

# The Accuracy of Alternative Supervisory Methodologies for the Stress Testing of Credit Risk

Michael Jacobs, Jr.<sup>1</sup>

Draft: May 25<sup>th</sup>, 2018

## Abstract

Banking supervisors have grappled with the problem of determining the optimal level of loss absorbing capital resources that institutions should hold to support their risk taking activities. Following the recent financial crisis traditional approaches such as regulatory capital ratios have been supplanted to supervisory *stress testing* as a primary tool for managing systemic risks. Financial institutions are mandated to perform stress testing to forecast performance over hypothetical multi-year stress scenarios, in the process developing models to support these forecasts. In parallel supervisors conduct their own stress tests and develop supporting models, to set financial institutions' minimum regulatory capital requirements in multiple jurisdictions, yet nothing is revealed to the public regarding the accuracy of such models. In this study we investigate a modeling framework that we believe to be very close to that employed by the regulators, which projects various financial statement line items for an aggregated "average" bank. We use various time periods, including the 2008 financial crisis, to assess the accuracy of alternative stress test modeling approaches, in particular simple single equation as compared to more complex multiple equation approaches, and in the latter case whether accounting for the correlation between line items has an influence on model results on both an in- and- out-of-sample basis. Our results show potentially inaccuracies in stress test model forecasts, even for models that fit the data exceptionally well in-sample, especially where more complex multi-equation models similar to those used by the Federal Reserve are misspecified and underperform simple models in explanatory power, due to incorrectly accounting for the dependency structure. We find that in the test sample, the 2-equation VAR model for IBTEIA and TAG performs best, and the single equation models for IBTEIA and TAG performs worse, while the single equation AR model for IBTEI has intermediate performance. Our results highlight the public policy need for reconsidering the existent regulations that fail to place limits on the use of regulatory stress tests, and the need for supervisory models to be subject to model validation and governance standards.

Keywords: Stress Testing, CCAR, DFAST, Credit Risk, Financial Crisis, Model Risk, Capital Adequacy, Income before Taxes and Extraordinary Items

JEL Classification: C31, C53, E27, E47, E58, G01, G17, C54, G21, G28, G38.

---

<sup>1</sup> Corresponding author: Michael Jacobs, Jr., Ph.D., CFA, CEO & Managing Partner, Jacobs Risk Advisory Services, Inc., 40 East 89<sup>th</sup> Street, New York, N.Y., 10128, 917-324-2098, [michael@michaeljacobsjr.com](mailto:michael@michaeljacobsjr.com).

# 1 Introduction

Following the financial crisis of the last decade (Acharya (2009), Demircuc-Kunt et al (2010)), regulators have utilized stress testing as a means to which to evaluate the soundness of financial institutions' risk management procedures. Stress testing has been used by supervisors to assess the reliability of credit risk models, as can be seen in the revised Basel framework (Basel Committee for Banking Supervision 2006; 2009 a,b,c,d,e; 1010 a, b) and the Federal Reserve's *Comprehensive Capital Analysis and Review* ("CCAR") program. The success of the CCAR gave rise to a new paradigm of prudential banking supervision based upon forward looking projections of institutions' performance under condition of economic and financial stress, and indeed in many jurisdictions some version of dynamic stress testing forms the basis of large banking regulation. Such stress tests use econometric models to forecast banking financial lines items, assets / liabilities, income and regulatory capital, across multi-period scenarios of economic stress. In this process, banks are mandated to forecast their performance under common regulatory as well as bank specific stress scenarios, and in parallel supervisors validate such by through comparison to supervisor' own stress testing models.

The process of evaluating institutions' stress scenario loss forecasts could potentially form the basis of the imposition remedies on the part of the supervisor. In the U.S., the Dodd-Frank Wall Street Reform and Consumer Protection Act mandates that the Federal Reserve Board ("FRB") supervise annual stress tests ("DFAST") conducted by the largest financial institutions. As a consequence of an institution failing such a stress test, the FRB may require certain remedial measures, such as the prohibition of certain capital distributions repurchases or mandates to enhance modeling methodologies or processes around stress testing. While the FRB utilizes internal estimates of modeled stress losses for benchmarking against institutions results, we note that there exist no strictures in the law around model publishing documentation of model development and testing results. The issue with this is that banks have no protection against being judged unfairly, in the case that the supervisory model's assessment is more pessimistic, but in fact the bank's model is more accurate. The issue with this situation is if a bank's own stress test model projections are significantly more optimistic than those of the FRB. That is, the bank's own estimates could be accurate, and yet there is currently no process to protect the bank from an inaccurate supervisory stress test assessment, and there is no mechanism to control for this type of error. Furthermore, the FRB's DFAST models are meant to estimate a typical bank's performance, and not calibrated to an individual institution's situation, so we have no reason to believe that these aggregated estimates are likely to be more accurate than a more granular model. Therefore, we argue that as with the advanced mathematical, statistical and quantitative techniques that are the primary means of risk measurement and management by banking institutions<sup>2</sup>, which leads to *model risk*<sup>3</sup>, the supervisors themselves should be subject to such guidance and standards within their own models used in stress testing.

---

<sup>2</sup> This is particularly in the field of credit risk (Merton, 1974).

<sup>3</sup> This can be defined as the potential that a model does not sufficiently capture the risks it is used to assess, and the danger that it may underestimate potential risks in the future (Board of Governors of the Federal Reserve System, 2011).

In this paper, we use both the 2008 financial crisis, as well as the last 2 years recent period, in order to measure the predictive accuracy of alternative stress testing model specifications, including those that we believe similar to the models specifications utilized by the U.S. banking supervisors to benchmark institutions own stress testing models. We estimate alternative stress test model specifications and use these estimates to forecast quarterly bank *income before tax and extraordinary items* (“IBTEI”), over a twenty-five year period beginning in the fourth quarter of 2001 and ending in the fourth quarter of 2017, comparing the stress model forecasts to actual bank results. We focus on IBTEI as it is recognized as the most important variable for determining capital adequacy in a dynamic stress test simulation. To that end, this paper shall proceed as follows. Section 2 reviews the supervisory background around stress testing. Section 3 reviews the available academic literature on stress testing. Section 4 reviews the competing econometric methodologies for modeling stressed bank revenues and losses. Section 5 discusses the experimental design and empirical implementation for estimating forecasts of IBTEI. Section 6 presents the data description, a discussion of the estimation results and their implications. Section 7 concludes the study and provides directions for future avenues of research.

## 2 Supervisory Background

Stress testing has been established as the basis of prudential banking supervision by both the U.S. Federal Reserve Board (“FRB”) in the 2010 Dodd-Frank Act, as well as by the European Banking Authority in Article 23 of European Union Regulation No. 1095/2010. We have also seen an endorsement of this paradigm by The Basel Committee on Banking Supervision (“BCBS”) within Basel II and Basel III and in its published supervisory guidance (BCBS, 2009e). The International Monetary Fund (“IMF”) has incorporated stress testing as a required component of its Financial Sector Assessment Program (“FSAP”), which requires member nations to submit to periodic FSAP examinations, as well as in various IMF working papers (Ong et al, 2014). Note that neither the BCBS guidance nor the IMF publications makes reference to specific requirements for model validation or standards of accuracy with respect to the supervisory models, other than some vague claims regarding senior executive oversight of the process (Board of Governors of the Federal Reserve System, (2000, 2016)). We note in this regard that while to some extent the absence of dialogue on model accuracy in the stress testing context stems from the hypothetical nature of the scenarios, which creates challenges in standard backtesting, nevertheless such models are still subject to supervisory standards on the management of model risk (Jacobs, 2013). This issue is of critical importance, since if there is no standard for the supervisors’ own models for stress testing, then it may be the case that banks are being required to hold excess capital in order to compensate for this potential inaccuracy, which could imply a drag on economic activity.

The specification of the supervisory models for stress testing has been kept confidential. The rationale for this is the desire to avoid a situation where banks reverse engineer the process, which could lead to amplified systematic risks. Nevertheless, while not revealing the exact model specification, supervisors have revealed details of its broad methodology for stress testing. We do know that the models are top-down and not loan specific, and also that they are not bank specific, but rather that they model a prototypical or representative bank. Therefore, there is no assumption made regarding the expected bank specific scenarios, only the reaction of various line items to macroeconomic conditions that are projected to prevail (FRB, 2016).

Based upon what is known publicly, the supervisory stress testing models are so-called “bottom-up” approaches, which estimate stressed losses at a line-items (not loan) level basis according to a specified economic scenario, which generates loss rates and income projections for various specified types of loan and business lines, perhaps using pooled proprietary data collected on individual loans and bank positions. In the FRB DFAST model, for each bank balance sheets are segmented into over 12 line items, and gross revenues into 22 components, that are each modelled; however, loan in each category are assumed to grow at a uniform industry average growth rate. It is claimed that the models are continuously re-developed and recalibrated as additional bank data is incorporated in order to improve model performance, but there is no detail around this process disclosed.

Alternatively, the New York FRB estimates a “top-down” model, the so-called CLASS approach (Hirtle et al, 2015). This rather detailed account of the earlier FRB stress testing modeling approaches report results that are supposedly highly aligned to those of the FRB’s DFAST models. Kupiec (2017) estimates a stress testing model that is similar in design to the FRBs CLASS models, albeit specified at a lower level of granularity, decomposing IBTEI into only five segments, net non-provision income-to-assets ratios are into 4 modeled components total bank provisions aggregated. Regarding the latter, the author argues that the 15 component segmentation of the CLASS model is unlikely to lead to superior accuracy in forecasting, as the variation in provisions relative to other components of INTEI are too large to matter. However, similarly to the CLASS models, Kupiec (2017) includes endogenous variables that measure bank risk characteristics. In comparing his approach to the CLASS model, Kupiec (2017) notes that standard measures of model fit are extremely optimistic (e.g., adjusted r-squared values on the order of 90%), which he attributes to a downward bias in standard error estimates due to not correcting for correlations amongst the 200 banks in the sample.

Hirtle et al (2015) compares actual to CLASS regulatory estimated capital projections for the 200 largest banks over the period June 2007 through December 2008, and even though these are in- and not out-of-sample forecasts, the CLASS model is found to be woefully inaccurate: the historical cumulative return over the test period was 0.13, while the CLASS model estimate is -0.05. Furthermore, a cross-sectional regression of the actual on CLASS model predicted bank performance for various stress testing components yielded r-squareds ranging between 0.025 and 0.12.

### **3 Review of the Literature**

Since the dawn of modern risk management in the 1990s, stress testing has been a tool used to address the basic question of how exposures or positions behave under adverse conditions. Traditionally this form of stress testing has been in the domain of sensitivity analysis (e.g., shocks to spreads, prices, volatilities, etc.) or historical scenario analysis (e.g., historical episodes such as Black Monday 1987 or the post-Lehman bankruptcy period; or hypothetical situations such as modern version of the Great Depression or stagflation). These analyses are particularly suited to market risk, where data are plentiful, but for other risk types in data-scarce environments (e.g., operational, credit, reputational or business risk) there is a greater reliance on hypothetical scenario analysis (e.g., natural disasters, computer fraud, litigation events, etc.).

Stress testing first appears in the supervisory realm under the auspices of the Basel I Accord within the 1995 Amendment on Market Risks (BCBS 1988, 1996). The contemporaneous publication of RiskMetrics™ (J.P. Morgan, 1996) established the practice of market risk management as an analytical discipline in its own right and subsumed several stress testing methodologies thereto developed in this context. Jorion (1996) discusses aspects of stress testing in a book addressing Value-at-Risk (“VaR”). Kupiec (1999) and Berkowitz et al (1999) analyze stress testing in a trading and treasury VaR context. Mosser et al. (2001) noted that most stress testing of the period relied upon transparent and easily identifiable historical market factors with respect to asset classes in the trading book.

