

Beverly Hirtle: The past and future of supervisory stress testing design

Remarks by Ms Beverly Hirtle, Executive Vice President and Director of Research of the Federal Reserve Bank of New York, at the 2018 Federal Reserve Stress Testing Research Conference, Federal Reserve Bank of Boston, Boston, 9 October 2018.

* * *

Good afternoon. It is a real pleasure for me to be here today at the fifth annual conference on stress testing research. As many of you know, I spent a considerable portion of my career working on stress testing issues, starting with the Supervisory Capital Assessment Program (SCAP) in 2009, through the initial implementation of the Comprehensive Capital Analysis and Review (CCAR) and the Dodd-Frank Act (DFAST) stress testing programs, as well as my personal research agenda on the impact of stress testing. And even now, though my responsibilities have changed, I remain actively engaged in debates and discussion as the CCAR and DFAST programs evolve, and as the details of stress test modeling and design evolve with those programs.

In fact, I want to make this evolution the focus of my remarks today. I will begin by briefly reviewing how stress testing emerged as a key supervisory policy tool, and talk about how the original goals of the SCAP, CCAR and DFAST stress tests affected modeling and design choices. I'll then talk about some of the consequences of those choices – in particular, what is missing or not well-captured in the current approach. Finally, I'll suggest some areas where new research could help push forward the frontier of stress test modeling, both of specific elements that go into the stress test calculation and of the broader impact of the stress testing programs implemented here in the United States and elsewhere. And of course, I need to note that my remarks today are my own and do not necessarily reflect the views of the Federal Reserve Bank of New York or the Federal Reserve System.

I believe that stress testing and the CCAR program were the most significant and impactful regulatory and supervisory changes coming out of the global financial crisis. Why do I feel that way?

To begin, it helps to remember the prevailing environment that generated the first broad-based supervisory stress tests – the SCAP – in early 2009. It was a period of significant uncertainty – uncertainty about the stability of funding markets, about losses at individual institutions and about losses in the financial system overall. There was a growing cottage industry in estimating global and U.S. financial sector losses on mortgages, mortgage-backed securities and other positions, with projected global losses ranging as high as \$4 Trillion.¹ Equity prices had plummeted, especially for banks, with declines running far ahead of falls in regulatory capital ratios. As of September 2008, the tier 1 capital ratios of the five largest U.S. bank holding companies ranged between 7.5 and 8.9 percent, well above the then-prevailing 4 percent minimum requirement. Many large banks continued to pay dividends well into 2008, even after government support was provided to the banking industry via the Troubled Asset Relief Program (TARP), something noted in the press at the time.² Some have argued that these continued dividend payments reflected intentional risk-shifting from equity to debt holders,³ though many institutions had stopped making share repurchases months before,⁴ which does not seem consistent with risk-shifting behavior. Alternatively, in an environment of considerable uncertainty, banks may have been concerned with the negative market signal that a dividend reduction might send, hesitating to be the first to take this step.

Supervisory stress testing emerged as a policy solution to address both backward-looking regulatory capital ratios and possible first-mover problems facing banks during periods of stress.

By their nature, stress tests are designed to ask “what if” in a forward-looking way. Stress tests aren’t a prediction of future events but rather hypothetical exercises intended to assess the robustness of a bank’s resources against various potential future bad outcomes. In this way, they provide additional insight into the true degree of capital adequacy inherent in a bank’s *current* capital position and the risks embedded in its balance sheet, including its ability to continue to make capital distributions in both baseline and stressed economic conditions.

Further, by requiring banks to raise additional capital if their stressed capital ratios fell below target levels, the SCAP addressed potential first-mover problems associated with banks having to individually decide to address capital shortfalls – though there was considerable concern at the time about how market participants and bank counterparties would view public supervisory assessments indicating that certain banks needed to raise additional capital. In fact, all but one of the banks required to raise new capital were able to do so, some of them in amounts greater than the SCAP requirements, suggesting that the SCAP created a “safe” environment for banks to enhance their capital positions. Following the SCAP, the CCAR program embedded supervisory stress test results into a broader assessment of a bank’s internal capital planning, providing supervisors both the empirical tools and the regulatory authority to limit distributions in the face of deteriorating economic and financial market conditions.

