

An Analysis of the Impact of Modeling Assumptions in the Current Expected Credit Loss (CECL) Framework on the Provisioning for Credit Loss

Michael Jacobs, Jr.¹

Abstract

The CECL revised accounting standard for credit loss provisioning is intended to represent a forward-looking and proactive methodology that is conditioned on expectations of the economic cycle. In this study we analyze the impact of several modeling assumptions - such as the methodology for projecting expected paths of macroeconomic variables, incorporation of bank-specific variables or the choice of macroeconomic variables – upon characteristics of loan loss provisions, such as the degree of pro-cyclicality. We investigate a modeling framework that we believe to be very close to those being contemplated by institutions, which projects various financial statement line items, for an aggregated “average” bank using FDIC Call Report data. We assess the accuracy of 14 alternative CECL modeling approaches. A key finding is that assuming that we are at the end of an economic expansion, there is evidence that provisions under CECL will generally be no less procyclical compared to the current incurred loss standard. While all the loss prediction specifications perform similarly and well by industry standards in-sample, out of sample all models perform poorly in terms of model fit, and also exhibit extreme under-prediction. Among all scenario generation models, we find the regime switching scenario generation model to perform best across most model performance metrics, which is consistent with the industry prevalent approaches of giving some weight to scenarios that are somewhat adverse. Across scenarios that the more lightly parametricized models tended to perform better according to preferred metrics, and also to produce a lower range of results across metrics. An implication of this analysis is a risk CECL will give rise to challenges in comparability of results temporally and across institutions, as estimates vary substantially according to model specification and framework for scenario generation. We also quantify the level of model risk in this hypothetical exercise using the principle of *relative entropy*, and find that credit models featuring more elaborate modeling choices in terms of number of variables, such as more highly parametricized models, tend to introduce more measured model risk; however, the more highly parametricized MS-VAR model, that can accommodate non-normality in credit loss, produces lower measured model risk. The implication is that banks may wish to err on the side of more parsimonious approaches, that can still capture non-Gaussian behavior, in order to manage the increase model risk that the introduction of the CECL standard gives rise to. We conclude that investors and regulators are advised to develop an under-

¹ Corresponding author: Michael Jacobs, Jr., Ph.D., CFA - Lead Quantitative Analytics & Modeling Expert, PNC Financial Services Group – Balance Sheet Analytics and Modeling / Model Development, 340 Madison Avenue, New York, N.Y., 10022, 917-324-2098, michael.jacobsjr@pnc.com. The views expressed herein are solely those of the author and do not necessarily represent an official position of PNC Financial Services Group.

standing of what factors drive these sensitivities of the CECL estimate to modeling assumptions, in order that these results can be used in prudential supervision and to inform investment decisions. .

JEL Classification numbers: G21, G28, M40, M48.

Keywords: Accounting Rule Change, Current Expected Credit Loss, Allowance for Loan and Lease Losses, Credit Provisions, Credit Risk, Financial Crisis, Model Risk.

1 Introduction

In the United States, the *Financial Accounting Standards Board* (“FASB”) is charged for the origination and issuance of the set of standards known as *Generally Accepted Accounting Principles* (“U.S. GAAP”). These standards represent a common set of guidelines for the accounting and reporting of financial results, the intent of which being to enforce standards established to insure provision of useful information to investors and other stakeholders. In this study we focus on the guidance governing the *Allowance for Loan and Lease Losses* (“ALLL”), which represent the financial reserves that firms exposed to credit risk set aside for possible losses on instruments subject to such risk. The recent revision to these standards, the *Current Expected Credit Loss* (“CECL”; FASB, 2016) standard, is expected to substantially alter the management, measurement and reporting of loan loss provisions amongst financial institutions and companies exposed to credit risk.

The prevailing ALLL loss standard for U.S. has used been the principle of *incurred loss*, wherein credit losses are recognized only when it is likely that a loss has materialized, meaning that there is a high probability that a borrower or loan has become materially weaker in terms of its risk characteristics. The key point here is that this is a calculation as of the financial reporting date and future events are not to be considered, which impairs the capability of managing reserves prior to a period of economic downturn. The result of this deferral implies that provisions are likely to be volatile, unpredictable and subject to the phenomenon of *procyclicality*, which means that provisions rise and regulatory capital ratios decrease exactly in the periods where we would prefer the opposite. Said differently, the incurred loss standard leads to an inflation in ALLL at the trough of an economic cycle, which is detrimental to a bank from a safety and soundness perspective, and also to the economy as a whole as lending will be choked off exactly when businesses and consumers should be supported from the view of systematic risk and credit contagion.

The realization by the industry of this danger motivated the FASB in 2016 to reconsider the incurred loss standard and gave rise to the succeeding CECL standard, according to which a loan’s lifetime expected credit losses are to be estimated at the point of origination. This paradigm necessitates a forward-looking view of the ALLL that more proactively incorporates expected credit losses in advance of the actual deterioration of a loan during an economic downturn. A potential implication of this is that under CECL the provisioning process should exhibit less procyclicality. This comes at a cost however, in that credit risk managers now need to make strong modeling assumptions in order to effectuate this forecast, many of which may be subjective and subject to questioning by model validation as well as the regulators. A further risk under the CECL framework is that the comparability of institutions, both cross-sectionally and over time, may be hindered as the CECL modeling specifications and assumptions are likely to vary widely across banks, from the perspective of prudential supervision and investment management. .

There are some key modeling assumptions to be made in constructing CECL forecasts. First, the specification of the model linking loan losses to the macroeconomic environment will undoubtedly drive results. Second, and no less important, the specification of a model that generates macroeconomic forecasts and most likely scenario projections will be critical in establishing the CECL expectations. As we know from other and kindred modeling exercises, such as *stress testing* (“ST”) used by supervisors to assess the reliability of credit risk models in the revised Basel framework (Basel Committee on Banking Supervision, 2006) or the Federal Reserve’s *Comprehensive Capital Analysis and Review* (“CCAR”) program (Board of Governors of the Federal Reserve System, 2009), models for such purposes are subject to supervisory scrutiny. One concern is that such advanced mathematical, statistical and quantitative techniques and models can lead to *model risk*, 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). We expect that the depth of review and burden of proof will be far more accentuated in the CECL context, as compared to Basel or CCAR, as such model results have financial statement reporting implications.

In this study, toward the end of analyzing the impact of model specification and scenario dynamics upon expected credit loss estimates in CECL, we implement a highly stylized framework borrowed from the ST modeling practice. We perform a model selection of alternative CECL specifications in a top-down framework, using FDIC FR-Y9C (“Call Reports”) data and constructing an aggregate or average hypothetical bank, with the target variable being net the *charge-off rate* (“NCOR”) and the explanatory variables constituted by Fed provided macroeconomic variables as well as bank-specific controls for idiosyncratic risk. We study not only the impact of the ALLL estimate under CECL for alternative model specifications, but also the impact of different frameworks for scenario generation: the Fed baseline assumption, a *Gaussian Vector Autoregression* (“VAR”) model and a *Markov Regime Switching VAR* (“MS-VAR”) model, following the study of Jacobs et al (2018a).

We establish in this study that in general the CECL methodology is at risk of not achieving the stated objective of reducing the pro-cyclicality of provisions relative to the legacy incurred loss standard, as across models we observe chronic underprediction of losses in the last 2-year out-of-sample period, which arguably is a period that is late in the economic cycle. Furthermore, the amount of such procyclicality exhibits significant variation across model specifications and scenario generation frameworks. In general, the MS-VAR scenario generation framework produces the best performance in terms of fit and lack of underprediction relative to the perfect foresight benchmark, which is in line with the common industry practice of giving weight to adverse but probable scenarios, which the MS-VAR regime switching model can produce naturally and coherently as part of the estimation methodology that places greater weight on the economic downturn. We also find that for any scenario generation model, across specification the more lightly parameterized credit risk models tend to have better out of sample performance. Furthermore, relative to the perfect foresight benchmark, the MS-VAR model produces a lower level of variation in the model performance statistics across loss predictive model specifications. As a second exercise, we attempt to quantify the level of model risk in this hypothetical CECL exercise an approach that uses the principle of *relative entropy*. We find that more elaborate modeling choices, such as more highly parametricized models in terms of explanatory variables, tend to introduce more measured model risk, but the MS-VAR specification for scenario generation generates less model risk as compared to the Fed or VAR frameworks. The implication is that banks may wish to err not on the side of more parsimonious approaches, but

also should attempt to model the non-normality of the credit loss distribution, in order to manage the increase model risk that the introduction of the CECL standard may give rise to.

AN implication of this analysis is that the volume of lending and the amount of regulatory capital held may vary greatly across banks, even when it is the case that the respective loan portfolios have very similar risk profiles. A consequence of this divergence of expected loan loss estimates under the CECL standard is that supervisors and other market participant stakeholders may face challenges in comparing banks at a point of time or over time. There are also implications for the influence of modeling choices in specification and scenario projections on the degree of model risk introduced by the CECL standard.

This paper proceeds as follows. In Section 2 we provide some background on CECL, including a survey of some industry practices and contemplated solutions. In Section 3 we review the related literature with respect to this study. Section 4 outlines the econometric methodology that we employ. Modeling data and empirical results are discussed in Section 5. In Section 6 we perform our model risk quantification exercise for the various loss model and scenario generation specifications. Section 7 concludes and presents directions for future research.

2 CECL Background

In Figure 1 we illustrate the procyclicality of credit loss reserves under the incurred loss standard. We plot NCORs, the provisions for loan and lease losses (“PLLL”) and the ALLL for all insured depository institutions in the U.S., sourced from the FDIC Call Reports (or the forms FR Y-9C) for the period 4Q01 to 4Q17. Note that these quantities are an aggregate across all banks, or an average weighted by dollar amounts, representing the experience of an “average bank”. NCORs began to their ascent at the start of the Great Recession in 2007, while PLLLs exhibit a nearly coinciding rise (albeit with a slight lead), while the ALLL continues to rise well after the economic downturn and peaks in 2010, nearly a year into the economic recovery. This coincided with deterioration in bank capital ratios, which added to stress to bank earnings and impaired the ability of institutions to provide sorely needed loans, arguably contributing to the sluggishness of the recovery in the early part of the decade.

In the aftermath of the global financial crisis there was an outcry from stakeholders in the ALLL world (banks, supervisors and investors alike) against the incurred loss standard. As a result of this critique, the accounting standard setters (both FASB and the International Accounting Standards Board – “IASB”) proposed a revamped expected loss (“EL”) based framework for credit risk provisioning. In July of 2014 IASB released its new standard, International Reporting for Financial Statement Number 9 (IASB, 2104; “IRFS9”), while FASB issued the CECL standard in June of 2016 (FASB, 2016).

While there are many commonalities between the two rules, namely that in principle they are EL frameworks as opposed to incurred loss paradigms, there are some notable differences between the two. Namely, in CECL we must estimate lifetime expected credit losses for all instruments subject to default risk, whereas IRFS 9 only requires this life-of-loan calculation for assets that have experienced severe credit deterioration and only a 1-year EL for performing loans. Another methodological difference is IFRS 9 contains a trigger that increases ALLL from 1 year EL expected losses to lifetime EL in the event that losses become of probable. There is also a difference in timing of when these standards take effect, for CECL 2020 for SEC filers and 2021 for non-SEC filers, whereas IRFS9 went into effect in January of 2018.

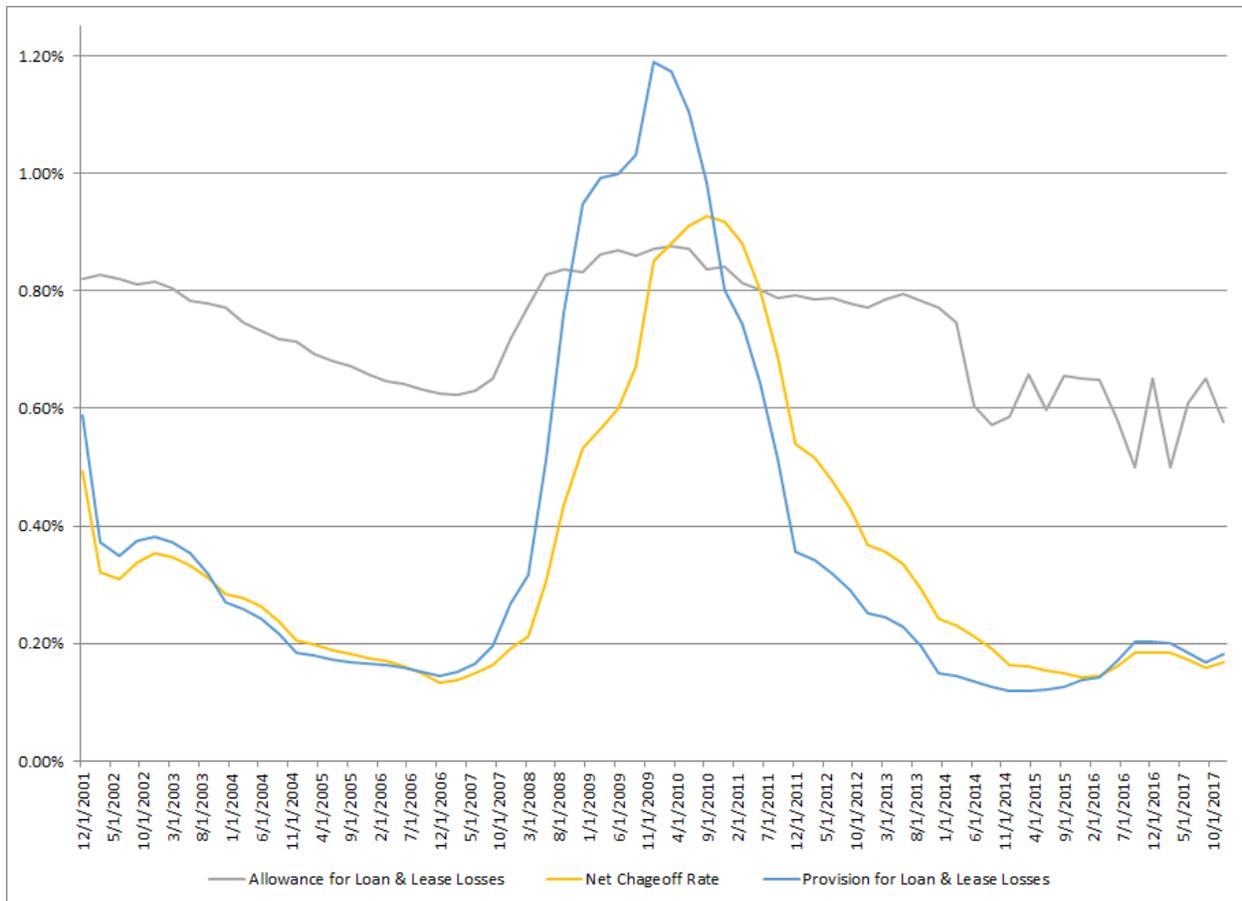


Figure 1: Net Charge-off Rates, Loan Loss Provisions and the ALLL as a Percent of Total Assets – All Insured Depository Institutions in the U.S. (Federal Deposit Insurance Corporation Statistics on Depository Institutions Report – Schedule FR Y-9C)

Focusing on CECL requirement, the scope encompasses all financial assets carried booked at amortized cost, *held-for-investment* (“HFI”) or *held-to-maturity* (“HTM”) instruments, which represent the majority of assets held by depository institutions (the so-called *banking book*), and such loans are the focus of this research. CECL differs from the traditional incurred loss approach in that it is an EL methodology for credit risk that uses information of a more forward looking character, and applied over the lifetime of the loan as of the financial reporting date. This covers all eligible financial assets, not only those already on the books, but also including newly originated or acquired assets. In the CECL framework, the ALLL is a valuation account, which means that it represents the difference between a financial assets’ *amortized cost basis* and the net amount expected to be collected from such assets.

In the estimation of the expected net collection amounts, the CECL standard stipulates that banks condition on historical data (i.e., risk characteristics, exposure, default and loss severity observations), the corresponding current portfolio characteristics to which history is mapped, as well as what FASB terms to be *reasonable and supportable forecasts* (i.e., forward-looking estimates of macroeconomic factors and portfolio risk characteristics) relevant to assessing the credit quality of risky exposures. However, much as in the *Basel Advanced Models Approach* or CCAR with respect to the banking supervisors, the FASB is not prescriptive with respect to the model specifications and methodologies that constitute reasonable and supportable assumptions. The intent of the FASB in specifying a principles based accounting standard

was to enable comparability and scalability across of range of institutions, differing in size and complexity. In view with this goal, the CECL standard does not mandate a particular methodology for the estimation of expected credit losses, and gives banks the latitude to elect estimation frameworks choose that are based upon elements that can be reasonably supported. For example, key amongst these elements to be supported is the forecast period, which is unspecified under the standard, but subject to this requirement of reasonableness and supportability. In particular, such forecast periods should incorporate contractual terms of assets, and in cases of loans having no fixed terms (e.g., revolving or unfunded commitments) such terms have to be estimated empirically and introduce another modeling element into the CECL framework.

Loan loss provisions are meant to provide banking examiners, auditors and financial market participants a measure of the riskiness of financial assets subject to default or downgrade risk. The incurred loss standard does so in backward looking framework, while CECL is meant to do such on a forward-looking basis. Presumably, the variation in ALLL under the legacy standard would be principally attributed to changes in the inherent riskiness of the loan book, such as *losses-given-default* (“LGDs”) or *probabilities of default* (“PDs”) that drive *Expected Loss* (“EL”). However, in the case of CECL, there are additional sources of variation that carry significantly greater weight than under the incurred loss setting, which create challenges in making comparisons of institutions across time or at a point in time.

