National Tax Journal
Vol 49 no. 4 (December 1996) pp. 665-73

MEASURING THE IMPACT OF TAX REFORM

ALAN J. AUERBACH *

Abstract - *This paper considers why so many questions about the economic effects of tax reforms remain unanswered, and draws implications for how economics can be used to evaluate and design tax changes. Tax reforms have proven difficult to assess for a variety of reasons, all related to the nonexperimental nature of empirical economic analysis. Though this makes continued reliance on theoretical predictions necessary, evaluations should distinguish clearly between theory and evidence. The paucity of evidence also militates against the enactment of major tax reforms, for dependence on such reforms limits our ability to adapt tax policy in response to new evidence.*

INTRODUCTION

The inability of economic research to provide clear and precise information about the economic impacts of tax policies has long frustrated policymakers. Although not necessarily written with this as their primary objective, the other papers in this symposium by Agell, Englund, and Södersten and Engen and Skinner (hereafter AES and ES) illustrate why such information has proved so hard to uncover.[1] In this brief paper, I will attempt to assess why so many questions remain unanswered and consider the implications for how economics can be used to evaluate tax reforms and how the reforms themselves should be structured.

WHY DON'T WE KNOW THE ANSWERS?

Economic theory provides researchers with powerful tools, suggesting the types of behavioral responses that are likely or feasible in response to a particular tax change. Indeed, much of what we "know" about the effects of tax reforms is really based primarily on economic theory, rather than direct observation. For example, we "know" that, other things equal, an excise tax on a particular commodity will reduce demand for that commodity, even if no such tax has been imposed in the past. Theory tells us that demand curves slope downward, and we have ample confirming evidence of this from other markets.[2] We may even "know" how responsive to the tax demand will be, based on estimates of the price elasticity of demand. Standard theory tells us that a tax-induced price increase should have the same impact on demand as

*Department of Economics, University of California, Berkeley, Berkeley, CA, 94720 and NBER, Cambridge, MA 02138.

National Tax Journal
Vol 49 no. 4 (December 1996) pp. 665-73

any other price increase of equal magnitude, and we have no evidence to the contrary.

Useful as it is, theory does not always offer clear predictions. Sometimes, the theory itself remains in dispute. In many instances, a wide range of outcomes is theoretically possible, and only convincing empirical evidence can narrow this range of possibilities. Here, the conclusions of AES and ES as they confront major policy questions are sobering. AES find it difficult to tease out the economic effects of Sweden's "tax reform of the century," while ES can offer only tentative conclusions about the effects of taxes on economic growth, impacts that over time potentially could exert an enormous influence on a country's standard of living. Written as they are by careful and informed scholars, the papers provide a catalog of the problems confronting empirical research, which I will attempt to summarize here.

Time Is Short

Sweden's 1991 tax reform took effect as the country was entering a severe recession. AES suggest that the recession cannot be treated as entirely unrelated to the tax reform—that certain elements of the reform acted to reduce aggregate demand and that monetary policy did not adequately compensate for this. But they do not argue that the recession was due primarily to the tax reform. This leaves them with the problem of distinguishing responses to the tax reform from changes in activity resulting from the independent macroeconomic downturn. As many of the reform's predicted changes were of a smaller magnitude than the changes likely to be associated with the recession alone, their problem is a very serious one. Ultimately, they are forced to base judgments of the reform, in large part, not on observed behavioral changes but on effects that, on the basis of theory, one should have expected the reform to produce. That is, they gain limited information from looking at what actually has happened to date.

With more time, one hopes, the cyclical factors will even out and more evidence can be uncovered. But the passage of time also inevitably brings with it other sources of behavioral change. Because economics (at least at the macroeconomic level) is not an experimental science, we will never get the clean experiment we desire.

For some changes, though, even a few more years may not be enough. Once we have enough data to control for cyclical variation, we might be able to determine the impact of a tax reform on, say, labor supply. But other questions, such as the effects of tax reform on the long-run growth rate, simply cannot be answered in the short run. As ES point out, it is important to distinguish between the short-run and long-run effects of tax policy on growth. In the short run, tax policy can affect growth by encouraging labor supply, increasing saving and capital accumulation, and improving the efficiency of the allocation of labor and capital in production. Over the longer run, once these effects have taken place, tax policy can influence the growth rate only to the extent that it has a permanent impact on the rate of technological progress. While economic research has succeeded in developing theories to trace out how these effects might occur, there is little empirical evidence that tax policy can influence a country's long-run growth rate. Indeed, there is some

