empirically, they do point to the possibility that output effects from public spending changes could be highly persistent.

Existing empirical evidence on public expenditure and growth

Much of the empirical literature testing for fiscal policy impacts on long-run GDP levels or growth rates has focused on taxes rather than public expenditures, and it suffers from various methodological weaknesses.[4] In particular it is now recognised that tests of the output effects of public expenditure decompositions (and other fiscal variables) must accommodate the total government budget (expenditures, revenues, deficits); a feature missing from much of the earlier literature.

As shown by Kneller *et al.* (1999), without accounting for the government budget constraint (GBC), evidence on the long-run output effects of public expenditures including *total* public expenditures is difficult to interpret and non-robust. However, where specific decompositions have been examined, education, health and/or transport & communication (T&C) spending have sometimes been shown to be positively associated with GDP growth – usually obtained from data covering 2-3 decades. For example, the meta-analysis of Nijkamp and Poot (2004), which covers many studies using a variety of methodologies, finds evidence of strong GDP growth

[3] Albertini *et al.* (2014) examine the short-run fiscal multipliers for different components of government spending at the zero lower bound. These authors calibrate a simple new Keynesian model and show that the multiplier at the zero lower bound is smaller when government spending is productive or when consumption and government purchases are substitutable.
[4] See Slemrod (1995), Tanzi and Zee (1997), Myles (2000) for earlier, more detailed reviews and Temple (1999) for a more wide-ranging review of growth empirics.

associated with higher education and infrastructure spending.[5] Nevertheless a general failure to incorporate the GBC in prior empirical estimates and a tendency to focus on specific spending categories while ignoring others, limits the reliability and generality of this evidence. Alfonso and Jalles (2014), for example, focus on social security and welfare, education and health, finding a negative impact for the first type of government spending and positive impact for the last two in a large sample of developing and developed countries.

Acosta-Ormaechea and Morozumi (2013) analyse the impact of five functions (transport and communications, defence, education, health and social protection) on growth for a sample of 56 low, middle and high income countries. They acknowledge the GBC by examining the impact of an increase in an expenditure component at the cost of each of the other four components analysed. They find that only education spending has growth-enhancing effects that are statistically significant. This happens specifically when an increase in education spending is financed by a fall in health or social protection spending. Teles and Mussolini (2014) incorporate the GBC and find, again for a mixed sample of developing and developed countries, that productive spending affects economic growth positively, but that this impact lowers as the public debt increases. Gemmell *et al.* (2011) focus on the impact of taxes and aggregated 'productive' and 'unproductive' public spending categories on OECD growth rates, using annual data over 3-4 decades. Their analysis highlights timing and persistence aspects of these relationships, however, and does not address the relationships between more detailed spending categories and long-run GDP levels or growth rates.

Similarly, Arnold *et al.* (2011) have examined the relationships between *tax* decompositions (personal income, corporate income, consumption and property taxes) and long-run GDP *levels* in the OECD and found that, not only is there evidence of long-run GDP level effects from some tax types, but that the way in which the GBC is captured in the regression specification affects interpretation of resulting parameter estimates, as argued by Kneller *et al.*(1999).[6]

An advantage of the 'levels specification' adopted by Arnold *et al.* (2011) is that it allows the data to identify the degree of persistence within the fiscal policy-GDP growth response, rather than impose a functional form derived from endogenous growth models embodying permanent effects. Below we adopt the Arnold *et al.* (2011) approach but adapted for public expenditure decompositions rather than taxes.

[5] They do not find strong evidence for fiscal variables in general however, which is again not surprising given the mixture of positive and negative effects expected for many fiscal aggregates, depending on composition and financing methods.

[6] Some of this evidence has been challenged by Xing (2011).

III. Methodology and data

Growth versus level effects

The empirical specification adopted by much of the empirical literature testing for public expenditure-GDP growth effects since Devarajan *et al*. (1996) is based on their endogenous growth model. Derived from the Barro (1990) framework, this model essentially generates an estimating equation in which the long-run (proxying the steady-state) growth of GDP in country i at time t is a function of the ratio of total government expenditure, E, to GDP and a vector of shares of j individual categories within total government expenditure, e_j. Sets of conditioning variables, consistent with a more general growth model or more pragmatic considerations, are often included; see Devarajan *et al.* (1996; p.322). When applied in a panel context, the Devarajan *et al.* endogenous growth model takes the form:

$$g_{i,t} = \Delta y_{i,t} = \beta_1 \left(E/Y\right)_{i,t} + \sum_k \beta_k \left(E_k / E\right)_{i,t} + \mu_i + \omega_t + \varepsilon_{i,t} \ldots \text{control variables} \tag{1}$$

where y_{it} is the log of GDP per capita in country i at time t, g_{it} is the growth rate (log difference: Δy_{it}), Y is GDP, E is total public expenditure and E_k is the k^{th} expenditure component, μ_i and ω_t are country and time fixed effects, and ε_{it} is a classical error term.[7]

A difficulty with equation (1) is that it is specific to the endogenous growth model with permanent growth effects of fiscal policy and no transitional dynamics. To allow for the possibility of non-permanent but potentially persistent Solow-type transitional dynamics as described above requires a more flexible functional form, particularly when working with annual data where mean reversion is likely to be important. We follow Arnold *et al*. (2007, 2011) and use an autoregressive distributed lag, ARDL(p, q), model, parameterised in error correction form.[8] This allows both the short-run dynamic and the long-run equilibrium relationships between GDP and fiscal variables to be separately identified.

Consider the following general ARDL(p, q) specification:

$$y_{i,t} = \sum_{j=1}^{p} \alpha_{i,j} y_{i,t-j} + \sum_{j=0}^{q} \beta_{i,j} X_{i,t-j} + \mu_i + \varepsilon_{i,t} \tag{2.1}$$

and

$$\Delta X_{i,t} = \zeta_1 \Delta X_{i,t-1} + \zeta_2 \Delta X_{i,t-2} + \ldots + \zeta_s \Delta X_{i,t-s} + u_{i,t} \tag{2.2}$$

[7] The dependent variable in some cases is per capita GDP growth, which simply involves a re-parameterisation when GDP is on the left-hand side, and the vector of control variables includes a measure of population growth on the right-hand side.

[8] Arnold *et al.* (2007) show that such a specification can be consistent with, and nest, alternative augmented Solow and (Uzawa-Lucas) endogenous growth models.

where the vector $\boldsymbol{X}_{i,t}$ in (2.1) includes both the fiscal variables of interest (e.g. the level and mix of expenditures, E/Y and E_k/E respectively), and control variables. The α, β, and ζ are parameters to be estimated; where the ζs capture the autoregressive process in $\Delta X_{i,t}$.

A number of parameterisations of (2.1) are possible (see Wickens and Breusch, 1988, for discussion), but it is readily shown that (2.1) can be rearranged in error correction (ECM) form to become:

$$g_{i,t} = \Delta y_{i,t} = \phi_i\left(y_{i,t-1} - \beta_i X_{i,t}\right) + \sum_{j=1}^{p-1} \alpha_{i,j}^{*} \Delta y_{i,t-j} + \sum_{j=0}^{q-1} \beta_{i,j}^{*} \Delta X_{i,t-j} + \mu_i + \varepsilon_{i,t} \qquad (3)$$

where $\phi_i = -(1 - \sum_{j=1}^{p} \alpha_{i,j})$ captures the error correcting component, and $\beta_i = (\sum_{j=0}^{q} \beta_{i,j}/\phi_i)$ captures the long-run equilibrium relationships between y and $\boldsymbol{X}$, with short-run effects measured by $\beta_{i,j}^{*}$ – the parameters associated with the ΔX variables in (3). The error correction term, ϕ_i, is a measure of the speed at which the model returns to equilibrium after a shock to exogenous variables. Equation (3) allows ϕ_i, and β_i, to vary across countries, though alternative econometric approaches (PMG, MG, DFE etc.) involve differing homogeneity restrictions, as discussed further below.

The degree of persistence of any fiscal effects on output *growth* in the model is, of course, determined by the model's convergence properties.[9] Note that, although the per capita GDP *growth rate*, $g_{i,t}$, appears on the left-hand side of (3), this regression captures the impacts of fiscal and other variables on the long-run *level*, *not the growth rate*, of GDP, since (3) is merely a re-parameterisation of (2.1).

The government budget constraint

Because the GBC describes a 'closed system', whether forms such as (2.1) or (3) are adopted, the estimating equation needs to recognise that expenditures, $E_{i,t}$, must be financed by revenues, $R_{i,t}$, and/or the budget surplus/deficit, $D_{i,t}$, each with potential output effects, since $D_{it} = R_{it} - E_{it}$ (see Kneller *et al.*, 1999; Gemmell *et al.*, 2011). As a result, in addition to the variables E/Y and $\Delta(E/Y)$ within the vectors $\boldsymbol{X}$ and $\Delta\boldsymbol{X}$ in (3), we should think of (R/Y) and (D/Y) as also potentially having effects on output with any net output effect depending on the particular 'financing' combinations assumed. Hence a decomposition of $\beta_i \mathbf{X}_{i,t}$ in equation (3) includes (setting $\beta_i = \beta$ for convenience):

$$\ldots \beta_1 (E/Y)_{i,t-1} + \beta_2 (R/Y)_{i,t-1} + \beta_3 (D/Y)_{i,t-1} + \sum_k \beta_{4,\mathrm{k}} (E_k/E)_{i,t-1} \ldots. \qquad (4)$$

[9] Though the parameter ϕ can be interpreted as a speed of adjustment, it is not equal to the more familiar rate of conditional convergence in this model. With short-run dynamics included here, the convergence rate varies both over time and across countries.

and similarly for the short-run output effects of expenditures, revenues and deficits, captured within an equivalent decomposition of $\Delta\mathbf{X}_{i,t-1}$ in (3). However, since introducing all three variables would be perfectly collinear in a regression, one may be chosen arbitrarily to omit. Omitting $(D/Y)_{i,t-1}$, (4) becomes:

$$\ldots\ (\beta_1 - \beta_3)(E/Y)_{i,t-1} + (\beta_2 - \beta_3)(R/Y)_{i,t-1} + \sum_k \beta_{4,\mathrm{k}}(E_k/E)_{it-1} \ldots. \tag{5}$$

This demonstrates the correct interpretation of the output effects of expenditure and revenue variables, within this model. Estimates of the fiscal parameters in (5) capture the effects of increases in total expenditure, or decreases in revenues, financed by changes in the budget deficit (the omitted category in this example). These effects can be seen to depend on the signs and relative sizes of β_1, β_2 and β_3. In comparison, the interpretation of the coefficients on the individual expenditure *share* components of interest, $\beta_{4,k}$, remains unaffected. It measures the effect on GDP of a change in the share of spending on each category holding constant the other variables in equation (5) including total expenditure.

