November 6, 2003

Growth effects of government expenditure and taxation in rich countries: A comment

By Jonas Agell, Henry Ohlsson and Peter Skogman Thoursie*

Abstract
In a recent article Stefan Fölster and Magnus Henrekson [2001] argue that "…the more the econometric problems that are addressed, the more robust the relationship between government size and economic growth appears". But in failing to control for simultaneity in a valid manner the regressions reported by Fölster/Henrekson are flawed. Moreover, using theoretically valid instruments we find that the estimated partial correlation between size of the public sector and economic growth is statistically insignificant and highly unstable across specifications. A policy-maker who wants to promote growth is well-advised to look for other evidence than cross-country growth regressions.

JEL classification: E62; H20; H50; 057
Keywords: Economic growth; public sector; cross-country regressions; panel regressions

* Agell: Department of Economics, Stockholm University, SE-10691 Stockholm and CESifo (email: JA@ne.su.se); Ohlsson: Department of Economics, Uppsala University, SE-75120 Uppsala (email: Henry.Ohlsson@nek.uu.se); Skogman Thoursie: Department of Economics, Stockholm University, SE-10691 Stockholm and The Swedish National Social Insurance Board (email: PT@ne.su.se). We are grateful for helpful comments from Per Pettersson Lidbom.

1. Introduction

During the 1980s and 1990s the literature on cross-country growth regressions literally exploded. One of the most controversial issues studied in this literature is whether a large public sector is growth promoting or growth retarding. In spite of intense research efforts, a number of recent overviews conclude that this line of research is unlikely to come up with very definite and reliable answers. Temple (1999, p. 145) concludes that the question of government size and growth is likely to remain largely unsettled, and that "...microeconomic evidence on labor supply and investment responses to changes in tax rates is likely to be more informative." Atkinson (1995) observes that the inherent neglect of microeconomic detail makes it difficult to draw any useful lessons from cross-country growth regressions. Agell, Lindh and Ohlsson (1997, 1999) argue that because of severe problems of econometric methodology and data quality, cross-country growth regressions will not shed light on the substantive policy issues. Slemrod (1995, 1998) points to the econometric problems of endogeneity and simultaneity, which may occur along several dimensions. In discussing the lessons for policy from cross-country growth regressions, Mankiw (1995, p. 307-308) writes: "Using these regressions to decide how to foster growth is also most likely a hopeless task. Simultaneity, multicollinearity, and limited degrees of freedom are important practical problems for anyone trying to draw inference from international data. Policymakers who would like to promote growth would not go far wrong ignoring most of the vast literature reporting growth regressions."

A recent paper in this journal by Henrekson and Fölster (2001), henceforth FH, argue otherwise. FH report many regressions, which according to them show that there is a robust negative relationship between economic growth and government size in a panel of rich countries. Though FH acknowledge that the previous literature has identified a number of econometric pitfalls, they seem to think that their own approach addresses most of the issues.

They summarize "Our general finding is that the more the econometric problems are addressed, the more robust the relationship between government size and economic growth appears. Our most complete specifications are robust even according to the stringent extreme bounds criterion" (FH, p. 1501). They also offer their econometric analysis as a guide for policy. In their conclusion, FH indicate that their results imply that an increase in the size of the public sector by ten percentage points can be expected to lower the growth rate with 0.7-0.8 percentage points.

This paper evaluates these claims.[1] In our view the new results reported by FH give no guidance on the magnitude of the growth-retarding effects of a large public sector. Most of their regressions do not address the difficult issues of identification and simultaneity at all, and the specifications that try to do something about it are flawed. The absence of clear negative relations in the aggregate data is of course no proof that a large public sector is of no consequence for economic growth. In effect, it merely suggests that cross-country panel regressions is too imprecise a business to shed light on the important policy issues.

