Statement before the Joint Economic Committee

"On the Dynamic Scoring of Fiscal Policy"

Kevin A. Hassett, Ph.D.
Director of Economic Policy Studies
& Resident Scholar
American Enterprise Institute

July 28, 2015

The views expressed are those of the author. The American Enterprise Institute does not hold institutional positions on any issue.

Introduction

Mr. Chairman and members of the committee: it is a great privilege to have this opportunity to appear before you today. I am an economist who is the director of economic policy studies at the American Enterprise Institute, a think tank here in Washington. Much of the research I have undertaken as a professional economist examines taxation and the consequences of tax policy. I come here today, specifically, to provide testimony on what is known as dynamic scoring in tax policy circles.

At the outset it is important to emphasize that the economics profession has made tremendous strides in the modeling of the impact of fiscal policy on the economy over the past few decades, and there is an ample amount of evidence to point to that suggests that a carefully designed fundamental tax reform could lead to a significant improvement in the wellbeing of Americans. Yet talk of tax reform has not produced truly significant action since the 1980s. I believe that one reason we have made such little progress is that scoring methods do not account for the impact that sound proposals would have on the overall economy. In my testimony today, I discuss the challenges facing those who would hope to do better.

What is dynamic scoring?

Perhaps one way to understand the concept of "dynamic scoring" of tax legislation is to examine the two words. "Scoring" refers to the process of estimating the effects of a given policy proposal. In the U.S. Congressional context, the JCT staff scores proposed legislation, focusing their estimates on the legislation's effects on government revenue. These estimates provide Congress with a guide to thinking about the revenue implications of proposed changes. Historically, static scores of tax proposals have often relied on enormous micro data files, giving the relevant staffs an impressive ability to account for compositional issues when evaluating policies. These microsimulation results, however, have, until this year, been the final word.

"Dynamic" refers to estimating these budgetary effects in a way that allows the proposed legislation in question to alter the overall level of economic activity. That is, the "dynamic" in dynamic scoring refers to allowing the estimate of the effect of the proposal being scored to include a causal effect from the proposal to the overall level of economic activity (i.e., GDP), which in turn could have an impact on revenue. Until the rule change enacted this year, scoring practice did allow for significant compositional changes in response to tax changes, but not dynamic changes.

For many proposals, the conventional scoring process is a sound way to achieve the objectives of scoring. For example, a targeted tax on a minor pollutant might reduce that pollutant, and raise some revenue, but have little impact on the overall economy. For major proposals that seek to draw on the best evidence in the academic literature on fundamental tax reform, however, the conventional scoring process can lead to wildly misleading estimates of the revenue impact of changes. The purpose of a fundamental reform, after all, is to improve the functioning of the economy. But dynamic scoring is also challenging. Most importantly, economists need to find a way to link the static estimates drawn from large microeconomic databases with macroeconomic models that, traditionally, have far less microeconomic detail.

How wrong is the current approach?

The potential for conventional scoring to mislead is significant. For example, in 1997, the JCT organized a symposium of the world's leading economists to analyze the impact of a fundamental tax reform that enacted a consumption tax. The average of the estimates at the symposium implied that such a reform would increase GDP far off in the then future, in the year 2010, by about 5 percent. If that effect turned out to be correct, then GDP would have been about $750 billion higher in that year, and federal tax revenue might have been about $150 billion higher. Assuming that effect carried forward to today, we might expect GDP to be almost a trillion dollars higher in 2015 had we adopted such a reform back then. Yet conventional scoring would not allow for these effects, which are very significant economically, since by construction it does not permit estimates to incorporate effects on the overall level of economic activity. Perhaps we did not adopt the reform back then, in part, because a trusted and professional staff did not incorporate such an effect into the analysis of a specific proposal. Had they done so, we might be a trillion dollars better off today.