In the applications below, each regression includes one of the $k = 1 \ldots K$ expenditure share elements, E_k/E, where the included k^{th} expenditure category is rotated across the K different categories. Including each expenditure share in turn (rather than all $K-1$ expenditure shares simultaneously) saves on degrees of freedom in our panel regression model which requires a large number of parameters to be estimated due to the assumption of short-run parameter heterogeneity across countries; see below.

By repeating regressions on equation (3) but including a different expenditure category each time (from (5)), it can be shown that the per capita GDP impacts of a bilateral switch between any two expenditure categories, l and m, can be obtained. This uses the property of derivatives whereby, $dE_l/dE_m = (dE_l/dE)/(dE_m/dE)$ and the fact that total expenditure is held constant in these regressions. It can be shown that the output effect associated with a switch between two expenditure shares l and m, is a form of weighted sum of relevant $\beta_{4,l}$ and $\beta_{4,m}$ where the β_4s are obtained from the relevant versions of regressions on equation (3) – including (E_l/E) and (E_m/E) respectively.

Allowing for heterogeneous fiscal-output effects

We estimate an equation of the form in (3) above using the pooled mean group (PMG) methodology of Pesaran *et al.* (1999), and compare this with the equivalent results using fixed effects (FE), and mean group (MG), estimators. The PMG estimator provides a useful intermediate alternative between estimating separate regressions for each country (the MG case) and the fixed effects estimator which imposes homogeneity on all slope coefficients and error variances across countries. The PMG estimator allows the intercepts, short-run coefficients and error variances to differ freely across groups, but constrains the long-run coefficients to be the same.

Acceptance of homogeneity of the long-run responses implies that the results from the PMG estimator are more efficient than those from the alternative MG estimator which permits long-run heterogeneity. Pesaran *et al.* (1999) demonstrate that the PMG's short-run parameter heterogeneity yields more reliable estimates of the long-run responses, and can affect estimated speeds of convergence towards long-run equilibrium.

A disadvantage of the PMG estimator is that, unless the available time series are long, a degrees of freedom problem is soon reached. For the dataset available here this requires some restrictions on lag lengths and/or the set of right-hand-side (RHS) variables. For this reason we restrict the RHS variables to include each expenditure category (share) separately in turn, with up to 2 lags. Though this lag length is relatively short, inclusion of the lagged dependent variable ensures that adjustment to equilibrium can be highly persistent.

We also include two non-fiscal control variables –the private non-residential investment rate, and employment growth. These variables appear in many empirical growth models capturing the role of private sector inputs. Also, recognising that some public expenditures may impact on GDP partially through their impact on private investment, we omit the latter variable from some specification to explore this possibility.

An important aspect of equation (3) when testing for the output effects of public expenditures is that it allows for the degree of persistence of these effects to differ across expenditure types, at least to the extent that these can be captured by the 30+ years of our data. In addition, though we initially adopt the PMG assumption of long-run parameter homogeneity across countries, we also test this against the mean group (MG) and fixed effects (FE) alternatives with a Hausman test.

Addressing endogeneity

A major concern when running regressions of the form in (3) is , of course, the potential for simultaneity between GDP per capita and the right-had-side variables - especially the fiscal and investment variables (a point stressed by Slemrod (1995) in a cross-section growth regression context). In addition, country-specific, time-varying factors such as changes to political and institutional settings may influence both fiscal policy and GDP per capita, and which are not adequately captured by time fixed effects. These effects may be compounded by persistence in the annual fiscal series and potentially long lags in the impacts of fiscal variables on GDP per capita, generating serial correlation in the error process that must be dealt with adequately. As a result the fiscal variables of interest here may be endogenous, with well-known problems of interpretation of the short-and long-run OLS parameters.

In a series of papers in the mid-1990s, Pesaran and associates (see, for example, Pesaran and Smith, 1995; Pasaran, 1997; Pesaran and Shin, 1999) demonstrated that, under a number of conditions, estimates of the long-run parameter vector, $\boldsymbol{\beta}$, from OLS regressions of models such as (3),are consistent.[10] Further, based on simulation results, Pesaran *et al.* (1999) demonstrate that even in small samples standard *t*- and *F*-tests on the long-run parameters from the ECMs are valid, given suitable specification of the lag structures of dependent and independent variables.[11]

Further, where serial correlation is a concern, ‘appropriate modification of the orders of the ARDL model’ (Pesaran and Shin, 1999, p.386) is sufficient to deal with both the serial correlation in the error process and/or regressor endogeneity (see also Kanas and Kouretas, 2005). We discuss these aspects further in section IV when we present our empirical results.

The updated dataset

Results reported below are based on an extension of the Bleaney *et al.* (2001) dataset, which uses the IMF’s Government Financial Statistics (GFS) fiscal data for 17 OECD countries to construct measures of total expenditure and individual expenditure shares, distortionary and non-distortionary taxes, and budget surpluses/deficits.[12] The original data, available from the early 1970s to 1995, has been updated to 2007 or 2008, providing around 30-35 annual time-series observations each for most of the 17 countries.[13] We exclude data after 2008 to avoid long-run estimates being unduly affected by the large shock (associated with the global financial crisis and its aftermath) to fiscal, GDP and other variables from their equilibrium values at the end of our period.

Data on GDP and the two control variables – the private investment/GDP ratio and employment growth – were obtained for the same period from OECD sources. An important difference from previous studies of taxes/public spending and GDP is that our investment control variable is *private non-residential* investment (PNRI) instead of total investment (gross fixed capital formation). Since all regressions include various public expenditure variables, the use of

[10] Indeed, where the variables are I(1), they argue that the estimates are super-consistent.
[11] See, for example, Pesaran (1997, pp. 182-185), Pesaran and Shin (1999, pp. 381-387; 404-405).
[12] The 17 OECD countries are: Australia, Austria, Canada, Denmark, Finland, France, Germany, Iceland, Luxembourg, Netherlands, New Zealand, Norway, Spain, Sweden, Turkey, UK and US. Other OECD countries are excluded due to missing values for several years of the time series from the early 1970s to 2008; see Appendix A.
[13] This is not straightforward, however, because of changes in the GFS methodology, which moved from a cash accounting basis to an accruals accounting basis for fiscal data from the late 1990s onwards. ; see Appendix A.

PNRI avoids the possibility of 'double counting' much public investment which otherwise would contribute to both the investment and public expenditure data.

In addition, before running all regressions we de-mean all data series as recommended by Lee *et al*. (1997) and Pesaran *et al.* (1999) to deal with potential cross-sectional error correlation. That is, all variables are measured as deviations from their sample means in each year, acting analogously to time dummies, thus removing common shocks in the data.

IV. Empirical results

Pre-testing for lag lengths

As discussed in section III, following Pesaran and Shin (1999), regressions on equation (3) require knowledge of the appropriate lag structure if long-run parameter estimates are to be interpretable as capturing exogenous or weakly exogenous effects; see below. Before running our PMG regressions, our procedure therefore follows the pre-test proposed by Pesaran and Shin (1999) to select the preferred lag structure for the dependent, and each independent, variable based on the Schwarz Information Criterion (see Pesaran and Shin, 1999). This uses up to two lags for each variable and country; in practice we find that two lags are required relatively rarely.[14] The results of PMG regressions on equation (3) reported below are based on the model-selected lag structures.

Testing for total public expenditure effects

Before examining the GDP effects of the share of particular expenditure categories it is worth noting the impact of the implicit financing category in the government budget constraint methodology because it changes the interpretation of the included parameters. As the parameter estimates on the shares of individual spending categories in total expenditure are unaffected by this we omit them from the regressions for the moment.

Table 1 shows the long-run and short-run parameters for six PMG regressions in which total spending is included, financed by combinations of three alternative fiscal variables (non-distortionary taxes, deficits and distortionary taxes) in columns 1-4. Column 5 repeats the regression in column 2, but with the private investment control variable omitted.[15] Results for alternative (MG and DFE) estimators are then reported in Table 2.

[14]For example, in each of regressions 1 – 3 of Table 1, there are a maximum of 204 possible parameters (17 countries *x* 6 included variables *x* 2 lags). Regression 4 involves 5 variables, hence 170 possible parameters. For those four regressions, two lags were selected in only 11%, 8%, 9% and 4% respectively of the possible cases.

[15] In general the 'non-distortionary' taxes in this context are consumption taxes. The term 'less growth distortionary' may be more appropriate in this case since these taxes can distort investment decisions in models where labour supply effects are allowed for. The method of aggregating the GFS functional classification into these sub-aggregates is described in Bleaney *et al.* (2001). 'Other revenues' refers to the GFS categories 'taxes on international trade and transactions', 'other taxes' and 'non tax revenues'. It is unclear how distortionary these might be. This category of

[Table 1 about here]

In interpreting the parameters in Table 1, it should be borne in mind that the data have been de-meaned; hence the parameters estimate the impact of a deviation from OECD mean levels of the variable in question on deviations of GDP per capita from the OECD mean, in a given year. The role of different implicit financing of fiscal variables excluded from regressions has been discussed by Gemmell *et al.* (2011) so is only briefly reviewed here.

Regression 1 in Table 1 reveals that increasing total public spending, financed by the (omitted) non-distortionary taxes has a long-run output effect that is positive and statistically significantly different from zero. However, the same spending increase can alternatively be financed by either a reduced budget surplus (regression 2), or by an increase in distortionary taxes and/or increases in any/all taxes (regressions 3 and 4), or by some combination. Regressions 2-4 in Table 1 show that each of those cases produces a negative, statistically significant long-run parameter on the public expenditure variable. That is, the assumed (or imposed) method of *financing* a public expenditure increase is crucial for identifying the likely impact of that expenditure on long-run levels of GDP per capita.

The evidence in Table 1 suggests that, within our dataset, use of 'non-distortionary' forms of taxation (such as VAT) to fund a given unit of additional public spending has been associated, *ceteris paribus*, with increases in GDP per capita levels in the long-run, whereas other forms of financing such spending have been associated with decreases in GDP per capita in the long-run.[16] Table 1 results for the total public expenditure parameters also tend to suggest that the adverse long run effects on GDP per capita from distortionary tax financing (at –0.024) are significantly greater than those associated with other forms of financing (around –0.012). However this should be interpreted cautiously. As we show below, specifications that also allow for different effects on GDP per capita from differences in specific public expenditure *shares* (see Table 3) suggests that the magnitude of the *total* expenditure parameter can be somewhat sensitive to the included expenditure share variables.

Recognising that some fiscal-GDP effects could be mediated through private investment (as opposed to via factor productivity), regression 5 in Table 1 examines whether omitting investment from the control variables affects conclusions regarding the GDP impacts of public

---

other revenues amounts to 13.4% of total revenues in our sample, and non-tax revenues account for 9.6% of total revenues.

