Kupiec, Paul H.

Working Paper

On the accuracy of alternative approaches for calibrating bank stress test models

AEI Economics Working Paper, No. 2018-04

Provided in Cooperation with:
American Enterprise Institute (AEI), Washington, DC

Suggested Citation: Kupiec, Paul H. (2018) : On the accuracy of alternative approaches for calibrating bank stress test models, AEI Economics Working Paper, No. 2018-04, American Enterprise Institute (AEI), Washington, DC

This Version is available at:
https://hdl.handle.net/10419/280581

On the Accuracy of Alternative Approaches for Calibrating Bank Stress Test Models

Paul H. Kupiec
American Enterprise Institute

AEI Economics Working Paper 2018-04
April 2018

On the Accuracy of Alternative Approaches for Calibrating Bank Stress Test Models

By Paul H Kupiec[1]

This draft: April 24, 2018

ABSTRACT

Multi-year forecasts of bank performance under stressful economic conditions determine large institution regulatory capital requirements and yet the accuracy of these forecasts is undocumented. I compare the accuracies of alternative stress test model forecasts using the financial crisis as the stress scenario. Models include specifications that mimic the Federal Reserve CLASS model and alternatives that use Lasso, the AIC and an abridged set of explanatory variables. A simple single-equation Lasso model has, by far, the best forecast accuracy. Large differences in model forecast accuracy are undetectable from estimation sample statistics highlighting the need for new stress test model evaluation methods.

Key words: stress test models, Lasso, Dodd-Frank stress tests

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

Disclosure Statement:

I, Paul H. Kupiec, have no conflicts of interest to disclose.

On the Accuracy of Alternative Approaches for Calibrating 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 economic conditions. While details differ, many countries now use "dynamic stress tests" in their large bank supervision processes. These stress tests use complex econometric models to forecast bank income and calculate regulatory capital over one or more hypothetical multi-year economic stress scenarios. Banks are required to project their performance under the specified stress conditions and regulators evaluate bank 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 Wall Street Reform and Consumer Protection Act requires the Federal Reserve Board (FRB) to perform annual stress tests on the largest financial institutions. Should an institution "fail" its stress test, the FRB may take actions including prohibiting bank dividends and capital repurchases or requiring the bank to improve its stress test modeling practices.

In the US, the FRB uses its own stress test model 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 determining which stress test model— the FRB's model or a bank's own internal model— produces a more accurate forecast of bank performance over a hypothetical stress scenario.

The dearth of statistical methods available for reliably determining the accuracy of alternative stress test models is an important issue. In this paper, I analyze the performance of alternative stress test modelling approaches by comparing model forecasts to actual bank performance over a three-year period beginning in September 2008. The results demonstrate that there can be an extremely large divergence in the out-of-sample forecast accuracies of alternative models, and these differences are not predictable based on within-sample regression diagnostics.

My baseline stress test model is built to mimic the design and estimation process that was used to specify the Federal Reserve CLASS model (Hirtle, et.al., 2015). This complex model includes an extensive set of macroeconomic factors and bank-specific characteristics as explanatory variables. It uses a traditional step-wise regression method for variable selection and estimation.

My alternative stress test models are designed to be parsimonious relative to CLASS-style models in order to reduce the chance of overfitting the data. These alternative models include only macroeconomic factors as explanatory variables and make use of the Akaike (1973) information criterion (AIC) and soft threshold coefficient constraints imposed by Tibshirani's (1996) least absolute shrinkage and selection (Lasso) machine learning algorithm. I consider the forecasting accuracy of multi-equation and single equation Lasso model specifications.

My forecast comparison focuses on bank income before tax and extraordinary items (INBFTXEX) because bank income (or loss) is the most important factor driving changes in bank capital adequacy over a three-year stress scenario.[2] Forecasts are constructed for an

[2] 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 to an estimate of INBFTXEX, estimates of retained earnings and bank risk-weighted assets. The INBFTXEX projection process is the part of the capital adequacy calculation that is directly linked to the macroeconomic stress scenario.

"average" or representative bank using quarterly bank regulatory data from March 1993 through June 2008. The forecasts are compared to the actual performance of the representative bank over the first 12 quarters of the financial crisis. The exercise simulates the accuracy of a regulatory stress testing exercise if one had been conducted in the summer of 2008, just prior to the onset of the 2008 financial crisis.

The analysis yields many interesting results. Among the approaches I examine, models that are specified and calibrated using a complex CLASS-style approach generate the least accurate forecasts even though these models fit exceptionally well within the estimation sample. In contrast, a simple single-equation stress test model that uses only macroeconomic factors as explanatory variables produces, by far, the most accurate forecasts. Finally, the combination of the AIC and the Lasso for variable selection and model calibration significantly improves forecast accuracy.

I analyze the forecast errors of the CLASS-style model and show that the magnitude of the error is linked to the complexity of the model. Model forecast errors are magnified when CLASS model specifications are selected from an expanded set of explanatory variables and when the model decomposes INBFTXEX into multiple separately-modeled components. The disaggregation of INBFTXEX and the inclusion of bank-specific characteristics as explanatory variables improves the within-sample fit of CLASS-style stress test models, but these features severely degrade the model's out-of-sample forecast accuracy.

These results also suggest that overfitting the data is a serious concern as stress test models with superior estimation sample fits can have demonstrably poor forecast accuracy out-of-sample. The existing literature on stress testing has paid little attention to these important

issues nor has it discussed any inherent cost-benefit tradeoffs between simple and complex approaches for stress test modeling.[3]

The forecast comparison highlights the need to develop reliable methods for assessing the forecasting accuracy of stress test models. True out-of-sample stress test model forecast errors are rarely observed and, when they are observable, sample sizes are small. Moreover, in the present context, differences in model forecast errors are not covariance stationary, a condition that precludes statistical tests using the Diebold-Mariano (1995) statistic. The regulatory stress testing process would benefit from research that prioritizes the development of new statistical methods for accessing stress test model accuracy.

The remainder of this paper is organized as follows. Section 2 provides background information on the regulatory use of stress tests. Section 3 reviews supervisory stress testing practices with a focus on the use of stress tests in the United States. Section 4 considers important issues related to stress test model specification. Section 5 discusses the data and experimental design used to compare model forecasts. Section 6 reviews the stress test model estimates. Section 7 compares the forecast accuracy of alternative stress test models over the first three years of the 2008 financial crisis. Section 8 focuses on statistical methods for comparing forecast accuracy. Section 9 discusses policy issues and 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

[3] Bidder, Giacomini and McKenna (2016) discuss an alternative approach that may improve stress test accuracy.

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 Basel II and Basel III international capital standards and published guidance on "best practice" standards for regulatory stress testing processes (BCBS, May 2009). Stress tests have been a mandatory component of the IMF's Financial Sector Assessment Program for nearly 20 years.[4]

The benefits of the supervisory use of dynamic stress tests depend, at least in part, on the accuracy of supervisory stress test projections. Yet neither the national laws that require supervisory stress tests, nor guidance issued by international standard setting bodies ever mentions the accuracy of supervisory stress test models as an important 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.

On the specific issue of Federal Reserve stress tests, the accuracy of the FRB's stress test models remains opaque to the public and to the banks required to undergo periodic stress test assessments. In its 2016 official stress testing methodology document (Board of Governors of the Federal Reserve, 2016, p. 13), the FRB describes its own model validation process as "A central

[4] IMF rules require each member country to submit to a Financial Stability Assessment on a periodic basis and the IMF has actively promoted stress testing as an important tool for monitoring financial stability. A collection of IMF staff working papers on stress testing can be found in Ong (2014).

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 guidance on minimum acceptable accuracy standards for stress test models can in part be traced to the hypothetical nature of a stress test exercise. Econometric models are used to simulate a bank's performance over one or more scenarios characterized by severely stressful macroeconomic conditions. Because the stress scenario conditions almost never materialize, it is impossible to compare a bank's actual stress scenario performance to the stress test model's forecast.

If supervisory stress tests are used to set a bank's minimum capital adequacy requirement and mandate improvements in bank stress test processes, it is important to understand whether supervisory actions are predicated on accurate stress test projections. The accuracy of the bank stress test models is a critical issue for the supervision of large complex banking institutions and yet there is little research on this topic.

3. Overview of Supervisory Stress Test Models

The details of the models that bank supervisors use to assess the internal stress tests conducted by banks are kept confidential. FRB officials (Tarullo (2016)) have defended opacity on the grounds 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 fully transparent, it does publish a summary 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." (Board of Governors of the Federal Reserve, 2016, p. 3.)

Both the FRB and the EBA use highly disaggregated approaches to conduct their stress tests.[5] 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 independently. For each institution, each loan category is assumed to grow at the industry average growth rate (Board of Governors of the Federal Reserve, 2016, p.11).[6] Bank net revenues, gross of losses and impairments, are decomposed into 22 separate components, each of which has its own separately estimated model specification.[7]

The complexity of banks' own internal stress test models appears to mirror the complexity of the FRB's supervisory stress test model. 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.