The sources of variation in loan loss provisions that are common between the former and CECL frameworks are well understood the credit risk modeling practice. These are the portfolio characteristic factors driving PDs and LGDs, at the obligor or loan level (e.g., risk ratings, collateral values, financial ratios), or at the industry ort sector level (e.g., geographic or industry concentrations, business conditions). Such factors are estimated from historical experience, but then applied on a static basis, by holding constant characteristics driving losses constant at the start of the forecasting horizon. Market participants and other stake

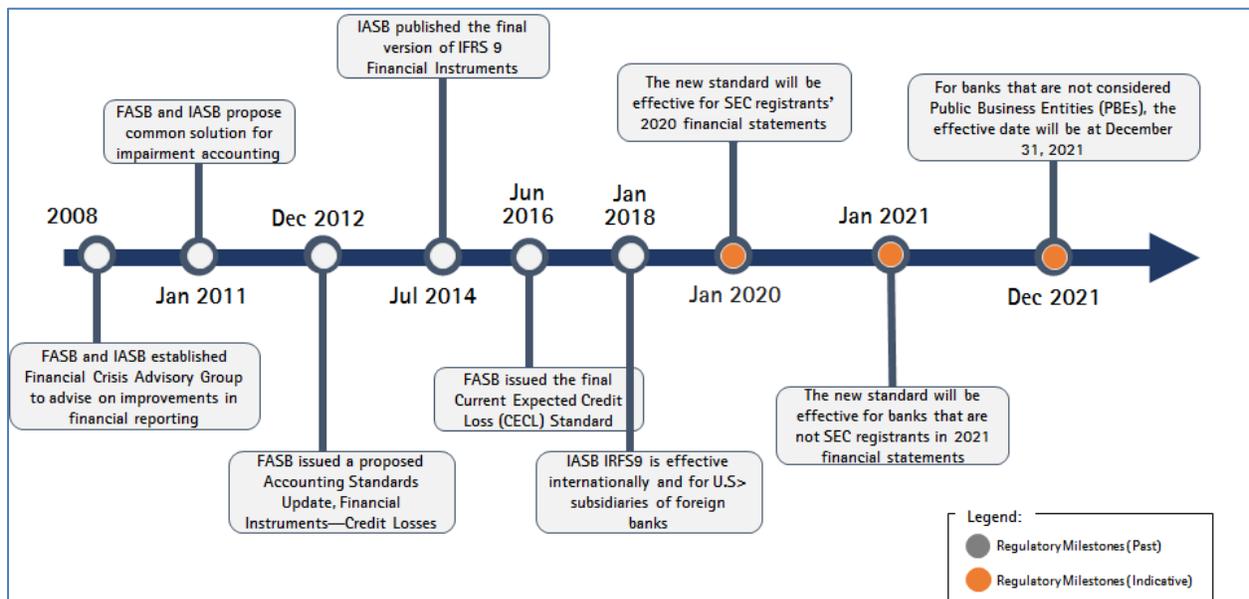


Figure 2: The Accounting Supervisory Timeline for CECL and IFRS9 Implementation

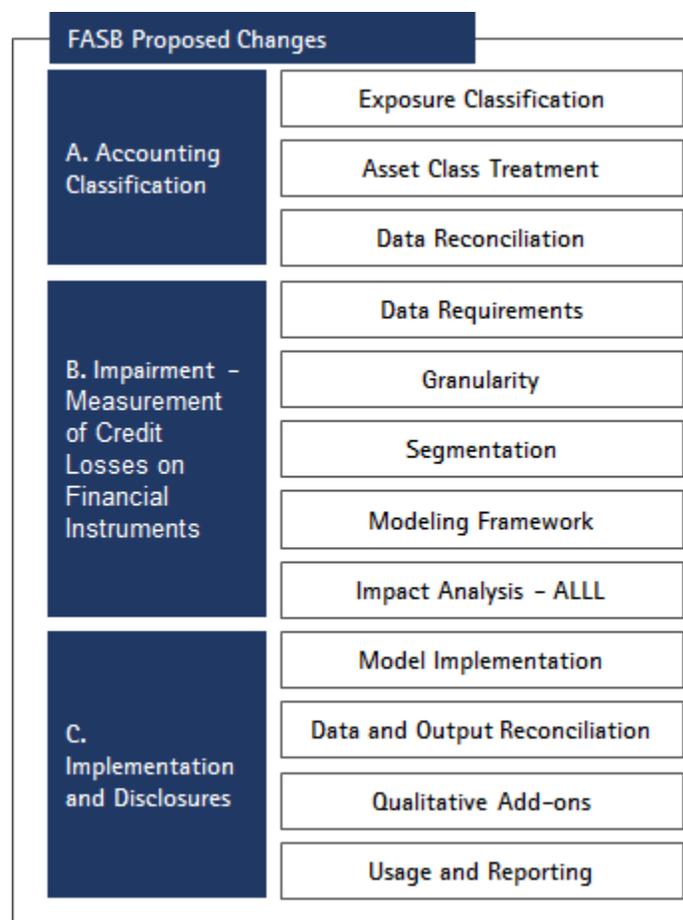


Figure 3: The CECL Accounting Standard – Regulatory Overview

holders are rather comfortable with understanding the composition of credit risk and provisions based upon these factors and the models or methodologies linking them to credit loss.

Modeling expected losses under CECL differs from other applications, such as decisioning or regulatory capital, is that this framework necessitates the estimation of credit losses over the lifetime of a financial asset, and such projections must be predicated upon reasonable and supportable expectations of the future economic environment. This implies that models for the likely paths of macroeconomic variables will likely have to be constructed. Another set of models embedded in the CECL framework introduces an additional complication, that not only makes challenging the interpretation of results, but also introduces a compounding of model risk and potential challenge by model validation and other control functions. This subjective and idiosyncratic modeling choice is not only uncommon in current models supporting financial reporting, but also in other domains that incorporate macroeconomic forecasts. Note that in CCAR, base projections were generally sourced from the regulators, and hence modeling considerations were not under scrutiny².

We conclude this section with a discussion of some of the practical challenges facing institutions in implementing CECL frameworks. In Figure 2 we depict the regulatory timeline for the evolution of the

² Several challenges are associated with macroeconomic forecasting related to changes in the structure of the economy, measurement errors in data as well as behavioral biases (Batchelor and Dua, 1990).

CECL standard.. In the midst of the financial crisis during 2008, when the problem of countercyclicality of loan loss provision came to the fore, the FASB and the IASB established the *Financial Crisis Advisory Group* to advise on improvements in financial reporting. This was followed in early 2011 with the communication by the accounting bodies of a common solution for impairment reporting. In late 2012, the FASB issued a proposed change to the accounting standards governing credit loss provisioning (FASB, 2012), which was finalized after a period of public comment in mid-2016 (FASB, 2016); while in the meantime the IASB issued its final IRFS9 accounting standard in mid-2014 (IASB, 2014). The IRFS9 standard was effective as of January, 2018 while CECL is effective in the U.S. for SEC registrants in January, 2020 and then for non-SEC registrants in January, 2021; however, for banks that are not considered *Public Business Entities (PBEs)*, the effective date will be at December 31, 2021.

In Figure 3 we depict some high level overview of the regulatory standards and expectations in CECL. The first major element, which has no analogue in the legacy ALLL framework, is that there has to be a clear *segmentation* of financial assets, into groupings that align with portfolio management and which also represent groupings in which there is homogeneity in credit risk. This practice is part of traditional credit risk modeling, as has been the practice in Basel and CCAR applications, but which represents a fundamental paradigm shift in provisioning processes. Second, there are changes to the framework for measuring impairment and credit losses on financial instruments, which has several elements. One key aspect is enhances data requirements for items such *troubled debt restructurings* (“TDRs”) on distressed assets, and lifetime loss modeling for performing assets. This will require a definition of model granularity based on existing model inventories (i.e., for Basel and CCAR), data availability and a target level of accuracy. Moreover, this process will involve the adoption of new modeling frameworks for provision. Finally, institutions will face a multitude of challenges around implementation and disclosures. This involves an enhanced implementation platform for model and reporting (e.g., dashboard), as well as revised accounting policies for loans and receivables, foreclosed and repossessed assets and fair value disclosures.

The new CECL standard is expected to have a significant business impact on the accounting organizations of financial institutions by increasing the allowance, as well as operational and technological impacts due to the augmented complexity of compliance and reporting processes:

- **Business Impacts**
 - Significant increase in the ALLL of 25 – 100%, varying based on portfolios
 - Potential reclassification of instruments & additional data requirements for lifetime loss calculations
 - Additional governance and control burdens due to new set of modeling frameworks & implementation platforms
 - More frequent consolidation of modeling and GL data, as well as results from multiple sources
 - Enhanced reporting of the ALLL and other factors
- **Operational Impacts**
 - Increased operational complexity due to augmented accounting requirements
 - Additional modeling and other operations resource requirements to support modeling, risk reporting and management
 - Alignment between modeling and business stakeholders
 - Operational governance increases for data quality, lifetime calculation, modelling and GL reconciliation

- **Technological Impacts**
 - Increased computational burden for different portfolios (e.g., high process times for portfolios based on granularity, segmentation and selected model methodology)
 - Expansion of more granular historical data capacity
 - Large computational power for more frequent (quarterly) runs of the ALLL estimate
 - Augmented time requirements to stabilize the qualitative and business judgement overlays across portfolios

3 Review of the Literature

The procyclicality of the incurred loss standard for the provisioning of expected credit losses has been extensively discussed by a range of authors: Bernanke and Lown (1991), Kishan and Opiela (2000), Francis and Osborne (2009), Berrospide and Edge (2010), Cornett, McNutt, Strahan, and Tehranian (2011), and Carlson, Shan, and Warusawitharana (2013).

In a study that is closest in the literature to what we accomplish in this paper, Chae et al (2018) notes that CECL is intended to promote proactive provisioning as loan loss reserves can be conditioned on expectations of the economic cycle. They study the degree to which a single modeling decision, expectations about the path of future house prices, affects the size and timing of provisions for first-lien residential mortgage portfolios. The authors find that while CECL provisions are generally less pro-cyclical as compared to the current incurred loss standard, the revised standard may complicate the comparability of provisions across banks and time.

We note some key studies of model risk and its quantification, to complement the supervisory guidance that has been released (Board of Governors of the Federal Reserve System, 2011). In the academic literature, Jacobs (2015) contributes to the evolution of model risk management as a discipline by shifting the focus on individual models towards aggregating firmwide model risk, noting that regulatory guidance specifically focuses on measuring risk individually and in aggregate. The author discusses various approaches to measuring and aggregating model risk across an institution, and also presents an example of model risk quantification in the realm of stress-testing, where he compares alternative models in two different classes, Frequentist and Bayesian approaches, for the modelling of stressed bank losses. In the practitioner realm, a whitepaper by Accenture Consulting (Jacobs et al, 2015a), it is noted that banks and financial institutions are continuously examining their target state model risk management capabilities to support the emerging regulatory and business agendas across multiple dimensions, and that the field continues to evolve with organizations developing robust frameworks and capabilities. The authors note that to date industry efforts focused primarily on model risk management for individual models, and now more institutions are shifting focus toward aggregating firm-wide model risk, as per regulatory guidance specifically focusing on measuring risk individually and in the aggregate. They provide background on issues in MRM, including an overview of supervisory guidance and discuss various approaches to measuring and aggregating model risk across an institution. Glasserman and Xu (2013) develop a framework for quantifying the impact of model risk and for measuring and minimizing risk in a way that is robust to model error. This robust approach starts from a baseline model and finds the worst case error in risk measurement that would be incurred through a deviation from the baseline model, given a precise constraint on the plausibility of the deviation, using relative entropy to constrain model distance leads to an explicit characterization of worst-case model errors. This approach goes well beyond the effect of errors in parameter estimates to consider errors in the underlying stochastic assumptions of the model and to characterize the greatest vulnerabilities to error in a model. The authors apply this approach to problems of portfolio risk measurement, credit risk, delta hedging and counterparty risk measured through credit valuation adjustment. Skoglund (2018) studies the quantification of the model risk inherent in loss projection models used in the

macroeconomic stress testing and impairment estimation, which is of significant concern for both banks and regulators. The author applies relative entropy techniques that allow model misspecification robustness to be numerically quantified using exponential tilting towards an alternative probability law. Using a particular loss forecasting model, he quantifies the model worst-case loss term-structures to yield insight into what represents in general an upward scaling of the term-structure consistent with the exponential tilting adjustment. The author argues that this technique can complement the traditional model risk quantification techniques where specific directions or range of reasons for model misspecification are usually considered.

There is rather limited literature on scenario generation in the context of stress testing. One notable study that examines this in the context of CCAR and credit risk is Jacobs et al (2018a), who conduct 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.

4 Time Series VAR Methodologies for Estimation and Scenario Generation

Stress testing is concerned principally concerned with the policy advisory functions of macroeconomic forecasting, wherein stressed loss projections are leveraged by risk managers and supervisors as a decision-support tool informing the resiliency institutions during stress periods³. Traditionally the way that these objectives have been achieved ranged from high-dimensional multi-equation models, all the way down to single-equation rules, the latter being the product of economic theories. Many of these methodologies were found to be inaccurate and unstable during the economic tumult of the 1970s as empirical regularities such as Okun's Law or the Phillips Curve started to fail. Starting with Sims (1980) and the VAR methodology we saw the arrival of a new paradigm, where as opposed to the univariate AR modeling framework (Box and Jenkins, 1970; Brockwell and Davis, 1991; Commandeur and Koopman, 2007), the VAR model presents as a flexible multi-equation model still in the linear class, but in which variables can be explained by their own and other variable's lags, including variables exogenous to the system. We consider the VAR methodology to be appropriate in the application of stress testing, as our modeling interest concerns relationships and forecasts of multiple macroeconomic and bank-specific variables. We also consider the MS-VAR paradigm in this study, which is closely related to this linear time-invariant VAR model. In this framework we analyze the dynamic propagation of innovations and the effects of regime change in a system. A basis for this approach is the statistics of *probabilistic functions of Markov chains* (Baum and Petrie, 1966; Baum et al., 1970). The MS-VAR model also subsumes the *mixtures of normal distributions* (Pearson, 1984) and *hidden Markov-chain* (Blackwell and Koopmans, 1957; Heller, 1965) frameworks. All of these approaches are further related to *Markov-chain regression models* (Goldfeld and Quandt, 1973) and to the statistical analysis of the *Markov-switching models* (Hamilton 1988, 1989). Most closely aligned to our application is the theory of *doubly stochastic processes* (Tjostheim, 1986) that incorporates the MS-VAR model as a Gaussian autoregressive process conditioned on an exogenous regime generating process.

³ Refer to Stock and Watson (2001) for a discussion of the basic aspects of macroeconomic forecasting (i.e., characterization, forecasting, inferences and policy advice regarding macroeconomic time series and the structure of the economy.)

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 or idiosyncratic factors. This gives rise to the *VARMAX* (p, q, s) (“vector autoregressive-moving average with exogenous variables”) representation:

$$\mathbf{Y}_t \Phi(B) = \mathbf{X}_t \Theta(B) + \mathbf{E}_t \Theta^*(B) \tag{1}$$

Which is equivalent to:

$$\mathbf{Y}_t - \sum_{j=1}^p \Phi_j \mathbf{Y}_{t-j} = \sum_{j=0}^s \Theta_j \mathbf{X}_{t-j} + \mathbf{E}_t - \sum_{j=1}^q \Theta_j^* \mathbf{E}_{t-j} \tag{2}$$

Where $\Phi(B) = \mathbf{I}_k - \sum_{j=1}^p \Phi_j B^j$, $\Theta(B) = \sum_{j=0}^s \Theta_j B^j$ 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 $\mathbf{E}_t = (E_{1t}, \dots, E_{kt})^T$ ⁴. The autoregressive parameter matrices Φ_j represent sensitivities of output variables to their own lags and to lags of other output variables, while the corresponding matrices Θ_j are model sensitivities of output variables to contemporaneous and lagged values of input variables⁵. It follows that the dependency structure of the output variables \mathbf{Y}_t , as given by the autocovariance function, is dependent upon the parameters Φ_j , and hence the correlations amongst the \mathbf{Y}_t as well as the correlation amongst the \mathbf{X}_t that depend upon the parameters Θ_j . In contrast, in a system of univariate *ARMAX* (p, q, s) (“autoregressive-moving average with exogenous variables”) models, the correlations amongst the elements of \mathbf{Y}_t are not taken into account, hence the parameter vectors Θ_j have a diagonal structure (Brockwell and Davis, 1991).

⁴ 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 modelling in which a simple step (or pulse indicator) function taking the values of 0 or 1 indicates the effect of output due to unusual intervention events in the system.

⁵ Note that the VARMAX model (1)–(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*. A related model is obtained by multiplying the dynamic simultaneous equations model form by the inverse of the lag 0 coefficient matrix and is referred to as the *reduced form model*, which has a *state space representation* (Hanan, 1988).

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 $\Theta_j^* = \mathbf{0}_{k \times k} \quad j > 0$:

$$\mathbf{Y}_t - \sum_{j=1}^p \Phi_j \mathbf{Y}_{t-j} = \sum_{j=1}^s \Theta_j \mathbf{X}_{t-j} + \mathbf{E}_t \quad (3)$$

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 VARMAX 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, 2017). Finally, the VARX framework is more practical and intuitive than the more elaborate VARMAX model, and allows for superior communication of results to practitioners.

