

Do Smaller Governments Raise the Level or Growth of Output? A Review of Recent Evidence^{*}

by NORMAN GEMMELL^{**} and JOEY AU^{***}

Abstract

Theoretical developments, improved methodologies and more extensive data have helped generate a dramatic increase in the literature testing for the impact of government size and fiscal policy on economic growth in recent years. We review a range of the more recent evidence and examine (1) the consistency or robustness of the results; (2) how these results differ from the earlier literature and (3) their usefulness as a guide to policy reform in practice. We find that the last decade has produced more robust evidence and more plausible orders of magnitude on the impact of fiscal policy on growth. However, the value of this evidence remains limited as a basis for quantifying macro-economic responses to fiscal policy reform in practice. (JEL: H10; H30; O40; O50. Keywords: *fiscal policy; government expenditure; taxation; economic growth*)

Acknowledgements

We are grateful for helpful comments on an earlier draft of this paper from Professor Noel Gaston and participants at the First Annual Workshop, on *Economic and Policy Developments in East Asia*, held at Bond University, Gold Coast, Australia (October 2011).

1. Introduction

Over the past 50 years the number of countries in the OECD has expanded beyond the original set of 20 countries in Europe and North America to include such newer

* This article is a revised version of ‘Government size, fiscal policy and the level and growth of output: A review of recent evidence’ published in the *Journal of the Asia-Pacific Economy*, 18 (2), 203–229. That paper had a more Asia-Pacific focus. We are grateful to the JAPE Special Issue editor (Noel Gaston) and publisher (Taylor & Francis) for permission to re-use material here.

** Victoria Business School, Victoria University of Wellington, New Zealand. Corresponding author. Email: Norman.Gemmell@vuw.ac.nz. The views expressed in this article are those of the authors and do not necessarily represent those of the New Zealand Treasury.

*** Chief Economist’s Unit, The Treasury, New Zealand.

members as the Czech and Slovak Republics, Chile, Korea and Mexico. Membership currently stands at 34 with many of those newer members having lower per capita income levels than longer established members but experiencing faster growth and, in some cases, overtaking the income levels of those original members. In joining the OECD, members sign up to Article 1 which states the aim of the OECD is “to promote policies designed ... to achieve the highest sustainable economic growth and employment and a rising standard of living in Member countries”.¹

In these largely market-driven mixed economies a key policy area of interest, in terms of its impact on income levels and growth rates, has been the extent and nature of government intervention. In particular, interest has focused on the twin questions of whether economic prosperity is fostered by larger or smaller governments, and by more interventionist or more *laissez faire* government policies. As a consequence, policy advice to OECD governments has often hinged on perceived answers to these questions. For Scandinavian economies, for example, the 1990s saw considerable debate over the long-run growth consequences of government intervention and the appropriate policy advice (see, for example, KORPI 1996, AGELL, LINDH and OHLSSON 1997).

Similarly, the rapid growth in the second half of the 20th century of numerous developing countries, initially in the Asian region – the so-called ‘Asian economic miracle’ – stimulated debate over the role of government in that process; see, for example, YOUNG (1992), RODRIK (1994). Much evidence on these ‘government and growth’ issues has since been collected and analysed, often distinguishing between high and low income, or high and low growth, economies. This paper provides an update on this research on OECD countries which has expanded rapidly in recent years.

There have, of course, been a number of reviews of the empirical literature on the impact of government size on income levels and growth rates, of which SLEMROD (1995) is probably the most comprehensive.² His review largely focused on cross-country evidence at the aggregate or macro-level up to the mid-1990s. The question addressed by SLEMROD (1995, p. 373) was simple: what is “the evidence about the influence of government tax and expenditures on economic prosperity and growth”?³ In answering this question, SLEMROD (1995) was generally sceptical of the evidence, arguing that:

“...the empirical findings are not robust to various reasonable specifications and, of most concern, do not address identification problems...” (p. 380) and “[t]his review of existing cross-country literature suggests that there is no persuasive evi-

¹ See <http://www.oecd.org/general/conventionontheorganisationforeconomicco-operationand-development.htm>.

² See also MYLES (2000; 2009a, b, c).

³ The distinction here between effects on income ‘levels’ or ‘growth rates’ is essentially a distinction between short-run, *temporary* effects on income growth which lead to permanent changes in income *levels*, versus persistent effects on income growth that thus have ‘permanent’ (or at least long-lasting) effects on income *growth rates*.

dence that the extent of government has either a positive or negative impact on either the level or the growth rate of per capita income... top-down studies find a negligible effect” (p. 401).

Much has changed since SLEMROD’s review, with a variety of traditional and new methodologies subsequently being applied to more extensive and different datasets. In particular, within the so-called ‘top-down’ or aggregate/macro-level approaches, cross-section studies have largely given way to panel analyses across or within countries, and country-specific time-series evidence. This is the motivation for the current review which returns to SLEMROD’s original question. Specifically, to help answer it we pose three more detailed questions.

- (1) How robust or consistent are the results from more recent studies?
- (2) Do the results from these studies justify a less sceptical verdict than emerged from SLEMROD’s review? and
- (3) Are these results sufficient to be useful as a guide to fiscal policy reform in practice?

To summarise our assessment, we argue that theory on the output effects of fiscal policy has developed considerably since the time of SLEMROD’s writing and this theory has increasingly been taken seriously in specifying empirical tests. Together with improvements in econometric methods and new datasets, these developments have allowed better interrogation of panels of data. Our conclusion is that, despite some remaining ‘big issues’, results from aggregate-level studies are now much more robustly in favour of identifiable effects of government size on income levels and growth rates, provided these conclusions are carefully circumscribed in ways that we will make clear. In addition, plausible orders of magnitude are beginning to emerge, though their value remains limited as an evidence-base for quantitative predictions of the output effects from fiscal policy reforms in practice.

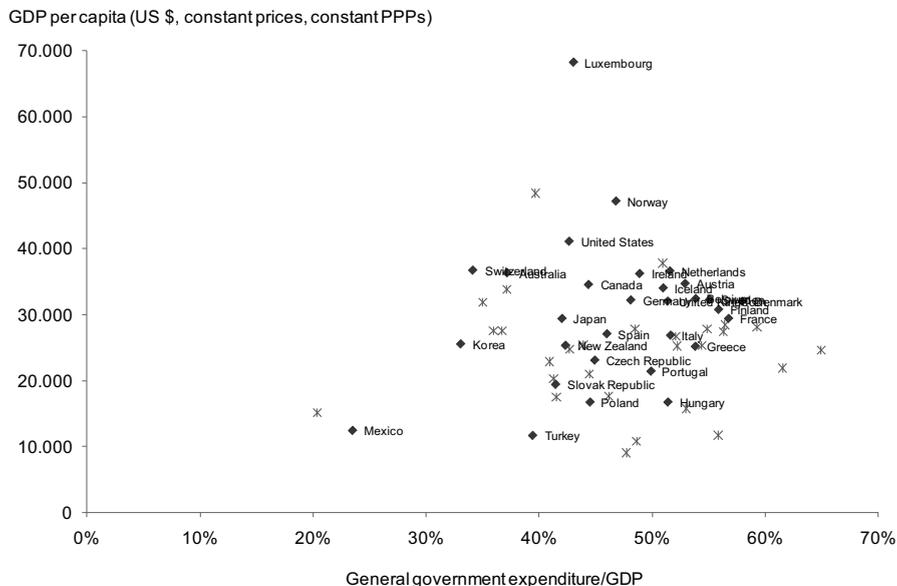
2. Reasons to be sceptical of the evidence

SLEMROD (1995) was sceptical of the cross-section evidence available up to the early 1990s, largely for two reasons. First, while most arguments at the time proposed that bigger governments were harmful for income levels (‘prosperity’ in SLEMROD’s terms) or growth rates, unconditional correlations across countries, and trends within countries over time, typically revealed *positive*, rather than negative, correlations or were unclear. Second, SLEMROD argued that, based on analysis of simple but plausible conceptual relationships among the variables involved, cross-section evidence could not be expected to reveal the empirical relationships sought by studies based on macro-level data. In this and the next section we explore each of those arguments in turn.

To begin, it is helpful to be reminded of the evidence on which Slemrod’s conclusions were based. He examined cross-plots of country levels of GDP per capita in 1990 with country ratios of government expenditure to GDP, G/Y, and tax revenue to

GDP, T/Y, in 1990. He considered a large sample of developing and developed countries and an OECD sub-sample; see SLEMROD (1995, Figures 3–6). SLEMROD found a wide array of observations in the GDP per capita, G/Y, space with a likely positive but weak association across all countries. For OECD countries, there was probably less support for a positive association between GDP per capita and G/Y but any attempt to fit a linear relationship to those variables appeared likely to be sensitive to outlying observations such as Japan, Turkey and Greece. Similar conclusions emerged from SLEMROD’s examination of cross-country patterns for GDP per capita versus the tax ratio, T/Y.

To replicate and update SLEMROD’s (1995) analysis, Figure 1 plots cross-country GDP per capita levels against G/Y for a sample of 30 OECD, using values for 1995 and 2009. It is clear from the data that the earlier patterns (or lack of) identified by SLEMROD (1995) are essentially unaltered by an updated dataset.⁴ If instead the general government (total) *tax revenue* to GDP ratio is examined against GDP per capita levels across the same OECD sample (not shown), it proves similarly difficult to



Note: country names for 1995 ‘x’ observation names have been omitted for clarity.
 Source: OECD National Accounts.