After many years of SCAP and CCAR stress tests, the empirical approach embedded in the calculations might feel pre-ordained, but in fact there were many design choices made along the way. These include the focus on regulatory capital ratios, the horizon of the stress test and the calculation of the maximum impact of the stress test within this horizon. It’s helpful to review these design choices in light of the goals of the supervisory stress testing regime and to highlight which banking sector vulnerabilities are well addressed, and not so well addressed, by the tests.

In reviewing the design choices, it’s also helpful to distinguish between choices about *what* the SCAP, CCAR and DFAST stress tests were designed to capture and *how* the models measure key elements of the stress test calculations.

The first and most significant design choice involved *what* the tests would measure: the SCAP and subsequent U.S. supervisory stress tests measure the impact of the stress scenario on regulatory capital ratios. Other choices were possible, of course. The tests could have focused on modeling the market value of assets or equity (similar to the Capital Shortfall approach developed by Acharya et al.⁵), or attempted to measure the value or cost of debt, or the impact on credit ratings. But regulatory capital ratios were a natural choice, allowing the output of the stress testing exercise to map into the broader regulatory capital regime while addressing the slow adjustment of those ratios during downturns.

The focus on regulatory capital ratios resulted in a cascade of additional design choices. To begin, since regulatory capital ratios are based on accounting definitions of earnings and capital, estimating net income and its flow through to regulatory capital became the focus of the stress test calculations. This stands in contrast to approaches that concentrate solely on projecting losses, which were the focus of much of the public discussion prior to the release of the SCAP results. In practical terms, the net income approach involves recognizing that banks will have non-credit-related expenses and earn revenues even during periods of stress. Modeling so-called pre-provision net revenue (PPNR) was one of the innovations introduced in the SCAP and represented one of the key challenges in that exercise – and, I will argue, is a continuing challenge today.

Given the choice to model net income and regulatory capital ratios, another set of design choices concerns the horizon over which net income is calculated and how cumulative changes in net income over this horizon are incorporated into regulatory capital ratios. The SCAP assumed a two-year forward horizon, while the subsequent CCAR and DFAST stress tests have assumed a 9-quarter forward horizon.⁶ Other supervisory stress test regimes have made different assumptions – for instance, the European stress tests use a three-year forward horizon.⁷ An

alternative choice would be to project losses and income over the full lifetime of a bank's loans and other assets. Estimating lifetime losses on particular fixed portfolios of loans and securities was arguably the prevalent approach during the financial crisis period.⁸ But in a stress testing context, the lifetime loss approach presents practical difficulties given the varying maturities of different types of assets, as well as the greater uncertainty inherent in more distant future points. At the time of the SCAP, we judged that a two-year horizon was long enough to capture significant losses while remaining within what could reasonably be projected conditional on the macroeconomic scenario.⁹

Closely related to the scenario horizon question is how cumulative losses and revenues are incorporated to calculate post-stress capital ratios. The DFAST and CCAR stress tests adopt a "walk-through-time" approach, in which net income and capital ratios are calculated on a quarter-by-quarter basis throughout the stress test horizon.¹⁰ An alternative approach would be to apply cumulative losses and revenues in a single, instantaneous shock – this was what was done in the SCAP. While the two approaches might be assumed to produce roughly consistent results, the walk-through-time approach may better capture true vulnerabilities, especially in cases where negative net income quarters come early in the stress test horizon but revenues accrue most strongly in subsequent quarters. The instantaneous shock approach implicitly recognizes revenues from later in the horizon as offsets to losses that occur early in the horizon and thus may overstate true capital resources. The tradeoff is that the walk-through-time approach implies a degree of precision about the quarter-by-quarter pattern of losses and revenues that might be beyond the capabilities of existing modeling technology (this was the judgment at the time of the SCAP).

But the walk-through-time approach requires further assumptions, particularly regarding the evolution of the balance sheet. These assumptions range from how to capture maturing loans and securities; what to assume about prepayments; what new lending occurs over the horizon (including replacement of maturing or prepaid positions); and whether to assume that the bank is actively managing its exposures by additional hedging, changes in the credit composition of its portfolios, or restriction of credit supply.

