

DOUGLAS ELMENDORF

Harvard University

GLENN HUBBARD

Columbia University

HEIDI WILLIAMS

Dartmouth College

Dynamic Scoring: A Progress Report on Why, When, and How

ABSTRACT By design, official budget estimates for legislative proposals generally exclude the proposals' likely effects on labor, capital, productivity, and output, as well as any feedback from such effects to the federal budget. Policymakers would benefit from knowing the expected sizes of those effects, and advances in research and in the estimating agencies' tools and experience have made such analysis more feasible. If Congress requested that those effects be included more often in official budget estimates—so-called dynamic scoring of legislation—the advantages and disadvantages would vary across policy areas. For some areas, the estimated budgetary impact of the currently excluded effects would be significantly different from the impact of the included effects. But dynamic scoring would be substantially more time-consuming than conventional scoring, and in some areas, the research base is insufficient for credible estimation.

Economists who study the growth of productivity and output often emphasize the role of public policies that affect labor supply, savings, and federal investments in infrastructure and research. Yet, the official

Conflict of Interest Disclosure: Heidi Williams wrote this paper while she was a visiting scholar at the Congressional Budget Office (CBO), where she is currently a consultant. The authors did not receive financial support from any firm or person with a financial or political interest in this paper. Deborah Lucas, one of the discussants, is a former employee and current consultant for CBO.

Brookings Papers on Economic Activity, Fall 2024: 93–134 © 2025 The Brookings Institution.

budget estimates provided to Congress when changes in policies are being considered generally exclude—by design—their effects on labor, capital, productivity, and output, as well as any feedback from such effects to the federal budget.

The exclusion of those economic impacts is a long-standing convention of budget estimates produced by the Congressional Budget Office (CBO) and the staff of the Joint Committee on Taxation (JCT). Over time, CBO and JCT have demonstrated a growing capacity to analyze such economic impacts and their feedback to the budget, and the agencies have published more analyses of this sort. But those analyses have been limited in number and generally provided separately from the official estimates on which congressional and public attention is focused during policy debates. The official budget estimates for proposed legislation continue to be produced almost entirely through conventional scoring that excludes those economic impacts rather than so-called dynamic scoring that includes them.

This paper analyzes the economic and institutional issues presented by the choice between conventional and dynamic scoring, as well as by some intermediate approaches. Thoughtful observers of the congressional budget process have a range of views about the best approach, and the agencies appropriately defer to Congress on this choice.

The current estimating convention has been supported by various lines of reasoning. One is that producing dynamic estimates takes too much time and work to be feasible given the large number of proposals for which estimates are needed and the short time frames typically available for preparing estimates. We take the practical constraints faced by the estimating agencies very seriously. However, CBO and JCT have invested—to their credit—in the tools and experience needed for dynamic analysis, and those investments have reduced to some extent the additional time and work needed to generate estimates. Further targeted investments could make more progress along these lines, but producing dynamic estimates would nonetheless be more time-consuming than producing conventional estimates in most cases.

Another line of reasoning is that the effects of policy changes on outcomes such as labor and productivity may be inherently more uncertain than other effects—and, perhaps relatedly, estimates of those effects might be more subject to external pressure. But in our assessment, uncertainty about the excluded effects is not necessarily greater than uncertainty about effects that are included, in part because, for some important effects, CBO and JCT can draw on substantial bodies of evidence and receive professional feedback on their use of the evidence. Moreover, including these additional effects would reduce the risk of bias in estimates.

A further line of reasoning in support of the current estimating convention is that the excluded effects are small for most proposals. The impact of most proposals on the aggregate economy is indeed small. However, that information itself can be useful for policymakers, and even a small effect on the aggregate economy can still be meaningful for a proposal's budget estimate. We document an especially stark example of the latter point: For a potential change in the authorized number of high-skilled immigrants, the budgetary impact of the change in population that is excluded from conventional estimates is much larger than the budgetary impact of the included factors and has the opposite sign.

Yet another line of reasoning is that CBO's and JCT's credibility might be damaged by changing their estimating methodology, and especially by reduced transparency due to the complexity of including additional effects. Maintaining the agencies' credibility is indeed crucial for the policymaking process. But in our assessment, the challenges of change and complexity related to dynamic scoring are similar to those with conventional scoring, for which the agencies have made many methodological advances involving greater complexity. In addition, excluding from budget estimates behavioral responses that people expect to occur also can hurt the agencies' credibility.

We want to underscore that official budget estimates from CBO and JCT are not the be-all and end-all of policymaking, even though estimates are sometimes blamed for blocking or inducing certain outcomes. The agencies provide considerable other information through additional material included with official estimates, published reports on policy issues, and informal exchanges with Congress. Other government agencies, analysts outside government, and other private actors also communicate with policymakers and the public. Still, official budget estimates are often consequential in the legislative process, so the conventions underlying those estimates warrant ongoing attention.

We note also that the exclusion of certain economic impacts is not the only aspect of the estimating process that deserves scrutiny and consideration of alternatives. Other effects of proposed policies are sometimes excluded from estimates because of so-called scorekeeping guidelines that have been developed by CBO, the Office of Management and Budget (OMB), and the House and Senate Budget Committees to help align the estimating process with legislative structures and procedures (CBO 2021a). For example, Hall (2024) explained why estimates of proposed activities to reduce fraud often exclude the impact on program spending or tax collections, and he proposed changes to the scorekeeping guidelines to address the issue.

In the following sections of the paper, we elaborate on the advantages and disadvantages of dynamic scoring, describe potential processes and criteria for Congress to make decisions on the subject, and explain how dynamic scoring might be undertaken. Because discussions of dynamic scoring often focus on tax policies, and because we have greater familiarity with the procedures used by CBO, we focus on spending policies and CBO's methodology, but we include substantial references to JCT's work as well. Many of these issues apply also to OMB, the Treasury Department's Office of Tax Analysis, and unofficial budget estimators outside government. We then present three case studies that illustrate some of the issues raised in the first parts of the paper. These case studies cover immigration policy, federal investment in research and development (R&D), and federal permitting of investment in infrastructure.

I. Why?

We begin with the crucial question of “why” or “why not” to do dynamic scoring; the subsequent sections turn to the following questions of “if so, when” and “if so, how.”

I.A. Budget Estimates

CBO and JCT provide to Congress the official estimates of the effects of legislative proposals on the federal budget. CBO produces public estimates for bills after they have been approved by congressional committees; the process of producing those cost estimates is often described as “scoring.” For bills that would alter the tax code, CBO's cost estimates use changes in estimated revenues provided by JCT, which publishes its revenue estimates directly as well. In this paper we use the term “budget estimates” to emphasize that changes in both spending and revenues are relevant. In addition, the agencies give private estimates to members of Congress and their staffs for proposals that are being developed and have not been released publicly. In a typical year, CBO publishes roughly seven hundred public estimates, and it and JCT provide thousands of private estimates. For more information on this process, see CBO (2023c) and JCT (2023). Budget estimates are used by Congress to evaluate alternative proposals, to iterate on policy development, and to enforce budget plans. CBO has been providing cost estimates since its founding roughly half a century ago, and JCT has been in place for nearly a century.

In addition, CBO provides projections of budgetary and economic outcomes under current law that are known as “baseline projections”; budget

estimates show effects relative to those baseline projections. And CBO and JCT publish analyses of actual or potential legislative proposals beyond budget estimates; we say more about those analyses shortly.

Both baseline projections and budget estimates typically focus on nominal cash flows for the current fiscal year and each of the ten subsequent years, a period that is often called the “budget window.” Baseline projections are extended in less detailed form beyond the budget window, and budget estimates are sometimes extended as well when Congress is especially interested in long-term effects and when the agencies expect that a proposal’s long-term effects would be notably different from its effects over the coming decade.

The budget estimates and baseline projections depend on assessments made by analysts at CBO and JCT and on procedures that have been codified in law or agreed to formally or informally by the relevant congressional committees. The agencies provide objective, nonpartisan analysis at a high professional standard. As a result, their estimates are important not only to Congress but also to press coverage of policy deliberations and to the broader public.

1.B. Behavioral Responses

CBO’s and JCT’s official budget estimates generally include the impact of behavioral responses to the proposed changes in law. For example, estimates for changes in benefit programs include shifts in take-up rates, and estimates for changes in income tax rates include shifts in the use of tax deductions. The agencies try to account for the behavior of households, businesses, implementers of legislation in the executive branch, and state and local governments; they do not attempt to predict future changes in federal law.

However, CBO and JCT do not include in official budget estimates any behavioral responses that would change the overall economy, which include effects on labor supply, saving and investment, productivity, and aggregate demand. As CBO (2023c, 9) explained: “According to long-standing practice, CBO’s conventional cost estimates reflect the expectation that nominal gross domestic product [GDP] . . . would not change.”¹

At the same time, CBO and JCT produce some analyses beyond budget estimates that do include effects on labor, capital, productivity, and demand. For example, for some bills, JCT publishes what it terms “macroeconomic

1. JCT’s cost estimates reflect the expectation that nominal gross national product (GNP), rather than nominal GDP, would not change. This difference and other complications that arise in applying the general principle are not crucial for our analysis.

analyses,” which generally appear several weeks after the official estimates for the bills and which include the budgetary feedback from expected changes in broad economic measures. As another example, in 2013, CBO released a cost estimate prepared by the agencies for a comprehensive immigration bill being debated in the Senate (CBO 2013a) and simultaneously released a report with “estimates of the overall economic impact of the legislation and of the incremental federal budgetary effects of changes in the economy that the cost estimate does not reflect” (CBO 2013b, 1). And as a further example, CBO publishes reports that analyze the economic and budgetary effects of some potential legislative proposals.² We return to each of these examples later.

In this paper we use the term “dynamic analysis” for reports that address broad economic changes outside of official budget estimates and the term “dynamic scoring” for official budget estimates that include those economic changes. The distinction is important because official budget estimates are often the focus of congressional and public attention during policy debates.

For economists, the deliberate exclusion of some behavioral responses from official budget estimates can seem quite odd. A change in a benefit program might induce an individual to stop or start taking up those benefits *and* to change their work effort; from an economic perspective, there is no fundamental difference between those responses, and they may well be directly related. Similarly, a change in the tax code might induce an individual to adjust their tax deductions *and* their saving behavior, and there is no fundamental economic difference between those responses.

Both the behavioral responses that are currently included in official estimates *and* the behavioral responses that are currently excluded can affect the budgetary impact of policy changes. Therefore, including in official estimates all of the foreseeable behavioral responses would seem to improve the accuracy of the estimates. For any given policy change, some of the behavioral responses may be seen by some policymakers as central or collateral advantages of the policy change and may be seen by other policymakers as disadvantages of the change. But regardless of the desirability or undesirability of certain behavioral responses, including them in budget estimates could help to make the estimates better—with caveats that we turn to shortly.

2. In addition, when legislative proposals are enacted, CBO’s subsequent baseline projections incorporate those proposals’ economic effects and the budgetary feedback from those effects.

Moreover, including in official estimates all of the foreseeable behavioral responses gives policymakers information about proposed policies that may be independently relevant to their decisions. For example, policymakers may want to know how a potential change in a benefit program would affect take-up for its own sake—for example, how a change in subsidies for health insurance would affect insurance coverage—in addition to the impact of take-up on the budget. And policymakers may want to know how a policy change would affect labor, capital, and productivity, as well as the follow-on effects on output, inflation, interest rates, and other economic outcomes.

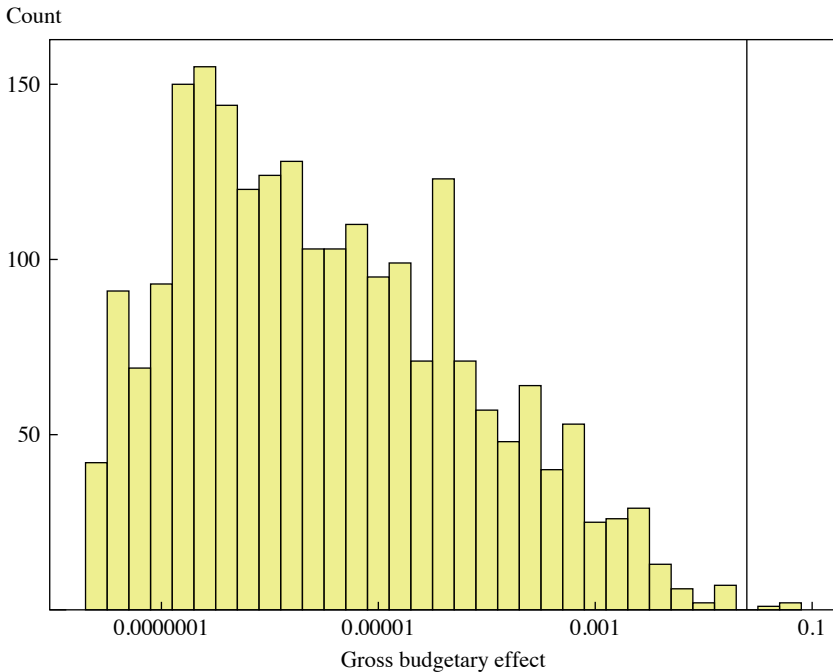
Why, then, are some people who have produced and used budget estimates skeptical about dynamic scoring? Four concerns have been most central to this long-running debate—the feasibility of dynamic scoring, the uncertainty of dynamic estimates, the magnitude of excluded effects, and the credibility of estimates. We take up these concerns in turn.

1.C. Feasibility of Producing Dynamic Budget Estimates Given Time and Resource Constraints

Including additional behavioral responses in budget estimates unavoidably requires more calculations and review, and doing that work would be especially challenging because of the large number of proposals to be analyzed, the many iterations through which proposals are developed, and the short time frames for producing estimates that Congress allows. These practical constraints are real, and they should be central to discussions about dynamic scoring.

One approach to this feasibility challenge would be to do dynamic scoring only for major legislation. A decade ago, one of us (Elmendorf) wrote: “All of the estimates of macroeconomic effects described above involved significant conceptual and practical challenges and required a great deal of analysts’ time to complete. The agencies can devote that much time to only a very small share of the thousands of proposals they examine each year. . . . For example, when CBO [2015a] examined three ways of reducing spending for the Supplemental Nutrition Assistance Program (commonly known as ‘food stamps’), the agency found that even the sign of some policies’ net effect on labor supply was unclear without detailed analysis” (Elmendorf 2015, 110–11). He argued therefore that dynamic scoring should be applied only to proposals for which conventionally estimated changes in revenues, spending, or deficits exceed 0.25 percent of projected output over the ten-year budget window (about \$900 billion currently).

Indeed, the House of Representatives currently requires CBO and JCT to provide dynamic estimates, “to the extent practicable,” for all bills

Figure 1. Gross Budgetary Effects and the Major Legislation Rule

Source: CBO.

Note: This figure plots the distribution of estimated gross budgetary effects for the universe of CBO's cost estimates as drawn from the XML files of estimates from the 116th, 117th, and 118th Congresses through January 2024; see <https://www.cbo.gov/cost-estimates/xml>. The *x*-axis shows gross budgetary effects relative to GDP and uses a log scale for readability. CBO generally computes gross budgetary effects by summing the absolute value of the budgetary effects of provisions and their interactions (Edelberg 2016); a "provision" corresponds to a line in a table summarizing a CBO estimate. The gross budgetary effects shown here ignore some rare complications that are difficult to calculate from publicly available data. The vertical line denotes the threshold for the House rule in the 118th Congress.

that exceed the 0.25 percent threshold in any year of the budget window (which ranges from about \$70 billion to about \$100 billion currently) or are deemed "major" by one of a small number of congressional leaders (Lynch and Gravelle 2023). Some version of the House rule has been in effect for much of the past decade. Yet, the threshold is high enough that dynamic scoring has occurred very rarely. Among the roughly 2,700 of CBO's published budget estimates between 2019 and January 2024 for which we could readily code budgetary effects, very few had effects large enough to warrant a dynamic estimate, shown on the far right of figure 1. JCT has published roughly a dozen dynamic analyses over the past decade, which is still a very limited number.

An alternative, possibly complementary, approach to the feasibility challenge would be to standardize or simplify some estimation processes in order to reduce the time and effort needed to produce dynamic estimates. It is said at CBO that estimates are produced by analysts rather than models, because analysts build models, translate legislative language into model inputs, and review model outputs to ensure that the modeling is suitable for the proposals at hand. Thus, standardizing and simplifying estimation processes while maintaining the quality of estimates would not be easy. However, CBO and JCT have invested in the tools of dynamic analysis and gained experience in doing such analysis, which helps in this regard.

Examples of the agencies' dynamic analysis include reports from JCT on the economic effects of four major pieces of tax legislation, which we return to shortly (JCT 2015b, 2015c, 2017, 2018b), and reports from CBO on repealing the Affordable Care Act, increasing federal investment, responding to the pandemic, funding more federal infrastructure, and raising the minimum wage (CBO 2015b, 2016, 2020, 2021c, 2023e). In addition, CBO has introduced more dynamic complexities into its baseline projections, including a painstaking collation of data on recent immigration and careful modeling of the likely economic and budgetary effects over the coming decade (CBO 2024d).

The agencies' new and refined tools and greater experience have made future dynamic estimates somewhat more tractable because the models and learning can be used repeatedly, just as the models and learning underlying conventional estimates are used repeatedly. And as the agencies have applied their technical frameworks to more policy issues, they have refined some of their analytic approaches to be more standardized, analogous to the way that conventional estimating can become standardized with practice. For example, CBO's (2016, 2021c) reports cover similar ground related to the economic effects of federal infrastructure investment, but the later report presents an explicit practical framework that can be deployed in future analyses. Further development of that framework could give analysts standardized algorithms for capturing effects that are now excluded from budget estimates. Similarly, CBO has encapsulated its expertise in estimating the budgetary impact of economic changes into a budgetary feedback model (Frentz and others 2020), which has been further boiled down to an online workbook (CBO 2023b) and back-of-the-envelope formulas (CBO 2024a).

