

Kupiec, Paul H.

Working Paper

Inside the black box: The accuracy of alternative stress test models

AEI Economics Working Paper, No. 2017-03

Provided in Cooperation with:

American Enterprise Institute (AEI), Washington, DC

Suggested Citation: Kupiec, Paul H. (2018) : Inside the black box: The accuracy of alternative stress test models, AEI Economics Working Paper, No. 2017-03, American Enterprise Institute (AEI), Washington, DC

This Version is available at:

<https://hdl.handle.net/10419/280556>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



Inside the Black Box: The Accuracy of Alternative Stress Test Models

Paul H. Kupiec
American Enterprise Institute

AEI Economics Working Paper 2017-03
March 2017

© 2017 by Paul H. Kupiec. All rights reserved.

The American Enterprise Institute (AEI) is a nonpartisan, nonprofit, 501(c)(3) educational organization and does not take institutional positions on any issues. The views expressed here are those of the author(s).



The Leverage Ratio Is Not the Problem

Paul H. Kupiec
American Enterprise Institute

AEI Economics Working Paper 2017-17
August 2017

© 2017 by Paul Kupiec. All rights reserved.

The American Enterprise Institute (AEI) is a nonpartisan, nonprofit, 501(c)(3) educational organization and does not take institutional positions on any issues. The views expressed here are those of the author(s).

Inside the Black Box: The Accuracy of Alternative Bank Stress Test Models

By Paul H Kupiec¹

December 27, 2017

ABSTRACT

Many jurisdictions use regulatory forecasts of bank performance over multi-year economic stress scenarios to set institutions' regulatory capital requirements and yet little is known about the accuracy of these forecasts. I use the 2008 financial crisis to assess the accuracy of alternative stress test modeling approaches. Stress test models calibrated using traditional econometric approaches similar to those that have been used by the Federal Reserve produce large forecast errors even though these models fit the estimation data exceptionally well. In contrast, machine learning calibration methods produce vastly superior stress scenario forecasts. The large forecast errors generated by commonly employed stress testing approaches highlight the need to develop techniques to assess the accuracy the regulatory stress tests models.

Key words: stress test accuracy, LASSO, Federal Reserve CLASS models

¹ Resident Scholar, American Enterprise Institute. 1789 Massachusetts Avenue, NW, Washington, DC 20036
Email: paul.kupiec@aei.org office: 202-862-7167

Inside the Black Box: The Accuracy of Alternative Bank Stress Test Models

1. Introduction

The success of the 2009 Federal Reserve Supervisory Capital Assessment Program spawned a new paradigm of bank regulation built around forecasts of bank performance under stressful conditions. While details differ, many countries now use “dynamic stress tests” in their large bank supervision process. These stress tests use econometric models to forecast bank income and regulatory capital over one or more hypothetical multi-year economic stress scenario. Banks are required to project their performance under the specified stress conditions and regulators evaluate banks estimates by comparing bank forecasts to projections from a supervisory stress test model.

In evaluating a bank’s stress scenario performance, forecasts of significant losses can trigger remedial supervisory actions. For example, in the US, the Dodd-Frank Act² requires the Federal Reserve Board (FRB) to run annual stress tests on the largest financial institutions. Should an institution fail to maintain its required minimum regulatory capital position throughout the stress test simulation, the FRB may prohibit the bank from paying dividends or making capital repurchases. If the FRB finds a bank’s internal stress testing processes to be deficient, it can require the bank to improve its models and processes.

In the US, the FRB uses its own stress test model estimates to evaluate the results produced by individual banks’ stress test models. The Dodd-Frank Act includes no requirement that the FRB document and disclose the accuracy of the models it uses in its stress test evaluations. Moreover, there is no widely-accepted method for comparing stress test models and determining which stress test model—the FRB’s model or the bank own internal model—produces a more accurate forecast of bank performance over a hypothetical stress scenario.

² Dodd-Frank Wall Street Reform and Consumer Protection Act.

The lack of statistical methods and legal processes to protect banks against a mistake in a supervisory stress test assessment is an important issue. The FRB’s stress test models are designed to simulate the performance of “an average bank.” They are not fine-tuned for each individual bank. It would be surprising if a bank’s own well-designed stress test model did not produce a more accurate forecast than a supervisory model calibrated for an average bank, and yet there is currently no mechanism to protect the bank from the consequences of an inaccurate supervisory stress test assessment.³

In this paper, I take advantage of a true historical stress scenario to assess the forecast accuracy of competing stress test models using the 2008 financial crisis. I compare accuracy of eight different stress test models by comparing stress test model predictions to actual bank performance over the 2008 financial crisis.

I estimate alternative stress test model specifications for an “average” or representative bank using quarterly bank regulatory data from March 1993 through June 2008, and use these estimates to forecast quarterly bank income before tax and extraordinary items (INBFTXEX) over a three-year period beginning in the third quarter of 2008.⁴ My analysis focuses on INBFTXEX because, by far, INBFTXEX is the most important determinant of a bank’s capital adequacy over a dynamic stress test simulation.⁵

³ The FRB recognizes that projections from the supervisory model may differ from bank’s own individual projections. In Federal Reserve Board (2012) the FRB states (p.11), “Further, because the [FRB] projections are based on a set of standardized models applied to all 19 BHCs, they will differ from projections that the individual BHCs will make of their own performance under the same set of hypothetical adverse conditions.” However, there is no discussion of processes or procedures that are in place that are used to determine which model is more accurate. Instead, the FRB adopts an approach that subjectively “averages” its supervisory projections with those of the bank to produce its final stress test projections. No information is provided to substantiate the accuracy of the FRB’s subjective “averaging process”.

⁴ My stress test models are so-called “top-down” models that use publicly available accounting and regulatory data as modeling inputs. They are similar to the Federal Reserve’s CLASS models and have similar within-sample goodness-of-fit statistics. The FRB’s Dodd-Frank stress test model is reportedly a so-called “bottom-up” model that use granular proprietary internal bank data on asset characteristics and performance as inputs into their stress model projections. The relationship between these modeling approaches is discussed in Section 3 of this paper.

⁵ Extraordinary gains and losses, by their very nature, should not be predictable using macroeconomic factors. Retained earnings are INBFTXEX, less tax (or plus tax refunds) and capital distributions. Capital adequacy calculations require, in addition, an estimate of retained earnings and bank risk-weighted assets. The most important part of the capital adequacy projection process is the stress test estimate of INBFTXEX.

The model specifications I consider vary according to the level of disaggregation and the set of explanatory variables used to calibrate the models. The most complicated stress test model I consider is a model designed to mimic the approach used to formulate the Federal Reserve's CLASS model.⁶

My selection of models is designed to investigate the benefits of alternative characteristics of stress test models. One important characteristic is the level of disaggregation used to generate stress test forecasts. To assess the potential benefits of using a disaggregated approach, I compare the forecast accuracy of a two equation model (asset growth, and the ratio of INBFTXEX to assets) to a six equation model that decomposes the INBFTXEX-to-asset ratio into five separate components and an additional model for forecasting asset growth.

Another common stress testing practice that has unknown implications for forecasting accuracy is the inclusion of time-varying bank-specific characteristics as explanatory variables. Time-varying bank characteristics are endogenous. They change in response to stress scenario conditions, and yet regulatory stress tests projections typically keep these values fixed over the stress scenario horizon. I investigate the benefits of including bank-specific characteristics as explanatory variables in addition to a constant, a lagged dependent variable, and a comprehensive set of macroeconomic factors.

The third dimension of stress test modeling I explore is the method used to calibrate stress test models. For each set of explanatory variables and level of model aggregation, I apply two different calibration schemes. One approach uses traditional econometric model selection criteria to calibrate the model. This approach mimics the approach used to specify the Federal Reserve CLASS model. A parsimonious model is produced in a step-wise fashion by eliminating explanatory variables that are not statistically significant and whose inclusion reduces a regression's adjusted R-squared statistic or increases a regression model's mean square error.

⁶ The CLASS model is described in detail in Hirtle et. Al. (2015).

My second method for calibration uses a machine learning algorithm to specify and calibrate stress test models. I use Tibshirani's (1996) least absolute shrinkage and selection operator (LASSO) to generate competing model calibrations and compare them to those produced by the traditional econometric model selection approach.

For each alternative stress testing approach, I forecast a representative bank's performance over 12 quarters beginning in September 2008. The forecasts are true out-of-sample forecasts where, following the initial forecast quarter, lagged dependent variables are set equal to lagged forecast values. The analysis simulates stress test estimates of bank performance that would have been produced by alternative model specifications if a supervisory stress test exercise were conducted in the summer of 2008 using a hypothetical stress scenario that exactly matched the macroeconomic data that materialized over the next three years.

The analysis yields many interesting and important results. The first surprising finding is that the traditional econometric approach for calibrating stress test models produces highly inaccurate forecasts over the three-year stress test horizon, even though these models fit the data exceptionally well within the estimation sample. For example, only one of the traditional model specifications has a smaller forecast error than a naïve forecast of zero quarterly profit over the 3-year stress scenario.