However, in the case of the banking book (e.g., corporate/C&I or consumer loans), this approach of asset class shocks does not carry over as well, as to the extent these are less marketable there are more idiosyncracies to account for. Therefore, stress testing with respect to credit risk has evolved later and as a separate discipline in the domain of credit portfolio modeling. However, even in the seminal examples of CreditMetrics™ (J.P. Morgan, 1997) and CreditRisk+™ (Wilde, 1997), stress testing was not a component of such models. The commonality of all such credit portfolio models was subsequently demonstrated (Koyluoglu and Hickman, 1998), as well as the correspondence between the state of the economy and the credit loss distribution, and therefore that this framework is naturally amenable to stress testing. In this spirit, a class of models was built upon the CreditMetrics™ (J.P. Morgan, 1997) framework through macroeconomic stress testing on credit portfolios using credit migration matrices (Bangia et al., 2002). Nevertheless, prior to the financial crisis supervisory guidance for stress testing were rather unformed in the banking book as compared to other areas such as interest rate, counterparty or country risk (Board of Governors of the Federal Reserve System 1996, 1999, 2002).

In the decade following the financial crisis there is a great expansion of the literature on stress testing. Foglia (2009) survey the existing credit risk stress testing literature on. Inanoglu and Jacobs, Jr. (2009) address the aggregation of risk types of capital models in the stress testing and sensitivity analysis of economic capital. Jacobs, Jr. (2010) extends Jacobs and Inanoglu (2009) to the validation of models for stressed capital. Schuermann (2014) analyze the predominance of stress testing as a supervisory tool in terms of rationales for its utility, outlines for its execution, as well as guidelines and opinions on disseminating the output under various conditions. Jacobs, Jr. (2013) surveys of practices and supervisory expectations for the stress testing of credit risk in the context of a ratings migration methodology in the CreditMetrics™ framework. Rebonato (2010) proposes a Bayesian casual network model, for stress testing having the capability to cohesively incorporate expert knowledge in the model design and methodology of the stress testing process. . Another recent study features the application of a Bayesian regression model for credit loss implemented using Fed Y9 data, wherein regulated financial institutions report their stress test losses in conjunction with Federal Reserve scenarios, which can formally incorporate exogenous factors such as such supervisory scenarios, and also quantify the uncertainty in model output that results from stochastic model inputs (Jacobs, Jr. et al., 2015). Jacobs (2015) presents an analysis of the impact of asset price bubbles on standard credit risk measures and provides evidence that asset price bubbles are a phenomenon that must be taken into consideration in the proper determination of economic capital for both credit risk management and measurement purposes. The author also calibrates the model to historical equity prices and in in stress testing exercise project credit losses on both baseline and stressed conditions for bubble and non-bubble parameter estimate settings.

Jacobs (2017) extends Jacobs (2015) by performing a sensitivity analysis of the models with respect to key parameters, empirically calibrates the model to a long history of equity prices, and simulates the model under normal and stressed parameter settings. While the author find statistically significant evidence that the historical S&P index exhibits only mild bubble behavior, this translates in underestimation of potential extreme credit losses according to standard measures by an order of magnitude; however, the degree of relative underestimation of risk due to asset price bubbles is significantly attenuated under stressed parameter setting in the model. Finally, Jacobs et al (2018) conducts an empirical experiment using data from regulatory filings and Federal Reserve macroeconomic data released by the regulators in a stress testing exercises, finding that the a Markov Switching model performs better than a standard Vector Autoregressive (VAR) model, both in terms of producing severe scenarios conservative than the VAR model, as well as showing superior predictive accuracy. Jacobs (2018) proposes a challenger approach to industry practice in the machine learning class of models, the Multivariate Adaptive Regression Splines (“MARS”) model. He empirically tests those models using Federal Reserve Y-9 filing and macroeconomic data, gathered and released by the regulators for CCAR purposes, and validates the champion MARS model through a rigorous horse race against the VAR model, and finding it to exhibit greater accuracy in model testing, as well as superior out-of-sample performance, according to various metrics across all modeling segments. Finally, in a paper most similar to this one, Kupiec (2017) uses the 2008 financial crisis to assess the accuracy of alternative stress test modeling approaches, including some that closely resemble Federal Reserve models. Results show large inaccuracies in stress test model forecasts, even for models that fit the data exceptionally well, while complex multi-equation models similar to those used by the Federal Reserve produce the least accurate forecasts. Simple models have less explanatory power within the estimation sample but produce more accurate out-of-sample

## **4 Time Series VAR Methodologies for Estimation and Scenario Generation**

In macroeconomic forecasting, there are 4 basic tasks that we set out to do: characterize macroeconomic time series, conduct forecasts of macroeconomic or related data, make inferences about the structure of the economy, and finally advise policy-makers (Stock and Watson, 2001). In the stress testing application, we are mainly concerned with the forecasting and policy advisory functions, as stressed loss projections help banking risk manager and banking supervisors make decisions about the potential viability of their institutions during periods of extreme economic turmoil. Going back a few decades, these functions were accomplished by a variety of means, ranging from large-scale models featuring the interactions of many variables, to simple univariate relationships motivated by stylized and parsimonious theories (e.g., Okun’s Law or the Phillips Curve). However, following the economic crises of the 1970s, most established economic relationships started to break down and these methods proved themselves to be unreliable. In the early 1980s, a new macro-econometric paradigm started to take hold, VAR, a simple yet flexible way to model and forecast macroeconomic relationships (Sims, 1980). In contrast to the univariate autoregressive model (Box and Jenkins (1970); Brockwell and Davis. (1991); Comman-deur and Koopman (2007)), a VAR model is a multi-equation linear model in which variables can be explained by their own lags, as well as lags of other variables. As in the CCAR / stress testing

application we are interested in modeling the relationship and forecasting multiple macroeconomic variables, the VAR methodology is rather suitable to this end.

Let  $\mathbf{Y}_t = (Y_{1t}, \dots, Y_{kt})^T$  be a  $k$ -dimensional vector valued time series, the output variables of interest, in our application with the entries representing some loss measure in a particular segment, that may be influenced by a set of observable *input variables* denoted by  $\mathbf{X}_t = (X_{1t}, \dots, X_{rt})^T$ , an  $r$ -dimensional vector valued time series also referred as *exogenous variables*, and in our context representing a set of macroeconomic factors. This gives rise to the VARMAX( $p, q, s$ ) (“vector autoregressive-moving average with exogenous variables”) representation:

$$\mathbf{Y}_t \Phi(B) = \mathbf{X}_t \Theta(B) + \mathbf{E}_t \Theta^*(B) \quad (4.1)$$

Which is equivalent to:

$$\mathbf{Y}_t - \sum_{j=1}^p \Phi_j \mathbf{Y}_{t-j} = \sum_{j=0}^s \Theta_j \mathbf{X}_{t-j} + \mathbf{E}_t - \sum_{j=1}^q \Theta_j^* \mathbf{E}_{t-j} \quad (4.2)$$

Where  $\Phi(B) = \mathbf{I}_r - \sum_{j=1}^p \Phi_j B^j$ ,  $\Theta(B) = \sum_{j=0}^s \Theta_j B^j$  and  $\Theta^*(B) = \mathbf{I}_r - \sum_{j=1}^q \Theta_j^* B^j$  are *autoregressive lag polynomials* of respective orders  $p$ ,  $s$  and  $q$ , respectively, and  $B$  is the *back-shift operator* that satisfies  $B^i X_t = X_{t-i}$  for any process  $\{X_t\}$ . It is common to assume that the input process  $\mathbf{X}_t$  is generated independently of the noise process  $\mathbf{E}_t = (E_{1t}, \dots, E_{kt})^T$ <sup>4</sup>. The autoregressive parameter matrices  $\Phi_j$  represent sensitivities of output variables to their own lags and to lags of other output variables, while the corresponding matrices  $\Theta_j$  are model sensitivities of output variables to contemporaneous and lagged values of input variables<sup>5</sup>. It follows that the dependency structure of the output variables  $\mathbf{Y}_t$ , as given by the autocovariance function, is dependent upon the pa-

---

<sup>4</sup> In fact, the exogenous variables  $\{\mathbf{X}_t\}$  can represent both stochastic and non-stochastic (deterministic) variables, examples being sinusoidal seasonal (periodic) functions of time, used to represent the seasonal fluctuations in the output process  $\{\mathbf{Y}_t\}$ , or intervention analysis modeling in which a simple step (or pulse indicator) function taking the values of 0 or 1 to indicate the effect of output due to unusual intervention events in the system.

<sup>5</sup> Note that the VARMAX model (3.1)-(3.2) could be written in various equivalent forms, involving a lower triangular coefficient matrix for  $\mathbf{Y}_t$  at lag zero, or a leading coefficient matrix for  $\mathbf{E}_t$  at lag zero, or even a more general form that contains a leading (non-singular) coefficient matrix for  $\mathbf{Y}_t$  at lag zero that reflects instantaneous links amongst the output variables that are motivated by theoretical considerations (provided that the proper identifiability conditions are satisfied (Hanan (1971), Kohn (1979))). In the econometrics setting, such a model form is usually referred to as a *dynamic simultaneous equations model* or a *dynamic structural equation model*. The related model in the form of equation (4.4), obtained by multiplying the dynamic simultaneous equations model form by the inverse of the lag 0 coefficient matrix, is referred to as the *reduced form model*. In addition, (4.3) has a state space representation of the form (Hanan, 1988).

rameters  $\mathbf{X}_t$ , and hence the correlations amongst the  $\mathbf{Y}_t$  as well as the correlation amongst the  $\mathbf{X}_t$  that depend upon the parameters  $\Theta_j$ . In contrast, in a system of univariate  $ARMAX(p, q, s)$  (“autoregressive-moving average with exogenous variables”) models, the correlations amongst the  $\mathbf{Y}_t$  is not taken into account, hence the parameter vectors  $\Theta_j$  have a diagonal structure (Brockwell and Davis, 1991).

In this study we consider a *vector autoregressive model with exogenous variables* (“VARX”), denoted by  $VARX(p, s)$ , which restricts the *Moving Average* (“MA”) terms beyond lag zero to be zero, or  $\Theta_j^* = \mathbf{0}_{k \times k}$   $j > 0$ :

$$\mathbf{Y}_t - \sum_{j=1}^p \Phi_j \mathbf{Y}_{t-j} = \sum_{j=1}^s \Theta_j \mathbf{X}_{t-j} + \mathbf{E}_t \quad (4.3)$$

The rationale for this restriction is three-fold. First, in MA terms were in no cases significant in the model estimations, so that the data simply does not support a VARMA representation. Second, the VARX model avails us of the very convenient DSE package in R, which has computational and analytical advantages (R Development Core Team, 2018). Finally, the VARX framework is more practical and intuitive than the more elaborate VARMAX model, and allows for superior communication of results to practitioners.

## 5 Experimental Implementation

Bank balance sheet and income statement data submitted to the supervisors tends to have unit roots, hence for stress modeling they are transformed to ratio forms (having assets in the denominator) that are more likely to be stationary, thereby avoiding the issue of spurious regression. On the other hand, asset balances are usually modeled as growth rates, and estimates are convoluted to form projections of ultimate interest, such as the critical IBTEI target variable.