But even when standardized and simplified approaches have been created, analysts still need to assess whether they are applicable to the proposal at

hand or require adjustment. Thus, producing dynamic estimates will continue to be more time-consuming than producing conventional estimates in most cases, and shifting to dynamic scoring broadly would require Congress to accept longer cycle times in those cases.

1.D. Uncertainty of Dynamic Budget Estimates

Budget estimates are unavoidably uncertain. CBO and JCT derive expected behavioral responses to policy proposals using a broad array of evidence, including external economic research, analyses done in-house, and institutional knowledge and anecdotal information from government agencies and others. CBO also has standing panels of outside advisers and has outside experts review some work products. However, many proposals would change policies in ways that have not happened before or have happened only under different circumstances, have not been examined closely in the research literature, or rely on data that have not previously been collected.³ Therefore, analysts at CBO and JCT often need to produce estimates with limited evidence. Meanwhile, the procedural rules of Congress rely on point estimates rather than ranges, so the agencies focus on point estimates.

If behavioral responses to policy changes are sufficiently uncertain, including those shifts in estimates could reduce the estimates' accuracy. To illustrate this concern, consider a simple example in which an outcome of potential interest has equal probabilities of taking the values -1 , 0 , 1 , or 2 . If the outcome is excluded from the estimating process by rule, the estimator is effectively predicting zero all the time, and the root mean square error (RMSE) of the estimates is 1.2 , as shown in the first block of table 1.⁴ If the estimator correctly predicts the average outcome but has no ability to distinguish between cases where the outcome will be low or high, the estimator predicts 0.5 all the time, and the RMSE dips to 1.1 (as in the second block). If the estimator incorrectly predicts the average outcome to be zero but correctly orders the outcomes and limits the variability of the estimates, the estimator might predict (in order of the outcomes above) -1 , -0.5 , 0.5 , and 1 , and the RMSE falls to 0.6 (as in the third block); if the estimator also correctly predicts the sign of the average outcome, the RMSE falls further. But if the estimator incorrectly predicts the average outcome, correctly

3. As just one example, consider the uncertainty in estimating the budgetary impact of a proposal to make eligibility for subsidized school lunches depend on personal information that is not currently being collected.

4. In evaluating its estimates, CBO is attentive to both their "centeredness" and "spread" (Hall 2015), and sometimes the agency measures accuracy through the RMSE, which combines bias and variance (CBO 2023d).

Table 1. Errors from Different Illustrative Estimation Processes

<i>Estimation process</i>	<i>Outcome</i>				<i>RMSE</i>
	<i>-1</i>	<i>0</i>	<i>1</i>	<i>2</i>	
Excluding the variable by rule					
Estimate	0	0	0	0	
Error of estimate	1	0	-1	-2	1.2
Predicting average outcome but not variation					
Estimate	0.5	0.5	0.5	0.5	
Error of estimate	1.5	0.5	-0.5	-1.5	1.1
Not predicting average outcome but ordering correctly and limiting variance of estimates					
Estimate	-1	-0.5	0.5	1	
Error of estimate	0	-0.5	-0.5	-1	0.6
Predicting average outcome and order but having a high variance of estimates					
Estimate	-3	-1	2	4	
Error of estimate	-2	-1	1	2	1.6

Source: Authors' calculations.

Note: The root mean square error (RMSE) for the estimation process of excluding the variable is calculated as the square root of $0.25 * \{[0 - (-1)]^2 + [0 - 0]^2 + [1 - 0]^2 + [2 - 0]^2\}$; the other RMSEs are calculated similarly.

orders the different outcomes, and generates estimates that are more variable than the outcomes, the estimator might predict (in the order above) -3, -1, 2, and 4, and the RMSE rises to 1.6.

These different possibilities show, in the words of CBO, that including additional responses will probably improve the accuracy of estimates “if the estimates are for policies that are likely to have significant feedback effects; if the estimates of those feedback effects are centered around the actual effects; and if the spread of those estimates is small” (Hall 2015, par. 12). That is, including additional effects will probably improve accuracy if their mean is different from zero, if the estimator knows the sign of the mean, and if the variance of the estimates is held down.

How do these considerations apply to dynamic scoring? For many potential policy changes, budgetary impacts of factors excluded from conventional estimates are significant relative to the budgetary impacts of included factors and have signs that are widely understood. Examples include changes in income tax rates, shifts in immigration rules that change the population, fiscal stimulus during recessions, and changes in federal investment. Indeed, economists have paid more attention to some responses that are excluded from conventional estimates than many responses that are included, so CBO and JCT can draw on larger bodies of evidence and receive more professional feedback.

At the same time, relevant parameter values cannot be extracted in a simple manner from the research literature on many topics. Estimates of parameters can vary widely, deciding which estimates to weight more heavily can be difficult, and some observers use motivated reasoning in interpreting evidence and could try to pressure the agencies to adopt their interpretation. In addition, the time dimension of dynamic analysis raises further issues. Many estimates in the research literature are for behavioral responses at a point in time and may fit well into cross-sectional models such as micro-simulation models, whereas dynamic analysis requires time paths for responses in dynamic models, and linking those approaches is complicated. Moreover, uncertainty in modeling economic effects over time goes beyond specific parameters to the choice of model itself; reflecting that uncertainty, for example, JCT deploys three models in its macroeconomic analyses and varies the weights it attaches to the models depending on the nature of the legislation being analyzed (JCT 2018a).

However, it is important to recognize that these challenges arise also with the behavioral responses included in conventional estimates, such as the effects of increasing the minimum wage on employment and earnings or the effects of changes in health policy on insurance take-up.⁵ CBO and JCT have responded to those challenges by evaluating evidence, reporting their evaluation, soliciting professional feedback, and defending their modeling choices to critics. The agencies have done the same for some behavioral responses included in dynamic analysis, and Congress and the public would benefit from the agencies continuing to do so in the future.

For example, a dozen years ago CBO undertook a comprehensive review of the response of labor supply to changes in fiscal policy. In three linked papers (CBO 2012c; McClelland and Mok 2012; Reichling and Whalen 2012b), the agency documented the estimated elasticities in economic studies and explained the elasticities it chose to use in its analyses (including lower, central, and higher estimates of substitution and income elasticities for primary and secondary earners at different wage levels). Similarly, CBO reviewed research on the response of output to fiscal stimulus under different economic conditions and explained how it applies that research (Reichling and Whalen 2012a; Whalen and Reichling 2015; Seliski and others 2020; Demirel and Wilson 2023), and CBO recently drew on a

5. For example, how CBO and JCT might formally take on board the forecasting literature's lesson that accuracy can be improved by "shrinking" estimates is a question that applies to conventional as well as dynamic scoring. See CBO (2019, 30–35) for one application of CBO's approach to analyzing uncertainty.

growing body of evidence on immigration as it modeled the effects of the recent immigration surge (CBO 2024d).

Note also that CBO's evaluation of the evidence did not point to substitution or income elasticities of labor supply that equal zero (although they might be offsetting for some policy changes) nor fiscal policy multipliers that equal zero. Excluding behavioral responses that are not zero greatly increases the chance that the resulting estimates will not be in the middle of the distribution of possible outcomes, which is a natural target for estimators.

Finally, evidence regarding the economic effects of changes in various public policies has expanded considerably over time and continues to do so. For the illustrative examples we discuss later in the paper, we highlight important new research from recent years.

1.E. Magnitude of Policy Effects on Labor, Capital, Productivity, and Other Economic Measures

For almost all policy proposals, their effects on labor supply, saving and investment, productivity, and other economic measures such as output, inflation, and interest rates would be very small relative to the economy's overall amounts or typical levels. However, other considerations related to magnitude are important.

First, even if expected changes in those measures are very small in macroeconomic terms, their budgetary effects could be large relative to other budgetary effects of a proposal and thus could significantly affect a budget estimate. For this reason, we do not refer to the excluded effects as "macroeconomic"; they are excluded because of the assumption that proposals do not affect the macroeconomy, but the rationale for including them is not to improve macroeconomic forecasts but instead to more accurately capture the budgetary impact of proposals.

For example, as we explain later, expanding immigration would increase both benefit payments and tax revenue. Conventional estimates include the former but not the latter, leading to estimated increases in deficits; dynamic estimates include the additional revenue as well and show estimated decreases in deficits. Note that this stark difference between conventional and dynamic estimates is unrelated to the number of people affected by a proposal and thus unrelated to a proposal's macroeconomic impact.

Second, a finding that expected changes in labor, capital, productivity, output, and other economic measures are small—in macroeconomic terms or relative to some other benchmark—can be important information for policymakers. Advocates for policies tend to expect large effects on desirable outcomes and small effects on undesirable ones, while opponents tend

Table 2. Comparing Conventional and Dynamic Estimates for Tax Legislation

<i>Legislative proposal</i>	<i>Budget window</i>	<i>Increase in deficit</i>		
		<i>Conventional estimate (billions)</i>	<i>Dynamic estimate (billions)</i>	<i>Difference (percent)</i>
Tax Relief Extension Act of 2015 (S. 1946)	2016–2025	\$97	\$87	10
A Bill to Amend the Internal Revenue Code of 1986 to Modify and Make Permanent Bonus Depreciation from 2015 (H.R. 2510)	2016–2025	\$281	\$267	5
Act to Provide for Reconciliation Pursuant to Titles II and V of the Concurrent Resolution on the Budget for Fiscal Year 2018 (often referred to as the Tax Cuts and Jobs Act)	2018–2027	\$1,456	\$1,071	26
Protecting Family and Small Business Tax Cuts Act of 2018 (H.R. 6760)	2019–2028	\$631	\$545	14

Source: JCT.

Note: The increases in the deficit are in nominal dollars and are drawn from reports published by JCT (2015b, 2015c, 2017, 2018b). The “Difference” column equals the conventional estimate less the dynamic estimate, divided by the conventional estimate.

to expect the opposite. Objective analysis by CBO and JCT can help policymakers and observers sort out the competing claims.

For example, changes in the tax code have effects on labor supply, saving, and investment that are often asserted to be large by some observers and small by others, and that disagreement in turn generates disagreement about the budgetary impact of tax changes. Objective estimates of those effects, even recognizing the uncertainty, are valuable. In this vein, a comparison of JCT’s conventional and dynamic estimates for four significant tax bills—as in table 2—shows that the macroeconomic effects were expected to offset between 5 percent and 26 percent of the conventionally estimated revenue loss.⁶ Both the larger and smaller offsets provide important information.

6. Table A-1 of a Congressional Research Service report (Lynch and Gravelle 2023) lists five significant bills for which macroeconomic analyses were released by the estimating agencies (the agencies have published macroeconomic analyses for other bills that we do not address here). Of those five bills, we exclude from the table the Restoring Americans’ Healthcare Freedom Reconciliation Act of 2015 because that bill had substantial spending changes that offset some of the tax changes, so comparing the deficit changes with and without the dynamic aspects is not useful for our purpose. In that bill, the dynamic effects make the estimated reduction in tax revenue 27 percent smaller.

I.F. Credibility of Estimates

The credibility of CBO's and JCT's estimates is crucial to the agencies' service to Congress and the country. Without agreed-upon numbers for budget discussions, policymaking would be less coherent and the budget process less effective, as emphasized to us by our discussant Deborah Lucas. The current level of agreement about numbers from CBO and JCT depends fundamentally on the credibility of the agencies' objectivity and methodologies—and maintaining that credibility may be especially challenging in the current era of amplified polarization and heightened suspicion about expertise. Therefore, any potential changes in methodologies should be evaluated in part by whether such changes might diminish or enhance that credibility.

All else equal, CBO's and JCT's credibility is fostered by consistency over time in scoring methodology and by simplicity of scoring methodology. Members of Congress and their staffs should understand the key features of budget estimates, and they should not be expected to become economic experts because they are appropriately focused on the spending programs or tax code items that are their primary interests as well as the many other aspects of their jobs. Including additional behavioral responses in budget estimates would represent a change in estimating approach and would make those estimates more complex than otherwise.

However, "all else" is not generally equal. *Excluding* from budget estimates behavioral responses that members of Congress and their staffs expect to occur hinders the agencies' credibility, and when the agencies can draw on empirical evidence to include additional responses (of the conventional or dynamic sort), that inclusion can strengthen their credibility.

Indeed, the history of CBO's and JCT's analysis is *not* one of stasis in methodology or particular focus on simplicity. Instead, the agencies have made ongoing advances in their methodologies for producing conventional estimates, usually involving greater complexity, in order to improve the accuracy of estimates and expand the information presented. One prominent example is CBO's work in the early 2000s to strengthen its modeling and experience in analyzing health care policies, including requesting and receiving additional funding from Congress to bolster the agency's staff. As a result, the methodology of estimates during the debate about the Affordable Care Act was notably different and more complicated than the methodology of estimates of the Clinton administration's health reform plan fifteen years earlier. Another important example is CBO's expansion in the 2010s of its use of so-called fair value accounting for the government's financial activities, as a complement to the accounting approach specified by the

Federal Credit Reform Act of 1990; see CBO (2012a, 2018). In addition, CBO's construction of its baseline projections has become more complex when developments have warranted it. For example, the agency's projections now incorporate much more primary data on immigration and rigorous analysis of immigration's effects on the economy and budget than previously.

As CBO and JCT have adopted increasingly sophisticated methodologies, the agencies have fostered transparency of their analytic approaches by devoting considerable attention to explaining what they do. Official budget estimates sometimes include substantial verbal descriptions, the agencies' leaders and staffs spend considerable time talking with members of Congress and their staffs, and the agencies publish separate reports documenting key methodologies. CBO's and JCT's openness about their methodologies is important for encouraging professional feedback and for showing that estimating approaches are grounded in evidence and are not "arbitrary or politically motivated" (CBO 1995, 12).

Similar transparency would be essential for the additional behavioral responses and modeling that would be used in dynamic scoring. The agencies have already published numerous descriptions of their analytic approaches to labor, capital, productivity, output, and other economic measures, as well as of the application of those approaches to specific policy changes. For CBO, the key models include a Solow-type growth model, a Keynesian multipliers model, and a small-scale policy model that combines the two preceding models; those models are described by CBO (2014b, 2021b), Shackleton (2018), and Lasky (2022), among other references. For JCT, the key models include a macroeconomic equilibrium growth model, an overlapping generations model, and a dynamic stochastic general equilibrium model; those models are described by JCT (2015a, 2018a, 2020), Auerbach and others (2017), and Moore and Pecoraro (2020, 2023), among other references.

For future explanations of specific dynamic estimates, we can envision three types. Sometimes the additional behavioral responses would make little difference to the estimated budgetary impact and would not be of independent interest to policymakers; in those cases, little explanation of the additional responses would be needed. In other cases, the additional responses would matter consequentially for budgetary impact or would matter to policymakers for other reasons in ways that would follow standard patterns (for example, an increase in deficits would raise interest rates and thus debt service on existing debt); in those cases, the explanation would be provided by reference to existing published material. And in

still other cases, the additional responses would matter and be somewhat idiosyncratic; in those cases, a bespoke and perhaps complex explanation might be required.

II. When?

In this section we address questions related to “when” dynamic scoring—or intermediate alternatives to conventional scoring—might be deployed. All of the considerations we discuss can evolve over time, which highlights the value of regularly reassessing these questions.

II.A. How Do Alternatives for Specifying the Deployment of Dynamic Scoring Compare?

If Congress wants to receive more information from CBO and JCT about the effects of potential policies on labor, capital, productivity, and other economic measures, we see four ways this objective could be reached—through more dynamic analyses apart from official budget estimates, through dynamic scoring for major legislation, through dynamic scoring for legislation in designated policy areas, or through a scoring approach that is intermediate between conventional and dynamic scoring. We take up these possibilities in turn.

First, Congress could ask CBO and JCT to increase the number and range of dynamic analyses they provide beyond the official scoring process. As we explained earlier, the agencies have published a substantial number of dynamic analyses covering different topics in recent years, and in doing so they have created new modeling tools and gained considerable experience in dynamic analysis. Many other dynamic analyses could be done in the future and would provide important information for policymakers and the public even if the analyses are not included in official budget estimates. Moreover, once an agency has completed a dynamic analysis in a policy area, the benefits and costs of including such analysis in budget estimates can be ascertained more readily.

Second, Congress could build on the current House rule requiring dynamic scoring for major legislation by adopting a similar approach in the Senate and by adjusting the definition of “major” so that more bills are covered. Because official budget estimates often receive particular attention during policy debates, this approach would give dynamic analysis more weight than if that analysis was made available just as ancillary information. Using dynamic scoring for proposals whose budget effects exceed a certain threshold has the advantage of focusing CBO’s and JCT’s efforts

on proposals that generally would have larger effects on the economy as well as the budget. Also, using a dollar threshold provides a transparent trigger for dynamic scoring, and different thresholds could be chosen to trigger dynamic scoring more or less often. However, when a bill with budgetary effects large enough to trigger dynamic scoring includes many disparate components—which is often the case for major bills—doing dynamic scoring for all of the components may be infeasible owing to a lack of prior investment by the agencies (because the policy areas that might arise in proposals exceeding thresholds are difficult to predict).

Third, Congress could adopt dynamic scoring for legislation in designated policy areas, which again would give dynamic analysis more weight than if that analysis was made available only as ancillary information. This approach would have the advantage relative to the “major legislation” approach of ensuring comparable estimates for alternative policies that Congress might consider for meeting a desired goal. Such comparability would be reinforced by applying dynamic scoring to both spending policies and tax policies in the chosen areas—for example, for federally funded R&D and for tax provisions affecting the after-tax price of private R&D. Moreover, CBO and JCT could focus their capacity-building investments on the chosen areas, which could enhance the quality and timeliness of subsequent estimates. However, applying dynamic scoring only to certain policy areas would be complicated to implement and explain when legislation combines proposals from different areas. In addition, focusing on certain policy areas would require decisions about whether proposals fall within those areas, which would not be straightforward in all cases; in other contexts in which classification issues arise (such as whether to apply the Federal Credit Reform Act), the agencies and congressional staffers have relied on written guidelines, frequent communication, and experience.

