Ex Ante Policy Evaluation: A Unified Approach

John A. Graves

*Vanderbilt University*

August 2019

Evaluation of policy change is the lifeblood of applied economic research. Often, evaluation is retrospective: a policy changes, data are collected, and researchers evaluate impact on welfare-relevant outcomes. This exercise is frequently conducted in service of a greater goal—namely, to make predictions based on theory about economic behavior, and to use policy change as a lens through which these predictions are tested. Knowledge produced through this exercise is then (ideally, but not always) applied to future policy to improve efficiency and human welfare.

The primary focus here is on a related but inverse exercise: ex antepolicy evaluation. Under this approach, empirical estimates serve as inputs *to* rather than outputs *from* the policy evaluation process. Typically, a model of economic behavior is conceptualized around an existing theoretic scaffold (e.g., utility maximization in a model of discrete choice). This framework is then coupled with data on individuals, firms, and other decision-making entities. Combined, these elements constitute a simulation model of human behavior or some other policy-relevant decision-making process (e.g., the progression of a disease or the adoption of a new technology). The model is further adorned with structural and behavioral parameters from existing evidence or, in cases where the empirical literature has not yet rendered reliable estimates, from judgment calls consistent with the underlying theory. Policy reforms are propagated through the model by simulating responses to changes in price, choice sets, health technologies, tax schedules, etc. Modelers then aggregate an array of policy- and welfare-relevant outputs (e.g., changes in demand for goods and services, quality-adjusted life expectancy, changes in health insurance coverage and premiums, changes in federal tax revenues and outlays, etc.). These outputs are (again, ideally, but not always) disseminated to inform policymakers’ decisions on the design and/or adoption of a particular policy.

In the United States, the Congressional Budget Office (CBO) is the most important practitioner of ex ante evaluation. It would be difficult to overstate the CBO’s importance in shaping the trajectory of federal policymaking over the last generation. For example, one Senator famously characterized the recent history of U.S. health reform as Congress “sending health legislation off to the Congressional Budget Office to die.”[[1]](#footnote-1)

Ex ante policy evaluation also plays an important role in determining the availability, pricing and reimbursement for goods and services worldwide. Entities such as the National Institute for Health Care Excellence (NICE) in the United Kingdom commission simulation analyses of the cost and quality-adjusted life year impact of new health technologies. Results from these simulation studies are used to determine coverage policy within the National Health Service. Microsimulation models have also been developed by government and non-government agencies to project the consequences of changes to tax and transfer policy, education policy, and food policy, among others (TK CITES).

Despite the important role played by simulation modeling in the policymaking process, scant formal attention has been paid to the theory, design and integration of ex ante evaluation within the broader economic research enterprise. The one exception is health technology assessment, where rigorous standards for conduct, methods and reporting have been put forth by the Panel on Cost-Effectiveness in Health and Medicine.

To the extent there are explicit linkages between ex post and ex ante evaluation, they are usually relegated to simple, back-of-the-envelope counterfactual policy evaluations that appear at the tail end of empirical and theoretic research manuscripts. Rarely do these exercises draw on a formal approach to comparative welfare analysis (Hendren 2019). Even more rarely do they grapple rigorously with the role of estimation and structural model uncertainty in guiding both policy decision-making and the direction of future research (Hendren 2019).

These shortcomings extend to ex ante policy modeling as well. For example, microsimulation models often produce an array of welfare-relevant outputs and leave it to policymakers to weigh these factors when making decisions (Finkelstein, Hendren and Shepard 2018; Finkelstein, Hendren and Luttmer 2019). This is particularly true in U.S. health policy, where federal policy decisions based on costs and cost-effectiveness are prohibited through both legislation and administrative rulemaking. But even absent a specific legislative or regulatory decree, the CBO and other modelers have generally shied away from producing comparative (overall) welfare assessments.

Compounding these shortcomings are three related challenges. First, while simulation models often draw on standard economic theory and a shared empirical evidence base, the underlying evidence is estimated with uncertainty (and possibly with bias) and is not always in uniform agreement. Models also differ in their structure, underlying data sources and assumptions. It should come as no surprise that models produce varying projections of the same reform proposal.