Although some of the implementation details differ, the SCAP and DFAST/CCAR supervisory stress tests made broadly similar assumptions on these issues. The guiding assumption is that a bank's balance sheet and the nature of its exposures do not change over the stress test horizon. The supervisory estimates do not embed behavioral responses to the stress scenario, including additional risk mitigation or changes in business strategy.¹¹ Further, maturing positions are assumed to be replaced on a like-for-like basis, maintaining the overall credit characteristics of the starting-date portfolios.¹² Importantly, banks are assumed to maintain credit supply even as economic conditions in the hypothetical scenario deteriorate. This is a policy judgment reflecting the objectives of the supervisory stress test regime.¹³ The implication of these choices is that the stress test results are conditional not just on the stress scenario but also on the policy choices embedded in the exercise, rather than being an historically unbiased projection of banks' behavior under stressed economic and financial market conditions.

These assumptions have been implemented in different ways in different supervisory stress test exercises: in SCAP, the balance sheet was assumed to be fixed over the stress test horizon; in the recent DFAST and CCAR stress tests, supervisory models have been used (these models actually tend to result in growth in the balance sheet over the stress test horizon)¹⁴; and in the currently proposed Stress Capital Buffer (SCB) approach, the balance sheet would once again be held fixed.

My discussion thus far has focused on *what* the stress tests measure. But there is a second set of design choices about *how* the stress scenario impacts will be modeled. To begin, capital ratios are measured separately for individual institutions, with no attempt to capture dynamic

interactions, spillovers or specific cross-firm exposures. Similarly, outcomes for the banking sector do not interact dynamically with the macroeconomic scenarios, which are inputs to the calculations. As a result, the calculations are “stand alone” in the sense that the outcomes for any individual institution are independent of those for others in the set of stress-tested firms. Results for the banking system are therefore simply the sum of results for individual institutions, rather than capturing interactions among them.

The stand-alone approach reflects the microprudential objectives of the various U.S. supervisory stress test programs, along with practical considerations. While the SCAP and CCAR/DFAST stress tests programs have strong macroprudential elements¹⁵, in each case the results have also been used to make firm-specific supervisory judgments about capital adequacy. This individual firm application pushes towards measurement approaches that capture more detailed, firm-specific characteristics. At the same time, fully dynamic models that capture interactions and exposures among firms or that build in feedback from banking sector outcomes to the macroeconomic scenario can be informationally demanding, computationally intensive and, at the stage of model development prevailing during the SCAP and early phases of the CCAR/DFAST regime, generally would have involved higher level, “industry average” assumptions about the performance of banks under the hypothetical stressed conditions. Thus, the stand alone approach was in many ways the only practical choice at the time the initial supervisory stress tests were conducted.

A final set of design choices affecting how stressed capital ratios are measured involves the actual models that produce the estimates of net income and capital. The models used in the CCAR and DFAST supervisory stress tests are developed and implemented by supervisors – they are almost entirely independent of stress test models used by banks. This is something that has evolved over time. The SCAP results were based on a variety of sources, including projections made by the banks, simple supervisory models, and historical data on bank performance. The blending of sources reflected that the SCAP was the first broad-based supervisory stress testing program, conducted in a crisis environment, with significant learning along the way.

Over time, supervisory estimates have become increasingly independent from the projections made by banks. Supervisors and economists in the Federal Reserve System have developed models to capture the impact of the stress scenarios on various elements of net income, including different categories of loans and securities, interest and non-interest income, non-credit expenses and balance sheet evolution, as well as the flow through of net income to regulatory capital. This was a policy choice intended to provide consistency of treatment across banks involved in the stress tests and to counteract any incentives banks might have to understate the impact of the stress scenario. Bank-specific elements – which, as noted above, are critical to the microprudential policy objectives of the CCAR and DFAST programs – are introduced by using detailed data provided by each participating institution. Thus, results for individual banks vary based on the characteristics of their balance sheet and business focus, but not due to differences in the way particular net income or capital elements are modeled.

