The Importance of Designing Rules for Ex-Post Evaluation[1]

Benjamin S. Kay and Marco Migueis

December 2024

Introduction

Ex-ante cost-benefit analyses and other impact assessments are now a standard part of the rulemaking process. However, some of the relevant impacts of rules are difficult or impossible to measure at the time a rule is written. Still, many of these impacts could be measured after implementation. Unfortunately, failure to plan for post-implementation evaluation at the time rules are written can make such impacts hard to measure in practice. This article argues that the rulemaking process should be reformed to avoid this problem. Designing rules with ex-post evaluation in mind would facilitate such evaluations, ultimately leading to better rules. The need for ex-post evaluation

Governments all over the world implement rules and regulations to safeguard the interests of their citizens. Citizens benefit from simple rules when complexity is unnecessary, inexpensive rules when costly ones are pointless, and rules that avoid adverse unintended consequences. To account for the tradeoffs posed by rules, the US Congress has mandated that federal agencies execute impact analyses for certain rulemakings.[2] In addition, by outstanding executive order federal agencies (but not independent regulatory agencies) must generally complete cost-and-benefit analyses before the issuance of major new rules.[3] The ex-ante impact analyses required by these statutes and policies help ensure that the costs of rules are justified and that some of the worst rules are never implemented.

However, rules may be more costly, generate fewer benefits, and have more undesirable consequences than analysts' ex-ante expectations. Ex-ante impact estimates are inherently inexact because future economic conditions and the reactions of the regulated are uncertain. Additionally, behavioral and political reasons like bureaucratic optimism, ideologically motivated policymaking, and regulatory capture by special interests can lead ex-ante impact analyses to overestimate the net benefits of rules.

[1]We thank Julianna Sterling for excellent research assistance. Also, we thank Mark Manuszak, Robert Stewart, and the participants of the Board's Supervision and Regulation Policy Research and Analytics seminar for their helpful comments and suggestions. The views expressed in this note are ours and do not reflect official positions of the Federal Reserve Board or the Federal Reserve System.
The authors have no conflicts of interests and did not receive funding from any entity to prepare this study.
[2] Two key laws here are the Regulatory Flexibility Act of 1980 and the Unfunded Mandates Reform Act of 1995. The Regulatory Flexibility Act requires agencies to consider the impact of their rules on small entities. The Unfunded Mandates Reform Act requires agencies to assess the costs and benefits of any regulation that has a material impact on state, local, or tribal governments. Both laws define agencies according to Title 5, Section 551(1) of the US Code; however, the Regulatory Flexibility Act includes independent regulatory agencies, while the Unfunded Mandates Reform Act excludes them.
[3] See Executive Order No. 12866, 58 Federal Register 51735 (October 4, 1993).

Ex-post evaluation can help ensure that rules are both efficient and effective. By assessing the impact of rules after implementation, ex-post evaluations can help determine whether rules achieve their intended objectives and justify the costs they impose. At the rulemaking stage, impact analysis is limited to the data and knowledge available to policy analysts before implementation. Ex-post evaluations have the benefit of hindsight. Once a rule is implemented, regulatory data shows the consequences of the rule, market prices reflect the rule's effects, and regulators can observe—rather than guess—how the rule was executed. Therefore, ex-post evaluations can provide a clearer picture of whether a rule is achieving its goals, identify undesirable consequences, and recommend improvements to streamline and simplify the rule. In some cases, ex-post evaluations can even identify rules that should be repealed. In other cases, an ex-post evaluation demonstrating the efficacy and low costs of a regulation will justify expanding the scope of those policies—applying an old requirement to a new set of firms, increasing an existing requirement, or substituting one rule for another based on evidence of superior outcomes in a different context..

In recognition of the benefits of ex-post evaluations, it is the current policy of the US executive branch that agencies should perform retrospective analysis to assist periodic review of the appropriateness of regulations.[4] Similarly, President Obama issued an executive order recommending that independent agencies perform retrospective analyses of regulations—a recommendation that has remained in place without amendment by subsequent administrations.[5] However, a 2024 Government Accountability Office assessment of the policies and procedures for regulatory analysis of independent federal bank regulators found that these regulators conducted few retrospective reviews of the effects of their existing rules.[6] By proposing approaches to improve the ability of government to perform useful ex-post evaluations, we hope to contribute to more frequent and effective use of this method of policy evaluation.[7]

Ex-post evaluations help build public trust in government by demonstrating that policymakers are evaluating the effectiveness of their rules. Increased trust would allow policymakers to more easily amend, repeal, or replace rules that fail at their goals or that are unjustified by their costs, potentially making regulations more efficient and effective.

Facilitating ex-post evaluation

While ex-post evaluations are essential for ensuring a better regulatory environment, conducting them can be challenging. There are difficulties in determining the appropriate counterfactual, collecting data on impacts, and attributing these impacts to the rule itself. Fortunately, these challenges can often be overcome if regulations are designed and implemented with ex-post evaluation in mind. Below, we discuss organizational choices and implementation choices that can support ex-post evaluation of a rule's impacts.

[4] See Executive Order No. 13563, 76 Federal Register 3821 (January 21, 2011).
[5] See Executive Order No. 13579, 76 Federal Register 41587 (July 14, 2011).
[6] See Government Accountability Office (2024).
[7] See Herrick and Sarewitz (2000), Kovacic (2006), and de Jong et al. (2018) for a discussion of the benefits of ex-post evaluation of policies in the context of environmental policy, competition policy, and infrastructure projects, respectively.

