Implementing Dodd-Frank Act Stress Testing

Margaret Ryznar, Frank Sensenbrenner***
*& Michael Jacobs, Jr.****

Abstract

In recent years, the question of how to prevent another crippling recession has become a prominent one. The answer provided by the Dodd-Frank Act is stress testing, which examines through economic models how banks would react to a bad turn of economic events, such as negative interest rates. The first of its kind in the legal literature, this Article offers a model for stress testing that banks should use in complying with Dodd-Frank. Specifically, this Article finds that the Bayesian model that takes into account past outcomes, namely the Federal Reserve's previous stress test scenarios, is the most accurate model in stress testing.

I. Introduction

As even Hollywood has taken to explaining these days[1], the credit crisis was a driver of the Great Recession.[2] In other words, banks had made risky loans and were low on capital when the loans defaulted. It became dire enough that bank credit cards and ATMs would have eventually stopped working. In fact, Secretary of the Treasury Henry Paulson estimated that in the last recession, the ATMs were three

* Associate Professor, Indiana University McKinney School of Law. The authors thank Jessica Dickinson and Susan deMaine for excellent research assistance.

** Risk Analyst, U.S. Commodity Futures Trading Commission; PhD in Finance, University of Sydney. The research for this article was conducted and the article was written by Frank J. Sensenbrenner in his personal capacity and not in his official capacity as an employee of the Commodity Futures Trading Commission (CFTC). The analyses and conclusions expressed in this research paper are those of Frank J. Sensenbrenner and do not reflect the views of other employees of the CFTC Division of Clearing and Risk, other CFTC staff, the Commission itself, or the United States Government.

*** Principal Director, Accenture Consulting. PhD in Finance, Graduate School and University Center of the City University of New York.

1. *See, e.g.*, The Big Short (2015) (starring Christian Bale and Brad Pitt).

2. *See, e.g.*, Kenneth W. Dam, *International Law and the Financial Crisis: The Subprime Crisis and Financial Regulation: International and Comparative Perspectives*, 10 Chi. J. Int'l L. 581, 581-84 (2010) (describing several causes of the 2008 financial crisis and the resulting impact on the economy).

days away from not working. Although the level of risk in the financial sector had been significant,[3] the regulators and the firms were all using the same risk models that did not measure it accurately, suggesting that the firms were not being sufficiently scrutinized.

The question on everyone's mind since these events has been how to prevent another banking institution failure.[4] The Federal Reserve ("Fed") has decided to implement the solution of stress testing, under the authority of the Dodd-Frank Act. The purpose of stress testing is to ensure that a bank has adequate capital to survive a financial crisis by not tying up its money in bad loans or risky investments.[5]

The idea of stress testing is not new.[6] In fact, stress testing is at least as old as fire drills, the classic stress test. The fire alarm rings, forcing people to leave their warm offices and stand outside in the cold, waiting for the drill to finish. It may be irritating to participants, but serves an important purpose: to ensure that everyone is ready in case there is ever a fire. Can management get people out of the building fast enough? Are the exits clearly marked? Do people know what to do?

Similarly, bank stress testing simulates bad economic conditions in economic models to ensure that a bank has enough money to survive another financial crisis. What if unemployment rises to 10%? What if

---

3. *But see* Wulf A. Kaala & Richard W. Painter, *Initial Reflections on an Evolving Standard: Constraints on Risk Taking by Directors and Officers in Germany and the United States*, 40 SETON HALL L. REV. 1433 (2010) ("[T]he concepts of 'whether there is any such thing as excessive risk, and if so, how excessive risk is to be defined, is another issue. Viewpoints on these questions will have a substantial impact on how a policy maker—or a group of policy makers in a particular country—approaches regulation of risk in the banking sector.'").

4. *See, e.g.*, Alan J. Meese, Essay, *Section 2 Enforcement and the Great Recession: Why Less (Enforcement) Might Mean More (GDP)*, 80 FORDHAM L. REV. 1633 (2012); Michael Greenberger, *Closing Wall Street's Commodity and Swaps Betting Parlors: Legal Remedies to Combat Needlessly Gambling Up the Price of Crude Oil Beyond What Market Fundamentals Dictate*, 81 GEO. WASH. L. REV. 707 (2013); Cody Vitello, *The Wall Street Reform Act of 2010 and What It Means for Joe & Jane Consumer*, 23 LOY. CONSUMER L. REV. 99 (2010).

5. James E. Kelly, *Transparency and Bank Supervision*, 73 ALB. L. REV. 421, 421 (noting that following the 2008 financial crisis, attention has focused on the role of systemic risk in financial institutions and markets). "The stress tests allow the Fed to tailor its regulatory efforts based on realistic information, and enable financial institutions to simulate how they might fare under highly adverse economic conditions." Derek E. Bambauer, *Schr?dinger's Cybersecurity*, 48 U.C. DAVIS L. REV. 791, 836-37 (2015).

6. "Although stress analysis is a parvenu in the bank regulatory regime, it has a long history in the engineering field from as early as the sixteenth century. These early stress testing methodologies evolved into professional norms on the part of engineers to remain focused on worst-case scenarios when designing and building structures, materials, and systems. Financial firms have adopted an extensive suite of stress testing techniques alongside their risk management systems."

Robert Weber, *A Theory for Deliberation-Oriented Stress Testing Regulation*, 98 MINN. L. REV. 2236, 2324-2325 (2014).

the stock market craters? Would the bank have enough money not tied up in loans or bad investments?[7] Stress testing uses hypothetical future scenarios set by the Federal Reserve to inform ex ante regulation. For the 2016 stress tests, for example, banks had to consider their preparedness for negative U.S. short-term Treasury rates, as well as major losses to their corporate and commercial real estate lending portfolios.[8]

Although stress testing is not new, what is new is the Federal Reserve's role in setting bad case scenarios and requiring banks to use them in their stress tests, the results of which must be reported annually.[9] The Dodd-Frank Act facilitated stress testing by empowering agencies to prevent another crisis. Thus, the Federal Reserve Bank imposes stress testing on banks that have over $10 billion in assets, in order to ensure the stability of the American financial sector. The $10 billion threshold implicates many banks in the United States, including BMO Harris, Key Bank, and smaller regional banks, in addition to the well-known big banks such as Bank of America and Goldman Sachs.

When a bank fails its stress test, it is headline news. A failed stress test raises red flags about whether a bank has enough capital to stay solvent in a crisis. Without enough capital, the bank would stop paying dividends, which would impact, for example, retirement portfolios with bank stock.

There have been a few banks who have failed their stress tests recently. Citigroup failed twice, and Goldman Sachs and Bank of

---

7. "Stress tests help financial institutions, as well as their regulators and other stakeholders, understand how an institution or system will respond to severe, yet plausible, stressed market conditions such as low economic output, high unemployment, stock market crashes, liquidity shortages, high default rates, and failures of large counterparties. The results of stress tests shed light on the tension points and weak links in portfolios and institutions that could create extraordinary, but plausible, losses."

Robert F. Weber, *The Corporate Finance Case for Deliberation-Oriented Stress Testing Regulation*, 39 J. CORP. L. 833, 833-34 (2014).

8. Board of Governors of the Federal Reserve System, *2016 Supervisory Scenarios for Annual Stress Tests Required under the Dodd-Frank Act Stress Testing Rules and the Capital Plan Rule* (January 28, 2016), available at http://www.federalreserve.gov/newsevents/press/bcreg/bcreg20160128a2.pdf.

9. Making public the results of bank stress tests is just another form of transparency.

"In the recent crisis transparency concerns have generally been focused on two groups: (i) financial institutions and markets, where disclosure, efficiency, fairness, and detection of systemic risk are emphasized as benefits of greater transparency, and (ii) the federal government, particularly the Department of Treasury, the Federal Reserve, and the banking supervisors, as stewards and regulators, where fairness, accountability and taxpayer protection tend to be emphasized."

James E. Kelly, *Transparency and Bank Supervision*, 73 ALB. L. REV. 421, 424-25 (2010). *See also* TIMOTHY GEITHNER, STRESS TEST: REFLECTIONS ON FINANCIAL CRISES (2014).

America would have failed if they had not amended their capital distributions, which changed the results. However, there are no guiding models for stress testing. This Article contributes by filling this void.

In their stress tests, banks have to measure two major types of risk: market risk and credit risk. Market risk is the risk that the banks will lose money on trading stocks and bonds, while credit risk is the risk that their customers will default on their loans. There is an additional risk in these stress test models, and that is the risk that the model does not accurately reflect all possible outcomes. This could lead to a failed stress test.

Any sort of model requires justification of why certain variables are in the model and what values are used for the variables. Otherwise, the model does not accurately reflect reality, which is called "model risk." Model risk is managed by model validation, which is the effective and independent challenge of each model's conceptual soundness and control environment.

Also, some models look only at the data, as opposed to historical experience or the judgment of experts who may bring experiences that do not exist in the data. For example, models might be missing input from loan officers, even when this input is helpful – a loan officer issuing mortgages for 30 years might have a lot of good qualitative perspective. A Bayesian methodology allows incorporation of these views by representing them as Bayesian priors.

This Article shows that the Bayesian model that takes into account past outcomes, namely the Federal Reserve's previous stress test scenarios, requires a more significant buffer for uncertainty – by 25% – as opposed to simply modeling each year's scenario in isolation. This means that if modelers do not take previous results into account, they can underestimate losses significantly – by as much as 25%. This could be the difference between a successful stress test and a failed stress test. Part II of this Article begins by laying out the legal framework. Part III suggests models for banks to use in stress testing.

This Article uses the previous Federal Reserve scenarios as priors. This is because of the belief in the industry that the Federal Reserve adapts its scenarios to stress certain portfolios, but remains consistent with its prior scenarios in terms of economic intuition. This Article uses two sources of data: the hypothetical economic scenarios released by the Federal Reserve annually and the consolidated financial statements of banks, which detail credit losses by type of loan.

II. LEGAL FRAMEWORK FOR STRESS TESTING

The legal framework on stress testing has exploded in the last decade, significantly since being required by the Dodd-Frank Act, which was the Congressional reaction to the Great Recession.[10] Stress testing has now become the primary way to regulate banks, despite several issues it raises, considered in this Part.