The symposium relied on very complicated computer models, which might defy the intuition of any but the most sophisticated of economists. But the key idea is not so elusive. A simple example of a dynamic scoring model that can provide intuition for the scale of the expected error for a typical static score comes from former CEA Chair N. Gregory Mankiw, in a 2006 paper that he co-authored with Matthew Weinzierl. Starting with the Ramsey growth model, a standard in macroeconomic textbooks, Mankiw and Weinzierl show that the following holds:

$$\Delta R_{Dynamic} = 1 - \frac{\alpha t_K + (1-\alpha) t_L}{(1-t_K)(1-\alpha)} \Delta R_{Static}$$

$\Delta R_{Dynamic}$ refers to the revenue change under the dynamic scoring procedure and ΔR_{Static} refers to the revenue change under the static procedure. Variable α refers to capital's share of income, t_L refers to the tax rate on labor income, and t_K refers to the tax rate on capital income. To borrow an example from Mankiw and Weinzierl (2006): suppose you wanted to compare the dynamic and conventional revenue estimates for a proposal to lower the tax rate on capital income t_K to be 25 percent, the same rate as the tax rate t_L on labor income in this hypothetical world, where the capital share of income, α, is 1/3. If you input these values into the algebraic expression above, the result is that it reduces to:

$$\Delta R_{Dynamic} = \frac{1}{2} \Delta R_{Static}$$

That is, as Mankiw and Weinzierl (2006) note, according to this calculation, in the long-run steady-state the revenue impact of the capital income tax cut is only half of the impact estimated by "static" conventional scoring. The growth effect of this capital income tax pays for 50 percent of the revenue cost of the cut. Or, to put it bluntly, the back of the envelope estimate suggests that the static score can be expected to be off by a factor of 2. The authors also demonstrate that this offset effect is much bigger for capital income taxes, which discourage growth. The intuition for this result is quite straightforward. If we want more output in the future, we will need to have more inputs. If we cut capital taxes, people invest more today, giving us more inputs tomorrow.

To be sure, this "back-of-the-envelope" method of dynamic scoring lacks the nuance and sophistication of the best dynamic scoring models in the literature. Nevertheless, it illustrates the power of even basic dynamic scoring models to shed light on macroeconomic effects of immediate relevance to policymakers.

It makes little sense to ignore any impact a proposal might have on the overall economy when analyzing its impact. If the proposal were expected to have zero macroeconomic effect, as assumed under traditional scoring rules, then in at least some cases there would be little reason to support the proposal.

What should dynamic scoring be applied to?

My testimony will focus on issues related to dynamic scoring of tax legislation, but it is important to stipulate that taxes are not the only policy lever that can affect the overall economy, and arguments in favor of a more rational approach to scoring may also, in the fullness of time, extend to other topics.

In a world of conventional scoring, no tax cut can be estimated as likely to "raise all ships" by raising the level of overall macroeconomic growth. This is because conventional scoring, by construction, only permits changes to the composition rather than the level of economic activity. Thus, current practice focuses one-hundred percent on questions of distribution, treating tax cuts as mere alterations in who gets the benefits of a fixed level of aggregate economic activity. Such a focus has no economic merit. Policymakers, of course, should consider issues of distribution when considering policy alternatives. But to look at distribution only, without regard to economic efficiency, is to deny the basic tradeoff between the two, and frankly, to deny the value of economic analysis whatsoever.

But in a world where all macroeconomic forecasts are uncertain, how can point estimates from dynamic scoring be considered reliable?

Many branches of government must make forecasts in order to fulfill their statutory mandates, even though those forecasts are by their nature uncertain. The Federal Reserve, for instance, must formulate monetary policy in the face of macroeconomic conditions that remain uncertain in perpetuity (albeit to varying degrees). Its members regularly document their own forecasts, and Federal Reserve policy is set with an eye toward the impact that interest rate changes will have on the economy. The reliance of the Federal Reserve on economic models is not controversial.