We now consider the MS-VARX generalization of the VARX methodology with changes in regime, where the parameters of the VARX system $\mathbf{B} \triangleq (\Phi^T, \Theta^T)^T \in R^{p+s}$ will be time-varying. However, the process might be time-invariant conditional on an unobservable regime variable $s_t \in (1, \dots, M)$, denoting the state at time t out of M feasible states. In that case, then the conditional probability density of the observed time series \mathbf{Y}_t is given by:

$$p(\mathbf{Y}_t | \Psi_t, s_t) = \begin{cases} f(\mathbf{Y}_t | \Psi_{t-1}, \mathbf{B}_1) & \text{if } s_t = 1 \\ \vdots \\ f(\mathbf{Y}_t | \Psi_{t-1}, \mathbf{B}_M) & \text{if } s_t = M, \end{cases} \quad (4)$$

Where \mathbf{B}_m is the VAR parameter matrix in regime $m \in (1, \dots, M)$ and Ψ_{t-1} are the observations $\left\{ \mathbf{y}_{t-j} \right\}_{j=1}^{\infty}$. Therefore, given a regime s_t , the conditional $VARX(p, s | s_t)$ system in expectation form can be written as:

$$E[\mathbf{Y}_t | \Psi_{t-1}, s_t] = \sum_{j=1}^p \Phi_j \mathbf{Y}_{t-j}(s_t) + \sum_{j=1}^s \Theta_j(s_t) \mathbf{X}_{t-j} \quad (5)$$

We define the innovation term as:

$$\mathbf{E}_t = \mathbf{Y}_t - E[\mathbf{Y}_t | \Psi_t, s_t] \quad (6)$$

The innovation process ε_t is a Gaussian, zero-mean white noise process having variance-covariance matrix $\Sigma(s_t)$:

$$\mathbf{E}_t \sim NID(\mathbf{0}, \Sigma(s_t)) \quad (7)$$

If the $VARX(p, s | s_t)$ process is defined conditionally upon an unobservable regime s_t as in equation (4), the description of the process generating mechanism should be made complete by specifying the stochastic assumption of the MS-VAR model. In this construct, s_t follows a discrete state homogenous Markov chain:

$$\Pr\left(s_t \mid \left\{ \mathbf{y}_{t-j} \right\}_{j=1}^{\infty}, \left\{ s_{t-j} \right\}_{j=1}^{\infty} \right) = \Pr\left(\left\{ s_{t-j} \right\}_{j=1}^{\infty} \mid \boldsymbol{\rho} \right) \quad (8)$$

Where $\boldsymbol{\rho}$ denotes the parameter vector of the regime generating process. We estimate the MS-VAR model using MSBVAR the package in R (R Development Core Team, 2019). Finally note that in the remainder of the document outside this section we will use the acronyms VAR and MS-VAR instead of VARX and MS-VARX to refer to our competing modeling methodologies.

5 Data and Empirical Results

In this paper the data are sourced from the *Statistics on Depository Institutions* (“SDI”) report, which is available on the *Federal Deposit Insurance Corporation’s* (“FDIC”) research website⁶. This bank data represents all insured depository institutions in the U.S. and contains income statement, balance sheet and off-balance sheet line items. We use quarterly data from the 4th quarter of 1991 through the 4th quarter of 2017. The models for CECL are specified and estimated using a development period that ends in the 4th quarter of 2015, leaving the last 2 years 2016 and 2017 as an out-of-sample test time period. The model development data are Fed macroeconomic variables, as well as aggregate asset-weighted average values bank financial characteristics for each quarter, the latter either in growth rate form or normalized by the total value bank assets in the system.

The Federal Reserve’s CCAR stress testing exercise requires U.S. domiciled top-tier financial institutions to submit comprehensive capital plans conditioned upon prescribed supervisory, and at least a single bank-specific, set of scenarios (base, adverse and severe). The supervisory scenarios are constituted of 9 quarter paths of critical *macroeconomic variables* (“MVs”). In the case of institutions materially engaged in trading activities, in addition there is a requirement to project an instantaneous market or counterparty credit loss shock conditioned on the institution’s idiosyncratic scenario, in addition to supervisory prescribed market risk stress scenarios. Additionally, large custodian banks are asked to estimate a potential default of their largest counterparty.

Institutions are asked to submit post-stress capital projections in their capital plan starting September 30th of the year, spanning the nine-quarter planning horizon that begins in the fourth quarter of the current year, defining movements of key MVs. In this study we consider the MVs of the 2015 CCAR, and their Base scenarios for CECL purposes:

- Real Gross Domestic Product Growth (“RGDP”)
- Real Gross Domestic Investment (“RDIG”)
- 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 Bond Rate (“BBBCR”)
- Dow Jones Index (“DJI”)
- National House Price Index (“HPI”)
- Nominal Disposable Personal Income Growth (“NDPIG”)

⁶ These are available at <https://www5.fdic.gov/sdi/main.asp?formname=standard>.

- Mortgage Rate (“MR”)
- CBOE’s Equity 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 VAR and MS-VAR models for the purposes of scenario generation⁷:

- 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)

We considered 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 on 5 of the 9 most commonly used national Fed CCAR MVs as input variables in the VAR model:

- Unemployment Rate (“UNEMP”)
- BBB Corporate Bond Yield (“BBBCY”)
- Commercial Real Estate Price Index (“CREPI”)
- CBOE Equity Volatility Index (“VIX”)
- BBB Corporate Credit Spread (“CORPSR”)

Similarly, we identify the following balance sheet items, banking aggregate idiosyncratic factors, according to the same criteria:

- Commercial and Industrial Loans to Total Assets (“CILTA”)
- Commercial and Development Loans Growth Rate (“CDLGR”)
- Trading Account Assets to Total Assets (“TAATA”)
- Other Real Estate Owned to Total Assets (“OROTA”)
- Total Unused Commitments Growth Rate (“TUCGR”)

This historical data, 65 quarterly observations from 4Q01 to 4Q17⁸, are summarized in Table 1 in terms of distributional statistics and correlations, and in Figures 1 through 10 of this section both historically as well as 3 sce-

⁷ We perform this model selection in an R script designed for this purpose, using the libraries “dse” and “tse” to estimate and evaluate VAR and MS-VAR models (R Core Development Team, 2019).

narios generation models (Fed, VAR and MS-VAR) for the period 1Q16-4Q17. Across all series, we observe that the credit cycle is clearly reflected, with indicators of economic or financial stress (health) displaying peaks (troughs) in the recession of 2001–2002 and in the financial crisis of 2007–2008, with the latter episode dominating in terms of severity by an order of magnitude. However, there are some differences in timing, extent and duration of these spikes across macroeconomic variables and loss rates. These patterns are reflected in the percent change transformations of the variables as well, with corresponding spikes in these series that correspond to the cyclical peaks and troughs, although there is also much more idiosyncratic variation observed when looking at the data in this form. First, we will describe main features of the distributional statistics of all variables, then the correlations with the variables and the NCORs, followed by the dependency structure within the group of input macroeconomic variables, and finally then the same for the input bank idiosyncratic variables, and finally the cross-correlations between these two groups. The summary statistics are shown in the top panel of Figure 1. The mean NCOR is 0.35%, ranging from 0.13% to 0.93%, and showing some positive skewness as the median is lower than the mean at 0.26%; furthermore, there is relatively high variation relative to the mean, with a coefficient of variation of 0.63. The mean UNEMP is 6.31%, ranging from 4.10% to 9.90%, and showing some positive skewness as the median is lower than the mean at 5.70%; furthermore, there is relatively high variation relative to the mean, with a coefficient of variation of 0.80. The mean BBBCY is 5.54%, ranging from 3.10% to 9.40%, and showing almost no skewness as the median is very close to the mean at 5.50%; furthermore, there is relatively low variation relative to the mean, with a coefficient of variation of 0.23. The mean CREPI is 201.12, ranging from 138.70 to 278.70, and showing almost no skewness as the median is very close to the mean at 198.50; furthermore, there is relatively low variation relative to the mean, with a coefficient of variation of 0.21. The mean VIX is 28.56, ranging from 12.70 to 80.90, and showing positive skewness as the median is lower than the mean at 22.70; furthermore, there is relatively low high relative to the mean, with a coefficient of variation of 0.46. The mean CORSPR is 2.99%, ranging from 1.40% to 7.20%, and showing almost no skewness as the median is very close to the mean at 3.00%; furthermore, there is moderate variation relative to the mean, with a coefficient of variation of 0.38. The mean CILTA is 0.11, ranging from 0.09 to 0.13, and showing almost no skewness as the median is very close to the mean at 0.11; furthermore, there is rather small variation relative to the mean, with a coefficient of variation of 0.09. The mean CDLGR is 0.11, ranging from -0.09 to 0.16, and showing positive skewness as the median is lower than the mean at 0.02; furthermore, there is very large variation relative to the mean, with a coefficient of variation of 7.21. The mean TAATA is 0.05, ranging from 0.03 to 0.08, and showing no skewness as the median is equal to the mean at 0.05; furthermore, there is rather moderate variation relative to the mean, with a coefficient of variation of 0.20. The mean OROTA is 0.0015, ranging from 0.0005 to 0.0040, and showing positive skewness as the median is lower than the mean at 0.0009; furthermore, there is rather high variation relative to the mean, with a coefficient of variation of 0.73. The mean TUCGR is 0.0062, ranging from -0.0889 to 0.0587, and showing positive skewness as the median lower than the mean at 0.0133; furthermore, there is very high variation relative to the mean, with a coefficient of variation of 4.53.

⁸ We leave out the last 2 years of available data, 1Q16–4Q17, in order to have a holdout sample for testing the accuracy of the models. 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 Historical Y9 Credit Loss Rates, Banking System Idiosyncratic Variables and Macroeconomic Variables (FDIC SDI Report and Federal Reserve Board 4Q91-4Q15).

Summary Statistics	Output Variable		Input Variables - Idiosyncratic Variables					Input Variables - Macroeconomic Variables					
	Count	Net Change off Rate	Commercial & Industrial Loans ¹	Construction & Development Loans ²	Trading Account Assets ³	Other Real Estate Owned ⁴	Total Unused Commitments ⁵	Unemployment Rate	BBB Corporate Bond Yield	Commercial Real Estate Price Index	CBOE Market Volatility Index	BBB Corporate - 5 Year Treasury Spread	
Count	65		65	64	65	65	64	65	65	65	65	65	
Mean	0.35%		0.1054	0.0073	0.0486	0.0015	0.0062	6.31	5.54	201.12	26.56	2.99	
Standard Deviation	0.23%		0.0096	0.0528	0.0095	0.0011	0.0281	1.73	1.29	40.64	12.09	1.15	
Minimum	0.13%		0.0871	-0.0933	0.0327	0.0005	-0.0889	4.10	3.70	138.70	12.70	1.40	
25th Percentile	0.17%		0.0986	-0.0353	0.0418	0.0007	-0.0015	5.00	4.50	167.30	18.90	2.30	
Median	0.26%		0.1054	0.0227	0.0496	0.0009	0.0133	5.70	5.50	198.50	22.70	3.00	
75th Percentile	0.44%		0.1120	0.0356	0.0527	0.0023	0.0227	7.70	6.40	234.00	30.80	3.50	
Maximum	0.93%		0.1296	0.1390	0.0756	0.0040	0.0587	9.90	9.40	278.70	80.90	7.20	
Coefficient of Variation	0.6714		0.0909	7.2075	0.1955	0.7223	4.5269	0.27	0.23	0.20	0.46	0.38	
Skewness	1.3235		0.1477	-0.2094	0.4624	0.9052	-1.2510	0.7939	0.7670	0.2092	1.9496	1.4919	
Kurtosis	0.6717		-0.2365	-0.0594	0.1591	-0.5947	2.0542	-0.7194	0.3860	-1.1247	5.5055	4.1348	
Output Variable	Net Change-off rate												
Correlations	Input Variables - Macroeconomic Variables	Commercial & Industrial Loans ¹	-0.5854										
		Construction & Development Loans ²	-0.8245	0.3869									
		Trading Account Assets ³	0.6256	-0.4280	-0.3394								
		Other Real Estate Owned ⁴	0.6905	-0.5857	-0.8368	0.0575							
		Total Unused Commitments ⁵	-0.6389	0.1296	0.6055	-0.0728	-0.5876						
		Unemployment Rate	0.9121	-0.6162	-0.8037	0.2955	0.9365	-0.4144					
		BBB Corporate Bond Yield	0.5206	-0.2511	-0.3118	0.4836	0.1325	-0.1790	0.0663				
		Commercial Real Estate Price Index	-0.7998	0.5631	-0.5793	-0.13259861	-0.6551	0.2539	-0.5374	-0.4076			
		CBOE Market Volatility Index	0.6401	-0.2199	-0.5174	0.0468	0.4452	-0.2706	0.3846	0.6017	-0.2187		
		BBB Corporate - 5 Year Treasury Spread	0.6939	-0.4547	-0.5544	0.0812	0.6453	-0.2115	0.6123	0.4680	-0.2516	0.8087	

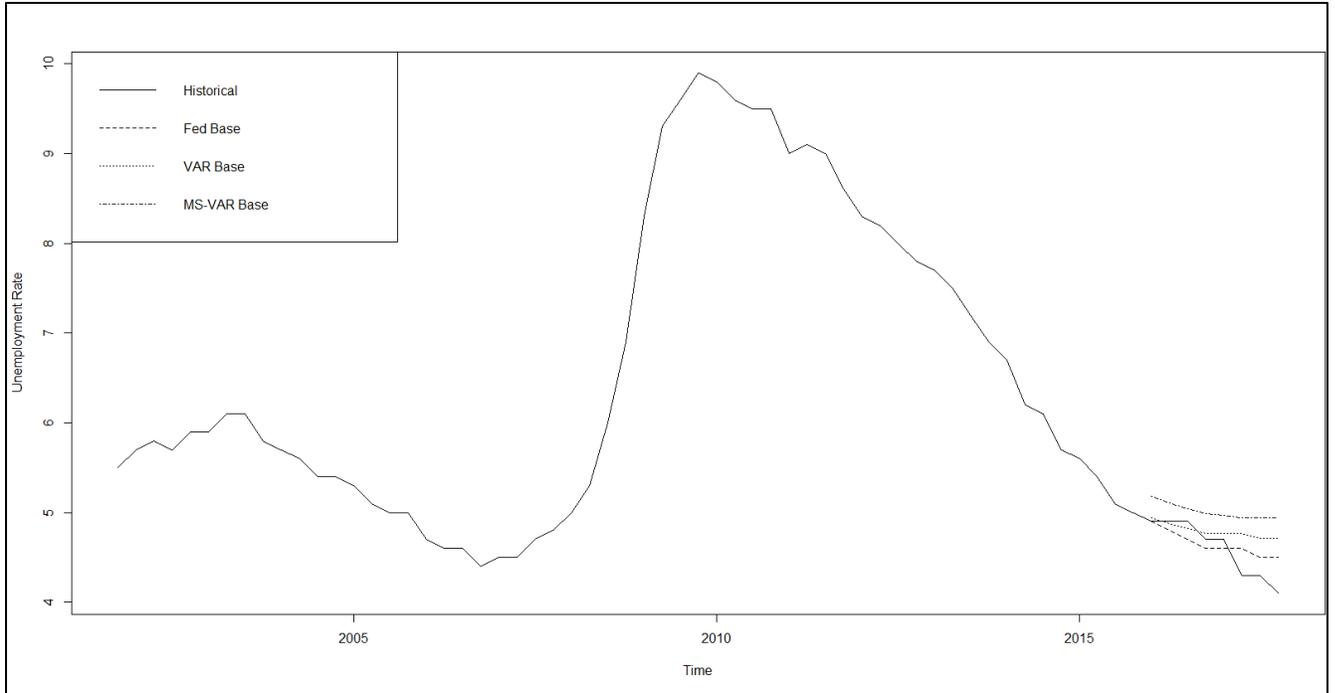


Figure 1: Time Series and Base Scenarios - Unemployment Rate (Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

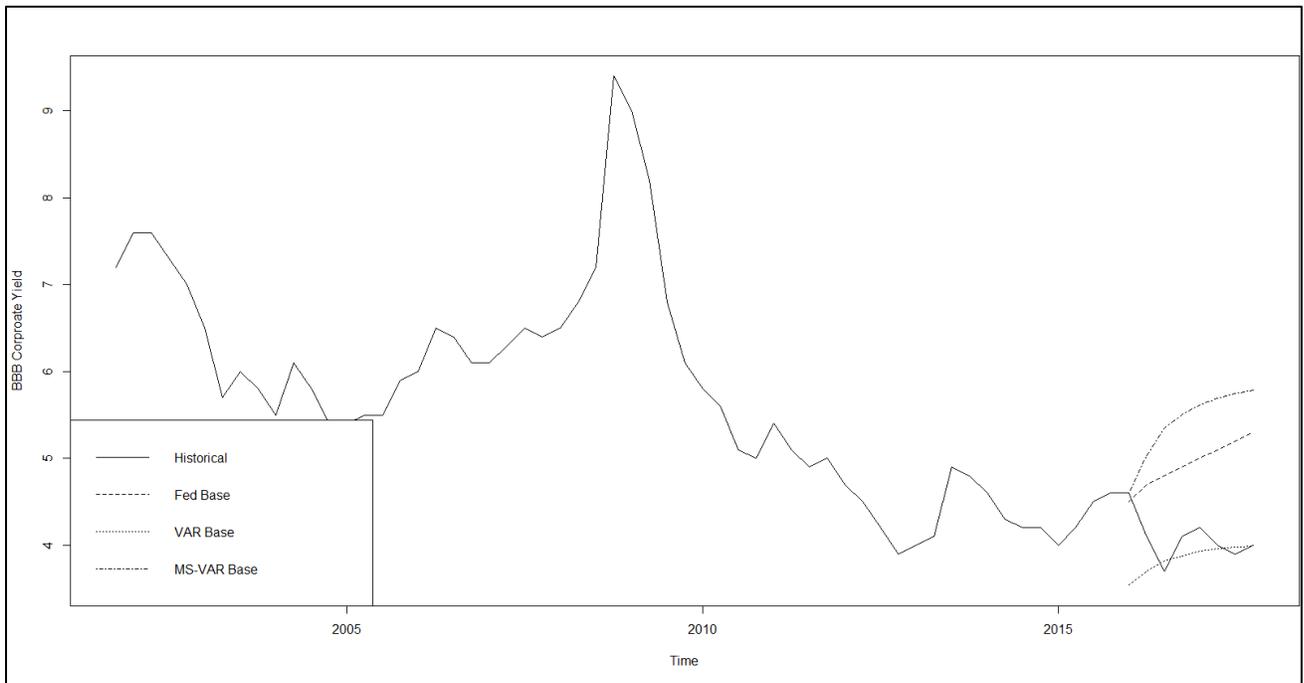


Figure 2: Time Series and Base Scenarios - BBB Corporate Yield (Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