A. *Introduction of Stress Testing*

Since the Great Depression, there have been several types of regulation of banks: geographic restrictions, activity restrictions, capital or equity requirements, disclosure mandates, and risk management oversight.[11] "These regimes have been employed successively and in tandem to combat new problems and to make use of technological innovation in modernizing regulatory tools."[12]

Stress testing is another category of regulation, which examines the performance of the regulated entity in hypothetical, challenging circumstances. Immediately after the beginning of the Great Recession, in February 2009, several regulators that included Treasury, the Office of the Comptroller of the Currency, the Federal Reserve, the Federal Deposit Insurance Corporation, and the Office of Thrift Supervision revealed the details of Treasury's Capital Assistance Plan (CAP), which required stress testing (Supervisory Capital Assessment Program (SCAP)) of, primarily, the 19 largest U.S. banking enterprises.[13] In other words, to receive government assistance in the wake of the financial crisis, banks had to subject themselves to stress testing.[14] The results showed that several of these banks would need more capi-

10. "[O]n July 21, 2010, President Obama signed into law a package of financial regulatory reforms unparalleled in scope and depth since the New Deal: The Dodd-Frank Wall Street Reform and Consumer Protection Act (Act or Dodd-Frank Act) was a sweeping reaction to perceived regulatory failings revealed by the most severe financial crisis since the Great Depression. The Dodd-Frank Act was intended to restructure the regulatory framework for the United States financial system, with broad and deep implications for the financial services industry where the crisis started."

Heath P. Tarbert, *The Dodd-Frank Act—Two Years Later*, 66 CONSUMER FIN. L.Q. REP. 373, 373 (2012).

11. Mehrsa Baradaran, *Regulation by Hypothetical*, 67 VAND. L. REV. 1247 (2014).

12. *Id.* at 1247.

13. MICHAEL P. MALLOY, BANKING LAW & REG. s 14.04 (2015).

14. "Prior to the crisis, surprisingly few financial institutions put themselves through stress testing exercises. In the depths of the crisis, however, regulators conducted stress tests on the largest U.S. banks with what seems to have been great success both for the health of the institutions and the marketplace." Eugene A. Ludwig, *Assessment of Dodd-Frank Financial Regulatory Reform: Strengths, Challenges, and Opportunities for a Stronger Regulatory System*, 29 YALE J. ON REG. 181, 188 (2012).

tal to withstand worse-than-expected economic conditions.[15] However, the banks eventually recovered.[16]

The regulators' continued interest in stress testing was then reinforced by the passage of the Dodd-Frank Act §165(i), legislation which required periodic stress tests conducted by the Federal Reserve on the regulated banks and by the banks themselves. The stated aim of the Dodd-Frank Act is "to promote the financial stability of the United States by improving accountability and transparency in the financial system, to end 'too big to fail', to protect the American taxpayer by ending bailouts, to protect consumers from abusive financial services practices, and for other purposes."[17]

To prevent financial instability in the United States, the Dodd-Frank Act generally sought to enhance the supervision of non-bank financial companies supervised by the Board of Governors and certain bank holding companies.[18] To advance this goal, the Act requires stress testing of financial institutions of a certain size because the biggest banks pose the most significant harm to the American economy.[19] The result was 31 bank holding companies participating in stress test-

---

15. "Through the Supervisory Capital Assessment Program [SCAP], the Treasury required each of the nineteen largest U.S. banks, representing some two-thirds of all U.S. bank assets, to simultaneously undertake a Treasury-specified assessment of the bank's capital two years into the future under two different scenarios—one baseline and one more adverse—in order to identify whether the bank had sufficient capital under each. The methodology of the Stress Test was publicly disclosed so that its credibility could be independently evaluated. Banks that reported a capital shortfall would be required to raise new capital in that amount, which the Treasury would provide if the market would not. Importantly, the Treasury publicly announced the results of the Stress Test, and the corresponding determination of capital adequacy. The Stress Test revealed that ten of the banks had inadequate capital, while nine had sufficient capital. Of the banks that had to raise new capital, the size of the shortfall ranged from $0.6 billion to $33.9 billion."

Ronald J. Gilson & Reinier Kraakman, *Market Efficiency After the Financial Crisis: It's Still A Matter of Information Costs*, 100 VA. L. REV. 313, 358 (2014).

16. "In a speech reflecting on the SCAP one year after its conclusion, Chairman Bernanke pointed to several factors as evidence of its success: the majority of the nineteen firms that needed capital following the stress tests were able to raise the capital in the market (as opposed to receiving assistance from Treasury); most of the nineteen firms had repaid TARP money that they had received during the crisis; share prices of the nineteen firms had generally increased; and banks' access to debt markets and interbank and short-term funding markets had improved."

Ludwig, *supra* note 14, at 188.

17. Pub.L. 111-203, Title I, § 165.

18. 12 U.S.C. § 5365(i) (2010). Pub.L. 111-203, Title I, § 165(i).

19. Additionally, "Congress demonstrated sensitivity to the regulatory burden on smaller institutions in the passage of Dodd-Frank." Heidi Mandanis Schoone, *Regulating Angels*, 50 GA. L. REV. 143, 156 (2015).

ing in 2015, which represented more than 80 percent of domestic banking assets.[20]

The Dodd-Frank Act authorized the Federal Reserve and other agencies to implement regulations to prevent another financial crisis.[21] The Federal Reserve included stress testing in its January 2012 proposed rules that would implement enhanced prudential standards required under Dodd-Frank Act §165, including stress testing,[22] as well as the early remediation requirements established under Dodd-Frank Act §166.[23] In October 2012, the Federal Reserve issued a final rule requiring financial companies with total consolidated assets of more than $10 billion to conduct annual stress tests, effective November 15, 2012.[24] The biggest banks, those with over $50 billion in assets, must conduct semi-annual stress tests.[25]

Stress testing under the Dodd-Frank Act is based on hypotheticals set by the Federal Reserve Bank.[26] Specifically, financial system mod-

---

20. Board of Governors of the Federal Reserve System, *Dodd-Frank Act Stress Test 2015: Supervisory Stress Test, Methodology and Results* (Mar. 2015).

21. *See, e.g.*, Greg Bader, *Stress Tests: Making Sure the Large Banks Can Weather Another Storm*, 18 WESTLAW JOURNAL BANK & LENDER LIABILITY 1 (2013) (providing an overview of stress testing and potential reasons why Congress included it as part of the Dodd-Frank Act). "But the Dodd-Frank Act is an empty vessel: it authorizes agencies to regulate without giving them much guidance as to how to regulate." Eric A. Posner & E. Glen Weyl, *An FDA for Financial Innovation: Applying the Insurable Interest Doctrine to Twenty-First-Century Financial Markets*, 107 NW. U. L. REV. 1307, 1309 (2013). There has been significant criticism of the Dodd-Frank Act. *See, e.g.*, Steven A. Ramirez, *Dodd-Frank as Maginot Line*, 15 CHAP. L. REV. 109, 119, 130 (2011) (suggesting that the Dodd-Frank Act will not necessarily prevent a future financial crisis).

22. "The Dodd-Frank stress test implemented Dodd-Frank sections 165(i)(1) and (i)(2), which required both supervisory and company-run stress testing over a wider set of institutions than those covered by the CCAR. Institutions subject to the Dodd-Frank stress test include those bank holding companies with assets of $50 billion or more that had participated in SCAP (and who had also participated the previous year in CCAR), as well as bank holding companies with between $10 billion and $50 billion in assets, and state member banks and savings and loan holding companies with over $10 billion in assets."

Gilson & Kraakman, *supra* note 15, at 358-361.

23. Other government agencies that have proposed or issued rules on stress testing include the Federal Deposit Insurance Corporation and the Office of the Comptroller of the Currency. MALLOY, *supra* note 13, at s 14.04.

24. *Id.*

25. Dodd-Frank Act, 12 U.S.C. § 5365(i) (2010). "Note that there is a redundancy in the stress testing; one should keep in mind, however, that in conducting its stress tests, the Federal Reserve has an eye toward macroprudential responsibilities, whereas the bank largely focuses on microprudential issues." Ludwig, *supra* note 14, at 186 n.20. *See also* Daniel K. Tarullo, Keynote Address, *Macroprudential Regulation*, 31 YALE J. ON REG. 505 (2014); Claude Lopez, Donald Markwardt, & Keith Savard, *Macroprudential Policy: What Does It Really Mean*, 34 NO. 10 BANKING & FIN. SERVICES POL'Y REP. 1 (2015).

26. "The genesis of the current supervisory stress tests and CCAR dates back to early 2009, when supervisors conducted simultaneous stress tests of the 19 largest U.S.

eling allows the introduction of variables that approximate various adverse economic developments, allowing an assessment of results if the system were under stress.[27] The Board of Governors must provide at least three different sets of conditions under which the evaluation shall be conducted, including baseline, adverse, and severely adverse.[28] In other words, the economy imagined by the hypotheticals is in differing levels of strain, allowing the banks to test their readiness for a range of different economies. The Federal Reserve must publish a summary of the results of these tests.[29]

The Federal Reserve has other discretionary powers under the Dodd-Frank Act as well. It may require additional tests, may develop other analytic techniques to identify risks to the financial stability of the United States, and may require institutions to update their resolution plans as appropriate based on the results of the analyses.[30]

In 2010, the Federal Reserve had also initiated the annual Comprehensive Capital Analysis and Review (CCAR) exercise, which involves quantitative stress tests and a qualitative assessment of the largest bank holding companies' capital planning practices, which requires the bank to submit its detailed capital plans.[31] CCAR is separate from the Dodd-Frank stress tests, impacting only the largest banks—those with over $50 billion in assets.[32] CCAR has become a

---

> BHCs (the Supervisory Capital Assessment Program or SCAP) in the midst of the financial crisis. The SCAP stress test assessed potential losses and capital shortfalls at the 19 large BHCs under a uniform scenario that was, by design, even more severe than the expected outcome at that time."