The absence of controversy regarding that reliance reveals a logical problem facing those who would dispute the usefulness of dynamic scoring for fiscal policy. For example, many tax reforms influence the economy by changing the cost of capital, a variable that depends on expected tax rates, depreciation rates, inflation and the interest rate. The Fed tracks the economic impact of interest rate changes in part through a model of the cost of capital, which influences business investment. An identical change in the cost of capital can be generated either through a change in the interest rate, or through a change in tax rates. The argument that it is acceptable to model the effects of an interest rate change in one corner of the government with such a model, but not model the effects of a tax rate change that has the same impact on the cost of capital in another corner of government is frankly noneconomic. Though the context of dynamic scoring and the context of monetary policy certainly are very different, in both cases the proper response is for the forecast to incorporate a nonpartisan staff's best estimate into the analysis.

While there is model uncertainty, for many policies that would require a dynamic score, the wide range

of plausible *ex ante* effects of the policy will not include zero, the static assumption. The assertion that uncertainty implies economists should adopt an answer known with certainty to be incorrect is not logical. Moreover, the uncertainty economists face when evaluating fiscal policy is not greater than the uncertainty they face evaluating monetary policies. If we can use models for one application, we can use models for both.

The final possible argument against dynamic scoring is that the congressional economic staff is not up to the task. I wholeheartedly disagree with this. The staffs of the CBO and the JCT are easily as impressive and accomplished as the staff of the Fed.

As does the Fed in its analysis of economic conditions, so should the staff of the JCT and others tasked with the dynamic scoring of proposals incorporate sensitivity analysis, a range of perspectives, and the best thinking of the academic community. If there are many available models for a specific question, the staff should evaluate the broad range of them, and then come to a considered judgment regarding the relative weights of the different results. Such a process already occurs when distributional changes are being modeled, and elasticity assumption are made. Over the years, the staffs of the CBO and the JCT have reliably met the high professional standards one would require of a staff engaged in this process. These staffs will be even more effective if we give them freedom to apply their macroeconomic expertise when circumstances warrant it. Asking them to pick a number that includes dynamic effects is no more of a stretch than asking them to pick a number in the first place.

At the same time, though, streamlining and systematizing the dynamic scoring process seems necessary if a significant number of proposals are to be dynamically scored in a timely manner.

In the context of dynamic scoring, one way of reconciling the need to account for the uncertainty inherent to the forecast with the need to have a process that is to some extent streamlined would be for the point estimate to be presented with a 95 percent confidence interval, much as the results of academic studies typically are. The presentation of these confidence intervals would allow policymakers to temper their interpretation of the point estimate in accordance with the level of uncertainty around it. For instance, suppose that two different tax reform proposals are each estimated to be budget-neutral and have a net zero effect on the federal budget. But suppose one of the reforms has a 95 percent confidence interval of +/- $500 billion and another has a 95 percent confidence interval of +/- $10 billion. Assuming policymakers possess a basic level of risk-aversion, it would be rational for lawmakers to prefer the +/- $10 billion as the more attractive of the two, even though the point estimate of 0 is the same for both of the budget-neutral proposals.

Yet even the construction of such confidence intervals, a critic might argue, leaves ample room for questionable judgment and even outright partisan gaming. It would be easy to imagine, for instance, that individuals opposed to a policy would want the confidence interval to express a wider range of possibilities. More broadly, how can one even have confidence in the confidence intervals?

Here another solution emerges from the context of central banking and the formulation of Federal Reserve Policy, in the form of proposals for monetary policy to be formulated on the basis of publicized rules. The most publicized of these comes from Stanford's John Taylor, who has proposed that the Federal Reserve follow the eponymous “Taylor rule” in formulating Fed policy—or explain, in writing, its decision to depart from the Taylor rule when it does choose to make such a departure. (The Fed would still be able to do whatever it wants, regardless of the Taylor rule, so long as it were willing to provide an explanation of why its chosen policy differs from that implied by the Taylor rule). One could imagine the JCT staff having a similar rules-based process for constructing its point estimates

and confidence intervals.