A second finding is that the common practice of including time-varying bank risk characteristics as explanatory variables causes a large reduction in the predictive accuracy of traditional stress test models. A final important finding is that the use of machine-learning calibration techniques, such as the LASSO approach, can substantially improve forecast accuracy over the traditional modelling approach.

The results highlight the need to develop methods to assess the accuracy of supervisory stress test models before they are used to evaluate the capital adequacy of individual banks. Under current conditions, there is no safeguard that prevents a regulator from imposing unnecessary supplementary

capital requirements and requiring mandatory stress testing modelling “improvements” when the identified “shortfalls” are entirely attributable to the inaccuracy of the regulator’s own stress test models.

The remainder of this paper is organized as follows. Section 2 provides additional background on the importance of stress test model precision. Section 3 discusses supervisory stress testing practices with a focus on the US use of stress tests. Section 4 discusses important considerations for stress test model specification. Section 5 reviews the paper’s experimental design and the data used in the analysis. Section 6 discusses the process used to specify and calibrate the models. Section 7 discusses individual stress test model estimation results. Section 8 reviews the within- sample prediction accuracy of the alternative models. Section 9 reports the results from out-of-sample forecasts where the alternative models are used to predict bank performance over the first three years of the financial crisis. Section 10 concludes the paper.

2. Background

Banking regulators, including both the US FRB and the European Banking Authority (EBA), have adopted dynamic stress testing as a component of their large financial institution supervision process. In the case of the FRB, the use of stress tests was mandated in the 2010 Dodd-Frank Act. In Europe, Article 23 of European Union Regulation No. 1095/2010 requires the EBA to develop “an adequate stress testing regime”.

The Basel Committee on Banking Supervision (BCBS) and the International Monetary Fund (IMF) have promoted supervisory emphasis on stress testing. For example, the BCBS has included requirements for supplemental stress tests in various parts of its Basel II and Basel III international capital standards and published guidance on “best practice” standards for regulatory stress testing processes.⁷

⁷ “Principles for sound stress testing practices and supervision,” Bank for International Settlements (May 2009).

Stress tests have been a mandatory component of the IMF's Financial Sector Assessment Program (FSAP) for nearly 20 years.⁸

Prudence suggests that the propriety of using dynamic stress tests hinges on the accuracy of supervisory stress test projections, and yet the national laws that require supervisory stress tests and the guidance issued by the BCBS or IMF⁹ never mention the accuracy of supervisory stress test models as an issue of concern. There are no guidelines or recommendations to ensure that supervisory stress test models meet a specified minimum level of precision before stress tests become a mandatory part of the supervisory assessment process.¹⁰ Still, it is hard to imagine that international standard setting bodies would recommend the use of stress tests to set bank capital requirements in instances where supervisory models are wildly inaccurate.

On the specific issue of Federal Reserve stress test model validation, the accuracy of the FRB's models are opaque to the public and the banks required to undergo stress test assessments. In its 2016 official stress testing methodology document,¹¹ the FRB describes its own model validation process as "A central oversight group consisting of senior-level Federal Reserve experts closely scrutinized the models and assumptions used in the supervisory stress test and model outputs."

The absence of explicit supervisory concern regarding the accuracy of stress testing models can in part be traced to the hypothetical nature of the stress tests. Econometric models are used to simulate a bank's performance over one or more hypothetical scenarios characterized by severely stressful macroeconomic conditions. Because these severe stress scenarios almost never materialize, it is impossible to compare a bank's actual stress scenario performance to the stress test model's forecast.

⁸ IMF rules require each member country to submit to an FSAP examination on a periodic basis and the IMF has actively promoted stress testing as an important tool for monitoring financial stability.

⁹ The IMF has many staff Working Paper on Stress Testing. A collection of these can be found in Ong (2014).

¹⁰ Ong and Pazarbasioglu (2013) mention the importance of supervisor model accuracy (p. 33) where they state, "Given the importance of projected pre-provision profits in determining banks' loss absorption capacity in stress scenarios, the credibility of these estimates are key in the overall perception of any stress test exercise." No further guidance is provided on suitable methods for establishing stress test credibility.

¹¹ Board of Governors of the Federal Reserve (2016).

If supervisory stress test models are used to set a bank's minimum capital adequacy requirement and evaluate a bank's internal stress test processes, it is important to understand whether supervisory actions are predicated on accurate stress test projections. Does the regulator really know that the bank will perform poorly in the stress scenario or that the bank's own stress test estimates are inaccurate? Are mandated improvements in bank capital and modeling processes justified, or are banks being required to hold unnecessary capital to buffer supervisory modeling errors? The accuracy of the supervisory stress test models is an important regulatory issue, especially in regards to the supervision of large complex banking institutions, and yet there is little research on this topic.

3. Overview of Supervisory Stress Test Models

Regulators keep confidential the details of the models they use to assess the stress tests of regulated banks. Supervisors defend this practice by arguing that full transparency would enable regulated banks to "game" stress test processes, and transparency could create a "herd mentality" in which every bank adopted the same stress test model.

While the FRB does not make its stress test models transparent, it does publish an extensive discussion of its stress test modeling approach. The FRB's own model uses an "industry average" approach. The models are estimated using pooled institution data, and are not bank specific,

"The estimated model parameters are the same for all BHCs and reflect industrywide, portfolio-specific, and instrument-specific response to variation in the macroeconomic and financial market variables. This industrywide approach reflects both the challenge in estimating separate, statistically robust models for each of the 33 BHCs and the desire of the Federal Reserve not to assume that historical BHC-specific results will prevail in the future."¹²

¹² Federal Reserve Board (2016) p. 3.

Both the FRB and the EBA¹³ use highly disaggregated approaches to conduct their stress tests.¹⁴ Consider, for example, the FRB model. According to public descriptions, for loss estimation purposes alone, the FRB segregates bank assets into more than 12 categories and models the performance of each category separately. For each institution, each loan category is assumed to grow at the industry average growth rate.¹⁵ Bank net revenues, gross of losses and impairments, are decomposed into 22 separate components, each of which has its own separate model specification. The FRB also emphasizes that its stress test models are undergoing continuous revision as the FRB receives and processes additional confidential data from bank holding companies, and as FRB staff discover changes that improve component model performance.

A Federal Reserve Bank of New York Working paper provides perhaps the most detailed description of a modeling approach that the Federal Reserve has developed for use in its stress test assessments. Hirtle et. al. (2015) describe the Federal Reserve's CLASS model, a model that reportedly produces very similar estimates to the estimates produced by the FRB's Dodd-Frank stress test models.

One class of models I analyze closely mimics the design and estimation approach used to calibrate the Fed's CLASS models. The primary difference between the models I analyze and the CLASS model is the level of aggregation. My analysis decomposes INBFTXEX into five components whereas the CLASS model uses 22 separate components to model INBFTXEX.¹⁶ The CLASS model disaggregates bank net non-provision income-to-assets ratios into 7 components whereas my disaggregation uses four

¹³ The European Banking Authority (2016). See The ClearingHouse (2016) for a comparison of the FRB and EBA stress test methodologies.

¹⁴ The FRB estimates stress test losses on a granular level for a specified economic scenario, providing loss rate and income projections for various specified types of loan and business lines, using pooled proprietary data on individual loans and other positions collected from their regulated banks.

¹⁵ Federal Reserve Board (2016) p.11. The FRB provides very little discussion of the individual asset category growth rates it uses in its models. The Federal Reserve CLASS models (Hirtle. et. al., p. 24) assume that assets grow at 1.25 percent per quarter throughout the stress period.

¹⁶ Brown, McGourty and Schuermann (2015) report that it is not uncommon for large bank holding companies to use between 50 and 150 models to project their FRB mandated stress test performance and formulate capital plans.

components. I use a single model for total bank provisions whereas the CLASS model decomposes provisions into 15 separately modeled categories.¹⁷

Each of the 22 separate CLASS models includes endogenous variables that measure time-varying bank risk characteristics. I assess the accuracy of alternative model specifications that both include and exclude time-varying bank characteristics.

Hirtle *et. al.*, report that the CLASS models has very good within-sample fits. Many of the individual equations having adjusted R^2 of nearly 90 percent. However, in many cases, the reported CLASS model adjusted R^2 statistics are overstated, as the models use individual bank data and require clustered standard errors to correct for correlated residuals. Correlation among CLASS models errors alters the interpretation attached to standard adjusted R^2 statistics because the correlation reduces the effective number of degrees of freedom in the estimation sample. While the authors correct their regression coefficient standard error estimates for this issue using a clustering estimator, they do not make any corrections in their reported adjusted R^2 statistics.¹⁸

Hirtle *et. al.*, (p. 32) also report that the Fed's CLASS estimates are similar to those produced by the FRB's stress test models. When FRB model estimates of pre-provision net bank revenue are regressed on the corresponding CLASS model estimates, the CLASS coefficient estimate is 0.845 and the regression R^2 is 0.869. Provision expense estimates are also closely related, but on average the CLASS

¹⁷ Separating provisions into 15 separate categories adds complexity, but it is unclear that the complexity will improve stress test forecasts. For example, within the estimation sample, the average total provision-to-asset ratio is 10.6 bps, with a standard deviation of 5.5 bps. Assuming the decomposition of provisions into 15 separate categories did, for arguments sake, improve the accuracy of the model's total provision estimate, the variation in the provision ratio is too small relative to the other components to make huge difference in the overall accuracy of the model.