National Tax Journal
Vol 49 no. 4 (December 1996) pp. 665-73

evidence to the contrary, that even in countries experiencing exceptional rates of growth, this growth cannot be attributed to induced changes in technology (Young, 1992). But research of this sort requires decades to be able to distinguish between transitory and long-run effects on growth.[3]

Tax Reforms Are Complex

Even when evidence is available, it is often difficult to determine what this evidence tells us about a tax reform's impact. A large-scale tax reform such as the Swedish reform of 1991 or the U.S. Tax Reform Act of 1986 (TRA 86) might have many provisions likely to influence some particular aspect of behavior. For example, if we wish to determine whether real estate investment's response to TRA 86 was consistent with theoretical predictions, and learn something of the relevant behavioral elasticities, we must take account not only of changes in depreciation provisions, marginal tax rates on ordinary income, and treatment of capital gains, but also the effects of passive loss restrictions and the changes in interest rates induced by the reform. We must account for a multitude of effects, many which are hard to evaluate. While we may have reasonably clear predictions of how a change in marginal tax rates influences investment, the incentive effects of changes in passive loss rules or similar restrictions are more difficult to quantify.

An even better illustration of this type of difficulty comes from the Swedish reform, which simplified the tax treatment of business investment by eliminating various incentive schemes, most notably the investment funds system, and compensated for this by reducing the statutory marginal tax rate. Because of the complexity of the previous tax regime, it was difficult to know how the tax system influenced investment before the reform (Auerbach, Hassett, and Södersten, 1995). Thus, it was hard to predict how the tax reform should have influenced investment. This type of problem—that we cannot estimate the effects of a tax change without knowing the effects of current provisions—seems likely to plague any major tax reform where one of the objectives is to simplify the tax system.

Sometimes, the data themselves are simply too limited to allow us to perform a serious evaluation of tax changes. For example, the typical cross-country study cited by ES considers the effects on growth of variations in the gross domestic product (GDP) shares of different types of taxes, without being able to account for the important distinction (noted by ES) between the effects of marginal tax rates and those of average tax rates or, to put it another way, between substitution effects and income effects. Even where better data are available, it is not always easy to measure these effects separately. For example, in studying the impact of TRA 86 on labor supply, it is much easier to estimate the change in an individual's marginal tax rate than the direct and indirect impacts of all provisions on that individual's purchasing power. Yet, without being able to estimate these impacts, we cannot tell how much of the observed labor supply response is due to the change in labor supply incentives.

Tax Policy Is Endogenous

Particularly in cases where long-run effects are at issue and time series from a single country are inadequate, researchers often resort to cross-country analysis as a substitute, arguing that one can evaluate the long-run effects of

National Tax Journal
Vol 49 no. 4 (December 1996) pp. 665-73

policies by comparing the relative performance of countries with and without the policy. Here, though, we face what ES refer to as the "Achilles' heel" of cross-country regressions—reverse causality. If countries are otherwise equal, why do they have different policies? If they are not otherwise equal (at least to the extent that we can control for observable differences), how can we be sure that the unobservable differences are not responsible both for the differences in growth and differences in tax policy, or perhaps for differences in growth which, in turn, are responsible for differences in tax policy? The answer, of course, is that we cannot be sure, nor can we often be even confident that the apparent effects of tax policy are being properly interpreted.

Careful researchers can improve their chances of full immersion in the river Styx by limiting their study to countries with similar nontax characteristics, or by controlling for unobservable differences among countries by looking at *differences* in growth rates among countries following the adoption of particular tax policies by some (the so-called *differences-in-differences* approach). But, because they do not use all available data, these approaches sacrifice some of the available information. Moreover, they do not offer a complete solution to the problem of endogeneity. There is no guarantee that our choice of comparison countries eliminates all important unobservable differences. The differences-in-differences approach also still has problems, which are discussed below.

Other Things Are Happening

For effects of a short-run nature, such as induced changes in labor supply or investment, a major problem confronting researchers, already discussed, is the confounding influences of contemporaneous macroeconomic fluctuations in wages and employment, interest rates and profitability, etc. To control for these macroeconomic effects, which strike all individuals and firms, many studies focus on the differential effects of tax changes in cross sections of individuals or among groups. With some luck, a reform will be expected to affect the behavior of some individuals or businesses more than others, and by comparing the changes in the behavior of these groups, we may be able to identify the effects of the tax change more generally.