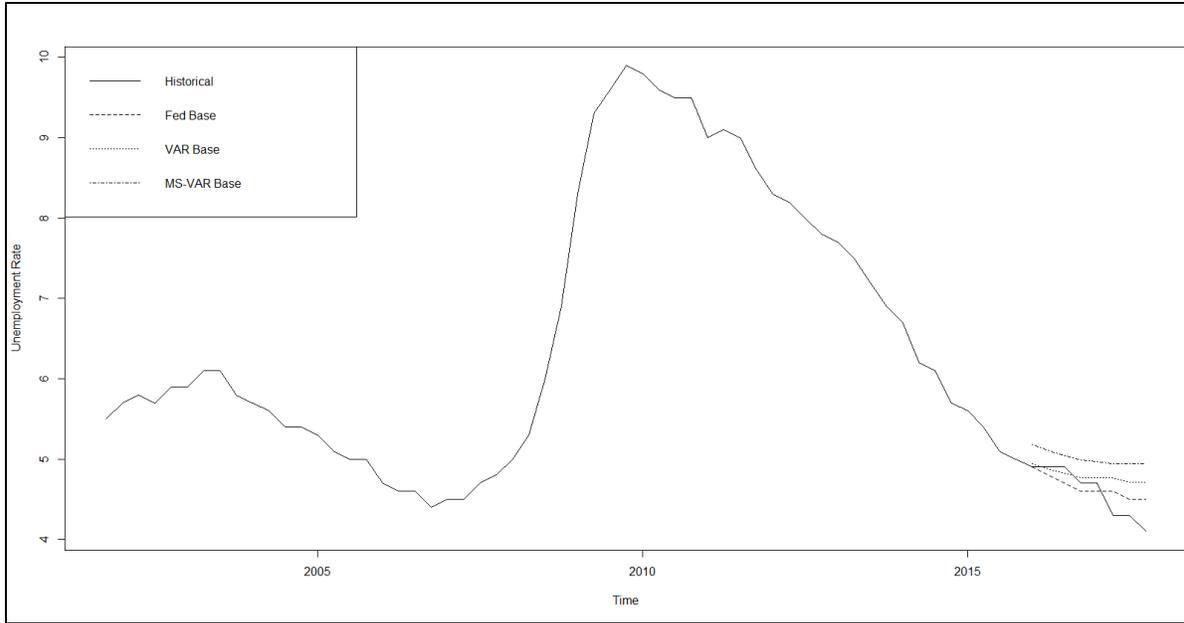


Figure 3: Time Series and Base Scenarios - Commercial Real Estate Index (Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

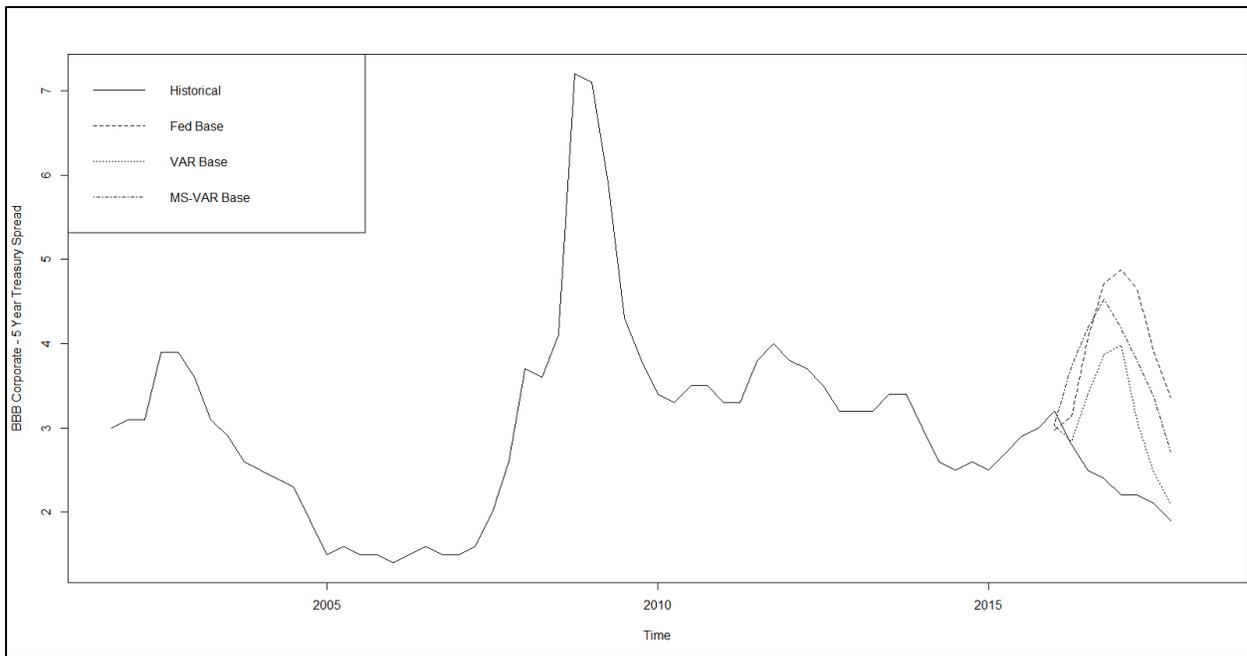


Figure 4: Time Series and Base Scenarios - BBB Corporate Bond minus 5 Year Treasury Bond Spread (Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

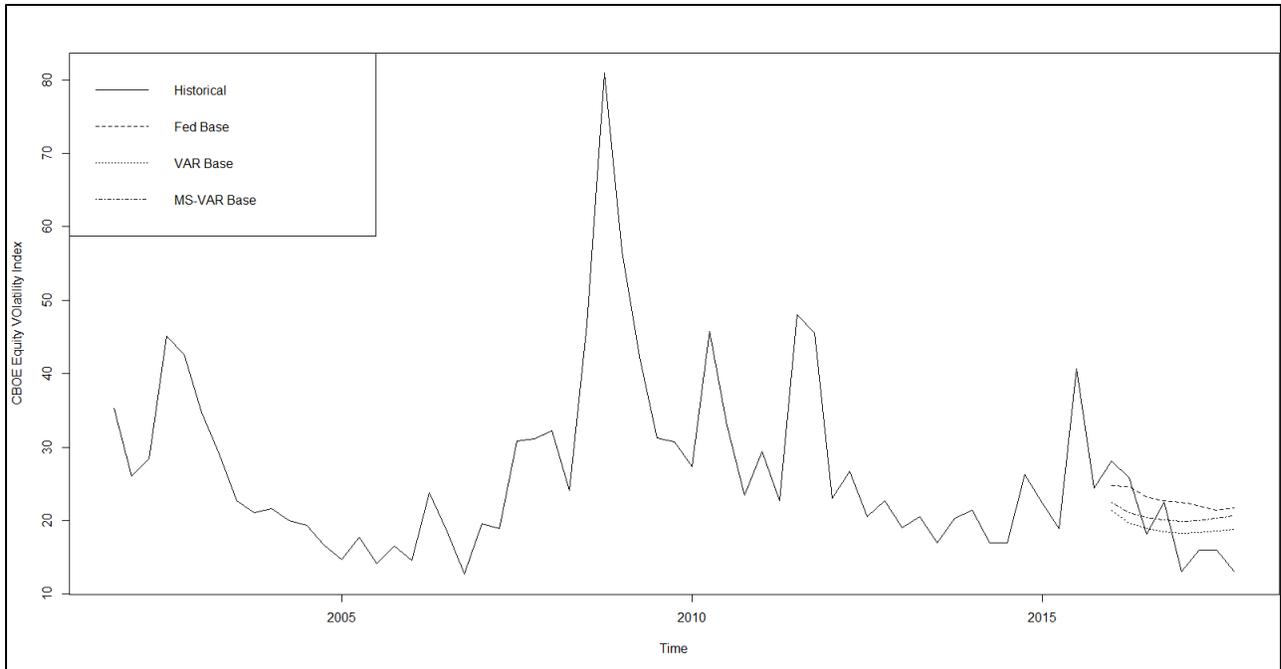


Figure 5: Time Series and Base Scenarios - CBOE Equity Market Volatility Index (Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

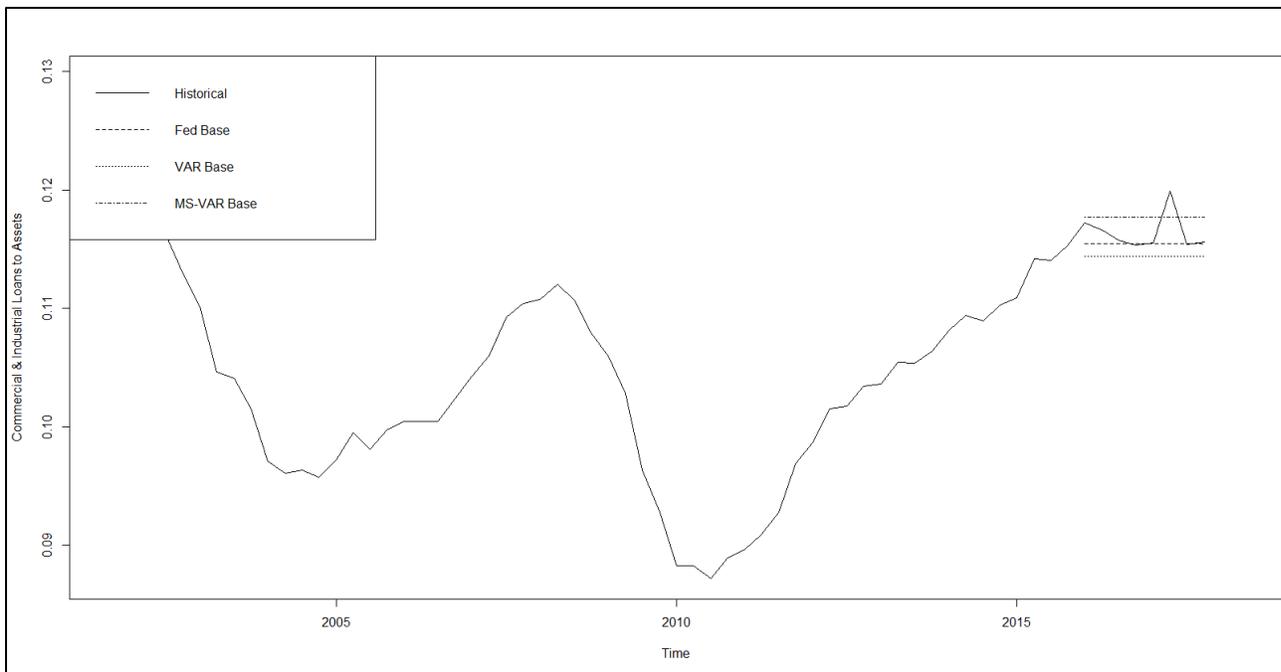


Figure 6: Time Series and Base Scenarios - Commercial and Industrial Loans to Total Assets (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

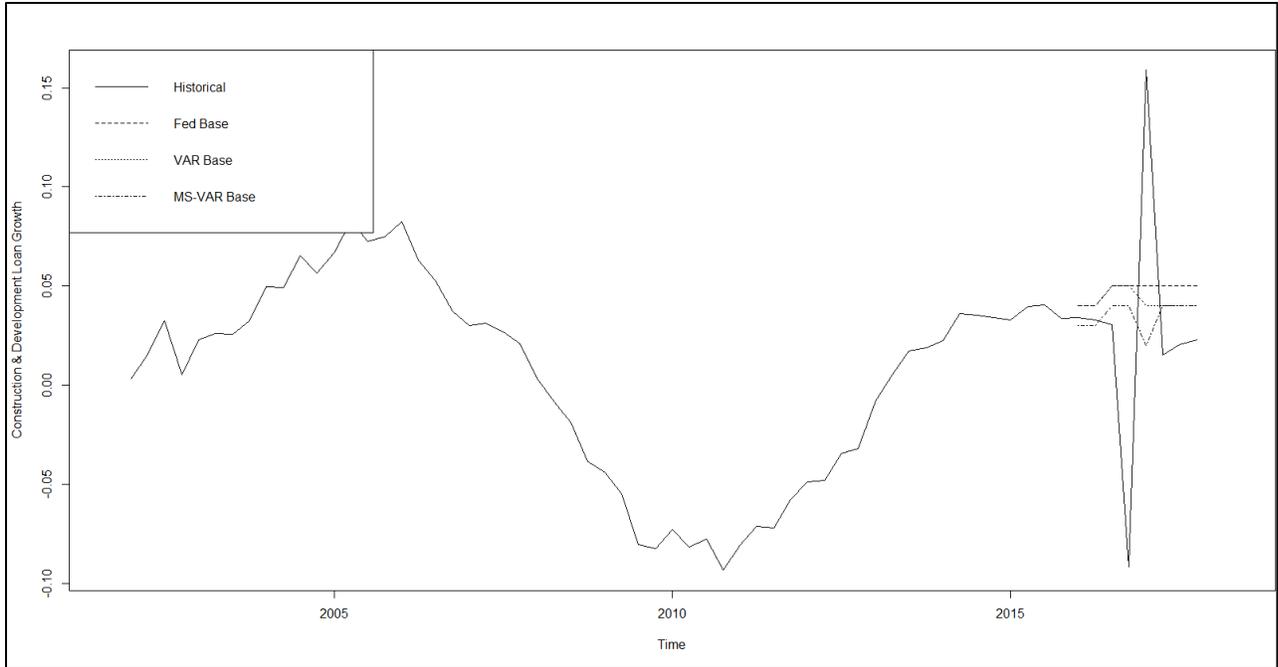


Figure 7: Time Series and Base Scenarios – Commercial and Industrial Loan Growth (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

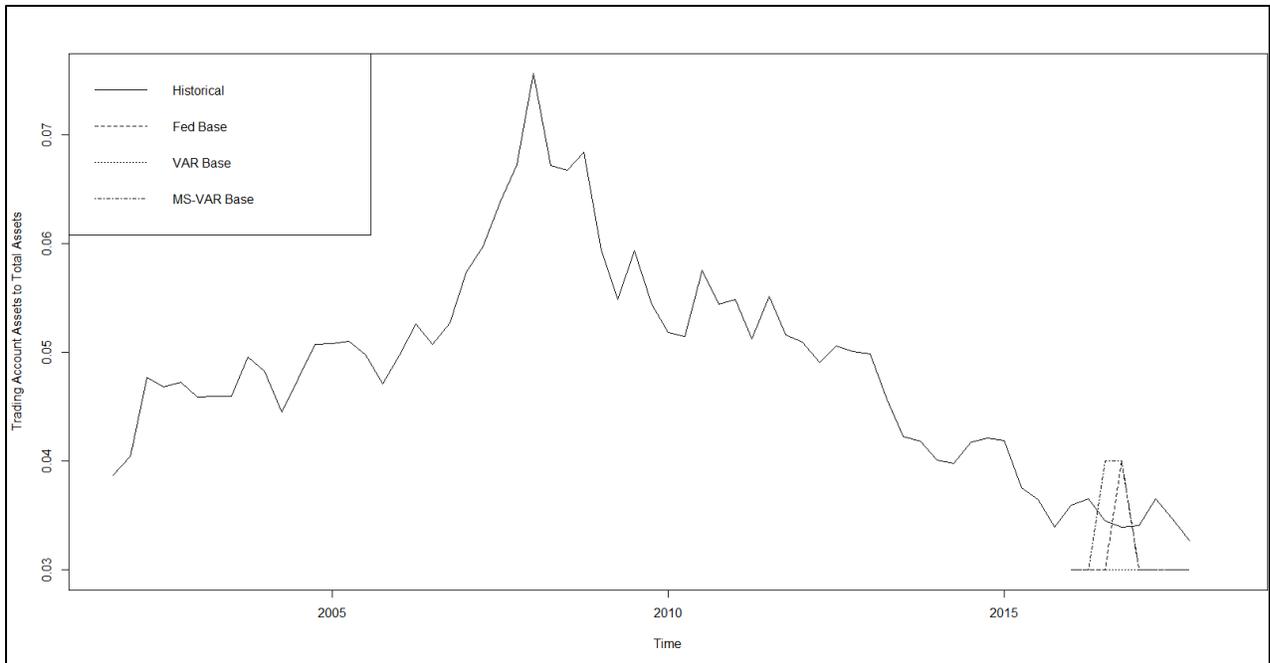


Figure 8: Time Series and Base Scenarios – Trading Account Assets to Total Assets (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

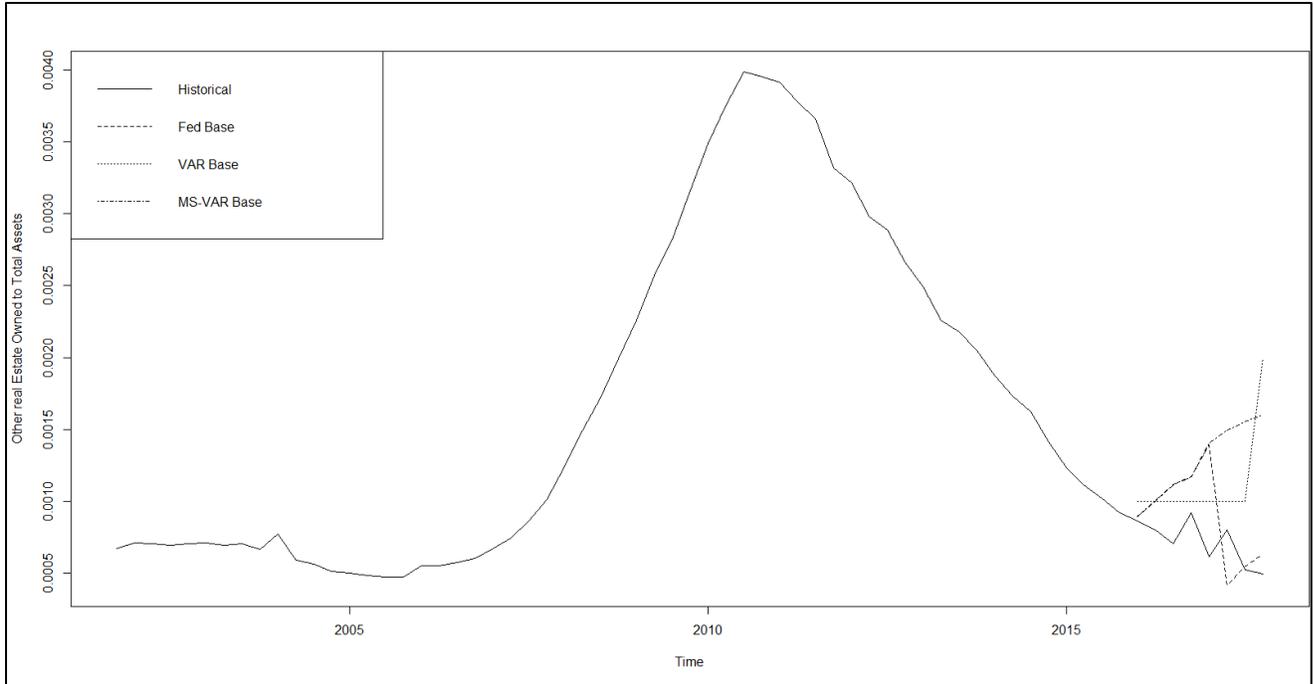


Figure 9: Time Series and Base Scenarios – Other Real Estate Owned to Total Assets (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