¹⁸ Hirtle *et. al.*, report that many CLASS models were estimated using data on the largest 200 banking firms with errors clustered for each quarter. The paper reports very high adjusted R^2 values for these models compared to much more modest adjusted R^2 values for models estimated on average quarterly data (instead of individual bank data). The models estimated using individual bank data should include fixed effect terms, but do not. The resulting regression errors will be correlated over time as well as in each cross section, and thus the CLASS estimates should have used a different robust coefficient standard error estimator. Regardless, when standard errors are clustered, the "text-book" adjusted R-square statistic is biased upward since the data contain far fewer independent degrees of freedom than the number of observations. Corrections can be made to account for correlation, but the authors do not mention using a correction in the paper.

model slightly underestimate provisions relative to FRB model estimates. The authors conclude (p. 33), “[These strong statistical relationships provide] ...encouraging evidence that CLASS provides a reasonable proxy as to how more detailed stress tests might have performed prior to the financial crisis or if applied to a broader range of firms.”

Hirtle et. al., (p. 33) also provide a comparison between CLASS regulatory capital projections and the actual regulatory capital positions for the 200 largest banks over the period June 2007 through December 2008. These comparisons are made within the estimation sample and are not an out-of-sample assessment of forecast accuracy. And yet even within the estimation sample, the CLASS model estimates are disappointingly inaccurate. From June 2007 through December 2008, the actual 9-quarter average cumulative bank return on assets was 0.13 while the CLASS model projected a return of -0.05. A cross sectional regression of the actual performance of the 200 largest banks on their CLASS model projections for several stress test component model estimates produced R^2 statistics that range between 0.025 and 0.12.¹⁹

4. Critical Issues in Stress Test Model Specification

Because regulatory data on bank income and assets have a common time-trend, bank stress test models are often estimated in ratio form, with bank assets or balance sheet values used to deflate bank income and expense components.²⁰ Typically, bank income and expense ratios are modeled as stationary time series without pre-whitening or any formal testing to confirm this maintained hypotheses. The Fed’s CLASS model follows this convention and treats bank income and expense ratios as stationary time series even though many of the stress test dependent variables exhibit near unit-root autoregressive coefficient

¹⁹ Hirtle et. al., Table 8 page 59. Also note that these poor R^2 values are for a predicted and actual regression that is estimated after the authors winsorized the data at the 2 percent and 98 percent levels. If the “outliers” were retained, the model fit would have been worse yet. The authors put a positive spin on their results arguing that the CLASS estimates are useful because (p. 34) “...CLASS income projections are significantly positively correlated across firms’ ROA and its major components...”.

²⁰ To model some income components, some approaches may use an alternative bank balance sheet “stock” variable (e.g. total loans, or a total loan type—total 1-4 family residential mortgages) as a denominator. For example, the CLASS model uses multiple denominators among its 22 models.

estimates.²¹ I adopt the CLASS model convention and model bank income and expense ratios as stationary time series. I model bank asset balances in terms of asset growth rates which do appear to be stationary over the estimation sample period.²²

When following a traditional econometric modelling approach, stress test model specifications are chosen on the basis of model fit within the estimation sample. Stress test models are selected to be parsimonious, to produce a high adjusted R^2 and to minimize the regression residual standard error estimate. Variables that have only weak statistical significance are typically dropped from model specifications during stress test development. Often, researchers require that model coefficient signs and magnitudes be consistent with economic priors.²³

Focusing on stress test model fit within the estimation sample can create important issues regarding out-of-sample forecast accuracy. There is the risk of overfitting the data. That is, including variables that have spurious explanatory power within the estimation sample. It is also possible that the inclusion of variables that measure bank characteristics, while improving estimation sample fit, may negatively impact stress test forecast accuracy.

Including variables that account for time-varying bank characteristics has intuitive appeal. The practice imparts model flexibility to adapt, at least in part, to individual bank characteristics.²⁴ But bank characteristics are endogenous variables that change as banks respond to changing macroeconomic conditions. The inclusion of bank characteristics may improve model fit during the estimation sample, but it may also reduce the impact multipliers of the macroeconomic factors and lagged dependent variables that drive bank stress scenario forecasts. Moreover, variables that measure time-varying bank

²¹ While many of these ratios exhibit near unit-root behavior, it is doubtful that any of them follow true unit root processes. For example, many ratios display mean reversion. Moreover the ratios are theoretically all bounded from below and, practically speaking, all have finite upper bounds as well.

²² Fed stress tests reportedly do not model bank asset growth rates as a function of macroeconomic conditions even though bank asset growth clearly varies over the business cycle. Instead, FRB and CLASS models assume that bank assets continue to grow at average historical rates during the stress scenario.

²³ These criteria also guide the model specification of the CLASS models. See Hirtle et. al., p. 20.

²⁴ When the bank is evaluated using the supervisory model, the bank's stress scenario projections are based on the bank's own risk characteristics and not the characteristics of the "average" bank that are used to estimate the model.

characteristics should be forecast in a stress scenario, but current supervisory practice is to hold a bank's characteristics constant throughout the scenario.²⁵

Both the FRB and the EBA have chosen to use complex models that project income and loss using many individual econometric models for specific bank income and expense categories. Complexity creates an illusion of precision, but complex models need not produce the most accurate forecasts. Individual model errors may compound when many separate income and expense model projections are aggregated to forecast bank net income.²⁶

5. The Experimental Design

The financial crisis that began in 2008 is used to compare the accuracy of alternative stress test modeling approaches by comparing forecasts with actual performance when the true underlying macroeconomic conditions are used to seed the stress test scenario. My analysis uses quarterly data from March 1993 to June 2011 on all US insured depository institutions' income statements, balance sheets, and off-balance sheet items reported in the "Reports of Conditions and Income" regulatory filings.²⁷ Stress test models are estimated using data from March 1993 through June 2008 and the stress scenario forecasts cover the period September 2008 through June 2011.

The representative bank's performance and time-varying characteristics are measured by asset-weighted averages of individual bank values. Let x_{it}^j represent the value of bank characteristic j for bank i at time t . Let a_{it} represent the total assets of bank i at time t and A_t represent the total value of all

²⁵ For example, the FRB's stress tests reportedly keeps bank characteristics constant throughout the stress scenario.

²⁶ For example, Green and Armstrong (2015) find that, after comparing the forecasting results from alternative published studies, complex approaches typically have significantly lower forecast accuracy.

²⁷ The Call Report data is publicly available as the Federal Deposit Insurance Corporation's *Statistics on Depository Institutions*. The Call reports includes data on all insured depository institutions (commercial banks, state and federally chartered thrifts, savings banks, savings associations, and insured US branches for foreign chartered institutions). The analysis in this paper uses aggregate call report data. Such data will be referred to as "bank" data even though the aggregate data includes data from other insured depository institutions.

insured depository institutions' assets at time t , $A_t = \sum_{\forall i} a_{it}$. Then, for every t , and bank characteristic j , the average system-wide variable- j -to-asset ratio is defined as,

$$\sum_{\forall i} \left(\frac{x_{it}^j}{a_{it}} \right) \frac{a_{it}}{A_t} = A_t^{-1} \sum_{\forall i} x_{it}^j \quad (1)$$

The list of bank income and balance sheet variables used in the analysis along with sample summary statistics are reported in in Table 1.

The macroeconomic factors used in the analysis are derived from data provided by the Federal Reserve Bank of St. Louis in its FRED economic database. The single macroeconomic factor that is not sourced from FRED is the change in the real house price index as calculated by the American Enterprise Institute International Center on Housing Price Risk. The definitions of the macroeconomic factors are reported in Table 2 along with summary statistics for the sample period March 1993 through June 2011.

Stress test model accuracy depends in part on the explanatory variables used in the model. I consider two alternative sets of explanatory variables. One set includes a constant, a lagged dependent variable, and the macroeconomic factors listed in Table 2. A second set of explanatory variables adds bank balance sheet and off-balance sheet characteristics. The list of bank characteristics used in the analysis and their corresponding sample summary statistics are reported in Table 1.

The overall analysis assesses the accuracy of eight alternative stress test models that differ according to: (1) the level of disaggregation used to model the representative bank's income-to-asset ratio; (2) the explanatory variables included in the model; and (3), the method used to select variables and calibrate model coefficient estimates. The eight INBFTXEX model identifiers and distinguishing characteristics are reported in Table 3.

6. Model Estimation Process

Alternative model specifications are estimated using quarterly data over the sample period 1993Q1 through 2008Q2. Explanatory variables that measure bank risk characteristics are endogenous

variables. Because these variables are contemporaneously correlated with the regression error terms, I use the lagged values of these variables as instruments in the respective regression models.

6.1. The Traditional Approach for Variable Selection and Model Calibration

The preferred parameterization of each alternative stress test model specification is selected using a step-wise algorithm. The model is estimated including all potential explanatory variables including a constant and a lagged dependent variable. In subsequent estimations, variables that exhibit weak statistical significance are excluded from the model.²⁸ The final model is the specification that produces the smallest regression mean square error estimate.