[5] The European Banking Authority (2016). The ClearingHouse (2016) provides a summary comparison of the FRB and EBA stress test methodologies.
[6] The FRB provides no details regarding 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.
[7] The FRB 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 Staff Report on the CLASS model provides perhaps the most detailed description of a regulatory stress test model (Hirtle, et. al., 2015). Bhanot, Hirtle, Kovner, and Vickery (2014) suggest that the CLASS model has been used in earlier Federal Reserve stress tests exercises. The CLASS model is estimated using publicly available regulatory data whereas the FRB's Dodd-Frank stress test models use confidential supervisory data that is not available to the public.

The CLASS model disaggregates bank net non-provision income-to-asset ratios into 6 components that are modeled independently: net interest income, trading income, non-interest non-trading income, compensation expense, expenses related to premises and fixed assets, and other non-interest expense. Each individual component model includes endogenous variables that measure time-varying bank characteristics.

The CLASS model does not estimate bank loan and lease loss provision expenses directly. Rather, it uses a judgmental algorithm to project quarterly loan and lease loss provisions during a stress scenario. The algorithm, described in the appendix, uses forecasts of loan and lease write-off amounts generated from individual models for 15 separate categories of loans and leases.[8] CLASS model loan and lease loss provision estimates are not calibrated to reproduce actual historical loan and lease loss provisions.

[8] First lien and junior lien residential mortgages, home equity lines of credit (HELOC), construction loans, multifamily and non-farm non-residential commercial mortgages, credit cards, other consumer loans, commercial and industrial (C&I) loans, leases, loans to foreign governments, loans to depository institutions, agriculture loans, other real estate loans, and all other loans. [8] Quarterly loan and lease write-off amounts scaled by loan and lease initial balances.

The within-sample fits of the individual models that comprise the Federal Reserve CLASS model are all exceptionally good. Many individual CLASS regression models have adjusted R^2 statistics close to 90 percent.

4. Issues in Stress Test Model Specification

Because regulatory data on bank income and assets have a common time-trend, bank stress test models are typically estimated in ratio form, with bank assets or balance sheet values used to deflate bank income and expense components.[9] 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 estimates. I adopt the CLASS model convention and model bank income and expense ratios as stationary time series.[10]

CLASS stress test models are calibrated using a traditional step-wise econometric modelling approach. Stress test model specifications are chosen on the basis of model fit within the estimation sample. Stress test model specifications are selected to be parsimonious, to produce a high adjusted-R^2 and to minimize the regression residual standard error estimate among specifications that have statistically significant coefficient estimates. Variables that have only weak statistical significance are typically dropped from model specifications during stress

[9] 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.

[10] Subsequent results for the Lasso stress test models show that these series can be modeled as stationary time series.

test development. CLASS models may also require that model coefficient signs and magnitudes be consistent with economic priors (Hirtle, et. al., p. 20).[11]

Focusing on stress test model fit within the estimation sample can create important issues regarding out-of-sample forecast accuracy. There is the well-known risk of model overfitting and including variables that have spurious explanatory power in the estimation sample.[12] It is also possible that the use of explanatory variables that measure bank-specific characteristics, while improving model fit within the estimation sample, may negatively impact forecast accuracy.

The inclusion of time-varying bank characteristics as stress test explanatory variables has intuitive appeal. Their presence imparts model flexibility. When an individual 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. As a consequence, the models seemingly adapt, at least in part, to individual bank performance characteristics. The use of time-varying bank characteristics in every CLASS regression model is a testament to the intuitive appeal of this approach.

However, including time-varying bank characteristics as explanatory variables can have a detrimental impact on forecast accuracy. Individual banks may not have the same coefficient value attached to the specific characteristics included in the models. For example, a bank that imposes very strict underwriting standards for mortgage approvals may have a coefficient on a mortgage share variable that differs substantially from the coefficient value for an average bank using average underwriting. Regional differences in housing markets conditions could also

[11] The CLASS model specification searches do not make use of any information criteria like the AIC, BIC or Mallows Cp statistics.
[12] See, for example, Forster and Sober (1994), Forster (2000), Forster (2001), Leinweber (2007), Green and Armstrong (2015), and Hastie, Tibshirani and Friedman (2009).

create differences among bank coefficient estimates even if banks used similar underwriting standards. Whether or not forecasting accuracy is improved by the inclusion of bank-specific explanatory variables is an open issue.

Moreover, and more importantly for the present analysis, bank balance sheet and exposures characteristics are endogenous variables that change as banks respond to changing macroeconomic conditions. Consequently, while bank characteristics should be modeled and forecast over the stress scenario, current supervisory practice is to hold a bank's characteristics constant throughout the scenario.[13]

The inclusion of bank-specific explanatory variables may improve stress test model fit during the estimation sample, but it may reduce the impact multipliers of the macroeconomic factors and lagged dependent variables that drive bank stress scenario forecasts. Since bank-specific variables are held constant over the stress scenario, the inclusion of these variables may reduce the model sensitivity of out-of-sample forecasts to macroeconomic factor variability.

A final issue concerns the level of aggregation used in the stress test model. The CLASS, FRB, EBA and indeed the internal stress test models of many banks are complex. The models 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. Moreover, if the CLASS modelling approach has been applied more generally, the individual component models used in the stress test are estimated independently without taking account of error correlations. Individual model errors are likely to be correlated and model forecast errors may compound when many separate income and expense

[13] The FRB's stress tests keep bank characteristics constant throughout the stress scenario.

model projections are aggregated to forecast bank net income. There is no guarantee that a disaggregated approach will produce more accurate stress scenario forecasts.[14]

5. Data and Experimental Design

I use the financial crisis that began in 2008 to compare the accuracy of alternative stress test modeling approaches for a representative bank. I compare stress test model forecasts with actual performance when true underlying macroeconomic conditions are used to seed the stress test scenario. The representative bank is constructed as the asset-weighted average of data on all insured institutions.

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 as reported in the "Reports of Conditions and Income" regulatory filings.[15] Stress test models are estimated using data from March 1993 through June 2008 and the stress scenario forecasts are for the period September 2008 through June 2011.

The representative bank's performance and time-varying characteristics are measured by asset-weighted banking system average values. Let w_{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

[14] An analog from astronomy is the historical debate between the Copernican and the Ptolemaic model of planetary motion. The simple Copernican system eventually displaced the much more complex Ptolemaic model because the Copernican model was both simpler and more accurate it its forecasts of planetary motion. See Forester and Sober (1994).

[15] The Call Report data is from the Federal Deposit Insurance Corporation's *Statistics on Depository Institutions.* These reports include 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 asset-weighted average of these data are used to construct the "representative bank" even though the aggregate data includes data from other insured depository institutions.

total value of all 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,

$$\sum_{\forall\, i} \left(\frac{w_{it}^{j}}{a_{it}}\right) \frac{a_{it}}{A_t} \;=\; A_t^{-1} \sum_{\forall\, i} w_{it}^{j} \tag{1}$$

The list of variables that measure bank income, balance sheet and exposure characteristics used in the analysis along with their sample summary statistics are reported in in Table 1.

The macroeconomic factors used in the analysis are derived from data provided in the Federal Reserve Bank of St. Louis 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 used in the analysis are reported in Table 2 along with summary statistics for the sample period March 1993 through June 2011.

5.1 The CLASS-style stress test model

My CLASS-style model decomposes bank INBXTEX into five components: net interest income, noninterest income, noninterest expense, security gains and losses, and provisions for loan and lease losses. I model bank provisions for loan and lease losses directly rather than employing the algorithm used by the CLASS model because the CLASS algorithm does not accurately reproduce the historical loan and lease loss provisions reported by banks.

To remove the effect of the common time trend, each of the five components that comprise INBFTXEX are modeled in ratio form, with the representative bank's total assets in a quarter used as the common denominator. Each of the 5 component CLASS models include both bank-specific characteristics (Table 1) and macroeconomic factors (Table 2) as explanatory

variables. Because bank-specific characteristics are endogenous, I use the lagged values of these variables in the respective regression models.

The preferred parameterization of each component model was selected using the same step-wise algorithm employed by the Federal Reserve's CLASS model. The model was estimated including all potential explanatory variables including a constant and a lagged dependent variable. In subsequent estimation steps, variables that exhibit weak statistical significance are excluded from the model. The final model specification selected was the model that included statistically significant explanatory variables and produced the smallest regression mean square error estimate. In one case, 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.[16] I retained a constant term in each specification regardless of statistical significance.[17]

Forecasts of INBFTXEX require stress scenario estimates of the representative bank's asset growth rate. Following the CLASS model approach, I use the historical average asset growth rate to project bank assets in each quarter of the stress scenario.

5.2 The alternative stress test models

The alternative stress test models I consider include only macroeconomic factors as explanatory variables. One alternative models the ratio of INBFTXEX to assets as a single equation. A second disaggregates the ratio into the same 5 component ratios used in my CLASS-

[16] The preferred CLASS model for non-interest income include some variables with relatively weak t-statistics because excluding these variables increases the regression model standard error estimate.
[17] In the bank provision model, the constant is not statistically significant, but is retained regardless.

style model. Both models forecast bank asset balances using a separate regression model that projects the representative bank's asset growth rate as a function of macroeconomic factors.