Second, despite [recent efforts](https://www.cbo.gov/publication/55116) at greater transparency, the opacity of microsimulation models makes it difficult for researchers to know whether and how their work can inform modeling efforts. Finally, the development, execution, and maintenance costs of microsimulation models are considerable. Combined, these factors contribute to high barriers to conducting rigorous ex ante policy evaluation and a muddled sense of how the economic research enterprise could be further refined to improve both modeling efforts and policy decision making.

This study outlines an approach to ex ante policy evaluation that addresses many of the above shortcomings. The first contribution—which is related to recent and ongoing work on *ex post* policy evaluation by Hendren and Sprung-Keyser (2019)—is the linkage and development of theories related to the Value of Information (VOI) and the Marginal Value of Public Funds (MVPF) for ex ante policy modeling. Intuitively, VOI quantifies the opportunity cost of policy decision making under uncertainty. At a given policy efficiency threshold (e.g., a MVPF value of 0.8, above which a policy might be desirable but below which it may not), modeling uncertainty may or may not affect “optimal” policy choices (i.e., choices that maximize relative comparisons of benefits to costs). If decisions based on comparative assessments of MVPF are insensitive to varying parameter values, then the value of uncertain information is low—i.e., it is not worth additional research effort to reduce parameter uncertainty since the same decision would be made today as it would if we had better information. If decisions are sensitive to this uncertainty, however, then VOI methods quantify the opportunity cost of making policy decisions based on *current* information versus if we had perfect information on uncertain parameters. Variation in modeled outputs can be further decomposed to identify the relative degree to which specific parameters contribute to the overall value of perfect information. These assessments, in turn, can provide guideposts for refining and prioritizing future research to focus on domains where the value of information is high.

To showcase these methods, I turn to a specific application germane to ongoing debates in U.S. health policy: whether or not to further expand health insurance coverage and if so, through what means (expansions of public or private insurance plans). To do this, I develop a generalized ex ante modeling approach based on a discrete time and choice framework. I demonstrate that this overall modeling framework can encompass many existing approaches to health policy microsimulation, including elasticity-based and utility maximization-based models.[[2]](#footnote-2) Critically, however, the approach also facilitates simple yet powerful counterfactual policy assessments based on reduced form estimates. That is, the framework provides researchers with a simple tool to investigate the coverage and cost impacts of reform alternatives without the need for a detailed individual-level microsimulation model.

As a proof of concept, I demonstrate how difference-in-differences evidence on the impact of Medicaid expansion on coverage take-up (Graves, McWilliams and Hatfield 2019), combined with estimates on take-up of subsidized private health insurance derived from regression-discontinuity estimates (Finkelstein, Hendren and Shepard 2019) can be harnessed to model the coverage and cost impact of further expansion of coverage via public programs versus via increased subsidies for private coverage. I then embed calculations of the MVPF within this framework and conduct probabilistic sensitivity analyses to assess how estimates of benefits and costs vary when allowing *all* model and MVPF parameters to vary. The parameter values and outcome results of this probabilistic sensitivity exercise are primary data inputs into a “metamodel”—that is, a statistical regression model that isolates the degree to which outcomes change as individual model parameters vary. This metamodel is then used to produce VOI estimates that isolate and rank-order model parameters in terms of their importance to the overall degree of uncertainty in the simulation model.

The first major contribution is a This framework has roots in modeling methods commonly used for health technology assessment, and in the "sufficient statistics" approach to welfare evaluation developed in the public finance literature. I focus in particular on a model for health reform policies that affect insurance (and uninsurance)..

Second, within this framework I tie together diverse approaches to assessing uncertainty and the welfare impacts of policy. Specifically, I draw linkages between the marginal value of public funds (MVPFs), a summary measure of the costs and benefits of public policies (Hendren 2017), and value of information (VOI) methods for ex ante modeling.[[3]](#footnote-3)

The remainder of this paper proceeds as follows…

*Comparative Policy Assessment*

The framework outlined above provides a generalized approach for modeling policy change. Critically, it is structured in such a way as to allow for both “structural” parameterization (i.e., explicitly specifying utility functions or elasticity equations that inform choice probabilities, as in most existing microsimulation models) and “direct” parameterization via reduced form estimates.

But this only solves part of the problem. Often, policymakers want to make comparisons *between* different policies. These comparisons are all the more difficult when models output an array of welfare-relevant outcomes, rather than a single, scalar-valued summary estimate (Hendren and Sprung-Keyser 2019).