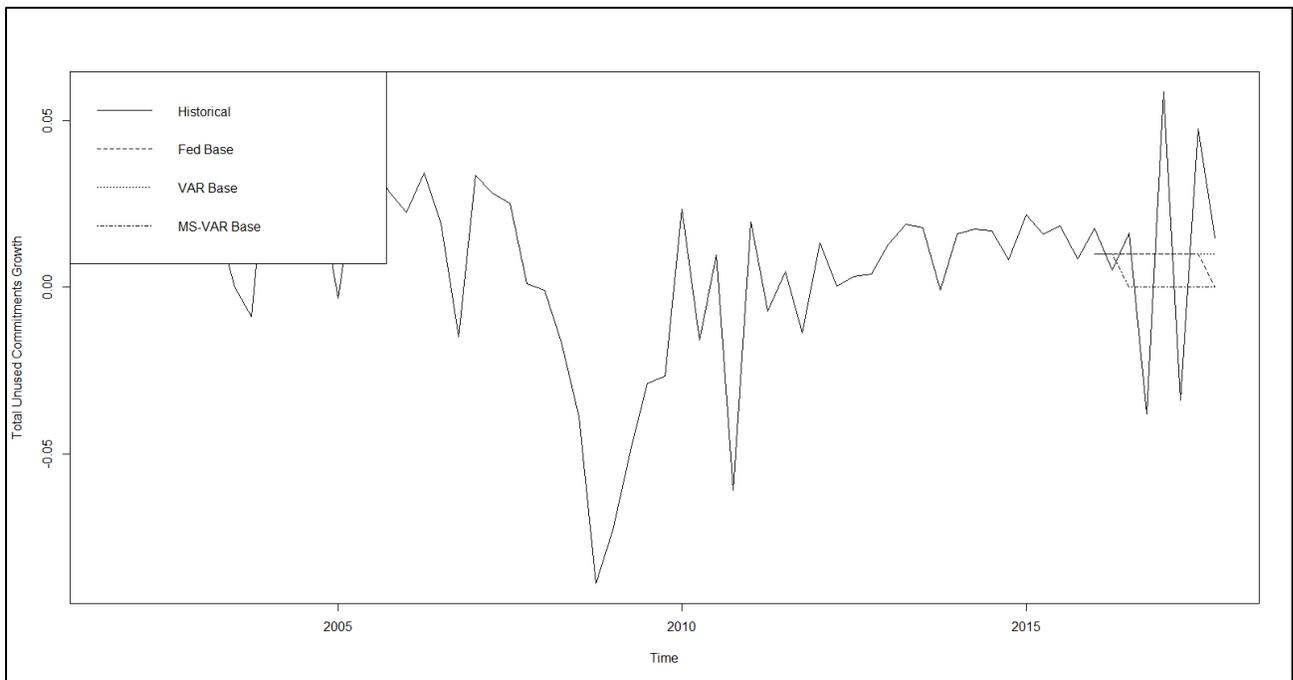


Figure 10: Time Series and Base Scenarios – Total Unused Commitment Loan Growth (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

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. Let us first consider the correlations of the macroeconomic and idiosyncratic variables to the NCOR target variable. The macroeconomic variables that are indicative of declines in economic activity (UNEMP, BBBCY, VIX and CORPSPR) all have substantial positive correlations with NCOR (0.01, 0.52, 0.64 and 0.70, respectively), while the single macroeconomic factor that reflects an improving environment (CREPI) has a large negative correlation with NCOR (-0.80). In the case of the idiosyncratic variables, we do not have a strong prior expectation from economic theory, but the signs of the correlations can be rationalized. Consistent with banks seeking to reduce loan exposures during turbulent economic periods, we find that CILTA, CDLGR and TUCGR are inversely related to NCOR, having respective correlations of -0.58, -0.82 and -0.65. On the other hand, we find TAATA to increase in downturns, having as correlation of 0.62, which can be explained by banks putting on hedging trades, as well as the illiquidity of markets which can augment the size of trading portfolios as banks sit on more marketable assets. We observe a similar phenomenon with OROTA, having a positive correlation with NCOR of 0.69, in that in downturn periods banks are stuck with larger amounts of distressed or otherwise non-marketable real estate assets.

Considering the correlations within and across the sets of macroeconomic and idiosyncratic variables, we note that signs are all economically intuitive, and while magnitudes are material they are not so high in order to result in concerns about multicollinearity. First, considering the cross-correlations between the macroeconomic and idiosyncratic variables, we note that as expected pairs of variables that have the same relationship to the credit environment are positively correlated, and vice versa for those with opposite relationships. Highlighting some of the stronger relationships, the UNEMP/CDLGR, CREPI/ROTA and CORPSPR/ROTA pairs have correlations of -0.80, -0.66 and 0.65, respectively. Second, considering the inter-group correlations between the idiosyncratic variables, we note that as expected pairs of variables that have the same relationship to the credit environment are positively correlated, and vice versa for those with opposite relationships. Highlighting some of the stronger relationships, the TUCGR/CDLGR, TUCGR/ROTA and CILTA/ROTA pairs have correlations of 0.61, -0.59 and -0.59, respectively. Finally, considering the inter-group correlations between the macroeconomic variables, we note that as expected variables that have the same relationship to the credit environment are positively correlated, and vice versa for those with opposite relationships. Highlighting some of the stronger relationships, the CORPSPR/VIX, CORPSPR / UNEMP and VIX/BBBCY pairs have correlations of 0.81, 0.61 and 0.60, respectively.

A critical modeling consideration for the MS-VAR estimation is the choice of process generation distributions for the normal and the stressed regimes. As described in the summary statistics of Jacobs (2018a), we find that when analyzing the macroeconomic data in percent change form, there is considerable skewness in the direction of adverse changes (i.e., right skewness for variables where increases denote deteriorating economic conditions such as UNEMP). Furthermore, in normal regimes where percent changes are small we find a normal distribution to adequately describe the error distribution, whereas when such changes are at extreme levels in the adverse direction we find that a log-normal distribution does a good job of characterizing the data generating process.⁹ Another important modeling consideration with respect to scenario generation is the methodology for partitioning the space of scenario paths across our 10 macroeconomic and idiosyncratic variables for the Base scenario. In the case of the Base scenario, we take an average across all paths in a given quarter for a given variable. The scenarios are shown in Figures 10 to

⁹ This is similar to the findings of Loregian and Meucci (2016) and Jacobs (2017) in the context of modelling U.S. Treasury yields. We observe that this mixture well characterizes the empirical distributions of the data in this paper.

14 where we show for each macroeconomic variable the Base scenarios for the VAR and MS-VAR models¹⁰, and also compare this to the corresponding Fed scenarios, along with the historical time series.

The VAR (1) estimation results of our CECL models are summarized in Table 2. We identify 14 optimal models according to the criteria discussed previously, 7 combinations of macroeconomic variables (4 bivariate and 3 trivariate specifications), as well as versions of these incorporating idiosyncratic variables:

- Model 1: Macroeconomic - UNEMP & BBBCY; Idiosyncratic - none
- Model 2: Macroeconomic - UNEMP & BBBCY; Idiosyncratic - CILTA & CDLGR
- Model 3: Macroeconomic - UNEMP & CREPI; Idiosyncratic - none
- Model 4: Macroeconomic - UNEMP & CREPI; Idiosyncratic - TAAA & CDLGR
- Model 5: Macroeconomic - UNEMP & CORPSPR; Idiosyncratic - none
- Model 6: Macroeconomic - UNEMP & CORPSPR; Idiosyncratic - OROTA
- Model 7: Macroeconomic - CREPI & VIX; Idiosyncratic - none
- Model 8: Macroeconomic - CREPI & VIX; Idiosyncratic - TAAA & OROTA
- Model 9: Macroeconomic - CORPSPR , UNEMP & VIX; Idiosyncratic - none
- Model 10: Macroeconomic - CORPSPR , UNEMP I & VIX; Idiosyncratic - TUCGR
- Model 11: Macroeconomic - BBBCY , UNEMP & CREPI; Idiosyncratic - none
- Model 12: Macroeconomic - BBBCY , UNEMP & CREPI; Idiosyncratic - CDLGR
- Model 13: Macroeconomic - BBBCY , UNEMP & CORPSPR; Idiosyncratic - none
- Model 14: Macroeconomic - BBBCY , UNEMP & CORPSPR; Idiosyncratic - TAAA

As we can see in Table 2, all of the candidate models satisfy our basic requirements of model fit and intuitive sensitivities. In terms of model fit, as measured by Adjusted R-Squared (“AR2”), this metric ranges from 87% to 97% across models, so that fits are good by industry standard and broadly comparable. The best fitting model is number 2, the bivariate macroeconomic specification with UNEMP and BBBCY, with idiosyncratic variables CILA and CDLG, having an AR2 of 97.7%. The worst fitting model is number 7, the bivariate macroeconomic specification with CREPI and VIX, with no idiosyncratic variables, having an AR2 of 87.5%. The autoregressive coefficient estimates all show that the NCORs display significant mean reversion, having values ranging from -0.89 to -0.65 and averaging -0.77. The coefficient estimates of the sensitivity to the macroeconomic factor UNEMP exhibits a moderate degree of responsiveness in the NCOR, having values ranging widely from 0.0003% to 0.23% and averaging 0.04%. The coefficient estimates of the sensitivity to the macroeconomic factor BBBCY exhibits a moderate degree of responsiveness in the NCOR, having values ranging narrowly from 0.02% to 0.04% and averaging 0.03%. The coefficient estimates of the sensitivity to the macroeconomic factor BBBCY exhibits a high degree of responsiveness in the NCOR, having values ranging narrowly from 0.20% to 0.43% and averaging 0.28%. The coefficient estimates of the sensitivity to the macroeconomic factor CORPSPR exhibits a moderate de-

¹⁰ Estimation results for the VAR and MS-VAR model are available upon request. The models are all convergent and goodness of fit metrics in with industry standards. Signs of coefficient estimates are in line with economic intuition and estimates are all significant at conventional levels. We use the `dse`, `tseries` and `MSBVAR` libraries in R in order to perform the estimations (R Development Core Team, 2019).

gree of responsiveness in the NCOR, having values ranging widely from 0.0002% to 0.05% and averaging 0.06%.

The coefficient estimates of the sensitivity to the macroeconomic factor VIX exhibits a moderate degree of responsiveness in the NCOR, having values ranging narrowly from 0.04% to 0.06% and averaging 0.06%. The idiosyncratic variable CILA appears in only one model, number 2 the bivariate macroeconomic specification with UNEMP and BBBCY, exhibiting a high sensitivity in the NCOR of -0.92%. The coefficient estimates of the sensitivity to the idiosyncratic variable CDLG exhibits a high degree of responsiveness in the NCOR, having values ranging moderately from -0.78% to -0.54% and averaging -0.66%. The coefficient estimates of the sensitivity to the idiosyncratic variable TAAA exhibits a high degree of responsiveness in the NCOR, having values ranging narrowly from 3.59% to 4.95% and averaging 4.43%. The coefficient estimates of the sensitivity to the idiosyncratic variable OREOTA exhibits a high degree of responsiveness in the NCOR, having values ranging widely from 0.13% to 9.22% and averaging 4.60%. The idiosyncratic variable TUCG appears in only one model, number 10 (the trivariate macroeconomic specification with VIX, UNEMP & CORPSPR), exhibiting a moderate sensitivity in the NCOR of -0.60%.

The model performance metrics of the 14 models are presented in Table 3. We tabulate these statistics for the development sample 4Q01-4Q15, and out-of-sample 1Q16-4Q17, and the latter is evaluated under the three scenario generation models (Fed, VAR as MS-VAR) as well as perfect foresight (i.e., assuming we anticipated the actual paths of the macroeconomic and idiosyncratic variables). We also depict graphically the in- and out-of-sample fits in the model accuracy plots of Figures 11 through 24. We consider the following industry standard model performance metrics:

- The model performance metrics of the 14 models are presented in Table 3. We tabulate these statistics for the development sample 4Q01-4Q15, and out-of-sample 1Q16-4Q17, and the latter is Generalized Cross-Validation (“GCV”)
- Squared Correlation (“SC”)
- Root Mean Squared Error (“RMSE”)
- Cumulative Percent Error (“CPE”)
- Akaike Information Criterion (“AIC”)

We observe from Table 3 and from Figures 11 through 24 that there is a great diversity of out-of-sample model performance across econometric specifications as well as across models for baseline economic forecasts, both in absolute terms and in comparison to the perfect foresight benchmark. We further note that in summary, out-of-sample performance across models is not assessed as being favorable by industry standards, even in the perfect foresight scenario. Measures of model (GCV, SC, RMSE & AIC) fit tend to be poor, and there is chronic underprediction of NCORs according to the CPE metric. For example, the average SC in-sample is 93.7% ranging in 88.9% to 87.1%, but out-of-sample it averages 17.2% and ranges in 0.06% to 87.3%. On the other hand, in sample the CPE averages 0.04% and ranges in -0.09% to 0.21%, while out of sample the average is -43.0% and ranges in -158.9% to 48.9%.

While this is not conclusive, if we assume that we are near the end of an economic expansion, these results imply that most banks models will under-provision as we enter a downturn, which is evidence that the

Table 3: CECL Vector Autoregressive Model In- and Out-of-Sample Performance Measures Compared - Historical Y9 Credit Loss Rates, Banking System Idiosyncratic and Macroeconomic Variables (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