[16] The control variables in these regressions generally perform as expected: larger investment ratios and employment growth (proxying labour force growth) are associated with higher GDP per capita, though the former is not statistically robust, sometimes taking a negative sign. The 'lagged residual' term captures the error correction parameter, ϕ_i, in (3) and, at around –0.03 to –0.09, it implies a fairly high degree of persistence.

expenditure. Comparing regressions 2 and 5 (the equivalent, budget deficit-financed cases), the estimates suggest that parameters on fiscal variables, including total public expenditure, are little affected by the omission; that is, public expenditure effects are largely orthogonal to private investment effects.[17]

Overall, while the results in Table 1 yield similar estimates of fiscal variable impacts on GDP (when the appropriate interpretation is recognised as discussed above), our preferred specification is regression 2 which omits the budget surplus. This facilitates interpretation because deficit-funded tax or expenditure changes have an intuitive economic interpretation and because across all countries a one unit change in a budget surplus/deficit is a more uniform metric whereas this is less true for a one unit change in distortionary or non-distortionary tax. This is because the composition of these tax variables can differ across countries, for example, the shares of personal and corporate income taxes, or VAT and excise taxes, typically differ across countries, and each may differ in their GDP impacts. We therefore base our further testing of public expenditure effects using this specification that omits the budget surplus.

As noted above, the PMG estimator provides a greater degree of flexibility, compared to dynamic fixed effects models, in estimating long-run effects by allowing short-run parameters to differ across countries. The Mean Group estimator allows for further flexibility via heterogeneous *long-run* effects across countries but at a further cost in terms of degrees of freedom. We therefore test for possible sensitivity of our Table 1 results to the PMG specification. Based on regression 2 in Table 1, Table 2 presents equivalent results for dynamic fixed effects and mean group models (and repeats PMG results for comparison; non-fiscal variables are included in regressions but omitted from the table to save space).

Of immediate note in Table 2, the parameters on each fiscal variable take the same signs and are of similar orders of magnitude. However standard errors are generally larger using the MG or DFE estimators such that parameter estimates for those cases are generally not statistically significant at conventional levels. That this is the case for the MG estimates is no surprise, given the averaging involved across heterogeneous long-run parameter estimates. However the comparison between PMG and DFE suggests that allowing for heterogeneous *short-run* parameters across countries (in the PMG) enables more precise *long-run* parameter estimates to be obtained. This mirrors results obtained by Gemmell *et al.* (2011, pp.F43-44) who also find

[17] This is perhaps not surprising, given the limited contribution to long-run output apparently due to private investment in Table 1. To investigate this issue further we ran similar regressions with original, rather than de-meaned, data. This yielded a larger, statistically significant investment parameter. This may reflect a high degree of co-movement in private investment across countries such that, after de-meaning the data, there is little country-specific variation in the investment data, or that any such variation is largely unrelated to countries' long-run per capita GDP levels (though it may still have a strong effect on short-run growth rates).

that the DFE, unlike the PMG, requires a lag structure of up to eight lags to identify long-run parameters.[18]

The far right-hand column of Table 2 also reports results from Hausman tests of the PMG restrictions against those associated with the (less restricted) MG estimators for each fiscal variable and the regression specification as a whole. The final row compares the PMG regression with the DFE equivalent. The Hausman tests fail to reject the null hypothesis of non-systematic differences in parameters, supporting the PMG restriction of homogeneous long-run effects across countries. This also holds for the regression as a whole ($\chi^2(5) = 7.20$; p-value = 0.21). By contrast, the Hausman test rejects the null hypothesis that differences between the PMG and DFE are not systematic ($\chi^2(5) = 16.51$; p-value = 0.00); that is, the additional short-run restrictions of the DFE model are rejected.

Finally, though results in Table 2 appear to support long-run parameter homogeneity, in Appendix Table B1 we report MG parameter estimates for total public expenditure by country, as well as cross-country means and medians. From Table 2, the mean long-run MG parameter for total public spending is –0.020 (t = 1.65). The appendix table shows that, when long-run parameters are allowed to differ, three countries, Sweden, Australia and New Zealand, take relatively large (absolute) parameter values for total expenditure (–0.20, 0.07, –0.06 respectively). However, for 13 of the 17 countries parameter estimates are negative and generally close to the arithmetic mean (median) long-run estimates of –0.020 (–0.009).

Public Expenditure Composition and GDP

To explore the potential long-run effects of public spending composition on GDP, we again focus on the Table 1 specification (PMG; regression 2) in which changes in total spending are implicitly funded from an increased budget deficit (reduced budget surplus). Table 3 shows the results from repeating this PMG regression, but adding the shares of each GFS public spending category – transport & communications (T&C), education, health, etc. – in total expenditure (excluding interest payments). To save space the table shows only the parameters on total public expenditure and the spending decompositions of interest.

Due to a lack of degrees of freedom, it is not possible to include all (*n-1*) detailed categories in one regression though, as noted earlier, it would be possible to construct the growth trade-off associated with any bilateral expenditure share switch. The parameter on each expenditure share in regressions (1)–(9) should therefore be interpreted as the impact on long-run per capita GDP

[18] Recall that the MG parameters reported are *unweighted* averages of the country-specific values. As a result this average can be heavily influenced by a few extreme values, even when these are imprecisely estimated.

levels of switching spending into the included expenditure category (say, T&C) and away from remaining expenditure categories on a *pro rata* basis, holding total spending constant as a ratio of GDP. A significant positive (negative) parameter therefore indicates that the category in question has a greater (smaller) impact on long-run GDP than the remaining expenditure categories.

Note first that all regressions (except the regression for T&C) reveal net negative *total* spending growth effects when funded from increased budget deficits (consistent with Table 1 results for this form of financing). Whilst the positive short-term GDP impacts of fiscal stimulus packages have been important in the recent macroeconomic debates (discussed further below), these initial results are consistent with a possible adverse impact on GDP per capita over the longer run if higher deficit-financed expenditure levels persist.

Switching into some spending categories might be expected to have limited effects on GDP either where there are negligible short-run impacts from the relevant public expenditures or where the 'favoured' category has similar (possibly substantial) GDP impacts to those categories where spending shares decrease. For example, as in Barro (1990), if public spending levels and shares have each been set in a growth-maximising manner there should be no evidence of output benefits or costs from reallocating expenditures. However, clearly some public spending in OECD countries is allocated to meet non-growth objectives such as social welfare provision or redistribution, and lack of knowledge of the growth effects of different spending types may also inhibit growth-maximising policy choices.

[Table 3 about here]

We find, in Table 3, that most of the expenditure shares exhibit small positive or negative long-run GDP effects that are not statistically different from zero. However, we also find that some changes in the mix of public expenditures have significant effects on GDP over the long-run. In interpreting regression parameter estimates it is important to remember that these represent *the combined impact* of one additional unit of a particular type of spending (say, education), financed by reductions in other spending categories

Table 3 provides evidence of potential *positive* GDP effects from changes in T&C, education and (less reliably identified) housing spending shares. The parameter for health spending is positive but small and with a relatively large standard error. Negative long-run associations with GDP are observed for spending shares for social welfare, defence, economic services, general public services and recreational services, but only for general public services is the estimate statistically significant at 10% or less.[19]

The parameters in Table 3 can be interpreted as follows. Consider the spending share with the largest estimated GDP impact in Table 3: T&C at 0.022. Across the sample, the T&C spending share averages 5.5% (standard error = 3.3%), and GDP per capita growth averages 2.0% (standard error = 2.4%). Hence a permanent 1 percentage point increase in the T&C share in total spending (e.g. from 5.5% to 6.5%) is associated, on average, with a long-run level of GDP per capita that is 2.2% higher than the counterfactual of an unchanged T&C spending share.[20] Thus, considering a 20 year horizon, GDP compounded at 2.0% growth per year would be expected to rise from 100 to 148.6 after 20 years in the absence of any T&C change. The 1 percentage point T&C share increase to 6.5% is thus predicted to raise GDP to around 150.8 instead of 148.6 after 20 years.

Equivalently this implies that the GDP expected after 20 years with no T&C change is reached approximately one year earlier when the T&C spending share increase permanently to 6.5% from 5.5%. This increase in GDP seems a plausible order of magnitude arising from a persistent non-marginal reallocation (about one-third of the sample standard error) towards a directly growth-enhancing spending category, and away from all other spending categories pro rata which, according to our estimates, have much less positive or zero/negative impacts on GDP.

These results align with a number of findings within the current literature. For example, Nijkamp and Poot (2004), in their meta-analysis of studies of fiscal policy and growth, report the positive impact of infrastructure and education, especially public education *investment* spending, on GDP as finding broad support in the literature. They conclude (p.91): 'On balance, the evidence for a positive effect of conventional fiscal policy on growth is rather weak, but the commonly identified importance of education and infrastructure is confirmed'. Our results are also consistent to those found more recently by Alfonso and Jalles (2014) and Acosta-Ormaechea and Morozumi (2013), who show evidence of a positive impact of education and health and negative effect for social security and welfare.

Vandenbussche *et al.* (2006) and Aghion *et al.* (2009) provide evidence that the impact of education attainment and public education spending on growth rates depend on the proximity of

[19]As with Table 1 results, we have also examined whether the PMG results in Table 3 are preferred to either a MG or DFE approach. Results are reported in Appendix Table B2, which again confirms that the PMG estimates are preferred, on a Hausman test, to either the MG or the DFE estimates. Both these other approaches tend to yield noisy parameter estimates for public expenditure variables.

[20]Recall that our results relate to a change *relative to the average of the OECD sample*; hence we assume here that only one country experiences the simulated T&C spending share increase.

a country to the technological frontier and the composition of the educational investment. Education in general, and higher-level education in particular (e.g. research-related and tertiary), have larger positive effects closer to the frontier.

Are expenditure-GDP effects endogenous?

Perhaps the most frequently cited reason for skepticism regarding the validity, or interpretation, of aggregate growth regressions is the possibility that estimated relationships represent correlations but not causation.[21] For our results so far, we cannot discount the possibility that the evidence arises from simultaneous relationships between GDP and fiscal variables. That is, as well as direct impacts of fiscal variables on GDP, changes in GDP may be inducing changes in these fiscal variables. It is also possible that fiscal policy changes will be associated with country-specific time-varying variables, such as political conditions, that also influence GDP levels or growth rates.

The arguments that higher GDP levels or growth rates induce changes in *total* government expenditure or taxation are well known. Economic downturns reduce taxable capacity and lead to increases in certain types of public expenditure such as unemployment benefits and social insurance payments. Though these may be at the expense of other types of expenditure, this is often insufficient to prevent total spending from rising in downturns. Short-run contractions of less cyclically-dependent expenditures, such as public investment, are typically more difficult to achieve when social expenditures increase; see Sanz and Velázquez (2004) and Sanz (2010).

The effect on expenditure components is more ambiguous. As already noted, social welfare expenditures might be expected to rise in response to an economic downturn implying a negative correlation with GDP. On the other hand, the share of more ‘productive’ expenditures would be expected to rise when faster GDP growth generates additional revenues, and demands for welfare-related expenditures, such as social insurance, weaken. This would have contrasting effects on the shares of these different components of expenditure in total expenditure and for total expenditure as a ratio to GDP. In addition, it is sometimes argued that over the longer-term some public expenditures such as education and health, display income-elastic qualities so that higher income levels induce greater consumer demand for such services typically delivered via public spending (Slemrod, 1995).

