



## **Learning from other Domains to Advance AI Evaluation and Testing**

We are grateful to the author of the enclosed expert report, which forms part of a broader series commissioned by Microsoft.

These reports were commissioned as part of Microsoft's effort to draw lessons from other domains to strengthen testing and evaluation as a cornerstone of AI governance.

The insights contained in each report reflect the authors' independent analysis and expertise. The views expressed are those of the authors alone.

We thank the authors for their intellectual generosity and thoughtful engagement throughout this project.

### **Full Series**

**Civil aviation:** *Testing in Aircraft Design and Manufacturing* by Paul Alp.

**Cybersecurity:** *Cybersecurity Standards and Testing—Lessons for AI Safety and Security*, by Stewart Baker.

**Financial services (bank stress testing):** *The Evolving Use of Bank Stress Tests*, by Kathryn Judge.

**Genome editing:** *Governance of Genome Editing in Human Therapeutics and Agricultural Applications*, by Alta Charo and Andy Greenfield.

**Medical devices:** *Medical Device Testing: Regulatory Requirements, Evolution and Lessons for AI Governance*, by Mateo Aboy and Timo Minssen.

**Nanoscience:** *The regulatory landscape of nanoscience and nanotechnology, and applications to future AI regulation*, by Jennifer Dionne.

**Nuclear energy:** *Testing in the Nuclear Industry*, by Pablo Cantero and Gerónimo Poletto Antonacci.

**Pharmaceuticals:** *The History and Evolution of Testing in Pharmaceutical Regulation*, by Daniel Benamouzig and Daniel Carpenter.

## The Evolving Use of Bank Stress Tests

Kathryn Judge

### Introduction

“Stress testing” of banks, or more accurately, banking organizations, is one of the most important regulatory innovations to emerge from the 2008 financial crisis. The stress testing of an individual bank entails assessing how that bank will fare under a given adverse scenario, usually meant to replicate the types of developments that would occur during a deep or prolonged recession, such as a significant rise in unemployment, heightened market volatility and large declines in asset values. Typically, the regulator will provide one or more scenarios, the bank then provides the requisite data and the regulator uses its model to assess how the bank will fare in the face of the scenarios provided. If the regulator determines that the bank would not remain in good enough health, in the sense of remaining well capitalized in the face of the hypothetical adverse scenario, it usually requires the bank to increase its capital by foregoing distributions to shareholders via dividends or share repurchases. A bank’s capital refers to the amount of equity it uses to fund its operations; higher capital increases a bank’s capacity to absorb losses. Banks also engage in bank-run stress tests using their internal risk-management tools as part of their obligations, providing separate insights into the quality of a bank’s risk management regime.

The benefits of regulatory stress testing are many, as are the challenges. They vary depending on the conditions under which the stress tests are run. The original regulatory stress tests occurred in early 2009, when the financial system remained mired in the Great Financial Crisis (GFC). This allowed bank regulators to devise an adverse scenario based on the actual and specific hardships the economy was already facing. Bank regulators were willing to subject banks to a rigorous and realistic stress test because Congress had already given the Treasury Department \$700 billion in funding that the Treasury Department could use to recapitalize any bank revealed to have deficiencies that it could not remediate by raising additional funds from market-based sources. This was critical to the rigor and credibility of the exercise, as bank regulators are hesitant to undertake any public exercise that might reveal adverse information or otherwise exacerbate financial fragility unless they have the tools to contain the fallout. With the benefit of the temporarily more expansive authority and other wise design choices, the original round of stress tests helped provide credible information about bank health and enhanced the functioning of the financial sector, helping to further pave the road to recovery.

Most regulatory stress tests today occur during times of peace. In the United States, the Federal Reserve now conducts annual stress tests of the country’s largest banking organizations and bi-annual tests of many other large banks. Like crisis-time stress tests, these exercises aspire to provide regulators and market participants credible information about the health of banks and to enhance the health of individual banks and the banking system. At the same time, the different context changes the mechanisms and associated risks. One of the biggest benefits of peace-time stress testing as currently operationalized is to make capital requirements—a core element of bank regulation—more forward-looking. The flip side is that as stress tests have been incorporated into capital requirements, they have become less dynamic and variable. This reduces the amount of new information they generate and dampens the role they could play in getting banks to contemplate different types of shocks to the economy. There are also signs that the tests may be

revised in the years ahead to further bifurcate the role they play in setting capital requirements and the role they play in producing useful information.