Variable selection and calibration of these alternative models is accomplished using the AIC and Tibshirani's (1996) Lasso algorithm. The Lasso is designed to reduce the chances of overfitting the data and including spurious explanatory variables. The Lasso has the potential to produce forecasts with lower mean-square-errors than ordinary least squares forecasts, but at a cost of producing biased estimates of the model coefficients. As a consequence, the statistical significance of individual Lasso coefficient estimates is not an important consideration.

Lasso minimizes the model's means square error while simultaneously imposing a penalty on the sum of the absolute values of the model's coefficient estimates. If y_t represents an observation on the dependent variable, x_{it} represents an observation on the i^{th} explanatory variable and β_i represents the coefficient on the i^{th} explanatory variable, Zou, Hastie and Tibshirani (2007) show that the Lasso coefficient estimates satisfy,

$$\hat{\beta} = \arg\min_{\beta}\left[\left(\sum_{t=1}^{T}\left(y_t - \sum_{i=1}^{p} x_{it}\beta_i\right)\right) + \gamma \sum_{i=1}^{p}|\beta_i|\right] \quad (2)$$

The Lasso penalty on the L^1 norm of the coefficients, γ, imposes a soft-threshold condition on the coefficient estimates. Coefficient estimates are set to zero unless the estimate's absolute value exceeds a threshold set by the penalty rate. If the penalty rate is set to zero, Lasso produces ordinary least squares estimates. If the penalty rate is set too high, Lasso will set all coefficient estimates to zero.

I estimate the Lasso model using the Wu and Lange (2008) coordinate decent algorithm over the global grid of relevant Lasso penalty rates using data from March 1993 through June 2008. For each Lasso penalty rate, I calculate the AIC estimate following Zou, Hastie and

Tibshirani (2007). I select the Lasso penalty rate and corresponding Lasso coefficient estimates that minimize the AIC.

My Lasso model specifications include only macroeconomic factors as explanatory variables.[18] I exclude bank-specific balance sheet and exposure characteristics to reduce the chances of overfitting as Flynn, Hurvich, and Simonoff (2017) show that the introduction of extraneous explanatory variables can compromise the performance of Lasso-based forecasts. Another reason to exclude these variables is the regulatory practice of holding the values of these variables constant over the stress scenario. This convention makes it highly unlikely that the addition of bank-specific variables will improve the accuracy of stress test forecasts.

My calibration and use of the Lasso differs from earlier applications in the stress testing literature. Kapinos and Mitnik (2015) use Lasso to identify a parsimonious number of macroeconomic variables, from which, principle components are extracted and used as stress test model explanatory variables. Chan-Lau (2017) uses Lasso in a stress test model of default probabilities. The Chan-Lau study uses a large cross sectional data set and calibrates Lasso using multiple holdout samples.

6. Model Estimation Results

The best-fitting CLASS model estimates are reported in Table 3. All CLASS component models fit the estimation sample data exceptionally well. All models include bank-specific explanatory variables as well as macroeconomic factors. Interestingly, no model includes the contemporaneous GDP growth rate as an explanatory variable. All models exhibit adjusted-R^2 statistics in excess of 83 percent.

[18] There is no intercept term in a Lasso model because means are removed from all variables.

Lasso requires that the data be normalized. All variables have their sample means removed and explanatory variables are normalized by their sample standard deviation estimates. Means and standard deviations are calculated over the period March 1993 through June 2008.

The 5-equation Lasso coefficient estimates are reported in Table 4. The estimates correspond with the penalty rate that produces the smallest AIC for each respective model. Each of these component Lasso models records its minimum AIC value with a positive Lasso penalty rate. The root mean-square error (RMSE) of each component Lasso model exceeds the RMSE of its CLASS-model counterpart suggesting that each of the CLASS models has the superior in-sample fit.

Table 5 reports optimal estimates for the single equation Lasso model and the representative bank's asset growth rate. Again the coefficient estimates correspond to the Lasso penalty rates that minimizes each model's AIC. The optimal Lasso penalty rate for the single equation model for INBFTXEX-to-assets is 0, indicating ordinary least squares is optimal. In contrast, the optimal growth rate model has a large penalty rate. The latter also has a relatively large RMSE suggesting a poor in-sample fit.

Table 6 compares the implied stress test model coefficient estimates for the ratio of INBFTXEX to assets from the 5-equation CLASS model, the 5-equation Lasso model, and the single-equation Lasso model. The ratio can be reconstructed from the 5 component models as,

$$\frac{INBFTXEX_t}{A_t} = \frac{NIM_t}{A_t} + \frac{NII_t}{A_t} - \frac{NIEX_t}{A_t} + \frac{SECGL_t}{A_t} - \frac{PrLLL_t}{A_t}, \tag{3}$$

where $INBFTXEX_t$ denotes interest before taxes and extraordinary items, NIM_t net interest income, NII_t non-interest income, $NIEX_t$ non-interest expense, $SECGL_t$ securities gains and losses, $PrLLL_t$ provision for loan and lease losses, and A_t total bank assets (all at quarter t).

Replacing each ratio with its model estimate, and aggregating across coefficients, yields the coefficients for the 5-equation models reported in Table 6.

The growth rate of GDP appears in both Lasso models in Table 6 whereas this macroeconomic factor is completely forced out of the optimal specification in the 5-equation CLASS-style model. Similarly, the 3-month and the 1-year Treasury yields are important explanatory variables in the Lasso models whereas both variables have diminished importance in the 5-equation CLASS-style model.[19] The results show that the inclusion of bank-specific explanatory variables in the CLASS-style model can reduce the model sensitivity to stress events by reducing the magnitude of impact multipliers attached to the macroeconomic factors.

The within estimation sample fit, as measured by the respective models' root-mean square regression errors, strongly favors the 5-equation CLASS model. The 5-equation Lasso model has a RMSE that is nearly twice as large as the RMSE of the 5-equation CLASS model; the RMSE of the single-equation Lasso model is over 1.5 times the RMSE of the CLASS model.

7. Stress Test Predictive Accuracy During the Financial Crisis

The accuracies of three alternative stress test models are compared using out-of-sample forecasts over the first 12-quarters of the financial crisis. The forecasts use the actual stress scenario values for the macroeconomic factors, and like the Federal Reserve approach, they hold the representative bank's time-varying characteristics fixed at June 2008 values.

The representative bank's asset value is forecasted using the applicable asset growth rate model. The CLASS-style model uses the historical average growth rate of 1.79 percent per

[19] The Lasso model explanatory variables are normalized so the impact multiplier of a 1-standard deviation shock is equal to the coefficient estimate. The impact of a 1 standard deviation shock to the 3-month Treasury rate is 0.014 in the 5-equation CLASS model compared to 0.211 in the single equation Lasso model, and 0.10 in the 5-equation Lasso model. The impact of a 1 standard deviation shock to the 10-year Treasury yield is -0.002 in the 5 equation CLASS model compared to -0.13 in the 5-equation Lasso model and -0.164 in the 1-equation Lasso model.

quarter. The Lasso models uses growth rate forecasts from the Lasso asset growth rate equation (Table 5). The representative bank's starting asset balance as of the end of June 2008 is arbitrary. To keep the analysis simple, I set this asset balance at $1 in June 2008.

Forecast accuracy is assessed by comparing each model's stress scenario forecast to the representative bank's actual performance. Let the superscript f represent a forecasted value. Let t denotes stress scenario time in quarters, and grw_t denote the quarterly asset growth rate from quarter $t-1$ to t.

The forecast of $\frac{INBFTXEX_t}{A_t}$ for the first stress test quarter constructed from a 5-equation stress test model is given by,

$$\frac{INBFTXEX_1^f}{A_1} = \left[\left[\frac{NIM_1}{A_1}\right]^f + \left[\frac{NII_1}{A_1}\right]^f - \left[\frac{NIEX_1}{A_1}\right]^f + \left[\frac{SECGL_1}{A_1}\right]^f - \left[\frac{PrLLL_1}{A_1}\right]^f\right]. \quad (4)$$

The corresponding first quarter forecast of INBFTXEX is,

$$INBFTXEX_1^f = \left[\left[\frac{NIM_1}{A_1}\right]^f + \left[\frac{NII_1}{A_1}\right]^f - \left[\frac{NIEX_1}{A_1}\right]^f + \left[\frac{SECGL_1}{A_1}\right]^f - \left[\frac{PrLLL_1}{A_1}\right]^f\right] \times \left(1 + grw_1^f\right), \quad (5)$$

where the growth rate forecast in (5) is selected appropriately for each model. The first period forecast of INBFTXEX from the single equation Lasso model is,

$$INBFTXEX_1^f = \left[\frac{INBFTXEX_1}{A_1}\right]^f \times \left(1 + grw_1^f\right), \quad (6)$$

where the asset growth rate forecast is from the Lasso growth rate model.

Many of the stress test models include a lagged dependent variable. The lagged dependent variable values are set equal to June 2008 values in the initial stress test quarter. For

all subsequent quarters in the stress scenario, lagged dependent variable values are set equal to the prior quarter's predicted value.