Tim P. Clark & Lisa H. Ryu, *CCAR and Stress Testing as Complementary Supervisory Tools*, Board of Governors of the Federal Reserve System (updated June 24, 2015), available at http://www.federalreserve.gov/ bankinforeg/ccar-and-stress-testing-as-complementary-supervisory-tools.htm#f5.

27. Dodd-Frank Act, 12 U.S.C. § 5365(i) (2010).

28. *Id.*

29. *Id.*

30. *Id.*

31. "In late 2011, following the Stress Tests, the Federal Reserve Board finalized a rule requiring U.S. bank holding companies with consolidated assets of $50 billion or more to submit annual capital plans for review in a program known as the Comprehensive Capital Analysis and Review ('CCAR'). The stress testing under CCAR is conducted annually. Each bank holding company's capital plan must include detailed descriptions of: 'the [holding company's] processes for assessing capital adequacy; the policies governing capital actions such as common stock issuance, dividends, and share repurchases; and all planned capital actions over a nine-quarter reporting horizon.' In addition, each holding company must report to the Federal Reserve the results of various stress tests that assess the sources and uses of capital under both baseline and stressed economic conditions."

Gilson & Kraakman, *supra* note 15, at 359-360.

32. "While DFAST [Dodd-Frank Act stress testing] is complementary to CCAR [Comprehensive Capital Analysis and Review], both efforts are distinct testing exercises that rely on similar

main component of the Federal Reserve System's supervisory program for the largest banks.

A bank holding company must conduct its stress test for purposes of CCAR using the following five scenarios: 1) supervisory baseline: a baseline scenario provided by the Federal Reserve Board under the Dodd-Frank Act stress test rules; 2) supervisory adverse: an adverse scenario provided by the Board under the Dodd-Frank Act stress test rules; 3) supervisory severely adverse: a severely adverse scenario provided by the Board under the Dodd-Frank Act stress test rules; 4) bank holding company (BHC) baseline: a BHC-defined baseline scenario; and 5) BHC stress: at least one BHC-defined stress scenario.[33]

If banks fail to meet the Federal Reserve's set capital levels, regulators can restrict their ability to pay dividends to shareholders so that the bank can accumulate additional capital. This is a decision ordinarily reserved for the banks' managers, illustrating the power of the regulator's role.[34]

Customers of banks have felt the consequences of this regulatory environment. Most notably, many banks have restricted their lending practices.[35] Indeed, the entire aim of these regulations is to diminish credit risk, part of which is ensuring that only credit-worthy people

---

processes, data, supervisory exercises, and requirements. The Federal Reserve coordinates these processes to reduce duplicative requirements and to minimize regulatory burden." Stress Tests and Capital Planning, Board of Governors of the Federal Reserve System (June 25, 2014), available at http://www.federalreserve.gov/bankinforeg/stress-tests-capital-planning.htm.

> "The main difference between the CCAR and the Dodd-Frank stress tests is the capital action assumptions that are combined with pre-tax net income projections to estimate post-stress capital levels. The Dodd-Frank test uses a standard set of capital action assumptions that are laid out in the Dodd-Frank test rules, while the CCAR analysis uses the bank holding company's planned capital actions to determine whether the company would meet supervisory expectations for capital minimums in stressful economic conditions."

Gilson & Kraakman, *supra* note 15, at 361.

33. Board of Governors of the Federal Reserve System, *Comprehensive Capital Analysis and Review 2015: Summary Instructions and Guidance* (updated October 31, 2014), available at http://www.federalreserve.gov/bankinforeg/stress-tests/CCAR/2015-comprehensive-capital-analysis-review-summary-instructions-guidance-stress-test-conducted.htm.

34. "Historically, bank regulators have restricted bank dividends as part of a larger effort to preserve banks' capital and make them more able to withstand losses." Robert F. Weber, *The Comprehensive Capital Analysis and Review and the New Contingency of Bank Dividends*, 46 SETON HALL L. REV. 43, 43 (2015).

35. Karen Gordon Mills & Brady McCarthy, *The State of Small Business Lending: Credit Access During the Recovery and How Technology May Change the Game*, 34-35 fig.24 (Harvard Bus. Sch., Working Paper No. 15-004, 2014), available at http://www.hbs.edu/faculty/Publication%20Files/15-004_09b1bf8b-eb2a-4e63-9c4e-0374f770856f.pdf [http://perma.cc/ES9U-MWWP]; Kelly Mathews, Comment, *Crowdfunding, Everyone's Doing It: Why and How North Carolina Should Too*, 94 N.C. L. REV. 276, 284-85 (2015).

are able to borrow. However, there have been several issues arise relating to stress testing.

B. *Issues Regarding Stress Testing*

The health of the financial sector has been left to stress testing, which has become the primary way to regulate banks. Some commentators want stress testing expanded to other firms.[36] However, there have been several issues emerge relating to stress testing since its rise as a major indicator of a financial institution's health.

> First, the various capital adequacy and liquidity ratio scenarios that were used in the initial round of stress tests were criticized as being too lenient and thus able to produce a false positive. Second, the macroeconomic indicator assumptions about the scenarios that these entities may face were also criticized as too optimistic, further exacerbating the problem of test validity. Third, choosing which institutions need to be tested is a tacit admission of their importance to the macroeconomic health of the country, and, as such, enshrines their status as 'too big to fail.'[37]

Another commentator has criticized regulation by hypothetical regime, namely by stress tests and living wills,[38] suggesting it must be either abandoned or strengthened because of its current flaws.[39] For example, there might be tension in the Federal Reserve Board's determination of the amount of stringency for the stress tests. On the one hand, the Federal Reserve Board is tasked with systemic risk regulation, but, on the other hand, the functioning of the markets is also a key concern.[40] Methodological issues include claims that the tests are

---

36. Kwon-Yong Jin, Note, *How to Eat an Elephant: Corporate Group Structure of Systemically Important Financial Institutions, Orderly Liquidation Authority, and Single Point of Entry Resolution*, 124 YALE L.J. 1746, 1785-1786 (2015) (suggesting expanding stress tests to the subsidiary level of financial institutions).

37. Behzad Gohari & Karen Woody, *The New Global Financial Regulatory Order: Can Macroprudential Regulation Prevent Another Global Financial Disaster?*, 40 J. CORP. L. 403, 432-33 (2015).

38. "The Dodd-Frank Act's "living will" requirement mandates that systemically important financial institutions develop wide-ranging strategic analyses of their business affairs, and submit comprehensive contingency plans for reorganization or resolution of their operations to regulators. The goal is to mitigate risks to the financial stability of the United States and encourage last-resort planning, which will allow for a rapid and efficient response in the event of an emergency."

Nizan Geslevich Packin, *The Case Against the Dodd-Frank Act's Living Wills: Contingency Planning Following the Financial Crisis*, 9 BERKELEY BUS. L.J. 29, 29 (2012).

39. Baradaran, *supra* note 11, at 1247; *see also* Jonathan C. Lipson, *Against Regulatory Displacement: An Institutional Analysis of Financial Crises*, 17 U. PA. J. BUS. L. 673, 724 (2015) ("While stress tests and living wills are laudable, they seem unlikely to overcome the uncertainty and complexity of Dodd-Frank's resolution regime, and the incentive effects it will have on potential pre-failure negotiations.").

40. Baradaran, *supra* note 11, at 1251.

not adverse enough and are too narrowly focused both on a single static point in time and single data point.[41] There have been some concerns caused by the consistently positive results delivered by stress tests. "When the government conducts what it claims to be a rigorous stress test of a bank and then gives that bank a clean bill of health, the market receives a signal not only that the bank's risks are well managed but also that the government itself will stand behind the bank if the assessment proves incorrect."[42] Commentators have also wondered whether the exercise of stress testing will be made moot by permanent stress testing that would continue to produce overly positive results.[43]

Criticism has also targeted the enforcement of any regulation. For example, there is the possibility of bias in enforcement of the laws.[44] Furthermore, there are separate critiques regarding over-regulation of the business environment.[45]

Finally, there has been some question about how much related to stress testing should be made public. Currently, the stress test models used by the Federal Reserve are not made public, as some commentators have wanted. However, the results of stress tests are made public, but that in itself is controversial too. Some have argued that people will avoid using banks that perform poorly in stress tests, preventing such banks from recovering from an unsatisfactory stress test.[46]

In 2015, the Federal Reserve started to make changes after issues were discovered internally with the model validation process, which seeks to ensure the quality of the economic models themselves. In 2014, the model validation function had conducted three reviews assessing its performance and that of the broader supervisory stress testing program. The model validation function noted several areas for improvement. First, its staffing methods were inconsistent with industry practice and depended on a select number of key personnel. Sec-

41. *Id.* at 1247.

42. *Id.*

43. Gohari & Woody, *supra* note 37, at 433.

44. Joan Macleod Heminway, *Save Martha Stewart? Observations About Equal Justice in U.S. Insider Trading Regulation*, 12 TEX. J. WOMEN & L. 247, 263 (2003).

45. *See, e.g.*, Karen Woody, *Conflict Minerals Legislation: The SEC's New Role as Diplomatic and Humanitarian Watchdog*, 81 FORDHAM L. REV. 1315 (2012) (noting that Dodd-Frank even extends to regulating conflict minerals for ethical reasons). For the argument that tax incentives might be better solutions to certain corporate issues than regulation, see Margaret Ryznar & Karen Woody, *A Framework on Mandating versus Incentivizing Corporate Social Responsibility*, 98 MARQ. L. REV. 1667 (2015).

46. *See generally* Gilson & Kraakman, *supra* note 15, at 313; Henry T. C. Hu, *Disclosure Universes and Modes of Information: Banks, Innovation, and Divergent Regulatory Quests*, 31 YALE J. ON REG. 565 (2014).

ond, there were risks identified that were related to changes to models that occur late in the supervisory stress testing cycle. Third, model inventory lacked several components either required or deemed useful by supervisory guidelines. Finally, limitations encountered by reviewers during model validation were not sufficiently identified for management in the validation reports submitted to management.[47]

In response, the Federal Reserve has devoted a full-time team to the validation process. Additionally, the Federal Reserve has established a committee of senior staff to oversee the model validation for the stress testing models.[48] Thus, the Federal Reserve has recommitted to stress testing as the means of regulation of banks, with several advantages. To the extent that stress testing improves the security of banks in the United States, it is useful. Also, to the extent that bank stress testing increases the use of American banks and confidence in the American markets, it is beneficial.[49] Nonetheless, there is room for improvement on the models that banks use. As one such improvement, this Article suggests the Bayesian model for stress testing.

