Dynamic Scoring: An Introduction to the Issues

Alan J. Auerbach[*]

The controversy over dynamic scoring in the United States is not new.[1] But the possibility that dynamic scoring might be adopted has moved closer to reality recently, with initial attempts by the Congressional Budget Office (CBO) and the Joint Committee on Taxation (JCT) shedding light on what the process would entail. Thus, it is an appropriate time to review the issues.[2]

What is Dynamic Scoring?

Working together, CBO and JCT provide two types of revenue and expenditure projections for Congress: (1) forecasts of expenditures and revenues based on assumptions about what "current" policy is, and (2) forecasts of the changes in expenditures and revenues that would result from proposed legislation. The first type of forecast is known as a *baseline*. The second type is called *scoring*. The two types of forecasts are interdependent. Projected effects of legislation depend on the starting point, so the baseline affects scoring. Updates of the baseline reflect the estimated impact of legislation, so scoring affects the baseline.

On the revenue side, where dynamic scoring has received the greatest attention, the baseline forecasts are provided to Congress by CBO, while scoring is done by the JCT.[3] In a sense, the debate over dynamic scoring may be traced to an inconsistency in the way that legislative changes are handled by the two agencies. Macroeconomic feedback effects are ultimately incorporated in CBO's baseline, but are not attributed by JCT to any source when the proposals are scored. This inconsistency is most easily understood in terms of relative informational requirements. One can update a baseline to account for macroeconomic changes without attributing these changes to particular legislation. Indeed, the models used to construct baselines are not specified at the level of detail that would be necessary to do so.

Under standard scoring procedures, JCT scores proposed legislation taking the macroeconomic forecast underlying the CBO baseline – including nominal GDP and other aggregates – as given. It is misleading to refer to these revenue estimates as "static," because they do take certain behavioral effects into account. For example, when JCT scores a change in the capital gains tax rate, it includes the revenue offset associated with changes in capital gains realizations. But missing is any further impact on revenues or expenditures that might result from induced changes in output, interest rates, or other elements of the macroeconomic forecast. Thus, current scoring estimates incorporate "micro" dynamic effects, but not "macro" ones.

Arguments in Favor of Dynamic Scoring

Arguments in favor of dynamic scoring generally emphasize its consistency with the principles of economic and statistical analysis.

1. *Dynamic scoring makes use of all available information.* Perhaps the simplest argument in favor of dynamic scoring is that failing to carry it out amounts to throwing away valuable information. Under current methods, the official score is not the expected impact on revenue.

2. *The lack of dynamic scoring biases the legislative process against tax cuts.* Proponents argue that tax cuts that encourage productive economic activity do not get adequate "credit" from the scoring methodology, and thus are viewed as more costly to enact. For example, a cut in marginal tax rates that increases labor supply is treated as having no such impact, and so is viewed less favorably by the budget process than it should be. Conversely, a distortionary tax increase is viewed with insufficient alarm, because some of its negative effects are ignored.

3. *Current methods impose constraints on scoring that are at odds with economic evidence.* While much of the criticism is directed at its lack of supply-side responses, the current approach has other anomalies. For example, the requirement that nominal GDP be held fixed means that

an increase in indirect taxes must lead to a dollar-for-dollar reduction in factor incomes, even though there is no evidence to suggest that this is what would happen.

4. *Advances in technology and economics strengthen the case for dynamic scoring.* Current scoring procedures date to a period of more primitive computing technology. Economic analysis and modeling have also advanced, with growing evidence of the importance of taxpayer responses to government policy changes and the development of a new generation of sophisticated economic models capable of gauging the magnitudes of these responses.

Arguments Against Dynamic Scoring

Arguments against dynamic scoring point to the technical difficulties of doing it correctly and the political reality of the context in which scoring is done.

1. *Dynamic scoring must rely more on assumptions and is susceptible to political pressure.* Often, there is econometric evidence about micro responses but not about the further macroeconomic effects. For example, there are a range of estimates about the impact of capital gains taxes on capital gains realizations in the short run and the long run. But there are few if any actual estimates of the impact on the economy as a whole.[4] With little economic evidence to support any one of a broad range of possible assumptions, it will be more difficult for choices to be made without being challenged by interest groups on one side or the other.