The stress scenario forecasts are true out-of-sample forecasts. The 5-equation model forecast of INBFTXEX for the q-th quarter in the stress scenario is given by,

$$INBFTXEX_q^f = \left[\left[\frac{NIM_q}{A_q}\right]^f + \left[\frac{NII_q}{A_q}\right]^f - \left[\frac{NIEX_q}{A_q}\right]^f + \left[\frac{SECGL_q}{A_q}\right]^f - \left[\frac{PrLLL_q}{A_q}\right]^f\right] \times \prod_{t=1}^{q}\left(1 + grw_t^f\right) \quad (7)$$

The q-th quarter single-equation Lasso model forecast is constructed in an analogous manner.

Figure 1 plots the stress scenario actual and forecasted quarterly values of the INBFTXEX-to-assets ratio from the 5-equation CLASS-style model, the 5-equation Lasso model, and the single equation Lasso model. Figure 1 shows the CLASS-style model forecasts to be inferior to both Lasso model forecasts. Among the Lasso models, the single-equation model produces the most accurate forecasts. The results do not suggest there are benefits to be gained by using a disaggregated stress test modeling approach.

Figure 2 decomposes the 5-equation CLASS-style model forecast errors by comparing the accuracy of the 5-equation CLASS-style model (Table 3), with two simplified models calibrated using the CLASS-style approach.[20] The first is a 5-equation model that excludes bank-specific explanatory variables. The second is a single equation model that excludes bank-specific explanatory variables but still uses the CLASS-style step-wise regression approach for variable selection. The forecasting accuracy of these alternative models are reported in Table 7.

[20] The individual coefficient estimates for these alternative models are unimportant for the discussion so they are omitted in the interest of brevity.

The plots in Figure 2 show that the inclusion of bank-specific explanatory variables is an important source of forecast error. When bank-specific explanatory variables are excluded from the 5-equation CLASS-style model, the stress test model root mean-square forecast error (RMSFE) drops from 0.842 to 0.182 (Table 7). The reduction in forecast RMSFE that occurs when moving from a 5-equation CLASS-style model without bank-specific variables to a single equation CLASS-style model without bank-specific variables is less dramatic (a RMSFE reduction of 0.047).

The results reported in Table 7 show that forecasts of the ratio of INBFTXEX to assets generated by the single equation CLASS-style model are inferior to forecasts from the single-equation Lasso model. The forecast accuracy improvement in this instance owes to the superiority of AIC variable selection process (as compared to the CLASS step-wise regression). The Lasso is not a source of improvement as the penalty rate (zero) corresponds to ordinary least squares estimation.

Figure 3 plots the actual and forecasted values of the representative bank's asset growth rate over the stress scenario. While the actual assets in the representative bank declined in many stress scenario quarters, both the Lasso growth rate model and the historical average growth rate used by the CLASS model predict positive asset growth in each quarter. The use of the historical average growth rate, while inaccurate, produces a better forecast than the Lasso growth rate model.

Figure 4 plots the actual and forecasted value of INBFTXEX for each stress test quarter for the 5-equation CLASS-style model and the single equation Lasso model. The simplest Lasso model closely tracks the representative bank's INBFTXEX over the entire stress scenario whereas the 5-equation CLASS-style model produces very large forecast errors.

The capital adequacy position of a bank hinges on the bank's cumulative performance over the entire 12-quarter stress period. Figure 5 plots the actual and forecasted values of the representative bank's cumulative INBFTXEX over the stress scenario. After recording an initial positive value of INBFTXEX in September 2008, the representative bank posts a large loss in December 2012, followed by small additional losses in the next two quarters. In September 2009, the representative bank again generates positive INBFTXEX and continues to record positive INBFTXEX in each remaining stress test quarter. The largest cumulative INBFTXEX loss, - 0.29, occurs in June 2009.

Figure 5 shows that the 1-equation Lasso model provides a reasonably accurate forecast of the representative bank's cumulative earnings over the stress scenario. This model forecasts INBFTXEX losses in the first four quarters of stress scenario, with profits thereafter. The model is too pessimistic in the first quarters of the stress scenario and slightly optimistic about the bank's projected recovery in the latter quarters of the stress scenario.

The largest Lasso model cumulative loss forecast, nearly -0.42, occurs in June 2009. Thus, from a capital adequacy standpoint, the single-equation Lasso model is conservative. It projects a worst case cumulative loss that is nearly 44 percent larger than the actual worst-case cumulative loss posted by the representative bank. Overall, the simple single-equation Lasso model performs remarkably well considering the exercise is a true out- of-sample forecast over a stress scenario that corresponds to the worst financial crisis since the great depression.

In contrast, the out-of-sample stress scenario forecasts from the 5-equation CLASS-style model are remarkably poor given this model's exceptionally strong within-sample goodness-of-fit statistics. The model projects INBFTXEX losses in each stress test quarter (Figure 4) with the

losses growing larger in latter quarters. As a consequence, the CLASS-style stress test model forecasts cumulative losses that compound and accelerate over the course of the stress scenario.

8. Identifying the Most Accurate Stress Test Model

Figures 4 and 5 provide a convincing visual case for the superior forecasting accuracy of the single-equation Lasso stress test model. However, the ability to compare the out-of-sample forecasting accuracy of these alternative stress tests models over an actual stress scenario is only possible because there is an actual stress event in the data. It is unclear how to reproduce these conditions when evaluating bank stress test models that are estimated using more contemporary data that do not include a stressful economic event.

Stress test models are built to forecast bank performance over periods of severe economic stress. Regulatory stress test scenarios are hypothetical events that (fortunately) never materialize, so it is impossible to observe the actual forecasting accuracy of stress test model projections over these hypothetical scenarios. Forecast errors are, for the most part, unobservable, so judgements about model accuracy must be based on statistics that measure the within-sample fit of a stress test model. The results of the prior analysis confirm what statisticians have long understood—that a model's within-sample fit can be a highly misleading indicator of the model's out-of-sample forecasting performance. Indeed, as Forster and Sober (1994) argue, a model that fits the estimation sample too well should be cause for concern.

By using the actual 2008 financial crisis as the stress scenario, my analysis identifies the true out-of-sample forecast errors produced by each model. If there is agreement about the economic loss function associated with forecast errors, the Diebold-Mariano (1995) statistic may provide a mechanism for identifying statistically significant differences in the forecast accuracy of the alternative stress test models.

Let e_{jt} represent the forecast error for model j at date t, and $g(e_{jt})$ represents the economic loss associated with a forecast error e_{jt}. The Diebold-Mariano statistic can detect differences in the expected economic loss rates that are generated by forecast errors from competing models. The null hypothesis is $E[g(e_{1t})] = E[g(e_{2t})]$.

The Diebold-Mariano statistic is given by,

$$DM = \frac{\sum_{t=1}^{T}[g(e_{1t}) - g(e_{2t})]}{\sqrt{T}}. \quad (8)$$

If the differences in the economic loss realizations are covariance stationary, for large T, under the null hypothesis, the DM statistic converges to a normally-distributed mean zero random variable with a variance equal to 2π times the spectral density of the loss differential at frequency zero. If the loss differential is covariance stationary, the variance of the DM statistic can be constructed as a weighted average of sample autocorrelation estimates of the loss differential series.

Figure 6 plots the economic loss differential for each quarter in the stress scenario when the economic loss is symmetric, measured by the square of the model's forecast error. Figure 6 plots the difference, $d_t = e_{t1}^2 - e_{2t}^2$, for each quarter t in the stress scenario, where model 1 is the 5-equation CLASS model (Table 3) and model 2 is the single-equation Lasso model (Table 5). A visual inspection of the time series in Figure 6 suggests that economic loss differential could be a nonstationary series. A formal Dickey-Fuller regression test cannot reject the null hypothesis that the loss differential series, d_t, contains a unit root,[21]

$$\begin{array}{ccccc} \Delta d_t = & -0.227 & -0.947\ d_{t-1} & +0.195\ t & \\ & (-1.67) & (-2.90) & (-3.13) & \end{array} \quad (9)$$

[21] T-statistics are in parenthesis. For a sample size of 25 observations, Fuller (1976) reports that statistical significance at the 5 percent level of the test requires a t-statistic on the lagged value of the series to be smaller than - 3.60. The current test statistic is based on 11 observations, so the true critical value is less than -3.60.

Since the economic loss differential series is not covariance stationary, the Diebold-Mariano test is not applicable in this particular situation. The loss differential series non-stationarity is not solely a consequence of the squared-error loss function. Figure 4 indicates that the forecast error for the 5-equation CLASS-style model grows continuously over the stress scenario. Any loss function that puts significant weight on errors that under-predict INBFTXEX will likely generate a non-stationary loss differential series. For example, if the economic loss function is given by, $g(e_{jt}) = e_{jt}$, the loss differential series is still nonstationary.[22]

9. Policy Implications

The very large forecasting errors exhibited by the CLASS-style model highlight the need for new approaches for evaluating the accuracy of regulatory stress test models. While my analysis does not evaluate the Federal Reserve's CLASS model, it does evaluate the forecasting accuracy of a similarly-constructed model with similar within-sample goodness of fit characteristics.