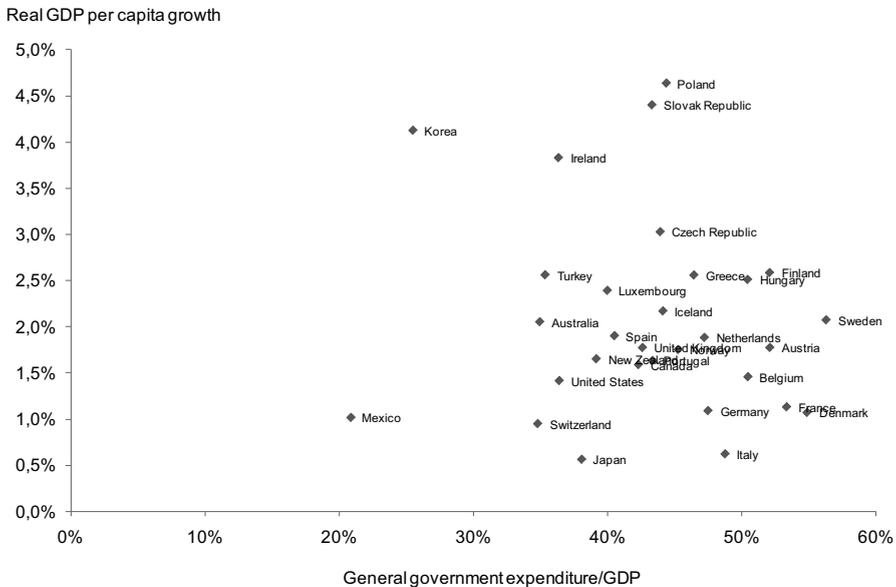
Figure 1: Government expenditure ratio versus GDP per capita: OECD countries, 1995 and 2009.

⁴ Similar patterns are found if 2007, rather than 2009, is used suggesting that the observed patterns are not specific to the post-2008 global recession period. Furthermore, averaging the data over the 1995-2009 period reveals similar patterns.

identify a positive or negative association.⁵ At a minimum, it does not confirm the priors of those who expect higher taxes to be associated with *lower* prosperity.

SLEMROD (1995) also states that he is “not aware of any serious academic study that purports to demonstrate a significant negative causal relationship between the extent of government involvement and the level of prosperity. There are, however, studies that purport to show a negative, and presumably causal, relationship between measures of government involvement and the *growth rate* of real per capita income...” (p. 377, emphasis added). The evidence on this univariate “growth rate/government size” association is also “hardly striking” (p. 378) according to SLEMROD’s 1970–1990 growth data.

For this case there was more, though still limited, support for a negative association across OECD countries than when a wider sample of countries was included. Figure 2 provides an update of the evidence for the OECD. This also appears to show no clear association between a country’s average annual growth rate of real per capita

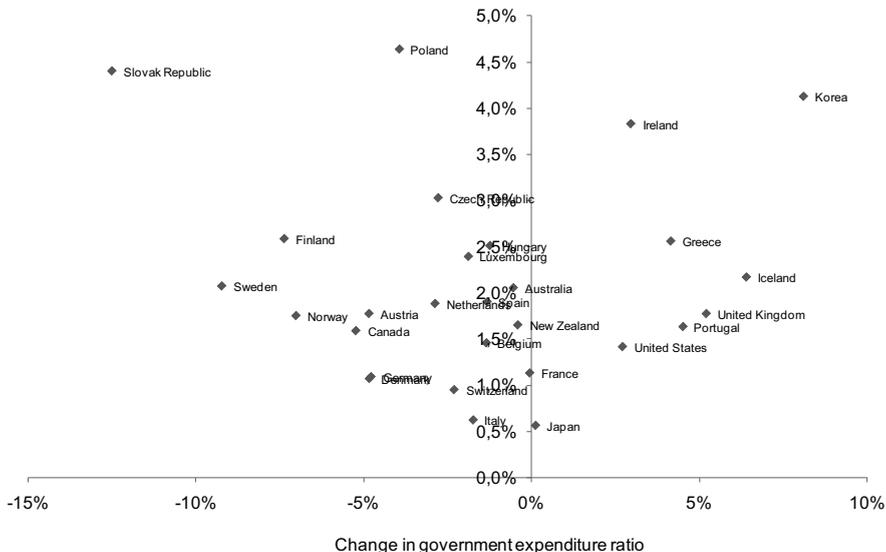


Source: OECD National Accounts, authors’ calculations.

Figure 2: Government expenditure ratio versus average real per capita GDP growth rate 1995–2009: OECD countries.

⁵ For the same sample of OECD countries, the T/Y ratio is heavily concentrated within a very small interval, largely falling within 15% to 30%.

Real GDP per capita growth



Source: OECD National Accounts, authors' calculations.

Figure 3: Change in government expenditure ratio 1995–98 to 2006–09 versus average real per capita GDP growth rate 1995–2009: OECD countries.

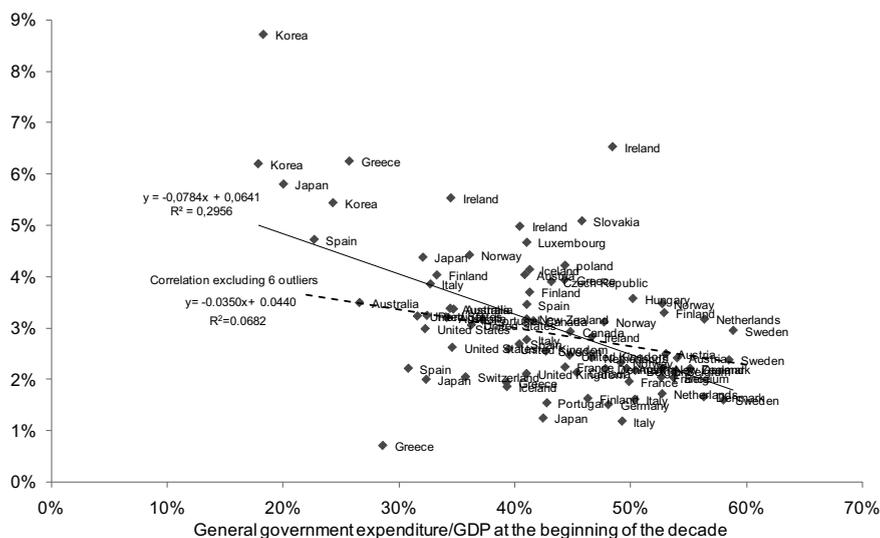
GDP over 1995–2009 and its average ratio of G/Y over the same period. These conclusions hold if T/Y is substituted for G/Y in Figure 2.⁶

Some modelling approaches, such as the growth accounting framework, predict a relationship between *changes* in government size and income growth rates. Figures 3 and 4 examine changes in G/Y and T/Y respectively for our sample of OECD countries, considering changes between average values of government size during 1995–1998 and during 2006–2009. The changes in G/Y and T/Y (in percentage points over the period) are plotted against annual per capita real GDP growth over 1995–2009. Once again, observations are widely dispersed and associations between these variables are unclear. Notably, a number of countries with very similar growth rates display large differences in the change (sign and magnitude) of the government share variables. While there is a suggestion in Figure 4 of a negative association between changes in T/Y and per capita GDP growth, the relationship is likely to be sensitive to outlying countries like the Slovak Republic.⁷

⁶ Statistically, weak negative associations are obtained in both cases, not significant at the 5% level.

⁷ The scatter in Figure 4 does reveal a statistically significant (at 5%) negatively signed relationship between the two variables.

Average growth in GDP over decade (geometric mean)



Source: OECD Economic Outlook database, NEW ZEALAND TREASURY (2011) and authors' calculations.

Note: Not all countries have a data point for each of the four decades due to data availability.

Figure 5: Government expenditure ratio versus average real GDP growth rate: OECD countries.

3. What relationships could be expected?

The theory on the relationship between government size and long-run income levels or growth rates has developed a great deal since the early 1990s, with the seminal paper by BARRO (1990) acting as a catalyst for much subsequent theoretical work. It is beyond the scope of the present paper to review this voluminous literature but it is useful to consider the relationship between output growth and government size that the BARRO (1990) and similar models predict. BARRO distinguishes between two types of public expenditures: those that are either 'productive' (enter private sector production functions) or are 'unproductive' (enter agents' utility functions). These expenditures must be financed by two types of taxation: those (such as income taxes) that are distortionary with respect to investment decisions, including human capital

when just the six observations (7.6% of the sample) on the extreme left of the Figure are omitted: 3 for Korea; 1 each for Japan, Greece and Spain. The regression fit (R^2) also falls from 0.30 to 0.07. This is not intended to suggest that such observations *should* be omitted but merely to demonstrate the sensitivity to inclusion/omission of small numbers of extreme observations.

investment, and taxes which do not distort investment decisions.¹⁰ Based on these distinctions, BARRO (1990) yields the decomposition of predicted long-run growth effects associated with the combinations of taxes and expenditures shown in Table 1.

Table 1: Predicted long-run growth effects: BARRO (1990) model

Financed by an increase in:		Public Spending:	
		Productive	Unproductive
Taxes	Distortionary	Positive (negative) at low (high) government size	Negative
	Non-distortionary	Positive	Zero

The table shows that the predicted long-run growth effect on output depends not only on the type of public spending but also the type of tax used to finance that spending. By recognising that, whatever the impact of public spending on output levels may be, the financing of that spending can also have output effects, these models formalise the role of the government budget constraint (GBC). More recent theoretical models tend to reinforce these conclusions.¹¹

The role of the GBC in these models has important implications for empirical testing because, as Table 1 demonstrates, the *combination* of particular tax and public spending categories matters for predicted growth outcomes. Therefore *which* taxes or expenditures are included in, or excluded from, regression models seeking to explain output levels or growth rates, can be expected to affect parameter estimates. However, many empirical tests for government-output effects have ignored this GBC aspect. Indeed, even where the GBC is recognised as important in principle for empirical testing, implementing it is not always straightforward.

Perhaps the most controversial aspect of aggregate level studies of government-growth effects is how to model the likely endogenous relationships involved. That is, as well as government spending or taxation potentially impacting on income levels or growth rates, higher income levels may impact on government size – for example, via increasing demand for government services or transfers at higher income levels. This is at the heart of SLEMROD'S (1995) criticism of the 'identification problems' with this literature, quoted above. In the following sub-section we use a simple illustration to demonstrate the ambiguous expected relationships that can arise.

