

Stress Test Dummies: A Fundamental Problem With CCAR (and How To Fix It)

Greg Baer | July 16, 2018

Reading and watching market analysis of recent 2018 CCAR stress test results, I've been struck by how analysts and investors clearly understand something that many policymakers do not: how extreme, and thus how extraordinarily improbable, this year's stress scenario was. They describe with wonder a scenario that includes a sudden increase in unemployment of 600 basis points and an immediate stock market crash of 62 percent. Using the confidence interval the Fed provides around quarterly monetary policy projections, we [calculate](#) that such a rapid increase in the unemployment rate (let alone all the simultaneous deterioration of all the other variables) has only about a fifty-fifty chance of occurring once in 10 thousand years.

Consider, then, the central, continuing problem with the Federal Reserve's CCAR stress test: it bases its capital assessment not on a bank's current financial condition or how its assets are most likely to behave in the future, but rather on how those assets would behave under a single stress event of unimaginably unlikely severity. The Federal Reserve also includes an adverse scenario but the severely adverse scenario (as the name would suggest) has always bound. And regulatory capital requirements produced by CCAR are now so high as to drive bank credit allocation decisions, which means that banks subject to CCAR are increasingly pressured to choose assets based on how well they perform under the Fed's single, apocalyptic scenario, not how they would perform under far more likely scenarios.

Of course, to some extent, the problems above are inherent in the nature of stress testing, as any prescribed stress will be unlikely to occur. Stress testing serves as a countercyclical capital buffer exactly because it presumes that current good times will turn bad at some point in the future, and rightly so. But the Fed multiplies this inherent problem exponentially by using a single scenario, and a single scenario of such extreme unlikelihood.

What to do? There is a very useful lesson to learn here from an unlikely source.

Beginning in 1979, the National Highway Transportation Safety Agency (NHTSA) began its New Car Assessment Program (NCAP). Each new car was tested for a 35 mph frontal impact with a rigid barrier. Thus, the famous crash-test dummies were born, as their post-crash health determined the score (one to five stars) of the car in which they were riding. The crash-test dummies become famous (even producing a mediocre Canadian rock group of that name), and car safety improved, as cars were engineered to better absorb frontal impacts through crumple zones and air bags, lessening the blow to drivers and passengers. The public felt better about car safety.

Then, in 2004, the NHTSA did something the Federal Reserve has not: it sought public comment on its scenario design. Below are excerpts from the comment letter filed by the Insurance Institute for Highway Safety, practically every word of which is directly relevant to CCAR:

This program has matured to the point where manufacturers are modifying vehicle designs to get better ratings, not to improve real-world crash protection. For example, manufacturers are making relatively minor changes to restraint system performance, changing airbag venting or modifying belt force limiters. Yet there is no evidence that this tweaking of restraint systems for optimum performance in a single test will improve protection in the range of serious frontal crashes that occur in the real world, especially because the rigid-barrier test with its relatively short crash pulse is not particularly representative of serious real-world frontal crashes.

Obviously, in CCAR, it is the banks that are the dummies being tested for a different sort of crash. Their health is evaluated against only a single type of impact each year: in 2018, a massive frontal crash involving an unprecedentedly sudden and severe increase in unemployment, with a simultaneous decrease in equity markets, disruption in capital markets, and decrease in housing prices.

The parallels continue. As part of its proposal, the NHTSA described one option it was considering: maintaining the same single test, but simply making it more severe. In other words, it proposed to do what the Federal Reserve has already done with CCAR scenarios. The Insurance Institute for Highway Safety's response is revelatory:

NHTSA has suggested it would consider altering the current NCAP star rating system, making it difficult to achieve a five-star rating by dividing the injury risk curves into smaller increments, resulting in a wider variation of star ratings. However, this would simply divide the four- and five-star performers into more groups; it would not provide consumers with additional meaningful information about crash protection. If such changes were adopted, it seems likely that automakers would strive to achieve good performance again. But increasingly this would mean manufacturers would do even more tweaking of restraint systems. The process would continue to be about passing the test, not improving real-world crash protection.... Test protocol changes must have some meaning outside of laboratory conditions to advance protection in real-world crashes.

[Another] NHTSA proposal for maintaining the flat-barrier test would be to increase the full-width barrier test speed to 40 mi/h. This idea is not well founded; it would increase the energy of an already high deceleration test by more than 30 percent, and NHTSA has not shown the relevance of doing so.

Raising the NCAP frontal barrier test speed from 35 to 40 mi/h probably sounds like a good idea to people with no technical background in crash protection, but it would drive restraint system designs in directions that, in the spectrum of real-world crashes, would not be beneficial. "Good" performance in a 40 mi/h rigid-barrier crash, which as noted above represents an extremely rare real-world event, almost certainly would result in poorer performance in the less severe crashes that occur far more often in the real world.