The results of this analysis show that the within-sample goodness of fit measures that are used to guide construction of CLASS-style models are highly unreliable. Forecasting accuracy appears to improve substantially when CLASS-style model specifications are simplified, and variable selection is based on within-sample criterion like the AIC, which penalizes over-parameterized models. The accuracy of the stress test models analyzed in this paper is not merely an issue of academic concern as Hirtle et. al. (2015, p. 32) report that the CLASS model estimates are very similar to those produced by the FRB's Dodd-Frank stress test model.

[22] The Dickey-Fuller regression t-statistic on d_{t-1} is -2.58, which is insufficient to reject the null hypothesis that the series contains a unit root.

True out-of-sample comparisons of forecast accuracy under historical stressful conditions, while suitable for an academic study, are unlikely to be a realistic approach for evaluating the performance of banks' contemporary stress test models. Moreover, even under the assumptions that allowed me to construct an out-of-sample stress test model forecast comparison, the existing statistical methods for comparing the forecast accuracy of competing models were inapplicable because of the time series characteristics of the model errors.

If the goal of stress testing is the accurate prediction of bank performance over prolonged stressful economic episodes, then there is a clear need to develop more reliable methods for calibrating stress test models. Some of the calibration methods that have been discussed in staff reports issued by important bank regulators appear to be problematic. Moreover, flaws in existing approaches can be disguised by spuriously tight model fits within the stress test estimation sample. Because most bank risk managers and bank regulators will never conduct a true out-of-sample test of the forecasting accuracy of their models under stressful conditions, many will remain ignorant of the magnitude of the inaccuracies that have unintentionally been built into their regulatory capital planning processes.

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 mimic those used to specify the Federal Reserve CLASS model. The results show that complex CLASS-style models produce stress test forecasts that severely underestimate the income before taxes and extraordinary items earned by a representative bank over the first three years of the financial crisis. In contrast, simplified stress test model specifications produce far more accurate forecasts.

The results challenge the notion that complex stress test modeling approaches are needed to produce an accurate assessment of actual bank capital needs. The results show that complex disaggregated stress test models that include a large number of explanatory variables, like the CLASS model, face an elevated risk of including spurious explanatory variables and ignoring component model error correlations that diminish the model's out-of-sample forecasting performance.

The analysis shows that the traditional step-wise regression approach for specifying stress test models can be significantly improved by using information theory and machine-learning calibration techniques to penalize non-parsimonious models. In particular, the use of the AIC and the Lasso for variable selection and model calibration can vastly improve forecast accuracy compared to the traditional CLASS-style modeling approach.

The relative accuracy of complex versus simplified stress test modeling processes is an important issue. The complex stress test models currently utilized by banks and regulatory authorities are costly to administer. According to United States Government Accountability Office (2016, p. 30), about half of the institutions required to perform Dodd-Frank stress tests reported annual stress test compliance expenditures of between $15-$30 million. One institution reported costs in excess of $90 million. The costs incurred by bank supervisors, while undoubtedly large, have not been reported. It is unclear that the magnitude of such expenditures are justified if simpler less costly stress test models are capable of producing more accurate forecasts.

The results highlight the need for national regulators, multinational standard setting bodies, and researchers to focus on developing reliable methods to assess the accuracy of supervisory stress testing models. The benefits of using costly complex stress testing models to

set minimum capital adequacy standards will remain a controversial issue, and appropriately so, until it can be demonstrated that complex regulatory stress test models achieve superior forecasting accuracy.

Appendix: CLASS Model Algorithm for Loan and Lease Loss Provisions

The CLASS model's provisions for loan and lease losses are driven by estimates for the write-off rates on 15 separate categories of loan and leases. Using the stress scenario assumptions for macroeconomic variables and the individual loss rate model coefficient estimates, the CLASS model forecasts quarterly loan and lease write-off amounts for each category and aggregates them by quarter.

Let LWO_t represent the aggregate CLASS model loan and lease write-off estimate for quarter t in the stress scenario. Let $SLWO_t$ represent the CLASS model estimate of the aggregate loan and lease loss write-offs over the next four stress scenario quarters,

$$SLWO_t = LWO_t + LWO_{t+1} + LWO_{t+2} + LWO_{t+3} \quad \text{(A.1)}$$

Let $ALLL_t$ represent the balance sheet value of the bank's allowance for loan and lease losses at stress test quarter t. Let $PrLLL_t$ represent the estimated value of the bank's loan and lease loss provisions for quarter t. This estimate is given by,

$$PrLLL_t = \begin{cases} SLWO_t - ALLL_t, \ if \ ALLL_t < SLWO_t \\ LWO_t, \quad if \ SLWO_t < ALLL_t \leq 2 \times SLWO_t \\ 2 \times SLWO_t - ALLL_t, \ if \ ALLL_t > 2 \times SLOW_t \end{cases} \quad \text{(A.2)}$$

References

Akaike, H., 1973. "Information theory and an extension of the maximum likelihood principle." In, *2nd International Symposium on Information Theory*, eds. B.N. Petrov and F. Csaki, Budapest: Akademiai Kiado, pp. 716-723.

Bhanot, M., B. Hirtle, A. Kovner, and J. Vickery, 2014. "The CLASS Model: A Top-Down Assessment of the U.S. Banking System," *Liberty Street Economics*, Federal Reserve Bank of New York, June.

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

Diebold, F. and R. Mariano, 1995. "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, Vol. 13, pp. 253-265.

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/

Flynn, Cheryl, Clifford Hurvich, and Jeffery Simonoff, 2017. "On the Sensitivity of the Lasso to the Number of Predictor Variables," *Statistical Science*, Vol. 32, No. 1, pp. 88-105.

Forster, Malcolm, 2000. "Key Concepts in Model Selection: Performance and Generalizability," *Journal of Mathematical Psychology*, Vol. 44, pp. 205-231.

Forster, Malcolm, 2001. "The new science of simplicity," in Simplicity, Inference and Modeling, Arnold Zellner, Hugo Keuzenkamp and Michael McAleer, editors. Cambridge: Cambridge University Press, pp. 83-117.

Forester, Malcolm and Elliott Sober, 1994. "How to Tell When Simpler, More Unified, or Less Ad Hoc Theories Will Provide More Accurate Predictions," *The British Journal for the Philosophy of Science*, Vol. 45, No. 1, pp. 1-35.

Fuller, W.A., 1976. Introduction to Statistical Time Series. New York: John Wiley and Sons.

Green, K and J. S. Armstrong, 2015. "Simple versus complex forecasting: The evidence," *Journal of Business Research*, Vol. 68, No. 8, pp. 1678-1685.

Haste, Trevor, Robert Tibshirani, and Jerome Friedman, 2009. The Elements of Statistical Learning: Data Mining, Inference and Prediction, Second Edition. New York: Springer-Verlag.

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.

Leinweber, D., 2007. "Stupid data miner tricks: Overfitting the S&P500," *Journal of Investing*, Vol. 16, No. 1, pp. 15-22.

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

Tarullo, Daniel, 2016. "Next Steps in the Evolution of Stress Testing," Remarks at the Yale University School of Management Leaders Forum (September 26), https://www.federalreserve.gov/newsevents/speech/tarullo20160926a.htm

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

The United States Government Accountability Office, 2016. Additional Actions Could Help Ensure the Achievement of Stress Test Goals, GAO-17-48, November.

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.

Zou, Hui, Trevor Hastie and Robert Tibshirani, 2007. "On the "Degrees of Freedom" of the Lasso," *The Annals of Statistics*, Vol. 35, No. 5, pp. 2173-2192.

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	$100*(asset_t/asset_{t-1} -1)$	-2.259	3.681	1.516	1.697	1.247
Income before tax and extraordinary items to total assets	$100*(nim_t+nonii_t+iglsec_t-elnatr_t-nonix_t)/asset_t$	-0.297	0.520	0.372	0.431	0.161
Net Interest Income to total assets	$100*nim_t/asset_t$	0.566	0.925	0.804	0.811	0.072
Noninterest income to total assets	$100*nonii_t/asset_t$	0.231	0.600	0.496	0.504	0.067
Noninterest expense to total assets	$100*nonix_t/asset_t$	0.631	0.937	0.794	0.806	0.063
Provisions for loan and lease losses to total assets	$100*elnatr_t/assets_t$	0.053	0.502	0.143	0.097	0.110
Securities gains and losses to total assets	$100*iglsec_t/asset_t$	-0.053	0.054	0.010	0.010	0.019
Total securities to total assets	$100*sc_t/asset_t$	14.625	23.616	18.975	18.964	2.129
Trading account assets to total assets	$100*trade_t/asset_t$	2.095	7.554	4.732	4.705	0.999
Other real estate owned to total assets	$100*ore_t/asset_t$	0.047	0.944	0.179	0.097	0.181
Income earned, not collected on loans to total assets	$100*oaienc_t/asset_t$	0.361	0.662	0.457	0.445	0.064
Total unused commitments to total assets	$100*uc_t/asset_t$	30.611	69.053	55.253	60.417	11.488
Notional value of derivative positions to total assets	$100*obsdir_t/asset_t$	213.936	1829.845	792.183	632.296	460.756
Total loans and leases to total assets	$100*lnlsgr_t/asset_t$	54.045	63.289	60.003	60.593	1.943
Construction and development loans to total assets	$100*lnrecons_t/asset_t$	1.660	4.891	2.873	2.877	1.037
Commercial real estate loans to total assets	$100*lnrenres_t/asset_t$	6.344	8.331	7.190	7.305	0.534
1-4 family residential loans to total assets	$100*lnrers_t/asset_t$	18.299	23.857	20.821	20.567	1.318
Commercial and industrial loans to total assets	$100*lnci_t/asset_t$	8.745	14.856	11.748	11.636	1.799
Consumer loans to total assets	$100*lncon_t/asset_t$	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