There are an array of challenges that impede both peace and crisis-time stress testing. The models that regulators use have become somewhat more sophisticated over time, but remain flawed. Among the significant drawbacks is that they often fail to incorporate the feedback loops that characterize periods of financial distress. Regulatory stress tests also present a possible regulatory conundrum. In the view of many regulators, the models used by the Federal Reserve—the bank regulator that runs the stress tests—must remain confidential for the tests to remain useful. The common concern is that full disclosure of the models would allow banks to game the tests, a common challenge in bank regulation. Yet the banks—primarily through efforts of leading trade groups—have argued that confidentiality and the refusal to submit the models they use to notice-and-comment rulemaking violate the Administrative Procedure Act.<sup>1</sup> In December 2024, the Federal Reserve announced it would make comprehensive changes to how stress tests are run to make the models more transparent and reduce the variability in the capital requirements they impose.<sup>2</sup> The changing administrative law landscape was specifically cited as a factor motivating the change, and the Fed is now facing a lawsuit claiming that the stress tests, as they had been run, ran afoul of the Administrative Procedure Act. The revisions have yet to be formally proposed, much less implemented, but the announcement suggests a major shift is underway.

### The Testing Landscape

The role of bank stress tests is best understood in relation to the broader bank regulatory and supervisory regime, and the deficiencies in that regime that stress tests are meant to overcome. Banks have long been heavily regulated. As illustrated all too vividly by the Great Depression and the deep recession that followed the GFC, banking crises inflict real and lasting harm on the real economy. To help avert the harm that bank failures can impose, the government intervenes through an array of ex ante and ex post mechanisms, ranging from deposit insurance to stepping in to prevent bank failures or mitigate the impact of a failure. This mutes the incentive of bank depositors and other creditors to monitor a bank's health as closely as they otherwise might, contributing to excess risk taking by banks.

Historically, the government sought to promote bank safety and soundness and counteract this moral hazard through a combination of regulation (*e.g.*, limits on scale, scope, activities restrictions, capital regulation and other rules) and supervision (which fills in the inevitable gaps in any rule-based regime and allows for a more individualized, dynamic and opaque form of oversight). As bank regulation became increasingly standardized across different jurisdictions, starting in the 1980s through the Basel Accords, capital regulation became increasingly central to efforts to keep banks healthy. Capital captures the amount of equity, and sometimes long-term or subordinated debt, that a bank uses to fund its operations. Higher levels of capital enhance the

---

<sup>1</sup> Complaint, *Bank Policy Institute v. Board of Governors of the Federal Reserve System*, No. 2:24-cv-04300-EAS-CMV (S.D. Ohio 2024), <https://fm.cnbc.com/applications/cnbc.com/resources/editorialfiles/2024/12/24/BPI-OHChamber-OHBankers-ABA-Chamber-Stress-Testing-Complaint-20241224.pdf>.

<sup>2</sup> Press Release, *Due to evolving legal landscape & changes in the framework of administrative law, Federal Reserve Board will soon seek public comment on significant changes to improve transparency of bank stress tests & reduce volatility of resulting capital requirements*, Board of Governors of the Federal Reserve System, <https://www.federalreserve.gov/newsevents/pressreleases/bcreg20241223a.htm>.

bank's capacity to absorb losses, reducing the probability of failure and reducing the misalignment that can otherwise arise between the risk-taking incentives of shareholders and what is socially optimal. A core drawback of capital regulation is that it tends to be backward-looking, dependent on stale asset values and otherwise outdated balance sheet metrics.

The stress testing regime that grew out of the GFC sits between supervision and regulation, and helps to overcome some of the shortcomings often associated with each. Whereas supervision occurs under a veil of opacity, with the results known only to supervisors and bank leadership, the results of stress tests are released to the public. This can be particularly valuable during periods of stress, when market participants may have lost faith in the capacity of bank supervisors and when there is often a high degree of uncertainty about the value of bank assets and the health of individual financial institutions. And whereas capital regulation is typically backward looking, stress tests by their nature force banks and their regulators to look ahead, and require banks to hold sufficient capital—at least in theory—to be able to continue to lend and support activity in the real economy even in the face of a significant economic shock or downturn.

Stress tests vary not only depending on the context in which they are run, but also by who is running them—the regulator or the bank. Both exercises are useful, and there is a distinct value in running them in parallel. Discrepancies between a bank's assessment of how it will fare in a given adverse scenario and the regulators' assessment can reveal shortcomings in the bank's risk management capabilities, deficiencies in the model that the regulator is using, or both.