Staffing rulemakings with an eye toward ex-post evaluation

The ex-post evaluation of a rule benefits from the early involvement of the staff responsible for it in the rulemaking. Such staff can identify opportunities to support ex-post evaluation through the design of the rule. The marginal costs of such an assignment are likely to be modest. Under current practices, staff responsible for the evaluation of rules are often already involved in the rulemaking process, making suggestions to reduce burdens and increase benefits. Tasking those staff with identifying opportunities for supporting ex-post evaluation is unlikely to meaningfully increase costs.

However, embedding impact analysis staff in the rulemaking process poses tradeoffs. The economic expertise of impact analysis staff can inform and improve the rule, but such involvement may compromise the independence of the evaluation process from the policy development process. The Securities and Exchange Commission (SEC) mitigates this problem by having different cost-benefit analysts participate in the drafting and evaluation of the same rule.[8] This approach requires more staff and therefore results in higher costs to the agency. But staff responsible for ex-ante impact analysis are often already involved in the rulemaking process. Assigning those staff the task of identifying opportunities to tweak the rule to facilitate ex-post evaluation is unlikely to reduce their independence further. Therefore, the incremental costs resulting from tasking impact analysis staff with identifying such opportunities are likely to be modest, especially compared with the compliance costs and regulatory burdens associated with major new regulations.

To the extent that designing for ex-post evaluation encourages that more ex-post evaluation be undertaken, expenditures will increase relative to the status quo where such formal ex-post analyses are relatively uncommon. But these expenditures are money well spent, as the costs of ex-post evaluations are likely low compared with the benefits that they may produce by improving the efficacy and efficiency of rules. The average cost of ex-post evaluations may also decrease if rules are explicitly designed to facilitate such analyses. In addition, a good natural experiment in policymaking, created by designing for ex-post evaluation, is fertile soil for outside researchers to study policy effects at no additional cost to taxpayers. Therefore, the incremental cost is zero in many cases.

Differentiated implementation and application of rules

Ex-post evaluation can be facilitated by differentiated application of rules across similar firms. By creating a treatment and a control group, the application of different rules to similar firms allows for better ex-post identification of rule effects.[9] Two types of differences in application are often adopted that can help understand rule effects: (1) tiered application of rules by firm size or other relevant

[8] Even with this separation of drafting and analysis roles, both sets of analysts ultimately answer to SEC leadership. Skeptics could still justifiably worry about the analysts' independence when they have the same bosses.

[9] The typical identification assumptions used in measuring the effects of regulations do not require that firms be "similar." Rather, the (partly non-testable) assumptions needed for causal identification are exchangeability (i.e., there is conditional mean independence of treatment and control units), positivity (i.e., the probability of a unit being treated is positive but below one), and stable unit treatment value (i.e., potential outcomes of each unit are unaffected by the treatment assignment of other units). Similarity of treated and control units is one way to satisfy these assumptions but is not necessary.

A potential objection to using differentiated application of regulations to similar firms to identify treatment effects is that, by changing the conditions of competition across firms, introducing a regulation for a subset of firms can also affect outcomes for the non-treated firms – which would violate the stable unit treatment assumption. This is a potential limitation to consider when applying the identification methodologies discussed below.

indicators and (2) staggered implementation.[10] In addition, pure randomization of regulatory policies across firms, while uncommon and raising several challenges, does happen in public policy, and we discuss this at the end of this section.

Many regulations apply requirements to firms above a certain threshold of activity and not to firms below that threshold.[11] Such tiered application of regulations can be appropriate because applying more restrictive regulations to certain entities, such as those that are larger and more complex, often provides substantially larger benefits than applying the same regulations to other entities.[12] Also, the costs of regulatory compliance generally scale less than proportionally with activity levels (Brock and Evans, 1985).[13] However, tiering of regulations incentivizes firms to remain below regulatory thresholds to avoid compliance costs; therefore, tiering can reduce firms' ability to compete and reduce economic efficiency.[14] In addition to these potential costs and benefits of tiering, tiering enables regulators to better understand the effects of rules by allowing for comparisons of relevant outcomes for firms above and below the tiering threshold. Consideration of this additional benefit should favor more tiering in future rulemakings.

Ex-post analyses that rely on rule application thresholds have provided several insights regarding the effects of financial regulation policy. Some examples include Acharya et al. (2018), who found that the introduction of stress testing for large banks resulted in these banks reducing credit to risky borrowers; Kay et al. (2014), who found that banks subject to the Durbin amendment saw their interchange income fall substantially relative to smaller, exempt banks (however, in-scope firms offset more than 90 percent of the lost interchange income through increases in deposit fees);[15] Alvero, Ando, Xiao (2023), who used bunching of banks around Dodd-Frank size thresholds to estimate the costs of the act's regulatory burdens; Bindal et al. (2017), who studied the effects of two thresholds of greater bank merger scrutiny ($10 billion and $50 billion in consolidated holding company assets); and Wood (2023), who assessed the effects on bank liquidity creation of 2005 amendments to the Federal Deposit Insurance Corporation Improvement Act (FDICIA) and the Community Reinvestment Act (CRA). In addition, many other papers use regulatory thresholds to study nonfinancial regulations.[16] These examples show that ex-post

[10] Differentiated application of regulations is often mandated by statute. This article focuses on the ex-post evaluation of regulations issued by the executive branch agencies, but similar approaches can be followed by Congress to evaluate the effectiveness of statutes ex-post.

[11] In the banking regulation context, regulations with thresholds of application abound (see Congressional Research Service (2017) for a listing of all the thresholds that were in place at that time). Several regulations include enhanced requirements for large firms, complex firms, or both, including capital, liquidity, and resolution requirements.

[12] Large banking organizations tend to pose proportionally more systemic risk to the financial system than smaller banking organizations (see, for instance, Lorenc and Zhang (2020)).

[13] Given the same regulatory requirements, a firm with $1 billion in activity may be able to use the same system—and incur a similar absolute cost—as a firm that performs $2 billion of the activity.