Even if we could filter model output into a single value, policymaking often boils down to a single choice to implement a policy or not. What are the opportunity costs of making the wrong choice given the available evidence? Models inevitably rely on a set of assumptions and rely on parameters that, even if well-identified, are estimated with some degree of imprecision uncertainty. Making policy decisions based on a single, scalar-valued outcome raises the question of whether the same policy decision would be made under a different and equally plausible set of modeling assumptions, or using a different set of parameter values that is also consistent (statistically) with the available data. This raises interesting and important questions for both modelers and policymakers. Is modeling output “good enough” with the available evidence, or should we invest in further research to reduce uncertainty? When it comes to sensitivity to model parameters, which have the most impact?

Moreover, decisions over competing policies often boil down to a single selection. Models, moreover, rely on assumptions and parameters estimated with uncertainty. This uncertainty could derive from any number of sources: estimation uncertainty, sampling uncertainty, modeling uncertainty. Even if a model could output a single welfare-valued output, it will have uncertainty around it. How should we think about this uncertainty and the opportunity costs associated with making the wrong decision. These are considerations we tackle in this section.

Fortunately, recent advances in the public finance and the health economic literatures offer a way to crack this nut. Intuitively, we need a single welfare-relevant output measure. Then, we need to define a threshold to make normative conclusions on whether a policy is “worth it.” With these in hand we can appeal to powerful methods in decision analysis to quantify the opportunity cost of making the wrong decision. This essentially amounts to conducting rigorous sensitivity analysis whereby we re-estimate this value while allowing model a parameters to vary over plausible ranges.

In a series of recent studies, Hendren (2016 and 2017) develops both the theory and applications for estimating the *marginal value of public funds* (MVPF)*,* a summary measure of the benefits of a policy relative to its costs. As I discuss in this section, this measure provides a critical TK to facilitate

The general form of the MVPF is simple and intuitive:

Depending on the policy context and assumptions, this measure can be further simplified. For example, in the case of a cash transfer, we can appeal to the envelope theorem and normalize both the benefits and mechanical cost of the cash transfer to simplify the MVPF to reflect only one quantity: the fiscal externality (FE), or

This measure has close analogues in the public finance literature to the marginal excess burden and the marginal cost of public funds. It also echoes the *cost-effectiveness ratio* measure used in health technology assessment:

The economic framework underlying the MVPF is powerful in that it crosswalks between welfare analyses and empirical estimates of policy impact (Finkelstein 2019).

*Comparative Policy Assessment based on MVPF*

The MVPF measures the marginal value of an additional dollar spent on a policy—it asks the question: how do the welfare benefits accrued by implementing the policy compare to the costs of adopting that policy? These costs could be mechanical (e.g., the dollar value of a subsidy or cash transfer) and also the result of economic frictions brought about through policy implementation (e.g., behavioral changes that result in changes in labor force participation, tax revenue, etc.). From a normative standpoint, however, the MVPF is agnostic: it simply measures the ratio of benefits to costs and does not make affirmative statements about whether a policy is “worth it.”

If a policy is considered in isolation, then some threshold value summarizing society’s willingness to adopt the policy is all that is needed for policy decision-making. This threshold value could simply be a MVPF of 1 or, if society values some redistributive consequence of the policy, could be set at a value less than one. For example, Finkelstein, Hendren and Shepard (2018) make comparative assessments of health insurance subsidization policies by specifying a social welfare function over CRRA utility and a defined coefficient of risk aversion. This results in a MVPF benchmark of 0.2. But researchers do not necessarily have to take a firm stance on the curvature and structure of the social welfare function to define a decision-making benchmark. A value tied to an existing policy with strong support across a variety of ideological perspectives could also suffice. For instance, Finkelstein, Hendren and Shepard (2018) also consider a benchmark (0.88) based on the MVPF of the Earned Income Tax Credit (EITC)—a popular means-tested cash transfer program.[[4]](#footnote-5)

Often, however, policymakers must choose among *competing* policy alternatives. For example, the policy objective of increasing insurance coverage could be achieved by subsidizing the purchase of private plans, via expansion of means-tested public insurance programs, or via some hybrid approach such as a public program “buy-in.” An estimate of the MPVF could be constructed for each of these policies. So the question naturally becomes: how can we make relative comparisons between them?