In theory, a bank should be better positioned than its regulator to assess how it will fare in a given scenario. It has higher quality information about its specific business lines and assets, and its risk management capabilities should constantly evolve and become more sophisticated in the domains where it is most active. Similarly, the technology a bank needs to undertake its own stress test should flow naturally from its risk management—a central component of any successful bank. The Federal Reserve updates its models every year, reflecting its own ongoing efforts at improving its procedures for the regulatory stress test, and the Fed has good incentives to try to create a robust model. At the same time, the models and assumptions that regulators use are standardized, even if applied in a way that is individualized to each bank.

Despite meaningful improvement over time, there remain significant deficiencies in how all bank stress tests are conducted. One of the most significant, and one which defines periods of widespread financial distress, are adverse feedback loops. This occurs when dysfunction in one market leads to liquidity hoarding by banks and other financial institutions contributing to dysfunction in other markets. The specifics may vary but can entail fire sales (the forced rapid sale of assets at a depressed value), contagion via interconnections or common exposures, higher spreads in secondary markets, and heightened volatility and reduced credit creation. Fear, uncertainty, and heightened risk aversion can further accentuate the dysfunction. These types of dynamics are very difficult to model. Although adverse feedback loops are common, the way that they arise and the nature of the interactions among different parts of a bank's operations, among banks, and between banks and nonbanks can vary significantly and be incredibly hard to predict. As a result, even when models seek to incorporate these dynamics—which they sometimes, but do not always seek to do—they typically rely on very coarse assumptions.

In failing to capture these dynamics, regulatory stress tests may yield overly optimistic results about the capacity of the banking system to withstand prolonged, adverse economic conditions. These challenges also mean that almost any model that a bank or regulator is using to undertake a stress test will involve significant and contestable assumptions. Notably, as the first round of stress tests illustrate, stress testing at times seeks to address these dynamics by short-circuiting them; when banks are adequately capitalized, they can more readily withstand losses without having them adversely impact their willingness or capability to provide other services.

In most countries, including the United States, stress testing is now an annual exercise often involving more than one scenario, although some are reducing the frequency to every other year. The way stress tests are run and the regulatory ramifications of stress tests have changed over time. Since they were first introduced in 2009, stress tests have been used to set individualized capital requirements, aiming to ensure each bank has adequate capital to remain healthy and active in the face of a given adverse shock—or requiring that they build up adequate capital to do so. As part of the 2024 stress tests, the Federal Reserve introduced a new addition to the stress testing process—an “exploratory analysis”—that tested how banks and the banking system would fare under a greater range of scenarios for purely informational purposes. By decoupling the stress test from a specific capital requirement, the exploratory analyses may make it easier for the Federal Reserve to be creative in the array of stress conditions it poses, providing it with potentially valuable new insights into the resilience of the system and sources of vulnerability.<sup>3</sup> In December 2024, the Federal Reserve announced its intention to undertake significant additional reforms to the stress testing process including proposed changes that are “not designed to materially affect overall capital requirements” suggesting continued interest in this type of exploratory stress testing.<sup>4</sup>

There has also been significant discussion in financial regulation about the use of reverse stress testing prompting the Federal Reserve to contemplate requiring banks to engage in reverse stress tests. In contrast to traditional stress testing, which provides insight into how a financial institution will fare in the face of a given set of adverse developments, reverse stress testing seeks to uncover the type and magnitude of adverse shocks that will cause a financial institution to fail. Used by regulators, reverse stress tests could help to reveal vulnerabilities in the banking system or the financial system more generally. Although there has been meaningful work on the topic—including exploration of how reverse stress tests may interact with stress testing using multiple scenarios—this is an example of the type of beneficial theoretical frontier that would be useful to reach, but where current technological limitations significantly impede its widespread usage.

## History of Stress Testing

Stress tests have long existed as a central part of the risk management toolkits of banks and other financial institutions undertaking internal efforts to identify and manage risks. Banks have

---

<sup>3</sup> Exploratory Analysis of Risks to the Banking System, Dodd-Frank Act Stress Test Publications, Board of Governors of the Federal Reserve System (June 2024) <https://www.federalreserve.gov/publications/June-2024-Exploratory-Analysis-of-Risks-to-the-Banking-System.htm#:~:text=The%202024%20exploratory%20analysis%20sheds,to%20the%20prior%20%20years.>

<sup>4</sup> See *supra* note 2.

also stress tested the balance sheets of their counterparties in setting limits and otherwise managing their exposures to other financial institutions. It was only in 2008, however, facing a financial system that was slow to heal despite significant government interventions, that horizontal, regulatory stress testing became a core component of the regulatory toolkit.