Fourth, Congress could choose a scoring approach that is intermediate between conventional and dynamic scoring. Following this path also would give dynamic analysis more weight than if that analysis was made available only as ancillary information, but the analysis would be more limited in some respects. Two sorts of intermediate approaches warrant particular attention:

- One intermediate approach that we discuss later in the paper is a “population change” approach to scoring that CBO and JCT have used in a few estimates of proposed changes in immigration policy. That approach adds to a conventional estimate the direct budget effects of changing the number of people in the country (through the impact on compensation and thus tax revenue) but not all the economic effects that likely would occur under a policy. The resulting

estimates are less comprehensive than dynamic estimates, but they are more tractable and easier to explain because they do not include macroeconomic complexities. The population change approach applies naturally to changes in immigration policy, and it also could be applied to changes in health policy that enable people to live longer—for example, see the analysis by CBO (2012b) of the effects of raising the tax on cigarettes—and to changes in the legal system that affect the number of people who are institutionalized.

- Another intermediate approach that deserves further development is to include specific additional types of behavioral responses but not all the responses that likely would occur under a policy. For example, we noted earlier that changes in benefit programs would induce changes in both benefit take-up and labor supply and that only the former are captured in conventional scoring. What if budget scoring added expected labor supply responses, but not other responses that are currently excluded? For policies that would have direct effects on labor supply and no significant direct effects on capital or productivity (or direct effects on capital and productivity that have the same sign as the direct effects on labor), adding labor supply responses to budget estimates would tend to improve accuracy relative to conventional scoring. Many potential changes in benefit programs meet this criterion. However, for policies that would have direct effects on labor and direct effects on capital or productivity with the opposite sign, adding only labor supply responses to budget estimates could reduce accuracy relative to conventional scoring. Shifts in the tax burden between labor and capital meet this criterion. Therefore, an intermediate alternative of this sort would need to be implemented carefully.

We close this section by noting that if Congress was interested in more dynamic analysis through any of these mechanisms, CBO and JCT might not be able to fulfill every request because of practicalities such as time constraints.⁷ Thus, the current congressional rule directing dynamic scoring

7. In writing “if Congress was interested in more dynamic analysis,” we are skipping over the notable complication that Congress rarely speaks with one voice. The committees that are most focused on CBO’s and JCT’s methodologies and allocation of effort are the House and Senate Budget Committees, the House Ways and Means Committee, the Senate Finance Committee, and the House and Senate Appropriations Committees. But other committees that rely on budget estimates sometimes have different views; resource levels for the agencies depend on the legislative branch subcommittees of the appropriations committees; and the leaders of the House and Senate sometimes have different views as well. In addition, of course, there can be partisan disagreements within each of these groups.

for major legislation sensibly includes the phrase “to the extent practicable” (Lynch and Gravelle 2023, 2).

II.B. What Specific Issues Arise with Appropriation Bills?

CBO produces budget estimates for appropriation bills—which determine annual funding for federal agencies, and which traditionally have covered most federal spending on investments like R&D and infrastructure—that focus on how quickly available funding would be spent (CBO 2023c). Those estimates do not include behavioral responses of the sort included in other budget estimates. In addition, certain sorts of feedback from appropriations to other parts of the federal budget are excluded from budget estimates by scorekeeping guidelines (as noted earlier) because including such feedback would complicate congressional processes.

Gullo and others (2025) comment on some of these issues in the context of R&D spending, and our discussant G. William Hoagland focuses on some of these issues in his remarks. Here we will only note that similar procedural issues have been addressed over time. For example, CBO sometimes has reported as a memo item but not added into the basic appropriation tabulations the estimated rise in tax collections from increased appropriations for enforcement, and on some occasions scorekeeping guidelines have been changed to make the information provided in estimates more useful to Congress. Analogously, the application of dynamic scoring to appropriation bills might take the form of memo items attached to basic tabulations that continue to exclude such effects, or it might engender changes in the scorekeeping guidelines.

II.C. What Factors Affect the Value of Investments in the Capacity for Dynamic Analysis?

As CBO and JCT invest in modeling and practice for dynamic analysis, the value of such investments in different policy areas depends principally on three factors in our assessment: the degree of interest from policymakers in considering policy changes; the strength of the research base about the impact of policies on labor, capital, and productivity; and the straight-forwardness of modeling policy alternatives.

Working on topics of enduring interest from policymakers is valuable because of the fixed costs of developing the tools and experience to conduct dynamic analysis on a policy topic. Given the limitations of CBO’s and JCT’s resources, paying those costs is most worthwhile for topics where the long-term payoff is greatest. For example, past investments in modeling

health proposals and the effects of fiscal policy have been in areas of sustained congressional interest.

Building capacity in policy areas with substantial research bases regarding effects that are excluded in conventional scoring is valuable because the credibility of estimates is enhanced when the agencies can point to clear evidence underlying their analytic approaches.

Investing in dynamic analysis for topics where modeling policy alternatives is straightforward is valuable because the time required for modeling varies greatly across topics. Dynamic analysis depends on both general macroeconomic modeling (for example, how changes in labor supply affect capital investment and interest rates) and modeling of specific policy provisions (for example, how various policy levers affect labor supply). The general macroeconomic modeling can be applied to many different types of policies, but the modeling of specific policy provisions may be bespoke. Therefore, the feasibility of dynamic analysis for a policy topic depends crucially on the complexity of modeling the leading policy alternatives.⁸

For topics where alternatives are denominated in standardized units such as dollars or numbers of people, and where effects scale roughly linearly over the relevant range, solid estimates made for some proposals can be scaled through simplified spreadsheets and rules of thumb to produce credible estimates for other proposals. Examples include federally funded research and immigration of higher-skilled people, and we return to those topics later.

For topics where the policy space is multidimensional because alternatives often involve different structures of policy, building simplified but still credible models is more challenging. For example, with health care and health insurance, the diversity and complexity of proposals makes conventional scoring quite difficult in many cases and dynamic scoring even more difficult. CBO and JCT provided a dynamic budget estimate for repeal of the Affordable Care Act (CBO 2015b) and have produced dynamic analyses of some other health proposals, including an increase in the tax on cigarettes (CBO 2012b) and an expansion of Medicaid eligibility (CBO 2024e). However, those examples do not demonstrate that the agencies could provide

8. Another challenge in doing dynamic analysis for some policies is capturing effects that occur far in the future. In select cases, CBO has provided numerical estimates or qualitative information about effects beyond the decade captured in the usual budget window and their feedback to the federal budget. For example, see the analyses of comprehensive immigration reform (CBO 2013a, 2013b) and a proposed change to Medicaid (CBO 2024e).

comparable analyses today for many other health proposals; indeed, the estimate for repeal of the Affordable Care Act quantified some influences on labor supply and saving but not others.

II.D. How Might the Foregoing Considerations Evolve over Time?

Advances in economic research and in CBO's and JCT's capabilities might affect the advantages and disadvantages of the different approaches discussed in the preceding subsections. For example, as the evidence base in the health area continues to expand and if CBO and JCT can make further improvements in their tool kits, the agencies may become better able to estimate the effects of policies on enhancing health and enabling a more productive workforce.

To improve both conventional and dynamic estimates, CBO and JCT could regularly publish "calls for research" on topics where a scarcity of evidence particularly hinders the agencies' work; one example is CBO's (2024b) article in the *Journal of Economic Perspectives*, which was solicited by one of the authors of the present paper (Williams) in her role as editor. In addition, there probably will be considerable "learning by doing" on the part of CBO and JCT, with positive spillovers among investments in different aspects of dynamic analysis. Therefore, it could be useful for the strengths and weaknesses of different approaches for budget estimates to be reviewed regularly, and decisions about the best approach might change over time.

III. How?

In this section, we draw on CBO's existing dynamic analyses to illustrate a process for "how" dynamic scoring might be done. JCT's dynamic analyses have been constructed somewhat differently, and we highlight one important aspect of JCT's approach below; for further information, see the references cited earlier.

III.A. Estimating Direct Effects of a Proposal on Labor, Capital, and Productivity

The starting point for a dynamic budget estimate is to assess the direct effects of a proposal on labor supply, saving and investment, and total factor productivity, leaving aside for the moment the financing of the policy. Those effects should be disaggregated into categories of labor with different marginal products and categories of capital with different gross marginal

products and depreciation rates. The effects depend on responses by households, businesses, implementers of legislation in the executive branch, and state and local governments; for example, an increase in federal highway funding could spur reinforcing or offsetting changes in state highway funding. The effects need to be estimated with a time dimension: Conventional budget estimates reflect the fact that federal funding affects outlays with a lag, and dynamic estimates should reflect also the fact that federal outlays may have lagged effects on labor, capital, and productivity. Capturing all the conceivable effects of a proposal on those economic factors is not possible, any more than capturing all the conceivable nondynamic effects is possible in a conventional estimate; the challenge for analysts remains to identify and include the effects that are likely to be most important and for which there is some basis for assessing the likely sign and magnitude.

As discussed earlier, investments in modeling and practice lower somewhat the marginal cost of future estimates because models and experience can be used repeatedly, in both conventional estimates and dynamic ones. But even when a model has been built and tested in a policy area, an analyst dealing with a new proposal in that area still needs to translate the proposal's legislative language into model inputs and review model outputs—and often undertakes this process many times as Congress iterates on alternative versions of the proposal. Therefore, producing dynamic estimates still will be more time-consuming than producing conventional estimates in most cases.

III.B. Estimating Changes in Output and Other Economic Factors

With the estimated effects of a proposal on labor, capital, and productivity in hand, an analyst can estimate the follow-on effects of the proposal on potential output, actual output, interest rates, inflation, and other economic factors. We address three issues that arise in such estimates—effects on aggregate demand, the impact of different methods of financing, and the challenge of aggregation across different elements of a proposal.

First, a proposal can cause actual output to deviate from potential output through its effect on aggregate demand. Some proposals might change aggregate demand roughly in line with potential output, and for other proposals, the Federal Reserve would adjust short-term interest rates to bring aggregate demand into balance with potential output and keep inflation close to the Fed's target. Still, output, interest rates, and inflation might all change, and shifts in those economic factors affect the federal budget through tax collections, debt service, and other channels. Moreover,

some proposals are *designed* to shift aggregate demand relative to potential output, such as fiscal stimulus when economic activity is weak. Dynamic estimates include shifts in aggregate demand to capture those economic impacts.⁹

Second, including the effects of financing a proposal is crucial to generating a useful assessment of its economic impact. For example, consider an increase in federal investment financed by an increase in borrowing. Taking account of the positive impact on growth of the additional government investment without taking account of the negative impact of crowding out some private investment would be misleading. And if the net impact on growth is an increase, including the resulting positive impact on the budget without including the negative impact on the budget of paying higher interest rates on outstanding debt also would be misleading—especially because the current high level of debt means that changes in rates can have large effects on the government’s fiscal position.¹⁰

For an illustration of both these points, consider CBO’s (2021c) analysis of a potential increase in infrastructure spending. In the first scenario considered, additional infrastructure spending is offset by a reduction in some other form of government spending that does not constitute investment.¹¹ CBO estimated that this policy would have little imprint on output in the short run, a growing positive effect on output over time, and little effect on interest rates; overall, the economic effects of the policy would reduce the budgetary cost by about one-third. In the second scenario, additional infrastructure spending is financed by borrowing. CBO estimated that this policy would boost output for a while but have little effect in the long

9. The economic impacts of shifts in aggregate demand depend on underlying economic conditions—as do other aspects of dynamic analyses. The macroeconomic models used by CBO and JCT are aligned with CBO’s baseline economic and budget projections and are updated as those projections evolve—just as all of the agencies’ models are updated regularly to reflect the baseline projections and other information.

10. Presumably a change in borrowing would be assumed to occur with the same maturity structure as outstanding debt. Some economic models imply that government budget deficits can be self-financing in certain circumstances so that additional debt imposes no burden. But in that case the whole purpose of budget estimates should be reconsidered; our maintained hypothesis in this paper is that deficits are not self-financing. Separately, long-term analyses of government borrowing face the question of how long debt can rise before changes in spending or taxes are needed to avoid economic instability—and assuming any future changes runs afoul of the principle that CBO and JCT should not predict policy changes. But even with the current high level of government debt, this question can be avoided or answered in a fairly neutral manner over the next few decades for most proposals.

11. In CBO’s analysis, the reduction in other government spending is implicitly assumed to have no economic effects.

run (because the additional borrowing would crowd out some private investment) and would raise interest rates in both the short and long runs; overall, the economic effects of the policy would *increase* the budgetary cost by about one-fourth. These patterns—varying between financing methods and over time for either method—are important for policymakers to understand.

We note also that, by long-standing convention, CBO's budget estimates do not include debt service from changes in government borrowing: A proposal that would raise spending by \$10 billion this year, leave spending unchanged in subsequent years, and make no change in taxes would be reported to increase deficits over the budget window by \$10 billion, not \$10 billion plus the interest costs that would be incurred later.¹² However, CBO's dynamic analyses do include the debt service that arises from changes in interest rates applied to the existing debt, which is what we are describing here; see CBO (2016, footnote 19) and CBO (2021c, footnote 5). Thus, a dynamic budget estimate of the hypothesized proposal would include the debt service stemming from higher interest rates applied to currently outstanding debt, but it would not include debt service on the \$10 billion.

Third, aggregating different elements of a legislative proposal represents a significant challenge to dynamic analysis. Bills often are constructed in pieces, and multiple analysts at CBO and JCT often work on the estimate for a single bill. To produce a dynamic estimate for the spending bills on which CBO focuses, the pieces usually do not interact in significant ways, so conventional estimates of the pieces can be constructed in a modular fashion and added together for an overall conventional estimate. Suppose that analysts estimated for each piece not only the conventional effects but also the direct effects on labor, capital, and productivity, the resulting change in government borrowing, and the follow-on effects on other economic measures. Those dynamic estimates of the pieces could be added together for an overall dynamic estimate. For example, if a bill paid for extra spending for a chosen purpose by cutting other types of spending, the sum across the bill's provisions of all the financing impacts would be zero, and the sum of the impacts of all the provisions on labor, capital, and productivity would match the effects of estimating the impact of the bill as a package.

By contrast, for the revenue bills on which JCT focuses, there are often significant interactions between provisions, so neither conventional

12. CBO (2024c) addressed the possibility of changing this convention.

nor dynamic estimates can be constructed piece by piece.¹³ As a result, conventional estimates of bills that change the tax code often need to be produced holistically, and dynamic estimates are the same in that regard.

III.C. Estimating Changes in the Budget

Given the estimated effects of a proposed policy on labor, capital, output, interest rates, and inflation, the impact on the government budget can be estimated using CBO's usual methodology for budgetary feedback, as noted earlier.

III.D. Reporting the Estimates

The first report of a budget estimate is often a private conversation between CBO or JCT analysts and staff members for a congressional committee that is developing a proposal. Because that estimate may influence the committee's further work, those staff members want to receive a close approximation to what the agencies would ultimately publish if the proposal was approved by the committee. Therefore, if that first report does not include the proposal's dynamic effects and those effects turn out to matter consequentially for the published estimate, then the iterative process of developing the proposal will have been based on potentially misleading information. But to include dynamic effects in early conversations would require CBO's and JCT's analysts to have a sense of those effects—which could be difficult to develop on tight timetables. This observation again highlights the value of standardized and simplified dynamic estimation approaches where such approaches are tractable and credible, because then more information on dynamic effects could be provided earlier in the proposal development process.

For published estimates, the estimating agencies often show conventional effects and dynamic effects separately and then added together, to maximize transparency. That approach has been taken by JCT for some time, and it was used, for example, in CBO's (2021c, table 3) dynamic analysis of infrastructure spending.¹⁴

13. One implication of these interactions is that the "stacking order" for tax provisions affects the estimated impact of each provision. For example, if a reduction in tax rates is combined with a strengthened minimum tax, the rate cut will cost more revenue and have larger dynamic effects if it is evaluated first than if it is evaluated second.

14. Because Congress's budget process relies on point estimates, CBO and JCT focus on point estimates as well. And because estimates are not usually derived from probabilistic models, prediction intervals do not flow naturally out of the estimating process. But the agencies do develop and report prediction intervals on occasion, especially for dynamic analyses, and that information can be useful for policymakers; for example, see CBO's (2013b) report on the economic effects of comprehensive immigration reform.

IV. An Illustrative Example: Potential Changes in Immigration Policy

Immigration policy is an important element of economic policy, ranging from specific issues such as green cards for STEM (science, technology, engineering, and math) workers to general topics like whether and how immigration rules should respond to slower growth of the domestic-born workforce.¹⁵ With conventional scoring, budget estimates of proposals to change immigration policy generally capture only the expected effects on federal spending and not the expected effects on federal revenue. By contrast, dynamic scoring and the intermediate alternative of population change scoring capture effects on revenue as well. In this section we compare these approaches to estimating the impact of immigration proposals.

IV.A. Conventional Scoring

Consider a proposal that would increase the number of immigrants. Under the conventional estimating approach of holding output fixed, the income earned by the new immigrants would be assumed implicitly to be offset by reduced earnings by other people, so there would be no estimated change in income or payroll tax revenue. However, the new immigrants would be modeled (appropriately) as receiving federal benefits for which they are legally eligible, such as the Affordable Care Act's subsidies for health insurance, so there would be an estimated increase in federal spending. As a result, essentially any proposal that would increase the number of immigrants would be estimated to increase budget deficits.