III. PROPOSED MODEL

Given the importance of stress testing,[50] there remains room for improvement. While the Federal Reserve stress testing methods have not been publicly disclosed, the banks have been more transparent in their stress testing. They have not used the Bayesian model in their stress tests, even though there are significant advantages to it.

47. Office of Inspector General, *The Board Identified Areas of Improvement for Its Supervisory Stress Testing Model Validation Activities, and Opportunities Exist for Further Enhancement* (October 29, 2015), available at https://oig.federalreserve.gov/reports/board-supervisory-stress-testing-model-validation-oct2015.htm.

48. Ryan Tracy, *Fed Finds Fault With Its Own Stress Tests*, WALL STREET JOURNAL (Dec. 6, 2015), available at http://www.wsj.com/articles/fed-finds-fault-with-its-own-stress-tests-1449311404.

49. "Accordingly, the SCAP strongly suggests that stress tests, if done properly, are able to build important market confidence and stability." Ludwig, *supra* note 14, at 188. There can be a similar argument made to Professor Coffee's that "[securities] [i]ssuers migrate to U.S. exchanges because by voluntarily subjecting themselves to the United States's higher disclosure standards and greater threat of enforcement (both by public and private enforcers), they partially compensate for weak protection of minority investors under their own jurisdictions' laws and thereby achieve a higher market valuation." John C. Coffee, Jr., *Racing Towards the Top?: The Impact of Cross-Listings and Stock Market Competition on International Corporate Governance*, 102 COLUM. L. REV. 1757, 1757 (2002).

50. "The way stress tests are conducted is critical. For example, it appears that the European stress tests have not been effective or helpful supervisory tools. In fact, one commentator has even gone so far as to call them 'farcical,' as the Irish banking system collapsed just four months after the country's banks passed their stress tests. Ludwig, *supra* note 14, at 188.

A. *Bayesian Modeling – A Theoretical Approach*

When implementing a model, the specification of a model, definition of parameters, quantities of interest, or specification of the parameter space all require justification that is an important part of the model validation procedure expected of financial institutions.[51] If the model is not appropriate for the purpose for which it has been put to use, model risk arises.[52] However, estimation of parameters, such as variable inputs, after these judgments are made typically proceeds without regard for potential non-data information about the parameters, in an attempt to appear completely objective.

Nonetheless, subject matter experts typically have information about parameter values, as well as about model specification. For example, a loan loss rate should lie between zero (0%) and one (100%), or a dollar loss estimate for an institution should be no more than the value of assets on an institution's books, or the definition of the parameter estimated. However, if we are considering a loss rate for a particular portfolio segment, we in fact have a better idea of the location of the rate.

The Bayesian approach allows formal incorporation of this information, i.e., formal combination of the data and non-data information using the rules of probability. In the context of stress testing, we may take an institution's or the regulators' base scenarios or their adverse scenarios to represent such non-data information as our Bayesian prior. Note that often when building a stress testing model, the developer would be given this information exogenously with respect to the reference data at hand.

The Bayesian approach is most powerful and useful when used to combine data as well as non-data information while incorporating powerful computational techniques such as Markov Chain Monte Carlo methods. Such models are widely discussed in the economics and finance literature and have been applied in the loss estimation settings. These applications invariably specify a "prior," which is convenient and adds minimal information – there is no such thing as an uninformative prior – allowing computationally efficient data analysis. However, this approach, while valuable, misses the true power of the Bayesian approach, i.e., the coherent incorporation of expert information.

---

51. *See* Board of Governors of the Federal Reserve System, Office of the Comptroller of the Currency, *Supervisory Guidance on Model Risk Management* (OCC 2011-12; FED-BOG SR 11-7), available at http://www.federalreserve.gov /bankinforeg/srletters/sr1107a1.pdf.

52. *See supra* Part I.

The difficulty in Bayesian analysis is the representation of expert information in the form of a probability distribution, which requires thought and effort, rather than mere computational power; therefore, it is not commonly followed. Furthermore, in "large" samples, data driven information will typically overwhelm non-dogmatic prior information, so the prior is irrelevant asymptotically, and economists often justify ignoring prior information on this basis.

However, there are many settings in which expert information is extremely valuable, such as when data is scarce, costly, or when its reliability is questionable. These cases can include model performance in unlikely but plausible scenarios (e.g., Lehman Brothers) or the construction of models on a low default portfolio. In the context of stress testing, it is the case that scenarios may be hypothetical and not supported by observed historical data. Such issues more frequently arise in loss estimation, where sufficient data may not be available for certain assets or for new financial instruments, or where structural economic changes, such as the growth of a new derivatives market, may raise doubts about the relevance of historical data.

Empirical analysis in our paper follows the steps in a Bayesian analysis of a stress testing model. Estimation of stressed losses rates for groups of homogeneous assets is essential for determining the amount of adequate capital under stressed scenarios. Since our goal is to incorporate non-data information in our Bayesian analysis, we utilize the supervisory stress testing scenarios to elicit and represent expert information which is used to make inferences in the context of a simple model of loss. In this regard, we are aware that many institutions are moving away from simple linear regression frameworks for CCAR or Dodd-Frank Act Stress Tests, toward models such as proportional hazards or rating migrations; nevertheless, regression-based techniques at an aggregated level are still rather prevalent in the industry, so that we think that there is value in using this as a starting point, and we can consider more advanced techniques for future directions of this research.[53]

The dynamic linear models (DLMs) can be regarded as a generalization of the standard linear regression model, where the regression coefficients are allowed to either change over time, or to be stochastic as in the formulation herein[54]. The Linear Regression Model (LRM)

53. Another reason why banks are moving toward these more advanced approaches is that loss at horizon can very well affect loss in later periods and, in turn, capital actions.

54. For a detailed exposition of the Bayesian approach, *see generally* JOHN GEWEKE, CONTEMPORARY BAYESIAN ECONOMETRICS AND STATISTICS (2005); *see also* CHRISTIAN P. ROBERT, THE

is the most popular tool for relating the variable Y, to a vector of explanatory variables X_t. It is defined as:

$$Y_t = X_t^T \beta + \varepsilon_t \quad t = 1, \ldots, n \quad (1)$$

where $\varepsilon_t \overset{i.i.d.}{\sim} N(0, \sigma^2)$ is a standard Gaussian disturbance term, Y_t is a random variable, and both X_t and β are p-dimensional vectors. In its basic formulation, the variables X_t are considered as deterministic or exogenous; while in stochastic regression, X_t are random variables. In the latter case we have in fact, for each t, a random $(p+1)$ dimensional vector $(Y_t, X_t)^T$, and we have to specify its joint distribution and derive the LRM from it. A way for doing this (but more general approaches are possible) is to assume that the joint distribution is Gaussian:

$$\begin{bmatrix} Y_t \\ X_t \end{bmatrix} \Bigg| \mu, \Sigma \sim N(\mu, \Sigma), \quad \mu = \begin{bmatrix} Y_t \\ X_t \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_y & \Sigma_{XY}^T \\ \Sigma_{XY} & \Sigma_{XX} \end{bmatrix} \quad (2)$$

From the properties of the multivariate Gaussian distribution, we can decompose the joint distribution into a marginal model for X_t and a conditional model for Y_t given $X_t = x_t$ as follows:

$$X_t | \mu, \Sigma \sim N(\mu_X, \Sigma_{XX}) \quad (3)$$

$$Y_t | x_t, \mu, \Sigma \sim N(\beta_1 + \beta_2^T x_t, \sigma_Y^2) \quad (4)$$

where $\mu_y = \beta_1 + \beta_2^T x_t \quad (5)$

$$\beta_2 = \Sigma_{XX}^{-1} \Sigma_{XY}^{-1} \quad (6)$$

$$\beta_1 = \mu_Y - \mu_x^T \beta_2 \quad (7)$$

$$\sigma_Y^2 = \Sigma_{YY} - \Sigma_{XY} \Sigma_{XX}^{-1} \Sigma_{XY}^T. \quad (8)$$

If the prior distribution on (μ, Σ) is such that the parameters of the marginal model and those of the conditional model are independent, then we have a partition in the distribution of (Y_t, X_t, μ, Σ); that is, if our interest is mainly on the variable Y_t, we can restrict our attention to the conditional LRM, which in this case describes the conditional distribution of the latter given (β, Σ) and X_t. Therefore, we may rewrite model (3)-(8) as follows:

$$Y | x, \beta, V \sim N(x\beta, V) \quad (9)$$

BAYESIAN CHOICE: FROM DECISION-THEORETIC FOUNDATIONS TO COMPUTATIONAL IMPLEMENTATION (2001).

where Y = $(Y_1,,,,Y_T)$ and X $=\begin{pmatrix} X_{1,1} & \cdot & \cdot & X_{1,p} \\ \cdot & \cdot & & \cdot \\ \cdot & & \cdot & \cdot \\ X_{T,1} & \cdot & \cdot & X_{T,p} \end{pmatrix}$. Equation (9)

implies a diagonal covariance matrix V $=\begin{pmatrix} \sigma^2 & 0 & \cdot & 0 \\ 0 & \sigma^2 & & \vdots \\ \cdot & & \ddots & \cdot \\ 0 & \cdot & 0 & \sigma^2 \end{pmatrix} = \sigma^2 I_n$.

Therefore, Bayesian inference in this LRM is conditionally independent, with the same variance σ^2. More generally, V can be any symmetric positive-definite matrix.