Other choices are possible, of course, principally the choice to make greater use of bank-generated stress test results as a way of capturing institution-specific factors, as is done in supervisory stress testing programs in some other jurisdictions. The European stress tests, for instance, are based on estimates made by the participating banks, subject to a variety of constraints on modeling assumptions and detailed review by supervisors.¹⁶

What I hope you take away from this review of the *what* and the *how* of stress testing is that the SCAP, CCAR and DFAST programs have involved a series of inter-connected design choices, most of which were intended to address the objectives of the stress test exercises, but none of which were inevitable or pre-ordained. Collectively, these design choices have implications for what the stress tests capture, as well as what they miss. I want to share a few thoughts about

the latter before turning to some suggestions for how additional research could help.

The most obvious place to start discussion of “what’s missing?” is to note that the SCAP, CCAR and DFAST programs are by design capital stress tests and do not directly assess other areas of institutional or financial sector vulnerability, such as liquidity, funding or firesale risks. The current supervisory stress testing regime does address these risks indirectly, to the extent that a banking sector with more robust capitalization is less likely to experience liquidity stresses, runs and the resulting firesales. And, through the Federal Reserve’s Supervisory Liquidity Analysis and Review (CLAR) program, large complex banking companies are subject to separate supervisory stress testing of their liquidity resources. But the current regime does not explicitly integrate capital and liquidity stress testing at the firm level, nor does it attempt to assess the probability or impact of a firesale of banking sector assets. Nor, given the focus on capital, does the regime address firm-specific or systemic issues related to resolution, when the shock to capital is large enough that the bank can no longer survive.

In large part, the focus on individual institutions – the “stand-alone” design choice – accounts for this outcome. As already noted, the CCAR/DFAST stress tests do not incorporate cross-bank exposures in a dynamic way, nor is there feedback between the outcomes for the banking sector and the macroeconomic scenario. A stress testing regime that incorporated these feedbacks, including the potential for liquidity pressures, bank runs, and firesale risk, would be considerably more complex to implement, and would likely involve a greater degree of abstraction and simplification – and thus less “precision” – than the current supervisory capital stress tests.

In that regard, the drive for precision and accuracy at the individual institution level – and the resulting complexity of the supervisory models and extensive firm-specific data inputs needed to run the models – has created other challenges. Perhaps the most notable of these is that generating the supervisory projections is resource- and time-intensive, which limits the number of scenarios that can be assessed during any particular CCAR cycle. Some have suggested that stress testing would be more effective if many supervisory scenarios were examined, instead of just a few. Any individual scenario could miss important risk exposures at individual banks or in the sector as a whole. Turning that thought around, a single scenario might not uncover true capital vulnerabilities at all institutions. The bank-generated scenarios that are part of the CCAR program are certainly an important channel to address concerns about idiosyncratic risk, as banks are meant to self-identify the scenarios that would be especially stressful for them. But running multiple common supervisory scenarios could provide a more robust assessment of the capital strength of the broader banking industry, at least as represented by the firms taking part in the CCAR and DFAST programs.

How can research help address these issues and, more generally, improve the overall supervisory stress testing regime? I want to suggest two general areas of research – one tactical and the other much bigger picture and strategic.

The tactical area is research that helps improve supervisory models. First, developing additional ways of projecting revenues and non-credit expenses in stressed environments is a particularly ripe area for additional work, in my view. As I noted, an important innovation in the SCAP was the recognition that a comprehensive stress test focused on net income needed to incorporate projections of interest and non-interest income and non-interest expense. While a number of models already existed for making such projections – for instance, those used by banks in their budgeting and planning processes – these were nearly all calibrated to produce projections assuming business as usual conditions, rather than the stressed environment assumed in the SCAP. The SCAP PPNR projections were thus based on stylized supervisory models that used available historical data on bank performance during recessions. Since then, the supervisory PPNR models have become considerably more sophisticated, but at their heart, they continue to rely on historical outcomes for revenues and expenses, rather than fully incorporating the fundamental drivers of those outcomes at the business line level. Additional research and data

collection on these fundamental drivers and their performance under stress could make the PPNR projections more robust to changing business strategies and focus.