2. *Dynamic scoring would require an impractical integration with the baseline process.* Baselines are produced much less frequently than legislation. To make dynamic scoring and the baseline consistent would require updating the baseline for each piece of legislation passed. This change in procedure would (1) require a much closer integration of the activities of two distinct organizations; (2) increase the work required to produce baselines several-fold; and (3) not be feasible without a substantial change in the type of model currently used to produce baselines.

3. *Dynamic scoring would have to account for all channels and expenditure-side changes.* Dynamic effects of revenue legislation would come not only through supply-side incentive effects, but also through budgetary effects. For example, tax cuts that encourage economic activity could still have negative macroeconomic effects through the crowding out of capital formation. Thus, all revenue provisions, not just those with significant incentive effects, would need to be evaluated. The same argument applies to changes on the expenditure side.

4. *Dynamic scoring requires assumptions about monetary and fiscal policy reactions.* The federal government faces a long-run budget constraint, so that any measure that upsets long-run budget balance must ultimately bring forth further fiscal changes. There is no such thing as a "permanent" tax cut if the tax cut induces reductions in revenue. It will be necessary to specify how such revenue losses are to be offset – predicting what further legislation Congress will pass in response – and the specification will influence the dynamic scoring of the legislation itself. The same type of problem arises with respect to monetary policy responses.

What is the Aim of Budget Scoring?

Arguments about dynamic scoring usually jump right to the specifics, but it is helpful to step back and ask what purpose is meant to be served by revenue estimates. A revenue estimate tells us the predicted impact of a change in policy on aggregate revenues. As conventionally reported, the revenue score does not even convey the full impact on the budget deficit, for it excludes associated changes in debt service, although this omission is not a major issue as long as it is recognized. Even taking debt service into account, though, a revenue score still conveys only some information about a policy's impact. Dynamic scoring does not necessarily provide better information regarding other effects of policy that are sometimes inferred from the revenue score.

Consider first the impact of a policy change on the level of macroeconomic activity. In the Keynesian tradition, it is customary to see policies that produce larger deficits as being more expansionary, other things being equal. This equation of deficits and fiscal thrust is not particularly accurate, for policies with similar revenue costs may differ substantially in their effects on aggregate activity. But it would be even less accurate if dynamic scoring were adopted, precisely because dynamic scoring will reflect predicted macroeconomic changes.

For example, suppose two tax cuts each would reduce revenues by $100 billion over a ten-year period, according to conventional revenue scoring. Let us assume that the first policy has little impact on economic activity other than reducing national saving, and as a consequence is estimated as having an even larger revenue cost when evaluated using dynamic scoring methods that incorporate the predicted decline in capital income taxes. Suppose that the second policy stimulates economic activity so that dynamic scoring reduces the projected revenue loss below $100 billion. Using the net impact on the deficit as a measure of a policy's impact on economic activity is clearly wrong in this case, for it would characterize the first policy as more significant that the second, when the opposite is true. Thus, the use of dynamic scoring makes measured effects on the deficit an even poorer measure of demand-side stimulus than under present scoring methods, which would see the policies as having similar demand-side impacts.

Revenue estimates are also sometimes interpreted as measuring the change in the burden of taxation. But there is a distinction between tax payments and tax burden, and this distinction increases with the size of taxpayer responses (JCT 2003). Consider again a reduction in capital gains taxes. To a first approximation, the taxpayer benefit equals the reduction in taxes due, ignoring any behavioral response. An induced increase in realizations is a voluntary event, and so must make the taxpayer even better off, but the revenue estimate is reduced by this response,

even under current methodology. Adding in further potential increases in revenue that might result under dynamic scoring would only widen the gap between tax burden and tax revenues further. Again, the use of dynamic scoring makes the revenue estimate less useful as a measure of something else, in this case the burden of taxation.