¹⁰ This latter category would include consumption taxes in the absence of labour supply effects.

¹¹ See, for example, the endogenous growth models of PERETTO (2003, 2007). Similar 'transitional' (as opposed to steady-state) growth effects of fiscal policy are obtained from neoclassical models; see, for example, TURNOVSKY (2004).

3.1 A simple model of government and growth

To fix ideas, consider a simple linear case in which government expenditure is income elastic – demands for the goods and services delivered via public expenditure rise faster than income, Y .¹² This demand relationship can be represented by equation (1) below where t indexes time, $g = G/Y$ and a, b are parameters.

$$(1) \quad g_t = \left(\frac{G}{Y}\right)_t = a + bY_t$$

The parameter b is positively signed if government expenditure is income-elastic such that G changes more than proportionately in association with a given percentage change in income, Y . In the absence of deficit financing, the GBC requires that all expenditure in each period is financed from tax revenues such that $g_t = r_t$, where r_t is the average tax rate, T/Y .

For simplicity assume a tax system where tax revenues rise with incomes.¹³ Where the tax system is progressive (average tax rates rise with income levels), as in most OECD countries, this tends to yield a relationship between tax revenue, R , and income, such that both effective marginal and average tax rates rise with incomes.¹⁴ At the aggregate level we can specify this as a reduced form relationship in which the effective marginal tax rate, m_t , is proportional (α) to the effective average tax rate, r_t .¹⁵ Hence we can write:

$$(2) \quad r_t = \left(\frac{R}{Y}\right)_t = \alpha m_t$$

where $0 < \alpha < 1$ in a progressive tax system. Since $g_t = r_t$, then, from equation (1), both average and marginal tax rates will rise with incomes if $b > 0$ in equation (1).

To capture the hypothesis that government spending and/or taxes affect income growth, further assume, in a simplification of the BARRO (1990) model, that changes in income are affected by these government size measures in the previous period. This can arise from the combined (positive and/or negative) effects of changes in

¹² Here we treat as given the institutional structures that determine which goods are delivered via market or non-market mechanisms.

¹³ For income taxes this is clearly built into the tax code where an income definition forms the tax base; however, an indirect relationship with income levels is also often observed for other taxes, such as expenditure taxes, where the tax base (e.g. consumption expenditure) moves in association with income.

¹⁴ For example, the familiar multi-step income tax used in many countries usually involves individual's marginal (and associated average) tax rates rising with income levels.

¹⁵ As noted above a statutory income tax involving a progressive multi-rate structure, generally leads to effective average and marginal tax rates rising as all income levels increase. In most such tax structures the relationship between effective average and marginal rates would not be strictly proportional as both would approach the top statutory marginal rate as incomes rise.

government expenditures and the (presumed negative) effect of changes in effective marginal tax rates. Hence:

$$(3) \quad \Delta Y_t = Y_t - Y_{t-1} = c + d_1 g_{t-1} + d_2 m_{t-1}$$

where c is an autonomous rate of change of income, determined outside this model. Using equation (2), equation (3) can be rewritten as:

$$(3') \quad \Delta Y_t = c + d_1 g_{t-1} + (d_2/\alpha) r_{t-1} = c + d g_{t-1}$$

where $d = \{d_1 + d_2/\alpha\}$ captures the *net* effect of government spending and taxes on the change in income. Though we might expect $d_2 < 0$ if taxes have a negative growth impact (via incentive effects), d_1 is harder to sign but will depend on the positive, zero or negative growth effects associated with various types of public expenditure. Infrastructure expenditure, for example, may boost income levels while others, such as welfare subsidies, may reduce them. Hence, the sign of d in equation (3') is ambiguous.

Using equations (1), (2) and (3') it can be shown that:

$$(4) \quad r_t = g_t = a' + b(1 + bd)Y_{t-1}$$

where $a' = \{a(1 + bd) + bc\}$. With the parameter d ambiguously signed, and $b > 0$, the parameter on Y_{t-1} , $b(1 + bd)$, is therefore also ambiguously signed.

Though simple, equation (4) illustrates a number of insights into the relationship between GDP and government size. Firstly, in the absence of deficit financing (strictly, in the absence of fiscal deficits rising or falling with income levels), any observed association between income levels and government expenditure/income ratios, g_t , can be expected also to be observed for the equivalent relationship with tax revenue/income ratios, r_t .

Secondly, since $b(1 + bd)$ in equation (4) may be positively or negatively signed but $b > 0$ in equation (1) then expected relationships between the government size variables (g_t , or r_t) and income levels depend crucially on the particular income lag structure chosen – Y_t or Y_{t-1} . However, consider the typical case where income is growing over time; that is, $Y_t > Y_{t-1}$. In this case any negative effect on growth from government size (if $d < 0$, in equation (3)) is insufficient to outweigh the positive impact of $c > 0$ in equation (3').

It can be shown that, for this (typical) case, even if $d < 0$, $b(1 + bd) > 0$ and hence we can expect to observe a positive association between g_t and Y_{t-1} in equation (4) as well as between g_t and Y_t shown in equation (1).¹⁶ Different countries may, of course, have different values of the parameter d depending, for example, on the form of public expenditure, the quality of the institutions delivering it. The key point here however is that, even for common negative values of d across countries, the negative association between (lagged) government spending and/or taxes and current income levels, for given lagged values of income (as captured in the structural relationships in equations (3) and (3')), would not be identified by either current or lagged relation-

¹⁶ When d is sufficiently negative to generate $b(1 + bd) < 0$, then Y_t will be lower than Y_{t-1} .

ships between government size and income levels as shown in the simple forms in equations (1) or (4).¹⁷

To illustrate how different possible causal impacts of government size on income changes (captured by d) might show up when cross-country relationships between income levels and government size are examined, we explore three cases: a negative, zero and positive value of d . Figure 6 illustrates these cases (which may be thought of as different countries) based on values of the parameters as follows: $a = 0.1$; $b = 0.001$; $c = 50$, and $d = -100, 0, +100$. Four time periods are considered with initial values of $Y_0 = 100$, $g_0 = 0.2$. For the case of no impact of government size on growth ($d = 0$), income is assumed to rise exogenously by 50 units per period. Both panels in Figure 6 plot resulting income levels against the expenditure share, $g (= G/Y)$ over the four periods. The left-hand panel compares the case of $d = 0$ with $d = 100$, and the right-hand panel compares $d = 0$ with $d = -100$.

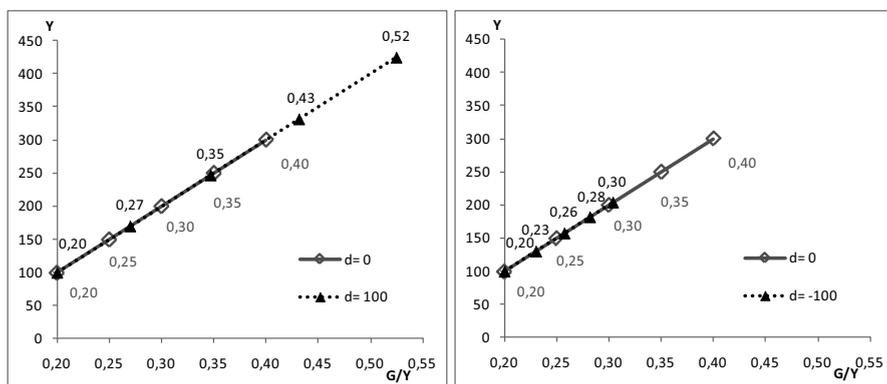


Figure 6: Illustrating government size and output levels

It can be seen that all three countries display identically positively sloping relationships between income levels and government size, as measured by the spending share in income. Simple within-country time-series, or cross-section (for a single period), or combined analysis in a panel would identify a positive association between government size and income levels. This is despite the fact that, in any causal sense, government size is variously harmful, irrelevant or enhancing for income levels over

¹⁷ The mis-specification of typical cross-country OLS regressions (in which income levels or growth rates are regressed on a measure of government size) can be seen by re-arranging (1) to give: $Y_t = -(a/b) + (1/b)g_t$ or, taking first differences, $\Delta Y_t = (1/b)\Delta g_t = (1/b)g_t - (1/b)g_{t-1}$. Hence regression testing for relationships between current government size and current income levels or changes are likely to pick up these endogenous demand-driven relationships between the variables depicted in (1) rather than, or as well as, the hypothesised causal relationships from government size to income levels or growth.

time for the three countries. Rather, what distinguishes the three countries is the *rate of progress* of income levels due to government size effects, d . A key insight from Figure 6 therefore is that a cross-country analysis of income levels is incapable of identifying any causal impact of government size on (subsequent) income levels. This is compounded when it is recognised that autonomous influences on government size, represented by the parameter a in equation (1), and contained within a' in equation (4), are likely to differ across countries such that there could be a series of positively sloping lines for different countries in Figure 6.

Figure 7 shows the implied relationships between *changes* in income levels and government size for the same three countries. As expected there is a difference in slope between the three countries, with negative, zero and positive slopes for the $d = -100, 0, 100$ examples respectively. Notice that the three lines are 'shifted' relative to each other, as well as revealing slope differences, even though all other parameters except d are common across the three countries. It might be thought that these slope differences (between cases of 'no government effect', and 'some government effect', on output) would allow these relationships to be identified in $(\Delta Y, g)$ space. However, once again, with different values of d likely across countries, any set of cross-section, time-series observations, such as in Figure 5, is unlikely to display a simple pattern. As in Figure 6, different autonomous sources of growth in different countries (parameter c values) would produce further diversity in the $(\Delta Y, g)$ space in Figure 7.