[14] France imposes numerous regulations on firms once they employ 50 or more workers, resulting in an unusually high number of firms with exactly 49 employees. Such "bunching" has been estimated to reduce French aggregate productivity by 3.4 percent of gross domestic product (Garicano et al., 2016).

[15] The Durbin amendment to the Dodd-Frank Wall Street Reform and Consumer Protection Act requires the Federal Reserve to limit fees charged to retailers for credit card processing.

[16] See, for example, Leonardi and Pica (2006) and Vaccro (2017).

evaluation of rules can produce insights at a limited marginal cost when rule applicability differs for comparable firms.

The use of tiering to identify the effects of regulations typically presents certain challenges. For the cleanest identification of the effects of a policy, its implementation would be unanticipated so that firms could not strategically avoid being in scope. This is seldom the case in the real world. Still, avoiding regulations by shrinking below application thresholds can be very expensive.[17] And despite the potential for firms to adjust their activities and have some control about whether they are above or below the threshold once a rule comes into effect, ex-post comparison of the behavior of firms above and below a rule's application threshold can help understand the incentives resulting from a rule.

An additional challenge in using the tiered structure of requirements to identify the effects of policies is that certain specific thresholds trigger multiple different requirements. For example, under the Dodd-Frank Act, growing above $50 billion in total assets would result in a bank being subject to several additional prudential requirements, including enhanced stress testing, liquidity, and risk management requirements.[18] Given the contemporaneous implementation of several of these requirements into regulations, identifying their distinct effects is challenging. Still, even in the case of multiple requirements being triggered by a single threshold and implemented in a short window, comparing firms above and below the threshold can provide relevant information. This is particularly the case when estimating the most direct and distinct effects of requirements (for example, the effect of liquidity rules on bank holdings of high-quality liquid assets). In addition, when new requirements are applied to a regulatory threshold that is already in place, identifying the marginal impact of a new requirement may be possible if the introduction of the new requirement was not foreseeable when the regulatory threshold was initially introduced. For example, while the 2010 Dodd-Frank Act mandates prudential liquidity requirements for larger firms, a key part of the liquidity framework, net stable funding ratio requirements, would be proposed in 2016, finalized in February 2021, and become effective in July 2021.

Tiered application of regulations can be part of a long-term regulatory strategy of expanding regulations that demonstrate positive measured net benefits and contracting or eliminating those that do not. Regulators can (1) introduce a regulation to a narrow set of firms, (2) undertake ex-post evaluation of the costs and benefits of the regulation, and (3) expand or narrow the scope of application of a rule based on the evaluation findings.[19] Arguably, certain expansions of banking requirements have followed this pattern. For example, the 2023 proposed long-term debt requirement for large banking organizations was partly inspired by the perceived successes of similar mandates for global systemically

[17] For example, consider that the average domestic banking organization subject to Category III requirements had $550 billion in assets as of 2023:Q4. Given a typical return on assets of 1 percent, these banks would have to give up $3 billion in returns to be subject to Category IV requirements instead. Given observed growth in balance sheets, the current regulations have not caused banks to downsize in this way.

[18] See Congressional Research Service (2017) and Labonte and Perkins (2017). The $50 billion threshold was generally increased to $250 billion in the Economic Growth, Regulatory Relief, and Consumer Protection Act of 2018.

[19] See List (2024) for a discussion of how to optimally generate policy-relevant evidence before scaling an intervention.

important banks implemented in 2016.[20] Nevertheless, regulators should more frequently consider whether to expand or shrink regulatory requirements and more directly tie decisions to the ex-post evaluation of the costs and benefits of rules.

Staggered implementation, either with deterministic or random scope of application, is another common feature of rules that can aid in ex-post evaluation. Like tiering, applying a new rule to similar firms at different dates supports ex-post evaluation by allowing for comparison of the behavior of the firms covered early and late. Therefore, regulators should consider making more frequent use of staggered implementation to better understand the efficacy of rules.

In some cases, staggered implementation has been explicitly introduced to test the effects of policies. For example, staggered implementation was used by the SEC in the decimalization of stock market prices in 2000 and by the Internal Revenue Service (IRS) to distribute stimulus payments during the Great Recession.[21,22] Thanks to these clever implementation designs, we now know far more about the consequences of these policies.[23] In other cases, staggering of regulations has been due to different factors, such as prioritization of the most relevant entities or readiness to implement. For example, the Financial Industry Regulatory Authority's (FINRA) Trade Reporting and Compliance Engine (TRACE) — which the SEC requires that broker dealers use to report their security trades—was initially introduced only for corporate bonds in 2002 and has gradually included more securities over the years. Such staggered implementation has allowed researchers to estimate the consequences of TRACE by comparing securities covered by and exempt from this requirement.[24]

Staggered implementation has also been used in bank regulation. Examples include the current expected credit losses (CECL) accounting standard,[25] stress-testing requirements for large banks,[26] and

[20] See Federal Reserve System (2016) and Office of the Comptroller of the Currency, Federal Reserve System, and Federal Deposit Insurance Corporation (2023).

[21] Between August 2000 and January 2001, the New York Stock Exchange, in a series of incremental steps, began trading and quoting all its listed securities in increments of a penny. See SEC (2000) for additional details.

[22] The primary component of the Economic Stimulus Act of 2008 was a $100 billion program of economic stimulus using tax rebates. The IRS implementation of this stimulus was randomized based on the last two digits of the filer's Social Security number. See Parker et al. (2013) for details.