Macreconomic 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 Enterpise 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: 5-equation CLASS-style stress test model estimates

Explanatory Variable	Net interest income to assets	t-statistic	Non interest income to assets	t-statistic	Non interest expense to assets	t-statistic	Securities gains (losses) to assets	t-statistic	Provision for loan and lease losses to assets	t-statistic
Intercept	0.4425	4.74	0.3647	2.34	-0.4929	-2.29	0.1555	4.05	-0.0589	-0.24
Lagged dependent variable					-0.3288	-2.72			0.4663	4.65
GDP growth rate										
civilian unemployment rate			-0.0218	-1.46					0.0455	7.32
10-year Treasury yield	0.0448	4.27	-0.0147	-1.67			-0.0181	-12.98	0.0137	3.66
3-month Treasury yield			0.0501	2.01	0.0427	1.96				
Moody's AAA yield	-0.0366	-3.05								
Moody's Baa yield	0.0142	4.21			0.0308	2.89				
Federal funds rate	-0.0207	-7.23	-0.0584	-2.59	-0.0411	-1.90				
Wilshire market index return			0.0008	1.75	0.0012	2.88				
Wilshire daily return standard deviation			-0.0212	-2.10					0.0113	1.77
Kansas City Fed FSI average							-0.0194	-6.89	0.0320	3.94
VIX daily average							0.0017	5.67		
VIX percent change							0.0000	-1.58		
Change in Real House Price Index										
Lagged securities to total assets					0.0129	3.07			-0.0165	-5.17
Lagged trading assets to total assets	0.0190	3.59					0.0036	1.92		
Lagged ORE to total assets					0.2881	4.77				
Lagged income earned and not collected on loans to total assets			-0.2673	-2.90						
Lagged total unused commitments to total assets			0.0047	4.66	0.0065	4.23			-0.0020	-5.59
Lagged total notional value of derivatives to total assets	-0.0003	-6.98	-0.0003	-4.53			-0.0001	-6.48	0.0002	4.79
Lagged total loans and leases to total assets									0.0077	2.19
Lagged construction and development loans to total assets	0.0537	4.59	0.0732	4.37			0.0229	5.88	-0.0511	-4.29
Lagged commercial realestate loans to total assets										
Lagged 1-4 family residential realestate loans to total assets							-0.0042	-2.43		
Lagged commercial and industrial loans to total assets	0.0146	4.13	0.0198	3.97	0.0183	4.07				
Lagged consumer loans to total assets	0.0166	2.51			0.0488	4.55			-0.0223	-3.05
Root mean square error (within sample)	0.0159		0.0236		0.0238		0.00601		0.0126	
Adjusted R-square	0.9438		0.8447		0.8335		0.8611		0.9483	

Table 4: 5-equation Lasso stress test model estimates

Explanatory Variable	Net interest income to assets	Non interest income to assets	Non interest expense to assets	Securities gains (losses) to assets	Provision for loan and lease losses to assets
Lagged dependent variable	0.043	0.021	0.005	0.002	0.028
GDP growth rate	0.007	0.006	0.001	0.002	-0.010
civilian unemployment rate	0.008	-0.027	0.010	0.006	0.005
10-year Treasury yield	0.000	-0.071	0.000	-0.033	0.026
3-month Treasury yield	0.000	0.116	0.063	0.008	-0.039
Moody's AAA yield	0.017	0.057	0.013	0.021	-0.021
Moody's Baa yield	0.003	-0.005	0.031	0.000	0.000
Federal funds rate	0.000	-0.119	-0.061	0.000	0.030
Wilshire market index return	0.001	0.005	0.012	0.000	0.004
Wilshire daily return standard deviation	0.000	-0.010	0.001	0.002	0.000
Kansas City Fed FSI average	0.000	0.001	0.000	-0.016	0.024
VIX daily average	0.003	0.006	0.016	0.011	-0.004
VIX percent change	0.000	0.000	0.000	-0.002	0.007
Change in Real House Price Index	0.000	0.000	0.017	0.001	-0.008
AIC	1.3153	1.4078	1.3924	1.43477	1.4449
optimal Lasso penalty	0.0010	0.0002	0.0004	0.0001	0.0003
Root mean-square error (within sample)	0.0208	0.0271	0.0274	0.0067	0.0161

Table 5: Single-equation Lasso and asset growth rate model estimates

Explanatory Variable	Lasso single-equation INBFTXEX-to-assets	Lasso asset growth rate
Lagged dependent variable	0.006	-0.0754
GDP growth rate	0.024	0.0000
civilian unemployment rate	-0.038	-0.0658
10-year Treasury yield	-0.164	0.0000
3-month Treasury yield	0.211	0.0000
Moody's AAA yield	0.138	0.0000
Moody's Baa yield	-0.015	-0.3048
Federal funds rate	-0.192	0.0000
Wilshire market index return	-0.010	-0.0262
Wilshire daily return standard deviation	-0.011	0.0000
Kansas City Fed FSI average	-0.085	0.0000
VIX daily average	0.038	-0.0305
VIX percent change	-0.011	0.2328
Change in Real House Price Index	0.008	0.0000
AIC	1.467	1.3498
optimal Lasso penalty	0.000	0.0410
Root mean-square error (within sample)	0.047	0.8841

Table 6: Alternative model estimates for INBFTXEX-to-assets

Explanatory Variable	Aggregated 5-equation CLASS-style model	Aggregated 5-equation Lasso model	Single-equation Lasso model
Intercept	1.515		
Lagged NIM to assets	0.000	0.043	
Lagged NII to assets	0.000	0.021	
Lagged NIEX to assets	0.329	-0.005	
Lagged SECGL to assets	0.000	0.002	
Lagged provisions to assets	-0.466	-0.028	
Lagged INBFTX to assets			0.006
GDP growth rate	0.000	0.024	0.024
civilian unemployment rate	-0.067	-0.028	-0.038
10-year Treasury yield	-0.002	-0.130	-0.164
3-month Treasury yield	0.007	0.100	0.211
Moody's AAA yield	-0.037	0.103	0.138
Moody's Baa yield	-0.017	-0.033	-0.015
Federal funds rate	-0.038	-0.088	-0.192
Wilshire market index return	0.000	-0.010	-0.010
Wilshire daily return standard deviation	-0.033	-0.010	-0.011
Kansas City Fed FSI average	-0.051	-0.039	-0.085
VIX daily average	0.002	0.008	0.038
VIX percent change	0.000	-0.008	-0.011
Change in Real House Price Index	0.000	-0.008	0.008
Lagged securities to total assets	0.004		
Lagged trading assets to total assets	0.023		
Lagged ORE to total assets	-0.288		
Lagged income earned and not collected on loans to total assets	-0.267		
Lagged total unused commitments to total assets	0.000		
Lagged total notional value of derivatives to total assets	-0.001		
Lagged total loans and leases to total assets	-0.008		
Lagged construction and development loans to total assets	0.201		
Lagged commercial realestate loans to total assets	0.000		
Lagged 1-4 family residential realestate loans to total assets	-0.004		
Lagged commercial and industrial loans to total assets	0.016		
Lagged consumer loans to total assets	-0.010		
Root mean-square error (within sample)	0.034	0.060	0.047

Table 7: Stress scenario RMSFEs for alternative forecasts of INBFTXEX-to-assets

Stress test model type	stress scenario RMSFE
5-equation CLASS-style model	0.842
5-equation CLASS-style model, no bank-specific variables	0.182
1-equation CLASS model, no bank-specific variables	0.135
1-equation Lasso model	0.084

RMSFE is an abbreviation for root mean-square forecast error.