As my colleague Francisco Covas has [demonstrated](#) and the IIHS would have predicted, the Fed's use of a single extremely severe scenario has led banks to choose portfolios that do well *in that scenario* at the expense of performance in other more likely possible scenarios.

The IIHS's proposed solution in 2004 probably will not come as a surprise: using more than one scenario. Or as the IIHS put it, "Significant Benefits Can Be Obtained In Other Crash Modes." It began by suggesting an offset test, and a test involving impact with a narrow object (e.g., a pole or tree).

And the NHTSA listened. Today, the NHTSA uses three different stress scenarios; the IIHS also conducts its own test, using five scenarios. Auto buyers consider all eight in deciding which car to buy. (Unlike with CCAR, the test results do not bind the automakers – e.g., by limiting their ability to sell cars with low ratings – but rather are information used by consumers in deciding safety risk.)

We have previously urged the Federal Reserve to seek comment on its stress scenarios, to use multiple scenarios, and to include bank-designed, bank-specific scenarios among them. We just had no idea that the NHTSA had thought of this idea thirteen years ago. (Perhaps the fact that the Fed’s test is called C...CAR should have prompted an earlier look to the auto industry.)

Two further notes.

First, the importance of the Federal Reserve doing likewise is even more important in finance than in auto safety. The NHTSA tests apply to all automobiles – there is no “shadow auto system.” Thus, the NHTSA had no concerns that making safety requirements higher for one set of automobiles, thereby increasing their cost to consumers, would cause consumers and safety risk to migrate to a group of automakers not subject to its standards. Presumably, had that been a concern, the NHTSA would have been even more concerned about the impact of a single scenario design.

Second, any concerns about the use of single scenario are greatly amplified when one realizes that the Fed, in addition to using only one *scenario* of its own devise, uses only one *model* of its own devise (kept secret, without public review let alone comment) to estimate each bank’s losses under that model, and has proposed to use its own or a Basel-designed model for every other component of its proposed new capital standard.

In a recent article, [The Quiet Revolution in Central Banks](#), I described how the Federal Reserve’s recent capital proposal would combine a Basel minimum standard, CCAR results (through a so-called Stressed Capital Buffer), a GSIB surcharge, and potentially a countercyclical capital buffer to establish what would almost certainly be the binding capital constraint for any bank subject to it. All of the component parts involve either the Federal Reserve or the Basel Committee, not the bank, modeling the risk:

The remarkable result of this process is that at no point is a bank’s view of the risk of a loan or any other asset relevant to the capital it must hold against that asset.... There is good reason to believe this would end poorly for the U.S. economy. Governmental attempts at direct or indirect credit allocation have a dismal history. Economic growth would suffer, as diversity in risk tolerance and judgment produce greater opportunities for businesses and individuals to obtain bank credit. Furthermore, because standardized risk measures and the Federal Reserve’s loss-forecasting models are necessarily crude and one-size-fits-all, relying on less data than the banks’ own models, much of the capital allocation they drive is likely to be misallocation. Lastly, systemic risk would increase as large banks were forced to concentrate in asset classes favored by the governmental models. (There is also the converse risk that non-banks would concentrate in asset classes disfavored by the governmental models; because those actors operate outside the purview of the Federal Reserve and do not qualify for lender-of-last-resort support, a collapse in prices for those assets would constitute its own systemic risk.)

In sum, leaving model variation aside for the moment and focusing only on scenario design, the Federal Reserve should learn a lesson from the NHTSA and seek public comment on at least three key questions: (1) how many stress scenarios to use (including the possibility of using some bank-designed, bank-specific scenarios), (2) by what standard those scenarios should be designed (e.g., with what probability they should be likely to occur); and (3) how their results should be combined (e.g., by averaging the losses and revenue under each scenario). Each subsequent year, it should then publish proposed scenarios for comment to ensure they meet the Fed’s own previously published design standard. Furthermore, to reduce volatility in outcome – imagine the NHTSA changing

crash-test standards each year, and what chaos that would have brought to car design — it should also propose averaging the previous 2-3 year results before using those results as minimum capital standard. This would reduce volatility significantly, and allow for better capital planning.

If the Fed took these steps, the banks would ride safer, and do a better job delivering for their customers.

Disclaimer: The views expressed do not necessarily reflect those of the Bank Policy Institute's member banks, the G30 or the G30 Working Group on Treasury Market Liquidity and are not intended to be, and should not be construed as, legal advice of any kind.