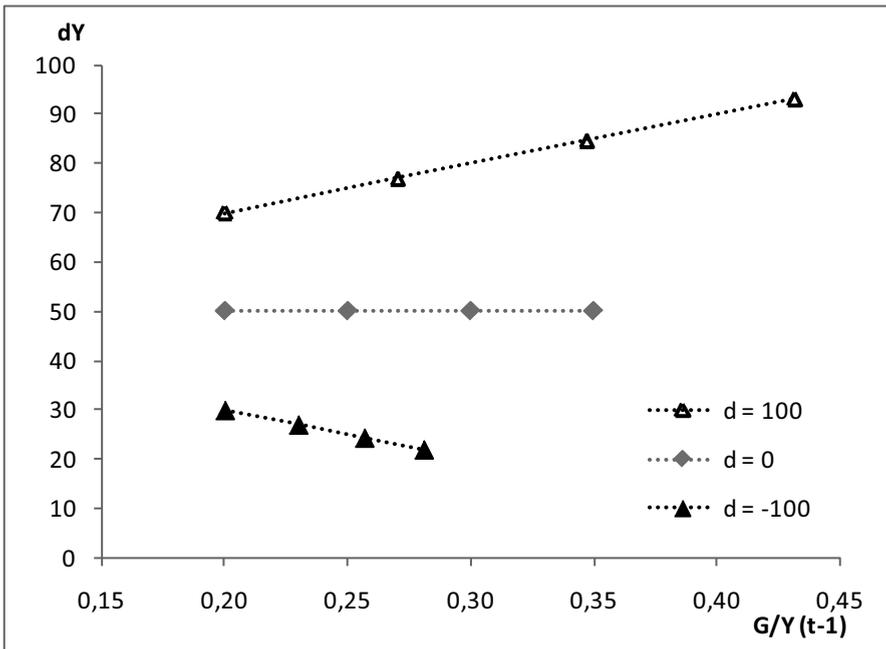


Figure 7: Illustrating government size and output growth

These graphs highlight a further insight from equation (4) – namely countries may differ in the values of any or all of the four critical parameters: a , b , c , and d (as well as further parameters in a more fully specified model).¹⁸ To the extent that they do, further combinations of line shifts and slopes in Figures 6 and 7 will occur, making identification of reliable aggregate cross-country relationships harder. While in principle we might hope to control for other non-government-related factors affecting income levels and growth rates, in practice it is likely to be difficult. This is not simply a matter of poor ‘signal to noise’ ratios due to random unobserved heterogeneity across countries, but also reflects differences in the underlying hypothesised relationships across countries (such as differences in the mix of growth-enhancing and growth-retarding types of public spending or the distortionary impacts of different types of taxes). Though the illustrations above have not included budget deficits, where these are relevant they can be thought of as simply another form of distortionary taxation.

To sum up, this simple illustration emphasises why cross-country, or pooled cross-country/time-series data on government size and income levels or growth rates can readily present confused, apparently inconsistent patterns, with what would appear to be ‘noise’. It also stresses the importance of careful theorising before seeking to identify relationships empirically. For example, reduced form relationships that do not pay careful attention to the lag structure that emerges from the theory are likely to be mis-specified.

While much of the early (approximately, pre-1990) empirical literature based on period averages of cross-country data yielded a confused picture, more recent studies have sought to deal with some of the problems identified above. For example, panel econometric methods are now better able to deal with cross-country heterogeneity, methods of controlling for fiscal and non-fiscal determinants of growth have improved, and improved tests for, and treatment of, endogeneity concerns between income and government size. In general, as Figures 6 and 7 illustrate, it is likely to be easier to identify the associations between government size and income *growth rates* rather than with income *levels*, provided care is taken in allowing for country-specific effects.

4. Evidence from recent studies

GEMMELL (2004) argued that empirical ‘fiscal policy and growth’ evidence could usefully be categorised as ‘first, second and third generation’ studies. First generation studies (approximately pre-1990) were generally motivated, at least formally, by little

¹⁸ The BARRO (1990) model, for example, effectively implies a value of d in (3) that is positive at low income levels, becomes zero then turns negative at high income levels due to the non-linear distortionary effects of rising marginal tax rates. SLEMMOD (1995) and others have argued that this may, in part, explain the failure of empirical studies to identify robust linear cross-country relationships between income levels and government size.

or no theory. To the extent that they were, this often took the form of various, broadly-defined, public choice arguments or hypotheses that predicted a negative association between government size and economic prosperity. Empirical methods were generally cross-section in nature and poorly specified regressions with results, not surprisingly, non-robust. In our view they are worth little as a guide to expected or observed relationships.

Second generation empirical studies (mainly during the 1990s) began to be informed by emerging endogenous growth theory and refinements of the neoclassical model, following BARRO (1990), KING and REBELO (1990), BAXTER and KING (1993) and others. These and other papers provided clearer theoretical foundations both for the basic hypothesis that government fiscal policy could affect long-run, steady-state growth rates, and identified specific channels through which this might occur. Alternative models predicted only transitional effects but these, it was shown, could be expected to last over several decades for plausible values of key parameters (TURNOVSKY 2004).

Empirical studies, however, continued to be based loosely at best on theoretical insights, generally continued to use cross-section econometric methods applied to limited datasets, and tended to use regression specifications with many fiscal variables apparently selected on an *ad hoc* basis rather than derived from specific theoretical frameworks. Again, results appeared non-robust, though KNELLER, BLEANEY and GEMMELL (1999) argued that at least some apparent non-robustness reflected failure to account properly for the GBC which led authors to misinterpret their own estimates.

GEMMELL (2004) defines the ‘third generation’ of long-run cross-country studies (generally post-2000) as those that meet three conditions. They: (i) recognise the implications for testing of the GBC; (ii) allow for potential growth differences from the tax/spending *decompositions* suggested by theory; and (iii) use panel or time-series rather than cross-section methods.¹⁹ GEMMELL, KNELLER and SANZ (2011) argue that this third generation evidence appears more robust than suggested by earlier approaches, but doubts regarding the reliability of even this evidence persist. This stems from a number of features that still afflict aggregate level evidence on fiscal policy and growth including: (a) limited evidence for public expenditures compared to tax (the opposite is generally true for evidence on short-run ‘fiscal shocks’ – see below); (b) conceptual and measurement problems regarding appropriate ‘tax rates’; and (c) difficulties distinguishing supply-side, from demand-side, effects of fiscal policy; see ROMER and ROMER (2010).

Below we summarise some recent evidence from research into the long-run impacts of fiscal policy on growth, not all of which would qualify as ‘third generation’.

¹⁹ Two earlier studies which broadly meet these criteria are MOFIDI and STONE (1990) and MILLER and RUSSEK (1997) – the former applied to cross-State data for the US. Note that almost all third generation evidence appears after the critical reviews of SLEMRD (1995), AGELL et al. (1997) and MYLES (2000).

Firstly, in the next sub-section, we briefly describe recent empirical approaches to measuring short-run impacts of fiscal policy.

4.1 Short-run Evidence from Vector Auto Regression Studies

Due to limited space, this paper focuses on the evidence concerning longer-run impacts of government on growth, but a number of developments in short-run modelling have had influences on the methodologies used to examine these long-run relationships.²⁰

BLANCHARD and PEROTTI (2002) introduced the structural Vector Auto Regression (SVAR) approach to the study of the output effects of fiscal policy and marked a watershed in this empirical literature. In particular, despite long neo-classical and Keynesian traditions supporting possible output effects over the short-run from fiscal policy stimuli or ‘shocks’, reliable evidence measuring these effects was limited prior to around the year 2000. As BLANCHARD and PEROTTI (2002, p. 1329) note: “the evidence from large-scale econometric models has been largely dismissed on the grounds that, because of their Keynesian structure, these models assume rather than document a positive effect of fiscal expansions on output.”²¹

Subsequent developments in SVAR modelling have been considerable, both in terms of methodological improvements and the scope of empirical estimates for individual and groups of countries. In addition, so-called ‘event studies’, such as war episodes, and ‘narrative approaches’ have begun to provide a range of possible approaches to the identification of the output responses to fiscal policy, both short-run and longer-run.²² While a review of the results of the numerous studies using these approaches is beyond the scope of the present paper, a number of features already stand out.

Firstly, because the short-run predictions of neo-classical and New Keynesian models differ in a number of respects, these approaches are, in principle, capable of testing their predictions empirically. Secondly, they provide better methods of testing or controlling for endogeneity (at least in a statistical sense) than has so far been possible in most of the aggregate long-run growth literature. They might therefore hope to provide insights into causal impacts rather than simple ‘associations’.

Thirdly, most of the evidence from the SVAR literature (and to some extent also for the narrative/event approaches) relates to effects from public expenditure shocks, with less evidence for taxation. To the extent that taxation impacts are examined the

²⁰ See DUNGEY and FRY (2009), ILZETZKI et al. (2010), and BEETSMA and GIULIODORI (2011) for more detailed reviews of methods and evidence.

²¹ Other evidence prior to 2000 includes time-series regression estimates of business cycle models for individual countries which estimate or simulate fiscal policy shocks (e.g. MCGRATTAN et al. 1997), and simple Granger-causality based tests that provide some, but limited, insights (e.g. SAUNDERS 1993).

²² See, for example, ROMER and ROMER (2010), RAMEY (2011), and BARRO and REDLICK (2011).

same methodological problems mentioned above regarding endogeneity and measurement of tax rates applies. Indeed for tax shocks, results are more often non-robust and/or counter-intuitive; see, for example, AUERBACH and GORODNICHENKO (2010) and FIELDING, PARKYN and GARDINER (2011) for US and New Zealand evidence respectively. Arguably problems of endogenous tax revenues and tax rate measurement are harder to overcome in this literature since the ability of SVAR approaches to model feedback to tax revenues from other variables in the VAR, requires an endogenous revenue variable, rather than a marginal or average tax rate, measure. However it is the latter that are most likely to capture the direct or exogenous output effects from discretionary *tax policy* changes.