Conventional scoring has been applied to most immigration proposals considered by Congress, and the resulting estimates have not provided a balanced picture of the likely impact of those proposals on the budget. Consider section 80303 of H.R. 4521, which was a provision in the America COMPETES Act to increase the availability of green cards for holders of advanced STEM degrees. Economic research shows that the federal income and payroll tax revenue generated by such immigrants would be much greater than the spending for federal benefits they would receive; for example, a 2017 National Academies of Sciences, Engineering, and Medicine report concluded that, on average, high-skilled immigrants contribute hundreds of thousands of dollars more in federal tax revenue than they receive in federal benefits over their working lives. Yet, CBO (2022) reported that this

15. This section draws in part on work by Elmendorf and Williams (2024), which was done in collaboration with Alex Arnon and Kent Smetters from the Penn Wharton Budget Model.

legislative provision would *cost* the federal government \$3.1 billion over a decade.

Note that conventional scoring of an immigration proposal relies on detailed modeling of how the proposal would affect the number and characteristics of people entering the country and of domestic residents who would change their legal status, which in turn could affect emigration, eligibility for federal benefits, and other outcomes like job changes and earnings.¹⁶ The result of this population modeling can be visualized as a table of changes in the numbers of people with different combinations of attributes—and the table offers much of the information needed to go beyond conventional scoring.

IV.B. Population Change Scoring

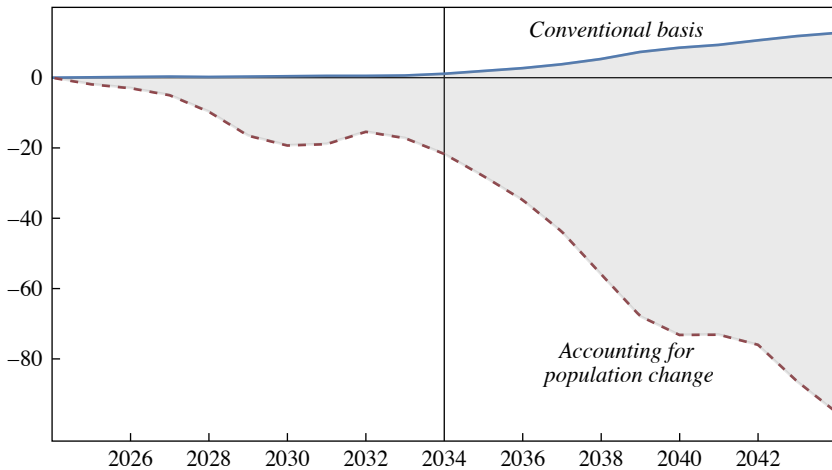
The estimating agencies do not always use conventional scoring for immigration proposals. In estimating the effects of S. 744, a proposal for comprehensive immigration reform, CBO (2013b, 2) explained: Following the “long-standing convention” of “assuming that total employment was unchanged would imply that any employment of the additional immigrants would be offset one-for-one by lower employment elsewhere in the population. Because that outcome would be highly implausible, CBO and JCT relaxed the assumption of fixed GDP and employment and incorporated into the cost estimate their projections of the legislation’s direct effects on the U.S. population, employment, and taxable compensation,” which primarily affected the amount of additional tax revenue that would have resulted from enacting the bill. In other words, this proposal would so clearly change output that relaxing the conventional assumption of fixed output was the only reasonable way to proceed—and the assumption was then relaxed in the simplest and most straightforward way.

This population change approach incorporates all the elements of a conventional estimate as well as the direct budgetary effects of changing the number of people in the country—in particular, the effects on taxable compensation and therefore on income and payroll tax revenue. But population

16. As an example, one of us (Williams) worked with Matthew Esche and Jeremy Neufeld on population modeling for a representative policy in the spirit of section 80303 and other recent proposals such as S. 2384 (the Keep STEM Talent Act of 2023). That representative policy would exempt from statutory limits employment-based green cards (EB-1, EB-2, and EB-3) for applicants who have earned a doctoral or master’s degree in STEM at a US research institution or foreign equivalent, as well as for an accompanying spouse and minor children. For the code and documentation for that population modeling, see Esche, Neufeld, and Williams (2024).

Figure 2. Estimated Effects on Deficits of an Increase in Green Cards to Holders of Advanced STEM Degrees, Conventional and Population Change Approaches

Effect on the primary deficit (billions of dollars)



Source: Reprinted from Elmendorf and Williams (2024, fig. 1), reflecting estimates produced by the Penn Wharton Budget Model.

change estimates do not incorporate all the likely economic effects of a proposal, as a fully dynamic approach would; we return to this issue shortly. To our knowledge, CBO has applied the population change approach to four legislative proposals over time: S. 2611 (CBO 2006), Senate Amendment 1150 to S. 1348 (CBO 2007), S. 744 (CBO 2013a), and H.R. 2131 (CBO 2014a); the first three proposals represented comprehensive immigration reform, and the fourth aimed to increase the number of visas for high-skilled noncitizens.

Two examples illustrate how markedly population change estimates can differ from conventional estimates:

- Both H.R. 2131 and section 80303 of H.R. 4521 would have increased the availability of green cards to holders of advanced STEM degrees. CBO's conventional estimate for section 80303 showed an increase in budget deficits of \$3 billion from 2022 to 2031, whereas CBO's population change estimate for H.R. 2131 showed a decrease in budget deficits of \$110 billion from 2014 to 2024.
- Elmendorf and Williams (2024) compared conventional and population change estimates for a section 80303-style proposal. Figure 2 summarizes their estimates of the effects on budget deficits over the

next twenty years. The conventional estimate (the solid line) shows a slight increase in deficits, while the population change estimate (the dashed line) shows a decrease in deficits that grows significantly over time as lawful permanent residents naturalize and sponsor family members, with an assumed nine-year lag, following Carr and Tienda (2013).

A natural follow-up question is how much conventional and population change estimates would differ for immigration proposals that involved a broad cross section of immigrants, whose average taxable compensation would be lower than that of high-skilled immigrants and whose average take-up of federal benefits would presumably be higher. CBO's recent update to its baseline economic and budget projections provides an answer. CBO (2024d) explained that a surge in immigration that occurred during 2022 and 2023 (and that CBO expected to continue to some extent for a few more years) comprised people who are much less educated, on average, than holders of advanced STEM degrees.¹⁷ CBO expected that about half of this group will receive authorization to work; some of the other half may work as well, but immigrants who work without authorization are less likely to pay taxes. CBO estimated that, over the next decade, the surge in immigration will increase federal revenue by \$1.2 trillion (mostly from additional income and payroll taxes) and increase federal spending by \$0.3 trillion (mostly from additional federal benefits), for a net reduction in deficits of \$0.9 trillion. Those figures suggest that legislative proposals that changed the number of less educated immigrants probably would receive starkly different population change estimates and conventional estimates.

A further question is whether population change estimates—estimates that include the effects of immigration proposals on revenue as well as spending—can be produced in reasonable time frames. One approach is for the estimating agencies to apply a microsimulation model for taxes to the estimated changes in numbers of people with different attributes that are generated currently for conventional scoring. Because the attributes needed to predict employment and compensation are similar to those needed to predict benefit eligibility, this approach would work fairly well—but still would be time-consuming. An alternative approach is for the estimating

17. CBO attributed the surge in immigration to people who are entering the country without inspection, who were allowed to enter the country lawfully through parole authority and are awaiting proceedings in immigration court, or who previously resided in the country under a temporary legal status and remained after that status expired.

agencies to calibrate and use a microsimulation model each time that baseline projections are updated, apply the results to create a “calculator” with changes in revenue per person for each combination of attributes in the population modeling, and use that calculator for all proposals until the baseline projections are updated again. This approach would allow for shorter turnaround times than running a microsimulation model for each request, and two of us (Elmendorf and Williams) are exploring this method in joint work with Theresa Gullo.

IV.C. Dynamic Scoring

Population change estimates capture more effects of immigration proposals than do conventional estimates, but they do not capture all the effects that would be included in dynamic estimates, such as potential changes in capital investment, interest rates, productivity, and more. CBO has produced dynamic analyses of two legislative proposals on immigration—S. 2611 (CBO 2006, appendix) and S.744 (CBO 2013b)—that were separate from the official budget estimates for those bills. CBO (2024d) used an analogous approach to update the agency’s baseline projections for the immigration surge.

Space constraints do not allow us to discuss all the economic effects of changes in immigration policy, so for illustrative purposes, we comment on just one effect that was featured in CBO’s baseline update—the impact of immigration on productivity. A number of papers present compelling evidence that when high-skilled immigrants move to the United States, their research becomes more productive (Prato 2025), they generate valuable innovations at disproportionately high rates (Bernstein and others 2022), and they start new firms that lead them to act more as *job creators* than *job takers* of domestic residents (Azoulay and others 2022); few papers link these innovation-related outcomes directly to productivity, although Peri (2012) and Prato (2025) are two exceptions. CBO (2024d, 19) estimated that 3 percent of people in the immigration surge are high-skilled STEM workers. That is a small percentage, but it represents a large absolute number—several hundred thousand people, which exceeds the statutory annual cap on H-1B visas. CBO estimated that those workers will contribute to innovation-related activities that raise total factor productivity roughly 0.3 percent by 2034 relative to what it would have been otherwise, boosting output and income for decades to come. However, CBO also estimated that lower average educational attainment and fewer years of work experience in the surge population relative to domestic workers would pull down productivity by a roughly offsetting amount in 2034.

Constructing a dynamic estimate of an immigration proposal that incorporates all the potential economic and budgetary effects requires considerable modeling, as demonstrated by the complexity of the analysis that CBO reported for its baseline update. Hence, a natural question is what share of the overall budgetary impact of a proposal could be captured by a simpler population change estimate. The answer would vary across legislative proposals, of course, but one data point is available: CBO estimated that the immigration surge will reduce deficits over the next decade by roughly \$600 billion through only the change in population and its direct effects on revenue and spending, and by an additional nearly \$300 billion when incorporating all of the dynamic effects. If the surge had been the result of a legislative proposal, a population change estimate would thus have captured about two-thirds of the budgetary impact captured by a dynamic estimate.

V. Illustrative Examples: Potential Changes in Federal Investment and Federal Permitting of Investment

Federal investments in infrastructure, education, and R&D are made with the intention of increasing physical capital, human capital, and productivity—and thereby spurring economic growth. CBO has examined the economic impact of federal investment but has never included that impact in official budget estimates. In this section, we explore potential dynamic scoring for federal investment and federal permitting of investment.

V.A. CBO's 2021 Report on Physical Infrastructure Spending

CBO has analyzed the effects of federal investments for some time (for example, see CBO 2013c, 2016), and it made a substantial further advance in its 2021 report, “Effects of Physical Infrastructure Spending on the Economy and the Budget Under Two Illustrative Scenarios” (CBO 2021c). The report was prepared in response to a request from then Senator Robert Portman, who wanted to bring a more comprehensive analysis to bear in developing the bipartisan Infrastructure Investment and Jobs Act (Bolton 2021).

The 2021 report effectively describes how CBO could produce a dynamic estimate of a proposal to change federal infrastructure spending if Congress requested it. The report is exemplary in its rigorous analysis and methodological transparency. Moreover, the modeling is presented in a modular and flexible fashion, so it could be applied readily to a variety of specific proposals. The report focuses on five key factors:

- **How state and local governments respond to additional federal funding.** Additional federal spending on infrastructure is expected to induce a partially offsetting reduction in infrastructure spending by state and local governments. For other types of federal investment, the expected response of state and local governments to a federal policy change could be different; moreover, other federal investments might spur an offsetting or reinforcing response by private actors as well.
- **How quickly funding leads to outlays.** Federal spending generally lags the authorization to undertake spending, to a degree that varies across types of spending. This factor is a standard aspect of CBO's conventional estimates, so including it in dynamic estimates would be straightforward.
- **How quickly outlays increase productivity.** CBO expects that additional federal infrastructure spending would increase productivity with a lag, with 40 percent of the effect occurring in the first year, 80 percent by the second year, and 100 percent by the seventh year. For other types of federal investment or for specific types of infrastructure spending, different lagged patterns might be appropriate and could be drawn from the research literature.
- **How much outlays increase productivity.** CBO expects that federal infrastructure has a gross annual return of 12.4 percent and depreciates by 3.2 percent annually, implying a net annual return of 9.2 percent. Again, different figures could be appropriate for different types of federal investment.
- **How outlays are financed.** As we described earlier, CBO examines two scenarios—one in which the additional infrastructure spending is financed by cuts in other spending that does not directly affect productivity, and one in which the additional spending is debt financed. With these two scenarios modeled, CBO could model an intermediate financing approach as well.

In sum, this modeling approach could be applied to a change in any federal investment for which there is a reasonable basis for determining these five factors. The most challenging factors to ascertain are the third and fourth regarding the effects of federal investment on productivity, and we turn to this challenge in the context of R&D spending shortly. More generally, the evidence base on the economic effects of federal investments will improve over time, and as it does, the feasibility and value of dynamic estimates will improve as well.

V.B. Application to Potential Changes in Federal Spending for Research and Development

For federal spending on R&D, the research literature is sufficiently developed that CBO's 2021 framework for analyzing federal infrastructure spending could be applied in a fairly straightforward way. One of us (Williams) has worked with Gullo, Benjamin Page, and David Weiner (Gullo and others 2025) to illustrate how that analysis could be done, so we refer interested readers to that paper and summarize the main points here.

For decades, economists have explored the effect of investments in research and innovation on economic growth; for example, see Romer (1990), Grossman and Helpman (1991), and Aghion and Howitt (1992). However, empirical evidence on the relationship between R&D and productivity growth has come slowly. Jones and Summers (2022) applied a somewhat indirect macroeconomic approach to infer that the marginal social return to R&D tends to be high (see their table 7). Azoulay and others (2019) and Myers and Lanahan (2022) estimated the spillovers from R&D funded by the National Institutes of Health and the Department of Energy, respectively, and found that the spillovers are substantial within a decade; however, neither of these papers quantified the effects of R&D on productivity.

Fieldhouse and Mertens (2024) and Dyèvre (2024) offered a key step forward, essentially stitching together the microeconomic evidence on returns to R&D with macroeconomic aggregates. Using quite different empirical approaches, these papers arrived at broadly similar quantitative estimates of the contribution of federal R&D to productivity. For example, Fieldhouse and Mertens used a narrative approach to classify postwar changes in appropriations for R&D at five major federal agencies and then used those data to make structural estimates of the economic returns to those appropriations. They concluded that the returns are between 140 and 210 percent, which implies that federal R&D has been responsible for at least one-fifth of overall postwar growth of total factor productivity in the United States.¹⁸

Gullo and others (2025) discuss how to apply this evidence in CBO's 2021 framework to analyze the economic and budgetary impact of changes in federal R&D investment. For example, to address some of the key factors identified by CBO, the research literature suggests that federal R&D crowds

18. The research literature distinguishes between defense and nondefense R&D. CBO (Campbell and Shirley 2018) argued that most defense R&D does not provide significant spillovers to the private sector, although there are exceptions, and therefore defense R&D does not substantially influence the agency's economic analysis. Recent research generally supports that position. More broadly, CBO (2014c) explored the impact of federal policies on innovation.

in private R&D, Fieldhouse and Mertens's (2024, fig. 6) estimated impulse response functions provide a basis for projecting time lags between R&D and productivity, and the Bureau of Economic Analysis has done extensive work on depreciation rates for R&D.

Because much of federal R&D can be denominated readily in dollars and probably has effects that scale roughly linearly over relevant ranges, modeling the effects of changing federal R&D in simplified ways could produce informative estimates. For example, one could imagine a spreadsheet with estimated effects on the economy and the federal budget of each dollar of additional funding through different agencies (say, the National Institutes of Health versus the National Aeronautics and Space Administration) and for different purposes (say, basic research versus applied research). A tool like this could be updated with each baseline and provide a tractable way to produce dynamic budget estimates on compressed timelines. In addition, investments in piloting, validating, and improving this approach could provide a template that could be applied over time in other policy areas.

V.C. Application to Potential Changes in Federal Permitting for Physical Infrastructure

For potential changes in federal permitting for infrastructure, CBO's 2021 framework seems appropriate for doing dynamic analysis—but the available research base for informing key parameters of *conventional* estimates (and therefore also dynamic estimates) is lacking. CBO's (2024b) description of the disconnect between the agency's modeling capability and the available research on this topic may spur additional research. As the research literature develops over time, the ability to produce rigorous budget estimates will improve.

In recent years Congress has been interested in the processes through which infrastructure projects are licensed, or “permitted,” before construction occurs. The National Environmental Policy Act (NEPA), which was enacted in the 1970s, can be viewed as an “umbrella” statute for compliance with a broad set of federal, state, local, and tribal permitting policies. NEPA covers both federally funded and federally subsidized projects (except in cases where the government exercises no discretion when awarding the funding, as with, for example, the Department of Housing and Urban Development's Housing Trust Fund, where the resources are dispersed to states to allocate). As discussed by Liscow (2025), the goal of NEPA is to force review of how proposed projects would affect the “environment,” including both nature and people. Potter, Datta, and Stapp (2022) argued that the NEPA

permitting process is hindering other national priorities such as a clean energy transition and therefore should be restructured.

To estimate the economic and budgetary effects of potential changes in NEPA, several factors are relevant: First, would the changes shorten the timeline on which infrastructure projects are built? Second, if so, how many more infrastructure investments would occur within a given time frame (including both the acceleration of projects that would occur anyway and the pursuit of new projects)? And third, how would the additional projects affect the economy and the federal budget?

The first and second questions matter for conventional budget estimates, and the third question would need to be addressed for dynamic estimates. For the third question, CBO's 2021 framework applies directly to federally funded investments, and that framework together with CBO's standard modeling of the effects of private capital can be applied to federally subsidized private investments. In contrast, for the first two questions, the available research base is much more limited—which hinders CBO's ability to produce conventional estimates as well as dynamic estimates.