[23] Chung et al. (2004) studied the staggered roll out of decimalization. The Nasdaq Stock Market and the New York Stock Exchange have also performed ex-post analyses of this policy. As predicted, these analyses found that decimal pricing reduced bid-ask spreads. There were concerns the rule would reduce market liquidity, but the analyses failed to find evidence that liquidity declined. Chung, Lee, and Rösch (2020) and Werner et al. (2022) look at a related policy experiment involving a stratified random sampling process that assigned different quote and trading price increments in 2016 called the tick size pilot program. Parker et al. (2013) studied the effect of the staggered IRS roll out of the economic stimulus payments of 2008. This paper helped establish the efficacy of fiscal stimulus spending, finding a marginal propensity to consume of 50 to 90 percent in the first three months after stimulus payments were received. These results imply that such payments are highly effective stimulus.

[24] See Bessembinder et al. (2006) and Goldstein et al. (2007).

[25] The CECL accounting standard was required for entities that meet the definition of an SEC filer as of January 2020, and all other entities were required to adopt CECL by January 2023 (see Financial Accounting Standards Board, 2016).

[26] Stress tests were initially implemented for the largest 19 banks in 2009 (under the Supervisory Capital Assessment Program) and later extended to all banks with more than $50 billion in total assets in 2013, following on the statuary mandate from the Dodd-Frank Act; see Board of Governors (2009, 2013).

the liquidity coverage ratio (LCR).[27] These staggered implementations have resulted in valuable analyses of the effect of rules. Some examples include Loudis et al. (2021), who assessed the effect of early CECL adoption on reserves during the COVID-19 stress and found that reserves were substantially more responsive to the economic outlook for early adopters than for other firms; and Banerjee and Mio (2018), who used the staggered implementation of liquidity regulation in the UK to assess its effect on bank balance sheets.

Use of staggered implementation to understand the effect of policies presents certain limitations. Because the units serving as controls are eventually treated, often within a year or two, an analysis of differences across firms under staggered implementation is more likely to provide insights for short-term outcomes than for outcomes that depend on the long-term state of the regulation. In the example of decimalization of stock prices, understanding the effect of this modification on the costs of trading stocks was possible as this effect is realized in the short term. Meanwhile, most banking regulations influence outcomes over long time horizons, and banks sometimes begin adjusting in anticipation of announced but not yet implemented regulations. Therefore, when relying on comparisons based on staggered implementation to estimate the effect of regulatory policies, humility about the quality of identification and the extent of external validity is needed.

The staggered implementations of TRACE and CECL have enabled ex-post policy analysis but were chosen for other reasons. We propose that staggered implementation should be more frequently adopted to support the evaluation of the effectiveness of rules. Like tiering, staggered implementation can inform iterative policy improvements. Costly and problematic implementations can be slowed or canceled. Conversely, implementation can be accelerated and the scope of application broadened when early data shows the effects are large and positive.[28]

A more modest alternative to staggering of regulation implementation with a random scope of application is the use of small-scale pilot programs, whereby the effects of a regulation could be studied by applying it to a small number of firms – potentially voluntarily – for a limited time. Such small-scale experimentation can reduce any costs relating to experimentation and provide insights regarding the different outcomes for treated and non-treated firms. The SEC has undertaken several pilot studies, including in the rollout of TRACE, a "up-tick" pilot study (which removed a short-sale price test), and a "tick-size" pilot study (which increased the minimum price increment for stock prices from one cent to give cents). These pilots resulted in several studies on the effectiveness of these regulatory changes.[29] However, the often voluntary nature of pilot studies can substantially limit their usefulness in identifying the effects of regulation (see Harris et al. (2021) for a discussion of the limitations of pilot programs in financial regulation).

[27] In an implementation that combines staggering and tiering, the LCR was initially introduced for the largest firms in 2015:Q1, and, later, a modified LCR was applied to a set of smaller yet large firms starting in 2016:Q1 (Office of the Comptroller of the Currency, Federal Reserve System, and Federal Deposit Insurance Corporation, 2014).
[28] This occasionally happens in medical trials where a treatment is so effective that the study is ended early over ethical concerns about denying potentially life-saving treatment to the control group. See Mueller, Montori, Bassler, Koenig, and Guyatt (2007) for a discussion of this issue.
[29] See Harris et al. (2021) for a description of these and other pilot programs and for several references to studies that were based on these pilot programs.

Beyond short-term randomized staggering of policy implementation, as in the examples of the decimalization of stock prices or the IRS stimulus check, a long-term strategy of evidence based regulation would study the effects of policy through a greater number of open-ended policy experiments based on randomization.[30] Randomized application of a treatment to comparable units is the gold standard for causally identifying treatment effects and, for this reason, has long been used in clinical trials. Research in experimental economics has attempted to illuminate significant economic questions by applying similar controlled trial methodologies in lab-like settings. While a fully controlled trial is generally unfeasible when dealing with the regulation of firms, regulators should seek opportunities to implement similar approaches.

Policy randomization can result in unequal treatment of regulated entities. Still, randomization is employed for important decisions of our society, such as the military draft, selection of juries, or clinical trials,[31] and these choices are generally perceived as fair when based on ex-ante fair mechanisms.[32] To the extent that we consider knowledge about the effectiveness of policies valuable, we should be willing to incur some costs to achieve it, including accepting some inequities among regulated entities for a period of time. This does not mean that arbitrary large amounts of inequity should be introduced just for the sake of policy experimentation. Rather, the potential benefits of policy experimentation should be weighted relative to the potential costs resulting from the inequities introduced. In general, policy randomization is likely to be more ex-ante justifiable the more uncertainty there is about the effects of policies (including the effects on the bottom line of regulated entities) and in situations where the potential costs of differentiated treatment are lower.