Fourthly, in order to test for short-run fiscal impacts many SVAR models apply methods that impose convergence to a 'no impact' long-run equilibrium. Such models are therefore incapable of identifying any long-run effects of fiscal policy imbedded in the data. Nevertheless, since most models test for the effects of 'one-off' or temporary expenditure or revenue shock (as opposed to permanent changes) such fiscal policy innovations would not be expected to have long-run impacts. They are, however, compatible with the long-run fiscal-growth literature which generally seeks to estimate effects from persistent changes in fiscal policy. Surprisingly, AUERBACH and GORODNICHENKO (2010) appears to be the only empirical study to demonstrate how different phases of the business cycle are associated with quantitatively different fiscal responses (large in recessions, small in expansions). There is, of course, no equivalent in the long-run response literature.

4.2 Long-run evidence from panel studies

As noted above, recent studies into the longer-run impacts of government on growth have increasingly used panel data methods and have generally focussed on the impact of taxes (as opposed to expenditures) on growth. Table 2, adapted from KNELLER and MISCH (2011), lists panel studies which have examined the long-run effects and identifies whether they explicitly consider the GBC and/or allow for dynamics; that is, they allow for short/long-run differences. KNELLER and MISCH's selection criterion required that the parameter estimates obtained by each study were capable of being used to make direct inferences regarding the growth impacts of tax reforms capable of being replicated in practice.²³

The table shows that a large number of studies continue to ignore the GBC in empirical specifications, making interpretation difficult and the actual tax reform harder to identify and replicate. An even larger number of panel studies still do not consider the dynamic effects of tax reforms such that the timeframe over which so-

²³ For example, where studies ignore the GBC and report parameter estimates for, say, an 'income tax' effect on output growth, this could not be used to investigate the growth effects of possible tax reform scenarios without knowing which other element(s) also changed (or were implicitly assumed to change) in the empirical exercise.

Table 2: Overview of recent panel studies

Paper	Data	GBC explicitly considered?	Short-run effects considered?
CROSS-COUNTRY			
Aiginger and Falk (2005)	All OECD, 1970–1999, 5-year averages	no	no
Angelopoulos et al. (2007)	23 OECD, 1970–2000, 5-year averages	no	no
Arin (2004)	G7, 1965–2000, 5-year averages/annual cyclically adj.	no	no
Arin et al. (2011)	UK, US, Scandinavia, 1960–2004, annual/3-year averages	no	no
Arnold (2008)	21 OECD, 1971–2004, annual	yes	yes
Romero-Ávila & Strauch (2007)	EU-15, 1971–2001, annual/cyclically adj.	no	no
Benos (2009)	14 EU, 1990–2006, annual	yes	no
Bleaney et al. (2001)	OECD, 5-year averages and annual data	yes	yes
Castro (2006)	EU-15, 1970–2000, 5-year averages	no	no
Colombier (2009)	21 OECD, 1970–2001	no	no
Doménech and García (2001)	OECD, 1960–1995, 5-year averages	no	no
Gemmell et al. (2011, 2013)	17 OECD, late 1970s–2004, annual	yes	yes
Kneller et al. (1999)	22 OECD, 1970–1995, 5-year averages	yes	no
Muinelo-Gallo and Roca-Sagales (2011)	45 middle/high income countries, 1972–2006, 5-year averages	yes	no
Padovano and Galli (2002)	25 developed countries, 1970–1998, 10-year averages	no	no
Widmalm (2001)	23 OECD, 1965–1990, 5-year averages	no	no
SUB-NATIONAL			
Bania et al. (2007)	49 U.S. states, 1962–1997, 5-year averages	no	no
Dahlby and Ferede (2008)	Canadian provinces, 1977–2006, 5-year averages	no	no
Denaux (2005)	North Carolina Counties, 1980–1995,	no	no
Denaux et al. (2005)	48 U.S. states, 1969–1988	no	no
Miyakoshi et al. (2007)	Japanese prefectures, annual	no	no
Reed, R. (2008)	48 U.S. states, 1970–1999, 5-year averages	no	no

Source: Adapted from KNELLER and MISCH (2011, p. 27).

Table 3: Summary of recent panel estimates of growth effects of taxation

Study	Personal income taxation											
	Tax measure	Angelo-poulos et al. (2007)	Artin (2004)	Dahby and Ferde (2008)	Gemmell et al. (2013)	Gemmell et al. (2013)	Gemmell et al. (2013)	Arnold (2008)	Arnold (2008)	Arnold (2008)	Arnold (2008)	
Log / level	'effective'	levels	IATR	top statutory levels	top statutory levels	top statutory levels	log-levels	log-level	log-level	log-level	% total tax	
	low	-0.174	-0.199	-0.065	-0.08	-0.04	-1.13	-0.2	-1.01	-0.96	% total tax	
Coefficient	high	-0.118		-0.054	-0.05	-0.02					log-level	
											log-level	
Corporate income taxation												
Tax measures	top statutory			top statutory	statutory	effective	% total tax					
	low	levels		levels	levels	levels	log-level	log-level	log-level	log-level	log-level	
Coefficient	high	0.047		-0.158	-0.04	-0.04	-2.01	-1.18	-2.04	-2.4		
				-0.108	-0.03	0.00						
Study	Distortionary taxation											
	Tax measures	Romero-Strach (2008)	Benos (2009)	Gemmell et al. (2011a)	Gemmell et al. (2013)	Bleaney et al. (2001)	Bleaney et al. (2001)	Kneller et al. (1999)	Arnold (2008)			
Log / level	IATR	IATR	IATR	IATR	EATR _L *	IATR	IATR	IATR	IATR	IATR	IATR	% total tax
	logs	levels	levels	levels	levels	log-level	log-level	log-level	log-level	log-level	log-level	log-level
Coefficient	-0.04	-1.516	-0.139	-0.07	-0.854	-0.427	-0.467	-0.463	-0.467	-0.446	-0.98	
	0.05	-0.077	-0.073	-0.03	-0.411	-0.393	-0.410	-0.410	-0.410	-0.427	-0.28	

* EATR_L is the Mendoza-Razin-Tezdar 'effective' average tax rate on labour. Source: Adapted from KNELLER and MISCH (2011, pp. 28–32).

called ‘long run’ effects occur is often unclear. Nevertheless, there are a number of exceptions, most notably ARNOLD (2008) and GEMMELL, KNELLER and SANZ (2011).

Results from some of those panel studies on the long-run impact of personal and corporate income taxes, and the broader category of ‘distortionary taxes’ are summarised in Table 3, as sourced and adapted from KNELLER and MISCH (2011). These consistently estimate the long-run effect of personal taxation on output levels or growth rates to be negative. Several measures of the personal income tax burden were examined including effective rates, implicit average rates, top statutory rates and the share of personal income tax in total tax revenue. For corporate taxation, the estimated effects on output in general are similarly negative and the estimates are often larger than for personal taxes. Nevertheless, some studies such as ANGELOPOULOS, ECONOMIDES and KAMMAS (2007) find somewhat perverse results for corporate taxation (e.g. incorrect signs).

For distortionary taxes overall, the effect of an increase in taxation on growth can appear relatively, perhaps implausibly, large. For example, BLEANEY, GEMMELL and KNELLER (2001) estimate tax-growth parameters at around -0.4 (an increase in the ratio of distortionary tax revenues to GDP of 1% is associated with a fall in long-run growth rates of 0.4 percentage points). However, this effect occurs in conjunction with a similarly sized effect of opposite sign from productive public expenditure increases (estimated distortionary tax-growth effects differ depending on which type of public spending is included in regressions). The combined effect of these two (tax, expenditure) changes appears to approximately cancel out in practice such that observed net effects on long-run growth are small. More recent estimates for similar specifications and samples reveal smaller parameters – around -0.07 to -0.15 (GEMMELL, KNELLER and SANZ, 2011, 2013). Overall, the greater consistency of the results from recent panel studies gives us more confidence that at least the *direction* of the long-run growth effects can be predicted with a reasonable degree of certainty. Furthermore, the size of the effect on output is also now more similar between different panel studies with differences potentially explained by differences in samples, measurement of fiscal variables or the assumed offsetting change.

4.3 Evidence from two recent studies

Of the research mentioned above, two recent studies offer evidence that seeks to shed light simultaneously on the short-run and long-run responses of output to fiscal policy. ROMER and ROMER (2010) use both ordinary least squares (OLS) and fiscal VAR methods applied to quarterly US data (1947–2005) to examine short-run fiscal responses and their persistence over longer periods. GEMMELL, KNELLER and SANZ (2011) use the methodology of long-run growth regressions applied to annual data for a panel of OECD countries (1970–2004), allowing for dynamic responses than can vary across countries in the short-run.

Each study benefits and suffers from the strengths and weaknesses of their respective approaches as described above, but both tell a broadly similar story – of statistically significant initial output responses to tax shocks that appear to persist over a number of years. However, the ‘motivation’ for tax changes and the GBC matter: for

example, whether fiscal changes involve attempts to reduce fiscal deficits, fund additional spending or achieve higher growth. We discuss the results from each of those studies by focusing on the key conclusions that emerge from them.

(1) Tax changes have observable impacts on aggregate output.

For the US, ROMER and ROMER (2010, p. 799) find that:

“results indicate that tax changes have very large effects on output. ... an exogenous tax increase of one percent of GDP lowers real GDP by almost three percent. Our many robustness checks for the most part point to a slightly smaller decline, but one that is still typically over 2.5 percent. In addition, ... investment falls sharply in response to exogenous tax increases.”