For example, to the extent that the JCT staff would construct its point estimate of a proposal's budgetary impact from a meta-analysis of the academic literature, the JCT staff could have a specified set of "best practice" procedures that it follows in performing meta-analyses. As with John Taylor's proposal for the Fed, any departures from that standard "best practice" set of procedures would be permissible—so long as it were accompanied by a written explanation of why the standard procedure did not seem appropriate to the staff in that particular instance. This would create transparency that could reassure policymakers and the public that the point estimates and confidence intervals rest on sturdy intellectual foundations. As a bonus, it would save the JCT staff the labor of producing lengthy explanations of each and every set of point estimates or confidence intervals, as the absence of any written explanation would serve as a tacit affirmation that the "best practice" procedures were followed. Only when the exception rather the rule were followed would the JCT staff need to provide detailed methodological explanations of its point estimates and confidence intervals. It is worth underscoring that this leaves discretion in the hands of the JCT staff rather than in any one model, in recognition of the necessity of human judgment in formulating views based on economic models rather than giving any one specific model the final word.

To be fair, the process of constructing a methodology for the construction of such a point estimate and set of confidence intervals will be complex, even daunting. But the JCT staff would be able to consult outside experts with the depth of knowledge and expertise necessary for the task. The CBO Panel of Economic Advisers serves as a model for the type of body of outside experts that would be well-suited to such a task. It has many of the nation's most prominent public finance specialists and, though the construction of point estimates based on the public finance literature may require a slightly different area of substantive expertise, it is an example of the type of resource the JCT staff could consult in constructing its point estimate processes.

What does dynamic scoring cost?

It is a testament to the quality of the JCT staff they are able to accomplish as much as they are today. Scoring policy proposals, whether through a dynamic or conventional process, is a complex task that demands substantial resources. The JCT staff faces constraints, in terms of both financial and human resources, that put a ceiling on how much work it can accomplish. It would be unreasonable to suppose that the JCT staff could handle any reform to the scoring process that would require substantially more resources on their part without giving them those additional resources.

Shifting from conventional to dynamic scoring is precisely an example of something that would require significantly more resources from the JCT staff. This is due largely to the mechanics of the way that the scoring process has worked in the past. The JCT staff, as one would imagine, has certain analytical assets and procedures that it uses now in its conventional scoring process. To date, these have served the JCT well. Yet shifting to dynamic scoring would require the JCT staff to integrate macroeconomic and other models that are not currently part of the JCT staff's existing stock of assets and procedures. Such integration would therefore require substantial amounts of additional financial resources.

A path forward toward dynamic scoring?

Advances in computing power have significantly increased the ability of economists to analyze complex models, and the Internet has enabled a high level of collaboration between scholars. Economic models are no longer black boxes sitting on a hidden disk on a mainframe. A move toward dynamic scoring should seek to be as open as possible, so that the large and thriving modeling community can provide scrutiny and feedback to professional staff. Even the most able and dedicated teams stand to benefit from the intellectual output and feedback of other able and dedicated teams and individuals, who may be able to fill technical and intellectual gaps that otherwise remain hard to fill. We believe that the scoring process, in particular, is one area where this type of collaborative interaction can add value. This does not mean that "the crowd" should be involved in every score, but rather, that "the crowd" be empowered to evaluate the methods used for the score.

This may sound like an abstract proposal for a fully transparent Wikipedia of the scoring process, but it is far from it. As one example of the potential for informed outsiders to play a role in providing resources to the JCT and its staff, I would like to mention an initiative at AEI we have named the Open Source Policy Center (OSPC). OSPC is a project a long-time in the making that already involves outside experts from all over the country and early beta testers and users. OSPC Managing Director Matt Jensen described the OSPC in a recent post:

The motivating principle behind the Open Source Policy Center is that policymakers and the public should have the best tools for understanding public policy choices, and that those tools should be completely transparent and collaborative in order to promote innovation and quality.

With that in mind, OSPC brings together an open-source community of economists, software engineers, and policy analysts who collaboratively produce open-source computational economic models and web applications that allow non-programmers to easily interact with those models.