Figure 1: Alternative stress test model forecasts of INBFTXEX-to-assets

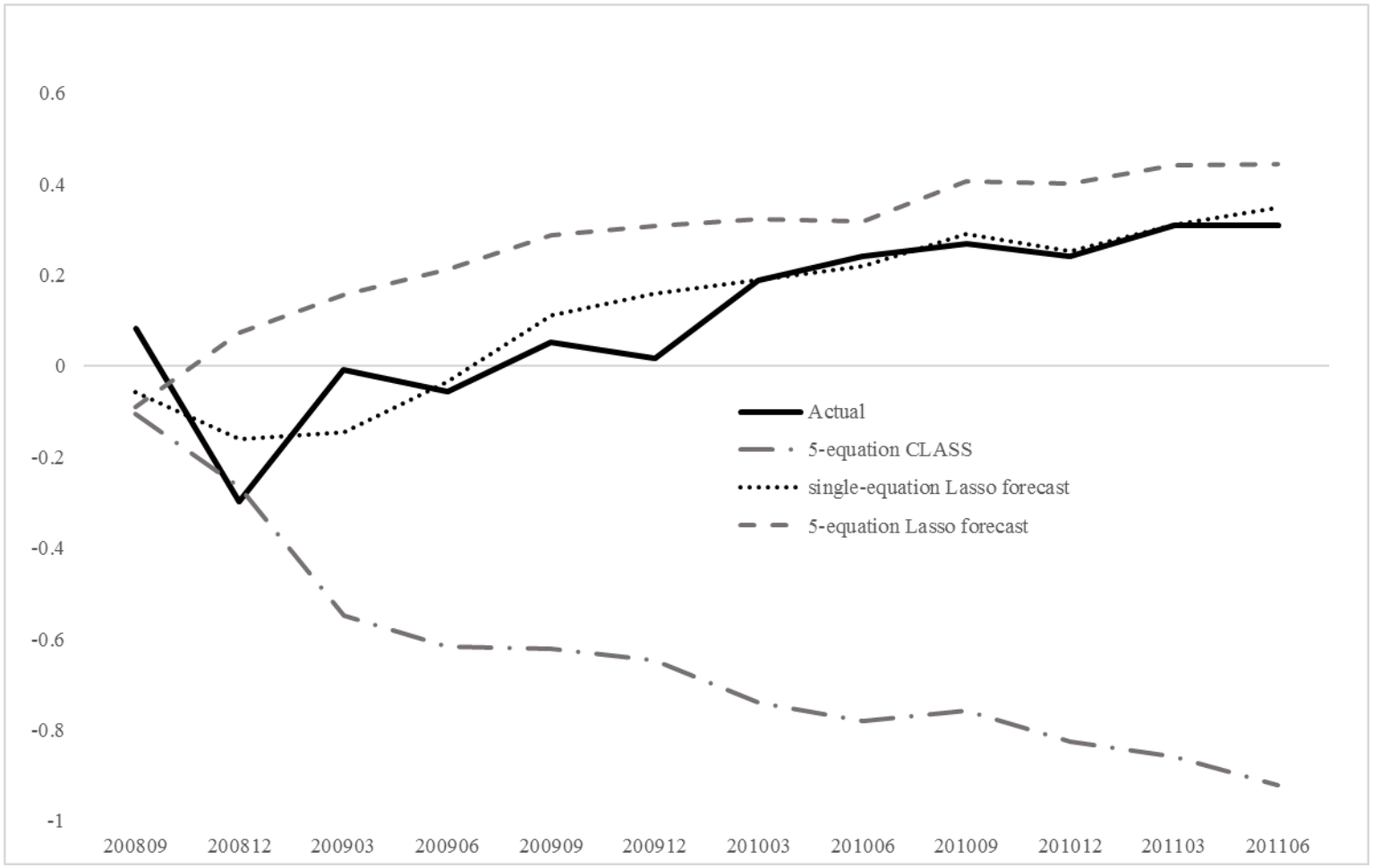

Figure 2: The impact of altering CLASS-style stress test modeling conventions on INBFTXEX-to-asset forecasts