2. The problem of simultaneity

As noted by e.g. Mankiw (1995) and Slemrod (1995, 1998) students of cross-country growth regressions have to address a fundamental simultaneity problem. Most of the right-hand side variables in cross country growth regressions are best thought of as being determined jointly with the economic growth rate. The size of the public sector affects growth via a standard supply side relation, while growth affects the size of the public sector via the income elasticity of the demand for public sector activities. As the observed data points can be interpreted as reflecting the intersection of the supply and demand relations, a given partial correlation is not

[1] Fölster and Henrekson (1999) is an earlier publication in the *European Journal of Political Economy* that reports very similar regressions to those reported by FH in the *European Economic Review*, and Agell, Lindh and Ohlsson (1999) is a response that partly overlaps with what we say here. We were surprised to see that this previous exchange is not acknowledged.

informative about the relation that is of our primary interest, the supply relation. For example, a negative correlation between government spending economic growth may reflect nothing more than the fact that the demand for unemployment benefits and social assistance increase in recessions. As discussed in great detail by Slemrod (1995), identification of the supply relation requires that we find an instrument set consisting of exogenous variables that shift the demand relation, but not the supply relation. In practice, this is bound to be difficult, and perhaps even impossible.

The punchline is that we should not expect cross-country growth regressions to be informative about causality. What is reported is a set of correlations between endogenous variables, and not something indicating how a change in a variable *X* impacts on variable *Y*. Such correlations void of causal interpretation can still be quite useful. As noted by Mankiw (1995) the fact that numerous studies have identified a positive correlation between growth and investment can be used to rule out theories that fail to generate this correlation.

We find it noteworthy that the vast majority of regressions reported by FH do not even try to address the simultaneity problem. Though they make reference to a number of difficult econometric issues, including simultaneity, this is not reflected in their empirical work, which is centred on ordinary fixed effect and weighted fixed effects OLS regressions. It is important to note that *none* of these regressions addresses the problems of simultaneity and identification.[2] Simply put, these regressions are potential mis-specifications, not up to the standards of the current research literature. This includes nearly all the regressions reported in Tables 2, 3, and 5, as well the robustness results reported in Tables 6 and 7.

An indication of a potentially severe mis-specification problem can be seen in Table 2 in FH (FH, p. 1508). In these specifications public spending and public consumption show up with highly significant negative coefficients, results very much stressed by FH. But the same

[2] The fixed effect specification is of course a traditional way of dealing with endogeneity problems that can be traced to unobservable time-invariant country fixed effects. However, it does not deal with reverse causation.

table also shows that many of the variables that have been shown to affect economic growth in numerous studies show up with the wrong sign, or get estimated with very low precision. Several studies have reported that a country’s physical investment share and indicators of a country’s human capital stock are positively associated with economic growth. Yet, of the twelve estimated coefficients on measures of physical investment and human capital shown in the basic fixed effects regressions reported in Table 2 (FH p. 1510), no less than seven have a negative sign, and none is estimated with a significant positive coefficient.

The specification problem that is paid the greatest attention by FH is heteroscedasticity, and they report scores of weighted least squares regressions, intended to address this issue. Today, most researchers would deal routinely with between-country heteroscedasticity by invoking a robust cluster estimator. An advantage of this class of estimators is that they do not require a precise modelling of the source of heteroscedasticity; they are robust to heteroscedasticity of arbitrary form. Cluster estimators tend to increase the reported standard errors by a fairly large amount, which reduces the statistical significance levels of the estimated coefficients. However, the non-robust weighing estimator invoked by FH – an estimator that *is* quite sensitive to the exact form of heteroscedasticity – shrinks all standard errors. As a result, the coefficients in front of their public sector variables tend to become more significant as FH move from ordinary least squares to weighted least squares. More generally, we found it hard to understand FH’s motivation for introducing their weighted least-squares procedure. FH, p. 1507, write: “Heteroscedasticity most often appears in a form where the error term is correlated with one of the independent variables or with the dependent variables.” Obviously, a correlation between the error term and one of the independent variables is the essence of an endogeneity problem, not heteroscedasticity.