Time Period	Statistic	Model														Summary Statistics - Scenario					
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	Minimum	Average	Maximum	Range		
Development Sample	Generalized Cross Validation	1.2E+10	1.2E+10	1.2E+10	9.2E+11	1.3E+10	1.0E+10	1.7E+10	9.9E+09	1.47E+10	1.0E+10	1.2E+10	1.2E+10	6.02E+11	1.0E+10	6.02E+11	1.27E+10	1.2E+10	1.3E+10	1.3E+10	1.3E+10
	Squared Correlation	93.14%	97.05%	90.00%	95.88%	88.87%	95.34%	89.95%	95.72%	93.98%	6.63E+04	6.33E+04	92.81%	96.82%	93.27%	95.31%	88.87%	93.68%	97.70%	97.05%	81.88%
	Root Mean Squared Error	6.44E+04	6.58E+04	7.2E+04	5.12E+04	8.20E+04	5.56E+04	7.88E+04	5.48E+04	6.63E+04	6.33E+04	6.67E+04	4.51E+04	6.21E+04	5.67E+04	4.51E+04	6.38E+04	6.48E+04	8.20E+04	3.68E+04	6.48E+04
	Cumulative Percentage Error	-0.018%	0.0238%	0.0243%	0.024%	-0.0016%	-0.0985%	-0.027%	0.020%	0.000%	0.120%	0.041%	-0.020%	0.217%	0.041%	-0.0985%	0.039%	0.024%	0.217%	0.217%	0.117%
	Alatake Information Criterion	-493.36	-589.47	-591.64	-581.07	-584.94	-589.18	-590.83	-579.95	-569.62	-561.36	-596.51	-611.67	-698.59	-576.33	-691.67	-589.11	-589.31	-589.62	-569.62	71.02
	Generalized Cross Validation	1.5E+08	1.5E+08	1.5E+09	8.41E+09	3.26E+10	1.0E+08	1.4E+08	2.87E+08	6.5E+09	1.0E+08	8.1E+09	2.7E+10	7.0E+08	4.2E+08	2.7E+10	2.0E+08	1.5E+08	1.2E+08	7.0E+08	2.0E+08
	Squared Correlation	0.39%	3.75%	5.21%	7.92%	50.39%	0.001%	10.07%	42.8%	30.34%	19.74%	1.0%	68.44%	2.6%	6.57%	0.06%	17.72%	7.24%	68.44%	68.38%	0.002%
	Root Mean Squared Error	0.0022	0.0020	0.0004	0.0013	0.0002	0.0012	0.0012	0.0024	0.0010	0.0015	0.0010	0.0002	0.0029	0.0029	0.0002	0.0013	0.0013	0.0030	0.0030	0.0029
	Cumulative Percentage Error	-85.7%	-90.4%	-94.6%	-90.7%	4.83%	-40.37%	38.12%	-87.2%	29.22%	33.57%	-32.6%	5.07%	-103.82%	-67.45%	-103.82%	-25.70%	-31.53%	38.12%	141.93%	40.02%
	Alatake Information Criterion	-79.04	-82.25	-106.49	-88.38	-118.99	-88.32	-88.82	-79.16	-92.31	-88.82	-91.26	-116.49	-73.96	-76.62	-118.99	-90.65	-88.57	-73.96	-45.69	45.69
	Generalized Cross Validation	4.02E+08	2.2E+08	6.12E+10	1.7E+08	4.9%	2.9E+08	6.47E+08	1.8E+07	7.2E+08	1.2E+07	1.6E+08	4.67E+09	2.0E+08	9.0E+08	6.12E+10	4.62E+04	2.3E+08	6.47E+08	6.47E+08	87.11%
	Squared Correlation	80.23%	87.95%	21.7%	28.95%	4.9%	4.9%	0.65%	0.1%	0.7%	0.44%	81.9%	2.0%	1.0%	9.72%	0.1%	23.1%	4.7%	87.95%	87.20%	87.11%
	Root Mean Squared Error	0.020	0.021	0.002	0.010	0.003	0.018	0.006	0.006	0.002	0.0050	0.013	0.010	0.010	0.0042	0.002	0.004	0.004	0.009	0.0060	0.005
	Cumulative Percentage Error	-80.0%	-58.89%	-7.41%	-48.72%	7.37%	-61.04%	-70.49%	-151.94%	-78.93%	-100.00%	-50.58%	6.13%	48.50%	-35.41%	-151.94%	-48.69%	-54.72%	48.50%	200.74%	200.74%
Alatake Information Criterion	-80.46	-81.22	-113.94	-83.24	-109.23	-82.07	-77.80	-64.25	-72.21	-67.20	-85.38	-93.68	-81.88	-69.97	-113.94	-81.25	-81.65	-64.25	-49.69	49.69	
Generalized Cross Validation	6.7E+08	4.2E+08	4.8E+10	1.69E+08	1.1E+09	1.1E+08	6.7E+08	2.0E+07	8.63E+08	1.37E+07	2.7E+08	4.5E+09	1.3E+07	4.6E+08	1.3E+07	5.71E+08	3.4E+08	2.0E+07	2.0E+07	45.64%	
Squared Correlation	23.24%	27.82%	11.7%	28.67%	9.58%	4.51%	0.1%	0.74%	0.34%	0.34%	45.74%	9.87%	12.45%	7.33%	0.10%	13.25%	9.72%	45.74%	45.64%	0.06%	
Root Mean Squared Error	0.0026	0.0029	1.6%	0.0018	0.003	0.012	0.0025	0.003	0.0034	0.0032	0.0019	0.0007	0.008	0.0051	0.0002	0.002	0.0022	0.003	0.0061	0.0061	
Cumulative Percentage Error	-101.23%	-80.19%	-1.6%	-47.15%	9.03%	-40.2%	-74.9%	-158.10%	-83.35%	-104.82%	-64.7%	3.7%	25.7%	-58.98%	-158.10%	-56.7%	-91.83%	25.70%	183.88%	183.88%	
Alatake Information Criterion	-76.25	-76.12	-116.14	-83.38	-108.71	-88.53	-76.91	-158.10%	-83.35%	-104.82%	-64.7%	3.7%	25.7%	-58.98%	-158.10%	-56.7%	-91.83%	25.70%	183.88%	183.88%	
Generalized Cross Validation	2.07E+08	1.5E+08	3.2E+09	1.17E+08	3.0E+09	6.33E+09	4.6E+08	1.4E+07	8.1E+08	1.41E+07	3.87E+09	1.7E+08	3.4E+08	7.8E+08	3.0E+09	4.2E+08	1.9E+08	1.5E+08	1.3E+07	6.47E+08	
Squared Correlation	46.24%	42.97%	0.63%	7.71%	4.62%	0.43%	0.25%	2.10%	0.23%	0.44%	0.2%	1.14%	65.05%	12.34%	14.65%	1.63%	85.05%	85.05%	84.81%	84.81%	
Root Mean Squared Error	0.0014	0.0017	0.0006	0.0015	0.0006	0.0091	0.0021	0.0052	0.0033	0.0007	0.0007	0.0019	0.0007	0.0021	0.0006	0.0022	0.0018	0.003	0.0047	0.0047	
Cumulative Percentage Error	-53.86%	-45.64%	13.03%	-13.20%	14.93%	-30.48%	-65.18%	-135.79%	-80.27%	-105.85%	-10.69%	37.41%	-72.95%	-28.88%	-135.79%	-41.45%	-38.06%	37.41%	173.20%	34.46	
Alatake Information Criterion	-85.91	-84.37	-100.61	-86.35	-101.00	-91.25	-79.31	-66.50	-72.85	-66.42	-97.20	-85.11	-79.71	-83.72	-101.00	-83.72	-84.74	-66.42	-66.42	34.46	
Generalized Cross Validation	2.20E+10	1.1E+04	2.4E+08	5.97%	1.2E+08	5.97%	1.2E+08	5.97%	1.2E+08	5.97%											
Root Mean Squared Error	0.006	0.0021	0.0018	0.0018	0.0018	0.0018	0.0018	0.0018	0.0018	0.0018	0.0018	0.0018	0.0018	0.0018	0.0018	0.0018	0.0018	0.0018	0.0018	0.0018	
Cumulative Percentage Error	-158.10%	-49.25%	-46.32%	-46.32%	-46.32%	-46.32%	-46.32%	-46.32%	-46.32%	-46.32%	-46.32%	-46.32%	-46.32%	-46.32%	-46.32%	-46.32%	-46.32%	-46.32%	-46.32%	-46.32%	
Alatake Information Criterion	-118.99	-83.31	-83.31	-83.31	-83.31	-83.31	-83.31	-83.31	-83.31	-83.31	-83.31	-83.31	-83.31	-83.31	-83.31	-83.31	-83.31	-83.31	-83.31	-83.31	

Summary Statistics - Total Out-of-Sample

CECL standard may fail to address the problem of procyclicality that was the intent of the new accounting guidance. Among the loss model specifications, we find that generally the loss prediction specifications having a lower parameterization, as well as the MS-VAR scenario generation model, tends to produce better measures of model fit and a lower degree of under-prediction. Furthermore, relative to the perfect foresight benchmark, the MS-VAR model produces a lower level of variation in the model performance statistics across loss predictive model specifications.

First, let us discuss the detailed out-of-sample performance of the GCV measure, which we consider to be our performance leading metric. First considering scenario generation models across predictive loss specifications, we observe that the MS-VAR model performs better than the Fed and VAR models, having GCVs averaging $1.89e-09$ and ranging from $3.09e-09$ to $1.41e-07$, outperforming the average GCV of $4.62e-04$ ($4.71e-08$) and ranging in $6.12e-10$ to $6.47e-03$ ($4.65e-10$ to $2.05e-07$) in the Fed (VAR) models. Moreover, the MS-VAR model is close to the perfect foresight benchmark, where the respective mean GCV is $1.94e-10$ and ranging in $2.70e-10$ to $7.06e-08$. Turning to the relative performance of loss predictive specifications according to GCV by scenario generation model, we note that across all scenario generation models, the simpler specifications all perform better under the GCV metric. Under the MS-VAR scenario model 10 (with VIX, UNEMP, CORPSPR & TUCG) is the worst performing, while the best performing model is number 5 (with UNEMP and CORSPR). On the other hand, under the VAR scenario model 8 (with VIX, UNEMP, CORPST & TUCG) is the worst performing, while the best performing model is number 3 (with UNEMP & CREPI). Furthermore, under the FED scenario model 13 (with BBBCY, UNEMP & CORPSPR) is the worst performing, while the best performing model is number 3 (with UNEMP & CREPI). Finally, under the perfect foresight benchmark model 13 (with BBBCY, UNEMP & CORPSPR) is the worst performing, while the best performing model is number 12 (with BBBCY, UNEMP, CREPI & CDLG).

Next let us discuss the detailed out-of-sample performance of the SC measure. First considering scenario generation models across predictive loss specifications, we observe that the MS-VAR model performs better than the Fed and VAR models, having SCs averaging 14.7% and ranging from 0.10% to 85.1%, outperforming the average SC of 23.2% (13.3%) and ranging in 0.15% to 87.1% (0.10% to 45.7%) in the Fed (VAR) models. Turning to the relative performance of loss predictive specifications according to SC by scenario generation model, we note that across all scenario generation models, the simpler specifications all perform better under the SC metric. Under the MS-VAR scenario model 1 (with UNEMP & BBBCY) is the worst performing, while the best performing model is number 11 (with UNEMP, BBBCY & CREPI). On the other hand, under the VAR scenario model 6 (with UNEMP, CORPSPR & OROTA) is the worst performing, while the best performing model is number 11 (with UNEMP, BBBCY & CREPI). Furthermore, under the FED scenario model 6 (with UNEMP, CORPSPR & OROTA) is the worst performing, while the best performing model is number 2 (with UNEMP, CORPSPR & TAATA). Finally, under the perfect foresight benchmark model 6 (with UNEMP, CORPSPR & OROTA) is the worst performing, while the best performing model is number 12 (with BBBCY, UNEMP, CREPI & CDLG).

Next let us discuss the detailed out-of-sample performance of the RMSE measure. First considering scenario generation models across predictive loss specifications, we observe that the MS-VAR model performs better than the Fed and VAR models, having RMSEs averaging 0.0022 and ranging from 0.0006 to 0.0053, outperforming the average RMSE of 0.0024 (0.0025) and ranging in 0.0002 to 0.0060 (0.0002 to 0.0063) in the Fed (VAR) models. Turning to the relative performance of loss predictive specifications according to RMSE by scenario generation model, we note that across all scenario generation models, the simpler specifications all perform better under the RMSE metric. Under the MS-VAR scenario model 3 (with UNEMP & BBBCY) is the worst performing, while the best performing model is number 10 (with

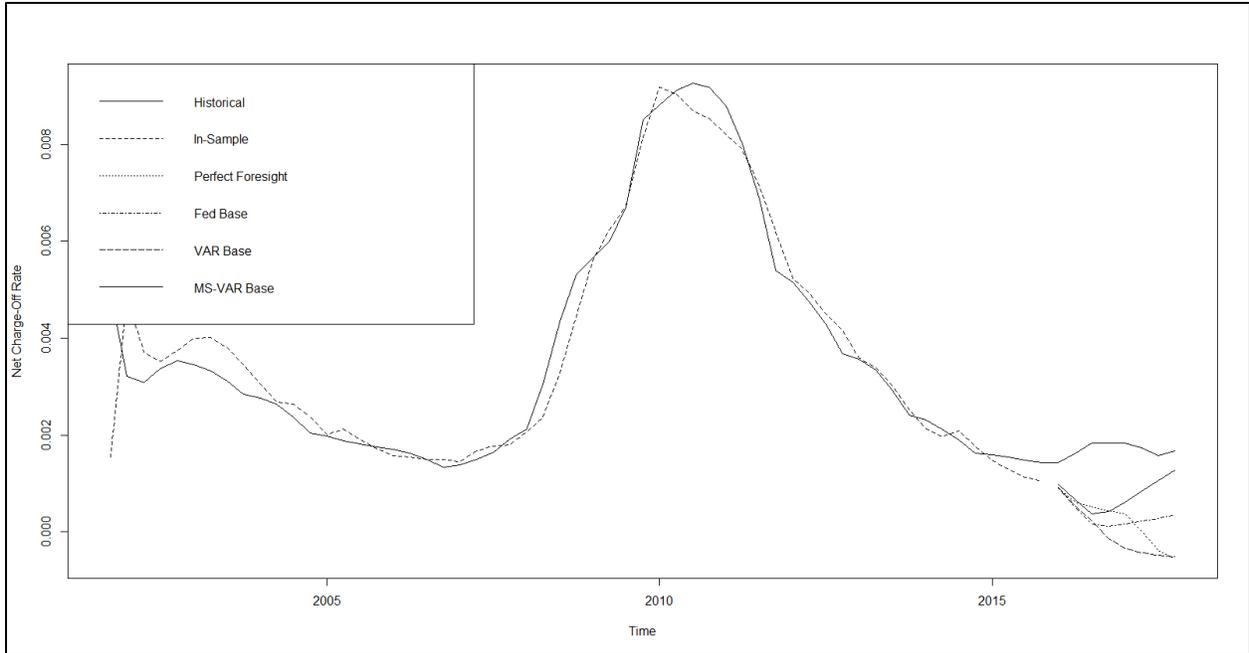


Figure 11: Net-Charge off Rate Model Accuracy Plots – Unemployment Rate and BBB Corporate Bond Yield (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

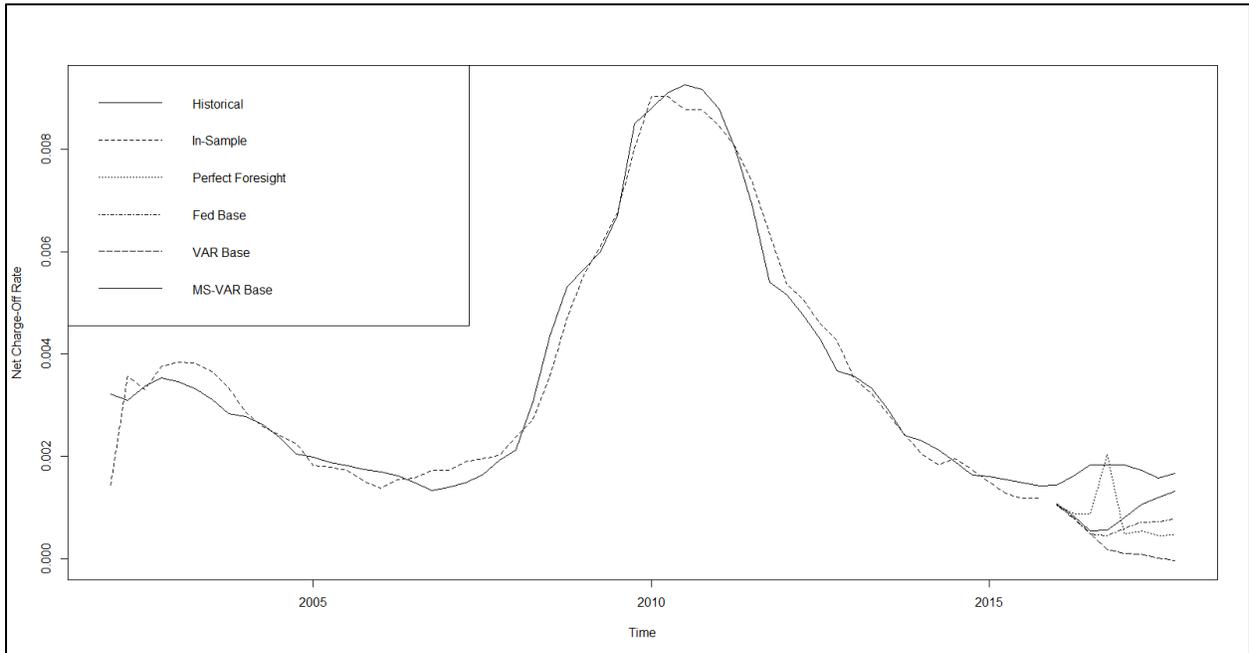


Figure 12: Net-Chargeoff Rate Model Accuracy Plots – Unemployment Rate, BBB Corporate Bond Yield, Commercial Development Loan Growth and Commercial & Industrial Loans to Total Assets (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

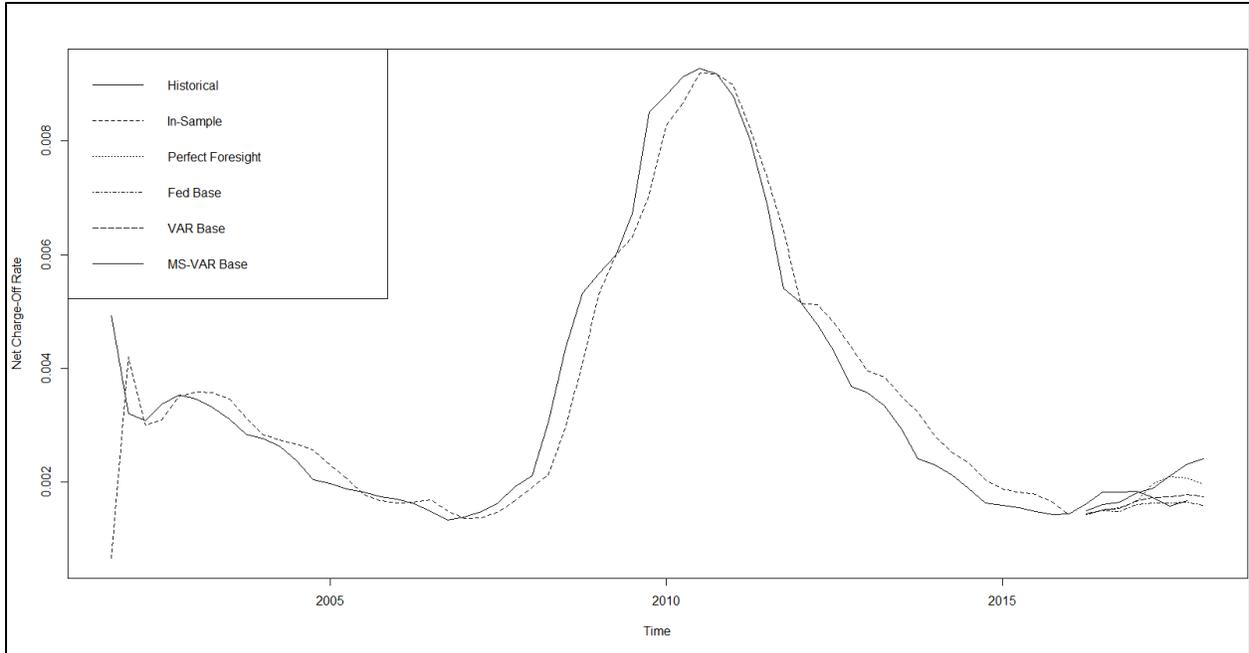


Figure 13: Net-Chargeoff Rate Model Accuracy Plots – Unemployment Rate and Commercial Real Estate Price Index (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