Pairwise comparisons of the MVPF between these alternatives are not sufficient to fully catalogue the relative benefits of one policy over another. To see this, the figure below plots MVPF values for two “competing” policies (A and B). To facilitate exposition, “benefit” (W) is plotted on the X axis while “cost” (C) is plotted on the Y axis. In this setup, the slope of a line connecting the origin to each policy point is equivalent to the inverse of the MVPF for that policy.



The figure also defines a threshold parameter to guide assessments of whether a given policy yields a favorable MVPF relative to a defined benchmark. This threshold is conceptualized similarly to the benchmarks discussed above, but in this setup would take the value of the inverse of the MVPF of the benchmark (e.g., 1.13 for an EITC-based benchmark [MVPF=0.88]). Whatever value takes, any policy that lies below the benchmark line would be considered worth adopting from a societal perspective, while those above the line would not.

As seen in the Figure, both policies (A and B) have favorable MVPF relative to the benchmark. But even though policy B yields greater benefit and has a MVPF less than , we cannot conclude that policy B should be adopted over policy A. That is because a pairwise comparison of MVPFA and MVPFB does not consider the incremental costs and welfare gains from adopting B over A. That is, by implementing policy A we can obtain total benefit at cost . But to obtain an additional benefit of we must incur additional costs . These costs may not be worth it for the welfare gain received. The ratio of these two incremental changes, i.e., the slope of the line connecting A and B, is well above We can call this ratio the *incremental value of public funds (IVPF).*

It’s worth pausing for a minute to consider *why* two policies might have different MVPF values. In the case of a cash transfer, two policies with the same fiscal externality (FE) would have the same MVPF value (i.e., their points would overlap on the plot). But if the FE differs (e.g., there are different ways finance the subsidy, each with its own externality cost) then we would just want to choose the policy with the lowest externality. But in cases where the same ends could be achieved through different means (e.g., cash subsidy vs. expansion of in-kind benefits), different MVPFs will result. This is because the “benefits” cannot be simplified (via the envelope theorem) to a value of 1 in the numerator of the MVPF; a different “benefit” (W) could be realized depending on the design and economic incentives created by the specific policy under consideration. Similarly, if the two policies target slightly different populations, the projection of estimates through the efficient welfare weight may result in different MVPF values [TK-check accuracy].

In short, all that can be inferred from pairwise comparisons of MVPF values is that policy B does not dominate policy A (i.e., it does not achieve higher benefit at lower cost). This point is further illustrated in the figure below. Here, we consider three “versions” of policy B—each with the same MVPF (i.e., they all lie on the same line from the origin with slope (1/MVPFB). Here we see that only two (Bii and Biii) would pass a traditional cost-benefit test with benchmark . Indeed, we could even conceive of another policy (B0) that is above to the left of point A. This policy would “pass” a cost-benefit test relative to , but would clearly be dominated by policy A (since A would achieve higher benefit at lower cost).



*Value of Information: Intuition*

The intuition behind value-of-information methods is illustrated in the figure below. Suppose the MVPF is estimated using costs and benefits , but there is uncertainty around these estimates. This uncertainty could derive from estimation, sampling, and/or modeling error in literature-based model parameter values, or because some model parameters are based on judgment calls/assumptions and there is a range of plausible values. In the figure below, the “base case” values of and contribute to the primary estimate of the MVPF, which is denoted by a black point.

Given that estimates of the MVPF will vary based on the specific input values for or , it is natural to ask how sensitive modeling output is to this uncertainty. Moreover, we might also be interested in identifying the relative degree to which specific parameters contribute to model sensitivity, with the hope that future research efforts could be directed at reducing parameter uncertainty. To do so, we can conduct a *probabilistic sensitivity analysis* (PSA) whereby we construct a series MVPF estimates by re-estimating the microsimulation model after sampling values of or from a range of plausible values, and/or from a distribution of values that are (statistically) consistent with the available evidence. Each (hypothetical) PSA-based MVPF value is plotted as a grey point in the figure.

The key intuition underlying VOI is that these questions are really only relevant if modeling uncertainty affects decision-making. If we would make the same decision today as we would if we had better information on parameter values, then any extra effort to reduce uncertainty is unnecessary; the value of new information is low. By contrast, if model sensitivity based on current information results in *different* decisions depending on the values used, then the value of information is high.