Though these effects relate to relatively short periods (ROMER and ROMER’s maximum impacts occur after 12 quarters and simulations are only reported for up to 20 quarters), there is no suggestion that the observed effects decay to zero after the simulated 5 years. They are also broadly consistent with the results for a sample of OECD countries reported by GEMMELL, KNELLER and SANZ (2011), reproduced here as Table 4. The effects shown there are effects on the level of GDP for a 1% of GDP increase in distortionary tax revenues, derived from growth regressions, for up to 20 years after the tax change. While the estimates for the US are somewhat larger than the ROMER and ROMER estimates, for OECD countries on average results are broadly similar. (Note ROMER and ROMER’s tax changes include all Federal tax changes but should be broadly comparable to the GEMMELL et al. ‘distortionary’ tax changes).²⁴

Table 4: GDP level effects of 1% distortionary tax changes (in percent)

Number of years after tax change:	1	5	10	15	20
OECD average	-0.6	-2.0	-3.2	-4.4	-5.5
Canada	-1.8	-2.3	-3.4	-4.6	-5.8
US	-1.5	-4.2	-5.8	-7.3	-8.4
France	-0.7	-0.7	-1.9	-3.0	-4.2
UK	-0.5	-1.6	-2.4	-3.6	-4.7
Australia	-1.0	-1.0	-2.1	-3.3	-4.5
New Zealand	-0.4	-1.6	-2.8	-4.0	-5.1
US:					
Romer & Romer (2010)	-1.0	-2.0*	-	-	-
Turnovsky & Chatterjee (2002)	-0.6	-1.3	-1.8	-2.2	-2.5

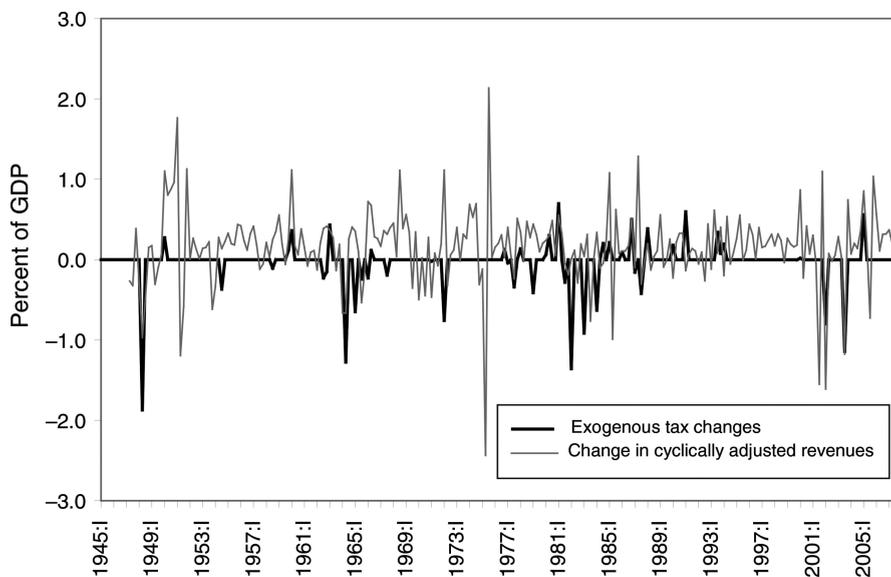
* Maximum estimated impact is 2.9% after 3 years.

Source: Adapted from GEMMELL et al. (2011, p. F51).

²⁴ The table also reports similar orders of magnitude for simulations of the output effects of capital tax rate changes for the US by TURNOVSKY and CHATTERJII (2002).

(2) *Failure to deal with endogeneity will lead to underestimates of tax-growth effects.*

ROMER and ROMER (2010) establish, firstly, that it is important to strip out exogenous changes in tax revenue from all observed changes, where the former are those legislated tax changes that were not motivated by prior changes in output (e.g. due to recessionary factors). Total tax revenues (ROMER and ROMER work with a cyclically adjusted total) may therefore be decomposed into those due to legislated tax changes and the sub-set of ‘exogenous’ legislated tax changes, with the remainder reflecting endogenous revenue responses. These series were created using their ‘narrative approach’ which involved scrutinising Federal Budget documents to identify both the magnitude of, and motivation for, legislated tax changes. Figure 8, reproduced from ROMER and ROMER (2010), shows that the ‘exogenous’ distinction matters a great deal for the resulting measures of tax change. They also show that a comparable measure capturing *all* legislated tax changes (as opposed to only those designated as exogenous) follows a quite different pattern from either of the other two revenue measures.



Source: ROMER and ROMER (2010, p. 779).

Figure 8: ROMER and ROMER’s (2010) tax change measures.

When the output effects associated with the ‘exogenous’ tax measures are compared with those obtained from the total ‘cyclically adjusted revenues’ measure, the estimated output effects are much larger, and identified more precisely, with the former. That is, failure to account for the endogeneity embedded in actual revenue data is more likely to lead to the false conclusion that taxes have only small, or no, effects

on output. For their OECD sample, GEMMELL, KNELLER and SANZ (2011, p. F59) also report some evidence of larger estimated tax effects on output growth when using methods that deal better with endogeneity aspects.

(3) Allowing for each element of the government budget matters.

As we noted earlier, the fact that the government budget is an accounting identity means that changes in one element must be matched by an equal and opposite net change across all other elements. Despite this, many authors of fiscal-growth studies interpret their results in terms of, for example, 'the effect of a tax change'. This only makes sense if the specific tax change, and the opposing changes are specified (except in the unlikely case where all possible opposing changes have identical effects). Thus, a reduction in tax rates which lowers tax revenue must be associated with a larger fiscal deficit or lower public spending, or both. The results of both ROMER and ROMER and GEMMELL et al. confirm that this particular distinction is especially important for the estimated growth effects of fiscal policy changes.

ROMER and ROMER (2010, p. 786 et seq.) report that when exogenous tax increases are 'motivated' by a need/desire to reduce fiscal deficits, as distinct from a motivation to raise long-run growth rates, the output effects are quite different. For the latter category, output effects are almost identical to those for exogenous tax changes overall; i. e. around a 3% reduction in output in response to a 1% increase in tax revenue. However, where deficit reduction is the motivation, around a 1–2% *increase* in output is observed when taxes rise by 1% with the impact tending, if anything, to increase, rather than decline, over time. This evidence therefore suggests strongly that reducing US fiscal deficits is growth-enhancing even though this involves raising taxes: deficits (together with pre-reformed taxes) are more growth-retarding than the combination of higher post-reform taxes and lower deficits.

This result is mirrored in GEMMELL, KNELLER and SANZ (2011, p. F41) for the OECD. Their results demonstrate that, when splitting the government budget into five elements (more/less distortionary taxes, productive/unproductive spending, and deficits), the estimated growth effects observed depend on the various *combinations* of budget change, as indicated in Table 4. For example, on average across the OECD sample, raising more distortionary (mainly income) taxes to fund cuts in less distortionary (mainly consumption) taxes or to increase unproductive government expenditure appears to be around twice as harmful for long-run growth rates compared to when the tax increase funds a reduction in fiscal deficits. The authors argue that a rough ranking of taxes and deficits (from largest to smallest) in terms of their growth effects is: more distortionary taxes → deficits → less distortionary taxes, where each option finances the same total and mix of public spending.

(4) Observed fiscal policy (tax/expenditure/deficit) changes overall do not have large effects on observed long-run GDP growth rates.

JONES (1995) and KARRAS (1999) argued that evidence of non-stationarity in investment/GDP ratios (JONES) and total tax/GDP ratios (KARRAS) in many countries is

at odds with evidence of stationarity in GDP growth rates, if the former were thought to cause the latter. Indeed, JONES argues that the former cannot plausibly explain the latter, unless ‘*by some astonishing coincidence all of the movements in variables that can have permanent effects on growth rates have been offsetting*’ (JONES 1995, p. 496). This might appear to be a damning critique of the argument that fiscal policy has persistent or ‘permanent’ impacts on GDP growth rates.²⁵ However there are at least two reasons why stationary long-run growth rates may be consistent with the evidence of persistent effects from fiscal policy from ROMER and ROMER (2010), GEMMELL, KNELLER and SANZ (2011), and others.

Firstly, OECD evidence suggests a clear tendency for so-called ‘productive’ public spending to be positively associated with long-run GDP growth rates, while distortionary taxes are negatively associated, such that in combination there are negligible long-run growth effects (GEMMELL, KNELLER and SANZ 2011, p. F50). With all changes in net tax revenues being offset by spending and/or deficit changes it is not surprising that the net growth effects largely ‘wash out’. This is quite consistent with a finding that reforms involving a particular type of tax funding of a particular composition of public expenditure can have a non-zero impact on GDP over the short- or long-run.

Secondly, the evidence suggests that many changes in fiscal variables, and especially exogenous changes, have often not persisted. Hence, even if in principle such fiscal changes would have persistent effects on GDP levels or growth rates, short-run reversals of these fiscal changes will similarly reverse the initial impact on GDP. ROMER and ROMER’S (2010) US data in particular show that substantial quarterly increases in tax revenues (as a percentage of GDP) are frequently followed by reverse movements soon thereafter, while their measure of exogenous tax changes display relatively few, intermittent and temporary changes in either direction (see Figure 8).

5. Limitations

The evidence reviewed in Section 4 would appear to suggest stronger support than previously acknowledged for at least some of the alleged relationships between government size or other aspects of fiscal policy and output at the macro-level. Nevertheless there are a number of reasons to exercise some caution in interpreting the results of even the more robust recent evidence as confirming non-zero output effects of fiscal policy, especially over the longer-run.

Issues of interpretation continue to pose difficulties with the current evidence principally for two reasons. Firstly, specifying the government budget constraint suitably in empirical studies can be problematic yet interpretation of parameters depends crucially on knowing how the GBC is affected by different specific fiscal policy

²⁵ We prefer the term ‘persistent’ to ‘permanent’ here since the generally limited periods of analysis used in most studies does not allow ‘persistent’ but nevertheless transitional effects (in the sense used in theoretical models), to be readily distinguished from permanent effects.

changes, Secondly, interpretation of results that may be subject to endogeneity problems continues to pose some difficulties for authors of aggregate level studies and allows some scepticism of their results to persist.

Sceptics, however, typically query whether estimated parameters on fiscal variables in output or growth regressions that demonstrate statistically significant non-zero estimates, genuinely capture non-zero effects. This fails to recognise that, to the extent that *tax* measures (often revenue-based) are subject to endogeneity, this is expected to bias the estimated parameter towards zero. ROMER and ROMER's (2010) evidence, for example, supports the view that endogeneity leads to *under*-estimates of the absolute value of tax effects on output and other macro-aggregates. For government expenditures the issue is less straightforward since the sign of any bias depends on the how particular expenditures are expected to change in response to rising income levels.