The sample selection process in the stress testing context is usually founded on the basis of fit to a historical sample, there are various constraints that are typically respected. These include intuitiveness of macroeconomic variables, the quality of the scenarios, parsimony as well as out-of-sample performance. Nevertheless, we have observed that in-sample fit tends to be either the primary or amongst the most favored criteria. The reason for this is that the most common model selection algorithms, such as feedforward or backward selection schemes, typically will produce a pool of candidate models with the highest adjusted r-squared models, and from there candidates are filtered based on other criteria. It is rarely the case that all these considerations are considered simultaneously and with optimally calibrated weightings.

This emphasis on in-sample performance has some profound implications for the stressed loss estimation. As we know from the classical variance-bias tradeoff, models that can closely track actuals in a given estimation period run the risk of not fitting well in a test sample, a phenomenon known as overfitting. While the attention paid to factors such as economic intuitiveness of var-



ables and scenarios acts to some extent as a check on this, there still remains this danger due to a preference for models with strong in-sample fit. Overfitting can lead us to include variables with spurious relationships to stress testing target variables, or that have counterintuitive relationships to these in stressed scenarios.

A case study in this issue of overfitting can be readily found in the FRB's stress test models. In that framework, the target variable estimates (income or loss projections) are constructed from a large suite of statistical models corresponding to bank's individual income statement or balance sheet components. As pointed out by Kupiec (2017), combining these independent estimates is not likely to improve the accuracy of the composite estimate of bank IBTEI, but exactly the opposite as model errors compound. This would be true on an in- as well as out-of-sample basis. Furthermore, it might be the case that a single equation model for IBTEI would outperform this multiple equation framework. However, in this research we take this reasoning a step further, as it might be the case that a multiple equation model for components of INTEI (e.g., a VARMAX model), that controls for the correlation in error terms, may have superior performance than either the single or multiple equation approach.

In specifying a model for stress testing, we are simultaneously choosing which variables to include as drivers of income and expense. We may choose macroeconomic variables, which are common to all banks, as well as firm-specific variable, which reflect bank risk profiles and appetites (e.g., the proportion of poorly rated loans in the banking book). While including explanatory variables that encompass bank risk characteristics gives rise to model of greater explanatory panel, and can capture nuances that models with only common factors are unable to, such an approach suffers from endogeneity bias. This is because bank managers will attempt to tune these variables as the macroeconomic environment changes. There are methods to correct for this issue, which Kupiec (2017) does not pursue while using such variables, but for the sake of simplicity we avoid bank-specific variables in this study. In practice, in most bank models where these internal variables are available, they are either kept constant in scenarios projections, or addressed through management overlays. We do not have a good sense of the effect on model performance of ignoring these variables, as they may or may not improve performance in development or test samples, and leave this question for future research.

Ideally we would like to test the performance of our model over a period of interest, such as the Financial Crisis that started in 2008 in the U.S., and motivated the revolution in supervision that made stress testing a centerpiece. Kupiec (2017) is able to do so and have the downturn as an out-of-sample testing period, as his data begins in 1993<sup>6</sup>. In our case, we could only source data starting in 1992, which leaves too short a duration for which to develop viable models.

The financial crisis that began in 2008 creates the ideal setting for testing the accuracy of alternative stress test models. The accuracy of forecasts from alternative model specifications can be compared to the actual performance of banks and the banking system when true macroeconomic conditions are used as the stress scenario. Therefore, we can only evaluate the candidate models over this period on an in-sample basis, and use the most recent 2 years (2016-2017) for out-of-sample evaluation, and in choosing a champion model we consider performance over both

---

<sup>6</sup> The quarterly data from the Statistics on Depository Institutions (SDI) that is publicly available on the Federal Deposit Insurance Corporation's public website starts in the 4<sup>th</sup> quarter of 1991. Kupiec (2011) states that the period begins in March 1993, but we cannot replicate this.

periods (i.e., if two models have similar recent out-of-sample performance, but one performs better over the Financial Crisis, then we will favor the latter). We shall leave out-of-sample testing over a stress testing for future research following the next crisis.

We perform an exercise of modeling a representative institution that is similar to the supervisory model albeit much simplified. The target variable is IBTEI and we estimate this in 3 ways:

- *Separate single-equation* VARMAX models for Total Asset Growth (TAG) and Income Before Taxes and Extraordinary Items to Assets (IBTEIA), with the IBTEI estimate taken as a ratio
- A *two-equation* VARMAX model for Total Asset Growth (TAG) and Income Before Taxes and Extraordinary Items to Assets (IBTEIA), with the IBTEI estimate taken as a ratio
- A *single-equation* VARMAX model for only IBTEI

Kupiec (2017) further decomposes the IBTEI into its components as described in the next section. We could perform a similar exercise as just outlined, incorporating single equations for each as well as a 6 equation model that models that full 6\*6 correlation matrix, but for now leave that exercise for future research.

## 6 Data and Empirical Results

First, we consider a description of the dependent variables in our models of IBTEI, which are IBTEI itself (or normalized by bank assets) and various components thereof, the latter including asset size normalized main income and loss items as well as asset growth. These data are sourced from the *Statistics on Depository Institutions* (“SDI”) report, which is available on the *Federal Deposit Insurance Corporation’s* (“FDIC”) research website<sup>7</sup>. This bank data represents all insured depository institutions in the U.S. and contains income statement, balance sheet and off-balance sheet line items.

In this paper we use quarterly data from the 4<sup>th</sup> quarter of 1991 through the 4<sup>th</sup> quarter of 2017. The models for stress testing are specified and estimated using a development period that ends in the 4<sup>th</sup> quarter of 2017, leaving the last 2 years 2016 and 2017 as an out-of-sample test time period. We also evaluate the models’ performance over the financial crisis period, which we define as the 3<sup>rd</sup> quarter of 2008 through the 2<sup>nd</sup> quarter of 2011. The model development data are aggregate asset-weighted average values bank financial characteristics for each quarter normalized by the total value bank assets in the system.

The list of bank income and balance sheet variables used in the analysis along with sample summary statistics are reported in in Table 6.1. In the general case we have:

- Net Interest Income (“NII”)

---

<sup>7</sup> These are available at <https://www5.fdic.gov/sdi/main.asp?formname=standard>

- Net Non-Interest Income (“NNII”)
- Securities Gains and Losses (“SGL”)
- Provision for Loan and Lease Losses (“PLL”)
- Non-Interest Expense (“NIE”)

Then we have the relation:

$$IBTEI = NII + NNII + SGL - PLL - NIE \quad (6.1)$$

While it is critical that we evaluate the stress test model performance of these components, the ultimate objective is the ability of the combined models to produce accurate in- and out-of-sample estimates of IBTEI. The forecast of IBTEI can be obtained as the product of a IBTEIA estimate, in combination with an estimate of total bank assets derived from a TAG model, and these may be either estimated independently or in a system of equations. Note that even if TAG and IBTEIA model estimates are unbiased, only in the case where they are statistically independent the product of the two estimates will differ from the product of the expected values of the individual series, hence our motivation for considering a correlated system of VAR equations for TAG and IBTEIA.

Second, we consider a description of the macroeconomic explanatory variables in our models of IBTEI. As part of the Federal Reserve's CCAR stress testing exercise, U.S. domiciled top-tier financial institutions are required to submit comprehensive capital plans, including pro forma capital analyses, based on at least one financial institution defined adverse scenario. The adverse scenario is described by quarterly trajectories for key *macroeconomic variables* (“MVs”) over the next nine quarters or for thirteen months to estimate loss allowances. In addition, the Federal Reserve generates its own supervisory stress scenarios, so that firms are expected to apply both financial institution and supervisory stress scenarios to all exposures, in order to estimate potential losses under stressed operating conditions. Firms engaged in significant trading activities (e.g., Goldman Sachs or Morgan Stanley) are asked to estimate a one-time trading-related market and counterparty credit loss shock under their own financial institution scenarios, and a market risk stress scenario provided by the supervisors. Large custodian banks are asked to estimate a potential default of their largest counterparty. Since CCAR is a comprehensive assessment of a firm's capital plan, the financial institutions are asked to conduct an assessment of the expected uses and sources of capital over a planning horizon of nine quarters. The Federal Reserve defines the stress supervisory scenario using 14 MVs:

- Real GDP Growth (“RGDPG”)
- Consumer Price Index (“CPI”)
- Real Disposable Personal Income (“RDPI”)
- Unemployment Rate (“UNEMP”)
- Three-month Treasury Bill Rate (“3MTBR”)
- Ten-year Treasury Bond Rate (“10YTBR”)
- BBB Corporate Rate (“BBBCR”)
- Dow Jones Index (“DJJ”)
- National House Price Index (“HPI”)
- Nominal Disposable Income Growth (“NDPIG”)
- Mortgage Rate (“MR”)
- CBOE’s Market Volatility Index (“VIX”)

- Commercial Real Estate Price Index (“CREPI”)

Our model selection process imposed the following criteria in selecting input and output variables across both multiple VARMAX and univariate ARMAX models<sup>8</sup>:

- Transformations of chosen variables should indicate stationarity
- Signs of coefficient estimates are economically intuitive
- Probability values of coefficient estimates indicate statistical significance at conventional confidence levels
- Residual diagnostics indicate white noise behavior
- Model performance metrics (goodness of fit, risk ranking and cumulative error measures) are within industry accepted thresholds of acceptability Scenarios rank order intuitively (i.e., severely adverse scenario stress losses exceeding scenario base expected losses)

A diverse set of macroeconomic drivers representing varied dimensions of the economic environment, and a sufficient number of drivers) balancing the consideration of avoiding over-fitting) by industry standards (i.e., at least 2-3 and no more than 5-7 independent variables). According to these criteria, we identify the optimal set focusing of 3 of the 9 most commonly used national Fed CCAR MVs as input variables in the ARMAX and VARMAX model: HPI, BBBCR and DJI.