There were a number of conditions that enabled stress tests to emerge as an important regulatory tool for which there was no readily available equivalent. As a threshold matter, the persistency of the challenges plaguing the financial system at the time made bank regulators more open to experimentation. Although the depths of the GFC had passed—thanks to significant government intervention, including Congress providing the Treasury Department \$700 billion to address deficiencies in the financial system and its use of some of those funds to recapitalize all of the largest banks—signs of dysfunction and distrust continued. This is consistent with a common trend in financial regulation of crisis leading to major regulatory reform and innovation, and in turn prompting important questions about what might drive such developments absent the type of widely felt hardships inflicted by banking crises.<sup>5</sup>

This may help explain why bank supervisors—who are traditionally loathe to make their work public—opted to undertake a system-wide stress test of the largest banks to see how they might fare if conditions deteriorated and decided to make the results and general methodology public. Also critical was the fact that Congress had already vested the Treasury Department with funds that it could use to further recapitalize any bank if it had capital shortfalls it could not address through raising additional funds on its own. A primary reason that bank supervision occurs under a cloak of opacity is to prevent bank regulators from having to sugarcoat bad findings, aware that negative information could spark a run or otherwise harm an already weakened financial institution. The dire circumstances also likely made banks more amenable to the process, despite the unusual nature of the process and what they now claim to be legal deficiencies in the regime.

The conditions also shaped the design of the original stress test along multiple dimensions. A paper by three economists at the New York Fed who were deeply involved in the original round of stress tests explains many of the key design decisions and why they were so important. “A key objective of the [initial supervisory stress test] was to curb the negative feedback loop of losses, actual and fearfully anticipated, curbing credit provision, which could in turn curtail real macroeconomic activity, making banks yet more reluctant to lend into a rapidly declining economy.”<sup>6</sup> Requiring banks to increase their capitalization to levels that would allow them to “withstand a severe macroeconomic scenario, the goal was to make the realization of that scenario less likely.”<sup>7</sup>

---

<sup>5</sup> The extent to which financial regulation is crisis driven is a topic of active scholarly debate. Although no one disagrees that crises play a major role, there is a robust discussion about how much regulatory change occurs outside such periods, with some evidence suggesting reforms occur both during and outside of periods of crisis. See Peter Conti-Brown & Michael Ohlrogge, *Financial Crises and Legislation* (September 23, 2022), <https://elischolar.library.yale.edu/journal-of-financial-crises/vol4/iss3/1/> and sources cited therein.

<sup>6</sup> Beverly Hirtle, Til Schuermann & Kevin Stiroh, *Macprudential Supervision of Financial Institutions: Lessons from the SCAP*, Federal Reserve Bank of New York Staff Reports No. 409 (November 2009), [https://www.newyorkfed.org/medialibrary/media/research/staff\\_reports/sr409.pdf](https://www.newyorkfed.org/medialibrary/media/research/staff_reports/sr409.pdf) at 6.

<sup>7</sup> *Id.*

To achieve this outcome, however, the tests had to be comprehensive along multiple dimensions. For one thing, the tests had to capture a sufficiently large swathe of the banking sector as to make it believable that having sufficient capital at the tested banks would suffice to preclude adverse feedback loops in the banking system generally. The Fed achieved this while still acting quickly by setting a size threshold and testing all bank holding companies above that threshold, allowing it to capture two-thirds of the assets in the banking system. Including more banks would have enhanced the credibility and impact, but could also have stretched already depleted regulatory resources beyond what they could bear. The initial stress tests also needed to be comprehensive across the range of further adverse developments that could materialize and the possibility of interactions across different types of assets. It achieved this by “assessing risk across the full range of the [bank’s] portfolios and activities” and projecting losses “based on a whole-firm view, rather than focusing on specific business line, risk exposures, or segment as in traditional horizontal examination.”<sup>8</sup> Again, however, scarce resources and time pressure meant there were limitations: operational risk, liquidity risk, and other risks were not incorporated into the stress test that would have made the exercise far more difficult to undertake. Another key factor related to process design which helped inform the assumptions used was the effort to harness multiple different perspectives and different types of expertise, each of which could raise different concerns and provide distinct insights.

The net result at the time was exceptionally positive. As then Fed Chair Ben Bernanke noted: “The SCAP stands out ... as one of the critical turning points in the financial crisis. It provided anxious investors with something they craved: credible information about prospective losses at banks.”<sup>9</sup> Subsequent analyses support his conclusion that the initial and subsequent rounds of stress tests provided the market credible and useful information about the health of individual banks and the banking system.<sup>10</sup>

The Dodd-Frank Act embedded stress tests as a permanent feature of bank regulation and supervision in the United States, and the Federal Reserve continued to use its authority to conduct tests that went beyond those mandated by the Dodd-Frank Act. The role they play in setting capital has evolved over time, merging into the layered capital requirements imposed on all of the largest banks through a “stress capital buffer.”