In most cases, the explanatory variables in the preferred model specification exhibit a high degree of statistical significance. However, in two cases the preferred econometric specification includes some explanatory variables that have only weak statistical significance because omitting these variables markedly increases the regression standard error estimate.²⁹ I retain a constant term in each specification regardless of statistical significance.³⁰

This model selection process mimics the process used to select the final specifications in the Federal Reserve's CLASS model (*op. cit.*, page 19):

The final specification of each equation was based on a specification search based on measures of overall model fit (R^2 and adjusted R^2) as well as statistical significance of the macroeconomic variable, and accordance with economic theory (e.g., that chargeoff rates are positively correlated with poor economic conditions).

6.2. The LASSO Approach for Variable Selection and Model Calibration

An alternative approach for selecting explanatory variables and calibrating model coefficients is the least absolute shrinkage and selection operator (LASSO) machine learning algorithm of Tibshirani

²⁸ Weak statistical significance was indicated by t-statistics with small absolute values.

²⁹ The preferred models for non-interest income and asset growth both include some variables with relatively weak statistical significance. Excluding these variables significantly increases the respective regression model standard error estimates.

³⁰ In the bank provision model, the constant is not statistically significant, but is retained regardless.

(1996). LASSO is focused on generating forecasts with a lower mean-square-error than ordinary least squares forecasts.³¹ To achieve this, LASSO abandons the goal of producing unbiased estimates of the exogenous variable coefficients.

Similar to ridge regression, the LASSO algorithm exploits the idea that there can be a tradeoff between bias and reduced mean square forecast errors. LASSO minimizes the model's mean square error while simultaneously imposing a penalty on the sum of the absolute values of the model's coefficient estimates. The LASSO penalty function imposes a soft-thresholding condition on variable coefficient estimates which sets coefficient estimates to zero unless the estimate's absolute value exceeds a threshold set by the penalty rate. If the penalty rate is set at zero, LASSO produces ordinary least squares estimates. If the penalty rate is set too high, LASSO will set all coefficient estimates to zero.

The "art" of the LASSO modeling process is selection of a penalty rate that produces minimum mean-square-error forecasts. When estimating LASSO models for large cross-sectional data, the penalty rate is selected by retaining multiple hold-out samples and selecting the LASSO penalty rate that produces the smallest average mean-square error across the hold-out samples.

In the case of stress testing models, the data are auto-correlated time-series with relatively few observations, so the multiple holdout sample approach is infeasible. My preferred model is identified by finding the LASSO penalty rate that minimizes the mean-square forecast error for a single 3-year holdout sample. To be precise, I estimate the LASSO model using the coordinate descent algorithm [Wu and Lange (2008)] over the global grid of potential LASSO penalty rates³² using data from March 1993 through June 2005. For each LASSO penalty rate, I use LASSO coefficient estimates and calculate the mean-square forecast error over the holdout sample (September 2005-June 2008). I select the LASSO penalty rate that minimizes the holdout sample mean-square forecast error, and re-estimate the model

³¹ There are other approaches that attempt to improve stress test accuracy. See for example Bidder, Giacomini and McKenna (2016).

³² The global grid search of penalty rates is from 0, which produces OLS estimates, to the penalty rate at which LASSO sets all coefficient estimates to 0.

using data from March 1993 through June 2008. I use the second-stage LASSO coefficient estimates to generate out-of-sample forecasts of bank performance over the subsequent three-year stress period.

This calibration approach differs from other studies that have used LASSO in a stress testing context. Kapinos and Mitnik (2015) use LASSO to identify a parsimonious number of macroeconomic variables to include in stress test models driven by principle components. They set the LASSO penalty rate so only a very few macroeconomic variables have non-zero coefficient estimates and then extract the principle components from the macroeconomic variables selected by LASSO. The principle components are used as the explanatory variables in their stress test models. Chan-Lau (2017) uses LASSO to calibrate a stress test model of default probabilities. His analysis uses a large cross section data set that permits the use of multiple holdout samples to calibrate the LASSO penalty rate.

7. Model Estimation Results

The data must be standardized before employing the LASSO algorithm. The respective estimation sample means are subtracted from all variables and all explanatory variables are normalized by their estimation sample standard deviation estimates. After the optimal LASSO penalty rate is identified using the holdout sample, the data are renormalized to reflect the means and standard deviations of the March 1993 to June 2008 sample that is used to estimate the LASSO model coefficients that are employed in the stress test forecast.

The best-fitting alternative stress test model estimates are reported in Tables 4-10. For example, Table 4 reports the best fitting models for the ratio of bank net interest income to assets. In each of these tables, columns 2 and 3 report the best fitting OLS estimates and associated coefficient p-values when traditional model selection procedures are used, and the model includes a constant, a lagged dependent variable and macroeconomic factors. Columns 4 and 5 report the best fitting model and associated p-

values when the explanatory variables set is expanded to include time-varying bank characteristics.³³ Column 6 reports the best³⁴ LASSO model estimate when the explanatory variable set includes a constant, a lagged dependent variable and macro factors. Column 7 reports the best LASSO model estimates when the explanatory variable set also includes time-varying bank characteristics. Coefficient p-value estimates play no direct role in the LASSO variable selection process, so they are not reported.

A detailed discussion of the statistical properties and accuracy of each individual component model is not important for purposes of this paper. What matters is the accuracy of the out-of-sample forecasts of INBFTXEX produced by a combination of component models. Still, it is appropriate to make a few observations about the individual models.

The stress test models estimated using traditional econometric model selection techniques fit the estimation sample data exceptionally well. Among these models, models that include time-varying bank characteristics have superior in-sample fits relative to models that include only a constant, a lagged dependent variable and macro factors. With one exception (asset growth rates) traditional models that include time-varying bank characteristics have adjusted-R² well in excess of 80 percent, and a number of models exhibit adjusted-R² in the mid-90 percent range. Overall these models have sample fit statistics comparable to the fit statistics reported for the Federal Reserve CLASS models.

Compared to the traditional stress test model estimates, the LASSO calibration approach identifies different explanatory variables. For example, the traditional models of the ratio of securities gains and losses to assets identify many variables as statistically important while one LASSO model identifies the mean as the only important explanatory variable. Similarly, the traditional models for noninterest income identify many explanatory variables as statistically significant while one LASSO

³³ Tables that show the step-wise results that lead to the traditional model specifications reported in these tables appeared in an earlier Working Paper version of this paper and are available upon request from the author.

³⁴The LASSO penalty rate minimizes the holdout sample mean-square forecast error.

models includes only the mean and a lagged dependent variable. Clearly, the LASSO approach may produce a far different model specification than one generated by traditional model selection techniques.

8. Within-Sample Predictions of Income Before Tax and Extraordinary Items

While the accuracy and specification of individual stress test component models can be of interest for any number of reasons, this paper focuses on model accuracy in predicting bank income before taxes and extraordinary items, out-of-sample, and under stressful conditions. Before examining out-of-sample stress predictions, I briefly review the predictive accuracy of candidate models within the estimation sample.

Within the estimation sample, the INBFTXEX prediction is the product of two estimates, an estimate of the ratio of INBFTXEX to assets, and an estimate of the bank's asset growth rate, where the product is multiplied by bank assets as of the end of the prior quarter.³⁵ Unlike out-of-sample forecasts, within sample predicted values are one-period ahead predictions—the bank's lagged asset values and time varying characteristics are actual values, not predicted values.

Because the predicted value of INBFTXEX is the product of two forecasted values, it is unlikely to be unbiased. Recall that LASSO estimates intentionally allow bias in order to reduce mean-square forecast error. But even in the traditional stress test approach that uses unbiased model estimators, the INBFTXEX estimate will be biased unless the two forecasted series are independent.³⁶ Regression test results reported in Table 11 indicate that asset growth and the ratio of INBFTXEX to assets are not independent, so even the traditional stress test models can be expected to produce biased estimates of INBFTXEX.³⁷

³⁵ The estimate of the ratio of INBFTXEX to assets could be produced by a single equation model or a five equation model.

³⁶ Recall that $Covariance(X Y) = E(X Y) - E(X) E(Y)$. Thus, unless the covariance between X and Y is zero, $E(X Y) \neq E(X) E(Y)$.

³⁷ The statistically significant coefficient estimates in Table 12 on asset growth rate and asset growth rate squared imply that the two forecasted series are correlated and thus not independent.

Figure 1 plots the representative bank's actual income before tax and extraordinary items over the sample estimation period when the individual bank is scaled by total assets in the banking system.³⁸ The last three observations in the sample period, December 2007- June 2008, are problematic in that they generate exceptionally large prediction errors for all of the models considered. Consequently, in Table 12, I report prediction error characteristics for both the full estimation sample and for a sample that excludes the last three observations. Model prediction errors are measured relative to the actual series value. Prediction errors are defined as the ratio of actual INBFTXEX less predicted INBFTXEX as a percentage of actual INBFTXEX.