This intuition is depicted visually by comparing scenarios A and B. We consider two decision-making thresholds ( and ) and a scenario with a high degree of sensitivity in MVPF values given the underlying uncertainty (Scenario A), and a scenario with a low degree of sensitivity (Scenario B).

The figure makes clear that the VOI depends as much on our selection of as on the underlying uncertainty. Under Scenario A, when decisions are benchmarked against the value of information is low: despite the fact that there is a large degree of modeling sensitivity to specific parameter values, we would *always* conclude that the policy has a favorable MVPF (i.e., it lies below the threshold line connecting the origin to . In other words, based on the available evidence underlying our model, there is no opportunity cost to making the wrong decision. By comparison, if we used we would make different decisions depending on the parameter values used. In that case, VOI methods quantify the opportunity cost (in welfare or dollar-valued terms) of making the wrong decision. We see in Scenario B, moreover, that in a scenario with a low degree of uncertainty and otherwise equivalent values, the value of information is low even with a threshold value of

*Value of Information: Theory*

will the value delivered by that additional dollar value deliver benefits that are worth the costs? But there are additional challenges when comparing the MVPFs across policy options. This MVPF does not consider the *incremental* value of one dollar spent on one policy vs. another.

Define as the societal willingness to pay (e.g., 0.8). Then the net value of public funds is

In a standard comparative assessment, the strategy that maximizes the NVPF would be selected as the “optimal” strategy (Hendren 2016, 2018; Neumann et al. 2016). That is, this strategy yields the largest welfare gain net of its costs for a given willingness to pay and among the strategies under consideration:

Normative comparisons between policies requires consideration of a social welfare function, since the appropriate benchmark for MVPF is not necessarily 1. This is because some policies serve a redistributive function and society may we willing to live with a MVPF less than 1 as a result of redistributive costs (Okun 1975). This function could be specified (e.g., a utilitarian social welfare function over CRRA utility functions and a defined coefficient of relative risk aversion) or could be benchmarked using an existing and generally supportive policy. For example, Finkelstein, Hendren and Shepard (2019) benchmark their assessments against the MVPF for the EITC (0.88).

Consider the figure below, which plots the marginal welfare benefit of each policy against its marginal cost. To facilitate interpretation the figure plots benefits on the X axis and costs on the Y axis. The MVPF is thus equal to the inverse of the slope of the line connecting the origin and each policy point.

The figure also plots a benchmark societal willingness to pay line. This line has slope , which is the inverse of the MVPF for the benchmark scenario.

We see that both A and B have favorable MVPF when compared against . It is tempting to conclude, then, that we could adopt either A or B. The only information we can take away from this exercise is that Policy B cannot dominate Policy A.



1. Most recently, Congressional attempts to repeal and replace the 2010 Affordable Care Act (ACA) were hampered by public outcry after the Congressional Budget Office (CBO) [projected](https://www.nytimes.com/2017/05/24/us/politics/cbo-congressional-budget-office-health-care.html) that upwards of 23 million people would become uninsured. The [twists](https://prescriptions.blogs.nytimes.com/2009/10/07/analysis-sees-baucus-bill-meeting-obamas-cost-and-deficit-targets/) [and](https://www.nytimes.com/2009/10/19/us/19iht-letter.html) [turns](https://www.nytimes.com/2009/10/06/health/policy/06health.html) of earlier debates over the ACA—and before it, the Clinton health plan, also were shaped by modelers' assessments of how reform would affect insurance coverage, premiums, health care spending, and government costs. [↑](#footnote-ref-1)
2. Historically, U.S. CBO relied on an elasticity-based model to simulate health reform policy, but recently (as of 2018) switched to a utility maximization framework. [↑](#footnote-ref-2)
3. A similar exercise for ex post policy evaluation is provided in Hendren (2019). [↑](#footnote-ref-3)
4. Similarly, the origin of the often-used $50,000 per quality-adjusted life year (QALY) threshold used in health technology assessments has been traced, in part, to the incremental cost effectiveness ratio for hemodialysis for end-stage renal disease—an explicit disease-based criterion used to determine Medicare eligibility for nonelderly adults in the United States (Grosse 2008; Neumann, Cohen and Weinstein 2014). [↑](#footnote-ref-5)