This historical data, 65 quarterly observations from 4Q01 to 4Q17<sup>9</sup>, are summarized in Table 1 in terms of distributional statistics and correlations, as in Figures 1 through 5 of this section. We will discuss main features of key Call Report and FED MVs.. TAG has averaged 1.26% per quarter, reaching a nadir of -4.76%, and peaking at 6.25%, having a high degree of relative variation as shown by the Coefficient of Variation (“CV”) of 1.34. TAG exhibits significant negative skewness and fat-tails, with the respective coefficients measuring -0.62 and 2.56. We can see these features graphically in Figure 1, where we plot both TAG as well as the level of aggregate total assets. IBTEIA has averaged 1.97% per quarter, reaching a nadir of -0.26%, and peaking at 3.99%, having a low degree of relative variation as shown by the CV of 0.54. IBTEIA exhibits insignificant negative skewness and thin-tails, with the respective coefficients measuring 0.36 and -1.20. We can see these features graphically in Figure 2, where we plot both IBTEIA as well as non-normalized IBTEI. Turning to the MVs, we shall discuss the 3 variables that entered our final models for TAG and IBTEIA: HPI, BBBCY and DJI. HPI has averaged 1.88, reaching a nadir of -8.20%, and peaking at 6.90%, having a high degree of relative variation as shown by the CV of 1.27. HPI exhibits significant negative skewness and fat-tails, with the respective coefficients measuring -1.61 and -4.99. We can see these features graphically in Figure 3, where we plot both the historical HPI time series as well as the Fed scenarios for HPI. DJI has averaged 1.48e+4, reaching a nadir of 5.01e+3, and peaking at 2.77e+4, having a low degree of relative variation as shown by the CV of 0.34. DJI exhibits significant positive skewness and thin tails,

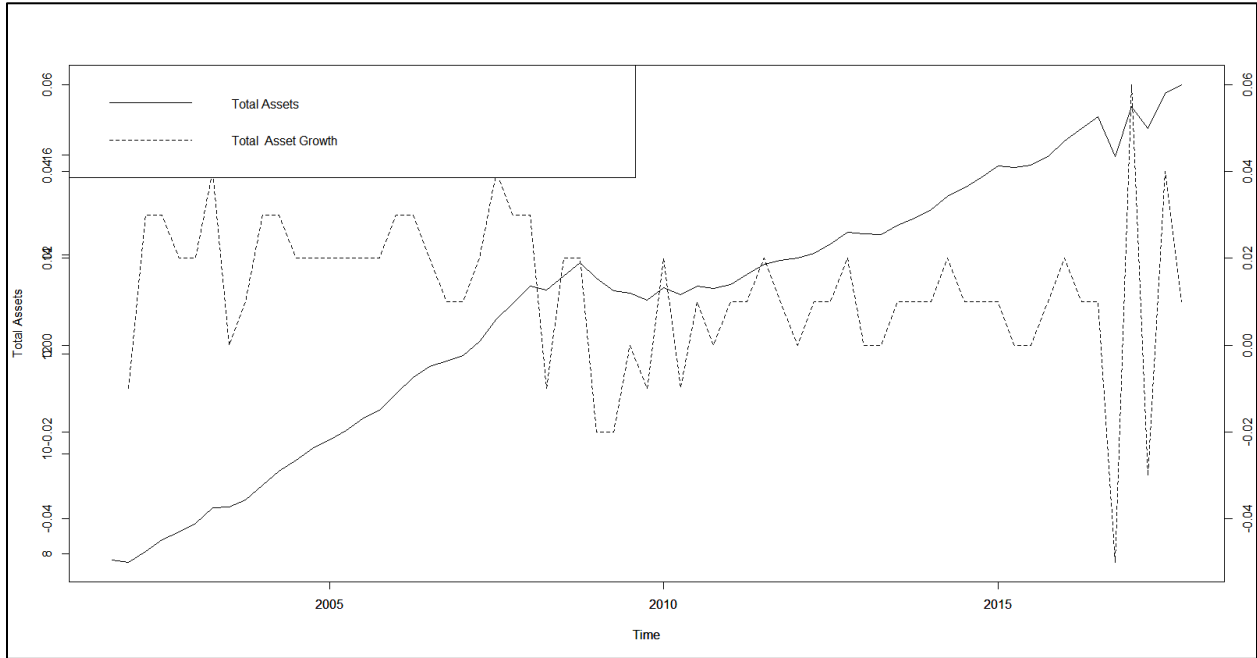
<sup>8</sup> We perform this model selection in an R script designed for this purpose, using the libraries “dse” and “tse” to estimate and evaluate VARMAX and ARMAX models (R Core Development Team, 2016).

<sup>9</sup> We leave out the last 2 years of available data, 1Q16-3Q17, in order to have a holdout sample for testing the accuracy of the models – refer to the Diebold-Mariano tests at the end of this section. We also choose to start our sample in 2001, as we believe that the earlier period would reflect economic conditions not relevant for the last decade, and also because in the financial industry this is a standard starting point for CCAR and DFAST stress testing models.

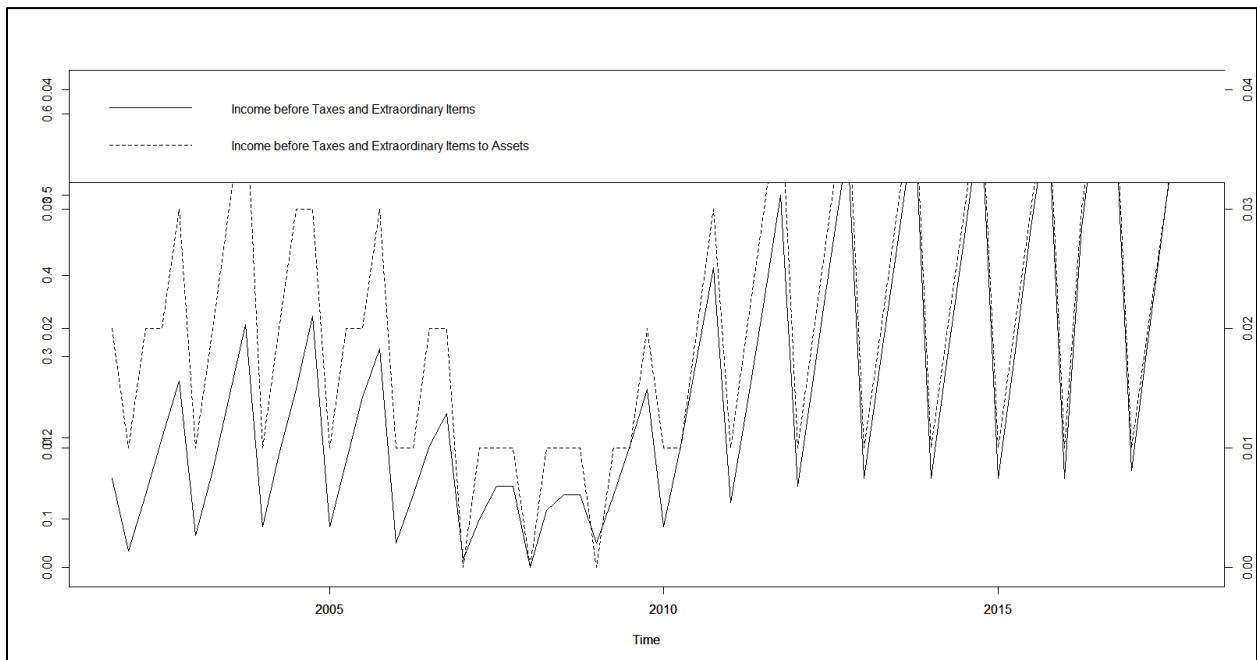
**Table 1: Summary Statistics and Correlations of Fed Macroeconomic Variables and Aggregate Y9 Bank Assets & Earnings**

	Output Variables - Banking Assets, Revenues and Expense										Input Variables - Macroeconomic Factors					
	Total Asset Growth	Income before Taxes and Extraordinary Items	Income before Taxes and Extraordinary Items to Assets	Net Interest Income to Assets	Net Non-Interest Income to Assets	Net Non-Interest Income to Assets	Securities Gains and Losses to Assets	Provision for Loan and Lease Losses to Assets	Non-Interest Expense to Assets	Real Gross Domestic Product Growth	Unemployment Rate	Residential Housing Price Index	Commercial Real Estate Price Index	Dow Jones Equity Price Index	BBB Corporate Bond Yield	CBOE Market Volatility Index
Count	64	65	65	65	65	65	65	65	65	65	65	65	65	65	65	65
Mean	0.0126	0.2607	0.0197	0.0190	0.0119	0.0003	0.0036	0.0078	1.9554	6.31	157.57	201.12	14,851.50	5.54	26.56	
Standard Deviation	0.0170	0.1670	0.0107	0.0083	0.0054	0.0004	0.0038	0.0072	2.3570	1.73	22.67	40.64	5,011.96	1.29	12.09	
Minimum	-0.0476	0.0354	0.0026	0.0057	0.0037	-0.0011	0.0005	0.0007	-8.20	4.10	118.20	138.70	7,773.60	3.70	12.70	
25th Percentile	0.0051	0.1350	0.0096	0.0136	0.0078	0.0001	0.0014	0.0028	1.10	5.00	137.70	167.30	11,144.60	4.50	18.90	
Median	0.0142	0.1959	0.0190	0.0207	0.0121	0.0002	0.0023	0.0048	2.30	5.70	156.90	198.50	13,284.10	5.50	22.70	
75th Percentile	0.0207	0.3527	0.0285	0.0270	0.0159	0.0004	0.0041	0.0110	3.20	7.70	177.80	234.00	19,413.20	6.40	30.80	
Maximum	0.0625	0.6323	0.0399	0.0328	0.0224	0.0014	0.0191	0.0299	6.90	9.90	194.40	278.70	27,673.20	9.40	80.90	
Coefficient of Variation	1.3442	0.6408	0.5453	0.4404	0.4598	1.6773	1.0414	0.9230	1.21	0.27	0.14	0.20	0.34	0.23	0.46	
Skewness	-0.8183	0.8161	0.3612	-0.0935	0.0556	-0.2217	2.1652	1.4064	-1.7243	0.2939	0.1524	0.2092	0.2844	0.7670	1.9495	
Kurtosis	2.5977	-0.4088	-1.1199	-1.3071	-1.0282	3.3698	4.7804	1.4131	5.5464	-0.7194	-1.3148	-1.1247	-0.5469	0.5860	5.5055	
Total Asset Growth																
Income before Taxes and Extraordinary Items	-0.7888															
Income before Taxes and Extraordinary Items to Assets	-0.7444	0.9276														
Net Interest Income to Assets	-0.3892	0.6325	0.7645													
Net Non-Interest Income to Assets	-0.1763	0.4881	0.6807	0.7732												
Securities Gains and Losses to Assets	0.1915	0.4028	0.4631	0.2652												
Provision for Loan and Lease Losses to Assets	-0.2698	-0.3604	-0.3253	-0.4784	0.4408	0.2594										
Non-Interest Expense to Assets	0.4248	-0.2544	-0.2709	0.5091	0.6448	0.2612	0.3451									
Real Gross Domestic Product Growth	0.3334	0.2364	0.3316	0.2515	0.2546	0.4376	-0.5980	0.1695								
Unemployment Rate	-0.6254	-0.4151	-0.3713	-0.2750	-0.2600	-0.2545	-0.3999	-0.3637	-0.1219							
Residential Housing Price Index	0.7266	0.2740	0.8240	0.1104291	0.1839	0.6039	-0.7205	-0.3017	0.1131	-0.7186						
Commercial Real Estate Price Index	0.3166	0.3351	0.2888	0.3209	0.5809	0.5929	-0.7186	-0.2725	-0.1515	-0.5374	0.7873					
Dow Jones Equity Price Index	0.2726	0.5544	0.3098	0.4733	0.2902	0.3441	-0.4209	-0.4335	0.1590	-0.3709	0.6103	0.8223				
BBB Corporate Bond Yield	-0.1215	-0.6187	-0.7010	-0.4490	0.3700	-0.3921	0.5732	0.5767	-0.6119	0.6123	-0.3472	-0.4076	-0.4750			
CBOE Market Volatility Index	-0.3563	-0.2466	-0.2914	-0.3881	-0.3571	0.4379	0.5334	0.2514	-0.6254	0.3846	-0.5256	-0.2187	-0.4643	-0.5256		