Consider a proposal aimed at reducing time to build by shortening the NEPA permitting process. To produce a conventional cost estimate, CBO would first need to quantify the additional spending on infrastructure that would occur within the budget window and the budgetary impact of that additional spending. For example, CBO would try to assess the extent to which organizations that opposed infrastructure projects could prevent an acceleration of permitting by shifting their attention to processes imposed by state, local, or tribal governments. Then CBO would need to assess the elasticity of infrastructure projects with respect to changes in time to build. This elasticity is conceptually straightforward: Shortening permitting time is analogous to shortening commercialization lags for new pharmaceutical drugs, which one of us (Williams) estimated with coauthors to have increased private investment in drug development (Budish, Roin, and Williams 2015). Unfortunately, we are aware of no empirical literature that could inform such an assessment; in the words of CBO's (2023a, 3) cost estimate for H.R. 1, the Lower Energy Costs Act: "CBO has insufficient information to determine the number of projects . . . that could be generated."

VI. Conclusion

In recent years Congress has asked its official estimators—CBO and JCT—to provide more information about the economic effects of policy changes. The agencies have responded by drawing on an expanding body

of economic research, improving their modeling tools, gaining experience at such analysis, and publishing reports regarding their methodologies and results. This work has reinforced the value of the agencies providing unbiased, evidence-based analyses of the effects of potential policies on labor, capital, productivity, output, and other economic measures.

Decisions by Congress about the extent to which such analysis should be included in official budget estimates—through dynamic scoring or intermediate alternatives to conventional scoring—involve trade-offs regarding time, resources, accuracy, and comparability. With labor force growth slowing and federal debt rising, interest in the intersection of public policy, economic outcomes, and budget impacts shows no signs of diminishing.

ACKNOWLEDGMENTS For valuable conversations and comments, we are grateful to Alexander Arnon, Alan Auerbach, Jan Eberly, Theresa Gullo, G. William Hoagland, Douglas Holtz-Eakin, Deborah Lucas, Donald Marron, Dylan Matthews, Alexander Mechanick, Benjamin Page, Kent Smetters, Jón Steinsson, Caleb Watney, David Weiner, David Wessel, and staff members of CBO and JCT. The case studies we discuss draw on collaborative work with Arnon, Gullo, Page, Smetters, Weiner, Matthew Esche, Zachary Liscow, and Jeremy Neufeld. We are solely responsible for the contents of the paper, and the views expressed here should not be attributed to any of the individuals acknowledged or any institution with which the authors are currently or formerly affiliated.

References

- Aghion, Philippe, and Peter Howitt. 1992. "A Model of Growth Through Creative Destruction." *Econometrica* 60, no. 2: 323–51. <https://doi.org/10.2307/2951599>.
- Auerbach, Alan J., Itai Grinberg, Thomas Barthold, Nicholas Bull, W. Gavin Elkins, Pamela Moomau, Rachel Moore, Benjamin Page, Brandon Pecoraro, and Kyle Pomerleau. 2017. "Macroeconomic Modeling of Tax Policy: A Comparison of Current Methodologies." *National Tax Journal* 70, no. 4: 819–36. <https://doi.org/10.17310/ntj.2017.4.06>.
- Azoulay, Pierre, Joshua S. Graff Zivin, Danielle Li, and Bhaven N. Sampat. 2019. "Public R&D Investments and Private-Sector Patenting: Evidence from NIH Funding Rules." *Review of Economic Studies* 86, no. 1: 117–52. <https://doi.org/10.1093/restud/rdy034>.
- Azoulay, Pierre, Benjamin F. Jones, J. Daniel Kim, and Javier Miranda. 2022. "Immigration and Entrepreneurship in the United States." *American Economic Review: Insights* 4, no. 1: 71–88. <https://doi.org/10.1257/aeri.20200588>.
- Bernstein, Shai, Rebecca Diamond, Abhisit Jirananphawiboon, Timothy McQuade, and Beatriz Pousada. 2022. "The Contribution of High-Skilled Immigrants to Innovation in the United States." Working Paper 30797. Cambridge, Mass.: National Bureau of Economic Research. <https://doi.org/10.3386/w30797>.
- Bolton, Alexander. 2021. "CBO Says Bipartisan Infrastructure Bill Would Add \$256B to Deficit over 10 Years." *The Hill*, August 5. <https://thehill.com/homenews/senate/566590-cbo-says-infrastructure-bill-would-add-256-billion-to-deficit-over-10-years/>.
- Budish, Eric, Benjamin N. Roin, and Heidi Williams. 2015. "Do Firms Underinvest in Long-Term Research? Evidence from Cancer Clinical Trials." *American Economic Review* 105, no. 7: 2044–85. <https://doi.org/10.1257/aer.20131176>.
- Campbell, Sheila, and Chad Shirley. 2018. "Estimating the Long-Term Effects of Federal R&D Spending: CBO's Current Approach and Research Needs." Blog Post, Congressional Budget Office, June 21. <https://www.cbo.gov/publication/54089>.
- Carr, Stacie, and Marta Tienda. 2013. "Family Sponsorship and Late-Age Immigration in Aging America: Revised and Expanded Estimates of Chained Migration." *Population Research Policy Review* 32: 825–49. <https://doi.org/10.1007/s11113-013-9300-y>.
- CBO (Congressional Budget Office). 1995. "Budget Estimates: Current Practices and Alternative Approaches." Washington: Congressional Budget Office.
- CBO. 2006. "S. 2611: Comprehensive Immigration Reform Act of 2006." Washington: Congressional Budget Office.
- CBO. 2007. "Senate Amendment 1150 to S. 1348, the Comprehensive Immigration Reform Act of 2007." Washington: Congressional Budget Office.
- CBO. 2012a. "Fair-Value Accounting for Federal Credit Programs." Washington: Congressional Budget Office.
- CBO. 2012b. "Raising the Excise Tax on Cigarettes: Effects on Health and the Federal Budget." Washington: Congressional Budget Office.

- CBO. 2012c. "How the Supply of Labor Responds to Changes in Fiscal Policy." Washington: Congressional Budget Office.
- CBO. 2013a. "Cost Estimate for S. 744, Border Security, Economic Opportunity, and Immigration Modernization Act." Washington: Congressional Budget Office.
- CBO. 2013b. "The Economic Impact of S. 744, the Border Security, Economic Opportunity, and Immigration Modernization Act." Washington: Congressional Budget Office.
- CBO. 2013c. "Federal Investment." Washington: Congressional Budget Office.
- CBO. 2014a. "H.R. 2131: Supplying Knowledge-Based Immigrants and Lifting Levels of STEM Visas Act (SKILLS Visa Act)." Washington: Congressional Budget Office.
- CBO. 2014b. "How CBO Analyzes the Effects of Changes in Federal Fiscal Policies on the Economy." Washington: Congressional Budget Office.
- CBO. 2014c. "Federal Policies and Innovation." Washington: Congressional Budget Office.
- CBO. 2015a. "The Effects of Potential Cuts in SNAP Spending on Households with Different Amounts of Income." Washington: Congressional Budget Office.
- CBO. 2015b. "Budgetary and Economic Effects of Repealing the Affordable Care Act." Washington: Congressional Budget Office.
- CBO. 2016. "The Macroeconomic and Budgetary Effects of Federal Investment." Washington: Congressional Budget Office.
- CBO. 2018. "How CBO Produces Fair-Value Estimates of the Cost of Federal Credit Programs: A Primer." Washington: Congressional Budget Office.
- CBO. 2019. "The 2019 Long-Term Budget Outlook." Washington: Congressional Budget Office.
- CBO. 2020. "The Effects of Pandemic-Related Legislation on Output." Washington: Congressional Budget Office.
- CBO. 2021a. "CBO Explains Budgetary Scorekeeping Guidelines." Washington: Congressional Budget Office.
- CBO. 2021b. "CBO's Policy Growth Model." Washington: Congressional Budget Office.
- CBO. 2021c. "Effects of Physical Infrastructure Spending on the Economy and the Budget Under Two Illustrative Scenarios." Washington: Congressional Budget Office.
- CBO. 2022. "Estimated Budgetary Effects of H.R. 4521, the America COMPETES Act of 2022." Washington: Congressional Budget Office.
- CBO. 2023a. "Estimated Budgetary Effects of Divisions A and B of H.R. 1, the Lower Energy Costs Act." Washington: Congressional Budget Office.
- CBO. 2023b. "Workbook for How Changes in Economic Conditions Might Affect the Federal Budget: 2023 to 2033." Washington: Congressional Budget Office.
- CBO. 2023c. "CBO Describes Its Cost-Estimating Process." Washington: Congressional Budget Office.
- CBO. 2023d. "CBO's Economic Forecasting Record: 2023 Update." Washington: Congressional Budget Office.

- CBO. 2023e. “The Budgetary and Economic Effects of S. 2488, the Raise the Wage Act of 2023.” Washington: Congressional Budget Office.
- CBO. 2024a. “How Changes in Economic Conditions Might Affect the Federal Budget: 2024 to 2034.” Washington: Congressional Budget Office.
- CBO. 2024b. “How Economists Could Help Inform Economic and Budget Analysis Used by the US Congress.” *Journal of Economic Perspectives* 38, no. 2: 3–24. <https://doi.org/10.1257/jep.38.2.3>.
- CBO. 2024c. “Adding Debt-Service Effects to CBO’s Cost Estimates.” Washington: Congressional Budget Office.
- CBO. 2024d. “Effects of the Immigration Surge on the Federal Budget and the Economy.” Washington: Congressional Budget Office.
- CBO. 2024e. “Extending Medicaid and CHIP Coverage for Children: Long-Run Budgetary Effects of the President’s 2025 Budget Proposal.” Washington: Congressional Budget Office.
- Demirel, U. Devrim, and Matthew Wilson. 2023. “Effects of Fiscal Policy on Inflation: Implications of Supply Disruptions and Economic Slack.” Working Paper 2023–05. Washington: Congressional Budget Office.
- Dyèvre, Arnaud. 2024. “Public R&D Spillovers and Productivity Growth.” Working Paper, June 27. https://www.arnauddyevre.com/files/arnaud_dyevre_jmp.pdf.
- Edelberg, Wendy. 2016. “Dynamic Analysis at CBO.” Presentation at the University of Chicago Booth School of Business, Chicago, February 29. <https://www.cbo.gov/publication/51286>.
- Elmendorf, Douglas W. 2015. “‘Dynamic Scoring’: Why and How to Include Macroeconomic Effects in Budget Estimates for Legislative Proposals.” *Brookings Papers on Economic Activity*, Fall: 91–133. <https://dx.doi.org/10.1353/eca.2015.0005>.
- Elmendorf, Douglas W., and Heidi Williams. 2024. “How Does Accounting for Population Change Affect Estimates of the Effect of Immigration Policies on the Federal Budget?” Penn Wharton Budget Model.
- Esche, Matthew, Jeremy Neufeld, and Heidi Williams. 2024. “Estimating Counterfactual Population and Status.” Manuscript, January 18. <https://drive.google.com/file/d/1HNcVldPV3r3Qtd0iYaQpsRauUsNOYVuW/view>.
- Fieldhouse, Andrew J., and Karel Mertens. 2024. “The Returns to Government R&D: Evidence from U.S. Appropriations Shocks.” Working Paper 2305. Dallas, Tex.: Federal Reserve Bank of Dallas. <https://doi.org/10.24149/wp2305r1>.
- Frentz, Nathaniel, Jaeger Nelson, Dan Ready, and John Seliski. 2020. “A Simplified Model of How Macroeconomic Changes Affect the Federal Budget.” Working Paper 2020–01. Washington: Congressional Budget Office.
- Grossman, Gene M., and Elhanan Helpman. 1991. “Quality Ladders in the Theory of Growth.” *Review of Economic Studies* 58, no. 1: 43–61. <https://doi.org/10.2307/2298044>.
- Gullo, Theresa, Benjamin Page, David Weiner, and Heidi L. Williams. 2025. “Estimating the Budgetary and Economic Effects of Research Investments.” Working Paper 33402. Cambridge, Mass.: National Bureau of Economic Research. <https://www.nber.org/papers/w33402>.

- Hall, Keith. 2015. "Answers to Questions About Dynamic Analysis." Blog Post, Congressional Budget Office, July 1. <https://www.cbo.gov/publication/50357>.
- Hall, Keith. 2024. "Beyond Dynamic Scoring: Enhancing CBO's Evaluation of Program Integrity Spending." Mercatus Special Study. Social Science Research Network. <http://dx.doi.org/10.2139/ssrn.4898038>.
- JCT (Joint Committee on Taxation). 2015a. "Macroeconomic Analysis at the Joint Committee on Taxation and the Mechanics of Its Implementation." JCX-3-15. Washington: Joint Committee on Taxation.
- JCT. 2015b. "A Report to the Congressional Budget Office of the Macroeconomic Effects of the 'Tax Relief Extension Act of 2015,' as Ordered to Be Reported by the Senate Committee on Finance." JCX-107-15. Washington: Joint Committee on Taxation.
- JCT. 2015c. "A Report to the Congressional Budget Office of the Macroeconomic Effects of H.R. 2510, 'Bonus Depreciation Modified and Made Permanent,' as Ordered to Be Reported by the House Committee on Ways and Means." JCX-134-15. Washington: Joint Committee on Taxation.
- JCT. 2017. "Macroeconomic Analysis of the Conference Agreement for H.R. 1, the 'Tax Cuts and Jobs Act.'" JCT-69-17. Washington: Joint Committee on Taxation.
- JCT. 2018a. "Overview of Joint Committee Macroeconomic Modeling." JCX-33-18. Washington: Joint Committee on Taxation.
- JCT. 2018b. "Macroeconomic Analysis of H.R. 6760, the 'Protecting Family and Small Business Tax Cuts Act of 2018' as Reported by the Committee on Ways and Means." JCX-79-18. Washington: Joint Committee on Taxation.
- JCT. 2020. "An Overview of a New Overlapping Generations Model with an Example Application in Policy Analysis." JCX-22R-20. Washington: Joint Committee on Taxation.
- JCT. 2023. "The Joint Committee on Taxation Revenue Estimating Process." Presentation, January 27. <https://www.jct.gov/getattachment/137bd134-a409-45ce-a582-a485a9b1733b/Revenue-Estimating-Process-January-2023.pdf>.
- Jones, Benjamin F., and Lawrence H. Summers. 2022. "A Calculation of the Social Returns to Innovation." In *Innovation and Public Policy*, edited by Austan Goolsbee and Benjamin F. Jones. Chicago: University of Chicago Press.
- Lasky, Mark. 2022. "The Congressional Budget Office's Small-Scale Policy Model." Working Paper 2022-08. Washington: Congressional Budget Office.
- Liscow, Zachary. 2025. "Getting Infrastructure Built: The Law and Economics of Permitting." *Journal of Economic Perspectives* 39, no. 1: 151-80.
- Lynch, Megan S., and Jane G. Gravelle. 2023. "Dynamic Scoring in the Congressional Budget Process." Washington: Congressional Research Service.
- McClelland, Robert, and Shannon Mok. 2012. "A Review of Recent Research on Labor Supply Elasticities." Working Paper 2012-12. Washington: Congressional Budget Office.
- Moore, Rachel, and Brandon Pecoraro. 2020. "Macroeconomic Implications of Modeling the Internal Revenue Code in a Heterogeneous-Agent Framework." *Economic Modelling* 87: 72-91. <https://doi.org/10.1016/j.econmod.2019.07.007>.

- Moore, Rachel, and Brandon Pecoraro. 2023. "Quantitative Analysis of a Wealth Tax for the United States: Exclusions and Expenditures." *Journal of Macroeconomics* 78: 103559. <https://doi.org/10.1016/j.jmacro.2023.103559>.
- Myers, Kyle R., and Lauren Lanahan. 2022. "Estimating Spillovers from Publicly Funded R&D: Evidence from the US Department of Energy." *American Economic Review* 112, no. 7: 2393–423. <https://doi.org/10.1257/aer.20210678>.
- National Academies of Sciences, Engineering, and Medicine. 2017. *The Economic and Fiscal Consequences of Immigration*. Washington: The National Academies Press. <https://doi.org/10.17226/23550>.
- Peri, Giovanni. 2012. "The Effect of Immigration on Productivity: Evidence from U.S. States." *Review of Economics and Statistics* 94, no. 1: 348–58.
- Potter, Brian, Arnab Datta, and Alec Stapp. 2022. "How to Stop Environmental Review from Harming the Environment." Institute for Progress.
- Prato, Marta. 2025. "The Global Race for Talent: Brain Drain, Knowledge Transfer, and Growth." *Quarterly Journal of Economics* 140, no. 1: 165–238.
- Reichling, Felix, and Charles J. Whalen. 2012a. "Assessing the Short-Term Effects on Output of Changes in Federal Fiscal Policies." Working Paper 2012–08. Washington: Congressional Budget Office.
- Reichling, Felix, and Charles J. Whalen. 2012b. "Review of Estimates of the Frisch Elasticity of Labor Supply." Working Paper 2012–13. Washington: Congressional Budget Office.
- Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98, no. 5: S71–S102.
- Seliski, John, Aaron Betz, Yiqun Gloria Chen, U. Devrim Demirel, Junghoon Lee, and Jaeger Nelson. 2020. "Key Methods That CBO Used to Estimate the Effects of Pandemic-Related Legislation on Output." Working Paper 2020–07. Washington: Congressional Budget Office.
- Shackleton, Robert. 2018. "Estimating and Projecting Potential Output Using CBO's Forecasting Growth Model." Working Paper 2018–03. Washington: Congressional Budget Office.
- Whalen, Charles J., and Felix Reichling. 2015. "The Fiscal Multiplier and Economic Policy Analysis in the United States." Working Paper 2015–02. Washington: Congressional Budget Office.

Comments and Discussion

COMMENT BY

G. WILLIAM HOAGLAND The question of why, when, and how to apply dynamic scoring in the legislative decision-making process deserves first a comment in the context of the country's broader fiscal challenges.