Neither of these incidental effects provides a convincing argument against the use of dynamic scoring, but they do emphasize the need for additional tools, including distributional and macroeconomic analysis. Indeed, the distinct information from different analyses can provide policy guidance. If a potential tax reduction is estimated to have little net revenue cost but a large reduction in tax burden, for example, this divergence points to the very distortionary nature of the tax being reduced, giving support to the contemplated policy change.

Distributional analysis has been subject to its own continuing controversy over appropriate methods – a cautionary tale in contemplating the introduction of dynamic scoring. Macroeconomic analysis has been a much more occasional element of policy analysis. However, were dynamic scoring adopted, it would require macroeconomic analysis as an input, and thus there would be no reason why such analysis could not be provided separately. Even if all legislative proposals were accompanied by analyses of distributional and macroeconomic effects, these evaluations would differ in one very important respect from the revenue score – they would provide information, but would not play any mandated role in the budget process. This brings up one final question about the desirability of dynamic revenue estimation: if it is done, should it be provided as informational analysis or as part of the official revenue score? Many of the arguments against dynamic scoring focus precisely on its use as an official measure, rather than on the provision of dynamic revenue estimates. Thus, there is likely to be far less disagreement about the provision of dynamic revenue estimates than about their official use.

How Should Dynamic Scoring (or Analysis) be Done?

Recent efforts undertaken by JCT and CBO to carry out dynamic analysis of actual legislation shed light on how dynamic scoring might be carried out. In March, 2003, CBO (2003a) produced an analysis of the President's 2004 budget (further explained in CBO 2003b) that included a range of estimates of the package's macroeconomic feedback effects. In May, 2003, JCT (2003) provided a dynamic analysis of the House version of the Jobs and Growth Reconciliation Tax Act of 2003. These studies' conclusions received the most attention, but summary results cannot be easily interpreted without a review of the underlying choices with respect to methodology, which reflect many of the issues that have arisen in the debate over dynamic scoring.

1. *Model specification.* Both agencies utilized several models. CBO used a "textbook" growth model, an infinite-horizon growth model, closed-economy and open-economy life-cycle growth models, and two traditional macroeconomic forecasting models. Only the life-cycle and infinite horizon models were forward-looking. Only the traditional forecasting models incorporated Keynesian demand-side effects. The infinite horizon model incorporated a strong implicit bequest motive absent from the other models. In short, the importance of several channels varied sharply across the six models. JCT took the same general approach, utilizing one forward-looking, life-cycle growth model and two more traditional macroeconomic forecasting models. Using multiple models offers insights into the impact of different assumptions, while allowing one to avoid choosing the "right" model and to side-step the problem that no one model has all the features necessary for a complete evaluation of complex legislation.

2. *Filling in the necessary fiscal details.* The government's intertemporal budget constraint requires that tax cuts or tax increases adopted by legislation have a compensating response at some future date. For forward-looking models, these future actions affect behavior during the

ten-year budget window considered by CBO and JCT, even if the actions occur after the ten-year period. Thus, further fiscal assumptions are needed to close the forward-looking models used. CBO assumed that the debt-GDP ratio would be stabilized after 2013 either by a reduction in government consumption or an increase in lump-sum taxes. JCT assumed made similar assumptions, except that it achieved the tax offset through an increase in individual income tax rates. No attempt was made to evaluate the likelihood of the respective future fiscal policies. A possible improvement would be to incorporate fiscal reaction functions, based on past behavior.

3. *Assumed Reaction by the Fed.* For the more traditional forecasting models that include money and monetary policy, some assumption is needed about the Fed's response to fiscal policy changes. JCT considered three possible monetary policy reactions: "aggressive" (the Fed completely counteracts demand stimulus), "neutral" (the Fed 'ignores' the fiscal and keeps the growth rate of the money supply constant), and (in a different model) an intermediate policy in which the Fed follows a Taylor rule and allows some but not all of the demand stimulus to work through the economy. In its analysis using traditional forecasting models, CBO adopted the models' assumed reaction functions, but also used simulations that tightened interest rates to choke off demand stimulus in order to estimate the magnitude of supply-side effects.