We describe the Bayesian inference with conjugate priors for the regression model for the case of inference on β and V. If both β and V are random, analytical computations may become complicated; a tractable case is when V has the form V = σ^2D, where σ^2 is a random variable and the $n \times n$ matrix D is known; e.g., D = I_n. Let $\phi = \sigma^{-2}$ be the precision parameter. Then a conjugate prior for (β,ϕ) is a Normal-Gamma distribution, with parameters $\left(\beta_0, N_0^{-1}, a, b\right)$:

$$\pi(\beta,\phi) \propto \phi^{a-1+\frac{p}{2}} \exp[-b\phi] \exp\left[-\tfrac{\phi}{2}(\beta-\beta_0)^T N_0^{-1}(\beta-\beta_0)\right] \quad (10)$$

that is: $\beta|\phi \sim N\left(\beta_0, (\phi N_0)^{-1}\right)$ (11)
$\phi \sim G(a,b)$ (12)

Here conditionally on ϕ, β has covariance matrix $(\phi N_0)^{-1} = \sigma^2 \tilde{C}_0$, where we let $\tilde{C}_0 = N_0^{-1}$, a symmetric pxp positive-definite matrix that "rescales" the observation variance σ^2. This is the version of the Bayesian regression model that we implement in this study.

B. *Bayesian Modeling – An Empirical Implementation*

Our empirical analysis for the implementation of Bayesian methodology to stress testing and model validation follows the CCAR program closely. As part of the Federal Reserve's CCAR exercise, U.S. domiciled top-tier Bank Holding Companies (BHC) are required to submit comprehensive capital plans, including pro forma capital analyses, based on at least one BHC defined adverse scenario which is to be defined by quarterly trajectories for key macroeconomic variables over the next nine quarters or longer, to estimate loss allowances.[55] The BHC scenarios are meant to test idiosyncratic risks per bank (e.g., a cybersecurity attack for a bank with a heavy retail presence). In addition, the Federal Reserve generates its own supervisory stress scenarios, so that firms are expected to apply both BHC and supervisory stress scenarios to all exposures, in order to estimate potential losses under stressed operating conditions. Separately, firms with significant trading activity are asked to estimate a one-time potential trading-related market and counterparty credit loss shock under their own BHC scenarios, and a market risk stress scenario provided by the supervisors. In the case of the supervisory stress scenarios, the Federal Reserve provides firms with global market shock components that are one-time hypothetical shocks to a large set of risk factors.[56]

Table 1 lists the macroeconomic variables used in supervisory stress testing scenarios as part of the Federal Reserve CCAR Program. Our analysis of using the macroeconomic stress scenarios to inform historical analysis is based on the collection of the last three Federal Reserve scenarios for the three macroeconomic variables over these nine quarter periods, i.e., Real Gross Domestic Product (year-to-year change), Unemployment Rate, and HPI - National Housing Price Index. We justify focusing on these three variables as they are the most commonly used for forecasting loan losses (especially as unemployment and GDP growth are perceived to be proxies for economic health), and the most accepted by regulators, as well as having good explanatory power for the target loss variables that we are considering in this study. We consider the two CCAR supervisory stress scenarios in 2011 and 2012, and the supervisory severely adverse scenario in 2013, focusing on aggregate bank gross charge-offs (ABCO) from the Fed Y9 report as a measure of loss. Our historical dataset covers the period from 2000 to 2013.

To the best of our knowledge, this is the first study of its kind which combines data to form the prior three supervisory exercises in stress testing within the Bayesian framework. The reason why we base the prior distributions upon the supervisory scenarios is a practical one, as often modelers will use the quality of the supervisory scenarios as a criterion in model development, for example, a common practice being testing the redeveloped model with the prior years' scenarios. Of

55. *See supra* Part II.

56. In addition, large custodian banks are asked to estimate a potential default of their largest counterparty. For the last two CCAR exercises, these shocks involved large and sudden changes in asset prices, rates, and Credit Default Swap spreads that mirrored the severe market conditions in the second half of 2008.

course, model developers have the option to use their own internally developed scenarios, or information gleaned from subject matter experts such as the lines of business, in order to form their priors.

Table 2 presents the summary statistics of and correlation between the macroeconomic variables for both our historical dataset and the Federal Reserve scenario we use in our empirical analysis. Figure 1 displays the time series and kernel density plots of these variables for both datasets. ABCO, a measure of bank losses, averages 70 basis points over the last 14 years since 2000, peaking at 2.72% toward the end of 2009. ABCO is extremely skewed toward periods of mild loss during early 2000s, having a mode of around 20-25 basis points. Year-on-Year Change in Real Gross Domestic Product (RGDPYY) historically averages 1.94%, while in the Federal Reserve stress scenarios it displays an average contraction of -0.73%, having mild positive and negative skews in historical and Federal Reserve scenario data, respectively.

Figure 1 shows that while the historical distribution of annual Real GDP changes is bimodal, having modes at around 5% and 9%, which represents the historical regime shift between expansionary and contractionary economic periods, annual real GDP changes in the Federal Reserve scenario distribution has a single mode at around zero. Unemployment has an historical average of 6.4% (ranging from 4% to 10%), while in the Federal Reserve scenarios it is centered at 11.9% (ranging from 10% to 14%). As with GDP, Unemployment displays a bimodal distribution, with modes of 4% and 9% (10% and 14%) considering the historical data (data from Federal Reserve scenarios). The historical average of HPI has an historical average of 150.2 (ranging from 101.6 to 199.0), while in the Federal Reserve scenarios it is centered at 129.0 (ranging from 112.8 to 142.4). As with the other macroeconomic variables, HPI displays a bimodal (unimodal) distribution, with modes of 140 and 180 (135) considering the historical data from Federal Reserve scenarios.

We estimate univariate, bivariate, and trivariate Bayesian as well as Frequentist models for each of the three macroeconomic variables by forming priors using univariate regressions. In our empirical implementation, our data sample is the history, which includes historical economic statistics, and the prior sample is formed from the macroeconomic statistics of the last three pooled Federal Reserve scenarios. Dependent variables for prior regressions are established by calculating the historical quantile of each macroeconomic variable based upon the scenario dataset, and using the historical value of ABCO at that quantile as the response variable.

Our empirical results for all regressions are presented in Table 3 and Figure 2. In order to conserve space, we only display the density plots and posterior distributions of trivariate Bayesian regressions for aggregate bank gross charge-offs versus each variable, identifying corresponding macro-sensitivity.

The results for estimating the posterior distributions on macro-sensitivities in univariate regression are presented in Panel A of Table 3. The historical data (Federal Reserve scenarios) estimate of the coefficient for real GDP changes is -0.2267 (-0.3953), resulting in a posterior estimate of -0.2500, which is slightly higher in absolute value. In the case of Unemployment, coefficient estimates are 0.3735, 0.4263, and 0.4 for historical, Federal Reserve scenario, and posterior, respectively. For HPI, historical data estimate of the coefficient is -0.0044 while for the Federal Reserve scenarios it is -0.0271, which translates into a posterior estimate of -0.0150 that is much higher in absolute value. Therefore, in general the posterior estimates are greater in absolute value, reflecting greater sensitivity as observed in the scenario dataset that informs the prior. Note that there is no loss of generality, as model developers could use their own priors formed from internal views on scenarios or expert opinion in lieu of prior Federal Reserve scenarios, and sensitivities may be reduced as well.

Panel B of Table 3 presents the results for bivariate regressions and, similar to univariate results, Federal Reserve scenarios have greater sensitivity than historical estimates. In the case of the pair of Real GDP changes and Unemployment, while the posterior estimate for Real GDP changes (-0.2147) is higher in absolute value, it is counter-intuitively lower for Unemployment (0.0062). One possible reason for this is that the Federal Reserve scenario dataset shows a correlation between Unemployment and Real GDP changes relative to the historical pattern that is such that the posterior estimate is pulled in an unintuitive direction. When we consider the Unemployment and HPI pair, we observe the posterior estimate for Unemployment lower in absolute value (0.3036 vs. 0.3738), which we also find to be counter-intuitive, and explain similarly as in the case of the Unemployment vs. Real GDP Changes pair. The posterior estimate for the HPI sensitivity is estimated to be larger in absolute value (|-0.0073| vs. |0.0001|). In the case of Real GDP changes and HPI, the historical coefficient data estimate for Real GDP Changes (HPI) is -0.2248 (-0.0038), and based on the prior scenarios having greater sensitivity of -0.3953 (-0.0271). For both macro variables, the posterior estimates are found to be higher in absolute value (—-0.2435— vs. —-0.0091—).

As with the univariate and bivariate regressions, we find for the trivariate model that in general the absolute value of macro-sensitivities are greater in magnitude in the Federal Reserve scenario regressions than in historical ones. The results for trivariate model estimation are presented in Panel C of Table 3. For Real GDP changes, the historical data (Federal Reserve scenarios) estimate of the beta coefficient is -0.1014 (-0.3953) and the posterior estimate is -0.1463, which is higher in absolute value. In the case of Unemployment, the historical data (Federal Reserve scenarios) estimate of the beta coefficient is 0.3345 (0.4263, which is indicative of greater sensitivity). The posterior parameter estimate for Unemployment is counter-intuitively lower (0.2594). Considering the housing index (HPI), coefficients are -0.0001 and -0.0065 for historical data and posterior estimates, respectively.

In Figure 3, we present the Bayesian and Frequentist models, the historical loss rates, and the forecasted scenario loss rates (Bayesian vs. Frequentist modeled in the case of the severely adverse scenarios and the Frequentist modeled for the base case). We observe that while the models tend to under- (over-) predict in the stress (recent benign) period, optically the Bayesian model actually performs worse in the stress period and better in the recent period. However, in the severe adverse scenario, the Bayesian modeled losses reach more extreme levels than those Frequentist modeled ones – this is a good property from a supervisory perspective, as it reflects greater conservativism. Additionally, we observe a steeper reversion to normal levels of loss in the Bayesian model than in the Frequentist one.

Figure 4 displays the estimated posterior distributions of the severely adverse loss rates for each quarter and of the cumulative 9-quarter losses. Figure 4 – Panel A shows that losses peak on average at around the third to sixth quarters, with the densities shifting leftward, and that the dispersion of the distributions also increases. This poses a dilemma revealed by the Bayesian approach that there is more parameter uncertainty just in the periods of its greatest importance for stress testing, the periods of peak stress in the economic scenarios. Figure 4 – Panel B shows that cumulative losses are centered in the low 40 percentages, with a fairly large relative variation (30%) with respect to the mean. Moreover, there is significant right-skewness, suggesting that the mean posterior loss rate may not be the most representative of the posterior distribution.