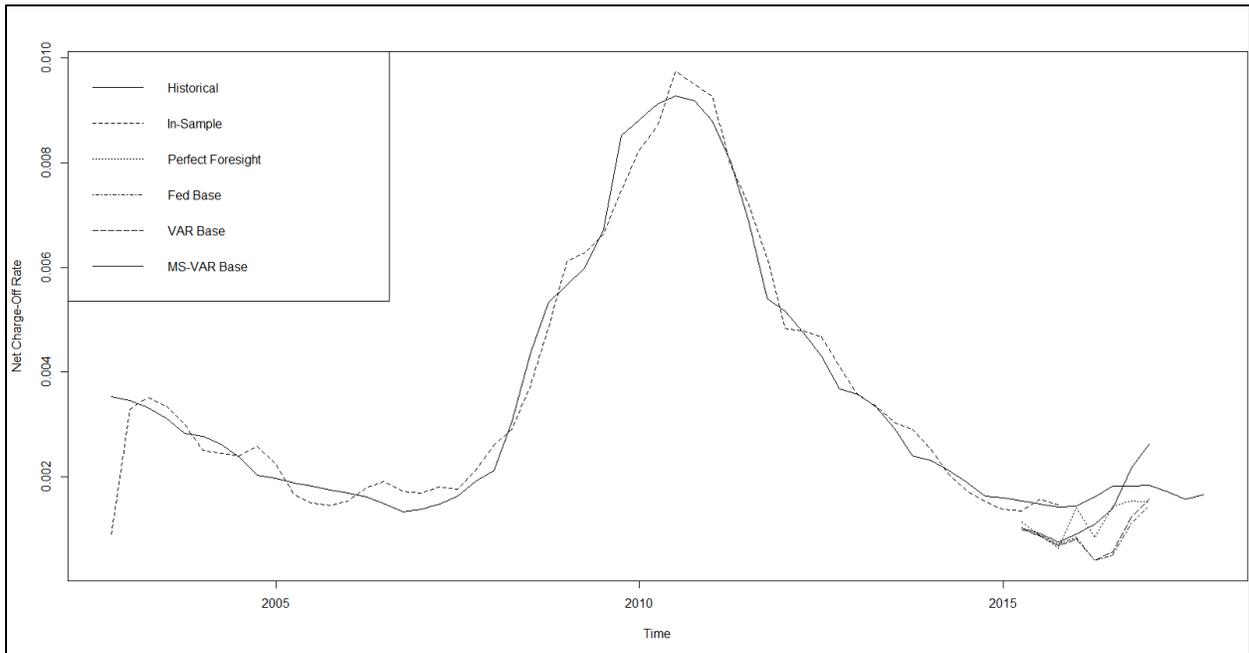


Figure 14: Net-Chargeoff Rate Model Accuracy Plots – Unemployment Rate, Commercial Real Estate Price Index, Commercial Development Loan Growth and Total Trading Account Assets to Total Assets (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

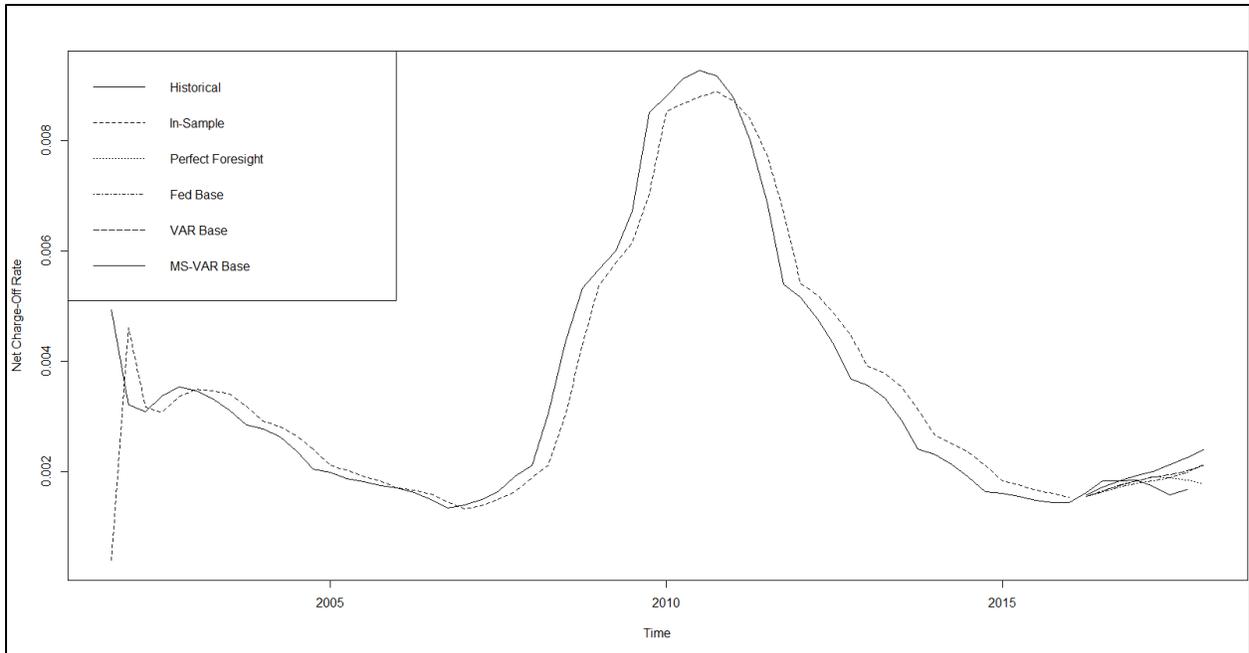


Figure 15: Net-Chargeoff Rate Model Accuracy Plots – Unemployment Rate and BBB Corporate - 5 Year Treasury Bond Spread (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

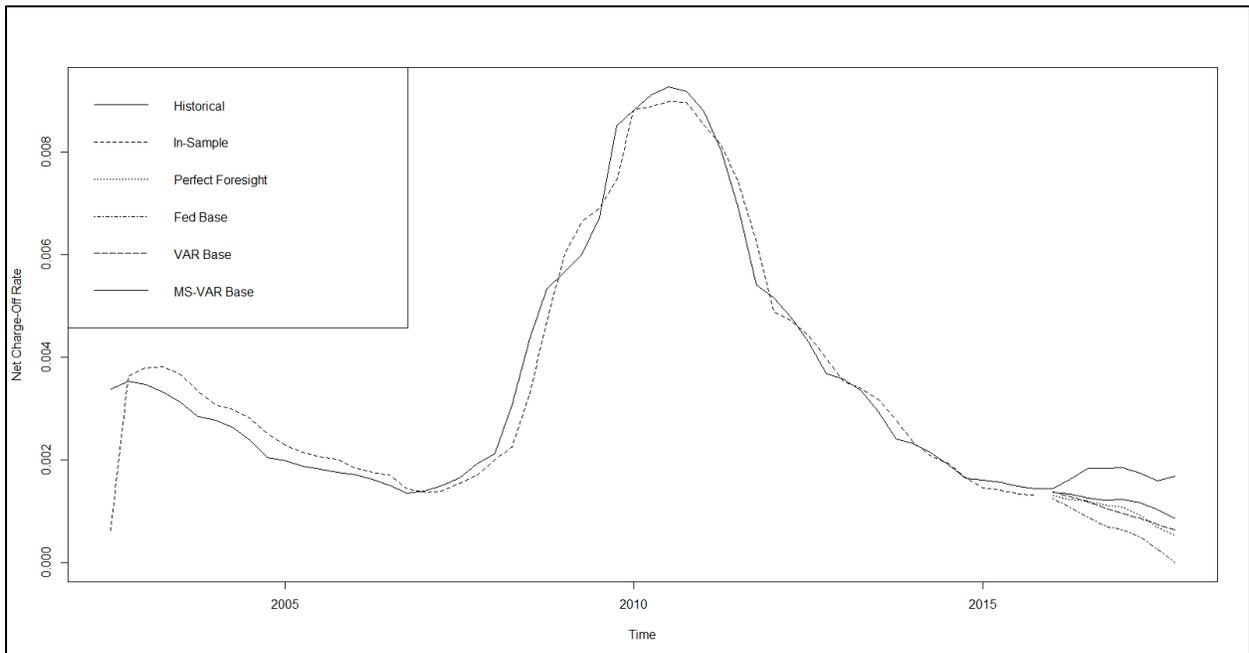


Figure 16: Net-Chargeoff Rate Model Accuracy Plots – Unemployment Rate, BBB Corporate - 5 Year Treasury Bond Spread and Other Real Estate Owned to Total Assets (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

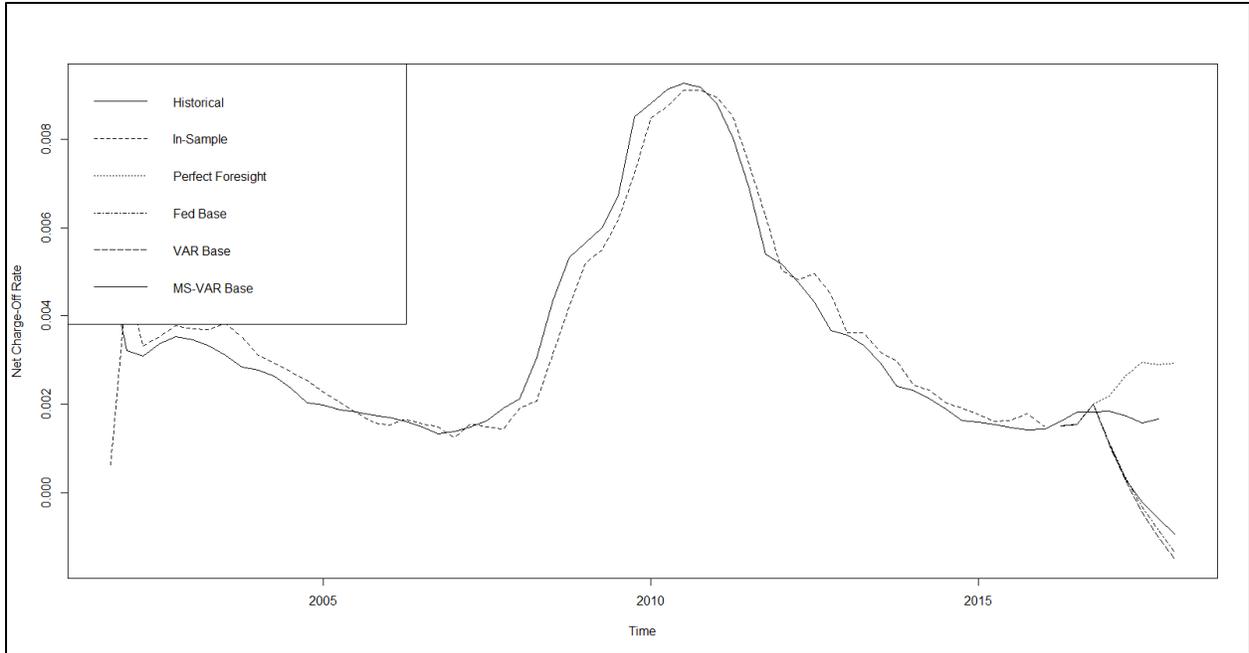


Figure 17: Net-Chargeoff Rate Model Accuracy Plots – Commercial Real Estate Price Index and CBOE Equity Volatility Index (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

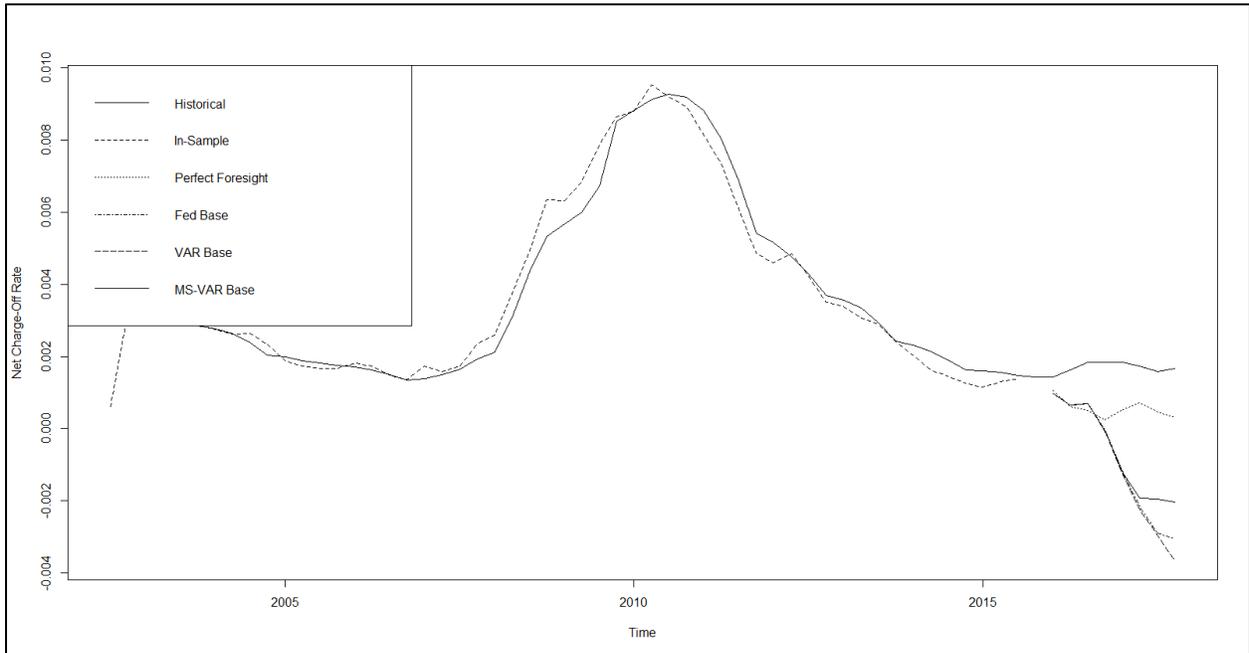


Figure 18: Net-Chargeoff Rate Model Accuracy Plots – Commercial Real Estate Price Index, CBOE Equity Volatility Index, Other Real Estate Owned to Total Assets and Total Trading Account Assets to Total Assets (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

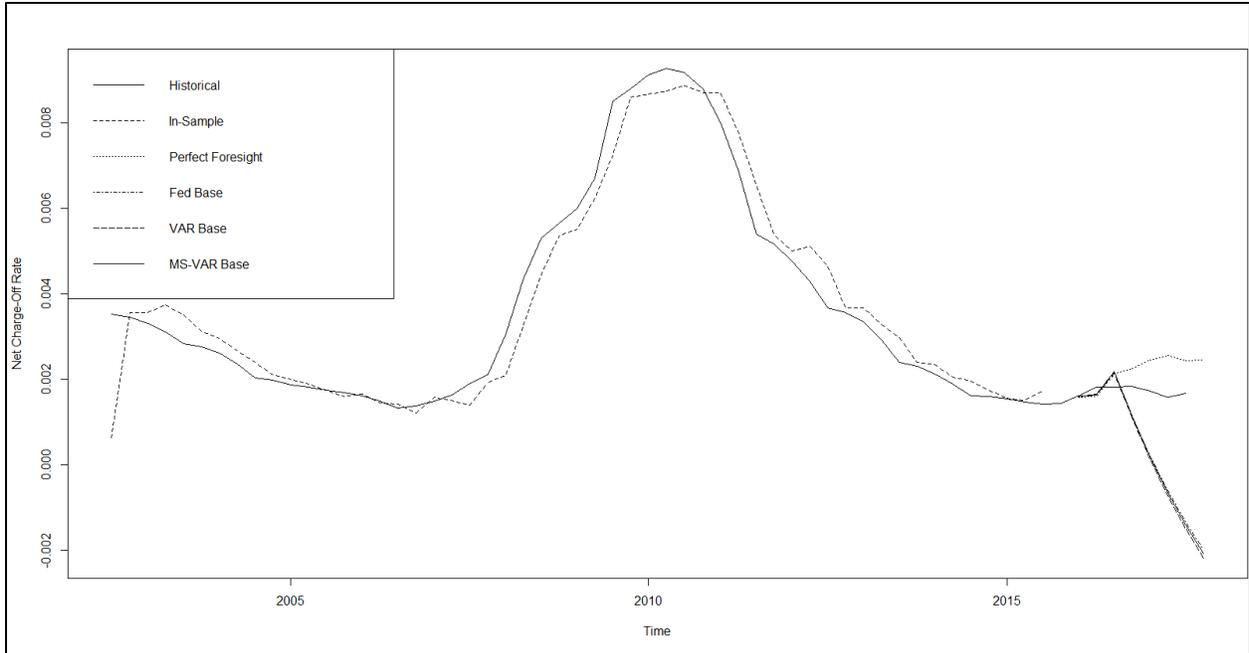


Figure 19: Net-Chargeoff Rate Model Accuracy Plots – Unemployment Rate, CBOE Equity Volatility Index and BBB Corporate - 5 Year Treasury Bond Spread (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

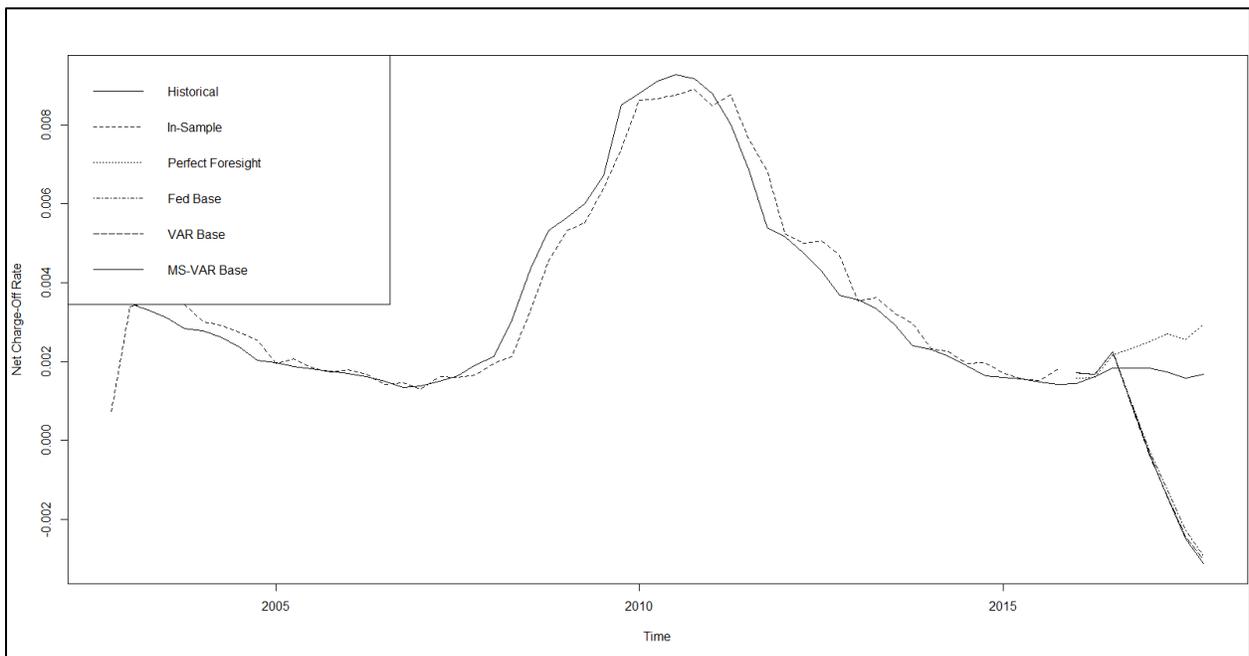


Figure 20: Net-Chargeoff Rate Model Accuracy Plots – Unemployment Rate, CBOE Equity Volatility Index, BBB Corporate - 5 Year Treasury Bond Spread and Total Uncommitted Loan Growth (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