3. New results

It is only in Table 3, columns 5 and 6, that FH report regressions that address the simultaneity problem. We now turn to our attempts of replicating these results. Before discussing this exercise in greater detail it should be noted that FH change the dependent variable in their TSLS-regressions. In all specifications where FH do not address the issue of simultaneity they use the growth rate of actual GDP, averaged over five-year periods, as the dependent variable. In their TSLS-regressions, however, FH invoke a measure of the growth rate of potential GDP. There is no motivation for this change of dependent variable. The results reported below, which differ significantly from those of FH, are based on data on potential GDP that FH sent us at an early stage of our replication project.[3]

Table 1, rows 1 and 7 show the TSLS-regressions as reported by FH. In addition to having the growth rate of potential GDP as the dependent variable, these first-difference regressions use an assorted set of instruments for the first differences of the tax and public spending variables. Both the instrumented tax and expenditure shares have large negative coefficients, with *t*-ratios of 3.4 and 5.6. Economically speaking the coefficient on the tax share suggests that a ten-percentage point increase in the tax-to-GDP ratio reduces the long-run growth rate with 1.2 percentage point, and the coefficient on the spending share suggests that a similar increase in the public spending-to-GDP ratio reduces the long-run growth rate with 1.1 percentage point.

Rows 2 and 8 show our unsuccessful attempt of replicating these results using the standard TSLS-routine provided by STATA. The point estimates of the coefficients in front of the tax and spending variables are more than 50 percent smaller than those reported by FH, with *t*-ratios of 0.86 and 1.31. It is worth repeating that these regressions are based on the data

[3] In response to our inability to replicate their results, FH indicated that the probable reason was that their own regressions had in fact relied on an alternative data set on potential GDP – a data set that they, however, could no longer recover. While FH promised to provide us with the missing data, they failed to do so.

that we actually received from FH, and not on the missing data set that FH later claimed to have used in obtaining the results shown in rows 1 and 7.

The remaining rows of Table 1 show the results when we proceed using the growth rate of actual rather than potential GDP as the dependent variable, i.e. we adopt the dependent variable used by FH in all their equations that do not instrument the tax and spending shares. This has some advantages. First, actual GDP growth is certainly the dependent variable that would be the preferred alternative by most students of cross-country growth, and – unlike series on potential GDP – it is not sensitive to the precise way of distinguishing between trends and cycles. Second, since we have been able to replicate the OLS-regressions of FH that use actual GDP, we can focus our discussion on issues of econometric methodology rather than missing data.

Table 1 about here

Rows 3 and 9 report the results when we use the econometric specification and instrument set of FH, but use actual rather than potential GDP growth as the dependent variable. The point estimates differ only marginally from those shown in rows 1 and 7, though the *t*-ratios are lower, and the tax share is no longer statistically significant at the five-percent level. However, a troublesome aspect of these regressions is that the Sargan test of overidentifying restrictions indicates that the instrument set is invalid. When we regress the TSLS residuals on all our exogenous variables, and test for the joint statistical significance of the instruments, we can reject the null that our instruments are uncorrelated with the residuals with *p*-values 0.004 (row 3) and 0.011 (row 9), see column 8 of the table. Thus, some of the many instruments used by FH are not exogenous, which will bias the results.

Some experimentation suggests that the departure from exogeneity can be traced to two characteristics of the econometric specification of FH. First, in spite of estimating a model in first differences FH introduce country dummy variables in the first-stage instrument regressions. This procedure, which is tantamount to saying that there are country-specific trends in the long-run growth rates of the tax and spending shares, need or need not be an appropriate one. However, adding country specific fixed effects to the set of instruments will in this case induce a correlation between the instruments and the TSLS-residuals. Rows 4 and 10 shows the results when we drop the country dummy variables from the first-stage instrument regressions. Now, the Sargan instrument validity test fails to reject the null that our instruments are uncorrelated with the residuals. At the same time the coefficient on the tax share in row 4 changes sign from negative to positive. The coefficient on the expenditure share in row 10 remains negative, but drops from –0.119 to –0.051, and since we lose explanatory power in the first-stage regressions when we drop the country dummies the TSLS standard errors increase to such an extent that the *t*-value drops to 0.41.