The prediction error statistics reported in Table 12 show that all of the stress test models are biased. On average, each model underestimates INBFTXEX over sample period. Omitting outliers, all but one model (5-eq OLS) still underestimates INBFTXEX, but without the outliers, the average underestimate is less than 1 percent, and most cases significantly under 1 percent. Still, the large in-sample prediction errors associated with the last three observations in the estimation sample foreshadow the potential for diminished accuracy in out-of-sample forecasts under stressful conditions.

Overall, the prediction error statistics reported in Table 12 suggest that the traditional models fit the estimation sample data better than the LASSO models. The 5-equation traditional model with time-varying bank characteristics has the smallest average prediction error over the entire sample, and the second smallest prediction error (in absolute value terms) after removing outliers. With or without the outliers, this model has the smallest root mean-square prediction error (RMSE). Compared to the traditional models, the LASSO models exhibit larger in-sample bias regardless of whether outliers are excluded, and the extra bias LASSO imparts does not appear to uniformly improve the LASSO model

³⁸ The representative bank's INBFTXEX-to-asset ratio must be multiplied by the level of assets in the representative bank to arrive at INBFTXEX for the representative bank. I multiply the INBFTXEX-to-asset ratio by the total assets in the banking system for date t . This normalization will scale up the representative bank models so they forecast the aggregate performance of the banking system.

RMSEs relative to comparable models that are specified using traditional econometric estimation techniques.

9. Stress Test Predictive Accuracy During the Financial Crisis

The accuracy of eight alternative stress test models are compared using out-of- sample forecasts over the first 12-quarters of the financial crisis. The out-of-sample forecasts use the actual stress scenario values for the macroeconomic factors, and like the regulatory approach, they hold the representative bank's time-varying characteristics fixed at June 2008 values. The representative bank's asset value is forecasted using an asset growth rate model with the initial value for the representative bank asset balance set equal to the aggregate value of assets in the banking system as of June 2008. With this initialization, the representative bank is scaled up to reproduce the aggregate performance of the banking system. Accuracy is assessed by comparing stress scenario forecasts to the banking system's actual aggregate performance.

If f denotes a forecasted value, A_0 designates the aggregate assets in the banking system as of June 2008, and the subscript denotes stress test time (in quarters), the first stress quarter forecast is calculated as,

$$\begin{aligned} & [nim_1 + nonii_1 + iglsec_1 - elnatr_1 - nonix_1^f] \\ &= \left[\frac{nim_1 + nonii_1 + iglsec_1 + elnatr_1 + nonix_1}{asset_1} \right]^f \times A_0 \times \left[\frac{asset_1}{asset_0} \right]^f \end{aligned} \quad (2)$$

Forecasts for the stress scenario's q^{th} -quarter are based on a true out-of-sample forecast of the bank's assets as of quarter q ,

$$\begin{aligned} & [nim_q + nonii_q + iglsec_q - elnatr_q - nonix_q]^f = \left[\frac{nim_q + nonii_q + iglsec_q + elnatr_q + nonix_q}{asset_q} \right]^f \\ & \times A_0 \times \left[\frac{asset_1}{asset_0} \right]^f \times \left[\frac{asset_2}{asset_1} \right]^f \times \dots \times \left[\frac{asset_q}{asset_{q-1}} \right]^f \end{aligned} \quad (3)$$

Table 13 reports statistics on stress scenario forecast errors for the 8 stress test models. Figure 2 plots the alternative cumulative INBFTXEX estimates and the actual cumulative INBFTXEX value for each quarter of the stress test scenario.

The statistics reported in Table 13 show that the model 1-eq LASSO, calibrated using LASSO, a single equation to forecast the ratio of INBFTXEX to assets, and excluding time-varying bank characteristics is, by far and away, the most accurate stress testing approach among those considered. The remaining LASSO models rank second (5-eq E-LASSO), third (1-eq E-LASSO) and fourth (5-eq LASSO) in terms of forecasting accuracy.

The traditional stress test models are all highly inaccurate. Indeed, only a single model (1-eq OLS) has accuracy better than a completely naïve forecast of zero profit for each period in the stress test. The model that most closely mimics the Federal Reserve's CLASS model, the 5-eq E-OLS model, is the least accurate model by a wide margin. It underestimates the banking system's cumulative stress scenario income by nearly 200 percent.

If dynamic stress tests are used as a supervisory tool, the results in Table 13 and Figure 2 caution against the use of traditional model selection procedures, complicated disaggregated models, and the use of time-varying bank risk characteristics as explanatory variables. The superiority of LASSO models suggests that the accuracy of the traditional approach for specifying stress test models likely suffers from overfitting the estimation sample data, especially when time-varying bank characteristics are included as endogenous variables.

After reviewing these results, a skeptical reader might try to dismiss the findings by arguing that Federal government crisis interventions like Federal Reserve special lending programs and the Treasury's Troubled Asset Relief Program are special factors that might easily explain the traditional model's exceptionally large stress test forecast error. But such a conjecture cannot explain why the LASSO model forecasts are not similarly impacted by these special factors. The LASSO forecasts are impacted by the

same special factors and yet they are far more accurate than the traditional stress test model specifications.

While the traditional models estimated in this paper have explanatory power and goodness-of-fit measures equal to those of the Federal Reserve’s CLASS models, it is difficult to imagine recommending that any of these stress testing models be used as a supervisory tool to set individual bank minimum regulatory capital requirements. These models are far too inaccurate for this important task. Still, based on publicly available information, it seems highly likely that regulators have used supervisory stress test models with similar levels of precision to determine individual institution minimum regulatory capital requirements.

10. Conclusion

Using the 2008 financial crisis as a “natural experiment,” I analyze the accuracy of alternative methods for forecasting bank performance over multi-year stress scenarios, including methods that produce estimates that are similar to estimates produced by the FRB’s stress test models. Comparing alternative stress test model forecasts to the actual performance of the banking system over the first three years of the financial crisis, I demonstrate some stress test modeling approaches that have been used in the past have exceptionally poor forecasting accuracy.

The analysis shows that the traditional approach for specifying stress test models can be significantly improved by using machine-learning calibration techniques that penalize non-parsimonious models. In particular, my analysis shows that the use of LASSO, calibrated using a single hold-out sample, can vastly improve stress test forecast accuracy compared to traditional stress test modeling approaches. The results also show that simple stress test models, when appropriately calibrated, can be as accurate, if not more accurate, than complex disaggregated models calibrated using traditional methods.

The results in this paper highlight an urgent need for national regulators, the Basel Committee on Banking Supervision, and other multinational standard setting bodies to focus on the accuracy of

supervisory stress testing models. The benefits of stress testing will be a controversial issue, and appropriately so, until stress test models achieve a minimum level of accuracy.

Under current practices, it is easy to imagine a situation where regulatory stress testing produces a significant misallocation of resources. Suppose an institution models its performance using the 1-eq LASSO model. We know that the Federal Reserve has used a complex and potentially very inaccurate stress testing model that produces forecasts that are not dissimilar to the 5-eq E-OLS model. Suppose stress test forecasts are tabulated, and the bank's own stress test indicates that the bank survives the scenario with positive income and adequate capital, while the FRB's model forecasts large bank losses and a capital deficiency. The FRB could well fail the bank and prevent it from paying dividends and making capital repurchases. Moreover, the FRB could require the bank to abandon its LASSO model in favor of spending considerable resources to build a much more complex stress test model—one that more closely mimics the FRB's model. At present, there is no law, regulation, guidance or judicial process in place to prevent such an outcome.

References

- Basel Committee on Banking Supervision, 2009. “Principles for sound stress testing practices and supervision,” Bank for International Settlements (May).
- Bidder, R., R. Giacomini, and A. McKenna, 2016. “Stress Testing with Misspecified Models,” Federal Reserve Bank of San Francisco Working Paper No 2016-26.
- Board of Governors of the Federal Reserve, 2012. “Comprehensive Capital Analysis and Review 2012: Methodology and Results for Stress Scenario Projections,” Washington D.C., available at <https://www.federalreserve.gov/newsevents/press/bcreg/bcreg20120313a1.pdf>
- Board of Governors of the Federal Reserve, 2016. “Dodd-Frank Act Stress Test 2016: Supervisory Stress Test Methodology and Results”, Washington D.C., available at www.federalreserve.gov/bankinforeg/stress-tests/2016-Supervisory-Stress-Test-Framework-and-Model-Methodology.htm
- Brown J, B. McGourty and T. Schuermann, 2015. “Model Risk and the Great Financial Crisis: The Rise of Modern Model Risk Management,” Wharton School Weiss Center Working Paper 14-02.
- Chan-Lau, J., 2017. “Lasso Regressions and Forecasting Models in Applied Stress Testing,” IMF Working Paper No. 17/108.
- The ClearingHouse, 2016. “Comparison between United States and European Union Stress Tests,” (May). <https://www.theclearinghouse.org/~media/files/research%20notes/20160518-tch-research-note-ccar-vs-eba-stresstests.pdf>
- The European Banking Authority, 2016. “2016-EU-Wide Stress Test: Methodological Note.” European Banking Authority, London, U.K., available at <https://www.eba.europa.eu/documents/10180/1259315/2016+EU-wide+stress+test-Methodological+note.pdf>
- Federal Deposit Insurance Corporation, *Statistics on Depository Institutions*. Available at <https://www5.fdic.gov/sdi/main.asp?formname=compare>
- Federal Reserve Bank of St. Louis. FRED economic data. <https://fred.stlouisfed.org/>
- Green, K and J. S. Armstrong, 2015. “Simple versus complex forecasting: The evidence,” *Journal of Business Research*, Vol. 68, No. 8, pp. 1678-1685.
- Hirtle, B. and A. Lehnert, 2014. “Supervisory Stress Tests,” FRB of New York Staff Report No. 696.
- Hirtle, B. A. Kovner, J. Vickery, and M. Bhanot, 2015. “Assessing Financial Stability: The Capital and Loss Assessment under Stress Scenarios (CLASS) Model,” Federal Reserve Bank of New York Staff Report No. 663 (July).