I began my budget career at the newly established Congressional Budget Office (CBO) in 1976. The federal deficit that year (\$74 billion) represented 4 percent of our national economy, and in the country's 200th anniversary, the accumulated level of public debt was 26.7 percent of GDP.¹ Nearly fifty years later, in 2024, the fiscal year just ended, the deficit reached \$2 trillion, 7 percent of GDP, and the public debt at \$28.2 trillion reached 99 percent of GDP (CBO 2024a).

Throughout these last fifty years, various Congresses and presidents have debated, advanced, enacted, and often failed to adopt policies that would balance the public's demand for government programs and services with the necessary resources to fund them.

In 1976, government receipts totaled 16.7 percent of GDP, expenditures 20.8 percent. Fast-forward, in 2024 receipts were 17.2 percent, while expenditures were 24.2 percent (CBO 2024a). The disconnect of receipts to expenditures over the period has resulted in pushing public debt to a level not seen since the peak years of World War II.

However, some parts of federal spending grew while others declined. Programs defined as mandatory grew from 8.8 percent of GDP in 1974 to

1. CBO, "Budget and Economic Data," under "Historical Budget Data," <https://www.cbo.gov/data/budget-economic-data#2>.

16.0 percent in 2024 (CBO 2024a), while programs defined as nondefense declined from 3.9 percent to 3.3 percent over the same time period.²

Finding it politically challenging, if not impossible, to reduce commitments to constituents for “mandatory” benefits (commitments largely made by previous holders of their offices) and unwilling to increase taxes to pay for those commitments, while having reduced and restrained nondefense programs, politicians’ magic elixir has become simply “increase growth.”

In a politically charged election year with continued projected growth in federal debt, both presidential candidates raised the issue of dynamic scoring even if they didn’t realize it, and they did so in a way that does not satisfy my concerns about the ability of either candidate to address the country’s unsustainable fiscal future.

In response to a question on how her administration would pay for her proposals, Kamala Harris responded on August 18: “Investing in a broad-based economy, everybody benefits, and it pays for itself in that way” (Peterson 2024, par. 19). In a speech on August 23 in Arizona, Donald Trump was asked how he would pay for the centerpiece of his economic agenda—extending the 2017 tax cuts. His response: “Growth, we’re gonna have tremendous growth” (Macri 2024, par. 2).

Understandably, the current congressional scorekeepers’ approach to analyzing legislation that includes individual and business behavioral responses but does not incorporate effects on the economy as a whole is incongruous with these campaign pronouncements. The result, conventional cost estimates are dismissed by partisans as incomplete or, worse, inaccurate. According to some proponents of dynamic scoring, a more fulsome, dynamic score of legislation that does not hold GDP constant will capture growth effects and therefore “pay for” new programs, tax cuts, or both.

Enter the current paper. Elmendorf, Hubbard, and Williams are to be commended on advancing the dynamic budget scoring debate. I am pleased that the authors acknowledge (at least in footnote) that they maintain the hypothesis that deficits are not self-financing.

The current paper extends, builds upon, and modifies Elmendorf’s (2015) seminal *BPEA* paper on why and how to do dynamic scoring. This paper adds the additional element of when to apply dynamic scoring. I would have added also what dynamic scoring should be applied to.

Overall, this paper is an important contribution to the literature of the federal budget and its processes. Nonetheless, the paper reaffirms, in my

2. The current paper focuses on dynamic scoring on a set of programs that fall into the category defined as nondefense, for example, immigration, research and development (R&D), and infrastructure.

mind, that the debate surrounding dynamic scoring will remain ongoing and controversial.

In this commentary, it is important to differentiate—as the paper does—between *dynamic analysis* and *dynamic scoring*. To this day, the agencies—after over two decades of debate—limit their dynamic scoring due to four criticisms raised in the paper: (1) time and resource constraints, (2) uncertainty, (3) magnitude of excluded effects, and (4) credibility of estimates.

The paper provides defensible responses to these criticisms—in particular, the recognition that whether it is dynamic or conventional scoring, all estimates have uncertainty. Hubbard’s comment on Elmendorf’s previous 2015 *BPEA* paper captures the uncertainty issue well: “The idea that the [dynamic] method’s uncertainty suggests the desirability of adopting a static-scoring answer that is *known with certainty to be incorrect* is not logical” (Hubbard 2015, 136; italics in the original). Apropos, the famous British statistician George E. P. Box is remembered for his salient observation that “all models are wrong, but some are useful” (Barroso 2018).

The paper’s focus is primarily on CBO’s spending estimates, not the Joint Committee on Taxation’s (JCT) revenue estimates. I welcome this focus. I have long argued that applying dynamic scoring only to federal tax legislation is incomplete, unfair, and inconsistent if it is not also applied to major spending legislation.

TWO GENERAL COMMENTS First, the authors provide clear evidence that the current congressional rule triggering a CBO or JCT dynamic analysis, the “major legislation” rule, is not effective or useful. That rule, in effect for nearly a decade and applied to all legislation estimated by CBO from 2019 to 2024, rarely has triggered a dynamic estimate. However, what was not clear from the paper was what is being proposed as an alternative to the rule. If, in the 119th Congress, the rule were to be dropped or revised, what would replace it, if anything? Or is the alternative, as seems to be inferred, a one-off, consult, negotiate, request between Congress and CBO or JCT?

It is certain, even if there is no congressional rule triggering a dynamic estimate, that it will be hard to put the genie back in the bottle. With or without a specific requirement, major tax legislation enacted over the last decade, including most recently the 2017 Tax Cuts and Jobs Act, included a dynamic score. The tax writing committees, in a bipartisan manner, will never abandon asking JCT for a dynamic score. As a result, dynamic scoring of revenue measures will continue to bias the consideration of tax legislation over spending legislation.

Second, what the paper does not adequately address is the treatment of applying dynamic scoring within the current rules, procedures, and limitations

placed upon CBO and JCT by the Congress itself. Phillip Swagel, in the *BPEA* conference discussion of Elmendorf's earlier critique of dynamic scoring, was reported to have raised the issue: "Economists too often overlook the significance of budget rules . . . including Senate procedures, which have a real effect on economic outcomes" ("General Discussion" 2015, 147–48). I raise in particular: (1) congressional scorekeeping rules, (2) the distinction and interaction between discretionary and mandatory spending programs, (3) jurisdictional disputes related to pay-as-you-go and legislative offset requirements, and (4) available agency resources.

ON THE QUESTION OF WHY? Clearly, including the effects of labor supply, savings and investment, capital, and productivity into cost estimates will improve estimates if, as cited in the paper, "the estimates are for policies that are likely to have significant feedback effects; if the estimates of those feedback effects are centered around the actual effects; and if the spread of those estimates is small" (Hall 2015, par. 12). Meeting those three factors—significant, centered, and small spread—in and of themselves, however, can create their own uncertainty.

Nonetheless, CBO and JCT should continue to provide dynamic analysis and when appropriate dynamic scoring. The question is when or for what?

I agree that in recent years the agencies have invested in dynamic analysis tools and gained much experience from such analyses. Those have often been targeted (one-off) investigations requested by specific members of Congress. As an example, at the request of the chairman of the Subcommittee on Health Care of the Senate Committee on Finance, CBO (2012) completed an exhaustive examination of the impact of raising excise taxes on cigarettes. CBO's (2021a) dynamic analysis of infrastructure spending on the federal budget, discussed in the paper, was in response to the specific request from then Senator Robert Portman.

These broad targeted analyses can provide CBO scorekeeping analysts with data and insights when conducting conventional scoring of individual bills.

Further, developing dynamic scoring for different areas of policies also makes perfect sense. It requires, however, that the agencies forecast or project, and prioritize what major policies the Congress will likely address in the future. This will not be an easy or exacting exercise.

Further, if we are abandoning the "major legislation" rule, who should determine when to apply dynamic scoring and for what policies? In order to maintain CBO's and JCT's credibility and quite frankly, their funding base, this ultimately must be a decision between the client (Congress) and the agencies.

The Budget Act of 1974, in creating the CBO, defines as its duties and functions first and foremost “assistance to Budget Committees,” second “assistance to Committees on Appropriations,” and to all other committees “to the extent practicable” (sect. 202).³ I am biased, but I conclude that the process for making decisions about dynamic scoring belongs first with the chairmen and ranking members of the Budget Committee, JCT, and the agencies, followed by the committee that “butters their bread”—the appropriations committees. Those committees will, of course, consult with the leadership of their respective chambers.

I note that the agencies’ combined resources have remained relatively flat over the last few years. Requiring additional dynamic analysis will put a strain on the agencies going forward unless their resources are to increase. Dynamic analysis will always take a back seat to “regular order” or CBO’s nearly 800 annual conventional cost estimates (CBO 2024b, appendix D) and an annual average of nearly 3,500 JCT bills and amendments. Note that in 2017 alone, JCT received over 7,000 requests.⁴

Finally, given Congress’s propensity to stack major legislation into the final days of a session, and given the time required to produce dynamic estimates, the inevitable tensions with CBO, JCT, and members of Congress need to be avoided. Dynamic scoring takes time.

ON THE QUESTION OF HOW? The paper’s research and development (R&D) discussion highlights one major concern. Setting aside whether R&D productivity effects are different if done via tax legislation or direct appropriations—current scorekeeping rules apply. Those rules can allow the tax writing committee to be scored with the macroeconomic benefits of increased revenues from providing R&D tax credits by using the additional dynamically scored revenues to offset other nondynamic tax provisions. However, increased direct appropriations for R&D funding would not provide the appropriations committees with any additional flexibility to increase appropriations in other nondynamic spending provisions.

The example argues for a rethink of the congressional scorekeeping rules if dynamic scoring is to apply fairly to both spending and revenue legislation, specifically guideline 3 (CBO 2021b). This rule prevents CBO from attributing appropriation increases in a particular account that might result in

3. Congressional Budget and Impoundment Control Act of 1974, Pub. L. 93–344, 88 Stat. 297.

4. Numbers of requests received by JCT by year as reported at JCT’s 2023 organizational meeting: 2016: 3,354; 2017: 7,063; 2018: 2,148; 2019: 2,654; 2020: 2,289; 2021: 4,259; 2022: 2,328 (Thomas Barthold, email to author, August 30, 2024).

reduced spending in a mandatory account being scored to the appropriations account. This matters critically in a time of statutorily defined discretionary spending caps.

However, the slippery slope that follows becomes what to apply dynamic scoring to for discretionary spending. Should an increase in discretionary Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) funding result in a reduction in mandatory Medicaid spending, and, if so, who should be scored with the reduction, the Senate Finance Committee or Senate Appropriations Committee?

While recently restricted, and directionally the opposite of what I just described, onetime changes in mandatory programs (CHIMPs) done in the appropriation process will be scored to the appropriators.⁵

The immigration population change model is a very good approach to providing *partial* dynamic analysis of immigration policy proposals. I totally agree that the conventional approach applied to most proposed immigration legislation does not provide a balanced analysis of such legislation on the federal budget. But other than immigration policy, few other policies come to mind that would benefit from the partial population change scoring approach. The exception might be health policy.

Which leads me to my disappointment that the paper leaves aside dynamic scoring in health care. Given the future growth of federal spending on health care, the population change model could apply here. As an example, the model could apply where the Medicare population with specific metabolic conditions qualify for—as an example—anti-obesity medications.⁶

Finally, a full dynamic analysis of immigration (or, for that matter, any other major federal policy area) would include the impact of the policy on state and local governments. A National Academies of Sciences, Engineering, and Medicine (2017) study found that first-generation immigrant adults and their dependents were more costly to state and local governments than the native-born, but that the second generation are among the strongest economic and fiscal contributors to the US economy, stronger than even the native-born population. This highlights the challenge of dynamic analysis that requires an extended time path to assess the policy's full impact.

5. When appropriation acts make changes to mandatory programs, those changes are reflected in cost estimates as discretionary costs or savings pursuant to congressional score-keeping guidelines. Because cost estimates for appropriation legislation are used when enforcing the caps, such changes affect the determination of whether funding exceeds the caps.

6. After the *BPEA* September 2024 conference, CBO released an analysis of the budgetary impact of anti-obesity medication on the federal budget. The analysis employs conventional, not dynamic scoring (CBO 2024c).

The immigration model described in the paper excluded state and local impact. Section 102 of the Unfunded Mandates Reform Act of 1995 requires CBO to analyze the impact of federal legislation on state, local, and tribal governments or the private sector.⁷

CONCLUSION In a polarized political environment, with multiple stakeholders and advocates advancing their own political agendas, broad systems thinking is required. A robust application of dynamic scoring applied to major public policy issues, as presented in this paper, is an important tool to inform and guide policymakers in managing this rich, complex system we live in today.

CBO has, should, and will continue to do dynamic analysis, as requested by members of Congress and as resources allow. Individual, one-off studies by CBO and JCT that incorporate variants of dynamic analysis will add to the general knowledge of the decision-makers and benefit conventional scorekeeping analyses.

However, when to apply dynamic scoring to specific legislation remains problematic. This is particularly true given current official scorekeeping rules. Those rules admittedly focus on deficits in the near term. Dynamic scoring by its very nature has a longer-term fiscal policy focus. Should those rules be changed or modified? It is unclear. Until they are changed or modified, however, dynamic scoring could delight some program advocates advocating for increased mandatory spending and reduced taxes while disappointing others whose programs are subject to the annual appropriation process (discretionary).

Further, in a politically charged environment, we cannot ignore the fact that some of those who advocate for dynamic scoring also may advance a specific progressive agenda that today will conflict with other more conservative, limited government agendas. At this time, CBO and JCT must not be drawn into this divisive environment, or they risk losing their credibility that has so aptly served the Congress for nearly fifty years. A bipartisan approach in the application of dynamic scoring is required. The debate will go on.

REFERENCES FOR THE HOAGLAND COMMENT

- Barroso, Guillem. 2018. ““All Models Are Wrong, But Some Are Useful”. George E. P. Box.” Blog Post, AdMoRe, May 3. <https://www.lacan.upc.edu/admoreWeb/2018/05/all-models-are-wrong-but-some-are-useful-george-e-p-box>.
- CBO (Congressional Budget Office). 2012. *Raising the Excise Tax on Cigarettes: Effects on Health and the Federal Budget*. Washington: Congressional Budget

7. Unfunded Mandates Reform Act of 1995, Pub. L. 104–4, 109 Stat. 48.

- Office. https://www.cbo.gov/sites/default/files/cbofiles/attachments/06-13-Smoking_Reduction.pdf.
- CBO. 2021a. “Effects of Physical Infrastructure Spending on the Economy and the Budget Under Two Illustrative Scenarios.” Washington: Congressional Budget Office. <https://www.cbo.gov/publication/57327>.
- CBO. 2021b. “CBO Explains Budgetary Scorekeeping Guidelines.” Washington: Congressional Budget Office. <https://www.cbo.gov/system/files/2021-01/56507-Scorekeeping.pdf>.
- CBO. 2024a. *An Update to the Budget and Economic Outlook: 2024 to 2034*. Washington: Congressional Budget Office. <https://www.cbo.gov/publication/60039>.
- CBO. 2024b. *The Congressional Budget Office’s Request for Appropriations for Fiscal Year 2025*. Washington: Congressional Budget Office. <https://www.cbo.gov/publication/60150>.
- CBO. 2024c. “How Would Authorizing Medicare to Cover Anti-Obesity Medications Affect the Federal Budget?” Washington: Congressional Budget Office. <https://www.cbo.gov/publication/60816>.
- Elmendorf, Douglas W. 2015. “‘Dynamic Scoring’: Why and How to Include Macroeconomic Effects in Budget Estimates for Legislative Proposals.” *Brookings Papers on Economic Activity*, Fall: 91–133.
- “General Discussion.” 2015. Conference discussion on “‘Dynamic Scoring’: Why and How to Include Macroeconomic Effects in Budget Estimates for Legislative Proposals” by Douglas W. Elmendorf. *Brookings Papers on Economic Activity*, Fall: 143–49.
- Hall, Keith. 2015. “Answers to Questions About Dynamic Analysis.” Blog Post, Congressional Budget Office, July 1. <https://www.cbo.gov/publication/50357>.
- Hubbard, Glenn. 2015. “Comment on “‘Dynamic Scoring’: Why and How to Include Macroeconomic Effects in Budget Estimates for Legislative Proposals.”” *Brookings Papers on Economic Activity*, Fall: 134–38.
- Macri, Giuseppe. 2024. “What to Know in Washington: Trump Says Growth to Cover Tax Cuts.” Bloomberg Government, August 26. <https://news.bgov.com/bloomberg-government-news/trump-says-growth-will-pay-for-tax-cuts>.
- National Academies of Sciences, Engineering, and Medicine. 2017. *The Economic and Fiscal Consequences of Immigration*. Washington: National Academies Press.
- Peterson, Matt. 2024. “Kamala Harris Tells a New Story About the Economy. Where It Matters Most.” *Barron’s*, August 23. <https://www.barrons.com/articles/kamala-harris-dnc-speech-economy-taxes-markets-a3916fc8>.

COMMENT BY

DEBORAH LUCAS Dynamic scoring was one of the first issues I was asked to engage in when I came to the Congressional Budget Office (CBO) to work for Dan Crippen as his chief economist. At the time, he strongly opposed the mounting pressures on CBO to adopt the practice, citing many

of the reasons that the authors mention in this paper. For me, observing that debate was an early lesson in the complexities and politics of the budget process, and also a first hint that in the realm of budgetary accounting, my economist's instincts might not always be reliable.

The invitation to discuss this thought-provoking paper has been a welcome opportunity to reengage in the debate. The bottom line is that I believe increasing the use of dynamic scoring would be ill-advised for a variety of reasons. Although the case made in its favor did not persuade me otherwise, the authors rightly raise concerns about the needlessly distorting effects of certain current scoring practices, particularly with regard to labor supply and immigration. However, I will make the case that those distortions and similar ones could be effectively addressed by simpler and less controversial alternatives. For one, CBO and the Joint Committee on Taxation (JCT), with the agreement of the budget committees, could reclassify any direct employment or investment effects as behavioral, and then routinely incorporate those effects into estimates of revenues and spending. As the authors note, there is precedent for CBO doing that.