Bank supervision is an area where policy randomization could be used to learn more. When appropriate baseline levels of supervision are in place, the incremental costs and benefits of banking supervision could be studied through randomized assignment of different supervisory approaches. Supervisors could test the efficacy of different frequencies of examination, sizes and compositions of examination teams, triggers for an examination, or examination focuses. In the long run, such experimentation could help regulatory agencies more efficiently deploy their limited resources.

Note that the socially optimal scope of application for regulations may be different from the optimal application for identifying a regulation's effects. Nevertheless, while employing a suboptimal scope of application is theoretically a potential cost of using the applicability of rules to learn their effects, empirical evidence demonstrating that existing scopes of application are efficient is generally lacking. Therefore, experimentation around the scope of application of regulations may be possible without ex-ante demonstrable loss of efficiency.

Notwithstanding our emphasis on expanding randomization, tiering, and staggering, ex-post evaluation may be harder to perform for certain policies. For example, as discussed earlier, the application of multiple regulatory requirements according to the same threshold can challenge ex-post evaluation. But, to the extent that this is a problem, it suggests that regulators should adopt rules that better allow

[30] See Abramowicz, Ayres, and Listokin (2011) for further discussion of the benefits of randomizing laws and regulations.

[31] See Silverman and Chalmers (2001), Erikson and Stoker (2010), Gramlich (2017), and Weldon (2017) for a discussion of public perspectives on fairness in these contexts.

[32] See Rawls (1971) for a discussion of the fairness of equality prior to resolution of uncertainty (i.e., being behind the veil of ignorance).

for testing of policy effects (such as by using more granular regulatory thresholds). Asking the public to live under untestable policies is not fair or reasonable. Introducing rules that are untestable is just experimenting on everyone but learning nothing from the process (Kling, 2020).

Dual reporting

Identifying the costs and benefits of rules requires more than clear identification of a treatment and control population. We also need good data, a concern sometimes neglected in the rule making process. For example, the definitions of capital and risk-weighted assets have changed several times in the post-Dodd-Frank Act era. However, regulatory reporting requirements usually ask firms for data under only the most recent definitions, making it difficult to measure whether observed changes in capital and assets are caused by changes in behavior or by changes in definitions.

New rules can avoid this issue by requiring that firms report data under both old and new definitions for a period. This is similar to how banks report the value of securities simultaneously under mark-to-market and amortized cost valuations. These parallel data were invaluable for understanding the banking problems of March 2023, as they helped regulators estimate which banks had large securities losses from changes in interest rates that were not reflected in regulatory capital. Similarly, when revising questions in its Current Population Survey, the Census Bureau asks the same people new and the old version of the same question to ensure that the data remain usable and consistent over time.[33]

Dual reporting for a period can also be applied when the scope of application of a rule is being reduced. Having data for the formerly covered firms can help assess whether regulatory simplifications and burden reductions deliver on their promises.

Not collecting overlapping data often leads to problems. Famously, the Census Bureau changed its annual health insurance survey contemporaneously with the implementation of the Affordable Care Act, greatly complicating ex-post evaluations of that law.[34] Another example was the introduction of the concept of discouraged workers and several other significant changes to the definitions of the U-4 through U-6 unemployment measures by the Bureau of Labor Statistics in 1994.[35] Such changes made comparisons with earlier data more difficult, which would complicate economic assessment of the effect of laws like the Family and Medical Leave Act of 1993 and the Violent Crime Control and Law Enforcement Act of 1994.

Dual reporting would temporarily increase compliance costs. Still, in cases where reporting under old definitions can rely on existing systems, marginal costs may be low In situations where costs of retaining legacy reporting are low, dual reporting is likely to provide higher potential benefits than costs.

Conclusion

Effective ex-post evaluation of rules is likely to provide substantial benefits. Therefore, we propose two rule design strategies to promote ex-post evaluation. First, the expanded use of tiering, randomization, and staggered implementation would allow for better identification of the effects of rules by creating a

[33] The Census Bureau frequently asks multiple versions of a question when introducing a potential question to the American Community Survey. In addition, extensive effort typically goes into the revision of existing questions to validate when and if the new and old series are comparable. See Census Bureau (2020).
[34] See Tavernise (2014).
[35] See Bureau of Labor Statistics (1995).

treatment and control population. Table 1 highlights several applications of these techniques in evaluating public policies as discussed in this paper, showcasing how they can yield insights across diverse contexts. Second, dual-reporting requirements, where overlapping new and existing data collections exist in a window around rule implementations, help distinguish the consequences of rules from the measurement changes of those rules. In addition, analysts responsible for undertaking ex-post evaluation of rules should be integrated into the rulemaking process early and consulted on how to ensure that ex-post evaluation can be effective. Once designing rules to facilitate ex-post evaluation becomes standard practice, economists and other regulators will surely find many other approaches to do so. With a broader array of approaches, many more ex-post evaluations would be possible.

Some policymakers may reject this suggestion, perhaps worrying that they have more to lose by clearly identifying the poor policies they have implemented than they could gain by demonstrating the successes of their best policies. In addition, policy makers worry about the costs of ex-post evaluations or the possibility that designing for ex-post evaluation could slow down the rulemaking process, delaying the public's access to the benefits of well-crafted rules. Nevertheless, policymakers should consider bolstering ex-post evaluation. The regulatory community has managed the discomforts of ex-ante impact analyses by normalizing them. Nearly every major rule includes such an analysis, which provides relevant information to policymakers and the public on the costs and benefits of proposed rulemakings. By taking steps to normalize ex-post evaluation, policymakers could help reduce the embarrassment and career harm associated with implementing rules that perform poorly under such evaluations. This, in turn, could encourage more honest and timely evaluations of rules. Such improved ex-post evaluation of rules would help the public know the benefits and costs of governmental policies and help eliminate or reform the policies that do not work.