Kapinos, P. and O. Mitnik, 2015. "A Top-down Approach to Stress-testing Banks," *Journal of Financial Services Research*, Vol. 49, No. 2, pp. 229-264.

Ong, L. L. and C. Pazarbasioglu, 2013. "Credibility and Crisis Stress Testing," *International Monetary Fund Working Paper No. WP/13/178*, International Monetary Fund, Washington D.C.

Ong, Li Lian editor, 2014. *A Guide to IMF Stress Testing: Methods and Models*. Washington D.C., International Monetary Fund, print.

Tibshirani, R., 1996. "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society B*, Vol. 58, pp. 267-288.

Wu, T. and K. Lange, 2008. "Coordinate Descent Algorithms for LASSO Penalized Regression," *The Annals of Applied Statistics*, Vol. 2, No. 1, pp.224-244.

Table 1: Bank variable definitions and data sources

Bank Variables	Variable definitions using variable names from FDIC SDI data	Minimum	Maximum	Average	Median	Standard deviation
Total asset growth	$100*(\text{asset}_t/\text{asset}_{t-1} - 1)$	-2.259	3.681	1.516	1.697	1.247
Income before tax and extraordinary items to total assets	$100*(\text{nim}_t + \text{nonii}_t + \text{iglsec}_t - \text{elnatr}_t - \text{nonix}_t)/\text{asset}_t$	-0.297	0.520	0.372	0.431	0.161
Net Interest Income to total assets	$100*\text{nim}_t/\text{asset}_t$	0.566	0.925	0.804	0.811	0.072
Noninterest income to total assets	$100*\text{nonii}_t/\text{asset}_t$	0.231	0.600	0.496	0.504	0.067
Noninterest expense to total assets	$100*\text{nonix}_t/\text{asset}_t$	0.631	0.937	0.794	0.806	0.063
Provisions for loan and lease losses to total assets	$100*\text{elnatr}_t/\text{asset}_t$	0.053	0.502	0.143	0.097	0.110
Securities gains and losses to total assets	$100*\text{iglsec}_t/\text{asset}_t$	-0.053	0.054	0.010	0.010	0.019
Total securities to total assets	$100*\text{sc}_t/\text{asset}_t$	14.625	23.616	18.975	18.964	2.129
Trading account assets to total assets	$100*\text{trade}_t/\text{asset}_t$	2.095	7.554	4.732	4.705	0.999
Other real estate owned to total assets	$100*\text{ore}_t/\text{asset}_t$	0.047	0.944	0.179	0.097	0.181
Income earned, not collected on loans to total assets	$100*\text{oaienc}_t/\text{asset}_t$	0.361	0.662	0.457	0.445	0.064
Total unused commitments to total assets	$100*\text{uc}_t/\text{asset}_t$	30.611	69.053	55.253	60.417	11.488
Notional value of derivative positions to total assets	$100*\text{obsdir}_t/\text{asset}_t$	213.936	1829.845	792.183	632.296	460.756
Total loans and leases to total assets	$100*\text{lnlsgr}_t/\text{asset}_t$	54.045	63.289	60.003	60.593	1.943
Construction and development loans to total assets	$100*\text{lnrecons}_t/\text{asset}_t$	1.660	4.891	2.873	2.877	1.037
Commercial real estate loans to total assets	$100*\text{lnrenres}_t/\text{asset}_t$	6.344	8.331	7.190	7.305	0.534
1-4 family residential loans to total assets	$100*\text{lnrers}_t/\text{asset}_t$	18.299	23.857	20.821	20.567	1.318
Commercial and industrial loans to total assets	$100*\text{lnci}_t/\text{asset}_t$	8.745	14.856	11.748	11.636	1.799
Consumer loans to total assets	$100*\text{lncon}_t/\text{asset}_t$	7.705	10.874	9.159	9.054	0.894

Notes: The data are calculated from Statistics on Depository Institutions (SDI) quarterly data as reported on the Federal Deposit Insurance Corporation's public website for the period March 1993 through June 2011. The exact variable definitions are reported in the SDI's "Read me" file. Variables with a "t" subscript represent the aggregate value for all reporting institutions in quarter "t" for the SDI variable of the same name.

Table 2: Macroeconomic factor variables and definitions

Macroeconomic Variable	Source Variable Series	Minimum	Maximum	Average	Median	Standard deviation
Nominal Quarterly GDP growth rate, SAAR	A191RP1Q027SBEA	-7.70	10.20	4.67	4.90	2.89
civilian unemployment rate, unadjusted +		3.80	10.00	5.82	5.50	1.64
10-year Treasury yield*	GS10	2.71	7.84	5.00	4.88	1.22
3-month Treasury yield*	GS3M	0.04	6.23	3.29	3.60	2.01
Moody's AAA yield*	AAA	4.57	8.55	6.39	6.39	1.01
Moody's Baa yield*	BAA	5.04	9.19	7.19	7.27	1.01
Federal funds rate*	FedFunds	0.09	6.52	3.45	4.06	2.09
Wilshire quarterly market index return	WILL5000INDFC	-22.92	21.51	2.44	3.29	8.69
Wilshire daily return standard deviation	WILL5000INDFC	-0.87	4.22	1.02	0.86	0.63
Kansas City Fed Financial Stability Index (FSI) quarterly average	KCFSI	-0.94	5.56	0.11	-0.29	1.13
VIX daily average	VIXCLS	11.03	58.60	20.57	19.92	8.02
VIX quarterly percent change	VIXCLS	-82.41	127.50	3.58	-0.49	30.70
Change in Real House Price Index	RHPI	-6.25	4.68	0.32	0.85	2.08

Notes: All data except the change in the real house price index are from the Federal Reserve Bank of St. Louis economic research department public database <https://fred.stlouisfed.org/>. The change in the real house price index is the American Enterprise Institute International Center on Housing Risk calculated quarterly as FHFA's all-transaction house price index divided by BEA's price index for personal consumption expenditures. + Indicates unemployment rate on the first day following quarter-end. * Indicates interest yield on the first day of each quarter.

Table 3: Stress test model specifications, explanatory variables and calibration methods

Model	Model Identifier	Level of INBFTXEX-to-asset disaggregation	Asset growth rate modeled separately	Includes lagged bank characteristics	Variable selection & model calibration
1	1-eq OLS	1 equation	yes	no	traditional OLS
2	1-eq E-OLS	1 equation	yes	yes	traditional OLS
3	1-eq LASSO	1 equation	yes	no	LASSO
4	1-eq E-LASSO	1 equation	yes	yes	LASSO
5	5-eq OLS	5 equations	yes	no	traditional OLS
6	5-eq E-OLS	5 equations	yes	yes	traditional OLS
7	5-eq LASSO	5 equations	yes	no	LASSO
8	5-eq E-LASSO	5 equations	yes	yes	LASSO

Table 4: Net interest income model estimates

Explanatory Variable	model calibration method					
	traditional		traditional		LASSO	LASSO
	OLS	t-statistic	OLS	t-statistic		
Intercept	0.0903	2.44	0.4425	4.74		
Lagged net interest income	0.7449	10.19			0.0346	0.0015
GDP growth rate	0.0030	2.04			0.0072	0.0031
civilian unemployment rate					0.0105	0.0066
10-year Treasury yield			0.0448	4.27	0.0064	0.0163
3-month Treasury yield					0.0000	0.0000
Moody's AAA yield	0.0151	2.87	-0.0366	-3.05	0.0153	-0.0009
Moody's Baa yield			0.0142	4.21	0.0074	0.0133
Federal funds rate			-0.0207	-7.23	-0.0019	-0.0160
Wilshire market index return					0.0021	0.0019
Wilshire daily return standard deviation					0.0012	-0.0021
Kansas City Fed FSI average					-0.0062	-0.0068
VIX daily average					0.0101	-0.0030
VIX percent change					0.0000	0.0012
Change in Real House Price Index					0.0022	0.0000
Lagged securities to total assets						-0.0175
Lagged trading assets to total assets			0.0190	3.59		0.0164
Lagged ORE to total assets						0.0112
assets						-0.0069
Lagged total unused commitments to total assets						0.0217
Lagged total notional value of derivatives to total assets			-0.0003	-6.98		-0.0988
Lagged total loans and leases to total assets						-0.0134
Lagged construction and development loans to total assets			0.0537	4.59		0.0565
Lagged commercial realstate loans to total assets						0.0176
Lagged 1-4 family residential realstate loans to total assets						0.0023
Lagged commercial and industrial loans to total assets			0.0146	4.13		0.0365
Lagged consumer loans to total assets			0.0166	2.51		0.0318
Mean square error	0.0005		0.0003			
Adjusted R-square	0.8857		0.9438			
optimal LASSO penalty rate					0.0047	0.0048
holdout sample MSE at optimal LASSO penalty					0.0132	0.0025