To explain the reasoning behind my conclusions, I first revisit the role and information content of budget scores—why they are critical to the budget process, and why the information they convey is, by design, very limited. I also mention some of the other scoring conventions that cause budget estimates to be misleading about economic cost, and hence why budget scores are inherently the wrong source for policymakers to rely on for more in-depth policy analysis. I then revisit the definitions and classifications that determine the current line between conventional and dynamic scoring, and propose changes to which variables fall into the “behavioral” versus “macro” bucket that would address some of the current weaknesses. Finally, I will elaborate on both the political and technical reasons for why I think the costs of routinizing dynamic scoring would far outweigh its benefits.

PURPOSE AND LIMITATIONS OF BUDGET SCORES Budget scores serve as a price system for federal policymakers. The scores are the visible hand that helps to guide the allocation of scarce public resources between competing uses. Like all price systems, the one created by budget estimates is imperfect. Nevertheless, the scores serve an important role by helping to inform, and constrain, the policy choices made by Congress.

A primary purpose of scoring is to provide agreed-upon inputs for enforcing caps on total expenditures or deficits. Budget scores also inform policymakers about how specific structural choices would affect a policy's budgetary cost. For example, the budgetary cost of a proposed new federal

insurance program could be reduced to meet a target spending limit in a variety of ways: with higher deductibles, higher premium rates, or tighter eligibility requirements. The bill's sponsors might request that CBO provide scores for the various alternatives, and those scores often influence which measures are included in the final proposed legislation. Informal interactions between CBO and the congressional staff about those estimates encourage more efficient policy choices, both in terms of budgetary cost and the wider economic consequences. Those communications are one of the most important ways that CBO provides information to Congress about the underlying economics of policy proposals.

A budget score is an estimate of the incremental budgetary effect of a legislative proposal, measured relative to a baseline. The budget itself is primarily an inventory of annual cash flows, projected over a ten-year horizon. It provides a measure, albeit an imperfect one, of how much money is expected to go in and out of the Treasury's coffers in the current year, and, were current law to continue, what those flows would be in subsequent years.¹ Select categories of expenditures are accounted for on an accrual basis, based on the logic that a legally binding spending obligation should be recognized in its entirety in the year the commitment is made, thereby making its budgetary cost "grant equivalent." Accrual accounting is mandated for most direct lending and loan guarantee programs and for capital leases. Although economically similar, deposit insurance, pension insurance, and tax-related commitments are accounted for on a cash basis.

Scoring rules and conventions often result in budgetary costs that align with how economists would evaluate annual costs (e.g., for purchases of perishable goods and services and most transfer payments). However, for certain types of activities the rules create a significant wedge between the two, and in some cases the rules require asymmetric treatment of similar activities. In addition to the distortions that the authors suggest dynamic scoring could alleviate, there are other wedges that dynamic scoring would leave largely unaddressed. These include: (1) upward bias in the budgetary cost of most types of investment; (2) asymmetries between cash and accrual; and (3) omission of unfunded mandates.

There are many examples of the upward bias in the budgetary cost relative to economic cost of most types of investment:

- The full cost of purchases of securities like stocks and bonds are treated as expenditures even when they are made at competitive

1. For programs accounted for on an accrual basis, the recorded budget cost can differ significantly from projected cash flows.

market prices. The transactions are economically cost neutral because future cash inflows are of equal value to the purchase price.

- The rent versus buy decision for long-lived durables (e.g., office buildings the government intends to occupy) is usually biased in favor of renting. New purchases require a large up-front expenditure, and there is no offset for the savings from the avoidance of future rental payments. However, there is also a bias that goes in the opposite direction: The cost of renting is overstated when it is equated to the undiscounted sum of projected rent payments over the budget window.
- Similarly, infrastructure investments may be disadvantaged by large up-front payments and revenues that arrive outside the budget window. However, if policymakers interpret the undiscounted sum of projected revenues as profit, the net value of the project may be overestimated.
- Some spending, such as for job training, public education, preventive medicine, research and development (R&D), and so on, could be considered to be an investment that will generate partially offsetting future government receipts. However, only the cash outlays enter into budgetary scores.

Creating a level playing field between programs accounted for on a cash versus an accrual basis is inherently challenging. Current scoring rules and practices fail to achieve that objective, and a number of people, myself included, have advocated for corrective changes.

For federal credit guarantee programs that are accounted for on an accrual basis, the costs reported in the budget are downward biased. For example, the budget score for a government loan guarantee is lower than what it would be for an identical guarantee purchased by the government from a private sector financial institution. The underlying reason is that the noncash cost of risk is omitted, and also that the scoring rules require administrative costs to be accounted for separately.

Some tax expenditures, which like loans and loan guarantees are contractually binding, are nevertheless scored on a cash rather than an accrual basis. For example, the reduction in revenue from the exemption of interest on state and local bonds from federal taxation is scored on a cash basis, whereas the economically equivalent cost of a concessional interest rate on a direct government loan is evaluated on an accrual basis.

A third notable distortion arises from the exclusion of unfunded mandates. When a government policy forces a nonfederal entity to take a costly and

uncompensated action, it creates an unfunded mandate. Economically, the mandate is equivalent to the government levying a tax and spending the proceeds. By law, unfunded mandates above a threshold must be reported by CBO annually. However, they are not formally scored nor do they appear in the budget.

The almost exclusive focus on cash flows in budgetary accounting is the root cause of the disconnect from economic cost in most of these examples.² As anyone familiar with private sector accounting principles knows, a statement of cash flows in itself provides very little information about whether an entity is creating or destroying value. A more complete picture emerges when the information on a balance sheet and income statement are brought into the picture.³ The US government produces balance sheets and income statements, for the government as a whole and for many individual agencies, but that information does not enter into budget deliberations.

The inadequacy of cash accounting for conveying information about economic cost is particularly glaring for capital expenditures, as the above examples illustrate. It is important to understand that the choice by Congress to nevertheless continue using cash basis accounting for investments is deliberate. Proposals to reduce the bias, such as incorporating a capital budget that better captures the economic effects of capital expenditures, have been rejected on multiple occasions.

Why would policymakers favor biased cost measures? All else equal, most undoubtedly would prefer to receive unbiased estimates. However, scoring rules were established with multiple objectives in mind. As some battle-hardened budgeteers emphasized when I would complain to them about some of these distortions, the overriding advantage of cash accounting is the far greater difficulty of gaming it. With regard to how investment activities are accounted for, some worry that a move away from cash basis accounting would enable excessive government investment that would crowd out more productive private activity. There is also the concern that advocates would repackage spending as investment, particularly in areas like education, job training, health care, R&D, and infrastructure programs. Returns on those activities would be hard to assess and disagreements not easily resolved by expert opinion.

2. To decide for yourself whether budget scores are useful for informing policy choices beyond revealing their budgetary impact, I suggest that you try reading a few budget estimates posted on CBO's website.

3. For firms, a comprehensive understanding of value creation only emerges when equity price changes are also taken into account, but clearly that information is unavailable for the government.

Notably, private sector accounting rules also in some respects have a conservative bias, presumably to prevent gaming. For instance, banks cannot freely increase the book value of loans that were previously marked down, even if objectively their value can be shown to have increased. Expected future profits are not capitalized on balance sheets.

Of course, the fact that there are other scoring rules and practices that cause budget costs to deviate from economic costs is not a reason in itself to oppose making changes to address agreed-upon problems. Rather, I emphasize these other distortions, and the noneconomic goals that have caused them to persist, for two reasons: (1) It supports the contention that the differences between budget scores and economic costs are fundamental and pervasive, making budget scores inherently of limited value for in-depth policy evaluation. CBO has more effective ways to communicate to policymakers about the economics of proposed legislation, including via reports prepared by the program divisions and through consultations between budget analysts and the congressional staff. Policymakers can and do draw on many other sources of information as well. (2) Maintaining the acceptance and credibility of the budget process and protecting against actual or perceived manipulation are legitimate objectives that have shaped the budget process. Changes to rules and practices aimed at reducing the disconnect between budgetary and economic cost must also be evaluated with these concerns in mind.

Perhaps the most important reason that budget scores have limited value for policy evaluation, and why that will always be true, is that they do not attempt to quantify or even enumerate a policy's benefits. Benefits are typically much harder to agree upon and to quantify, and current scoring rules largely save CBO from having to take a position on what they are or how big they might be. That helps protect the agency's credibility and appearance of neutrality. As budgeteers like to say, federal costs are determined by the budget process, benefits by the political process.

In subsequent conversations with the authors, they have emphasized that dynamic scoring maintains the focus on cost and does not cause CBO to quantify benefits. I don't think that is strictly true. In order to incorporate some of the effects that the authors would classify as dynamic, effectively CBO would have to take a stand on a policy's more far-reaching consequences, including those that would be considered to be its benefits. This is particularly a concern for productivity, where the effects (e.g., of an infrastructure bill) would remain highly controversial even if expert opinion could be used to justify the assumptions that are made. To implement the proposed vision of dynamic scoring, the budget agencies would unavoidably

and much more regularly have to take implicit or explicit positions on the size and consequences of various benefits, which I fear could damage the credibility of CBO and the budget process.

REVISITING DEFINITIONS, CLASSIFICATIONS, AND ALTERNATIVE SCORING RULES
Dynamic scoring differs from conventional scoring by incorporating feedback effects to the budget from a policy's effects on the macroeconomy or on variables classified as macroeconomic. It is "dynamic" because it relaxes the maintained assumption in conventional scoring that macroeconomic variables are static. In order to calculate a dynamic score, in addition to what analysts do already, they would have to estimate the macroeconomic consequences of a legislative proposal and then estimate how those macroeconomic changes will affect the budget.

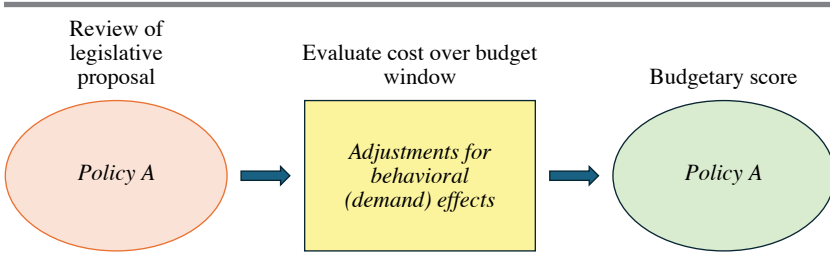
When I first came to CBO, the main proponents of dynamic scoring were Republicans who believed that the budgetary cost of tax cuts were overstated because the scores neglected the positive effects on economic growth and hence on future revenues. This is one variation of the familiar story that all policy advocates tell when arguing for including indirect cost savings in other parts of the budget.

An important distinction, and one that comes up frequently in the paper, is between "behavioral" and "macro" effects. Some behavioral effects are routinely incorporated into budget scores, whereas macro effects are excluded except on rare occasions when a proposal is large enough to trigger a dynamic score. I will suggest that relaxing the strict dividing line between the two categories could address some of the authors' concerns without taking the more problematic step of expanding the use of dynamic scoring.

With regard to behavioral effects, the ones most frequently taken into account are demand or supply changes that directly affect the cost of the policy itself. For example, for a bill that would increase the subsidies to health insurance premiums in a mandatory program, scorekeepers would take into account the increased demand for that insurance. Indirect behavioral effects that spill over into demand for other programs are included less often. Continuing with the example, people with health insurance are likely to see a doctor more often when they fall ill, and arguably that reduces their usage of other expensive government health services. However, the proposed insurance subsidy program may receive no credit for savings from that type of spillover.

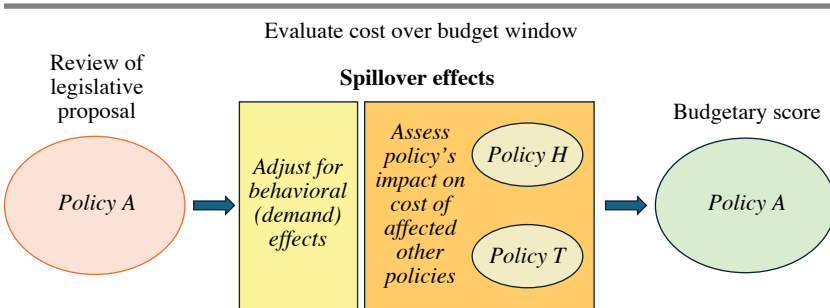
The health insurance example suggests another distinction that is useful for understanding alternative scoring rules: the difference between "direct" and "spillover" (or indirect) effects. Figure 1 shows a schematic of the scoring

Figure 1. The Scoring Process with Only Direct Behavioral Effects



Source: Author's illustration.

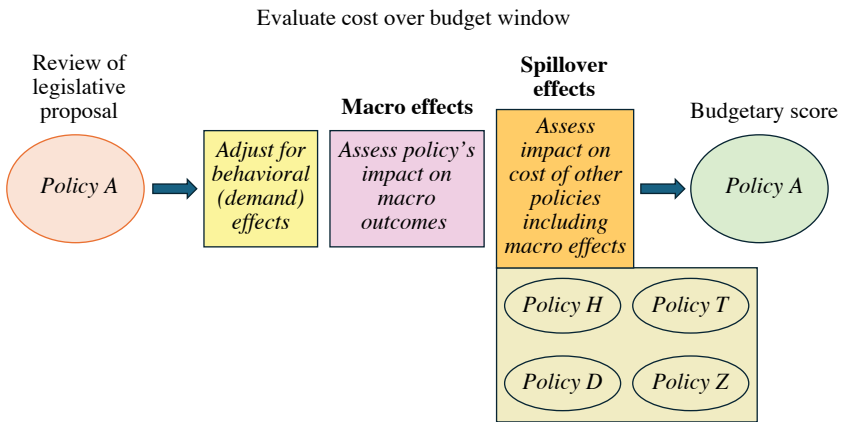
Figure 2. The Scoring Process with Direct and Spillover Effects



Source: Author's illustration.

process when applied to a policy that involves only direct behavioral effects. Such scores have the advantage of simplicity; they are relatively easy for budget analysts to produce and explain and for policymakers to understand.

Some types of policy proposals have significant spillover effects. The schematic in figure 2 illustrates the scoring process for a policy where spillovers significantly affect the costs of two existing policies. For example, most of the budgetary effects from immigration policies come from estimated spillovers to other programs. A policy that permits larger numbers of legal immigrants would cause an increase in demand for various social assistance programs, but beyond administrative expenses would involve few direct expenditures. Incorporating spillover effects tends to increase the accuracy of budget scores. The downside is that it increases complexity, making estimates harder to produce, audit, and explain. As noted earlier, estimating spillover effects also may require controversial judgments about the size and consequences of a policy's benefits.

Figure 3. The Dynamic Scoring Process

Source: Author's illustration.

Consistent with traditional scoring policies and practices that are aimed at preventing gaming and that promote credibility, it seems wise to limit the inclusion of spillover effects to those that meet a fairly high materiality standard and to err on the side of reporting a smaller or zero effect when the surrounding uncertainty is large. This already seems to be the de facto practice, but I am not aware that it has been codified, or that a conscious effort is made to apply that principle consistently across program areas.

Dynamic scoring expands spillover effects to include those arising by way of a macro channel. Adopting it would in some cases improve accuracy and remove biases. I return to the immigration example below, which is one such case. However, relative to only considering material direct spillovers, it adds considerable complexity and the costs that would entail. Figure 3 illustrates the dynamic scoring process.

A MORE MODEST PROPOSAL Could some of the shortcomings of conventional scoring that the authors point out be alleviated more simply and less controversially than by adopting dynamic scoring? I think the answer is yes, and that much could be accomplished just by shortening the list of variables constrained to be static in the scoring process. The proposal floated here is closely related to, but more general than, the “population change” approach mentioned in the paper. It might be described as a “direct effect” approach.

The paper indicates that the macro variables treated as static by the budget agencies include labor supply, savings, investment, productivity, output, inflation, and interest rates. To implement a direct effect approach,

a distinction would be made between the subset of these variables that have microeconomic as well as macroeconomic effects (labor supply, savings, and investment) and those that are fundamentally aggregate or equilibrium outcomes (GDP, total factor productivity, inflation, and interest rates).⁴ The suggestion is to reclassify variables in the first group as behavioral for the purposes of estimating budget scores. The change could be justified by the fact that those variables are directly determined by behavioral choices, similar to the demand effects that are routinely included. Although changes to labor supply, savings, and investment at the micro level affect the aggregate variables that would still be held static, for the vast majority of policy proposals, the aggregate changes would be small. Smaller still would be any predicted feedback to the rest of the budget. Hence, broadening the set of variables treated as behavioral could improve the accuracy of cost estimates with only minor changes to the conventional scoring process.

To give a concrete example, consider again a bill that increases the number of legal immigrants. As the authors note, under current scoring rules its budgetary cost would be biased upward. The score would include the spillover effects of higher demand for expenditure programs. However, because changes to labor supply are not currently recognized, it would exclude spillover effects to revenues arising from immigrants' labor and investment income. Adopting the proposed direct effect approach would recognize the direct effects on labor supply and capital income, thereby restoring symmetry between the treatment of expenditures and revenues.

With regard to total factor productivity, the suggested direct effect approach would continue to treat it as static. For that reason, it would not satisfy the authors' preference for dynamically scoring policies like infrastructure investment, R&D, or liberalized investment permitting, which they believe would produce budgetary savings from the higher economic growth the policies would generate. On this point we will have to agree to disagree. For a variety of reasons mentioned in the next section, I remain unconvinced that those dynamic effects should be included. However, like labor supply, productivity has a microeconomic dimension that would be included under a direct effect approach, for instance, because productivity influences projected wage earnings.