4. *Other assumptions.* Both the President's budget analyzed by CBO and the House bill analyzed by JCT contained complex provisions that had to be evaluated through preliminary analysis before being incorporated in the forecasting models. Perhaps most vexing were provisions to reduce the taxation of dividends and capital gains. None of the models used by either agency are specified at a level of detail necessary to evaluate the effects of such provisions, so both agencies produced "off-model" estimates of the impact of the tax changes on the effective tax rate on capital income and the incentive to invest, which were then fed into the various models.

5. *Summary of findings*. CBO's analysis found a range of possible dynamic effects on the economy and the budget, according to the model used and assumptions about future fiscal policy. The range of results included both positive and negative effects over both five- and ten-year windows on economic growth (CBO 2003a, Tables 15 and 17), and on the budget (CBO 2003a, Tables 16 and 18). Taking into account macroeconomic feedback effects changed the budget cost of the President's proposal between -10 percent and 29 percent over five years, and between -15 percent and 17 percent over ten years.[5] JCT (2003, Table 2) found a small positive impact for H.R. 2 on real GDP in all simulations over the first five years, but a zero or negative effect on GDP during the second five years, the latter largely because of cumulative crowding-out effects. Incorporating dynamic analysis (JCT 2003, Table 5) reduced the net revenue cost of H.R. 2 by 6-28 percent over the first five years, and by 3-23 percent over the second five years. The difference in budgetary impact between the two studies (CBO finding both positive and negative feedback effects, JCT finding all feedback effects to be deficit-reducing) may be largely due to the composition of the packages being evaluated. While JCT analyzed a tax cut proposal, CBO analyzed a package that included both tax cuts and spending increases. Evaluation of proposals' components, rather than just proposals in the aggregate, could provide useful information.

From these first attempts, it seems clear that dynamic analysis has value, but also that adjustments to estimates are smaller than some might have expected. The process to date offers some support to those on both sides of the debate. On the one hand, the ability of CBO and JCT to produce dynamic analyses of complex, realistic proposals lends credence to the argument that dynamic analysis and, indeed, dynamic scoring may be feasible. On the other hand, the many models used and the many assumptions needed will leave many with doubts about the quality of these estimates and how they would fit into the budget scoring process as currently structured.

References

Auerbach, Alan J. "Dynamic Revenue Estimation," Journal of Economic Perspectives, Winter 1996, 10(1), pp. 141-157.

Congressional Budget Office. An Analysis of the President's Budgetary Proposals for Fiscal Year 2004, March 2003a.

Congressional Budget Office. How CBO Analyzed the Macroeconomic Effects of the President's Budget, July 2003b.

Council of Economic Advisors. Economic Report of the President, 2004.

Feldstein, Martin, "The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act," Journal of Political Economy, June 1995, 103(3), pp. 551-72.

Joint Committee on Taxation. Methodology and Issues in Measuring Changes in the Distribution of Tax Burdens. JCS-7-93, June 14, 1993.

Joint Committee on Taxation. Macroeconomic Analysis of H.R. 2, The "Jobs and Growth Reconciliation Tax Act of 2003" Prepared by the Staff of the Joint Committee on Taxation, May 2003.

Notes

* Department of Economics, University of California, Berkeley, CA 94720-3880, and NBER, Cambridge, MA 02138. I am grateful to Kevin Hassett for comments on an earlier draft.

[1] See, for example, the discussion in Auerbach (1996).

[2] For another recent review, see the Economic Report of the President (2004), Chapter 5.

[3] On the expenditure side, both procedures are done by CBO.

[4] There may be evidence that incorporates some, but not all, macroeconomic responses. For example, the literature on responses of taxable income to marginal tax rates (e.g., Feldstein 1995) incorporates both "micro" shifting responses and "macro" labor supply responses.

[5] The range over ten years excludes the two traditional forecasting models, one of which was responsible for the largest positive budget impact over the five-year window.