Table 1. Selected Examples of Ex-Post Evaluations Using Randomization, Tiering, and Staggering

Category	Example	Details/Implications	Sources
Tiering	Dodd-Frank size thresholds for banking regulations	Size thresholds influence regulatory compliance, allowing study comparisons of firms near the threshold	Acharya et al. (2018), Alvero, Ando, and Xiao (2023)
Tiering	Stress testing for large banks	Stress testing led to reduced credit to risky borrowers, aiding in understanding regulation effects	Acharya et al. (2018)
Tiering	Durbin amendment effect on banks	The effect of the Durbin amendment on interchange income and banks' offsetting strategies	Kay et al. (2014)
Tiering	Bank merger regulatory thresholds ($10 billion and $50 billion)	Assessed the effects of increased scrutiny on bank mergers	Bindal et al. (2017)
Tiering	Bank liquidity creation effects of FDICIA/CRA amendments	Amendments to banking acts provided a basis to study liquidity changes	Wood (2023)
Staggering	Decimalization of stock market prices by the SEC	Consequences of decimalization on market health and function	Chung et al. (2004)
Staggering	IRS distribution of stimulus payments during the Great Recession	Economic effects of fiscal stimulus from transfer payments	Parker et al. (2013)
Staggering	FINRA's TRACE implementation for corporate bonds	The effects of disclosure on bond-market functioning	Bessembinder, Maxwell, and Venkataraman (2006), Goldstein, Hotchkiss, and Sirri (2007)
Staggering	CECL accounting standard adoption and its COVID-19 impact assessment	Assessed reserve responsiveness during economic stress periods	Loudis et al. (2021)
Staggering	Effect of liquidity regulation in the UK	Assessed the effect on the balance sheet of banks	Banerjee and Mio (2018)
Randomization	Stock tick size pilot program	Effects of tick and trading increments on stock liquidity	Chung, Lee, and Rösch (2020), Werner et al. (2022)

Randomization	Vietnam draft lottery for military service	Randomized draft assignments illuminated political, social, and economic outcomes	Erikson and Stoker (2009), Angrist (1990)
Randomization	Clinical trials for new medicines	Randomization provides gold-standard causal identification for treatment effects	Crofton and Mitchison (1948), Meldrum (2000)

References

Abramowicz, Michael, Ian Ayres, and Yair Listokin. "Randomizing Law." *University of Pennsylvania Law Review* 159, no. 4 (2011): 929–1005, https://www.jstor.org/stable/41149891.

Acharya, Viral V., Allen N. Berger, and Raluca A. Roman. "Lending Implications of U.S. Bank Stress Tests: Costs or Benefits?" *Journal of Financial Intermediation* 34 (2024): 58–90, https://www.sciencedirect.com/science/article/pii/S104295731830010X.

Alvero, Adrien, Sakai Ando, and Kairong Xiao. "Watch What They Do, Not What They Say: Estimating Regulatory Costs from Revealed Preferences." *Review of Financial Studies* 36, no. 6 (2023): 2224–73, https://academic.oup.com/rfs/article/36/6/2224/6873755?login=true.

Angrist, Joshua D. "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records." *American Economic Review* (1990): 313–36, https://www.jstor.org/stable/2006669.

Banerjee, Ryan N., and Hitoshi Mio. "The Impact of Liquidity Regulation on Banks." *Journal of Financial Intermediation* 35 (2018): 30–44, http://dx.doi.org/10.1016/j.jfi.2017.05.008.

Bessembinder, Hendrik, William Maxwell, and Kumar Venkataraman. "Market Transparency, Liquidity Externalities, and Institutional Trading Costs in Corporate Bonds." *Journal of Financial Economics* 82, no. 2 (2006): 251–88, https://www.sciencedirect.com/science/article/pii/S0304405X06000699.

Bindal, Shradha, Christa H.S. Bouwman, Shuting Sophia Hu, and Shane A. Johnson. "Regulatory Size Thresholds and Merger and Acquisition Behavior." *Mays Business School Research Paper* 2974303 (2017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2974303.

Board of Governors of the Federal Reserve System. *Supervisory Capital Assessment Program: Design and Implementation* (2009), April 24, www.federalreserve.gov/bankinforeg/bcreg20090424a1.pdf.

Board of Governors of the Federal Reserve System. *Dodd-Frank Act Stress Testing* (2013), www.federalreserve.gov/bankinforeg/stress-tests/dodd-frank-act-stress-testing.htm.

Brock, William A., and David S. Evans. "The Economics of Regulatory Tiering." *Rand Journal of Economics* (1985): 398–409, https://www.jstor.org/stable/2555566.

Bureau of Labor Statistics. *BLS Introduces New Range of Alternative Unemployment Measures* (1995), October, https://www.bls.gov/opub/mlr/1995/10/art3full.pdf.

Census Bureau. "How a Question Becomes Part of the ACS," in *Understanding and Using American Community Survey Data: What Federal Agencies Need to Know* (2020),

www.census.gov/content/dam/Census/library/publications/2020/acs/acs_federal_handbook_2020_ch03.pdf.

Chakravarty, Sugato, Stephen P. Harris, and Robert Wood. "Decimal Trading and Market Impact." Available at SSRN 266877 (2001), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=266877.

Chung, Kee H., Bonnie F. Van Ness, and Robert A. Van Ness. "Trading Costs and Quote Clustering on the NYSE and NASDAQ after Decimalization." *Journal of Financial Research* 27, no. 3 (2004): 309–28, https://onlinelibrary.wiley.com/doi/full/10.1111/j.1475-6803.2004.00096.x.

Congressional Research Service. "Tailoring Bank Regulations: Differences in Bank Size, Activities, and Capital Levels." R45051 (2017), December, https://crsreports.congress.gov/product/pdf/R/R45051.