The use of stress tests to determine capital requirements has meaningful benefits and drawbacks. By aspiring to ensure banks have adequate capital to not only avert failure, but also continue to lend and support economic activity during periods of widespread distress, the use of stress test results can produce capital requirements better calibrated to achieve desired aims than ones reliant on more coarse, backward-looking metrics. On the other hand, this has made it more vulnerable to legal attack, and even apart from the legal issues, may well have contributed to less dynamism in the adverse scenarios, reducing the information production that comes from stress testing. Banks also have complained about the accuracy of the results, and have been particularly

---

<sup>8</sup> *Id.* at 7.

<sup>9</sup> Bernanke, Ben S. 2013. “Stress Testing Banks: What Have We Learned?” Speech at the “Maintaining Financial Stability: Holding a Tiger by the Tail” financial markets conference sponsored by the Federal Reserve Bank of Atlanta, Stone Mountain, Georgia. <https://www.federalreserve.gov/newsevents/speech/bernanke20130408a.htm>

<sup>10</sup> [https://www.newyorkfed.org/research/staff\\_reports/sr744.html](https://www.newyorkfed.org/research/staff_reports/sr744.html)

inclined to challenge the quality of the models used when a bank's internal assessment suggests it would fare better than the prediction offered by the Fed's model in a given adverse scenario.

The legal issues clouding the stress testing landscape in the United States are significant, and have become more pressing as a result of the overall shift in the judiciary's approach to administrative authority. Exacerbating the legal pressure has been a long-time concerted effort by the largest bank trade groups to raise these issues. In July 2023, for example, the Bank Policy Institute and American Bankers Association submitted to the Federal Reserve a petition under "section 553(e) of the Administrative Procedure Act to engage in rulemaking to seek public comment on, and codify by rule, any and all models, formulas, or other decisional methodologies that the Board uses to calculate the "stress capital buffer requirement." The BPI and ABA further argue that all of the referenced materials should also be released pursuant to the Freedom of Information Act and they argue that any scenario on which a capital requirement is based should also go through notice-and-comment rulemaking

The thrust of the legal claim is that because these models are being used to set binding capital requirements, they are legislative rules that should go through the trans-substantive standard procedures for the issuance of such rules. In establishing the "stress capital buffer" regime the Federal Reserve did indeed go through a notice-and-comment process, but the open question is whether it must go further and undertake that type of process for all of the models and specific scenarios it uses. This would likely be counterproductive in efforts to use stress tests for a number of reasons. One of the most cited concerns is that making the models public would allow banks to reverse engineer the models, making it easier for them to abide by the letter while finding ways to evade the spirit of the regime. Another concern, common to rulemaking generally, is that the comment process would be one-sided, with the most informed comments coming from industry and their representatives, potentially skewing the outcome toward one that is more industry friendly. A final concern, also common to rulemaking, is that requiring the models to go through such a process would introduce rigidity and discourage learning and evolution—just the opposite of what should be the common aim in this domain.

The BPI and ABA had raised these issues previously and backed away, but in December 2024, they and other interested parties filed a lawsuit pressing some of these claims once again. Each of these groups has also been very engaged in pushing back against another recent regulatory initiative by the Fed, and the overall heightened willingness of the judiciary to construe legislative authority more narrowly—and procedural obligations more robustly—in recent years could incline them toward further action. This may help to explain the decision by the Federal Reserve to introduce the exploratory analyses in 2024, and it certainly explains the decision to undertake a more substantive review of the stress testing process.

### **International coherence**

Much bank regulation stems from coordinated international efforts. Stress testing is part of this overall effort, but also remains distinct in how it is implemented in each jurisdiction. The very fact that so many jurisdictions now use horizontal, regulatory-run stress tests of the largest banks in their jurisdictions is a sign of the extent of coordination among them. At the same time there are meaningful differences in how they are run.



Since the 1980s, the world's leading economies have joined together to promulgate an evolving set of regulatory and supervisory expectations through the Basel Committee on Banking Supervision. These include both standards that should be incorporated into laws and regulations (such as capital and liquidity requirements), and best practices. As part of these efforts, the Basel Committee issued in 2009, and updated in 2018, a set of guidelines for the use of stress testing by regulators and financial institutions.<sup>11</sup> As with most of the work product coming from the Basel Committee, these guidelines are the product of efforts undertaken by a consortium of 28 central banks, with the largest jurisdictions having the most influence in shaping the work product. That being said, because they were not even softly binding in a manner akin to the standards embedded in Basel III and earlier guidelines, there was contention and less effort to devise uniform expectations. While most jurisdictions have now incorporated regulatory stress testing into their regimes, the manner for doing so, the stress scenarios they use, and other dynamics continue to vary among jurisdictions.