A final observation on scoring rules pertains to the treatment of interest payments on the government debt. Currently debt is treated as an off-budget "means of financing," consistent with the observation that no new program

4. Other variables that are not on this list but are also constrained to be static also could be divided between the two groups.

directly causes additional debt issuance; that is a financing choice. The current treatment is effectively consistent with the on-budget accrual approach applied to loans made by the government. Economically the current treatment of interest is also the correct one. When the Treasury issues debt, the principal it borrows is equal in present value to the future interest and principal repayments made—it is a financially neutral transaction that has no cost beyond administrative expenses. The authors contemplate the possibility that, with a wider adoption of dynamic scoring, programs might be charged for the pecuniary externality of higher interest rates caused by additional debt issuance. As a practical matter, it seems clear from the very large body of academic literature on interest rate determination and sovereign debt that too little is understood about the relation between the size of the public debt and Treasury interest rates to credibly adopt a scoring rule that charges policies for causing interest rates to be higher than they otherwise would be.

FIRST DO NO HARM I want to elaborate on some of the political and technical reasons behind the conclusion that the costs of routinizing dynamic scoring would outweigh the benefits. To start with the political, having lawmakers with widely divergent world and economic views agree to use a common set of numbers in the federal budget is invaluable for transparency, discourse, and discipline. It is of paramount importance for maintaining an orderly, or at least less disorderly, budget process. However, that agreement is fragile, and if it breaks, it will be hard to restore. A crucial part of the job of the budget agencies and scoring rules is to protect it. Hence, an important consideration in evaluating any rule change is whether it might weaken the bipartisan acceptance of official budget scores.

Estimation procedures that require little or no judgment on the part of agency staff are more likely to be trusted as free from manipulation or biased perceptions. The fact that most budget scores do not require judgments about the implications of a policy beyond its narrow effect on government cash flows has helped to preserve acceptance of the process. Scoring is particularly straightforward when an appropriation and the amount spent are likely to be approximately equal, and when there are no significant spillover effects.

By contrast, routinizing dynamic scoring would require CBO to regularly make implicit or explicit assumptions about the uncertain macroeconomic consequences of a much larger number of policies. Even the decision about whether a policy does or does not have significant macroeconomic effects can have a political element, and in some cases is likely to be highly contentious. Take infrastructure spending. Some argue it will increase aggregate productivity and therefore government revenues, even paying for itself.

Others believe it is largely pork, money thrown into the ocean or into politicians' pockets. Because infrastructure legislation often delegates authority to agencies to choose and execute specific projects, the implications for productivity, either locally or in aggregate, are very hard to predict. Or consider job training, or health care, or education, or childcare, where at best the evidence is very mixed. For the most part, CBO hasn't needed to step into these debates.

I think the authors have in mind that the agencies could pick and choose when to incorporate macroeconomic effects, and that they would do so only if the evidence for inclusion is compelling. I was therefore surprised to see several examples presented where dynamic scores were significantly lower than conventional ones because of the productivity-enhancing effects attributed to infrastructure investment and R&D. The large differences from conventional scores are interpreted as demonstrating the merits of dynamic scoring. However, given the very mixed evidence and polarized prior beliefs about productivity effects, they might instead be seen as evidence for why the practice should be avoided.

Relatedly, the authors also seem to suggest that the trust CBO has established with policymakers—which has provided the agency with some latitude in how it constructs the baseline and in the methodologies that are used in program division analyses—would be likely to carry over to its models and assumptions used to produce dynamic scores. It is an open question whether the lack of pushback would carry over to the much higher stakes activity of budget scoring, and I would not count on it.

Returning one last time to the immigration example, I have suggested that a direct effect approach would eliminate the first-order distortion of underestimated revenue effects, with only a minor change in interpretation of the existing rules. A dynamic score would also eliminate that distortion, but it would require many additional assumptions that could make the estimate much more controversial. For instance, imagine that CBO's model predicts that admitting more highly skilled immigrants would increase total factor productivity and spur job creation that is in excess of direct labor supply effects. Feedback to revenues would then lower the score relative to what would be reported using a conventional or direct effect approach. However, even if reputable academic research supported CBO's modeling choices, opponents of the policy might claim that the jobs lost from immigrants displacing native workers were underestimated, or that wages and hence revenues would be lower because of the increased labor market competition. The considerable uncertainty about these highly politicized issues leaves the door open to controversy, or possibly even to a rejection

of the estimate by lawmakers. Although CBO regularly has to respond to criticism even for its simpler scores, more complex analyses tend to be harder and more costly of agency time to defend. The gains in this case from a more complex dynamic analysis are also unclear. Immigration has significant long-term budgetary consequences and also affects local government finances. Distortions from zeroing out all effects beyond the ten-year budget window, ignoring time value, and excluding state and local costs make it impossible to use the score of an immigration bill, dynamic or not, for serious cost-benefit analysis.

Turning to technical considerations, the authors express confidence that CBO's and the profession's models, and the available evidence to calibrate them, have improved to the point where they are reliable enough for dynamic scores to be credible at least for some applications. They are also optimistic about the prospects for further improvements, and that providing dynamic scores will become increasingly affordable for CBO over time. I am less sanguine on all of these fronts. I suspect that there will continue to be sharp and justifiable disagreements about the models and assumptions used to produce dynamic scores, and that the staff time it would take to consistently produce credible dynamic scores would continue to be considerable.

It is unclear that macroeconomic models, or macroeconomists, are suited to the task of producing credible dynamic scores. That is particularly true on the expenditure side. The ambition of macroeconomic models is to capture the relation between macro variables. Providing a mapping from microeconomic expenditure policies into macroeconomic outcomes is something few models are designed to do, except perhaps in a generic way that would abstract from any material differences across policies. It is suggested that CBO's experience estimating behavioral effects could help, but the informational requirements for estimating micro and macro spillover effects seem quite different. CBO could try to build and vet new models. However, there is little established precedent to draw on, and attempting to build high-quality capacity would be time-consuming. Whether dynamic scoring models could be reused without significant customization and still be credible is doubtful. CBO has relatively few macroeconomists (fifteen are listed in that division on CBO's website), and most of their time is taken up by the baseline and other responsibilities.

Another big question is about the completeness of the list of macroeconomic variables that dynamic scoring is envisioned by the authors to take into account. Notably absent from most of the discussion in the paper are the effects of higher debt and deficits. This clearly favors debt-financed tax cuts. I would argue that excluding the potentially large negative effects

of high debt ratios on economic growth has the effect of biasing up dynamic scores relative to conventional scores.⁵ Although in principle it could go in the opposite direction, notice that all of the examples provided show dynamic scoring lowering budget costs.

The proposed exclusion of debt and deficits as variables affecting dynamic scores probably reflects the difficulty of predicting their effects, and also the common practice among macroeconomists of assuming the effects away, except perhaps to predict slightly higher real interest rates. Relatedly, macro models generally assume that the economy is in a steady state, which rules out unsustainable debt ratios and the possibility of severely adverse effects on the economy.

Financial markets and academic financial economists (the tribe I belong to), along with some politicians and members of the public, are more likely to be concerned about the possibility of severely adverse effects from high debt ratios. High debt levels increase the likelihood of major disruptions in financial markets, fiscal dominance, high inflation, and higher distortionary taxes. While the size and frequency of these effects are hard to predict, the sign on economic growth is clear. These policy-related consequences are also a reminder that if a “current policy” rule is imposed on the models used to implement dynamic scoring, the distance between budget scores and economic cost might be increased rather than reduced.

Although I’ve focused on areas of disagreement, it is important to emphasize that the paper makes a valuable contribution that will be especially helpful to readers interested in learning more about the seemingly arcane world of the federal budget process and its implications for fiscal policy. This is an important topic that unfortunately is overlooked by the mainstream economics literature, and this paper helps fill that gap. Personally, it was very useful to be forced to rethink, and to be more precise in my thinking about, how the scoring process works, why it has taken its current form, and which rules remain especially problematic. The current budget process was largely codified in the 1974 Budget Act, and there have been few major changes made to it in the fifty years since then. How various scoring biases could be reduced without opening the door to bigger problems remains an open and important question, and there is a growing awareness of the need for policymakers, the budget agencies, and academics to devote more effort to trying to address it.

5. This may be why there is an unusual coalition in favor of dynamic scoring between conservative Republicans who favor tax cuts and liberal Democrats who favor greater government investment activity.

GENERAL DISCUSSION Phillip Swagel stated that members of Congress continue to want dynamic analyses and invite participants to share feedback with him on how to implement dynamic scoring. He noted that the Congressional Budget Office (CBO) is thinking hard about how to expand dynamic analyses to areas such as immigration, health care, labor policy, energy policy, and even regulation. He also thanked the authors for their paper and for highlighting CBO's work on dynamic effects of infrastructure investments, focusing on new roads and bridges that were a part of the Infrastructure Investment and Jobs Act (IIJA). He cautioned that, even after a substantive evidence base has been produced, predictions about implementation can prove difficult in many areas. Using rural broadband as an example, he noted that the executive branch is not using the standard approach in spending IIJA dollars dedicated for improving rural broadband access—a choice that CBO could not have anticipated and that would have confounded the accuracy of a dynamic analysis.

Swagel also tempered expectations about the political impact of dynamic scoring, arguing that legislators tend to have a good understanding of the economic effects of, for example, immigration. Despite a powerful result in the paper showing a dynamic analysis of high-skilled immigration, he observed that immigration policy is difficult politically; therefore, the limiting factor on whether a specific immigration proposal passes is not whether CBO publishes a dynamic score of said proposal.

Henry Aaron first addressed Deborah Lucas's discussion comment on spillover effects, offering the example of enacting a payroll tax on fringe benefits to help close Social Security's deficit. He contended that while it would be straightforward to estimate revenues if behavior were unchanged, it would be much harder—if not impossible—to estimate behavioral responses, including in employers' use of Medicare and Medicaid, which could well generate larger revenue effects. Aaron then brought up patent law, which like immigration policy, illustrates how non-budgetary legislation can have large impacts on revenues and expenditures.

Regarding the inclusion of spillover effects, Alan Auerbach pointed out that traditional scoring already does this, for example, by accounting for how a change in the standard deduction will affect the use of other deductions. Auerbach continued that effective dynamic scoring efforts will have to estimate fiscal policy responses to deficit-increasing legislation and monetary policy responses to inflation-increasing responses. Likewise, many analyses would require estimating responses, such as potential retaliation, from other governments, as in the case of a hypothetical 20 percent tariff on all imports. He concluded that requiring CBO to effectively tell legislators what it thinks

they will do in the future may be difficult and would certainly represent a marked departure from traditional approaches to scoring.

Brian Sack also wondered about incorporating responses from financial markets. He believed that CBO has not generally incorporated changes to risk premia, especially on Treasury securities whose supply may change substantively following large pieces of legislation. He argued that while it is not clear how risk and term premia change with supply considerations, the ten-year Treasuries used to trade with yields lower than swaps two decades ago, whereas currently, yields are nearly 50 basis points above swaps. Therefore, while interest expenses are large, it may be particularly important for dynamic scores to estimate these financial market responses.

Jeffrey Kling responded that CBO always tries to do a full counterfactual analysis relative to baseline, which would require incorporating both Federal Reserve and financial market responses. He also commented that, regardless of what legislators choose to do with the information CBO provides, it is CBO's goal to constantly provide better, more comprehensive, and more accurate information.

Glenn Hubbard echoed Kling, commenting that static scores already make implicit assumptions about government policy. In terms of the case studies, Hubbard suggested that there will need to be a process for figuring out the most important areas to focus on. He suggested this would come from a combination of members of Congress, CBO, and the Joint Committee on Taxation (JCT), as well as outside academic experts who can inform CBO and JCT of areas where there have been large increases in the knowledge base, such as immigration or permitting reform.

Wendy Edelberg first addressed the CBO baseline, remarking that CBO's forecast of a functioning economy under current law (and indeed any economic forecast over a period with rising debt) must implicitly include assumptions on the part of households and businesses about a closure rule off in the future. She then clarified the meaning of the word "benefits" in the context of dynamic scoring, drawing a distinction between positive feedback to the budget that dynamic scoring would consider and broader increases to welfare that would still be omitted. She offered the example of a large cut to Social Security, which might increase labor force participation or savings rates under a dynamic score, thereby generating positive feedback effects for the budget, despite potentially large overall negative effects on social welfare. Lastly, she questioned how genuine lawmakers' interest in dynamic scoring is, relaying that, in her experience, lawmakers preferred their own talking points to estimates from CBO that might differ on some dimension. On Edelberg's last point, George Akerlof expressed worry about

the extent to which dynamic scoring may be too apt to be gamed, if politicians have discretion over when or how it should be done.

Jason Furman suggested that CBO produce a dynamic score for a bill that would increase CBO's budget by \$5 million annually, allowing it to hire thirty more analysts dedicated to producing dynamic scores. He contended that this would be a rare example of spending that would fully pay for itself. He also lamented that, under the status quo, everyone acknowledges that dynamic effects are present but due to uncertainty about their magnitude, scorekeepers instead assume they are zero. He argued that there must be a better approach for the agency.

Douglas Elmendorf complimented the work of the CBO and JCT analysts, adding that some recent dynamic analyses that they produced persuaded him to reexamine the topic. He emphasized that many of the arguments he heard against dynamic scoring while at CBO—that it is uncertain, susceptible to gaming, and requires predicting what policy actors will do—seemed to also apply to both conventional scoring and CBO's baseline estimates. Therefore, he suggested, discussions about dynamic scoring should separate out what is particular to dynamic scoring rather than inherent to scoring efforts across the board.

Ben Harris inquired how scoring works—or should work—in the face of uncertainty. For example, the American Rescue Plan Act was passed before the subsequent Delta and Omicron COVID-19 waves, making the eventual utility of the bill nebulous at the time of the passage. He offered an additional hypothetical of a bill designed to avoid an asteroid impact that has a 1 percent chance of occurring—how would CBO score this? He suggested that current scoring practices might be biased against insurance-like programs that address unlikely but catastrophic outcomes.

William Gale proposed that, in addition to considering the three questions of why, when, and how, the paper might also consider the question of over what time period, given that some policies see quick dynamic effects while others, such as investment in children, take much longer to occur.

Jonathan Pingle added to previous comments about the uncertainty of dynamic scoring by noting that even static estimates for both the Inflation Reduction Act and the IIJA have been repeatedly revised. Therefore, he suggested focusing dynamic estimates on areas such as immigration where the magnitude of switching to dynamic scoring would not be overwhelmed by the inherent uncertainty even in static scoring.

Marc Goldwein raised the practical concern that legislative processes often end in a rapid amendment process, which is already difficult for CBO to keep pace with when producing static updates. He worried that shifting

to dynamic scoring for initial estimates would force policymakers to either evaluate amendments without any additional information or to compare the initial dynamic estimate with subsequent static scores of the amendments.

Mark Mazur referred to Joel Slemrod's hierarchy of behavioral responses to tax changes, which argues that timing responses come first, followed by avoidance responses through financial transactions, with real economic behavior coming last.¹ Mazur argued that dynamic scoring places substantial emphasis on real effects, giving them outsized priority. This effect may bias lawmakers toward tax cuts, given that, Mazur noted, the estimates for the macroeconomic effects of the Tax Cuts and Jobs Act have been revised down multiple times. This bias may be of particular concern for pursuing balanced solutions to the US budget deficits. On the hierarchy of responses, Elmendorf contended that dynamic scoring does not necessarily place real responses at the top but instead incorporates these responses, which would otherwise be left out entirely.

Koichiro Ito inquired if various elasticity estimates from the applied microeconomic literature might be helpful in estimating large macroeconomic responses for dynamic scoring.

Susan Athey spoke about her experience overseeing a few dozen economists evaluating several thousand mergers each year, arguing that a small team can complete many complex analyses quickly when that team has gained experience and developed heuristics. While this approach is not foolproof, she affirmed that her team was able to build its muscle through case studies and by choosing areas of policy focus to develop the necessary expertise. Athey welcomed help by outside contributors in writing case studies such as the one in the paper, noting that, over time, this can lead to marked improvements in the areas where scores are likely most wrong today. By taking this approach, a relatively small team of economists could regularly produce meaningfully better scoring estimates, even on a short time frame.

Elmendorf agreed with Athey that much of the discussion around dynamic scoring boils down to how much certain elements can be standardized and simplified to create dynamic estimates more quickly. As an example, for questions related to the federal budget, CBO already publishes rules of thumb, which roughly show how certain changes to GDP, inflation, and other economic factors flow through to budget outcomes. They even publish this in spreadsheet form, making it easy for anyone to adjust numbers on

1. Joel Slemrod, "A General Model of the Behavioral Response to Taxation," working paper 6582 (Cambridge, Mass.: National Bureau of Economic Research, 1998).

the go and see immediate results.² Ultimately, he argued, the question is whether this sort of simplified—but credible—approach can be expanded to other areas to make dynamic analyses more feasible.

Hubbard recalled his first exposure to dynamic scoring: On his first day working in the Office of Tax Analysis in 1991, he received a letter from Rep. Richard Armey criticizing the office for not doing dynamic scoring and concerned about his pedigree. Ever since then, he said, he has continued to believe that educating members of Congress is important and that it is the job of scoring to do so. Furthermore, scoring is inherently uncertain, so uncertainty itself is not a reason to avoid dynamic estimates. On the other hand, Hubbard identified a key shift since 1991: CBO and JCT are increasingly soliciting outside experts' advice to help develop heuristics and rules of thumb. He clarified that he and his coauthors do not necessarily believe that dynamic scoring is always appropriate, but that developing this ability in some capacity is absolutely worthwhile.

2. CBO, "Workbook for How Changes in Economic Conditions Might Affect the Federal Budget: 2023 to 2033" (Washington: Congressional Budget Office, 2023), <https://www.cbo.gov/publication/59027>.