Crofton, John, and D.A. Mitchison. "Streptomycin Resistance in Pulmonary Tuberculosis." *British Medical Journal* 2, no. 4588 (1948): 1009, https://pmc.ncbi.nlm.nih.gov/articles/PMC2092236.

de Jong, Gerard, Silvia Vigneti, and Chiara Pancotti. "Ex-Post Evaluation of Major Infrastructure Projects." *Transportation Research Procedia* 42 (2020): 75–84, https://doi.org/10.1016/j.trpro.2019.12.008.

Erikson, Robert S., and Laura Stoker. "Vietnam Draft Lottery Status and Political Attitudes." *Annual meeting of the Midwest Political Science Association, Chicago, IL* (2009), https://www.columbia.edu/~rse14/vietnam_rev_Feb2010.pdf.

Executive Office of the President. Executive Order 12866 of September 30, 1993: Regulatory Planning and Review. *Federal Register* (1993), October 4, www.archives.gov/files/federal-register/executive-orders/pdf/12866.pdf.

Executive Office of the President. Executive Order 13563 of January 18, 2011: Improving Regulation and Regulatory Review. *Federal Register* (2011), January 21, www.federalregister.gov/d/2011-1385.

Executive Office of the President. Executive Order 13579 of July 11, 2011: Regulation and Independent Regulatory Agencies. *Federal Register* (2011), July 14, www.federalregister.gov/d/2011-17953.

Federal Reserve System. "Total Loss-Absorbing Capacity, Long-Term Debt, and Clean Holding Company Requirements for Systemically Important U.S. Bank Holding Companies and Intermediate Holding Companies of Systemically Important Foreign Banking Organizations." *Federal Register*, Docket No. R-1523, RIN 7100-AE37 (2016), December, www.federalreserve.gov/newsevents/pressreleases/files/bcreg20161215a1.pdf.

Financial Accounting Standards Board. *Accounting Standards Update (ASU) 2016-13, Financial Instruments—Credit Losses (Topic 326)*. Financial Accounting Foundation (2016), https://www.fasb.org/projects/current-projects/financial-instruments%E2%80%94credit-losses-(topic-326)%E2%80%94purchased-financial-assets-401651.

Garicano, Luis, Claire Lelarge, and John Van Reenen. "Firm Size Distortions and the Productivity Distribution: Evidence from France." *American Economic Review* 106, no. 11 (2016): 3439–79, https://www.aeaweb.org/articles?id=10.1257/aer.20130232.

Goldstein, Michael A., Edith S. Hotchkiss, and Erik R. Sirri. "Transparency and Liquidity: A Controlled Experiment on Corporate Bonds." *Review of Financial Studies* 20, no. 2 (2007): 235–73, https://academic.oup.com/rfs/article/20/2/235/1573567.

Government Accountability Office. “Financial Services Regulations—Improvements Needed to Policies and Procedures for Regulatory Analysis.” GAO-24-106206 (2024), July, https://www.gao.gov/products/gao-24-106206.

Gramlich, John. “Jury Duty Is Rare, but Most Americans See It as Part of Good Citizenship.” *Pew Research Center* (2017), August 24, https://www.pewresearch.org/short-reads/2017/08/24/jury-duty-is-rare-but-most-americans-see-it-as-part-of-good-citizenship.

Harris, Larry, Charles Kahn, Robert McDonald, and Chester Spatt. “The Role of Pilot Studies in Financial Regulation.” October 2019. Available at the Social Science Research Network: https://ssrn.com/abstract=3629339.

Herrick, Charles, and Daniel Sarewitz. “Ex Post Evaluation: A More Effective Role for Scientific Assessments in Environmental Policy.” *Science, Technology, & Human Values* 25, no. 3 (2000): 309–31, https://www.jstor.org/stable/690129.

Kay, Benjamin S., Mark D. Manuszak, and Cindy M. Vojtech. “Competition and Complementarities in Retail Banking: Evidence from Debit Card Interchange Regulation.” *Journal of Financial Intermediation* 34 (2018): 91–108, https://www.sciencedirect.com/science/article/pii/S1042957318300184.

Kling, Arnold. “Micro Experiments and Macro Experiments,” *askblog* (2020), May 1, https://www.arnoldkling.com/blog/micro-experiments-and-macro-experiments.

Kovacic, William E. “Using Ex Post Evaluations to Improve the Performance of Competition Policy Authorities.” *Journal of Corporation Law* 31 (2006): 503–47, https://scholarship.law.gwu.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1786&context=faculty_publications.

Labonte, Marc, and David W. Perkins. *Bank Systemic Risk Regulation: The $50 Billion Threshold in the Dodd-Frank Act*. Congressional Research Service (2017), https://sgp.fas.org/crs/misc/R45036.pdf.

Leonardi, Marco, and Giovanni Pica. “Effects of Employment Protection Legislation on Wages: A Regression Discontinuity Approach.” *University of Milan* (2006), https://conference.iza.org/conference_files/TAM2006/pica_g1513.pdf.

List, John. A. “Optimally Generate Policy-Based Evidence before Scaling.” *Nature* 626 (2024): 491–99, https://www.nature.com/articles/s41586-023-06972-y.

Lorenc, Amy G., and Jeffery Y. Zhang. “How Bank Size Relates to the Impact of Bank Stress on the Real Economy.” *Journal of Corporate Finance* 62 (2020): 101592, https://doi.org/10.1016/j.jcorpfin.2020.101592.

Loudis, Bert, Sasha Pechenik, Ben Ranish, Cindy M. Vojtech, and Helen Xu. “New Accounting Framework Faces Its First Test: CECL during the Pandemic.” FEDS Notes (2021), December 3, https://www.federalreserve.gov/econres/notes/feds-notes/new-accounting-framework-faces-its-first-test-cecl-during-the-pandemic-20211203.html.