These arguments suggest the possibility that our previous evidence of positive impacts on GDP of T&C, education etc, and some negative impacts associated with, for example, social welfare spending or ‘economic services’, might reflect these reverse causation arguments. Or

[21] See Slemrod (1995) for a robust critique of this issue both in principle and with respect to cross-country empirical evidence.

they may simply be the outcome of fiscal variables being correlated with time-varying unobservable factors that affect GDP. Similarly, endogenous responses could also account for our estimated relationship between GDP and (deficit-financed) total public expenditure, if economic downturns induce additional total spending in association with worsening deficits.

Pesaran and Shin (1999) contend that, in the context of ARDL models, the problem of endogenous regressors can readily be handled by the PMG where the regressors are I(1) and not cointegrated among themselves. In this case, Pesaran *et al*. (1999) and Pesaran and Shin (1999) show that endogeneity can be corrected by an appropriate augmentation of the lag structure of the ARDL(p, q) model – to an ARDL(p, m) model, where $m \geq q$. Endogeneity of I(1) regressors can be accommodated by a projection of the errors on the regressors. An ARDL with sufficiently long lags overcomes the endogeneity problem if regressors are not cointegrated, and where the focus of the analysis is on the *long run* coefficients (see Pesaran and Shin, 1999, pp. 372-3; 384-5).

To explore these endogeneity issues for our case we first check whether our variables are I(0) or I(1) and whether they are cointegrated. Secondly, we consider the appropriate lag structure for our ARDL(p, m) model. We discuss each in turn below.

Testing the order of integration and cointegration

We begin by testing whether our variable as are I(0) or I(1), using the panel unit root tests of Harris-Tzavalis (1999), Breitung (2000) and Pesaran (2007). Results are reported in Appendix Table B3. The Harris-Tzavalis (1999) test assumes that all panels have the same autoregressive parameter and that the number of time periods is fixed. The Breitung (2000) test also assumes that all panels have a common autoregressive parameter, but Breitung shows that the test has power in the heterogeneous case, where each can take its own autoregressive parameter.[22] The Harris-Tzavalis test rejects the null hypothesis of non-stationarity only for the budget surplus and employment growth. The Breitung test rejects the null for these two variables along with the housing and economic services spending shares. Note, however, that in the regression specifications that are the main focus of our analysis, the budget surplus is omitted.

The second generation panel unit root test of Pesaran (2007) – the cross-sectionally augmented Im, Pesaran and Shin (CIPS) test – allows for heterogeneity in the autoregressive coefficient of the Augmented Dickey-Fuller regression with lagged cross-sectional mean and its first difference capturing the cross-sectional dependence. Pesaran (2007) shows that the CIPS

[22]Breitung's (2000) Monte Carlo simulations suggest that his test is substantially more powerful than other panel unit-root tests for the modest-size dataset he considered (N=20, T=30), which is similar to the sample size we have here.

test has satisfactory size and power even for relatively small samples. According to this test, all of our variables are I(1) except employment growth and the defence spending share. Overall, from these three tests it seems that our variables are best treated as non-stationary except for the budget surplus and employment growth. After taking first differences, however, the panel test (not reported) rejects the null of non-stationarity for each of the variables. From this we conclude that all the variables are I(1), except employment growth and budget surplus.

For the cointegration test, we implement the error-correction, based panel cointegration tests developed by Westerlund (2007), based on structural rather than residual dynamics, which therefore does not impose any common-factor restrictions. The idea is to test the null hypothesis of no cointegration by inferring whether the error-correction term in a conditional panel error-correction model is equal to zero. The tests (P_a and P_t) [23] are general enough to allow for a large degree of heterogeneity, both in the long-run cointegrating relationship and in the short-run dynamics, and dependence within as well as across the cross-sectional units. P_a and P_t test the alternative hypothesis that the panel is cointegrated as a whole.

First we test whether the variables included in the PMG regression 1.2 (the base estimation for subsequent regressions in Table 3) are cointegrated, including GDP per capita, our fiscal variables and investment. Appendix Table B3 shows that P_t rejects the null hypothesis of no cointegration for all of our PMG specifications at a 5% level whereas P_a marginally fails to reject the hypothesis at a 10% level (p-value 0.13). In contrast, these two tests do not reject the hypothesis of no cointegration among our regressors. Including expenditure shares in the cointegration test does not change this conclusion. Our model therefore appears to fulfill the first set of Pesaran and Shin (1999) conditions which allows the ARDL model to overcome endogeneity: namely, I(1) variables and non-cointegrated regressors.

Now we turn to the second condition to correct for the problem of endogenous regressors, namely, the appropriate modification of the orders of the ARDL.

Testing the ARDL lag structure

Pesaran *et al*. (1999) and Pesaran and Shin (1999) stress the importance of allowing a *sufficiently long* lag structure to deal with endogeneity problems in this ARDL context. Their proposed method of correcting for endogeneity – in the form of contemporaneous correlation among the error terms of (2.1) and (2.2) – is by ‘appropriate augmentation’ of the lag structure

[23] The two differ because they start from a weighted average of the individually estimated coefficients (P_a), or their respective *t*-ratios (P_t).

of the ARDL(p, q) model. Their method assumes zero cross-correlation among lagged errors; namely Cov($\varepsilon_{i,t-i}$, $u_{i,t-i}$) = 0, for i ≠ j.

In particular, they show that running an ARDL(p, m) model where m = *max*(q, *s*+1), where *s* is the appropriately selected lag length of the autoregression of $\Delta X_{i,t}$ in (2.2), yields consistent estimates of the long-run parameters of interest. Hence 'in the context of the ARDL model inference on the long run parameters is quite simple and requires *a priori* knowledge or estimation of the orders of the extended ARDL(p, m) model. Appropriate modification of the orders of the ARDL model is sufficient to simultaneously correct for the residual serial correlation and the problem of endogenous regressors.' (Pesaran and Shin, 1999, p.386).

These arguments for 'augmentation' may be thought of as applying to an initial ARDL(p, q) model where lags have been chosen *a priori* as described by Pesaran and Shin above. Thus, for example, where an initially selected ARDL (1, 1) suffers from endogeneity, this may be corrected by running an ARDL(p, m) model where m ≥ 1, based on estimates of the lag structure from regressions of (2.2).[24]

The PMG regressions discussed earlier in this section, were obtained using the Pesaran and Shin (1999) two-step strategy in which we first chose *p*, the number of lags for (log) GDP per capita, and *m* the number of lags for our independent variables (fiscal policy, investment and employment growth). Following Pesaran and Shin (1999), these were chosen using the Schwarz Criterion (SC) to identify the appropriate lag structure.[25] Hence, arguably our previous estimates should be regarded as obtained from an ARDL(p, m) model, with suitably long lag structures.

However, to allow for the possibility that a longer lag structure is required to deal with possible endogeneity problems, we also ran equivalent regressions to those reported in Table 3 but based on an ARDL(2, 2) model rather than the ARDL(p ≤ 2, m ≤ 2) reported there.[26] Results from the two models, for the expenditure variables of interest, are compared in Table 4. The key differences between the two specifications, in terms of the expenditure share parameter estimates, are as follows.

Firstly, the two expenditures previously found to have significant effects on GDP– T&C and education –have smaller parameter estimates in Table 4 but still statistically significant (at the 5% level). Secondly other expenditure share parameters – for the major spending categories of

[24] Computation of valid standard errors for the long-run parameters in our PMG models has been carried out by the so-called delta method.

[25]Pesaran and Shin (1999) claim, from Monte Carlo experiments on the ARDL(p, m) model, that the Schwarz Criterion is slightly superior to the Akaike information Criterion (AIC).

[26] Testing more than two lags is inhibited by the loss of degrees of freedom in our case.

health, housing, social welfare and defence– are generally larger (in absolute value) in Table 4 and are now statistically different from zero; negative in the case of social welfare and defence.[27]

[Table 4 about here]

Our interpretation of these results is that, allowing the ARDL model to select the appropriate lag length is preferable, while allowing for a longer fixed lag length of two periods accommodates the possibility that additional lagged effects may be erroneously omitted when selecting lags via the Schwarz Criterion. Results suggest that this latter approach, tends to *strengthen*, rather than weaken, the case for significant causal impacts from a number of public expenditure categories on long-run GDP levels.

Weak exogeneity tests

As noted earlier, Pesaran and Shin (1999) contend that the PMG overcomes endogeneity, assuming zero cross-correlation among lagged errors; namely Cov($\varepsilon_{i,t-i}$, $u_{i,t-i}$) = 0 for i ≠ j. We can therefore check formally whether our fiscal variables are 'weakly exogenous' or 'long-run forcing' for GDP. That is, whether changes in the fiscal variables can be shown to be statistically unrelated to the error correction term from the Table 3 regressions. Each fiscal variable may still react to its own lagged changes, lagged changes of other fiscal and non-fiscal variables, and to lagged changes in GDP growth.

As shown by Johansen (1992) and Boswijk (1995), weak exogeneity of the long-run parameters can be checked by estimating marginal models for each of our fiscal variables and using a variable addition test to assess the statistical significance of the error correction terms obtained from Table 3 regressions for each of those marginal models. If we can show that the fiscal variables are potentially long-run forcing for GDP, we could then interpret our results in Tables 1 and 3 as causal effects.

Specifically, following Calderón *et al.* (2014), we test for weak exogeneity of our vector of fiscal and investment variables, $\boldsymbol{X}_{i,t}$, in the following marginal models:

$$\Delta x_{i,t} = \beta_{i,1}\Delta\boldsymbol{X}_{i,t-1} + \beta_{i,2}\Delta\boldsymbol{X}_{i,t-2} + \alpha_{i,1}\Delta Y_{i,t-1} + \alpha_{i,2}\Delta Y_{i,t-2} + \delta_i \xi_{i,t}(\hat{\beta}) + \varepsilon_{i,t} \quad (4)$$

where $x_{i,t}$ represents each element of the vector of fiscal/investment variables $\boldsymbol{X}_{i,t}$; $\xi_{i,t}(\hat{\beta}) = (Y_{i,t-1} - \hat{\beta}\boldsymbol{X}_{i,t-1})$ are the estimated long-run equilibrium error correction (ECM) terms from the Table 3 regressions; and $\varepsilon_{i,t}$ is a random error term.[28] Hence the $\boldsymbol{X}$ vector includes total expenditure, distortionary taxes, non-distortionary taxes, non-residential private investment (all

[27]The smaller spending categories (general public, and recreation, services) appear non-robust, involving a change of sign.