Despite varying practices, however, the role of Basel gatherings and the many other international settings where policy makers meet (often with academics and members of industry) play an important role facilitating at least some degree of cross-jurisdictional learning, coordination, and communication. These settings and the relationships and discourse they engender do promote common language and learning, but have yet to consistently lead to common practices.

For example, in connection with updating its guidelines, the Basel Committee surveyed more than 50 financial institutions across more than 20 jurisdictions to learn more about how stress testing was evolving and how it was being used. The results highlighted the degree of variation in how regulators and banks use stress tests. For example, with respect to scenario design, the survey found that “[s]tress testing methodologies used by supervisors vary in terms of objectives, time horizon, risks analyzed, models, level of granularity, among other aspects.”<sup>12</sup> At the same time, in publishing the results of this survey, the Basel Committee promoted cross-jurisdictional learning and coordination.<sup>13</sup> As another example, in 2019, the International Monetary Fund hosted a conference which brought together the authors of more than 20 separate chapters in a forthcoming volume on stress testing—promoting learning across jurisdictions but also among academics, industry participants, and policymakers.<sup>14</sup>

Some variation across regimes, however, is likely to remain, and may be for the best. At times, this provides helpful lessons, even if learned the hard way. For example, while the initial round of stress tests coming out of the GFC were exceptionally helpful in the United States, the first round of European stress tests had little positive impact. Many attributed this to the fact that in contrast to the United States, European regulators did not have a readily available mechanism for addressing any deficiencies they might identify, and absent the capacity to threaten or promise a government-backed recapitalization, they were understandably less rigorous—and seen to be less

---

<sup>11</sup> Stress Testing Principles, Basel Committee on Banking Supervision, Bank for International Settlements (October 2018), <https://www.bis.org/bcbs/publ/d450.htm>.

<sup>12</sup> Supervisory and Bank Stress Testing: Range of Practices, Basel Committee on Banking Supervision, Bank for International Settlements (December 2017), <https://www.bis.org/bcbs/publ/d427.pdf> at 6.

<sup>13</sup> *Id.*

<sup>14</sup> IMF Conference on Rethinking Financial Stability: The FSAP at 20 (October 14-15, 2019), <https://www.imf.org/en/News/Seminars/Conferences/2019/09/20/rethinking-financial-stability-the-fsap-at-20>. I was a participant at the conference and contributed to the handbook that grew out of it.

rigorous—largely mitigating the intended aim. More generally, the existing regulatory architecture, the relative role of prudentially regulated banks in providing loans and other financial services, and other variations across jurisdictions cast doubt on whether the tests should converge across jurisdictions.

### **Lessons learned and recommendations**

Stress testing has been and should always remain a work in progress. Changes in the structure of the financial system, the relative capacity of other financial regulatory tools, and the capacity of the models used for stress testing to replicate real-life are among the constantly evolving dynamics that ensure stress testing—if it is to remain robust and useful—will continue to evolve and change. Yet regulatory stress tests of some sort seem likely to remain a staple of bank regulation and supervision going forward. Stress tests were one of, if not the most, important regulatory innovations to grow out of the GFC. They can help bank and other financial regulators enhance the safety and soundness of the financial institutions they oversee and the resilience of the broader financial system.

One reason for this is that regulatory stress tests have an important informational role to play. By enabling regulators to test how individual banks and the bulk of the banking system would fare in a given adverse scenario, stress tests provide regulators valuable new information about where vulnerabilities might lie, commonalities and differences in risk exposures across the financial system, and the capacity of that system to weather different types of shocks. Similarly, using stress test results to shape capital requirements helps overcome meaningful deficiencies in the typically backward-looking nature of traditional requirements. More generally, requiring banks and their regulators to regularly focus on bad states of the world should have beneficial effects in enhancing the capacity of banks and regulators to be prepared when bad times strike.

At the same time, significant challenges remain. Stress tests may well withstand further legal scrutiny, but that remains uncertain. Just as significant as the merits of the legal issue are the changes the Federal Reserve seems poised to undertake in a preemptive effort to accommodate the concerns raised by industry. The current challenges seem primed to bring about a major re-design in the regulatory stress testing process in the United States. How this is done will determine how much and which of the benefits of stress testing will be retained. Even if not forced to change, reducing the tight nexus between stress testing and capital—as done with the exploratory analyses—could promote more creativity and dynamism in the range of scenarios tested, helping to restore the information-production component of stress testing and helping to make them less routine. And technology, embodied for example in the models used, remains far from perfect, allowing meaningful room for further improvement.