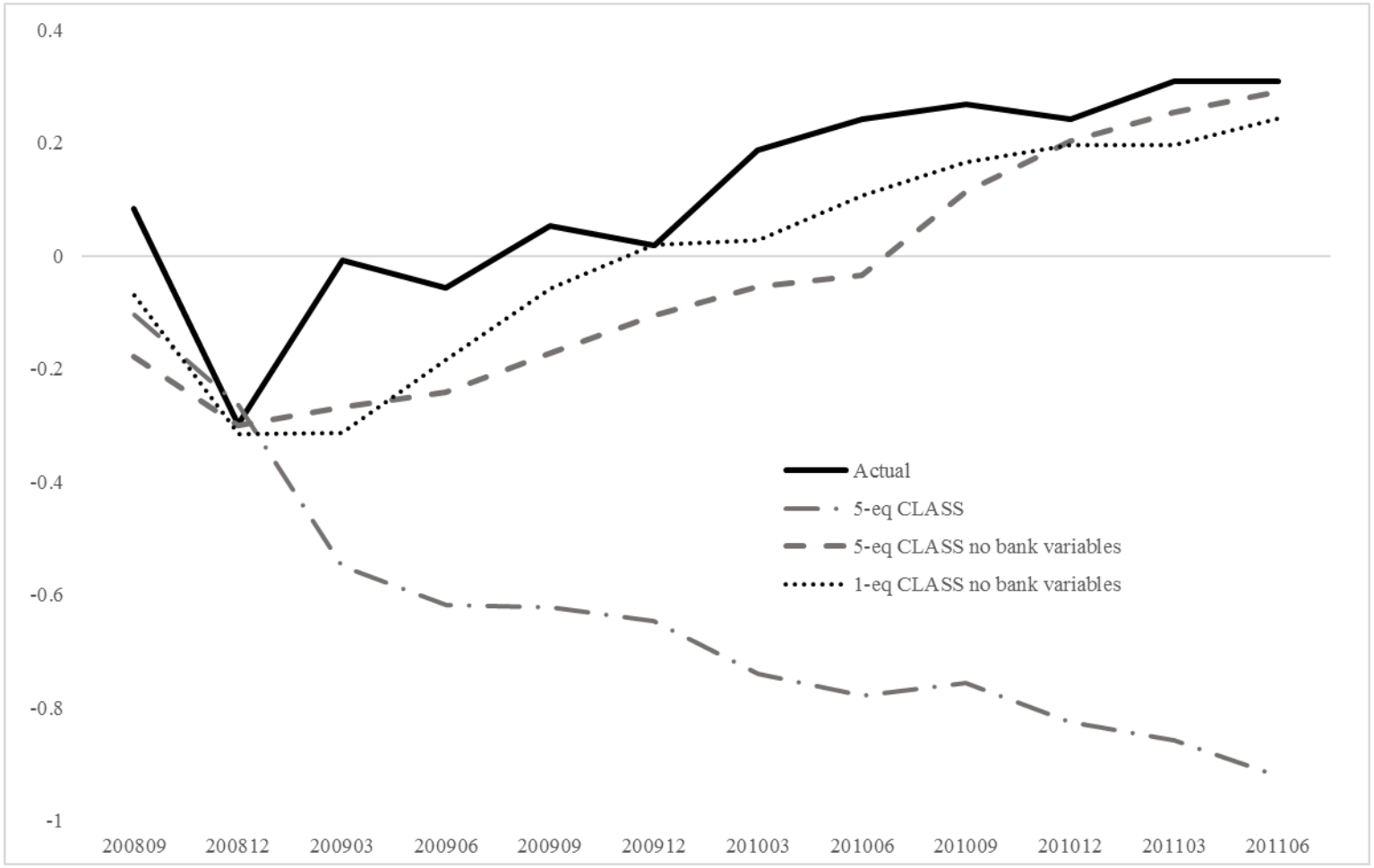

Figure 3: Alternative stress scenario asset growth rate forecasts

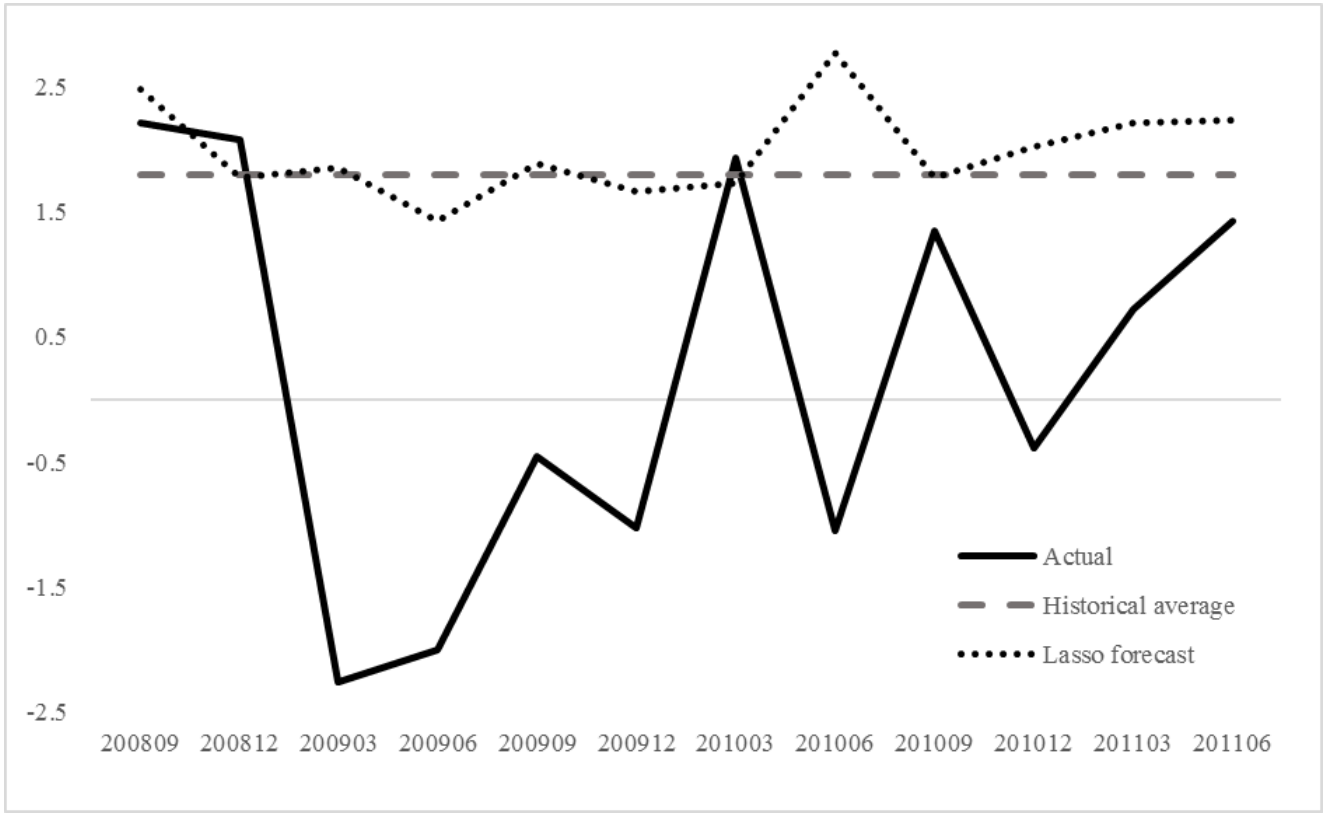

Figure 4: Stress scenario forecasts of quarterly INBFTXEX

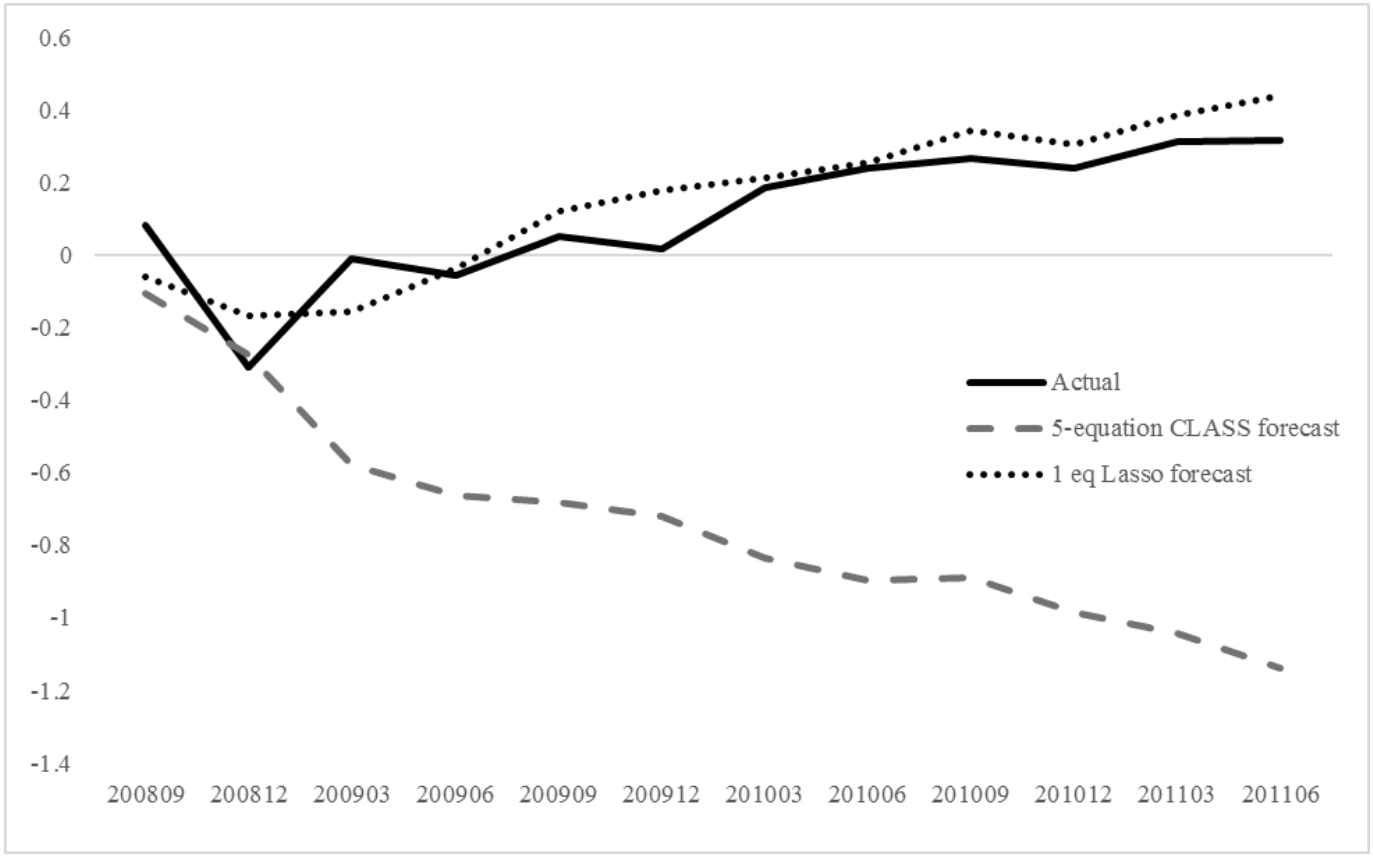

Figure 5: Stress scenario forecasts of cumulative INBFTXEX

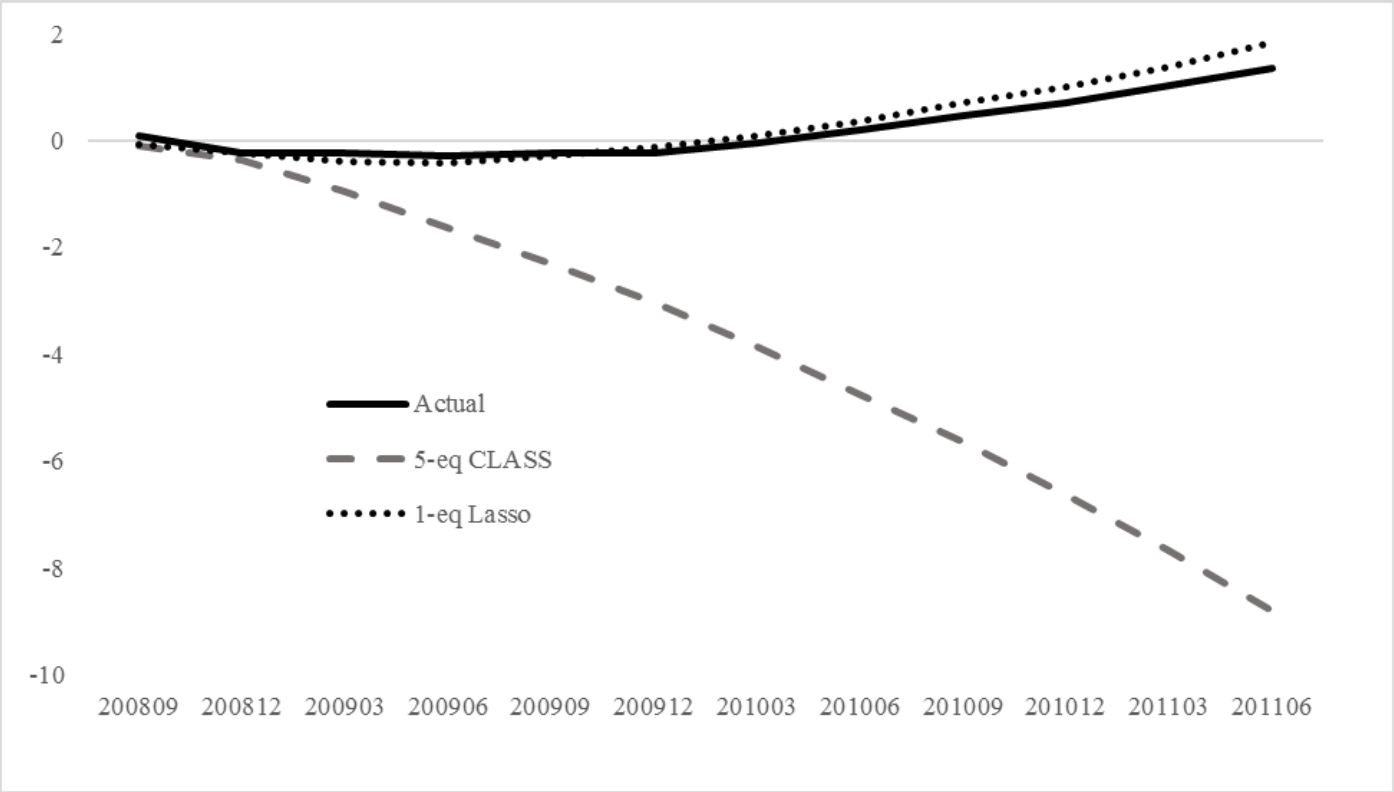

Figure 6: Forecast error loss differentials