For example, TRA 86 reduced the marginal tax rates on high-income individuals much more than on low- and middle-income individuals. Thus, we might consider the relatively large change in the marginal tax rates of high-income individuals as a *natural experiment*; by analogy to the terminology of true experiments, high-income individuals are the "treatment" group and other individuals are the "control" group. Eissa (1995) took this approach to estimate the labor supply response of married women, and Feldstein (1995) used it to estimate the responsiveness of taxable income. Their studies found larger increases in labor supply and taxable income for those with larger marginal tax rate reductions. Comparing the differences in behavioral effects to differences in marginal tax rate changes then provides us with an elasticity of response. Armed with this elasticity, we can estimate how the remainder of the taxpaying population responds to tax rate reductions, even though we cannot observe their responses directly.

Alas, the validity of this inference depends on how well the *natural*

National Tax Journal
Vol 49 no. 4 (December 1996) pp. 665-73

experiment resembles a true, well-designed experiment.[4] One problem is that, unlike in the case of a random trial, the control and treatment groups in the natural experiment are not drawn from the same population. Even if high-income people respond to marginal tax rate reductions, we are not necessarily justified in assuming that low-income people would respond in the same manner to comparable marginal tax rate reductions. These groups may differ in their preferences or in their ability to adjust hours of work or to shelter income from taxation. A second problem is that macroeconomic phenomena may hit the control and treatment groups differently. During this period of increasing wage inequality, high-income individuals faced relatively higher after-tax wages not only because of reductions in marginal tax rates, but also because of increases in their before-tax wages. While it may be possible to control for relative changes in wages, there may be other differential changes that are less easily observed. Finally, as discussed above, we need to take account not only of changes in marginal tax rates, but of other tax-related changes in income and incentives. If we fail to do so, and if these other effects differ between treatment and control groups (as is likely), we may attribute to marginal tax rate changes the effects of other provisions.

Some Effects Are Hard to Measure

Some of the potentially important effects of a tax reform are difficult to observe and measure. For example, both TRA 86 and Sweden's 1991 reform attempted to reduce the disparity of treatment among types of investment, supported by the argument that this would improve the allocation of capital and thereby reduce the deadweight loss from taxation.

Unlike changes in labor supply or investment, though, changes in deadweight loss are not immediately observable. To determine how well these reforms actually did in reducing deadweight loss, it would be necessary to measure not only the extent to which capital was reallocated, but also the extent to which before-tax returns rose as a result of these shifts to more socially productive uses. Yet, it is essentially impossible to observe before-tax returns on specific types of assets.

Summary

To determine the impact of a tax reform, it is necessary not only to develop theories of that tax reform's impact, but to test the theories. The lack of controlled experiments and of the ability to measure economic changes limits the scope for performing such evaluations. Thus, even potentially important economic effects may be difficult to uncover, particularly within a period of time when such information would be most useful.

ON THE METHODOLOGY OF TAX REFORM EVALUATIONS

All of the above limitations notwithstanding, economists have learned from tax reforms and can use what they have learned, in conjunction with some basic economic principles, to help guide policy decisions. However, in light of the situation, several useful principles should be applied to the analysis of tax reforms.

Distinguish Between Assumptions and Evidence

Given the many holes in our empirical evidence, it is natural and useful to fill in the gaps in our knowledge with theoretical predictions. For example,

National Tax Journal
Vol 49 no. 4 (December 1996) pp. 665-73

when AES argue that the Swedish tax reform's shift away from housing subsidies "did much to promote a less inefficient allocation of investment resources," they are assuming that housing investment generates no special positive externalities, such as a homeowner's increased commitment to community service. This is a quite reasonable assumption with which most economists would be comfortable, but it is an assumption nonetheless, rather than an empirical observation.

Another common assumption that would likely be somewhat more controversial is that investments in business equipment provide no positive externalities, for this is one of the channels through which the "new" growth theory posits that technological progress can be induced. Lacking evidence of this externality, we normally exclude it from our estimates, but we should make clear that we are doing so. In short, others cannot judge the strength of our conclusions without knowing the assumptions on which they are based.

Provide a Road Map of Reported Results

Large-scale tax reforms can affect behavior through many channels simultaneously. Indicating the relative importance of different channels helps others evaluate the conclusions. For example, during the 1995 budget showdown between President Clinton and Congress, the Congressional Budget Office (CBO) provided estimates of the deficit-reducing macroeconomic feedback effects of a seven-year balanced-budget policy (CBO, 1995). CBO was quite explicit that these effects came from one source: a reduction in the crowding out of private investment, which was predicted to lead to lower interest rates and less debt service and greater income because of capital deepening. The estimate assumed that there would be no change in the unemployment rate because of a successful coordination of monetary policy with this contractionary fiscal policy, an assumption that one might question in light of the finding by AES that the 1991 Swedish reform entailed considerable short-run adjustment costs because "the government took a rather careless attitude to the transition problem."