A second tactical research area concerns measuring model risk. To begin, how do we assess errors from models intended to capture performance under stressed conditions when those conditions have not yet been realized and might not be in the historical data? How can we assess the uncertainty or margin of error around loss and revenue projections derived from models that can be quite complex, often involving multiple estimation steps? Further, since the final stress test calculations are derived from the combination of projections from many different (often multi-step) models, how do we assess the error around the ultimate calculation of stressed capital ratios? How do we measure correlation in model errors in a tractable and practical way? How much model risk owes to the decision to develop complex models for many individual pieces of the net income and regulatory capital ratio calculations instead of using simpler, “top down” estimation approaches?

A final tactical research area involves this last point – the role of simpler, easier to estimate models of net income and its key components. Federal Reserve modeling teams have already developed a set of these “benchmark” models that produce loss and revenue estimates as a comparison to the projections from the more sophisticated and complex production models. Creating more of these models – including multiple benchmarks for an individual loss, expense or revenue component, benchmarks that cover the full range of detailed net income elements, and benchmarks at higher levels of aggregation – would enable more robust assessments of supervisory results both across institutions and over time. Further, growing deviations between benchmark and production projections could highlight emerging (or declining) areas of risk. In this regard, a related area of research could address optimal ways for supervisory to assess the signal when benchmark and production model results deviate significantly from one another. Finally, these simpler models could potentially form the basis of a more dynamic, system-focused stress test analysis that builds in the linkages and feedbacks not currently captured in the CCAR and DFAST stress testing programs.

This is a short list of what I see as some important tactical issues in supervisory stress test modeling. But I also want to highlight some bigger picture issues where additional research could help guide the evolution of the supervisory stress testing regime. Recall that I framed my discussion by arguing that supervisory stress testing emerged from the experience of the financial crisis to address both the backward-looking nature of regulatory capital requirements and potential first-mover problems that an individual bank might face in cutting dividends or raising new capital during periods of emerging stress. If that’s the case, how well has the stress testing regime, as incorporated into the SCAP and now the CCAR and proposed Stressed Capital Buffer approaches, addressed these problems? Is the approach working? Does it stand up?

It is easy to see that banks have more regulatory capital now than before the financial crisis. There is a growing body of literature that examines the impact of stress testing and other changes to regulatory capital requirements (Basel 2.0, 2.5 and 3.0) on lending to various categories of borrowers. This is an important area of research, but in my view, it is somewhat incomplete in its focus. To begin, studies that focus on lending by banks participating in the CCAR/DFAST stress tests ignore the possibility of substitution of this activity to non-stress-tested banks or to non-banks. This substitution effect could have significant consequences for both current economic activity and for systemic risk, channels that are important to understand in assessing the full impact of the programs. Some studies have looked more broadly at these substitution effects, but more work here – especially substitution into the non-bank sector – would be very helpful.

Further, examining how lending has been impacted in the current economic environment of strong economic activity does not quite provide the full picture, as it does not address whether

these programs will be successful in mitigating spillovers, credit contractions and other negative outcomes during periods of stress. In assessing the costs and benefits of stress testing and other post-crisis regulations, it seems critical to assess not just current impact but outcomes over the cycle – something that is a challenge during a period of recovery and growth.

A related set of questions concerns how stress testing affects the cyclical nature of capital requirements and what degree of cyclical nature is appropriate. The Basel countercyclical capital buffer is the primary supervisory tool to implement regulatory capital requirements that vary over the credit cycle, but stress testing has cyclical elements as well. Both stress scenario design and the starting condition of banks' balance sheets – the credit characteristics of the loan and securities portfolios, the extent of currently non-performing loans, exposures from trading and counterparty positions – can result in differences in stress test severity over the cycle. What is the optimal degree of cyclical nature for the stress testing regime? How does cyclical nature of stress testing interact with other cyclical elements, such as the countercyclical capital buffer and the incoming current expected credit loss (CECL) approach to loan loss provisioning? How should the pieces fit together?

Finally, one concern that has been raised about the Federal Reserve's approach to stress testing and integration into the CCAR program is that of "model monoculture" – the idea that banks will be incentivized to develop models that mimic the Fed's models rather than developing their own independent approaches.¹⁷ Significant commonality in modeling approaches in the banking system could result in banks adopting similar risk exposures and hedging techniques, exposing the sector to additional systemic risk and potential future instability. How do we measure this risk? What should we think about differences (and similarities) between bank-generated and supervisory stress test results, both for a particular stress test cycle and over time? What disclosure and transparency policies about supervisory models can address these concerns, while still supporting insight and credibility into the DFAST and CCAR stress testing programs?