[28] Results reported below from regressions on (4) also include a constant term.

as percentages of GDP), and the shares of each expenditure component in total public expenditure. The null hypothesis of weak exogeneity involves testing that the $\delta_i = 0$. This may take the form of a *t*-test on individual δ_i for each variable, or a Wald test that the $\delta_i s$ are jointly zero for all the suspected endogenous variables. Rejection of the null in each case implies rejection of weak exogeneity.

For each of the nine estimations in Table 3, we perform five different marginal models (for the four fiscal variables and investment – the potentially endogenous variables) country by country. We use the MG approach to average the coefficients on the ECM term across countries and then test its significance across the panel countries as a whole using the MG standard error (Pesaran and Smith, 1995). In addition, we perform the equivalent weak exogeneity test for each country separately.

Row 1 of Table 5 shows results for weak exogeneity tests on the fiscal variables and investment entering Table 1 regression (2) – the equivalent regression to those reported in Table 3 but excluding any expenditure share variables. Each cell contains the estimate (*t*-ratio) for the unweighted average MG error correction parameters in (4) for each country. Beneath these values, the cell indicates the number of individual countries for which the null hypothesis of weak endogeneity is rejected. It can be seen that, for three of the four fiscal/investment variables we cannot reject weak exogeneity across the sample of countries. While the null hypothesis is rejected across all countries combined for one variable, 'distortionary and other taxes', it is clear that for all four variables in row 1, the null hypothesis is accepted in all, or nearly all, (17) countries. Hence one or two countries appear to drive the overall result for distortionary taxes.

[Table 5 about here]

Rows 2-10 show results for the weak exogeneity tests for each expenditure share (based on Table 3). Of particular interest here is the *t*-ratio on each expenditure share variable (column 1). This reveals that, only in the case of the T&C and economic services share is the mean MG parameter in the marginal model statistically significantly different from zero (at 5%). More importantly perhaps, across the 9 expenditure shares there is typically only 1 or 2 of the 17 countries for which the individual country value of δ_i is statistically non-zero. These results offer fairly strong support to the view that the expenditure share variables can be considered weakly exogenous and hence 'long-run forcing' for GDP.

A similar conclusion emerges from Table 5 for the other variables in the regression. That is, the average MG parameter (across all countries) is not significantly different from zero in most cases, and this is also true for the vast majority of individual country estimates. Out of a total of

36 parameter sets (9 spending shares *x* 4 variables) for 17 countries, weak exogeneity is never rejected in more than 4 out of 17 countries, is rejected for 3 or 4 out of 17 countries only 8 times (out of a possible 36), with the remainder rejected in only 0–2 out of the 17 countries.

Finally, the parameter $\hat{\beta}$ in (4) which we have subjected to the weak exogeneity test is obtained from the previous model in which lag lengths were selected endogenously using the Schwarz Criterion, as described above. Given some uncertainly regarding the number of lags required to remove potential endogeneity, we re-ran Table 1 and Table 3 regressions on equation (3) imposing an ARDL (2, 2) structure. Weak exogeneity tests on the $\hat{\beta}$ parameter estimates obtained from these regressions are reported in Appendix Table B4. If anything, these reveal even stronger support for weak exogeneity that those in Table 4.

Based on the weak exogeneity results in the previous sub-section, whether it is necessary to impose two lags on the ARDL model to deal with endogeneity is unclear. However, to the extent that the longer lag structure is warranted, this would seem to strengthen, rather than weaken, the case for statistically significant effects on GDP (in plausible directions) from a number of public expenditure categories. We are inclined to be circumspect, based on Table 3, in claiming that exogenous positive impacts on GDP are associated at least with greater transport & communications and education spending. However, there is also some support for positive (negative) effects from reallocating spending towards health and housing (social welfare and defence).

Short-run dynamics

As the short-run dynamics in the PMG regressions are allowed to differ across countries, it is possible that the short run output effects of a given fiscal change differ across countries both in initial direction and in the speed with which the long run effect is realised. The trajectory followed by each country is determined by the estimated lag structure on each fiscal variable for that country together with its error-correction parameter, ϕ_i, in equation (3).

Given the noise associated with estimates of the short-run dynamics for each country, little reliance can be placed on these estimates and we do not discuss them here.[29] Instead, Figure 1 shows the 'average' time path of GDP per capita growth rates associated with a 1 percentage point change in year 0 in each of the larger expenditure categories: T&C, education, health, and social security, based on Tables 3 regressions. We use the average of the country short-run

[29]In addition, Pesaran and Shin (1999, pp.372, 384-5) warn that, even though the long-run parameters obtained from the PMG are consistent and converge to their asymptotic values relatively quickly, short-run parameters converge more slowly. These authors also conclude that an appropriate choice of the order of the ARDL model overcomes endogeneity of the long-run parameters, but short-run effects would still require explicit modelling of the contemporaneous dependence among the error terms of (2.1) and (2.2).

PMG parameters; hence obtained under the assumption of homogeneous long-run parameters. Note however that the estimated long-run parameters for health and social welfare were not statistically different from zero.

[Figure 1 about here]

As expected the figure suggests that GDP per capita *growth* converges on zero, as GDP per capita converges on its new long-run value following the 1% 'shock'. However for the two significantly positive spending types (T&C, education) there is considerable persistence with growth effects still evident after 20 years. In addition, the immediate impact (years 1 and 2) is noticeably more positive for T&C than for education – a result that seems plausible given the known immediate output effects of infrastructure-related investment compared to education spending which can be expected to take longer to impact on output. For social security, Figure 1 suggests the possibility of negative short-run growth effects – also a plausible result as spending is reallocated from other categories likely to be less dominated by consumption spending.

V. Conclusions

This paper has offered some new evidence for OECD countries on the impact of the size and composition of public expenditure on GDP per capita. Many previous regression analyses have been unclear about which specific elements of the government budget are omitted from their analyses, or interpretation of parameter estimates is made difficult where several elements are explicitly or implicitly omitted. Clearly however, the method of financing any spending increase matters, whether this involves higher taxes (with differing distortionary characteristics), higher deficits, or reductions in alternative spending categories. Previous studies have also generally tested endogenous growth model specifications with *permanent* growth effects, and endogeneity concerns remain regarding the reliability of these previous estimates.

Our examination of the impact of public expenditure on GDP for OECD countries has sought to deal explicitly with the financing aspects, allowed output effects to be persistent rather than permanent, and specifically addressed potential endogeneity concerns. Using longer time-series data than has been examined hitherto allowed us to apply more flexible (pooled mean group) methods that can accommodate heterogeneous, short-run responses across countries. We examined the growth impacts of changes in total government expenditure, alongside changes in the shares of spending devoted to various categories.

Both our initial PMG results and those that explicitly account for contemporaneous correlation, find robust long-run positive effects on GDP per capita levels for transport &

communication and education, with some evidence supporting positive (negative) effects for housing and health (social welfare) spending. Our results do not support positive long-run output effects from switching expenditure towards defense spending, despite the growth benefits sometimes claimed for this category of spending. In our analysis, the estimated effects of a switch into defense spending involve pro-rata reductions in other spending categories including T&C and education.

Though our growth model specification precludes permanent effects on the *growth rate* of GDP per capita, we nevertheless estimate relatively persistent effects on growth rates from simulated T&C and education expenditure changes that last for 20 years or more.

Our results also confirm that the assumed form of expenditure financing is crucial for estimated growth effects, whether for aggregate spending or for individual categories. In general we find evidence of negative long-run effects on output from deficit-financed increases in total public spending. Our interpretation is that such increases in public expenditure cannot be expected to be growth-enhancing unless the specific forms of that expenditure are considered carefully. Pro-rata expansions in particular would appear to be growth-retarding on average across OECD countries.

The effects of spending *share* changes are obtained here under the assumption that total spending remains unchanged and spending reallocations occur on this pro-rata basis. The evidence of positive long-run effects on GDP per capita from increases in transport & communication, and education spending *shares* appear to be robust and are consistent with the message from an increasing body of recent research. This points to the possibility that current levels of such spending in OECD countries are sub-optimal from a growth perspective, though clearly this may be consistent with a wider social welfare objective.

REFERENCES

Acosta-Ormaechea, S. and Morozumi, A. (2013). 'Can a government enhance long-run growth by changing the composition of public expenditure?', *IMF Working Paper*, WP/13/162.

Adam, C.S. and Bevan, D.L. (2005). 'Fiscal deficits and growth in developing countries', *Journal of Public Economics*, Vol. 89, pp. 571-597.

Agénor, P.R. (2008). 'Fiscal policy and endogenous growth with public infrastructure', *Oxford Economic Papers*, Vol. 60, pp. 57-87.

Agénor, P.R., and Neanidis, K. (2006). 'The allocation of public expenditure and economic growth'. *Economics Discussion Paper* No. EDP-0608. Centre for Growth and Business Cycle Research, Economic Studies, University of Manchester.

Aghion, P., Boustan, L., Hoxby, C. and Vandenbussche, J. (2009). 'The causal impact of higher education on growth: Evidence from the United States' in D. Romer and J. Wolfers (eds.), *Brookings Papers on Economic Activity: Conference Draft*. Brookings Institution, Washington, pp. 1-74.

Albertini, J., Poirier, A. and Roulleau-Pasdeloup, J. (2014). 'The composition of government spending and the multiplier at the zero lower bound', *Economics Letters*, Vol. 122, pp. 31-35.

Alfonso, A. and Jalles, J.T. (2014). 'Fiscal composition and Long-term growth', *Applied Economics*, Vol. 46, pp. 349-358.

Arnold, J., Bassanini, A. and Scarpetta, S. (2007). 'Solow or Lucas?: testing growth models using panel data from OECD countries,' *OECD Economics Department Working Paper*, No. 592, OECD.

Arnold, J., Brys, B., Heady, C., Johansson, Å., Schwellnus, C. and Vartia, L. (2011). 'Tax policy for economic recovery and growth', *Economic Journal*, Vol. 121, pp. F59-F80.

Barro R. (1990). 'Government spending in a simple model of endogenous growth', *Journal of Political Economy*, Vol. 98, pp. s103-s117.

Blankenau, W.F. and Simpson, N.B. (2004). 'Public education expenditures and growth', *Journal of Development Economics*, Vol. 73, pp. 583-605.

Bleaney, M., Gemmell, N. and Kneller R. (2001). 'Testing the endogenous growth model: public expenditure, taxation and growth over the long run', *Canadian Journal of Economics*, Vol. 34, pp. 36-57.

Bose, N., Haque, M.E. and Osborne, D. (2007). 'Public expenditure and economic growth. A disaggregated analysis for developing countries', *Manchester School*, Vol.75, pp. 533-556.

Boswijk, H. (1995). 'Efficient inference on cointegration parameters in structural error correction models', *Journal of Econometrics*, Vol.69, pp. 133-158.

Breitung, J. (2000). 'The local power of some unit root tests for panel data', in B. Baltagi, ed, *Nonstationary Panels, Panel Cointegration, and Dynamic Panels, Advances in Econometrics*, JAI Press, Amsterdam, pp 161-178.