Another illustration comes from attempts to relate tax cuts and subsequent economic growth. ES argue that "it is a difficult task to sort out" whether the strong U.S. economic growth in the mid-1980s and the Reagan tax cuts that began in 1981 were related via traditional Keynesian stimulus or supply-side reductions in marginal tax rates. But such distinctions are important if, for example, the next tax cut we consider takes place when the economy has much less room for demand-induced expansion than it did in 1982.

Reconcile Micro and Macro Effects

Predictions can be provided either at the micro or the macro level. For example, we might say that a tax cut will increase the labor supply of married women by two percent, or that it will increase GDP by 0.5 percent. In many cases, simply asking whether microlevel and macro-level predictions are consistent—i.e., following through on either the "bottom-up" or the "top-down" approach discussed by ES—can help determine whether the predictions make sense. Indeed, even in cases where only micro or only macro estimates are provided, we can perform our own translation in order to evaluate how realistic the predictions are.

National Tax Journal
Vol 49 no. 4 (December 1996) pp. 665-73

One can think of several instances where such attempts at reconciliation have been helpful. During the last decade, considerable empirical evidence (e.g., Venti and Wise, 1992) has been put forward suggesting that a large portion of individual contributions to Individual Retirement Accounts has come through new saving rather than saving that would have been done anyway or shifts of existing assets. But this apparent increase in saving seems to have had no positive effect on aggregate saving (Engen, Gale, and Scholz, 1994). The lack of confirming aggregate evidence does not disprove the existence of individual effects, but it does force those arguing in favor of such effects to reconcile the apparent inconsistency by explaining what other factors might have caused saving to decline.

Consider the debate over the effects of a capital gains tax cut on realizations. A central controversy has been over the extent to which observed short-run elasticities overstate the long-run increase in realizations in response to a permanent tax cut. While direct evaluation of these elasticity estimates has suggested that long-run elasticities are considerably smaller (e.g., Auerbach, 1988; Burman and Randolph, 1994), this conclusion gets further support from calculations that show just how extreme the underlying change in the frequency of trading would have to be to produce such large behavioral responses (Auerbach, 1989).

A final example comes from the recent debate over how much the shift from an income tax to a simple consumption tax (such as a value-added tax, a retail sales tax, or a flat tax) would increase GDP in the short run. We can judge at least the plausibility of claims of massive growth by asking what changes in labor supply and saving these increases would require.

Be Appropriately Humble

A tension exists between wishing to make results clear and comprehensible, on the one hand, and offering needed qualification, on the other. The conclusions that AES and ES offer to their evaluations show that one can provide information without oversimplifying or overstating one's knowledge.

IMPLICATIONS FOR THE DESIGN OF TAX POLICY

What do the findings of AES and ES imply about the design of tax policy? One sort of lesson is about the economic merits of certain tax changes themselves. AES conclude that Sweden, which started with a very complex tax structure and high marginal tax rates, improved economic efficiency through its reform aimed at simplification and base-broadening. But they also find that the changes in real behavior, such as saving and labor supply, were modest, and the transition costs considerable. Thus, the reform was beneficial, particularly for a country with Sweden's initial tax system, but not the panacea some might have predicted. ES conclude that tax policy can affect at least short-run economic growth by enough to make an important difference in a country's standard of living.

But equally important are the lessons from these two papers about the process of tax reform itself. First, major changes in tax policy are not permanent or even long-lived. AES conclude that "even a 'tax reform of the century' implemented with such force as TR 91 did not stay unaffected for very long." They point in particular to marginal rate increases that broke the

National Tax Journal
Vol 49 no. 4 (December 1996) pp. 665-73

spirit of the reform's trade of tax expenditures for lower rates. One notices a disturbing similarity to the evolution of tax policy in the United States since the passage of TRA 86 (Auerbach and Slemrod, forthcoming, 1997). This finding is important, because it counters one of the arguments often put forward in favor of large-scale tax reforms over more incremental ones—that the large-scale reforms overcome entrenched interests and permanently alter the political landscape, making reversion to earlier policies more difficult.

Second, the economic effects of policies are difficult to evaluate; long-run effects are the most difficult. ES conclude, for example, that we really do not know the extent to which tax policy can affect long-run growth. This, too, offers an argument against major tax reforms that produce marked changes in economic incentives, because such changes deprive us of the opportunity to learn from our mistakes.