Identifying and using the 'right' tax rates remains problematic. In the case of fiscal VARs, modelling of *tax* shocks is especially unsatisfactory because of their reliance on revenue-based tax measures. While modelling the endogenous response of tax *revenue* to an exogenous expenditure change may be the appropriate procedure to capture the feedbacks on fiscal budgets and output via demand pressures, it is not the appropriate way to capture responses to exogenous tax shocks. Perhaps it is not surprising then that several VAR studies have found perverse (i.e. positively signed) output effects from tax increases. Tellingly, ROMER and ROMER's (2010) fiscal VAR for the US obtains the expected negative sign using their measure of 'exogenous' tax revenue changes but not for their measure of total revenue changes.

In the case of long-run growth regressions, tax rates are also often measured from a revenue-based variable so hampering interpretation of results. However, as GEMMELL, KNELLER and SANZ (2011, 2013) demonstrate, results using better measures of effective marginal and average tax rates – for personal and corporate income taxes – continue to support previous evidence of long-run negative, if quantitatively small, tax-growth effects. Nevertheless, evidence on the output or growth effects of reliably exogenous effective marginal (or average) tax rates remains limited.

A surprising aspect of much of the VAR and growth regression evidence is the lack of an *international dimension* with many models either implicitly or explicitly 'closed economy'. This is despite the increasing evidence over the last several decades of global flows that can be expected to give rise to 'fiscal leakages' and increased co-determination of fiscal policy settings such as corporate tax rates across countries (see, for example, DEVEREUX, LOCKWOOD and REDOANO 2008). BEETSMA and GUILIODORI (2011) and GEMMELL, KNELLER and SANZ (2013) are two of the small number of recent studies that explicitly seek to incorporate these international aspects.

Another limitation arises as a consequence of the widespread use of 'tax revenue' or tax revenue/GDP measures. These measures are not direct discretionary policy instruments and hence provide little guidance regarding how *actual* fiscal instruments can be suitably reformed. Finance Ministers, if they are to enact tax policy reform typically need to know: '*if a single tax rate is cut by y%, or several tax rates are*

simultaneously cut by x , y , $z\%$, what will be the effects on output growth over the next 5 (or more/fewer) years?’ and ‘what are the impacts of alternative revenue-neutral tax rate and/or other tax parameter reforms?’.

Currently available parameter point estimates that could be applied to address such questions are not generally available or are not precise enough to provide narrow prediction ‘cones’ for effects on GDP. Moreover, most available parameter estimates are averages across episodes, countries or states, and their reliability in a given condition in a specific country, time or region can be unclear. The challenge for evidence-based fiscal policy reform is to quantify tax/spending reform impacts on output in ways that will allow policy-makers to compare alternative fiscal and other policy reforms in a quantitatively robust manner.

Finally, single country SVARs and ‘narrative’ studies such as ROMER and ROMER (2010) who combine both methods, are probably the most reliable methodologies where it is desired to identify fiscal-growth effects up to a horizon of five years or so. However, even here, most studies have largely ignored the possibility that these effects differ depending on the *degree* of disequilibrium in the economy. The recent paper by AUERBACH and GORODNICHENKO (2010) appears to be the first formally to try to quantify such differences (for the US). They argue, with empirical support, that this distinction may matter a great deal.

6. Conclusions

We began this paper with the objective of answering three questions relating to the relationships between government size, fiscal policy and output levels or growth rates:

- (1) how robust or consistent are the results from recent macro-level studies?;
- (2) do these results justify a less sceptical verdict than emerged from earlier reviews such as SLEMROD (1995) and MYLES (2000)?; and
- (3) are these results sufficient to be useful as a guide to fiscal policy reform in practice?

In summary, we have argued that the theory on the output effects of fiscal policy has developed considerably since the time of SLEMROD’s writing and this theory has increasingly been taken seriously in specifying empirical tests. Undoubtedly improvements in panel datasets (both across countries and within them), and methods to analyse them, have also helped to address some of the criticisms levelled by SLEMROD (1995) at the mainly cross-section empirical evidence available at that time. Partly as a result, the last decade or so appears to have produced more robust evidence on the impacts of taxes and public expenditures on GDP at the macro-level.

The last decade has also seen much greater application of time-series, mainly VAR, methods to examine shorter-run impacts of fiscal policy. Together with greater use of dynamic specifications in long-run panel regression studies, this research has now established clearer evidence of short-run output effects from fiscal policy that demonstrate some persistence – at least up to five years and probably longer.

Our overall conclusion is that, despite some remaining ‘big issues’, results from aggregate-level studies are now much more robustly in favour of identifiable effects of government size on income growth rates, provided these conclusions are carefully circumscribed in ways that we have discussed. In general, positive output effects from increases in at least some public expenditures and negative output effects from higher tax rates are supported. Plausible orders of magnitude are also beginning to emerge suggesting that the impacts of fiscal policy may be sizeable in the short-run (but depend, among other things, on the degree of excess capacity in the economy), and are probably moderate or small in the longer-run. Importantly, long-run *net* fiscal effects on output are often small or negligible when the more sizeable effects of taxes, expenditures or deficits measured *in isolation*, are instead considered in *combination*, as they should be.

Question marks remain over the reliability of estimates of tax effects on output versus public spending effects on output. Within the VAR literature on short-run output responses, evidence on expenditure shocks is both more extensive and seems to be more robust than equivalent evidence for tax shocks. In the long-run growth regression literature the reverse is probably true – it has generally proved easier to identify negative tax-growth impacts than to identify growth impacts from public expenditures. This may partly reflect the fact that the expected sign of the former is unambiguously negative, whereas expenditure effects on growth are ambiguously signed. A major weakness of almost all recent research is that its value remains limited as an evidence-base for robust quantitative predictions of the output responses to changes in the specific tax policy instruments that are typically available to governments undertaking fiscal reform in practice.

Much debate has surrounded how far existing evidence supports neo-Keynesian, neo-classical or endogenous growth model specifications. To some extent the answer to this question depends on the in/out-of steady-state issue mentioned above. However, much of the debate has also sought to shed light on whether, and how far, observed fiscal-growth effects represent demand-side or supply-side responses in the economy. The difficulties of this exercise are summed up by ROMER and ROMER (2010, p. 799) who acknowledge:

“our results are largely silent concerning whether the output effects operate through incentives and supply behavior or through disposable income and demand stimulus. The persistence of the effects is suggestive of supply effects. But other studies have found that monetary policy, which necessarily works through demand, also has highly persistent output effects. The speed of the effects is suggestive of demand effects. But rapid supply responses are not out of the question.”

We began by suggesting that a knowledge of the OECD evidence on the ‘government and growth’ debate would be beneficial for policy makers and advisers. Clearly the relevance of the evidence from individual OECD countries will be quite different and depend on a variety of institutional differences. However, some lessons would seem to have general applicability.

Firstly, cross-country correlations are likely to provide little assistance and can be positively misleading, especially where interest focuses on understanding the impact of changes in government size *within* a country.

Secondly, when looking for relationships between government size and income variables, the tendency for these two variables to be mutually reinforcing ('endogeneity') often appears to be important, and ignoring this can readily lead to false conclusions from the data. If newer OECD members with lower income levels experience the increased demands for some publicly supplied services in association with income growth that has been observed in some 'older' OECD countries, then this will similarly complicate identifying any causal relationships from government size (or growth) to GDP. A balance may need to be struck between a desire for more and better public services and the subsequent growth consequences of the resulting public spending and tax choices.

Thirdly, the fairly wide-ranging support for adverse effects of taxes on aggregate output (in addition to more extensive micro-level evidence on behavioural responses to taxes) suggest that policy makers in OECD countries should be cautious about raising potentially growth-retarding distortionary taxes. However, the small *net* growth impacts of larger public expenditures, deficits and tax revenues that have been estimated cautions against expecting a sustainable increase in government size (i. e. one that is not financed longer-term by increased deficits) to have substantial long-run effects, either growth-enhancing or retarding.

Finally, evidence is still scarce on how far the *institutions* of government, the *quality* of those institutions (including such things as corruption among, or performance of, government officials) and the degree of non-fiscal government interventions such as business regulation affect aggregate income levels or growth rates. Most individual OECD countries have developed their own specific institutions, as well as often being part of broader cross-country institutions such as the European Union, over a long period of time. At least some evidence suggests that institutions such as the 'rule of law' and government corruption are among the 'necessary but not sufficient' conditions required for public spending to affect private sector investment or productivity positively.²⁶ Some countries may therefore do more to achieve or maintain high economic growth rates by ensuring that key governance-related institutions are maintained or developed than by modest changes to the level or structure of public expenditures and taxes.

References

- AGELL, J., T. LINDH and H. OHLSSON (1997): Growth and the Public Sector: A Critical Review Essay, *European Journal of Political Economy* 13 (1), 33–52.
AIGINGER, K. and M. FALK (2005): Explaining Differences in Growth Among OECD Countries, *Empirica* 32 (1), 19–43.

²⁶ See, for example, GLAESER et al. (2004), MAURO (1995).