Calderón, C., Moral-Benito, E. and Servén, L. (2014). 'Is infrastructure capital productive? A dynamic heterogeneous approach', *Journal of Applied Econometrics* (forthcoming).

Dalgaard, C-J. and Kreiner, C.T. (2003). 'Endogenous growth: A knife edge or the razor's edge?', *Scandinavian Journal of Economics*, Vol. 105, pp.73-85.

Devarajan S., Swaroop V. and Zou H. (1996).'The composition of public expenditure and economic growth', *Journal of Monetary Economics*, Vol. 37, pp. 313-344.

Eicher, T.S. and Turnovsky, S.J. (1999).‘Convergence speeds and transition dynamics in non-scale growth models’, *Journal of Economic Growth*, Vol. 4, pp. 413–428.

Futagami, K., Morita, Y. and Shibata, A. (1993). ‘Dynamic analysis of an endogenous growth model with public capital’, *Scandinavian Journal of Economics,* Vol. 95, pp. 607-625.

Gemmell, N., Kneller, R. and Sanz, I. (2011). ‘The timing and persistence of fiscal policy impacts on growth: evidence from OECD countries’, *Economic Journal*, Vol. 121, pp. F33–F58.

Gemmell, N., Kneller, R. and Sanz, I. (2014). ‘The growth effects of tax rates in the OECD’, *Canadian Journal of Economics*, forthcoming.

Haque, M.E. (2004). ‘The composition of government expenditure and economic growth in developing countries’, *Global Journal of Finance and Economics,* Vol. 1, pp. 97-117.

Harris, R.D.F. and Tzavalis, E. (1999). ‘Inference for unit roots in dynamic panels where the time dimension is fixed’. *Journal of Econometrics*, Vol. 91, pp. 201–226.

Johansen, S. (1992). ‘Cointegration in partial systems and the efficiency of single-equation analysis’, *Journal of Econometrics,* Vol. 52, pp. 389-402.

Kanas, A. and Kouretas, G.P. (2005). ‘A cointegration approach to the lead-lag effect among size-sorted equity portfolios’, *International Review of Economics and Finance*, Vol. 14, pp. 181-201.

Kneller, R., Bleaney, M. and Gemmell, N. (1999). ‘Fiscal policy and growth: evidence from OECD countries’, *Journal of Public Economics,* Vol. 74, pp. 171-190.

Lee, K., Pesaran, M.H. and Smith, R. (1997).‘Growth and convergence in a multi-country empirical stochastic Solow model’, *Journal of Applied Econometrics*, Vol. 12, pp. 357-392.

Lee, Y. and Gordon, R.H. (2005). ‘Tax structure and economic growth’, *Journal of Public Economics,* Vol. 89, pp. 1027-1043.

Myles, D. (2000). ‘Taxation and economic growth’, *Fiscal Studies*, Vol. 21, pp. 141-168.

Nijkamp, P. and Poot, J (2004). ‘Meta-analysis of the effects of fiscal policies on long-run growth’, *European Journal of Political Economy*, Vol. 20, pp. 91-124.

Padovano, F. and Galli, E. (2002).‘Comparing the growth effects of marginal vs average tax rates and progressivity’, *European Journal of Political Economy,* Vol. 18, pp. 529-544.

Pesaran, H.M. (1997). ‘The role of economic theory in modelling the long run’, *Economic Journal*, Vol. 107, pp. 178-191.

Pesaran, M. H. (2007). ‘A simple panel unit root test in the presence of cross section dependence’, *Journal of Applied Econometrics,* Vol. 22, pp. 265-312.

Pesaran M.H. and Smith, R.P. (1995). ‘Estimating long-run relationships from dynamic heterogeneous panels’, *Journal of Econometrics,* Vol. 68, pp. 79-113.

Pesaran M.H. and Shin, Y. (1999). ‘An autoregressive distributed lag modelling approach to cointegration analysis’, in S. Strom (ed.), *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium*, Cambridge University Press, Cambridge, Ch.11, pp. 371-413.

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

Romero-Avila, D. and Strauch, R. (2008). ‘Public finances and long-term growth in Europe: Evidence from a panel data analysis’, *European Journal of Political Economy,* Vol. 24, pp. 172-191.

Sanz, I. (2010). ‘What do OECD countries cut first when faced with fiscal adjustments?’, *Southern Economic Journal*, Vol. 77, pp. 753-775.

Sanz, I. and Velázquez, F.J. (2004). ‘The evolution and convergence of the government expenditure composition in the OECD countries’, *Public Choice*, Vol. 119, pp. 61-72.

Semmler, W., Greiner, A., Diallo, B., Rezai, A. and Rajaram, A. (2007). ‘Fiscal Policy, Public Expenditure Composition and Growth’, *Policy Research Working Paper*, WPS4405, World Bank.

Slemrod, J. (1995). ‘What do cross-country studies teach about government involvement, prosperity and economic growth?’, *Brookings Papers on Economic Activity*, Vol. 2, pp. 373-431.

Solow, R. M. (1994). ‘Perspectives on growth theory’, *Journal of Economic Perspectives*, Vol. 8, pp. 45-54.

Tanzi, V. and Zee, H.E. (1997). ‘Fiscal policy and long-run growth’, *IMF Staff Papers*, Vol. 44, pp. 179-209.

Teles, V.K. and Mussolini, C.C. (2014). ‘Public Debt and the Limits of Fiscal Policy to Increase Economic Growth’, *European Economic Review*, Vol. 66, pp. 1-15.

Temple, J.R.W. (1999). ‘The new growth evidence’, *Journal of Economic Literature*, Vol. 37, pp. 112-156.

Turnovsky, S.J. (2004). ‘The transitional dynamics of fiscal policy: long-run capital accumulation and growth’, *Journal of Money, Credit and Banking*, Vol. 36, pp. 883-910.

Vandenbussche, J., Aghion, P. and Meghir, C. (2006). ‘Growth, distance to frontier and the composition of human capital’, *Journal of Economic Growth*, Vol. 11, pp 97-127.

Westerlund, J. (2007). ‘Testing for error correction in panel data’, *Oxford Bulletin of Economics and Statistics*, Vol. 69, pp. 709–748.

Wickens, M.R. and Breusch, T.S. (1988). ‘Dynamic specification, the long-run and the estimation of transformed regression models’, *Economic Journal*, Vol. 98, pp. 189-205.

Wickens, T.M. (2002). ‘Classification of GFSM data to the 2001 GFSM framework’, Government Finance Statistics Manual, 2001, Companion Material, IMF.

Xing, J. (2011) ‘Does tax structure affect economic growth? Empirical evidence from OECD countries’, *Oxford University Centre for Business Taxation Working Paper* No. 11/20.

TABLE 1

Testing expenditure levels & implicit financing: pooled mean group estimates

Dependent variable: Annual GDP per capita growth rate

Regression No.: *Fiscal variables financed by:*	1 *Non-distort. taxes*	2 *Budget Surplus*	3 *Distortionary taxes*	4 *Distortionary& non-distort. taxes*	5 *Budget Surplus*
Long run effects					
Budget surplus	0.022** (2.34)	-	-0.020** (4.64)	-0.013** (3.08)	-
Distortionary taxes	-0.060** (-4.56)	-0.012** (2.84)	-	-	-0.014** (3.11)
Non-distortionary taxes	-	0.023** (4.71)	0.051** (7.78)	-	0.029** (5.27)
Total Public Expenditure	0.020** (2.17)	-0.013** (5.29)	-0.024** (6.12)	-0.012** (3.44)	-0.012** (4.94)
Investment ratio	0.029** (2.79)	-0.007** (2.41)	0.021** (3.32)	-0.004 (1.08)	-
Employment growth	0.063** (4.72)	0.020** (4.93)	0.041** (7.03)	0.039 (6.37)	0.026** (5.12)
Lagged residual	-0.039** (4.80)	-0.092** (6.16)	-0.062** (5.12)	-0.077* (6.32)	-0.092** (5.49)
Short run effects (average)					
Short run. Total Public Expenditure. 1st Difference	-0.002 (1.83)	-0.001 (1.44)	-0.001* (2.07)	-0.001 (1.57)	-0.001** (2.93)
Short run. Total Public Expenditure. 2nd Difference	-0.000 (1.00)	0.000 (1.19)	0.000 (1.14)	0.000 (0.00)	0.000 (1.18)

Note:t-statistics in parentheses below parameters; *, ** = significant at the 5%, 1% respectively.

TABLE 2

Testing expenditure levels:comparing regression methods

Dependent variable: Annual GDP per capita growth rate

Regression No.	2.1	2.2	2.3	
Financed by: Budget Surplus	Pooled Mean Group	Dynamic Fixed Effects:	Mean Group:	Hausman Tests
Distort. tax & other revenues	-0.012** (2.84)	-0.015 (1.71)	-0.018 (0.74)	$\chi^2(1)$ 0.08 p-value 0.78
Non-distortionary taxes	0.023** (4.71)	0.024 (1.72)	0.094 (1.50)	$\chi^2(1)$ 1.30 p-value 0.25
Total public expenditure	-0.013** (5.29)	-0.007 (1.28)	-0.020 (1.65)	$\chi^2(1)$ 0.36 p-value 0.55
MG versus PMG				$\chi2(5)$: 7.20 p-value 0.21
PMG versus DFE				$\chi2(5)$:16.51 p-value 0.00

Note:t-statistics in parentheses below parameters; *, ** = significant at the 5%, 1% respectively. All regressions include the three non-fiscal control variables shown in Table 1. All coefficients are estimated long-run effects

TABLE 3

Testing public expenditure composition: pooled mean group estimates

Dependent variable: Annual GDP per capita growth rate

Regression:	1	2	3	4	5	6	7	8	9
Share of:	*T&C*	*Educat-ion*	*Health*	*Hous-ing*	*Social welfare*	*Defence*	*Econ. Serv.*	*Gen. pub. Serv.*	*Recreat--ion*
Expenditure SHARE	0.022** (9.47)	0.020** (4.92)	0.001 (0.80)	0.009 (1.56)	-0.001 (1.00)	-0.001 (0.57)	-0.005 (1.35)	-0.005** (2.86)	-0.005 (0.30)
TOTAL Expenditure	0.003 (1.80)	-0.014** (5.24)	-0.006 (1.90)	-0.016** (6.09)	-0.008 (4.00)	-0.004* (2.32)	-0.014** (3.17)	-0.011** (5.86)	-0.007 (1.99)
Short run effects (average)									
Expenditure SHARE (1STdif)	0.000 (0.40)	-0.001 (1.29)	0.000 (0.22)	-0.000 (0.04)	-0.001* (2.21)	-0.001 (0.61)	0.001 (0.67)	0.000 (0.38)	0.005 (1.23)
Expenditure SHARE (2NDdif)	0.001 (1.08)	-0.001 (1.42)	-0.000 (0.10)	-0.000 (0.17)	-0.000 (0.63)	0.001 (1.44)	0.001 (1.37)	0.001 (1.47)	-0.001 (1.00)

Notes: (1) All regressions in the Table take the same form as regression 2 in Table 1, augmented to include each individual expenditure share. That is, *included variables* are total public expenditure; distortionary taxes & ‘other’ revenues; non-distortionary taxes; investment ratio; employment growth, lagged GDP per capita; *excluded variable* are Budget surplus.(2) *t*-statistics in parentheses below parameters; *, ** = significant at the 5%, 1% respectively.