For example, a central argument for a shift to consumption taxation is that this shift will increase individual saving. This is what accepted economic theory tells us, but we have relatively little supporting evidence. Even evidence on the efficacy of saving incentives may not be that informative regarding a move toward a broad-based consumption tax, as such incentives may work for other reasons, such as employer promotion and education.[5] At present, housing equity aside, a considerable share of the saving that most people do is through employer-sponsored pension (including 401(k)) plans. If such plans lose their tax advantage relative to other saving and cease to be offered to the extent that they are now, can we be sure that private saving will not be adversely affected? Even if we believe not, our uncertainty should temper our willingness to experience the major transition problems associated with the shift to a major new tax system.

In short, learning by doing is an option not just for the private sector. As compensation for the potential gains sacrificed by delaying full implementation of the "best" tax policy, we have the increased certainty that it really is the best tax policy. Should the change not turn out to be for the best, we can also take comfort from the prospect that tax changes, even "permanent" ones (and, alas, even good ones), may not last very long.

ENDNOTES

I am grateful to Joel Slemrod for comments on an earlier draft of this paper.

1 A similar conclusion follows from the consideration of the effects of TRA 86 by Auerbach and Slemrod (forthcoming, 1997).

2 Indeed, the evidence extends beyond humans. See, for example, Kagel et al. (1981).

3 This problem arises in other contexts in macroeconomics as, for example, we attempt to determine whether GDP cycles around a trend or follows a random walk, or whether stock markets are efficient or exhibit mean reversion. In each case, the fundamental problem is the same: we can only distinguish between the alternatives after long periods of observation.

4 For further discussion, see Heckman (1996).

5 See the discussion in Bernheim (forthcoming,1997).

REFERENCES

Agell, Jonas, Peter Englund, and Jan Södersten. "Tax Reform of the Century—the Swedish Experiment." *National Tax Journal* 49 No. 4 (December, 1996): 641–62.

Auerbach, Alan J. "Capital Gains Taxation in the United States: Realizations, Revenue, and Rhetoric." *Brookings Papers on Economic Activity* 19 No. 2 (1988): 595–631.

Auerbach, Alan J. "Capital Gains Taxation and Tax Reform." *National Tax Journal* 42 No. 3 (September, 1989): 391–401.

Auerbach, Alan J., Kevin Hassett, and Jan Södersten. "Taxation and Corporate Investment: The Impact of the 1991 Swedish Tax

National Tax Journal
Vol 49 no. 4 (December 1996) pp. 665-73

Reform." *Swedish Economic Policy Review* 2 No. 2 (Autumn, 1995): 361–83.

Auerbach, Alan J., and Joel Slemrod. "The Economic Effects of the Tax Reform Act of 1986." *Journal of Economic Literature* 35 No. 1 (forthcoming, March, 1997).

Bernheim, B. Douglas. "Rethinking Saving Incentives." In *Fiscal Policy: Lessons from Economic Research*, edited by Alan J. Auerbach. Cambridge, MA: MIT Press (forthcoming, 1997).

Burman, Leonard, and William Randolph. "Measuring Permanent Responses to Capital-Gains Tax Changes in Panel Data." *American Economic Review* 84 No. 4 (September, 1994): 794–809.

Eissa, Nada. "Taxation and Labor Supply of Married Women: The Tax Reform Act of 1986 as a Natural Experiment." NBER Working Paper No. 5023. Cambridge, MA: National Bureau of Economic Research, 1995.

Engen, Eric M., William G. Gale, and J. Karl Scholz. "Do Saving Incentives Work?" *Brookings Papers on Economic Activity* 25 No. 1 (1994): 85–151.

Engen, Eric M., and Jonathan Skinner. "Taxation and Economic Growth." *National Tax Journal* 49 No. 4 (December, 1996): 615–40.

Feldstein, Martin. "The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act." *Journal of Political Economy* 103 No. 3 (June, 1995): 551–72.

Heckman, James J. "Comment." In *Empirical Foundations of Household Taxation*, edited by Martin Feldstein and James Poterba, 32–8. Chicago: University of Chicago Press, 1995.

Kagel, John H., Raymond C. Battalio, Howard Rachlin, and Leonard Green. "Demand Curves for Animal Consumers." *Quarterly Journal of Economics* 96 No. 1 (February, 1981): 1–15.

U.S. Congressional Budget Office. *An Analysis of the President's Budgetary Proposals for Fiscal Year 1996*. Washington, D.C.: Government Printing Office, April, 1995.

Venti, Steven, and David Wise. "Government Policy and Personal Retirement Saving." In *Tax Policy and the Economy* 6, edited by James M. Poterba, 1–41. Cambridge, MA: MIT Press, 1992.

Young, Alwyn. "A Tale of Two Cities: Factor Accumulation and Technical Change in Hong Kong and Singapore." *NBER Macroeconomics Annual*, 13–54. Cambridge, MA: MIT Press, 1992.