Table 5: Noninterest income model estimates

Explanatory Variable	model calibration method					
	traditional		traditional		LASSO	LASSO
	OLS	t-statistic	OLS	t-statistic		
estimate	t-statistic	estimate	t-statistic	estimate	estimate	
Intercept	0.5412	4.90	0.3647	2.34		
Lagged net interest income	0.3465	3.26			0.0265	0.0000
GDP growth rate	0.0023	1.17			0.0000	0.0028
civilian unemployment rate	-0.0439	-3.81	-0.0218	-1.46	0.0000	0.0000
10-year Treasury yield	-0.0757	-3.87	-0.0147	-1.67	0.0000	0.0000
3-month Treasury yield	0.0787	3.10	0.0501	2.01	0.0000	0.0000
Moody's AAA yield	0.0676	4.15			0.0000	0.0000
Moody's Baa yield					0.0000	-0.0073
Federal funds rate	-0.0810	-3.5078	-0.0584	-2.59	0.0000	-0.0133
Wilshire market index return	0.0007	1.2911	0.0008	1.75	0.0000	0.0033
Wilshire daily return standard deviation	-0.0174	-1.3056	-0.0212	-2.10	0.0000	-0.0069
Kansas City Fed FSI average					0.0000	0.0000
VIX daily average					0.0000	0.0000
VIX percent change					0.0000	0.0000
Change in Real House Price Index					0.0000	0.0039
Lagged securities to total assets	0.0009					0.0080
Lagged trading assets to total assets	0.7514					0.0132
Lagged ORE to total assets						0.0147
Lagged income earned and not collected on loans to total assets			-0.2673	-2.90		-0.0181
Lagged total unused commitments to total assets			0.0047	4.66		0.0745
Lagged total notional value of derivatives to total assets			-0.0003	-4.53		-0.0830
Lagged total loans and leases to total assets						-0.0186
Lagged construction and development loans to total assets			0.0732	4.37		0.0853
Lagged commercial realestate loans to total assets						0.0106
Lagged 1-4 family residential realestate loans to total assets						0.0407
Lagged commercial and industrial loans to total assets			0.0198	3.97		0.0953
Lagged consumer loans to total assets						0.0204
Mean square error	0.0009		0.00056			
Adjusted R-square	0.7514		0.845			
optimal LASSO penalty rate					0.0210	0.0047
holdout sample MSE at optimal LASSO penalty					0.0037	0.0047

Table 6: Securities gains and losses model estimates

Explanatory Variable	model calibration method					
	traditional		traditional		LASSO	LASSO
	OLS		OLS			
estimate	t-statistic	estimate	t-statistic	estimate	estimate	
Intercept	-0.0871	-5.61	0.1555	4.05		
Lagged net interest income					0.0000	0.0000
GDP growth rate	0.0013	2.44			0.0000	0.0000
civilian unemployment rate	0.0089	3.32			0.0000	0.0000
10-year Treasury yield	-0.0394	-7.22	-0.0181	-12.98	0.0000	-0.0063
3-month Treasury yield	0.0058	3.66			0.0000	0.0000
Moody's AAA yield	0.0280	5.45			0.0000	0.0000
Moody's Baa yield					0.0000	0.0000
Federal funds rate					0.0000	0.0000
Wilshire market index return					0.0000	0.0000
Wilshire daily return standard deviation					0.0000	0.0000
Kansas City Fed FSI average	-0.0287	-8.0689	-0.0194	-6.89	0.0000	0.0000
VIX daily average	0.0024	7.7454	0.0017	5.67	0.0000	0.0000
VIX percent change	-0.0001	-2.4705	0.0000	-1.58	0.0000	0.0000
Change in Real House Price Index					0.0000	0.0000
Lagged securities to total assets						0.0009
Lagged trading assets to total assets			0.0036	1.92		0.0024
Lagged ORE to total assets						-0.0022
Lagged income earned and not collected on loans to total assets						0.0010
Lagged total unused commitments to total assets						-0.0074
Lagged total notional value of derivatives to total assets			-0.0001	-6.48		-0.0271
Lagged total loans and leases to total assets						-0.0014
Lagged construction and development loans to total assets			0.0229	5.88		0.0247
Lagged commercial real estate loans to total assets						0.0011
Lagged 1-4 family residential real estate loans to total assets			-0.0042	-2.43		-0.0042
Lagged commercial and industrial loans to total assets						0.0055
Lagged consumer loans to total assets						0.0010
Mean square error	0.0001		0.00004			
Adjusted R-square	0.7910		0.861			
optimal LASSO penalty rate					0.0115	0.0109
holdout sample MSE at optimal LASSO penalty					0.0003	0.0001

Table 7: Bank provision model estimates

Explanatory Variable	model calibration method					
	traditional		traditional		LASSO	LASSO
	OLS		OLS			
estimate	t-statistic	estimate	t-statistic	estimate	estimate	
Intercept	0.1154	3.99	-0.0589	-0.24		
Lagged net interest income	0.8237	7.96	0.4663	4.65	0.0278	0.0194
GDP growth rate	-0.0050	-3.79			-0.0098	0.0000
civilian unemployment rate			0.0455	7.32	0.0025	0.0051
10-year Treasury yield	0.0400	2.85	0.0137	3.66	0.0000	0.0000
3-month Treasury yield	-0.0487	-3.09			-0.0043	-0.0052
Moody's AAA yield	-0.0320	-2.55			-0.0025	0.0000
Moody's Baa yield					0.0008	0.0000
Federal funds rate	0.0382	2.7489			0.0000	0.0000
Wilshire market index return	0.0007	1.5647			0.0022	0.0000
Wilshire daily return standard deviation			0.0113	1.77	0.0000	0.0000
Kansas City Fed FSI average	0.0391	3.7601	0.0320	3.94	0.0136	0.0000
VIX daily average	-0.0017	-2.2476			-0.0006	0.0000
VIX percent change	0.0003	2.6860			0.0048	0.0000
Change in Real House Price Index					-0.0124	0.0000
Lagged securities to total assets			-0.0165	-5.17		-0.0252
Lagged trading assets to total assets						0.0097
Lagged ORE to total assets						0.0060
Lagged income earned and not collected on loans to total assets						0.0026
Lagged total unused commitments to total assets			-0.0020	-5.59		-0.0356
Lagged total notional value of derivatives to total assets			0.0002	4.79		0.0538
Lagged total loans and leases to total assets			0.0077	2.19		0.0207
Lagged construction and development loans to total assets			-0.0511	-4.29		-0.0504
Lagged commercial realestate loans to total assets						0.0032
Lagged 1-4 family residential realestate loans to total assets						-0.0210
Lagged commercial and industrial loans to total assets						-0.0044
Lagged consumer loans to total assets			-0.02228	-3.05		-0.0231
Mean square error	0.0003		0.00016			
Adjusted R-square	0.8874		0.948			
optimal LASSO penalty rate					0.0004	0.0112
holdout sample MSE at optimal LASSO penalty					0.0085	0.0177

Table 8: Noninterest expense model estimates

Explanatory Variable	model calibration method					
	traditional		traditional		LASSO	LASSO
	OLS		OLS			
estimate	t-statistic	estimate	t-statistic	estimate	estimate	
Intercept	0.2303	3.95	-0.4929	-2.29		
Lagged net interest income			-0.3288	-2.72	0.0077	0.0000
GDP growth rate					0.0019	0.0000
civilian unemployment rate	0.0226	2.42			0.0103	0.0000
10-year Treasury yield					0.0021	0.0000
3-month Treasury yield	0.0834	3.91	0.0427	1.96	0.0153	0.0000
Moody's AAA yield					0.0147	0.0000
Moody's Baa yield	0.0454	6.7640	0.0308	2.89	0.0263	0.0000
Federal funds rate	-0.0734	-3.3839	-0.0411	-1.90	-0.0152	0.0000
Wilshire market index return	0.0015	3.0414	0.0012	2.88	0.0129	0.0000
Wilshire daily return standard deviation					0.0027	0.0000
Kansas City Fed FSI average					0.0000	0.0000
VIX daily average	0.0043	4.9600			0.0116	0.0000
VIX percent change					0.0000	0.0000
Change in Real House Price Index	0.0107	3.9790			0.0174	0.0000
Lagged securities to total assets			0.0129	3.07		0.0437
Lagged trading assets to total assets						0.0134
Lagged ORE to total assets			0.2881	4.77		0.0760
Lagged income earned and not collected on loans to total assets						0.0059
Lagged total unused commitments to total assets			0.0065	4.23		0.1062
Lagged total notional value of derivatives to total assets						-0.0078
Lagged total loans and leases to total assets						0.0030
Lagged construction and development loans to total assets						-0.0083
Lagged commercial real estate loans to total assets						-0.0116
Lagged 1-4 family residential real estate loans to total assets						0.0172
Lagged commercial and industrial loans to total assets			0.0183	4.07		0.0272
Lagged consumer loans to total assets			0.048759	4.55		0.0465
Mean square error	0.0009		0.00057			
Adjusted R-square	0.7394		0.833			
optimal LASSO penalty rate					0.0002	0.0335
holdout sample MSE at optimal LASSO penalty					0.0097	0.0061