TABLE 4

Testing public expenditure composition under alternative lag structures

Regression:	1’	2’	3’	4’	5’	6’	7’	8’	9’
Share of:	*T&C*	*Education*	*Health*	*Hous-ing*	*Social welfare*	*Defence*	*Econ. Serv.*	*Gen. pub Serv.*	*Recreat--ion*
				ARDL (2, 2) model					
Expenditure SHARE	0.003** (2.58)	0.004** (2.48)	0.012** (3.64)	0.037** (4.95)	-0.008** (3.90)	-0.007** (10.10)	-0.001 (0.37)	0.013** (3.97)	0.029* (2.12)
TOTAL Expenditure	-0.012** (7.19)	-0.012** (6.65)	-0.015** (5.66)	-0.016** (4.81)	-0.004 (1.88)	-0.002** (2.77)	-0.014** (6.17)	-0.003 (0.74)	-0.015** (7.19)
			ARDL (p, q) lag structure (p, m ≤ 2) chosen by Schwarz criterion (Table 3)						
Expenditure SHARE	0.022** (9.47)	0.020** (4.92)	0.001 (0.80)	0.009 (1.56)	-0.001 (1.00)	-0.001 (0.57)	-0.005 (1.35)	-0.005** (2.86)	-0.005 (0.30)
TOTAL Expenditure	0.003 (1.80)	-0.014** (5.24)	-0.006 (1.90)	-0.016** (6.09)	-0.008 (4.00)	-0.004* (2.32)	-0.014** (3.17)	-0.011** (5.86)	-0.007 (1.99)

TABLE 5

Weak exogeneity tests for fiscal and investment variables

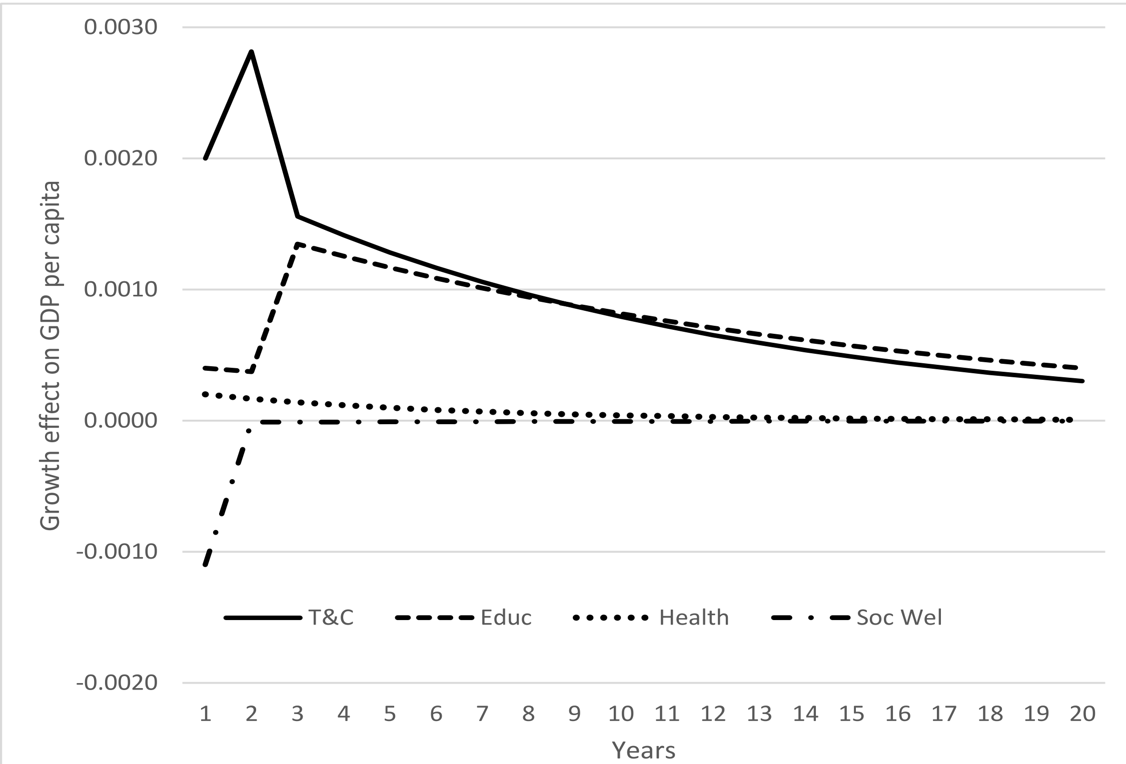
	Expenditure share	*Total expenditure*	*Distortionary & other taxes*	*Non-distortionary taxes*	*Investment ratio*
Table 1: regression 2	-	-0.928 (0.99) 0 countries	-3.429 (2.08)** 2 countries	-1.110 (1.01) 0 countries	-0.868 (0.74) 2 countries
Table 3 regressions:					
T&C	1.747 (2.46)** 1 country	0.522 (0.21) 2countries	-2.911 (1.99)** 1 country	0.714 (0.55) 1 country	0.756 (0.72) 1 country
Education	-0.735 (1.03) 0 countries	-2.382 (2.04)** 2 countries	-1.790 (1.79) 1 country	0.227 (0.30) 1 country	-1.013 (0.90) 1 country
Health	-0.948 (1.29) 0 countries	-3.901 (2.71)*** 2 countries	-3.023 (2.75)*** 4 countries	-1.002 (1.59) 4 countries	1.589 (1.56) 2 countries
Housing	0.427 (0.62) 1 country	-2.523 (2.19)** 0 countries	-3.685 (2.18)** 1 country	-1.675 (1.40) 1 country	-0.046 (0.06) 0 countries

Social Welfare	-0.435 (0.12) 4 countries	-3.294 (1.61) 0 countries	-1.356 (1.40) 1 country	0.535 (0.95) 1 country	-1.753 (0.88) 3 countries
Defence	0.179 (0.20) 1 country	-0.482 (0.21) 2 countries	-3.397 (1.47) 1 country	-2.328 (1.56) 2 countries	0.073 (0.05) 0 countries
Economic Services	-1.973 (2.06)** 2 countries	-3.489 (3.29)*** 2 countries	-2.270 (2.57)** 4 countries	-0.651 (1.38) 1 country	1.333 (2.06)** 3 countries
General Public Services	-0.669 (0.54) 1 country	-3.402 (1.94)* 1 country	-2.369 (1.52) 1 country	-0.175 (0.22) 2 countries	3.068 (3.50)*** 2 countries
Recreation	-0.124 (0.80) 1 country	-3.189 (2.46)*** 1 country	-2.782 (2.52)*** 3 countries	-0.609 (0.83) 4 countries	2.003 (1.71) 3 countries

Notes: *t*-statistics in parentheses below parameters. . *,**, *** = significant at the 10%, 5% and 1% level. Terms such as "2 countries" refers to the number of country-specific weak exogeneity tests where weak exogeneity is rejected at the 5% level.

FIGURE 1

Growth effect of a 1 percentage point change in the expenditure mix (in year 0)



Note: Based on the results from Table 3

Appendix A: The updated OECD dataset

The dataset used in this paper builds on that used by Bleaney *et al.* (2001), who used GFS fiscal data, covering consolidated *central* government functions only, based on the 1986 GFS Manual classification of fiscal variables (labeled '*old*' below). Like much National Accounting at that stage, these variables were measured based on a 'cash', as opposed to 'accruals', accounting method. We refer to this below as the '*old*' classification. The 2001 GFS Manual introduced a '*new*' classification system (mainly involving the reclassifying of other expenditures into general public services, and separating environmental protection from housing; see Wickens 2002). In line with new National Accounting practice, the 'new' GFS is based on accruals accounting and so is not directly comparable with the original Bleaney *et al.* dataset. In addition GFS data for central government on a cash basis has not generally been updated beyond about 1999 or 2000 for most countries in our sample. The most recent data available (typically updated to 2007 or 2008), based on the new classification and accrual accounting, is available for central and general (central plus local) government but has only been back-dated to 1990.

Annual differences between fiscal variables measured on cash or accrual bases can be quite substantial. For example, the financial year in which corporation tax (cash) payments are made in many OECD countries can be different by up to 2-3 years from the (accrual) accounting period to which the tax liability relates. As a result, up-dating our dataset beyond around 2000 requires a careful splicing of 'old' and 'new' data streams and is likely to involve a number of inaccuracies of unknown magnitude.

The currently available data is, in summary:(i) the latest GFS data on a cash basis for central government to update Bleaney *et al.*(typically to 1999 or 2000) and then (ii) the annual rate of change in 'new' fiscal variables for central government to update the series to the latest possible year. In some cases, where overlaps in the series suggest that the new and old GFSY do not correspond well, we supplement this with OECD sourced data which is based on a similar definition to the new GFS. Though in principle we would prefer to use a dataset capturing all levels of government, the unavailability of data on this basis prior to 1990 or 1995 would leave us with insufficient time-series observations. The updated dataset includes data from the early 1970s to 2007 or 2008 for 17 countries: Australia, Austria, Canada, Denmark, Finland, France, Germany, Iceland, Luxembourg, Netherlands, New Zealand, Norway, Spain, Sweden, Turkey, UK, USA. Other OECD countries are excluded due to gaps in their data series, usually for fiscal variables. For example, Central and Eastern European Countries (Czech Republic, Hungary, Poland and Slovak Republic) provide data starting only in the mid-nineties. The rest of the excluded OECD countries have missing values for several mid-series years: Belgium (1989-95), Greece (1982-91), Ireland (1970-81), Italy (1989-95), Mexico (1972-79; 2001 onwards), Portugal (1989-97) and Switzerland (1985-1990). Japan only has only data for 1990-93; 1995-98 and 2001-05.

Appendix B: Additional results

APPENDIX TABLE B1

Mean group parameter estimates for total public expenditure by country

Country	Parameter (*t*-value)	Country	Parameter (*t*-value)
Australia	0.073 (0.94)	Norway	-0.012 (5.01)**
Austria	-0.021 (2.22)	Spain	-0.002 (0.16)
Canada	0.011 (1.04)	Sweden	-0.198 (0.12)
Denmark	0.019 (4.70)**	Turkey	-0.019 (2.88)**
Finland	-0.005 (1.99)*	United Kingdom	0.003 (1.75)
France	-0.024 (2.34)*	United States	-0.005 (0.93)
Iceland	-0.001 (0.14)	Germany	-0.009 (0.10)
Luxembourg	-0.009 (0.78)	New Zealand	-0.058 (0.54)
Netherlands	-0.012 (2.28)*		
Mean long run	-0.020 (1.65)		
Median long run	-0.009		
Mean short run (1^{st} Difference)	-0.001 (0.90)		
Median short run (1^{st} Difference)	0.000		
Mean short run (2^{nd} Difference)	0.000 (0.20)		
Median short run (2^{nd} Difference)	0.000		
Observations	590		

Notes: *t*-statistics in parentheses; * and** = significant at the 5% and 1% levels respectively.