A core lesson from banking is that the function, quality, and impact of stress testing cannot be understood independent of the background regulatory regime. For one thing, despite having considerable expertise and experience in bank regulation and banking crises, by the time bank regulators started stress testing banks, banks had already been engaging in their own stress testing. This private and public expertise, and the capacity to implement in a setting that was already so heavily regulated played a significant role shaping how quickly and powerfully the regime could take hold. This was aided by the fact that banks and the financial industry were facing so many challenges during the GFC, and that they had their own reasons for wanting the stress tests to help

bring about a shift. Yet as recent developments reflect, stress tests are also subject to shifting regulatory winds generally, and those right now are blowing to reduce the rigor of regulation.

Another reason that the regulatory background setting is key is that many of the aims of stress testing in banking were developed in response to shortcomings in the existing regulatory and supervisory regimes. Among other things, this helps to explain at least some of the tensions among the information-producing aims and the capital regulation aims of the stress testing exercise. A core lesson may be the value of trying to achieve some consensus as to the objectives of stress testing, recognizing that some aims may only be achieved only at the expense of others. This lesson also raises questions about the marginal gain that stress tests provide relative to other available tools for promoting related aims.

### References

**David Aikman, Romain Angotti, Katarzyna Budnik** Stress testing with multiple scenarios: a tale on tails and reverse stress scenarios, ECB Working Paper Series No 2941, <https://www.ecb.europa.eu/pub/pdf/scpwps/ecb.wp2941~28e2ec1e42.en.pdf>

**Alves, Ricardo Jorge Santos.** 2013. “Information Value of EU-Wide Stress Tests: How did Markets React to Stress Test Results.” Master’s Dissertation for the ISCTE Business School, Instituto Universitario de Lisboa. Supervisor: Prof. Doutor José Joaquim Dias Curto. <https://repositorio.iscte-iul.pt/bitstream/10071/9900/1/INFORMATION%20VALUE%20OF%20EU-WIDE%20STRESS%20TESTS.pdf>

**Basel Committee on Banking Supervision,** Supervisory and bank stress testing: range of practices December 2017, <https://www.bis.org/bcbs/publ/d427.pdf>

**Basel Committee on Banking Supervision,** Stress testing principles, 2018, <https://www.bis.org/bcbs/publ/d450.htm>

**Bernanke, Ben S.** 2013. “Stress Testing Banks: What Have We Learned?” Speech at the “Maintaining Financial Stability: Holding a Tiger by the Tail” financial markets conference sponsored by the Federal Reserve Bank of Atlanta, Stone Mountain, Georgia. <https://www.federalreserve.gov/newsevents/speech/bernanke20130408a.htm>

**Borio, Claudio, Mathias Drehmann, and Kostas Tsatsaronis.** 2014. “Stress-Testing Macro Stress Testing: Does it Live Up to Expectations?” *Journal of Financial Stability* 12: 3-15. [https://ac.els-cdn.com/S1572308913000454/1-s2.0-S1572308913000454-main.pdf?\\_tid=47cb6dfe-7ce2-42d3-b95a-](https://ac.els-cdn.com/S1572308913000454/1-s2.0-S1572308913000454-main.pdf?_tid=47cb6dfe-7ce2-42d3-b95a-)

**Carmassi, Jacopo, and Richard Herring.** 2016. “The Corporate Complexity of Global Systemically Important Banks.” *Journal of Financial Services Research* 49 (2-3): 175-201. <https://link.springer.com/article/10.1007/s10693-016-0251-4>

**Dudley, William C.** 2011. “US Experience with Bank Stress Tests.” Remarks by Mr. William C. Dudley, President and Chief Executive Officer of the Federal Reserve Bank of New York, at the Group of 30 Plenary Meeting of 28 May 2011, Bern, Switzerland. <https://www.bis.org/review/r110704c.pdf>

**Ellahie, Atif.** 2012. “Capital Market Consequences of EU Bank Stress Tests.” <https://www.newyorkfed.org/medialibrary/media/research/conference/2012/FinancialServices2012/Ellahie.pdf>

**European Central Bank.** 2010. “Questions and Answers: 2010 EU-Wide Stress Testing Exercise.” Press Release 23 July 2010. <https://www.ecb.europa.eu/pub/pdf/other/euwidestresstestingexercise-qaen.pdf?860c12759915bf0e3a8cd5486cb595a4>