Second, it is a common – though potentially hazardous – procedure in the cross-country growth literature to base identification of structural relationships by promoting lagged endogenous variables to exogenous instruments. Apart from a few brief references to the literature on dynamic panel analysis FH do not pay any attention at all to the substantive issues. In fact, their way of dating instruments is theoretically invalid. FH instrument the first differences of the tax and public spending variables with once lagged levels of taxes and public spending (and with levels and first differences of population and initial GDP). It is a simple exercise in algebra to show that this will induce a bias in the first-stage regressions; see Agell, Lindh and Ohlsson (1999). Moreover, the first difference specification implies that the lagged growth rate of GDP will appear along with the other explanatory variables. FH treat this lagged growth rate as an exogenous right-hand side variable, while it in fact will be correlated with the error term.[4]

Rows 5 and 11 show the results when we instrument the first differences of the tax and public spending variables with theoretically valid tax and spending shares dated *t*-2, and rows 6 and 12 add variables dated *t*-2 as instruments for the lagged growth rate.[5] As a result the coefficient on the tax share remains positive, while the coefficient on the spending share takes a small positive number in row 11, and a small negative number in row 12. In all cases the coefficients on the tax and spending shares are estimated with very low precision, while the Sargan instrument test gives no indication of lingering endogeneity problems.

It goes without saying that the results of Table 1, which show little evidence of convergence towards the mean, in no way prove that the size of the public sector is of no consequence for the economic growth rate. Though we confidently believe that our new specifications improve on the ones of FH, there remain several unresolved issues, issues that are germane to the cross-country growth literature. First, in *all* the specifications of the table we have a problem of weak instruments, in the sense that the partial correlation between the instruments and the endogenous tax and spending shares is low. This can be seen from the penultimate column, which presents the first stage *F*-statistic for the null-hypothesis that the instruments do not enter the first stage regressions for the tax and spending variables. As shown by Staiger and Stock (1997) *TSLS* estimates and confidence intervals can be highly unreliable in situations when the first stage *F* statistic is less than ten. In the specifications shown in Table 1 the *F* statistic lies in an interval from 1.58 to 3.74!

---

[4] This is a well-known result in the literature on dynamic panel models, see e.g. Baltagi (2001). A failure to instrument lagged endogenous variables will give rise to a large bias in panels with a small time dimension, which certainly is the case of FH.

[5] A possible objection to using instruments dated *t*-2 is that one loses degrees of freedom, a precious commodity in cross-country growth regressions. But the loss of observations and degrees of freedom is hardly an argument (as implicitly argued by FH in note 15 on p. 1511) for using non-valid instruments.

Second, we have followed the common, and probably unavoidable, practice of using lagged endogenous variables as instruments for current fiscal policy. This is however a mechanical procedure, which gives us little insight about the economic and other factors that determine a country's long term tax and spending policies. Moreover, lagged endogenous variables create problems in the presence of a serially correlated error term.

Finally, it is important to check whether one's preferred specification checks with the results from other sources. Students of the relationship between economic growth and the size of the public sector often tend to focus attention to the stability of the coefficient in front of the public sector variable(s). But it is also important to check whether the other variables of the model show up with plausible coefficients. It is a clear sign of misspecification if some of these other variables go wild, or deviate substantially from prior empirical studies. We can see from Table 1 that in no specification do we estimate a positive and statistically significant coefficient in front of the physical investment variable. We can also see that several specifications produce a very large negative coefficient in front of the initial GDP variable, suggesting an exceptionally large convergence effect. These "crazy" coefficients strongly suggest that there is some remaining homework to take care of.

4. Conclusions

FH claim that they solve a number of difficult econometric issues that has been identified by recent surveys of the cross-country growth literature. We do not agree. While the fixed effect estimator adopted by FH offers an easy remedy to endogeneity problems that can be traced to unobservable (time-invariant) country heterogeneity, it does not solve the challenging problem of reverse causation because of mutual interactions between economic growth and fiscal policy. We have shown that standard tests of instrument validity in fact show that FH rely on invalid instruments. Moreover, using theoretically valid instruments we find that the estimated partial correlation between size of the public sector and economic growth is statistically insignificant and highly unstable across specifications. A policy-maker who wants to promote growth is well-advised to look for other evidence than cross-country growth regressions.