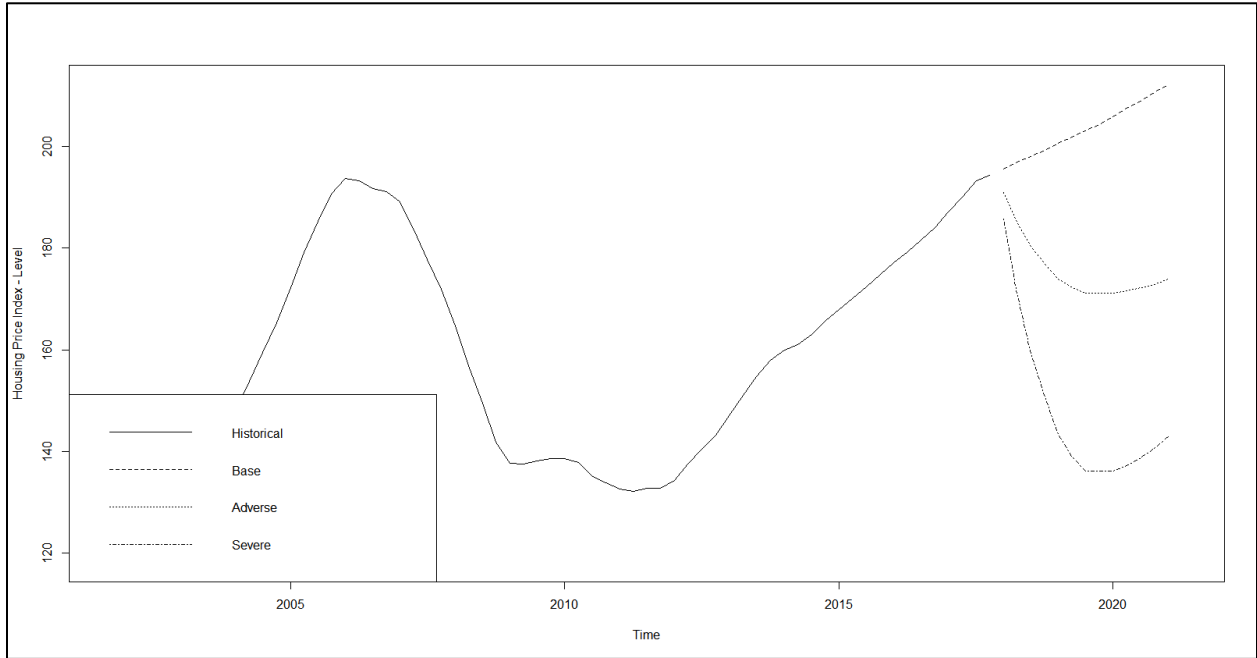
**Figure 1. Time Series Plots – Total Assets and Total Asset Growth (Call Report Data - Aggregate of Depository Institutions in the U.S.)**



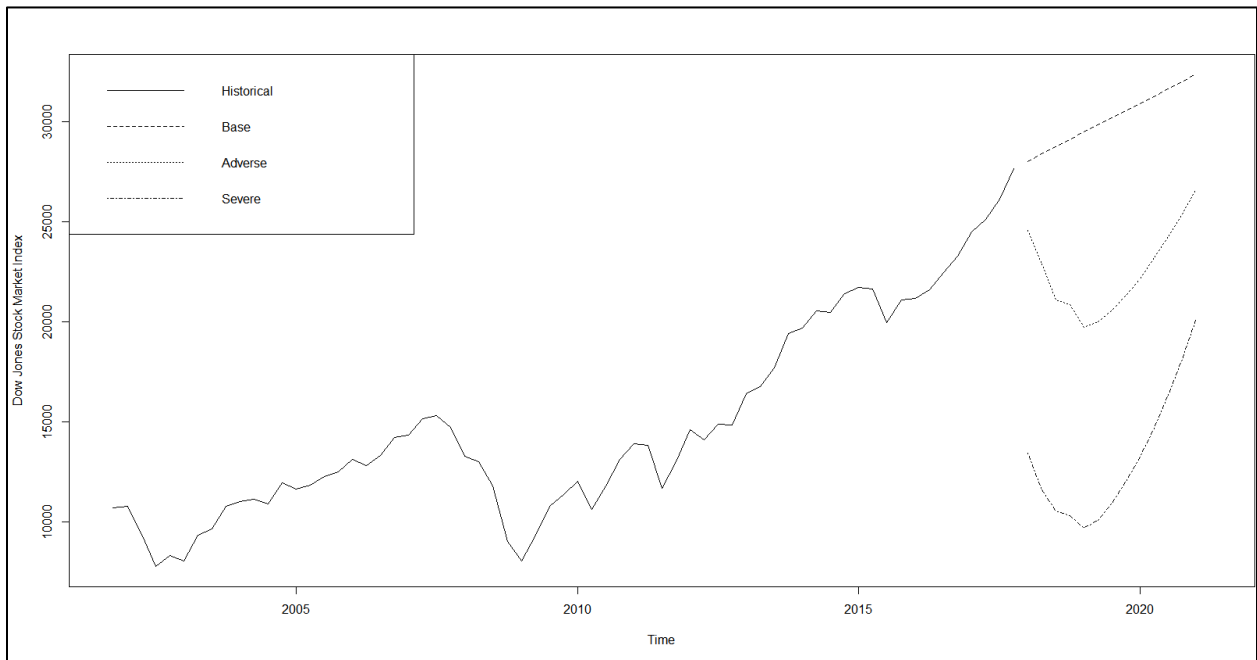
**Figure 2. Time Series Plots – Income before Taxes and Extraordinary Items to Assets & Income before Taxes and Extraordinary Items (Call Report Data - Aggregate of Depository Institutions in the U.S.)**



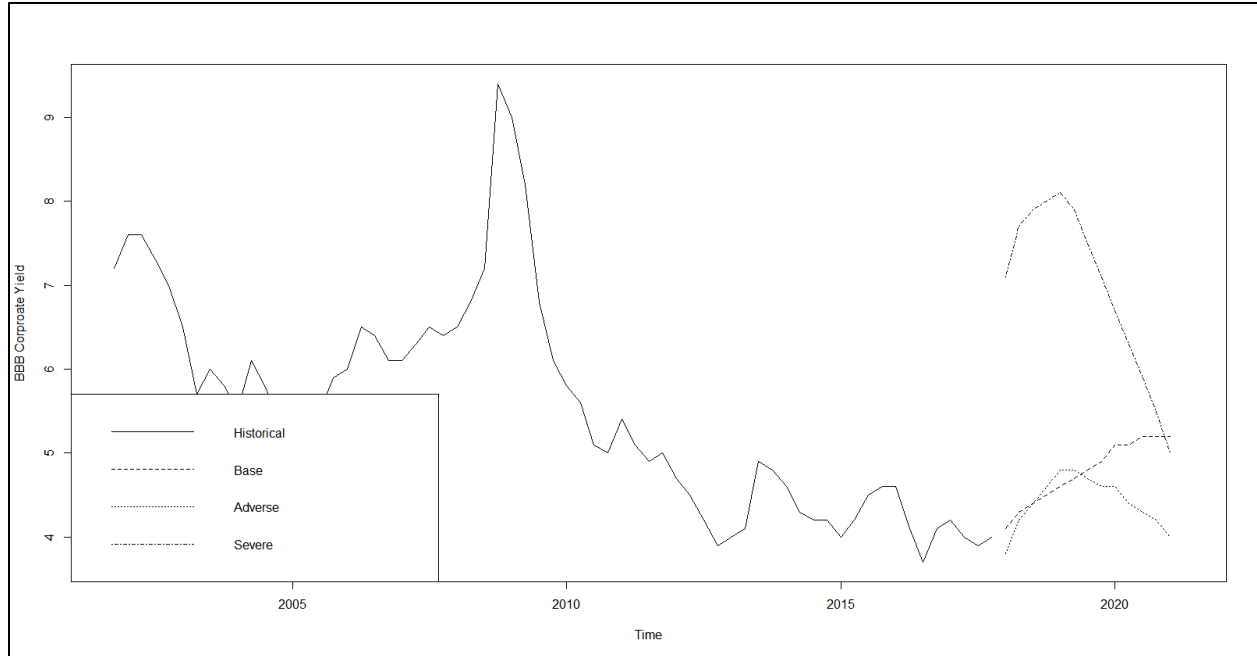
**Figure 3. Time Series Plots – Commercial Real Estate Price Index (History and Fed Scenarios)**



**Figure 4. Time Series Plots – Dow Jones Equity Price Index (History and Fed Scenarios)**



**Figure 5. Time Series Plots – BBB Corporate Bond Yield (History and Fed Scenarios)**

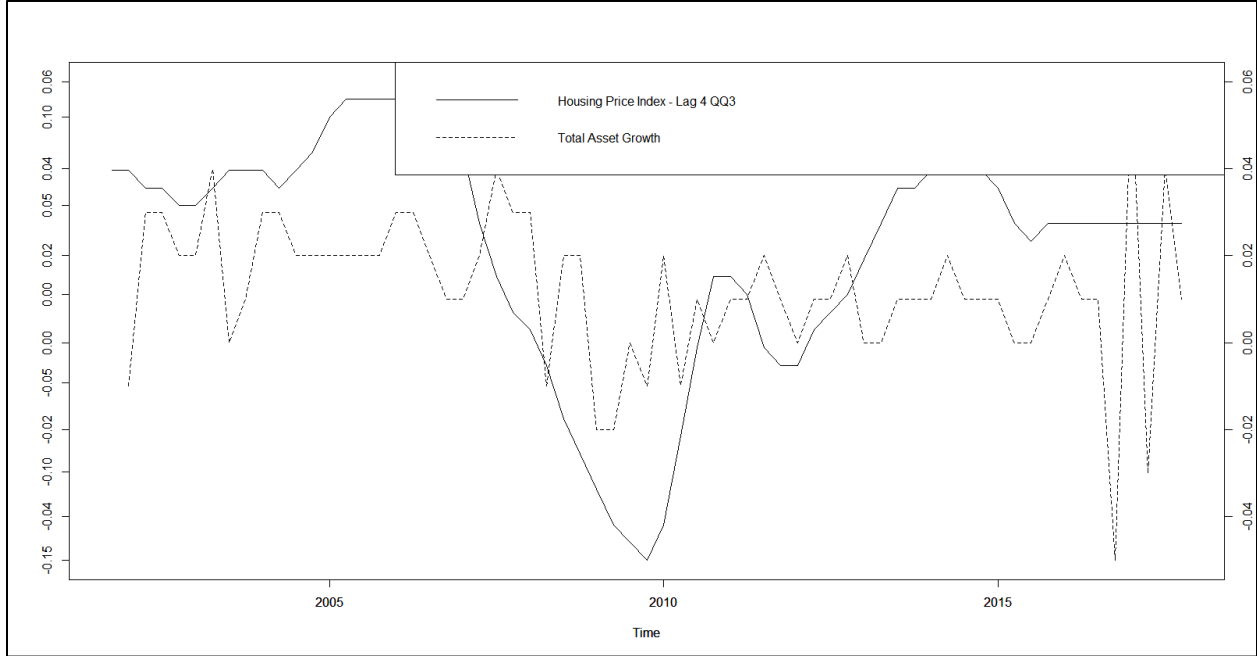


with the respective coefficients measuring 0.78 and -0.35. We can see these features graphically in Figure 4, where we plot both the historical DJI time series as well as the Fed scenarios for DJI. BBBCY has averaged 5.54%, reaching a nadir of 3.70%, and peaking at 9.40, having a low degree of relative variation as shown by the CV of 0.23. BBBCY exhibits significant positive skewness and fat tails, with the respective coefficients measuring 0.77 and -0.39. We can see these features graphically in Figure 4, where we plot both the historical BBBCY time series as well as the Fed scenarios for BBBCY.