- ANGELOPOULOS, K., G. ECONOMIDES and P. KAMMAS (2007): Tax-spending Policies and Economic Growth: Theoretical Predictions and Evidence from the OECD, *European Journal of Political Economy* 23 (4), 885–902.
- ARIN, P. (2004): Fiscal Policy, Private Investment and Economic Growth: Evidence from G-7 Countries, unpublished manuscript, Massey University, Auckland.
- ARIN, K. P., M. BERLEMANN, F. KORAY and T. KUHLENKASPER (2011): Nonlinear Growth Effects of Taxation: A Semi-parametric Approach Using Average Marginal Tax Rates, forthcoming in: *Journal of Applied Econometrics*, preprint available under DOI:10.1002/jae.2311.
- ARNOLD, J. (2008): Do Tax Structures Affect Aggregate Economic Growth: Empirical Evidence from a Panel of OECD Countries. OECD Economics Department Working Paper 643, Paris.
- AUERBACH, A. J. and Y. GORODNICHENKO (2010): Measuring the Output Responses to Fiscal Policy. National Bureau of Economic Research (NBER) Working Paper 16311, Cambridge (MA).
- BANIA, N., J. A. GRAY and J. A. STONE (2007): Growth, Taxes, and Government Expenditures: Growth Hills for U.S. States, *National Tax Journal* 60 (2), 193–204.
- BARRO, R. (1990): Government Spending in a Simple Model of Endogenous Growth, *Journal of Political Economy* 98 (5), 103–17.
- BARRO, R. and C. J. REDLICK (2011): Macroeconomic Effects of Government Purchases and Taxes, *Quarterly Journal of Economics* 126 (1), 51–102.
- BAXTER, M. and R. G. KING (1993): Fiscal Policy in General Equilibrium, *American Economic Review* 83 (3), 315–34.
- BEETSMA, R. and M. GIULIODORI (2011): The Effects of Government Purchases Shocks: Review and Estimates for the EU, *Economic Journal* 121 (550), F4–32.
- BENOS, N. (2009): Fiscal Policy and Economic Growth: Empirical Evidence from EU Countries. MPRA Paper 19174, University Library of Munich, Germany.
- BLANCHARD, O. and R. PEROTTI (2002): An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output, *Quarterly Journal of Economics* 117 (4), 1329–68.
- BLEANEY, M., N. GEMMELL and R. KNELLER (2001): Testing the Endogenous Growth Model: Public Expenditure, Taxation and Growth Over the Long Run, *Canadian Journal of Economics* 34 (1), 36–57.
- CASTRO, C. (2006): Política fiscal e crescimento econômico, *Polytechnical Studies Review* 3 (5/6), 87–118.
- COLOMBIER, C. (2009): Growth Effects of Fiscal Policies: An Application of Robust Modified M-Estimator, *Applied Economics* 41 (7), 899–912.
- DAHLBY, B. and E. FEREDÉ (2008): Tax Cuts, Economic Growth, and the Marginal Cost of Public Funds for Canadian Provincial Governments, unpublished manuscript.
- DENAUX, Z. S. (2005): A Cross-County Evaluation of Fiscal Policy, *Journal of Business and Public Affairs* 32 (1), 14–20.
- DENAUX, Z. S., D. J. CUNNINGHAM and R. C. ALLEN (2005): Spend More of Tax Less? Which Way to State Economic Growth?, *Journal of Business and Public Affairs* 32 (1), 21–5.
- DEVEREUX, M. P., B. LOCKWOOD and M. REDOANO (2008): Do Countries Compete Over Tax Rates?, *Journal of Public Economics* 92 (5–6), 1210–35.
- DOMÉNECH, R. and J. R. GARCÍA (2001): Estructura Fiscal y Crecimiento Económico en la OECD, *Investigaciones Económicas* 15 (3), 441–472.

- DUNGEY, M. and R. FRY (2009): Identifying Fiscal and Monetary Policy in a Small Open Economy VAR, *Economic Modelling* 26, 1147–60.
- FIELDING, D., O. PARKYN and P. GARDINER (2011): Explaining Some Puzzles in the Estimated Response of New Zealand GDP to Fiscal Shocks, paper presented to the New Zealand Association of Economists, Annual Conference, Wellington, New Zealand, July.
- GEMMELL, N. (2004): Fiscal Policy in a Growth Framework. In: T. Addison and A. Roe (Eds), *Fiscal Policy for Development: Poverty, Reconstruction and Growth*, Palgrave Macmillan, London, 149–76.
- GEMMELL, N., R. KNELLER and I. SANZ (2011): The Timing and Persistence of Fiscal Policy Impacts on Growth: Evidence from OECD Countries, *Economic Journal* 121 (550), F33–58.
- GEMMELL, N., R. KNELLER and I. SANZ (2013): The Growth Effects of Tax Rates in the OECD, Victoria University of Wellington Working Papers in Public Finance 02/2013, forthcoming in *Canadian Journal of Economics*.
- GLAESER, E. L., F. LOPEZ-DE-SILANES and A. SCHLEIFER (1994): Do Institutions Cause Growth?, *Journal of Economic Growth* 9 (3), 271–303.
- GWARTNEY, J. D., R. A. LAWSON and R. G. HOLCOMBE (1998): The Scope of Government and the Wealth of Nations, *Cato Journal* 18 (2), 163–90.
- ILZETZKI, E., E. G. MENDOZA and C. A. VEGH (2010): How Big (Small?) are Fiscal Multipliers?. National Bureau of Economic Research (NBER) Working Paper 16470, Cambridge (MA).
- JONES, C. I. (1995): Time Series Tests of Endogenous Growth Models, *Quarterly Journal of Economics* 110 (2), 495–525.
- KARRAS, G. (1999): Taxes and Growth: Testing the Neoclassical and Endogenous Growth Models, *Contemporary Economic Policy* 17 (2), 177–88.
- KING, R. G. and S. REBELO (1990): Fiscal Policy and Economic Growth. Developing Neoclassical Implications. National Bureau of Economic Research (NBER) Working Paper 3338, Cambridge (MA).
- KNELLER, R., M. BLEANEY and N. GEMMELL (1999): Fiscal Policy and Growth: Evidence from OECD Countries, *Journal of Public Economics* 74 (2), 171–90.
- KNELLER, R. and F. MISCH (2011): What Does Ex-post Evidence Tell us About the Output Effects of Future Tax Reforms?. Centre for European Economic Research (ZEW) Discussion Paper 11-029, Mannheim, Germany.
- KORPI, W. (1996): Eurosclerosis and the Sclerosis of Objectivity: On the Role of Values Among Economic Policy Experts, *Economic Journal* 106 (439), 1727–46.
- MAURO, P. (1995): Corruption and Growth, *Quarterly Journal of Economics* 110 (3), 681–712.
- MCGRATTAN E. G., R. ROGERSON and R. WRIGHT (1997): An Equilibrium Model of the Business Cycle with Household Production and Fiscal Policy, *International Economic Review* 38 (2), 267–90.
- MILLER S. and F. RUSSEK (1997): Fiscal Structures and Economic Growth: International Evidence, *Economic Inquiry* 35 (3), 603–13.
- MİYAKOSHI, T., Y. TSUKUDA, T. KONO and M. KOYANAGI (2007): Public Expenditure Composition and Economic Growth: Optimal Adjustment by Using Gradient Method. Osaka University Discussion Paper 07–17, Japan.
- MOFIDI, A. and J. STONE (1990): Do State and Local Taxes Affect Economic Growth?, *Review of Economics and Statistics* 72 (4), 686–91.

- MUINELO-GALLO, L. and O. ROCA-SAGALES (2011): Economic Growth and Inequality: The Role of Fiscal Policies, *Australian Economic Papers* 50 (2–3), 74–97.
- MYLES, D. (2000): Taxation and Economic Growth, *Fiscal Studies* 21 (1), 141–68.
- MYLES, D. (2009a): Economic Growth and the Role of Taxation-Theory. OECD Economics Department Working Papers 713, OECD Publishing.
- MYLES, D. (2009b): Economic Growth and the Role of Taxation-Aggregate Data. OECD Economics Department Working Papers 714, OECD Publishing.
- MYLES, D. (2009c): Economic Growth and the Role of Taxation-Disaggregate Data. OECD Economics Department Working Papers 715, OECD Publishing.
- NEW ZEALAND TREASURY (2011): Government and Economic Growth. Does Size Matter?, New Zealand Treasury Paper 11/01, Wellington.
- PADOVANO, F. and E. GALLI (2002): Comparing the Growth Effects of Marginal vs Average Tax Rates and Progressivity, *European Journal of Political Economy* 18 (3), 529–44.
- PERETTO, P. F. (2003): Fiscal Policy and Long-run Growth in R&D-based Models with Endogenous Market Structure, *Journal of Economic Growth* 8 (3), 325–47.
- PERETTO, P. F. (2007): Corporate Taxes, Growth and Welfare in a Schumpeterian Economy, *Journal of Economic Theory* 137 (1), 353–82.
- RAMEY, V. A. (2011): Identifying Government Spending Shocks: It's All in the Timing, *Quarterly Journal of Economics* 126 (1), 1–50.
- REED, R. (2008): The Relationship Between Taxes and US State Income Growth, *National Tax Journal* 61 (1), 57–80.
- RODRIG, D. (1994): King Kong Meets Godzilla: The World Bank and the East Asian Miracle. Centre for Economic Policy Working Paper 944, London.
- ROMER, C. D. and D. H. ROMER (2010): The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks, *American Economic Review* 100 (3), 763–801.
- ROMERO-ÁVILA, D. and R. STRAUCH (2008): Public finances and long-term growth in Europe: evidence from a panel data analysis, *European Journal of Political Economy* 24 (1), 172–191.
- SAUNDERS, P. J. (1993): A Granger Causality Approach to Investigating the Impact of Fiscal Policy on the US Economy: 1970.I – 1990.III, *Studies in Economics and Finance* 16 (1), 3–22.
- SLEMROD, J. (1995): What Do Cross-Country Studies Teach about Government Involvement, Prosperity, and Economic Growth?, *Brookings Papers on Economic Activity* 2, 373–431.
- TURNOVSKY, S. J. (2004): The Transitional Dynamics of Fiscal Policy: Long-run Capital Accumulation and Growth, *Journal of Money, Credit and Banking* 36, 883–910.
- TURNOVSKY, S. J. and S. CHATTERJII (2002): To Spend the US Government Surplus or to Increase the Deficit? A Numerical Analysis of the Policy Options, *Journal of Japanese and International Economics* 16 (4), 405–35.
- YOUNG, A. (1992): A Tale of Two Cities: Factor Accumulation and Technical Change in Hong Kong and Singapore, *NBER Macroeconomics Annual* 7, 13–64.