APPENDIX TABLE B2

Testing public expenditure composition: comparing regression methods

Regression:	1	2	3	4	5	6	7	8	9
Share of:	*T&C*	*Educat-ion*	*Health*	*Hous-ing*	*Social welfare*	*Defence*	*Econ. services*	*Gen pub. services*	*Recreat-ion*
Pooled Mean Group									
Expenditure share	0.022**	0.020**	0.001	0.009	-0.001	-0.001	-0.005	-0.005**	-0.005
	(9.47)	(4.92)	(0.80)	(1.56)	(1.00)	(0.57)	(1.35)	(2.86)	(0.30)
Total expenditure	0.003	-0.014**	-0.006	-0.016**	-0.008	-0.004*	-0.014**	-0.011**	-0.007
	(1.80)	(5.24)	(1.90)	(6.09)	(4.00)	(2.32)	(3.17)	(5.86)	(1.99)
Dynamic Fixed Effects									
Expenditure share	-0.001	-0.002	0.003	-0.005	0.002	-0.006	-0.000	-0.003	0.004
	(0.15)	(0.35)	(0.97)	(0.63)	(0.94)	(0.97)	(0.07)	(1.10)	(0.21)
Total expenditure	0.001	-0.002	-0.002	-0.001	-0.001	-0.002	-0.002	-0.002	-0.002
	(0.23)	(0.49)	(0.46)	(0.24)	(0.34)	(0.53)	(0.37)	(0.39)	(0.42)
Mean Group									
Expenditure share	-0.084	-0.029	0.019	-0.040	0.105	-0.040	-0.006	-0.008	-0.732
	(0.89)	(0.88)	(0.78)	(1.44)	(0.94)	(1.32)	(0.63)	(1.19)	(0.87)
Total expenditure	-0.052	-0.026	-0.016	0.015	-0.285	-0.022	-0.001	-0.011	-0.012
	(1.00)	(1.75)	(1.44)	(0.76)	(1.06)	(1.48)	(0.07)	(1.49)	(1.16)
<u>Hausman tests:</u>									
MG versus PMG									
Expenditure share	$\chi^2(1)$ 1.24 p 0.27	$\chi^2(1)$ 2.23 p 0.14	$\chi2(1)$ 0.53 p 0.47	$\chi^2(1)$ 3.28 p 0.07	$\chi^2(1)$ 0.90 p 0.34	$\chi^2(1)$ 1.69 p 0.19	$\chi^2(1)$ 0.01 p 0.92	$\chi^2(1)$ 0.20 p 0.65	$\chi^2(1)$ 0.75 p 0.39
Total expenditure	$\chi^2(1)$ 1.12 p 0.29	$\chi^2(1)$ 0.64 p 0.42	$\chi2(1)$ 0.85 p 0.36	$\chi^2(1)$ 2.57 p 0.11	$\chi^2(1)$ 1.06 p 0.30	$\chi^2(1)$ 1.54 p 0.21	$\chi^2(1)$ 6.12 p 0.01	$\chi^2(1)$ 0.01 p 0.92	$\chi^2(1)$ 0.31 p 0.58
All variables	$\chi2(6)$ 7.14 p 0.31	$\chi2(6)$ 6.97 p 0.32	$\chi2(6)$ 5.79 p 0.45	$\chi2(6)$ 7.50 p 0.28	$\chi2(6)$ 7.09 p 0.31	$\chi2(6)$ 8.73 p 0.19	$\chi2(6)$ 5.93 p 0.43	$\chi2(6)$ 4.88 p 0.56	$\chi2(6)$ 6.40 p 0.38
PMG versus DFE									
All variables PMG v DFE	$\chi2(6)$ 45.5 p 0.00	$\chi2(6)$ 53.2 p 0.00	$\chi2(6)$ 51.2 p 0.00	$\chi2(6)$ 58.3 p 0.00	$\chi2(6)$ 18.1 p 0.01	$\chi2(6)$ 42.51 p 0.00	$\chi2(6)$ 14.4 p 0.03	$\chi2(6)$ 17.4 p 0.01	$\chi2(6)$ 55.9 p 0.00

Notes: (1) All regressions take the same form as regression **3** in Table 1, augmented to include each individual expenditure share. That is,*included variables* = total public expenditure; distortionary; non-distortionary taxes; investment ratio; employment growth, lagged residual;*excluded variable* = Budget surplus.(2) p = p-value; *t*-statistics in parentheses below parameters; *, ** = significant at the 5%, 1% respectively. All coefficients are estimated long-run effects

APPENDIX TABLE B3

Panel Unit Root Tests

	Harris-Tzavalis (1999)	Breitung (2000)	Pesaran (2007)
	(H_0: Panels contain unit roots; H_a: Panels are stationary)		
Distortionary taxes	ρ 0.9860 p-value 1.000	λ -1.0962 p-value 0.136	Z_{tbar} -0.495 p-value 0.310
Investment	ρ 0.9141 p-value 0.999	λ 0.2616 p-value 0.603	Z_{tbar} -0.860 p-value 0.195
Total spending	ρ 0.9891 p-value 1.000	λ -0.1800 p-value 0.428	Z_{tbar} 2.251 p-value 0.988
Non-distortionary taxes	ρ 0.9950 p-value 1.000	λ -1.1298 p-value 0.129	Z_{tbar} 0.619 p-value 0.732
Deficit	ρ 0.7803 p-value 0.000	λ -3.0396 p-value 0.001	Z_{tbar} -0.721 p-value 0.236
Log of GDP per capita	ρ 1.0064 p-value 1.000	λ -0.9136 p-value 0.180	Z_{tbar} -0.929 p-value 0.176
Employment growth	ρ 0.2723 p-value 0.000	λ -3.7036 p-value 0.000	Z_{tbar} -6.365 p-value 0.000
T&C	ρ 0.9477 p-value 1.000	λ 0.1201 p-value 0.547	Ztbar 0.634 p-value 0.737
Education	ρ 0.9956 p-value 1.000	λ -0.6984 p-value 0.242	Ztbar -0.496 p-value 0.310
Health	ρ 0.9977 p-value 1.000	λ -0.3194 p-value 0.374	Ztbar 2.840 p-value 0.998
Housing	ρ 0.8981 p-value 0.999	λ -1.8060 p-value 0.036	Ztbar 1.201 p-value 0.885
Social Welfare	ρ 0.9823 p-value 1.000	λ -0.4212 p-value 0.336	Ztbar 0.883 p-value 0.811
Defence	ρ 0.9857 p-value 1.000	λ -0.9910 p-value 0.161	Ztbar -3.178 p-value 0.001
Economic Services	ρ 0.9288 p-value 1.000	λ -2.0372 p-value 0.021	Ztbar 1.893 p-value 0.971
Gen pub. Services	ρ 0.9439 p-value 1.000	λ -0.9533 p-value 0.170	Ztbar -0.074 p-value 0.471
Recreation	ρ 1.0101 p-value 1.000	λ 1.4533 p-value 0.926	Ztbar 2.234 p-value 0.987

Cointegration Tests

Test	*Robust p-value for GDP, fiscal variables and investment*	*Robust p-value among regressors (fiscal variables and investment)*
P_t	*0.02*	*0.60*
P_a	*0.13*	*0.93*

Cointegration test for Estimation 1.2 in Table1

APPENDIX TABLE B4

Weak exogeneity tests for ARDL (2,2) model

	Expenditure share	*Total expenditure*	*Distortionary & other taxes*	*Non-distortionary taxes*	*Investment ratio*
Estimation 1.2		-1.083 (1.33) 0 countries	1.751 (1.05) 2 countries	1.411 (3.01)** 0 countries	0.100 (0.09) 1 country
T&C	-1.986 (1.02) 2 countries	4.097 (0.89) 1country	-0.100 (0.04) 2 countries	1.600 (1.36) 1 country	1.449 (0.81) 1 country
Education	-0.797 (1.33) 0 countries	-1.145 (0.76) 2 countries	0.637 (0.42) 1 country	3.156 (1.67) 4 countries	0.392 (0.45) 1 country
Health	2.713 (1.86) 0 countries	-3.202 (1.61) 1 country	-1.348 (1.13) 2 countries	0.884 (0.62) 2 countries	0.064 (0.05) 3 countries
Housing	-1.851 (1.70) 3 countries	0.653 (0.32) 2 countries	1.897 (1.50) 3 countries	1.759 (2.39)* 3 countries	2.562 (2.03)* 2 countries
Social Welfare	-7.567 (1.91) 3 countries	1.239 (0.54) 2 countries	-0.070 (0.08) 0 countries	0.420 (0.53) 2 countries	2.429 (1.66) 1 country
Defence	-1.128 (2.49)* 1 country	1.275 (1.45) 2 countries	-1.185 (1.46) 4 countries	0.319 (0.51) 3 countries	-0.340 (0.54) 0 countries
Economic Services	5.117 (1.88) 2 countries	-1.873 (0.60) 0 countries	0.831 (0.54) 2 countries	3.001 (3.12)* 2 countries	1.553 (0.86) 1 country
General Public Services	-5.981 (0.75) 2 countries	-1.720 (0.56) 0 countries	3.691 (1.43) 1 country	-3.292 (1.13) 0 countries	1.868 (2.60)* 0 countries
Recreation	-0.153 (0.28) 0 countries	2.462 (0.69) 0 countries	0.701 (0.48) 2 countries	1.852 (1.21) 3 countries	0.849 (0.87) 1 country

Notes: *t*-statistics in parentheses below parameters. * = significant at the 5% level. Terms such as “2 countries” refers to the number of country-specific weak exogeneity tests where weak exogeneity is rejected at the 5% level.

About the Authors

Norman Gemmell is Professor of Public Finance at Victoria Business School, Victoria University of Wellington, New Zealand.
Email: norman.gemmell@vuw.ac.nz

Richard Kneller is Research Fellow in the Centre for Globalisation and Economic Policy and a professor of Economics at the University of Nottingham.
Email: richard.kneller@nottingham.ac.uk

Ismael Sanz is Professor of the Department of Applied Economics I at Universidad Rey Juan Carlos.
Email: ismael.sanz@urjc.es

VICTORIA
UNIVERSITY OF WELLINGTON
TE WHARE WĀNANGA
O TE ŪPOKO O TE IKA A MĀUI

Chair in Public Finance
Victoria Business School

Working Papers in Public Finance