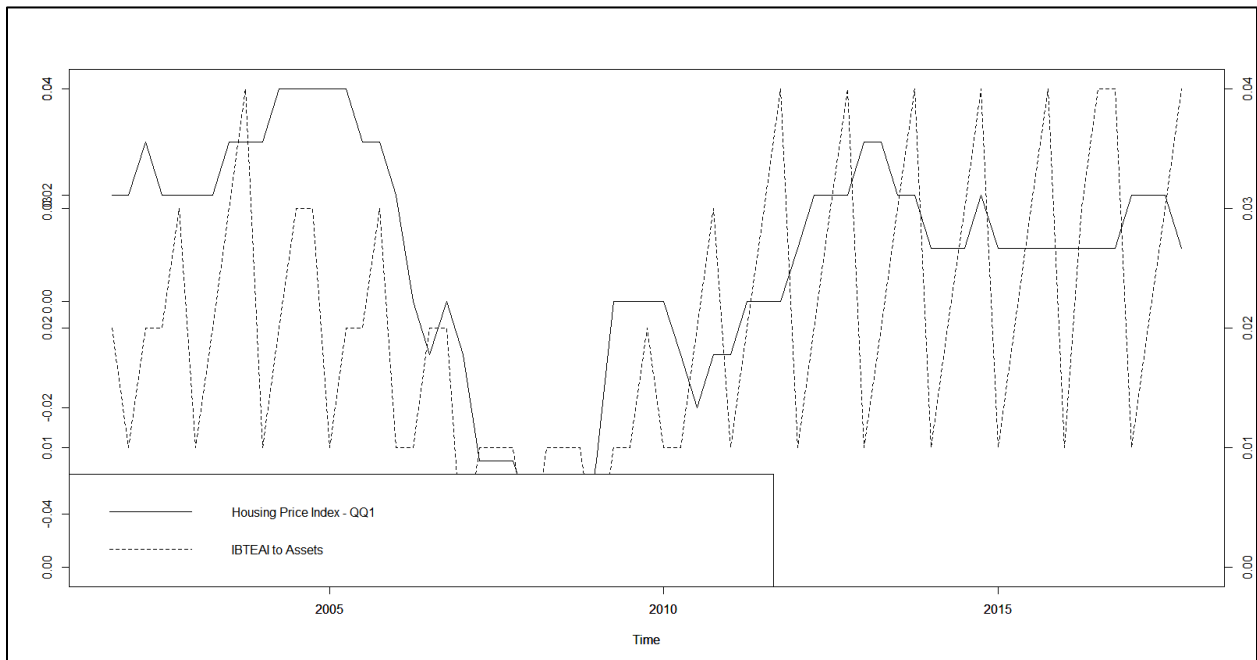
The correlations amongst all of the independent and dependent variables, in both their level and percentage change forms, are displayed in the bottom panel of Table 1. First, we will describe main features of the dependency structure within the group of input macroeconomic variables, then the same for the output loss rate variables, and finally the cross-correlations between these two groups. We observe that all correlations have intuitive signs and magnitudes that suggest significant relationships, although the latter are not large enough to suggest any issues with multicollinearity. The correlation matrix amongst the macroeconomic variables appear in the lower right quadrant of the bottom panel of Table 1. For example, considering some of the stronger relationships, the correlations between BBBCY / DJI , CREPI / DJI, CREPI / RESI are -75.9%, 82.2% and 78.7%, respectively. The correlation matrix amongst the Y9 earnings and revenue variables appear in the upper left quadrant of the bottom panel of Table 1. For example, considering some of the stronger relationships, the correlations between IBTEIA / IBTEI, IBTEIA / TAG and NNIIA / IIA are 92.8%, -74.4% and -77.3%, respectively. The correlation matrix



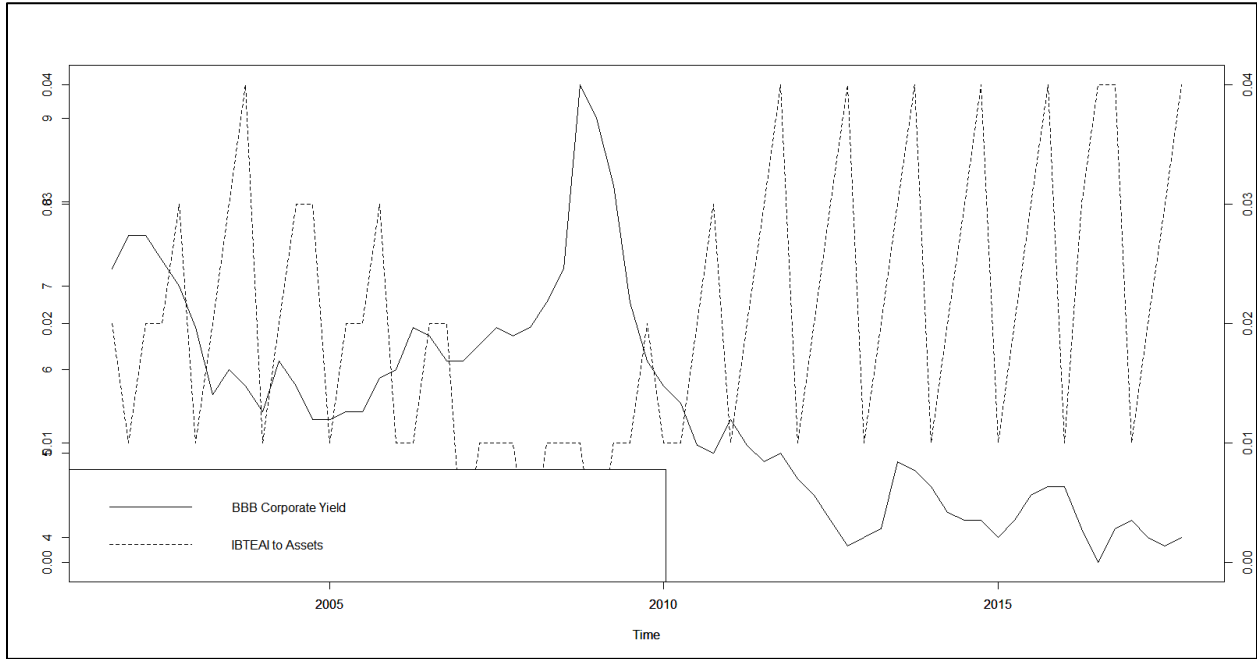
**Figure 6. Time Series Bi-plots – Residential Housing Price Index and Total Asset Growth**



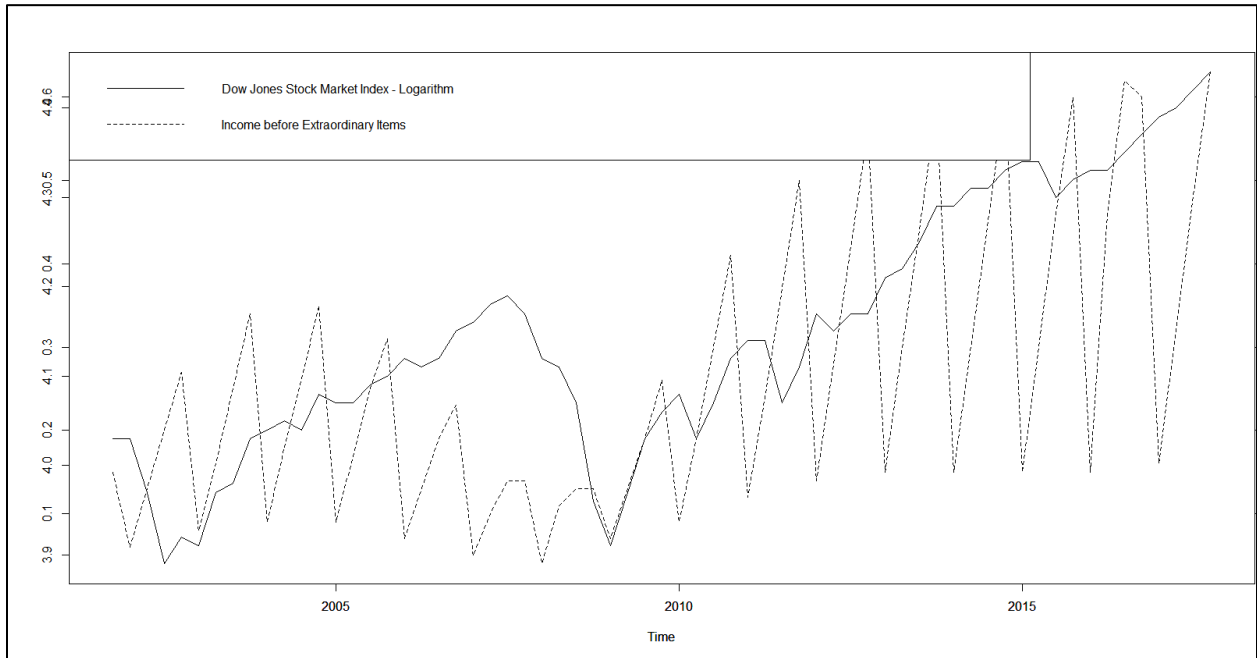
**Figure 7. Time Series Bi-plots – Residential Housing Price Index and Income before Taxes and Extraordinary Items to Assets**



**Figure 8. Time Series Bi-plots – BBB Corporate Bond Yield and Income before Taxes and Extraordinary Items to Assets**



**Figure 9. Time Series Bi-plots – Dow Jones Equity Price Index and Income before Taxes and Extraordinary Items**



amongst the Y9 earnings and revenue and macroeconomic variables appear in the lower left quadrant of the bottom panel of Table 1. For example, considering some of the stronger relationships, the correlations between TAG / RESI, IBTEIA / RESI, IBTEIA / BBBCY and IBTEIA / DJI are 89.7%, 82.4%, -70.1% and 55.4%, respectively. In Figures 6 through 9 we present time series bi-plots that illustrate the relationship amongst these latter variables discussed in Table 1.

The maximum likelihood estimation results for our set of models are shown in Table 2, ARMAX single equation models for IBTAEIA, TAG and IBTAEI; and a VARMAX two equation model for IBTAEIA and TAG. All of the models are convergent, coefficient estimates highly statistically significant as well as having economically intuitive signs. We may note further that magnitudes of coefficient estimates indicate strong sensitivity of target variables to input MV drivers, but that the degree of persistence as measured by the estimated autoregressive coefficients indicate rather mild persistence in the dependent variables. The values of the log-likelihood ratio statistics, far lower for the VARMAX than the ARMAX models for TAG and IBTEIA, indicate that we cannot support statistically the parameter restrictions of the single equation ARMAX models, which makes sense as we not the high correlation between these dependent variables.