**Financial Stability Oversight Council.** 2017. “2017 Annual Report.” Washington, D.C. [https://www.treasury.gov/initiatives/fsoc/studies-reports/Documents/FSOC\\_2017\\_Annual\\_Report.pdf](https://www.treasury.gov/initiatives/fsoc/studies-reports/Documents/FSOC_2017_Annual_Report.pdf)

**Frame, W. Scott, Kristopher Gerardi, and Paul S. Willen.** 2015. “The Failure of Supervisory Stress Testing: Fannie Mae, Freddie Mac, and OFHEO.” Federal Reserve Bank of Boston Working Paper No. 15-4. <https://pdfs.semanticscholar.org/048a/0dbc65109908151db462218bc28e85aa89ea.pdf>

**GAO Report.** 2009. Troubled Asset Relief Program, Status of Efforts to Address Transparency and Accountability Issues, GAO-09-296. <https://www.gao.gov/new.items/d09296.pdf>

**Geanakoplos, John.** 2010. “The Leverage Cycle.” Cowles Foundation Discussion Paper No. 1715R. <https://cowles.yale.edu/sites/default/files/files/pub/d17/d1715-r.pdf>

**Goldstein, Itay, and Yaron Leitner.** 2017. “Stress Tests and Information Disclosure.” *Journal of Economic Theory* 177: 34-69; Federal Reserve Bank of Philadelphia Working Paper No. 17-28. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3029761](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3029761)

**Goldstein, Itay and Haresh Sapra.** 2013. “Should Banks’ Stress Test Results be Disclosed? An Analysis of the Costs and Benefits.” *Foundations and Trends in Finance* 8 (1): 1-54. <http://finance.wharton.upenn.edu/~itayg/Files/stresstests-published.pdf>

**Gorton, Gary.** 2012. Misunderstanding Financial Crises: Why We Don’t See Them Coming.

**Greenwood, Robin, Samuel G. Hanson, Jeremy C. Stein, and Adi Sunderam.** 2017. “Strengthening and Streamlining Bank Capital Regulation.” Brookings Papers on Economic Activity, BPEA Conference Drafts, September 7-8, 2017. [https://www.brookings.edu/wp-content/uploads/2017/09/6\\_greenwoodetal.pdf](https://www.brookings.edu/wp-content/uploads/2017/09/6_greenwoodetal.pdf)

**Hirtle, Beverly, Til Schuermann and Kevin Strieth.** 2009. “Macroprudential Supervision of Financial Institutions: Lessons from the SCAP.” Federal Reserve Bank of New York Staff Report No. 409, November 2009. [https://www.newyorkfed.org/medialibrary/media/research/staff\\_reports/sr409.pdf](https://www.newyorkfed.org/medialibrary/media/research/staff_reports/sr409.pdf)

**Judge, Kathryn.** “Stress Testing in Times of War” in the Financial Stress Testing Handbook, J. Doyne Farmer, Alissa Kleinmijhuis, Til Schuermann & Thom Wetzer (Eds.), Cambridge University Press (2022).

**Morrison, Alan D., and Lucy White.** 2013. “Reputational Contagion and Optimal Regulatory Forbearance.” *Journal of Financial Economics* 110: 642-58.

[https://www.hbs.edu/faculty/Publication%20Files/White\\_Lucy\\_J2\\_Reputational%20Contagion%20and%20Optimal%20Regulatory%20Forbearance\\_21ac1359-176d-4c2b-9344-5cc6407948e2.pdf](https://www.hbs.edu/faculty/Publication%20Files/White_Lucy_J2_Reputational%20Contagion%20and%20Optimal%20Regulatory%20Forbearance_21ac1359-176d-4c2b-9344-5cc6407948e2.pdf)

**Peristiani, Stavros, Donald P. Morgan, and Vanessa Savino.** 2010. "The Information Value of the Stress Test and Bank Opacity." *Journal of Money, Credit and Banking* 46 (7) (October) 1479-500; Federal Reserve Bank of New York Staff Report No. 460, July 2010.  
[https://www.newyorkfed.org/research/staff\\_reports/sr460.html](https://www.newyorkfed.org/research/staff_reports/sr460.html)

**Sarin, Natasha and Til Schuermann,** Stress Testing Lessons from the Banking Turmoil of 2023 (April 29, 2024), <https://ssrn.com/abstract=4811157> or  
<http://dx.doi.org/10.2139/ssrn.4811157>

**Tarullo, Daniel K.,** Reconsidering the Regulatory Uses of Stress Testing, Hutchins Center Working Paper #92 (2024), <https://www.brookings.edu/articles/reconsidering-the-regulatory-uses-of-stresstesting>