In Table 4, we summarize the posterior conditional severely adverse loss distributions of the Bayesian regression model, for both quarterly and cumulative 9-quarter ABCOs, presenting both summary statistics and numerical Bayesian coefficients of variation (BNCV), which measures the relative variation in the posterior samples. The BNCV is defined as the ratio of the Bayesian 95th percentile credible interval (B95CI) – which is the Bayesian analogue of the classical 95th percentile confidence interval – to the mean of the posterior distribution:

$$BNCV_X^{95} = \frac{Q_{97.5}(\mathrm{x}) - Q_{2.5}(\mathrm{x})}{\frac{1}{N}\sum_{i=1}^{N} x_i} = \frac{B95CI_X}{\overline{x}} \quad (13)$$

where X is a random variable, x = $(x_1,..,x_N)$ is a vector of draws from the posterior distribution of X, $Q_{97.5}$ and $Q_{2.5}$ are the respective empirical 97.5th and 2.5th quantiles of x, and $\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$ is the posterior sample mean of x.

Figure 5 presents the posterior conditional severely adverse quarterly-loss distributions of the Bayesian regression model. Variability in the loss distribution displays a humped shape across the nine quarters, with the highest losses observed around the 5th and 6th quarters. In addition, the humped shape of the distributions become skewed for worst losses (97.5th percentile and maximum), while it is more symmetric for optimistic loss outcomes (minimum and 2.5th percentile). Results for the BNCV measure, shown in Table 4, display a humped shaped pattern in variability over the forecast horizon. The BNCV measure can be interpreted as the proportional model risk uncertainty buffer, stemming from the parameter uncertainty as inferred from the Bayesian regression model. This result is an important contribution of our research focusing on the model validation aspect of stress testing, as we have a quantity that can be used in model monitoring or backtesting, which is not purely based upon data but also incorporates prior views on model parameters.

In Table 5, we summarize the conditional severely adverse loss distributions of the Frequentist regression model, for both quarterly and cumulative 9-quarter aggregate bank gross charge-off rates. Classical Frequentist coefficients of variation (FCV) is the simple ratio of the Frequentist 95th percent confidence interval (F95CI) to the mean of the sampling distribution:

$$FCV_X^{95} = \frac{(\overline{x} + 1.96 \times s_X) - (\overline{x} - 1.96 \times s_X)}{\overline{x}} = \frac{F95CI_X}{\overline{x}} \quad (14)$$

where s_X is the standard error of the forecast mean. Losses average a range of 0.9% to 1.7% in the first two quarters, peaking at a mean ranging in 2.3%-3.6% and reverting to a mean of 1.2% to 1.7% in the final two quarters. We observe that the variability in the loss distribution displays a U-shape, peaking in the low loss early and end quarters: the standard error drops from a range of 0.34% -0.38%, to a range of 0.28%-0.34%, and then rises to a range of 0.41%-0.46%. This observation on the pattern in variability over the forecast horizon holds on a relative basis as well, by considering the FCV. The relative variability in the loss distribution displays a U-shape, peaking in the early and later quarters: the FCV decreases from a range of 77%-161%, to a range of 32%-56%, and then rising to a range of 94%-153% in the first two, middle and last two quarters, respectively. The FCV measure can be interpreted as the proportional model risk uncertainty buffer, stemming from the sampling error as inferred from the Frequentist regression model.

We contribute to the model validation literature by comparing the proportional *model risk buffer* measures obtained from our empirical implementation of the Bayesian to the Frequentist models. One common way to estimate a model risk buffer is as a measure of statistical uncertainty generated by a model, such as a standard error or a confidence interval; other means of quantifying this metric include sensitivity analysis around model inputs or model assumptions, i.e., varying the latter and measuring the variability of the model output. The model risk buffer is a valuable model validation tool, as it helps us to understand the potential expected variability in model output - e.g., when we perform model benchmarking or backtesting, we can gauge if new observation of actual results are lying in an expected range, and this can serve as a basis for remedial actions such as model overlays, or potentially re-developing a flawed model whose outcomes are not within the expected range.

The mean of the posterior distribution in the 9-quarter severely adverse loss generated by the Bayesian model is 43.2%, with a Bayesian 95th percent credible interval of 11.0%, resulting in a BNCV of 25.5%. The mean of the sampling distribution in the 9-quarter severely adverse loss generated by the Frequentist model is 20.6%, with a classical 95th percent confidence interval of 4.1%, resulting in a FCV of 20.0%. Therefore, our Bayesian analysis suggests that a quantitatively developed model risk uncertainty buffer to account for parameter uncertainty is 5% (20%) higher in absolute (relative) terms than that implied by the Frequentist model.

We compare the Bayesian and Frequentist stress testing models according to several measures of model performance, as commonly used in model validation exercises. First, we use the root mean squared error (RMSE), which measures the average squared deviation of model predictions from actual observations:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i - A_i)^2} \quad (15)$$

where P_i are predicted and A_i actual. Secondly, we calculate squared-correlation (SC) between model predictions and actual observations:

$$SC = \left[\frac{\sum_{i=1}^{N}\left[\left(P_i - \frac{1}{N}\sum_{i=1}^{N}P_i\right)\left(A_i - \frac{1}{N}\sum_{i=1}^{N}A_i\right)\right]}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(P_i - \frac{1}{N}\sum_{i=1}^{N}P_i\right)^2}\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(A_i - \frac{1}{N}\sum_{i=1}^{N}A_i\right)^2}}\right]^2 \quad (16)$$

Finally, we consider a measure widely used in model validations of stress testing models for CCAR or Dodd-Frank Act Stress Tests, the cumulative percentage error (CPE), which is favored by prudential regulators:

$$CPE = \frac{\sum_{i=1}^{N}(P_i - A_i)}{\sum_{i=1}^{N}A_i} \quad (17)$$

We estimate these model performance measures, in-sample, and across the entire historical period (2001-2013) and over the twelve quarter downturn period (2006-2009).[57] In addition, we estimate the sampling distributions of these measures using a bootstrap procedure, in order to test the statistical significance of the observed differences in model performance measures. We observe, in Table 6, that the Frequentist model outperforms the Bayesian model according to RMSE and SC measures (10.1% and 84.5% vs. 15.4% and 75.2%, in mean, for the entire sample; 13.9% and 92.0% vs. 19.6% and 87.0%, in mean, for the downturn sample). However, the Bayesian model outperforms, over the entire sample as well as during the stressed period, according to the CPE measure (7.9% and 5.9%, in mean, for the entire sample; -13.7% and -12.1%, in mean, for the downturn sample). It is not surprising that the Frequentist model performs better when RMSE and SC measures are used for validation, since it is a model which is purely calibrated to the historical data.

The reason for the Bayesian approach's superior performance, using the CPE-preferred measure of model validators and supervisors, is

57. A best practice would be to have out-of-sample or out-of-time model performance metrics, but paucity of data precludes that in this case, as it does in most CCAR and DFAST stress testing modeling validations. However, these in-sample measures represent the current state of the practice and are in line with supervisory expectations.

that this model constrains the regression coefficients to exhibit more sensitivity, so that when there are large losses, the model matches actuals to a great degree than when losses are toward the middle of the distribution – intuitively, we are able to better match the tails of the error distribution than its body. In contrast, the Frequentist regression model simply tries to minimize the total squared deviation over the entire sample, which is modeling the body but not the tail of the error distribution. Note, moreover, that we could also impose alternative priors – e.g., informed by external data, internal scenarios, or expert opinion – which could either accentuate this effect, or even work in the opposite direction and dampen sensitivities.

IV. CONCLUSION

Since the 2008 financial crisis, stress testing has become the primary means by government regulators in the United States to ensure that banks have enough capital to survive another financial crisis. However, there are no guiding models for stress testing despite the importance of stress testing to the financial regulatory landscape.

This Article fills this void by proposing that banks utilize the Bayesian model in their stress tests, which takes into account past outcomes – previous Federal Reserve scenarios serve as priors. This is because of the belief in the industry that the Federal Reserve adapts its scenarios to stress certain portfolios.

The Bayesian model requires a bigger buffer for uncertainty – by 25% – as opposed to simply modeling each year's scenario in isolation. This means that if modelers do not take previous results into account, they can underestimate losses significantly – by as much as 25%. This could be the difference between a successful stress test and a failed stress test.

Stress testing is an emerging field, and banks are constantly trying to forecast potential losses in a future recession to be able to manage their capital effectively. Therefore, more innovations in modeling credit risk will lead to better models – models that can incorporate expert judgment, as well as the relationship between certain types of losses. In turn, this will keep the American financial system secure from another crisis.