In Table 3 we tabulate the main results of this study, which are shown graphically in Figures 10 through 13, the calculation of a range of model performance metrics:

- Generalized Cross-Validation (“GCV”)
- Squared Correlation (“SC”)
- Root Mean Squared Error (“RMSE”)
- Cumulative Percentage Error (“CPE”)
- Aikaike Information Criterion (“AIC”)

We measure these quantities over four periods of importance:

- Development Sample (4Q91-4Q15)
- Test Sample (1Q16-4Q18)
- Full Sample (4Q91-4Q17)
- Downturn Sample (3Q08-2Q11)

Focusing upon the prediction of IBTEIA, we can make several observations that hold consistently across model performance. In the development sample, the single equation AR models for TAG and IBTAI perform best, followed by the 2-equation VaR model, and the single equation IBTEI AR model performs worse; and the margin of outperformance of the 2 equation AR over the VAR model is slimmer than the outperformance of the latter to the single equation AR model. However, in the test sample, we observe that the 2-equation VAR model for IBTEIA and TAG performs best, and the single equation models for IBTEIA and TAG performs worse, while the single equation AR model for IBTEI has intermediate performance. Over the full sample, the relative

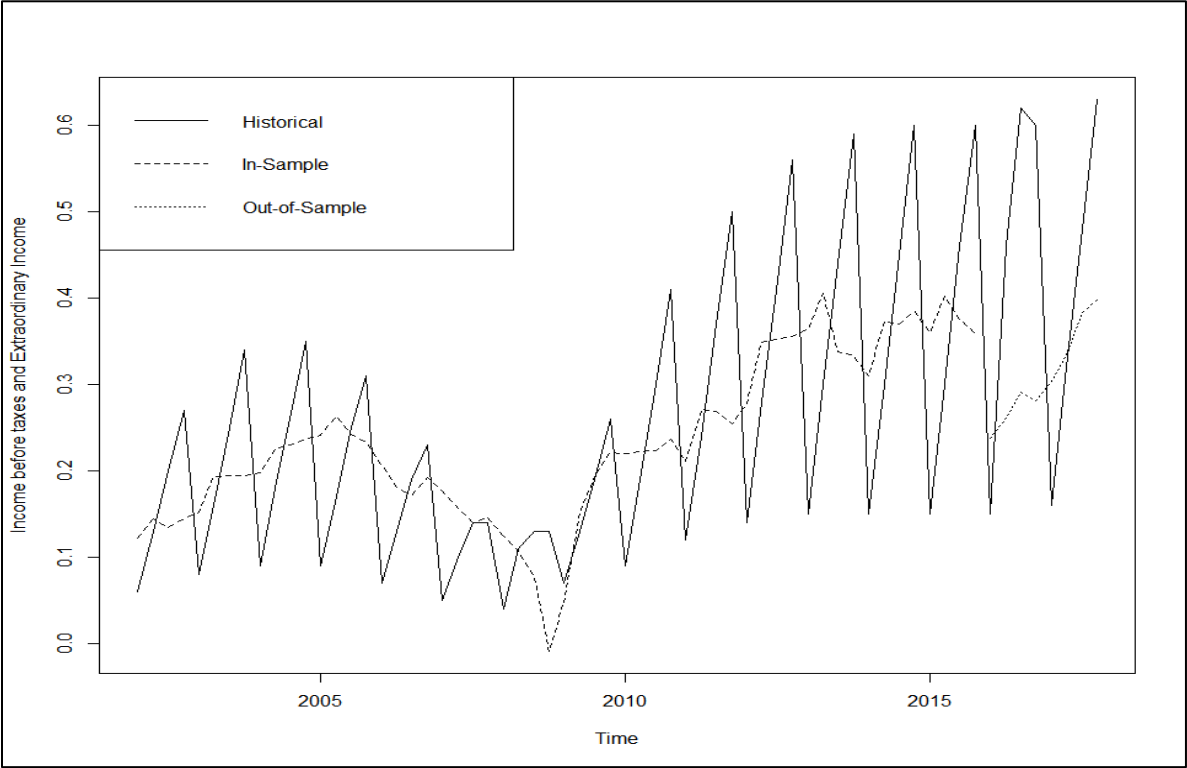
**Table 2: Vector Autoregressive vs. Single Equation Autoregressive Maximum Likelihood Model Estimation Results (Fed Macroeconomic Variables and Aggregate Y9 Bank Assets & Earnings)**

Model	Statistic	Coefficient Estimates			P-Values			
		Income before Taxes and Extraordinary Items to Assets	Total Asset Growth	Income before Taxes and Extraordinary Items	Income before Taxes and Extraordinary Items to Assets	Total Asset Growth	Income before Taxes and Extraordinary Items	
Multiple Vector AR(1) Model	Output Variables - Balance Sheet Items	Income before Taxes and Extraordinary Items to Assets	-0.1079	-0.0222		0.0689	0.0368	
		Total Asset Growth	-0.1243	0.0378		-0.0124	0.00E+00	
	Input Variables - Macroeconomic Factors	Housing Price Index -1	0.0032	0.0014		0.00E+00	4.88E-06	
		Housing Price Index -2	0.1603	0.0773		6.16E-04	2.77E-02	
		BBB Corporate Yield	-0.0050	0.0936		3.82E-02	4.93E-06	
		Dow Jones Equity Price Index						
Log-Likelihood		-352.63						
Single Equation AR(1) Models	Output Variables - Balance Sheet Items	Income before Taxes and Extraordinary Items to Assets	0.0384			0.0000		
		Total Asset Growth		8.86E-03			1.62E-08	
		Income before Taxes and Extraordinary Items			0.0919			0.0000
	Input Variables - Macroeconomic Factors	Housing Price Index	0.1610	0.0720		4.89E-04	4.12E-04	
		BBB Corporate Yield	-0.0384		-0.0606	0.0000		0.00E+00
		Dow Jones Equity Price Index			0.1194			0.00E+00
Log-Likelihood		-186.9326	-168.6210	-39.5529				

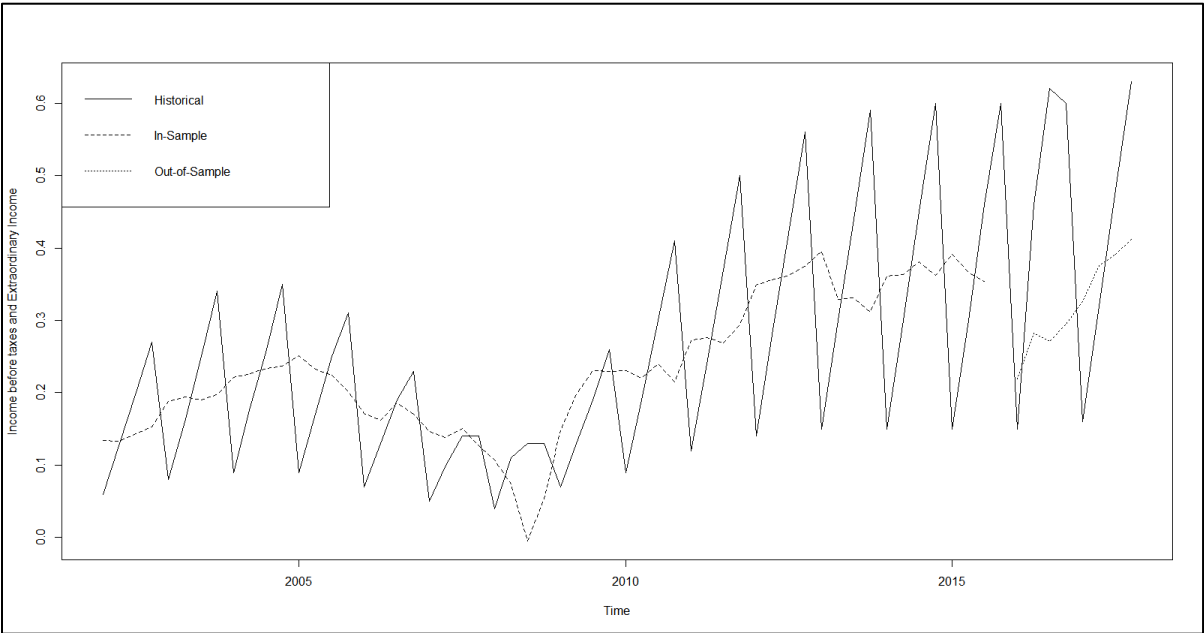
**Table 3: Vector Autoregressive vs. Single Equation Autoregressive Maximum Likelihood Model Performance Metrics (Fed Macroeconomic Variables and Aggregate Y9 Bank Assets & Earnings)**

Time Period	Statistic	Model						
		IBTEIA - Single Equation AR	TAG - Single Equation AR	IBTEI - Single Equation AR for IBTEIA & TAG (Ratio)	IBTEIA - Multiple Equation VAR	TAG - Multiple Equation VAR	IBTEI - Multiple Equation VAR for IBTEIA & TAG (Ratio)	IBTEI - Single Equation AR
Development Sample	Generalized Cross Validation	2.65E-08	4.96E-08	4.33E-06	2.80E-07	4.54E-08	4.34E-05	4.66E-05
	Squared Correlation	0.3016	0.2314	0.4179	0.3896	0.2564	0.5417	0.3070
	Root Mean Squared Error	0.0094	0.0124	0.1224	0.1038	0.0132	1.2919	0.1881
	Cumulative Percentage Error	3.48E-05	-1.29E-08	3.48E-05	7.17E-05	-5.81E-04	-9.94E-03	1.17E-04
	Aikaik Information Criterion	-297.17	-259.91	-16.59	-13.78	-248.59	-0.34	-14.47
Test Sample	Generalized Cross Validation	4.73E-06	2.36E-05	8.71E-05	4.35E-06	2.42E-05	8.43E-05	4.53E-04
	Squared Correlation	0.2500	0.0382	0.1979	0.3265	0.0009	0.4562	0.2588
	Root Mean Squared Error	0.0215	0.0430	0.2921	0.1652	0.0974	2.2995	0.1885
	Cumulative Percentage Error	-3.98E-01	0.8209	1.36E-02	3.23E-03	1.13E+00	-2.38E-03	6.45E-02
	Aikaik Information Criterion	-42.33	-31.47	-2.70	-8.10	-13.26	4.54	-5.64
Full Sample	Generalized Cross Validation	2.21E-08	6.31E-08	3.93E-08	6.86E-08	5.90E-08	1.15E-05	8.13E-08
	Squared Correlation	0.3005	0.1627	0.4444	0.3999	0.2169	0.5902	0.3552
	Root Mean Squared Error	0.0098	0.0162	0.1323	0.1057	0.0170	1.3051	0.1998
	Cumulative Percentage Error	9.02E-07	1.21E-07	9.92E-07	1.15E-06	1.24E-06	-4.51E-07	8.68E-06
	Aikaik Information Criterion	-325.95	-258.52	-13.82	-15.27	-244.80	0.61	-13.15
Downturn Sample	Generalized Cross Validation	3.34E-07	2.36E-06	6.80E-07	5.38E-08	2.11E-06	8.53E-06	3.73E-06
	Squared Correlation	0.3867	0.0634	0.2081	0.3113	2.98E-06	0.3869	0.1493
	Root Mean Squared Error	0.0078	0.0195	0.1405	0.0838	0.0553	0.4722	0.1665
	Cumulative Percentage Error	4.49E-02	8.17E-01	-6.03E-03	-1.20E-03	1.11E+00	-1.37E-03	-1.18E-02
	Aikaik Information Criterion	-83.57	-62.13	-2.76	-6.19	-45.46	-1.11	-1.98