Table 9: Income before tax and extraordinary items model estimates

Explanatory Variable	model calibration method					
	traditional		traditional		LASSO	LASSO
	OLS		OLS			
estimate	t-statistic	estimate	t-statistic	estimate	estimate	
Intercept	0.5073	4.67	2.1770	3.44		
Lagged net interest income			-0.1675	-1.63	0.0121	0.0000
GDP growth rate	0.0079	2.13			0.0281	0.0110
civilian unemployment rate	-0.0809	-4.18	-0.0405	-2.83	-0.0142	0.0000
10-year Treasury yield	-0.1862	-4.79	-0.0189	-1.40	-0.0548	0.0000
3-month Treasury yield	0.1663	3.72			0.0156	0.0000
Moody's AAA yield	0.1744	4.65			0.0586	-0.0052
Moody's Baa yield					-0.0223	-0.0157
Federal funds rate	-0.1598	-4.1793			-0.0061	-0.0013
Wilshire market index return					-0.0058	0.0000
Wilshire daily return standard deviation					-0.0001	0.0000
Kansas City Fed FSI average	-0.1684	-6.3094	-0.0916	-6.67	-0.0385	-0.0368
VIX daily average	0.0061	2.7536			0.0168	0.0000
VIX percent change					-0.0074	0.0000
Change in Real House Price Index					0.0264	0.0076
Lagged securities to total assets						0.0157
Lagged trading assets to total assets						0.0035
Lagged ORE to total assets			-0.1817	-1.77		-0.0458
Lagged income earned and not collected on loans to total assets			-0.3551	-2.58		-0.0288
Lagged total unused commitments to total assets			0.0088	3.82		0.0573
Lagged total notional value of derivatives to total assets			-0.0009	-7.72		-0.2224
Lagged total loans and leases to total assets			-0.0254	-2.93		-0.0550
Lagged construction and development loans to total assets			0.1574	4.69		0.2029
Lagged commercial real estate loans to total assets						0.0212
Lagged 1-4 family residential real estate loans to total assets						0.0580
Lagged commercial and industrial loans to total assets						0.1283
Lagged consumer loans to total assets						0.0281
Mean square error	0.0028		0.00148			
Adjusted R-square	0.6768		0.829			
optimal LASSO penalty rate					0.00056	0.0078
holdout sample MSE at optimal LASSO penalty					0.008135	0.0043

Table 10: Asset growth model estimates

Explanatory Variable	model calibration method					
	traditional		traditional		LASSO	LASSO
	OLS estimate	t-statistic	OLS estimate	t-statistic	estimate	estimate
Intercept	3.4001	3.64	67.0501	3.42		
Lagged net interest income			-0.2503	-1.95	0.0000	0.0000
GDP growth rate					0.0000	0.0000
civilian unemployment rate					0.0000	-0.6276
10-year Treasury yield	0.6372	1.77	-1.1511	-1.47	0.0000	0.0000
3-month Treasury yield					0.0000	0.0000
Moody's AAA yield	-0.9125	-2.32	1.0098	1.00	0.0000	0.0000
Moody's Baa yield					-0.0500	-0.0529
Federal funds rate			0.3577	1.78	0.0000	0.0000
Wilshire market index return					0.0000	0.0298
Wilshire daily return standard deviation	0.9955	3.0124			0.0550	0.1549
Kansas City Fed FSI average			0.8115	1.23	0.0000	0.0000
VIX daily average			0.0677	1.30	0.0000	0.5462
VIX percent change					0.0000	0.0000
Change in Real House Price Index	0.1314	1.9605	0.3532	2.77	0.0000	0.0000
Lagged securities to total assets			-0.5410	-2.13		1.5549
Lagged trading assets to total assets						-0.0844
Lagged ORE to total assets			-6.6810	-2.47		0.5766
Lagged income earned and not collected on loans to total assets			-5.7267	-1.79		0.2369
Lagged total unused commitments to total assets			-0.1922	-2.37		-0.1273
Lagged total notional value of derivatives to total assets			-0.0038	-1.42		2.6472
Lagged total loans and leases to total assets			0.9005	2.02		0.7663
Lagged construction and development loans to total assets						-1.9595
Lagged commercial realestate loans to total assets			-2.1023	-1.56		0.7482
Lagged 1-4 family residential realestate loans to total assets			-1.4605	-1.99		-0.1570
Lagged commercial and industrial loans to total assets			-2.4336	-3.25		-0.6994
Lagged consumer loans to total assets			-2.21689	-3.27		0.1348
Mean square error	0.8379		0.56194			
Adjusted R-square	0.1482		0.423			
optimal LASSO penalty rate					0.2025	0.1055
holdout sample MSE at optimal LASSO penalty					1.2286	3.9043

Table 11: Regression test for independence

<i>Independent variable: income before tax and extraordinary items to assets</i>	<i>Coefficient Estimates</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	0.407	0.029	13.94	0.000
asset growth rate	0.098	0.049	2.00	0.051
square of asset growth rate	-0.072	0.039	-1.86	0.068
cube of asset growth rate	0.013	0.008	1.65	0.104

Table 12: Within Sample Stress Test Model Prediction Errors

model	average prediction error	RMSE	average prediction error w/o outliers	RMSE w/o outliers
1-eq OLS	10.50%	101.62%	0.29%	10.16%
1-eq E-OLS	10.64%	80.82%	0.16%	7.84%
5-eq OLS	10.79%	94.32%	-0.07%	10.00%
5-eq E-OLS	9.48%	74.79%	0.14%	7.52%
1-eq LASSO	16.68%	144.01%	0.47%	9.44%
1-eq E-LASSO	11.24%	99.29%	0.24%	8.34%
5-eq LASSO	19.87%	180.00%	0.49%	12.34%
5-eq E-LASSO	14.63%	154.80%	0.98%	9.96%

RMSE denotes root mean-square prediction errors. Columns 5 and 6 report average prediction errors and RMSEs omitting the last 3 quarterly estimates. 2007Q4 produces an extremely large estimation error in all of the models.

Table 13: Alternative stress test model performance out-of-sample in stress scenario

model	average forecast error	forecast RMSE	cumulative model prediction error over stress horizon	model prediction error as percent of actual cumulative INBFTXEX
1-eq OLS	37,864,491	57,904,214	454,373,893	92.29%
1-eq E-OLS	43,433,998	55,998,898	521,207,981	105.87%
5-eq OLS	46,577,528	65,652,010	558,930,337	113.53%
5-eq E-OLS	77,206,584	83,877,901	926,479,003	188.18%
1-eq LASSO	3,212,422	32,770,239	38,549,063	7.83%
1-eq E-LASSO	19,560,226	34,677,763	234,722,711	-23.49%
5-eq LASSO	-11,913,956	36,130,794	-142,967,477	-29.04%
5-eq E-LASSO	-9,637,812	23,049,008	-115,653,741	-23.49%

Figure 1: Representative bank income before taxes and extraordinary items

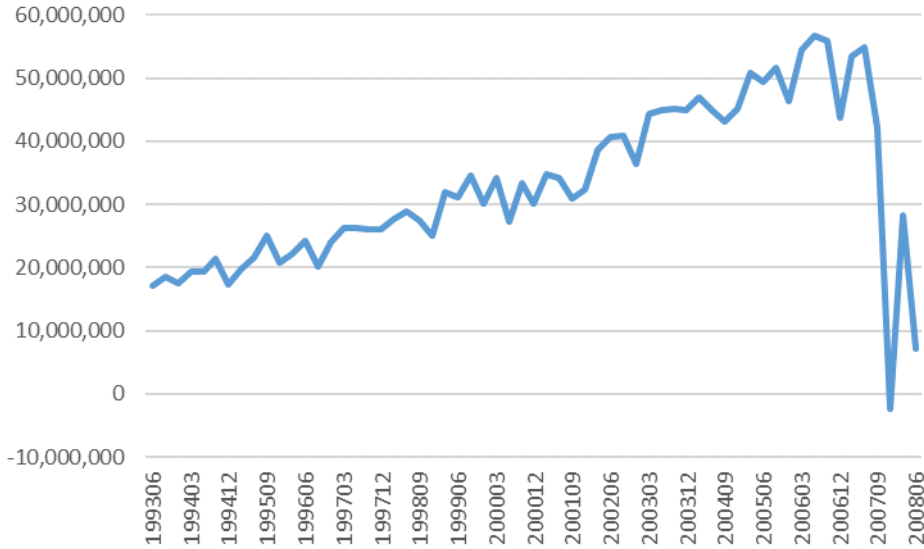


Figure 2: Stress scenario actual and predicted cumulative income before tax and extraordinary items