Table 1: Change in economic growth rate, panel regressions for 23/22 OECD countries over 5-year periods between 1970-1995, by two-stage least squares.

	Δ tax share	*Δ exp. share*	*Initial GDP*	*Δ Inv. share*	*GDP-growth*	*Country dummies as IV*	*Lag of tax/spend. share as IV*	*Sargan p-value*	*F-value first stage regression*	*Obs.*
Tax share										
1. FH, Table 3	-0.12		-0.027	0.071	Potential.	Yes	1	–	–	65
	(-3.40)		(-2.00)	(2.83)						
2. Replication	-.054		-0.004	0.088	Potential	Yes	1	0.677	2.13	61
	(-0.86)		(-0.19)	(2.29)						
3. Actual growth instead	-0.141		-0.165	-0.026	Actual	Yes	1	0.004	2.46	92
	(-1.83)		(-7.34)	(-0.048)						
4. No country dummies	0.015		-0.169	-0.025	Actual	No	1	0.42	2.02	92
	(0.08)		(-7.06)	(-0.49)						
5. Valid lag length	0.104		-0.205	-0.038	Actual	No	2	0.744	3.47	69
	(0.61)		(-6.92)	(-0.66)						
6. Instrument for initial GDP	0.041		-0.112	-.0.038	Actual	No	2	0.963	3.74	69
	(0.22)		(-1.10)	(-0.63)						
Expenditure share										
7. FH, Table 3		-0.11	-0.037	0.058	Potential	Yes	1	–	–	64
		(-5.59)	(-3.12)	(2.88)						
8. Replication		-0.053	-0.013	0.069	Potential	Yes	1	0.80	2.47	60
		(-1.31)	(-0.63)	(1.71)						
9. Actual growth instead		-0.119	-0.162	-0.080	Actual	Yes	1	0.011	2.75	87
		(-2.58)	(-7.08)	(-1.68)						
10. No country dummies		-0.051	-0.161	-0.070	Actual	No	1	0.212	1.58	87
		(-0.41)	(-6.85)	(-1.37)						
11. Valid lag length		0.023	-0.195	-0.067	Actual	No	2	0.277	1.74	65
		(0.17)	(-6.69)	(-1.14)						
12. Instrument for initial GDP		-0.016	-0.063	-0.097	Actual	No	2	0.794	1.77	65
		(-0.10)	(-0.67)	(-1.39)						

t-statistics in parentheses. All models include the first difference of the average annual growth rate of the labor force, the first difference of the annual growth rate of the average years of schooling in the total population and period dummy variables as regressors. In addition to the instruments as specified in the Table, all models include the size of the population and the first difference of the size of the population as instruments. When initial GDP is used as an exogenous variable (rows 1-5 and 7-11) the level of initial GDP is also used as an instrument. Column 9 reports the value of the F-statistic testing the hypothesis that the coefficients on the instruments are all zero in the first-stage regression. As regards rows 6 and 12, the value of the F-statistic corresponds to the instruments in the first-stage regression of the endogenous first difference of the tax and expenditure shares, respectively.

References

Agell, J., Lindh, T., Ohlsson, H., 1997, Growth and the public sector: a critical review essay. *European Journal of Political Economy* 13, 33-52.

Agell, J., Lindh, T., Ohlsson, H., 1999, Growth and the public sector: a reply. *European Journal of Political Economy* 15, 359-366.

Atkinson, A. B., 1995, The welfare state and economic performance. *National Tax Journal* 48, 171-198.

Baltagi, B. H., 2001, *Econometric analysis of panel data* (second edition). John Whiley & Sons, Ltd (Chichester).

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

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

Mankiw, N. G., 1995, The growth of nations. *Brookings Papers on Economic Activity*, Issue. 1, 373-431.

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

Slemrod, J., 1998, How costly is a large, redistributive public sector? *Swedish Economic Policy Review* 5, 87-105.

Staiger, D., Stock, J. H., 1997, Instrumental variables regression with weak instruments. *Econometrica* 65, 557-586.

Temple, J., 1999, The new growth evidence. *Journal of Economic Literature* 37, 112-156.