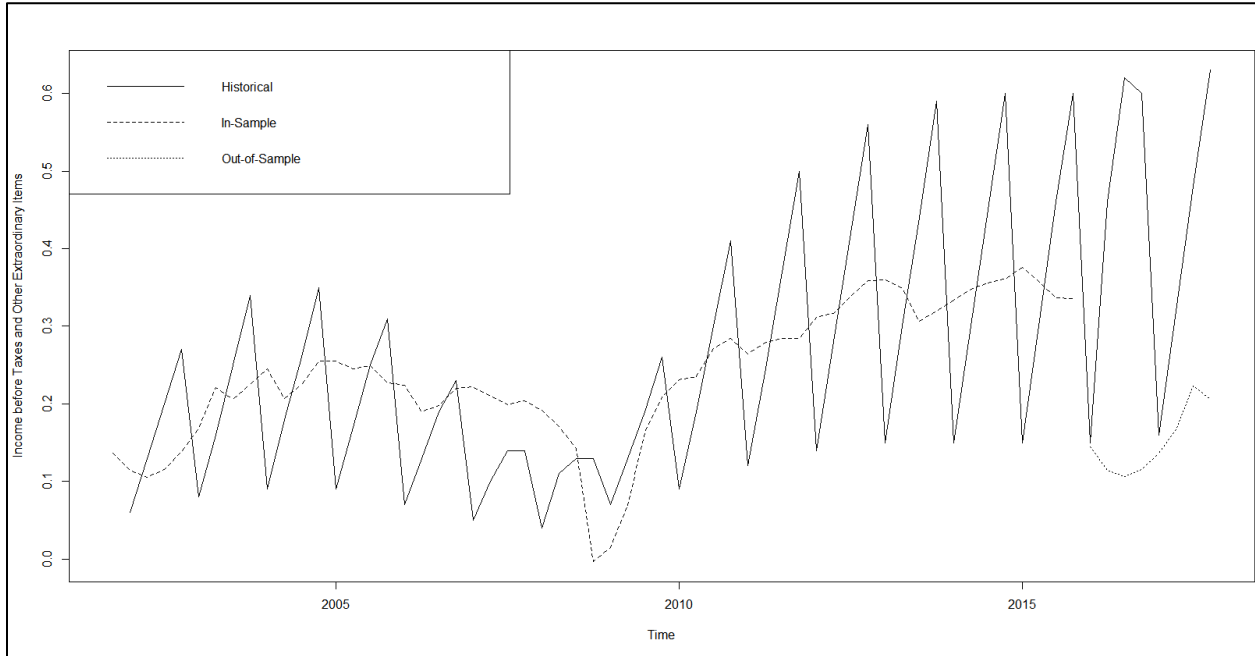
**Figure 10. Econometric Model Accuracy Plot for IBTEI – Single Equation AR Models for IBTEIA and TAG**



**Figure 11. Econometric Model Accuracy Plot for IBTEI – Two Equation VAR Models for IBTEIA and TAG**



**Figure 12. Econometric Model Accuracy Plot for IBTEI – Single Equation AR Models for IBTEI**



performance of the models is the same as over the development sample, although we observe that the VAR models shows slight increase in performance relative to the development sample, while the AR models degrade somewhat in performance relative to the development sample. Finally, over the downturn period, the VAR models also performs best, although now the single equation AR models is inferior to the 2 equation version, and while all measures deteriorate, the VAR model hold up best in this period.

## 6 Conclusion and Future Directions

In this study we have investigate a modeling framework for stress testing that we believe to be very close to that employed by the regulators, which projects various financial statement line items for an aggregated “average” bank. Using various time periods, including the 2008 financial crisis as well as the last 2 years recent period, we have assessed the accuracy of alternative stress test modeling approaches, in particular simple single equation as compared to more complex multiple equation approaches, and in the latter case whether accounting for the correlation between line items has an influence on model results on both an in- and- out-of-sample basis. We have estimated alternative stress test model specifications and use these estimates to forecast quarterly bank IBTEI, over a twenty-five year period beginning in the fourth quarter of 2001 and ending in the fourth quarter of 2017, comparing the stress model forecasts to actual bank results. We focused on IBTEI as it is recognized as the most important variable for determining capital adequacy in a dynamic stress test simulation. Our results have shown potentially inaccuracies in stress test model forecasts, even for models that fit the data exceptionally well in-sample, especially where

more complex multi-equation models similar to those used by the Federal Reserve are misspecified and underperform simple models in explanatory power, due to incorrectly accounting for the dependency structure. We have found that in the test sample, the 2-equation VAR model for IBTEIA and TAG performs best, and the single equation models for IBTEIA and TAG performs worse, while the single equation AR model for IBTEI has intermediate performance. Our results have highlighted the public policy need for reconsidering the existent regulations that fail to place limits on the use of regulatory stress tests, and the need for supervisory models to be subject to model validation and governance standards.

There are several directions in which this line of research could be extended, including but not limited to the following:

- More granular classes of credit risk models, such as ratings migration or PD / LGD scorecard / regression
- Alternative data-sets, for example bank or loan level data
- More general classes of regression model, such as logistic or semi-parametric
- Applications related to stress testing, such as regulatory or economic capital

## 6 References

- Acharya, V. V., & Schnabl, P. (2009). How banks played the leverage game. In V.V. Acharya, & M. Richardson (Eds.), *Restoring Financial Stability* (pp. 44-78). New York, N.Y.: Wiley Finance.
- Bangia, A. Diebold, F.X., Kronimus, A., Schagen, C., & Schuermann, T. (2002) Ratings migration and the business cycle, with application to credit portfolio stress testing. *Journal of Banking and Finance*, 26(2), 445- 474.
- BANK FOR INTERNATIONAL SETTLEMENTS. "Basel committee on banking supervision (BCBS)." *Internal convergence of capital measurement and capital standards:: a revised framework* (1988).
- \_\_\_\_\_. *Amendment to the capital accord to incorporate market risks* (1996).
- \_\_\_\_\_. "The committee on global financial systems." *Stress testing by large financial institutions:: current practice and aggregation issues* (April, 2000).
- \_\_\_\_\_. *International convergence of capital measurement and capital standards:: a revised framework* (2006).
- \_\_\_\_\_. *Principles for sound stress testing practices and supervision - consultative paper* (May – No. 155) (2009a).
- \_\_\_\_\_. *Guidelines for computing capital for incremental risk in the trading book* (2009b).
- \_\_\_\_\_. *Revisions to the Basel II market risk framework* (2009c).
- \_\_\_\_\_. *Analysis of the trading book quantitative impact study* (October) (2009d).
- \_\_\_\_\_. *Strengthening the resilience of the banking sector - consultative document* (De-

- \_\_\_\_\_ (*cember*) (2009e).
- \_\_\_\_\_ *Basel III:: a global regulatory framework for more resilient banks and banking systems (December)* (2010a).
- \_\_\_\_\_ *An assessment of the long-term economic impact of stronger capital and liquidity requirements (August)* (2010b).
- Berkowitz, J., & Jeremy, A. (1999). Coherent framework for stress-testing, *U.S. Board of Governors of the Federal Reserve System Working Paper, No. 99-29*.
- Board of Governors of the Federal Reserve System, 1996, “Joint Policy Statement on Interest Rate Risk”, Supervisory Letter 96-13, May 23<sup>rd</sup>.
- \_\_\_\_\_, 1999, “Supervisory Guidance Regarding Counterparty Credit Risk Management”, Supervisory Letter 99-03, February 1<sup>st</sup>.
- \_\_\_\_\_, 2002, “Interagency Guidance on Country Risk Management”, Supervisory Letter 02-05, March 8<sup>th</sup>.
- \_\_\_\_\_, 2011, “Supervisory Guidance on Model Risk Management”, Supervisory Letter 11-7, April 4<sup>th</sup>.
- \_\_\_\_\_, 2016, “Dodd-Frank Act Stress Test 2016: Supervisory Stress Test Methodology and Results”, Washington D.C.
- Box, G., & Jenkins, G. (1970). *Times series analysis: forecasting and control*. San Francisco,CA: Holden-Day.
- Brockwell P.J., & Davis R.A. (1991). *Time series: theory and methods*. New York, N.Y.: Springer-Verlag.
- Commandeur, J. J. F., & Koopman, S.J. (2007). *Introduction to state space time series analysis*. New York, N.Y: Oxford University Press.
- Demirguc-Kunt, A., Detragiache, E., & Merrouche, O. (2010). Bank capital: lessons from the financial crisis, *World Bank Policy Research Working Paper No. 5473*.
- Foglia, A. (2009). Stress testing credit risk: a survey of authorities’ approaches. *International Journal of Central Banking*, 5(3), 9-37.
- Hanan, E.J. (1971). The identification problem for equation systems with moving average errors. *Econometrica*, 39, 751-766.
- \_\_\_\_\_, (1988). *The statistical theory of linear systems*. John Wiley, N.Y.: New York.
- Hirtle, B. A. Kovner, J. Vickery, & Bhanot, M. (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)*.
- Inanoglu, H., & Jacobs, Jr., M. (2009). Models for risk aggregation and sensitivity analysis: an application to bank economic capital, *The Journal of Risk and Financial Management*, 2(1), 118-189.
- Jacobs, Jr., M. (2010). Validation of economic capital models: state of the practice, supervisory expectations and results from a bank study, *Journal of Risk Management in Financial Institutions*, 3(4), 334-365.
- \_\_\_\_\_. (2013). Stress testing credit risk portfolios. *Journal of Financial Transformation*, 37,



53-75.

- \_\_\_\_\_, Karagozoglu, A.K., & Sensenbrenner, F.J. (2015), Stress testing and model validation: application of the Bayesian approach to a credit risk portfolio. *The Journal of Risk Model Validation*, 9(3), 1-70.
- \_\_\_\_\_. (2015). The impact of asset price bubbles on credit risk measures. *The Journal of Financial Risk Management*, 4, 251–266.
- \_\_\_\_\_. (2017). Asset price bubbles and the quantification of credit risk capital with sensitivity analysis, empirical implementation and an application to stress testing. *The Journal of Risk Model Validation*, 11(4), 1–35.
- \_\_\_\_\_, & Sensenbrenner, F.J. (2018). A comparison of methodologies in the stress testing of credit risk – alternative scenario and dependency constructs. *Quantitative Finance and Economics*, 2(2), 294–324.
- \_\_\_\_\_. (2018). The validation of machine learning models for the stress testing of credit risk. *The Journal of Risk Management in Financial Institutions*, Forthcoming.
- J.P. Morgan. (1996). *RiskMetrics<sup>TM</sup> (Fourth Edition)*. New York, N.Y.: J.P. Morgan.
- \_\_\_\_\_. (1997). *CreditMetrics<sup>TM</sup> (First Edition)*. New York, N.Y.: J.P. Morgan.
- Jorion, P. (1996). Risk2: measuring the value at risk. *Financial Analyst Journal*, 52(6), 47-56.
- Kohn, R. (1979). Asymptotic results for ARMAX structures. *Econometrica*, 47, 1295–1304.
- Koyluoglu, H., & Hickman, A. (1998). Reconcilable differences. *Risk (October)*, 56-62.
- Kupiec, P. (1999). Risk capital and VaR. *The Journal of Derivatives*, 7(2), 41-52.
- \_\_\_\_\_. (2017). Inside the black box: the accuracy of alternative stress test models. *Working Paper, March*.
- Merton, R. (1974). On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance*, 29(2), 449–470.
- Mosser, P.C., Fender, I., & Gibson, M.S. (2001). An international survey of stress tests. *Current Issues in Economics and Finance*, 7(10), 1-6.
- Ong, L.L. (2014). A guide to IMF stress testing: methods and models. Washington D.C.: International Monetary Fund.
- R Development Core Team, 2018: “R: A Language and Environment for Statistical Computing.” R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
- Rebonato, R. (2010). *Coherent stress testing: a Bayesian approach*. New York, N.Y.: Wiley.
- Schuermann, T. (2014). Stress testing banks. *International Journal of Forecasting*, 30(3), 717-728.
- Sims, C.A. (1980). Macroeconomics and reality. *Econometrica*, 48, 1-48.
- Stock, J.H., & Watson, M.W. (2001). Vector autoregressions. *Journal of Economic Perspectives*, 15(4), 101-115.
- Wilde, T. (1997). *CreditRisk+: a credit risk management framework*. London, U.K.: Credit Suisse First Boston.