Meldrum, Marcia L. “A Brief History of the Randomized Controlled Trial: From Oranges and Lemons to the Gold Standard.” *Hematology/Oncology Clinics of North America* 14, no. 4 (2000): 745–60, https://zenodo.org/records/1260107.

Mueller, Paul S., Victor M. Montori, Dirk Bassler, Barbara A. Koenig, and Gordon H. Guyatt. “Ethical Issues in Stopping Randomized Trials Early Because of Apparent Benefit." *Annals of Internal*

Medicine 146, no. 12 (2007): 878–81, https://www.acpjournals.org/doi/full/10.7326/0003-4819-146-12-200706190-00009.

Office of Management and Budget. "Circular A-4: Regulatory Analysis." Executive Office of the President (2023), www.whitehouse.gov/wp-content/uploads/2023/11/CircularA-4.pdf.

Office of the Comptroller of the Currency, Federal Reserve System, and Federal Deposit Insurance Corporation. "Liquidity Coverage Ratio: Liquidity Risk Measurement Standards." *Federal Register*, vol. 79, no. 197 (2014), October 10, www.federalregister.gov/documents/2014/10/10/2014-22520/liquidity-coverage-ratio-liquidity-risk-measurement-standards.

Office of the Comptroller of the Currency, Federal Reserve System, Federal Deposit Insurance Corporation. "Net Stable Funding Ratio: Liquidity Risk Measurement Standards and Disclosure Requirements." *Federal Register* (2016), June 1, www.federalregister.gov/documents/2016/06/01/2016-11505/net-stable-funding-ratio-liquidity-risk-measurement-standards-and-disclosure-requirements.

Office of the Comptroller of the Currency, Federal Reserve System, Federal Deposit Insurance Corporation. "Net Stable Funding Ratio: Liquidity Risk Measurement Standards and Disclosure Requirements." *Federal Register* (2021), February 11, www.federalregister.gov/documents/2021/02/11/2020-26546/net-stable-funding-ratio-liquidity-risk-measurement-standards-and-disclosure-requirements.

Office of the Comptroller of the Currency, Federal Reserve System, Federal Deposit Insurance Corporation. "Long-Term Debt Requirements for Large Bank Holding Companies, Certain Intermediate Holding Companies of Foreign Banking Organizations, and Large Insured Depository Institutions." *Federal Register*, vol. 88, no. 64524 (2023), September 19, www.federalregister.gov/documents/2023/09/19/2023-19265/long-term-debt-requirements-for-large-bank-holding-companies-certain-intermediate-holding-companies.

Parker, Jonathan A., Nicholas S. Souleles, David S. Johnson, and Robert McClelland. "Consumer Spending and the Economic Stimulus Payments of 2008." *American Economic Review* 103, no. 6 (2013): 2530–53, https://www.aeaweb.org/articles?id=10.1257/aer.103.6.2530.

Rawls, John. *A Theory of Justice*. Belknap Press of Harvard University Press (1971), https://doi.org/10.2307/j.ctvjf9z6v.

Securities and Exchange Commission. "Commission Notice: Decimals Implementation Plan for the Equities and Options Markets." Exchange Committee on Decimals (2000), July 24, https://www.sec.gov/rules-regulations/2000/07/decimals-implementation-plan-equities-options-markets-exchange-committee-decimals-submitted-july-24.

Silverman, William A., and Iain Chalmers. "Casting and Drawing Lots: A Time Honoured Way of Dealing with Uncertainty and Ensuring Fairness." *BMJ* 323, no. 7327 (2001): 1467–68, https://pmc.ncbi.nlm.nih.gov/articles/PMC1121910.

Tavernise, Sabrina. "Census Survey Revisions Mask Health Law Effects." *New York Times* (2014), April 15, www.nytimes.com/2014/04/16/us/politics/census-survey-revisions-mask-health-law-effects.html.

U.S. Congress. Regulatory Flexibility Act, Public Law 96-354, 94 Stat. 1164 (1980), www.govinfo.gov/content/pkg/STATUTE-94/pdf/STATUTE-94-Pg1164.pdf.

U.S. Congress. Unfunded Mandates Reform Act of 1995, Public Law 104-4, 109 Stat. 48 (1995), www.govinfo.gov/content/pkg/PLAW-104publ4/pdf/PLAW-104publ4.pdf.

U.S. Congress. Economic Stimulus Act of 2008, Public Law 110-185, 122 Stat. 613 (2008), www.congress.gov/bill/110th-congress/house-bill/5140/text.

U.S. Congress. Dodd-Frank Wall Street Reform and Consumer Protection Act, Public Law 111-203, 124 Stat. 1376 (2010), www.govinfo.gov/content/pkg/PLAW-111publ203/pdf/PLAW-111publ203.pdf.

U.S. Congress. Economic Growth, Regulatory Relief, and Consumer Protection Act, Public Law 115-174, 132 Stat. 1296 (2018), www.congress.gov/bill/115th-congress/senate-bill/2155/text.

Vaccaro, Giannina. "Using Econometrics to Reduce Gender Discrimination: Evidence from a Difference-in-Discontinuity Design." 2nd IZA Workshop: Gender and Family Economics (2018), https://conference.iza.org/conference_files/Gender_2018/vaccaro_g23998.pdf.

Weldon, Kathleen. "Suppose They Gave a War and Nobody Came: Changing Opinions on the Draft." Roper Center for Public Opinion (2017), July 24, https://ropercenter.cornell.edu/blog/suppose-they-gave-war-and-nobody-came-changing-opinions-draft.