These are big questions to which I do not have answers. But they are the questions that challenge policymakers as the regulatory and supervisory regime put in place following the crisis is re-examined and, appropriately, evolves in the wake of that re-examination. A disciplined analytical approach to these topics is critical in weighing future design choices, such as those made during the initial implementation of SCAP, CCAR and DFAST stress testing. Research on these topics could make substantial and meaningful contributions.

I appreciate the opportunity to talk about these issues and hope that my discussion spurs some ideas and new research. Thank you. [1234567891011121314151617](https://www.federalreserve.gov/1234567891011121314151617)

¹ International Monetary Fund. "Global Financial Stability Report." April 2009. www.imf.org/en/News/Articles/2015/09/28/04/53/sores042109c

² David S. Scharfstein and Jeremy C. Stein. "This Bailout Doesn't Pay Dividends." *The New York Times*. October 21, 2008.

³ Viral V. Acharya, Irvind Gujral, Nirupama Kulkarni and Hyun Song Shin. "[Dividends and Bank Capital in the Financial Crisis of 2007–2009](#)," [NBER Working Papers](#) 16896, National Bureau of Economic Research, 2011.

⁴ Beverly Hirtle. "Bank Holding Company Dividends and Repurchases during the Financial Crisis." Federal Reserve Bank of New York *Staff Report* no. 666, March 2014.

⁵ Viral Acharya, Robert Engel and Matthew Richardson. "Capital shortfall: a new approach to ranking and regulating system risk." *American Economic Review*, 102 (3) (2012), pp. 59–64.

⁶ The 9-quarter, rather than two year, horizon provides a two year forward horizon after the one-quarter gap between the "as of" date of the stress test and when robust data on banks' balance sheets, off-balance sheet exposures, and income are available.

- ⁷ European Banking Authority. “2018 EU-Wide Stress Test: Methodological Note.” November 17, 2017. www.eba.europa.eu/documents/10180/2106643/2018+EU-wide+stress+test+-+Methodological+Note.pdf
- ⁸ See, for instance, International Monetary Fund. *Global Financial Stability Report*. April 2009. www.imf.org/en/News/Articles/2015/09/28/04/53/sores042109c
- ⁹ Board of Governors of the Federal Reserve System. 2009. “Supervisory Capital Assessment Program: Overview of Results.” May 7, 2009. www.federalreserve.gov/newsevents/files/bcreg20090507a1.pdf
- ¹⁰ The exception to this approach is for the global market shock on trading and counterparty positions, which is assumed to occur instantaneously in a single quarter.
- ¹¹ Sales, purchases and acquisitions that have been contractually agreed by but not consummated by the start of the stress test are incorporated into the projections, however.
- ¹² Board of Governors of the Federal Reserve System. “Dodd-Frank Act Stress Test 2018: Supervisory Stress Test Methodology and Results.” June 2018. www.federalreserve.gov/publications/files/2018-dfast-methodology-results-20180621.pdf.
- ¹³ Daniel K. Tarullo. “Stress Testing after Five Years.” Remarks at the Federal Reserve’s Third Annual Stress Testing Symposium. June 24, 2014. www.federalreserve.gov/newsevents/speech/tarullo20140625a.htm.
- ¹⁴ In the initial years of CCAR and DFAST stress testing, balance sheet projections made by the banks were used in the supervisory estimates.
- ¹⁵ Beverly Hirtle, “Structural and Cyclical Macroprudential Objectives in Supervisory Stress Testing.” Federal Reserve Bank of New York. June 22, 2018. www.newyorkfed.org/newsevents/speeches/2018/hir180622
- ¹⁶ European Banking Authority. “2018 EU-Wide Stress Test: Methodological Note.” November 17, 2017. www.eba.europa.eu/documents/10180/2106643/2018+EU-wide+stress+test+-+Methodological+Note.pdf
- ¹⁷ Til Schuermann. “The Fed’s Stress Tests Add Risk to the Financial System.” *Wall Street Journal*. March 19, 2013. www.federalreserve.gov/newsevents/files/bcreg20090507a1.pdf