The community's first priority is building simulation models of the federal individual income tax system. Later projects will move beyond taxes to model other economic policies, including spending programs such as Social Security, welfare programs, and health care programs. Our goals are to be able to both replicate the analysis performed by government agencies and expand and improve upon that analysis with more elaborate tools.

OSPC projects that have reached the alpha or beta stage include Taxcalc, the first-ever open-source microsimulation model of the US individual income tax code, and LOGUS, the first-ever open-source large-scale dynamic overlapping generations model of the US economy. To enable policymakers, journalists, students, and citizens to interact with the models and gauge first-hand the effects of policies, the OSPC community has also developed an easy-to-use web application called TaxBrain.

In addition, OSPC hopes to pave the way for others to adopt a more collaborative, transparent, and accessibility-driven approach to the development of policy-relevant economic models. Our intent is not just to build models, but to develop a technological approach and workflow that enables geographically-dispersed experts to develop models in an open environment.

To illustrate how initiatives like OSPC can add value and facilitate dynamic scoring, it would be helpful to understand how OSPC "bridges" between public-use individual income tax data and the type of macroeconomic models typically produced by academic economists. Many organizations—the Urban-Brookings Tax Policy Center and the National Bureau of Economic Research TAXSIM

program, for example—deploy the public-use income tax file that the IRS Statistics of Income program generously makes available. With this file, analysts can assess how changes to specific items of the tax code would affect the individuals in the sample. The inputs into the "microsimulation" models based on the public-use file tend to be proposals to specific statutes that a policymaker or Congressmen would know in detail (e.g., raising the maximum value of the Earned Income Tax Credit). One can extrapolate the output of these microsimulation models, generated based on the policy-change input, from the sample of the public-use file to the aggregate population as a whole. This extrapolated output can then be fed as an input into a dynamic macroeconomic model that models a substantial portion of aggregate economic activity— a model of the sort that academic economists tend to create. Thus, one has effectively "bridged" between the micro-level inputs – tax-law parameters -- and the macroeconomic output of a dynamic model. One can then feed the macroeconomic output from one time period into the individual level data of the public-use file in the simulation of the next time period, allowing dynamic effects to play out over time. And so "bridging" between the individual public-use file and dynamic models allows for precise dynamic simulations. As the programs are completely transparent, any individual can explore the impact of changing parameter assumptions, or even flipping from dynamic model to dynamic model. To the extent that analysts disagree about the likely impact of a policy, the software will help the analysts identify the source of their disagreement.

The value of our bridging approach is that it ties together the inputs of the microsimulation model, which tend to be specific statutory proposals that would be unfamiliar to an academic economist, with the output of the dynamic macroeconomic models of academic economists. This serves as a link between the ways policymakers think (e.g., in terms of a modification to a specific statute) and the ways that the world's best economists express themselves (e.g., in terms of dynamic models with general equilibria). You might think of bridging as analogous to translating economic knowledge expressed in two different languages into a single language understandable by speakers of both—to the benefit, therefore, of speakers of both languages, who can each access the insights of the other in the shared language.

We hope the OSPC evinces a level of transparency and technical rigor that serves as an example for how this type of collaboration can add value to both those inside the policy community and those outside of it. By making the bridge between the two types of models simple, our hope is that the best and most cutting edge modelers will see the value in making their models available to the broader policy community over time. There is no question that dynamic scoring will, if this process is successful, improve over time. But even today, folding the two types of models together in a systematic manner is quite possible.

Congress, and the United States more generally, would benefit from the dynamic scoring of more policy proposals. Much work remains to be done in fleshing out how exactly such a system of dynamic scoring should work in practice. Nevertheless, the obstacles to transitioning to a world where dynamic scoring becomes the norm are not insurmountable.

References

Mankiw, N. Gregory, and Matthew Weinzierl. 2006. "Dynamic scoring: a back-of-the-envelope guide." *Journal of Public Economics* 90 (8): 1415-1433.