TABLE 1

Federal Reserve Comprehensive Capital Analysis and Review Program (CCAR)

Macroeconomic Variables Used in Supervisory Stress Testing Scenarios

CCAR Date	Macroeconomic Variable (MV)	Code
2011	Real GDP (year-to-year)	RGDPYY
	Consumer Price Index	CPI
	Real Disposable Personal Income	RDPI
	Unemployment Rate	UNEMP
	Three-month Treasury Bill Rate	3MTBR
	Ten-year Treasury Bond Rate	10YTBR
	BBB Corporate Rate	BBBCR
	Dow Jones Index	DJI
	National House Price Index	HPI
2012	***All 2011 CCAR MVs +***	
	Nominal GDP Growth	RGDPG
	Nominal Disposable Income Growth	NDPIG
	Mortgage Rate	MR
	CBOE's Market Volatility Index	VIX
	Commercial Real Estate Price Index	CREPI
2013	***All 2011 & 2012 CCAR MVs***	

TABLE 2

Summary Statistics and Correlations
Macroeconomic Variables and Bank Charge-offs
Historical Data vs. Federal Reserve Scenarios
(2001 – 2013)

	Statistic	ABCO	RGDPYY	UNEMP	HPI
		Macroeconomic Variables			
Historical Data	Count	56	56	56	56
	Mean	0.70	1.94	6.39	150.21
	Standard Deviation	0.79	1.90	1.88	26.63
	Minimum	0.01	-4.09	3.92	101.60
	25th Percentile	0.11	1.43	4.81	136.23
	Median	0.22	2.10	5.77	144.45
	75th Percentile	1.29	3.12	8.05	166.60
	Maximum	2.73	5.27	9.91	199.00
	Skewness	1.10	-1.48	0.56	0.29
	Kurtosis	-0.11	2.84	-1.09	-0.64
Fed Scenarios	Count		21	30	30
	Mean		-0.73	11.90	129.02
	Standard Deviation		2.02	1.73	9.84
	Minimum		-4.31	9.58	112.80
	25th Percentile		-1.96	10.54	120.80
	Median		-0.58	11.49	128.85
	75th Percentile		0.72	13.65	136.90
	Maximum		2.14	14.31	142.40
	Skewness		-0.37	0.05	-0.25
	Kurtosis		-0.99	-1.62	-1.19
Correlations	ABCO	100%	-	-	-
	RGDPYY	-54.35%	100%	-	-
	UNEMP	88.55%	-37.91%	100%	-
	HPI	-14.91%	3.90%	-17.18%	100%

ABCO: Aggregate Bank Gross Charge-offs
RGDPYY: Year-on-Year Change in Real Gross Domestic Product
UNEMP: Unemployment Rate
HPI: National Housing Price Index

TABLE 3

Estimation of Stress Testing Macroeconomic Models Bayesian with Federal Reserve Scenario Priors vs. Historical Models GDP, Unemployment Rate, and Housing Index

	RGDPYY	UNEMP	HPI
Panel A: *Univariate Regressions - Posterior Distribution (β)*			
Historical Model	-0.2267	0.3735	-0.0044
	(-4.758)	**(14.00)**	**(2.233)**
Fed Scenarios	-0.3953	0.4263	-0.0271
	(-7.770)	**(9.844)**	**(-4.446)**
Bayesian Model	-0.2500	0.4000	-0.0150
	(-5.435)	**(13.79)**	**(-4.054)**
Panel B: *Bivariate Regressions - Posterior Distribution (β)*			
Historical Model - RGDPYY & UNEMP	-0.1013	0.3347	
	(-4.023)	**(13.14)**	
Bayesian Model - RGDPYY & UNEMP	-0.2147	0.0062	
	(-4.226)	**(2.138)**	
Historical Model - UNEMP & HPI		0.3738	0.0001
		(13.67)	(0.049)
Bayesian Model - UNEMP & HPI		0.3036	-0.0073
		(820.5)	**(-304.2)**
Historical Model - RGDPYY & HPI	-0.2248		-0.0038
	(-4.725)		(-1.123)
Bayesian Model - RGDPYY & HPI	-0.2435		-0.0091
	(-46.83)		**(-267.6)**
Panel C: *Trivariate Regressions - Posterior Distribution (β)*			
Historical Model - RGDPYY, UNEMP & HPI	-0.1014	0.3345	-0.0001
	(-3.985)	**(12.82)**	**(-0.058)**
Bayesian Model - RGDPYY, UNEMP & HPI	-0.1463	0.2594	-0.0065
	(-38.50)	**(720.6)**	**(-295.5)**

T-statistics are presented in parenthesis with significant ones identified in bold.

RGDPYY: Year-on-Year Change in Real Gross Domestic Product

UNEMP: Unemployment Rate

HPI: National Housing Price Index

TABLE 4
Posterior Conditional Distributions of Bayesian Regression Model Quarterly and Cumulative 9-Quarter Aggregate Bank Gross Charge-off Rates
Summary Statistics and Numerical Coefficients of Variation

Severe Losses

	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Quarter 5	Quarter 6	Quarter 7	Quarter 8	Quarter 9	Cumulative
Minimum	0.41	0.99	1.32	1.61	1.74	1.75	1.53	0.84	0.00	21.01
2.5th Percentile	0.65	1.39	2.29	2.95	2.94	2.47	2.08	1.16	0.19	37.80
25th Percentile	0.75	1.61	2.84	3.74	3.63	2.87	2.30	1.38	0.59	41.32
Mean	0.80	1.73	3.13	4.15	3.99	3.08	2.42	1.50	0.80	43.23
75th Percentile	0.86	1.85	3.41	4.56	4.35	3.29	2.54	1.62	1.00	45.12
97.5th Percentile	0.96	2.07	3.99	5.37	5.06	3.71	2.77	1.86	1.41	48.81
Maximum	1.12	2.46	4.77	6.42	6.04	4.39	3.31	2.24	2.01	55.24
Skewness	-0.02	0.00	0.02	0.02	0.03	0.03	0.01	0.02	0.02	0.01
Kurtosis	0.12	0.14	0.09	0.09	0.09	0.10	0.19	0.04	0.03	0.41
Standard Error	0.08	0.17	0.43	0.61	0.54	0.32	0.18	0.18	0.31	2.82
95th Percentile Bayesian Credible Interval	0.30	0.67	1.69	2.42	2.12	1.25	0.70	0.70	1.23	11.01
95th Percentile Bayesian Numerical Coefficient of Variation	37.61	38.97	54.10	58.42	53.26	40.57	28.85	46.31	153.48	25.49

TABLE 5
Distributions of Frequentist Regression Modeling Quarterly and 9-Quarter Cumulative Aggregate Bank Gross Charge-off Rates
Summary Statistics and Parametric Coefficients of Variation

Severe Losses

	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Quarter 5	Quarter 6	Quarter 7	Quarter 8	Quarter 9	Cumulative
Expected Value	0.92	1.71	2.78	3.56	3.49	2.84	2.38	1.71	1.20	20.58
Standard Error of Forecast	0.38	0.34	0.29	0.29	0.28	0.30	0.34	0.41	0.47	1.05
95th Percentile Frequentist Confidence Interval	1.47	1.33	1.14	1.13	1.11	1.19	1.34	1.60	1.83	4.11
95th Percentile Frequentist Coefficient of Variation	160.67	77.78	41.16	31.64	31.95	42.02	56.48	93.74	153.44	19.99

TABLE 6
Summary Statistics and T-Tests of the Mean for Resampled Model Validation Performance Measures Frequentist vs. Bayesian Regression Models for Stress Testing Bank Aggregate Gross Charge-offs

Model Performance Metric	Entire History (2001-2013)						Economic Downturn Period (2006-2009)					
	Root Mean Squared Error		Squared Correlation		Cumuluative Percentage Error		Root Mean Squared Error		Squared Correlation		Cumuluative Percentage Error	
Summary Statistic	Bayesian	Frequentist	Bayesian	Frequentist	Bayesian	Frequentist	Bayesian	Frequentist	Bayesian	Frequentist	Bayesian	Frequentist
Minimum	0.0001	0.0001	0.4540	0.6322	-0.7502	-0.6467	0.0053	0.0002	0.1922	0.5422	-0.2909	-0.2267
2.5th Percentile	0.0006	0.0002	0.6457	0.7582	-0.7370	-0.6350	0.0053	0.0002	0.7183	0.8364	-0.1920	-0.1628
25th Percentile	0.0142	0.0087	0.7247	0.8238	-0.2188	-0.1238	0.0142	0.0153	0.8399	0.9020	-0.1434	-0.1237
Mean	0.1539	0.1007	0.7519	0.8452	0.0588	0.0792	0.1967	0.1386	0.8700	0.9201	-0.1212	-0.1369
75th Percentile	0.2511	0.1728	0.7849	0.8722	0.4256	0.3338	0.3442	0.2516	0.9126	0.9444	-0.0980	-0.0888
97.5th Percentile	0.5628	0.4032	0.8290	0.9043	0.6498	0.5613	0.5628	0.4183	0.9606	0.9719	-0.0572	-0.0585
Maximum	0.5894	0.4183	0.8799	0.9314	0.7677	0.5969	0.5628	0.4183	0.9976	0.9962	0.0103	-0.0096
Skewness	1.0506	1.1006	-0.7553	-0.8699	-0.1222	-0.4050	0.7224	0.7281	-1.4791	-1.3374	-0.2249	-0.2882
Kurtosis	0.0114	0.1482	1.2594	1.3246	-0.7845	-0.3892	-1.0075	-0.7629	5.3417	4.4183	0.2731	0.2945
Standard Error	0.1673	0.1167	0.0470	0.0376	0.3898	0.3115	0.1973	0.1384	0.0623	0.0352	0.0342	0.0266
T-statistic	2,611.49		15,516.67		410.61		2,409.76		6,988.91		3,618.85	
P-value of t-test	0.0000		0.0000		0.0000		0.0000		0.0000		0.0000	

FIGURE 1
Time Series and Kernel Density Plots
Historical vs. Federal Reserve Scenario Macroeconomic Variables

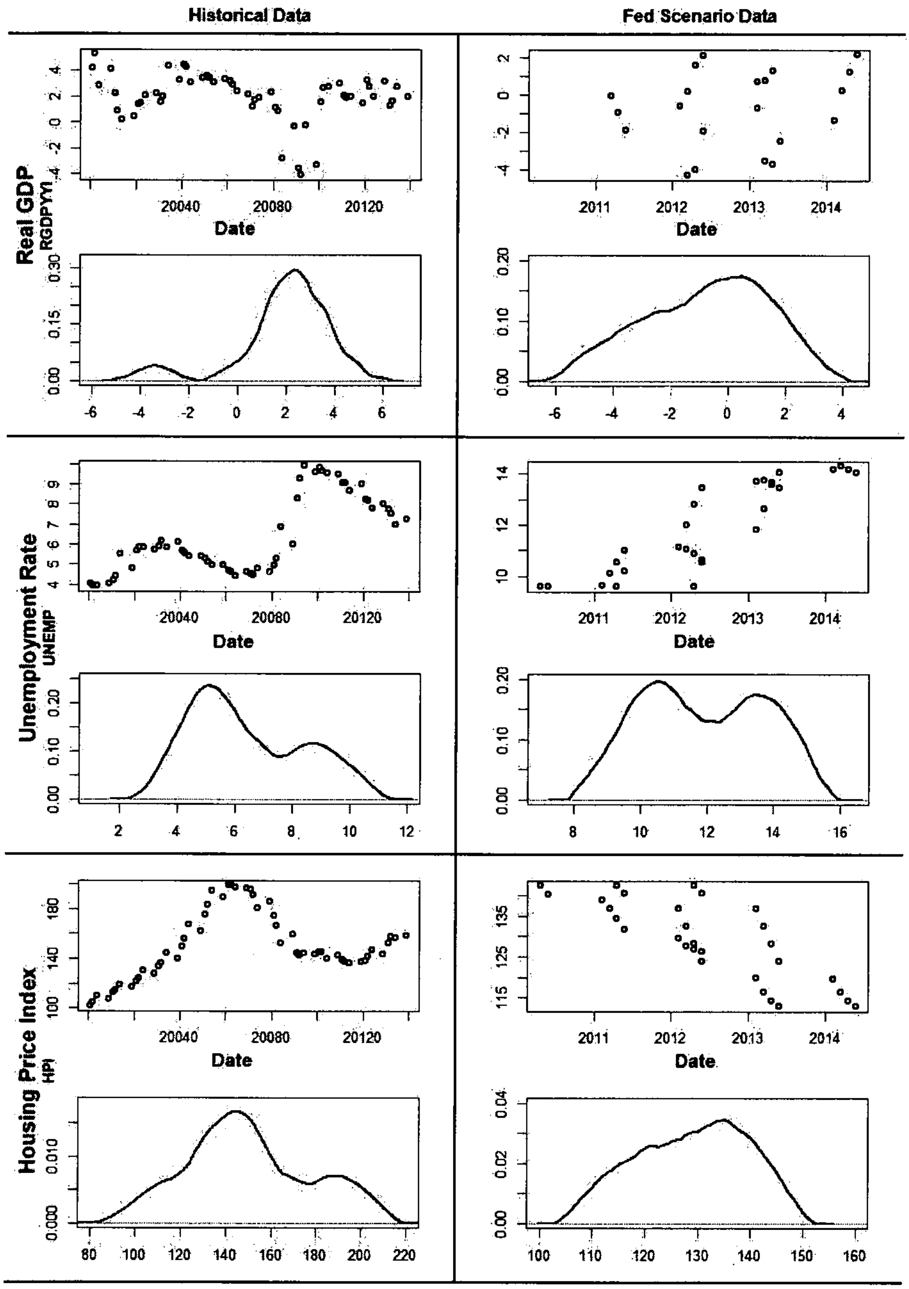


FIGURE 2

Density Plots and Posterior Distributions of Trivariate Bayesian Regressions for Aggregate Bank Gross Charge-offs (2001-2013)

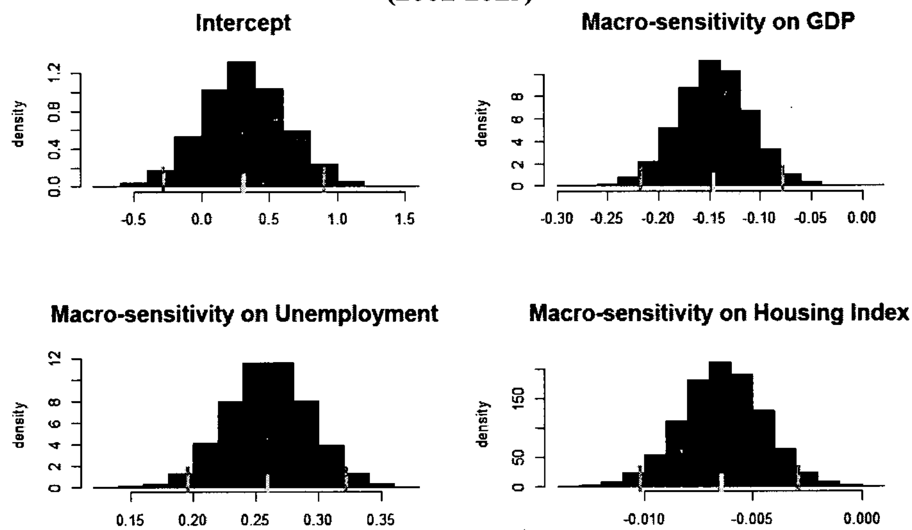


FIGURE 3

Historical, Scenario and Bayesian vs. Frequentist Regression Modeling Aggregate Bank Gross Charge-off Rates

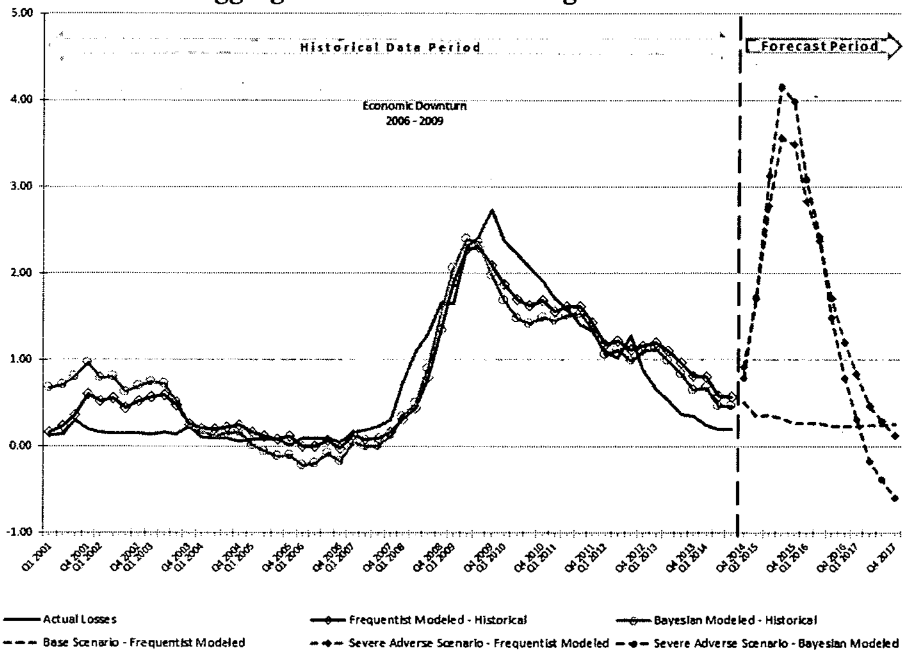
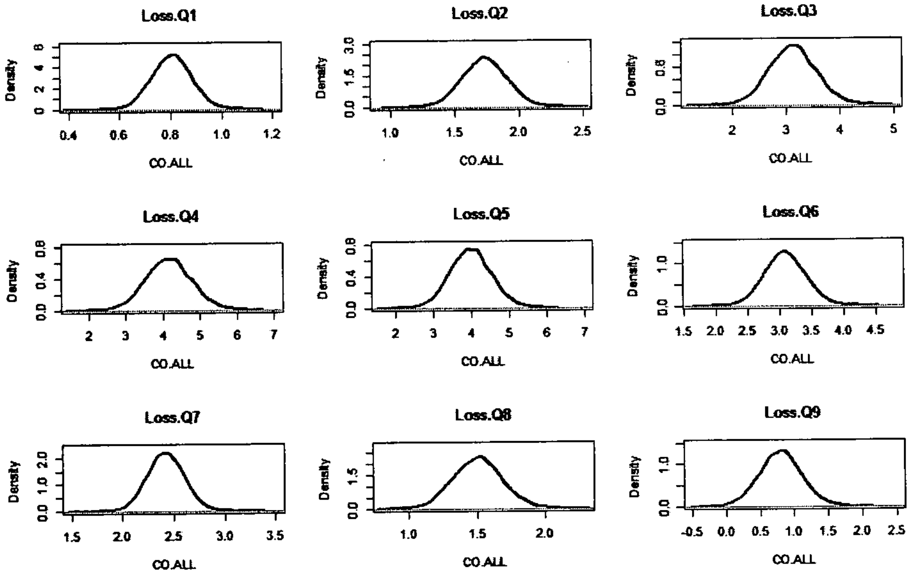


FIGURE 4

Posterior Conditional Distributions of Bayesian Regression Modeling Quarterly Aggregate Bank Gross Charge-off Rates



PANEL B

POSTERIOR CONDITIONAL DISTRIBUTION OF BAYESIAN REGRESSION MODELING CUMULATIVE 9-QUARTER AGGREGATE BANK GROSS CHARGE-OFF RATES

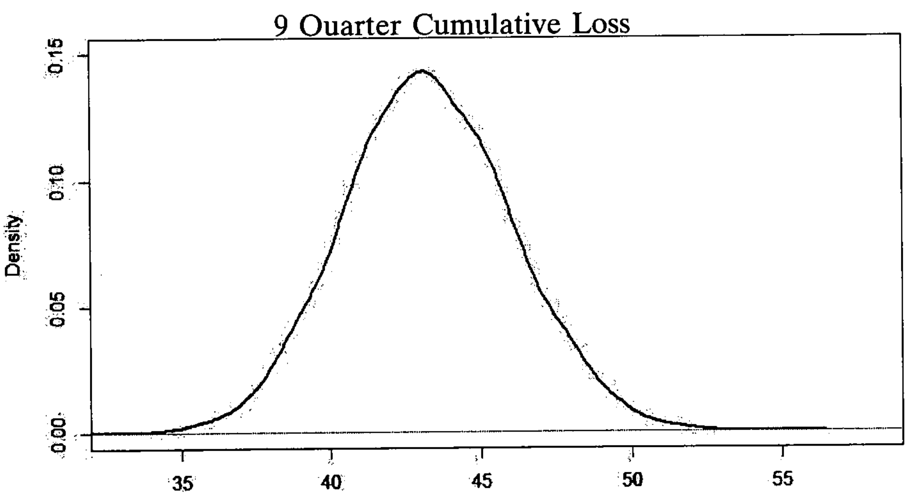


FIGURE 5

Posterior Conditional Distributions of Bayesian Regression Model Quarterly Aggregate Bank Gross Charge-off Rates