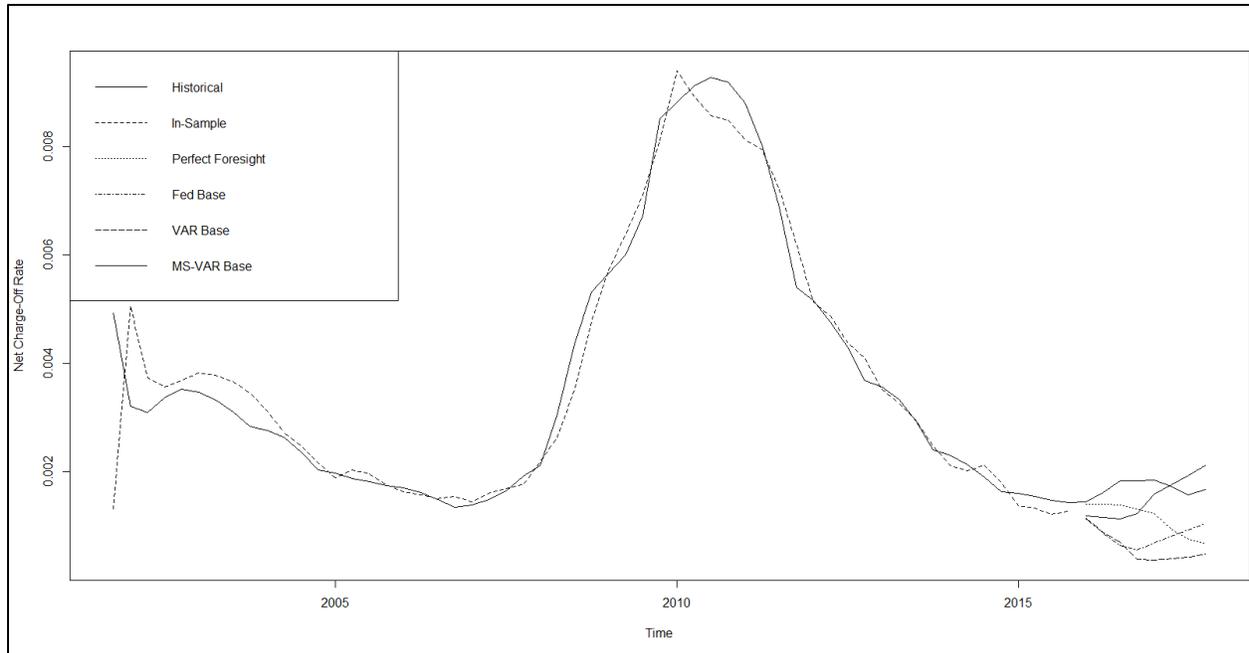


Figure 21: Net-Chargeoff Rate Model Accuracy Plots – Unemployment Rate, Commercial Real Estate Price Index and BBB Corporate Bond Yield (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

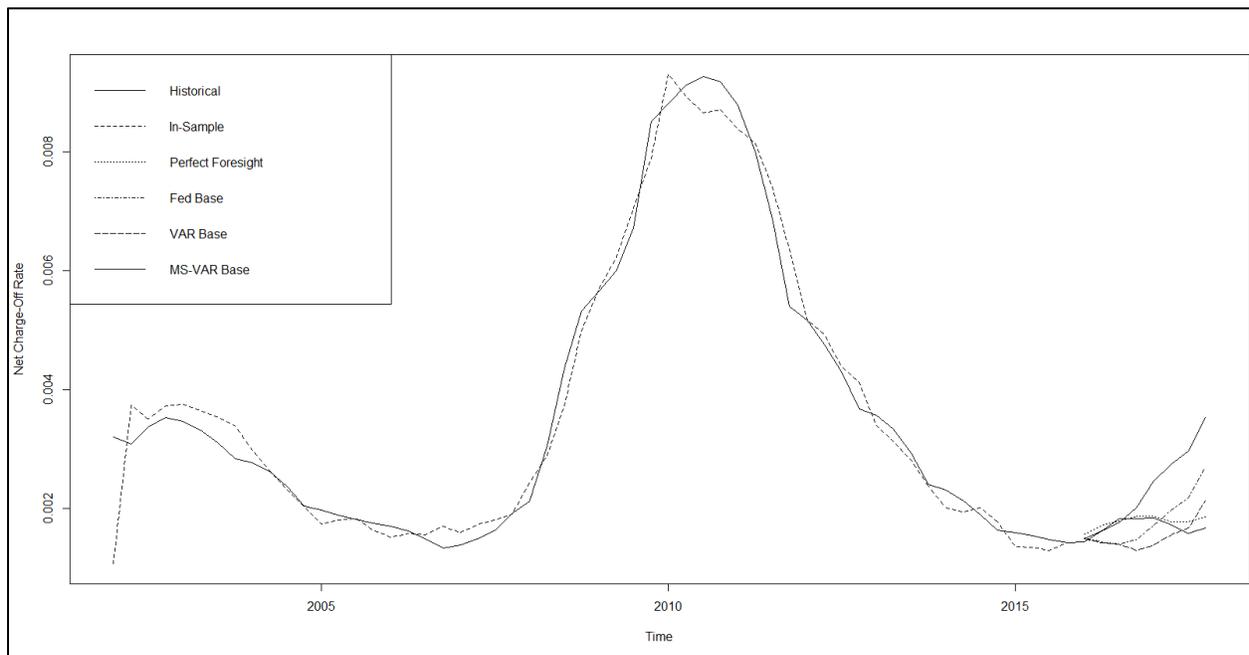


Figure 22: Net-Chargeoff Rate Model Accuracy Plots – Unemployment Rate, Commercial Real Estate Price Index, BBB Corporate Bond Yield and Commercial Development Loan Growth (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

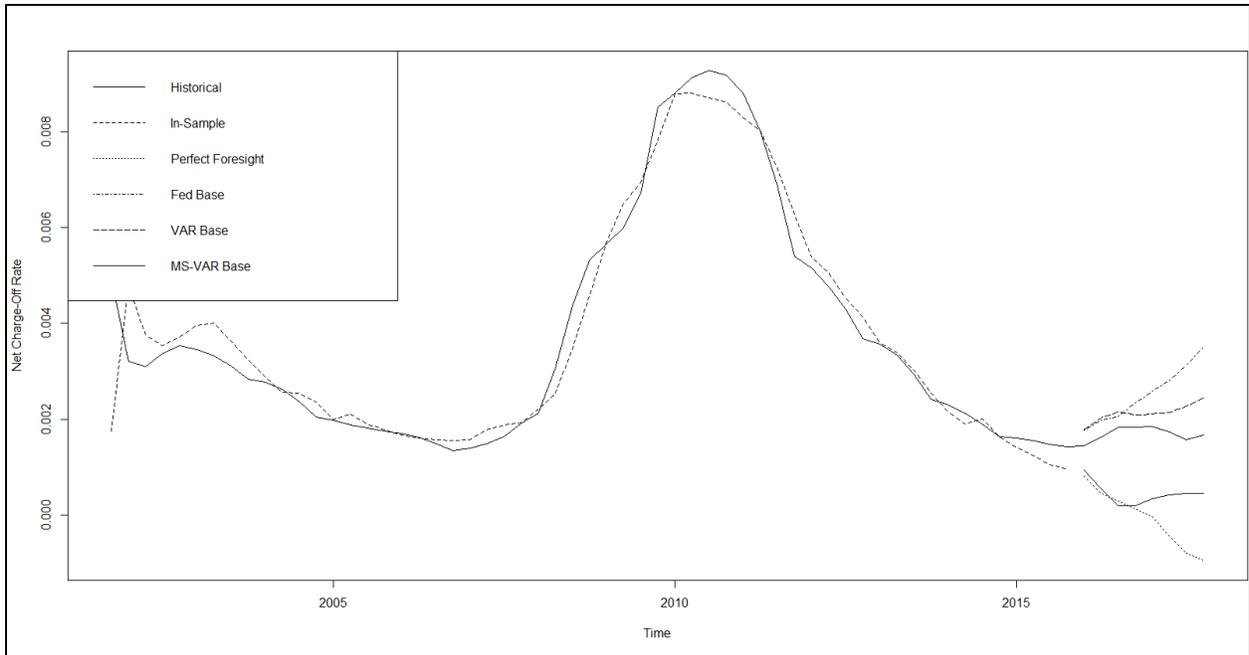


Figure 23: Net-Chargeoff Rate Model Accuracy Plots – Unemployment Rate, BBB Corporate - 5 Year Treasury Bond Spread and BBB Corporate Bond Yield (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

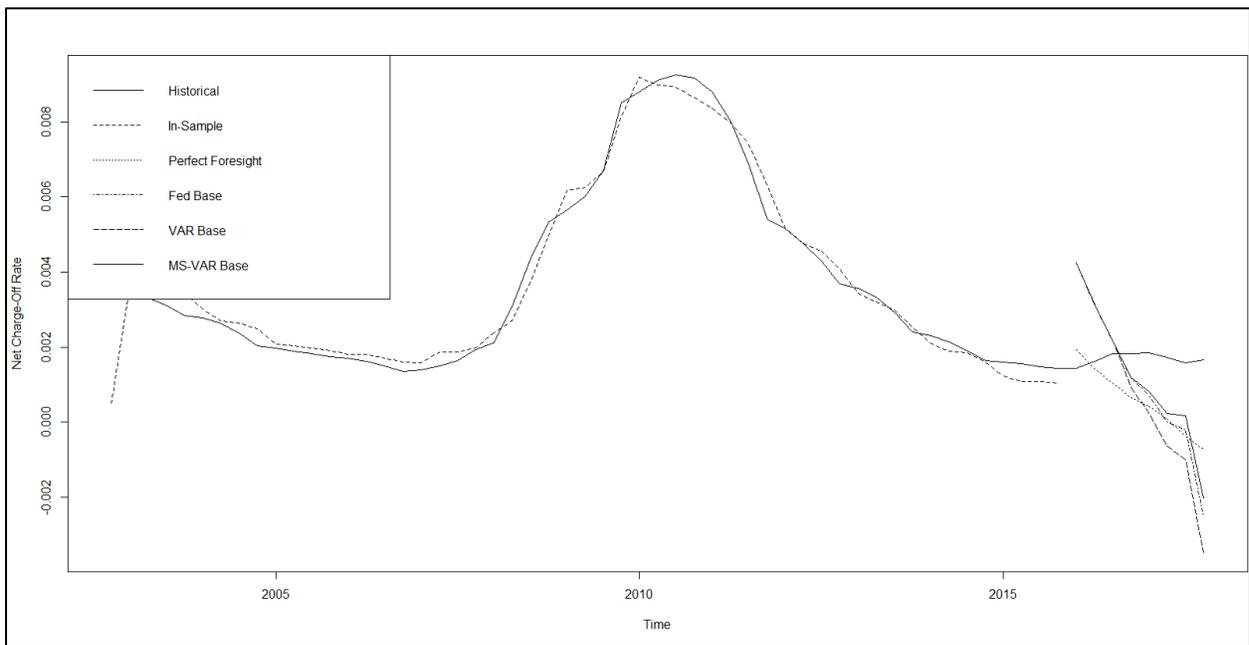


Figure 24: Net-Chargeoff Rate Model Accuracy Plots – Unemployment Rate, BBB Corporate - 5 Year Treasury Bond Spread, BBB Corporate Bond Yield and Total Trading Account Assets to Total Assets (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

VIX, UNEMP, CORPSPR & TUCGR). On the other hand, under the VAR scenario model 8 (with CREPI, VIX, OROTA & TAATA) is the worst performing, while the best performing model is number 3 (with UNEMP & BBBCY). Furthermore, under the FED scenario model 8 (with CREPI, VIX, OROTA & TAATA) is the worst performing, while the best performing model is number 3 (with UNEMP & CREPI). Finally, under the perfect foresight benchmark model 10 (with BBBCY, UNEMP & CORPSPR) is the worst performing, while the best performing model is number 3 (with UNEMP & BBBCY).

Next let us discuss the detailed out-of-sample performance of the CPE measure. First considering scenario generation models across predictive loss specifications, we observe that the MS-VAR model performs better than the Fed and VAR models, having CPEs averaging -41.5% and ranging from -135.8% to 87.4%, outperforming the average CPE of -48.7% (-56.5%) and ranging in -151.8% to 48.9% (-135.8% to 37.4%) in the Fed (VAR) models. Turning to the relative performance of loss predictive specifications according to CPE by scenario generation model, we note that across all scenario generation models, the simpler specifications all perform better under the CPE metric. Under the MS-VAR scenario model 10 (with VIX, UNEMP, CORPSPR & TUCGR) is the worst performing, while the best performing model is number 11 (with BBBCY, UNEMP & CREPI). On the other hand, under the VAR scenario model 8 (with CREPI, VIX, OROTA & TAATA) is the worst performing, while the best performing model is number 3 (with UNEMP & CREPI). Furthermore, under the FED scenario model 8 (with CREPI, VIX, OROTA & TAATA) is the worst performing, while the best performing model is number 12 (with BBBCY, UNEMP, CREPI & CDLG). Finally, under the perfect foresight benchmark model 13 (with BBBCY, UNEMP & CORPSPR) is the worst performing, while the best performing model is number 5 (with UNEMP & CORPSPR).

Finally let us discuss the detailed out-of-sample performance of the AIC measure. First considering scenario generation models across predictive loss specifications, we observe that the MS-VAR model performs better than the Fed and VAR models, having AICs averaging -83.7 and ranging from -101.0 to -66.42, outperforming the average AIC of -83.3 (-83.7) and ranging in -113.9 to -64.3 (-116.4 to -63.4) in the Fed (VAR) models. Turning to the relative performance of loss predictive specifications according to AIC by scenario generation model, we note that across all scenario generation models, the simpler specifications all perform better under the AIC metric. Under the MS-VAR scenario model 8 (with CREPI, VIX, OROTA & TAATA) is the worst performing, while the best performing model is number 3 (with UNEMP and CREPI). On the other hand, under the VAR scenario model 8 (with CREPI, VIX, OROTA & TAATA) is the worst performing, while the best performing model is number 3 (with UNEMP & CREPI). Furthermore, under the FED scenario model 8 (with CREPI, VIX, OROTA & TAATA) is the worst performing, while the best performing model is number 3 (with UNEMP & CREPI). Finally, under the perfect foresight benchmark model 9 (with VIX, UNEMP & CORPSPR) is the worst performing, while the best performing model is number 12 (with BBBCY, UNEMP, CREPI & CDLG).

6 The Quantification of Model Risk According to the Principle of Relative Entropy

Risk measurement relies on modelling assumptions, the errors in which exposing such models to model risk. In this paper we apply a tool for quantifying model risk and making risk measurement robust to modeling errors. As simplifying assumptions are inherent to all modelling frameworks, the prime directive of model risk management is to assess vulnerabilities to and consequences of model errors. Therefore, a well-designed model risk measurement framework is capable of bounding the effect of model

error on specific measures of risk, given a baseline nominal model for measuring risk, as well as identifying the sources of model error to which a measure of risk is most vulnerable, and furthermore isolating which changes in the underlying model have the greatest impact on this risk measure.

In this paper, consistent with the objectives of credit loss measurement in CECL, we focus on the both objectives through calculating an upper bound on the range of credit risk values that can result over a range of model errors within a certain distance of a nominal model, for a range of credit loss model and economic scenario generation models. This bound is somewhat analogous to an upper confidence bound, but whereas a confidence interval quantifies the effect of sampling variability, the robustness bound that we develop quantify the effect of model error. The simple example of a standard error estimate should help illustrate this idea, as a conventional measure of credit risk in a loan or bond portfolio. Measuring standard deviation prospectively requires assumptions about the joint distribution of the returns of assets, or default correlation, in a credit portfolio.

In light of the first objective listed above and our focus in the CECL context, we would want to bound the values of standard deviation that can result from a reasonable degree of model error. In practice, model risk is sometimes addressed by comparing the results of different models, but more often if it is considered at all, model risk is investigated by varying model parameters. Crucially, the tools applied here go beyond parameter sensitivity to consider the effect of changes in the probability law that defines an underlying model, enabling us to identify vulnerabilities to model error that are not reflected in parameter perturbations. For example, the main source of model risk might result from an error in a joint distribution of returns that cannot be described through a change in a covariance matrix. To work with model errors described by changes in probability laws, we need a way to quantify such changes, and to this end we deploy the *principle of relative entropy* (Hansen and Sargent (2007), Glasserman and Xu (2013).) In Bayesian statistics, the relative entropy between posterior and prior distributions measures the information gained through additional data. In characterizing model error, we interpret relative entropy as a measure of the additional information required to make a perturbed model preferable to a baseline model. Thus, relative entropy becomes a measure of the plausibility of an alternative model. It is also a convenient choice because the worst-case alternative within a relative entropy constraint is typically given by an *exponential change of measure*.

We quantify the model risk with respect to a champion, or null, model $y = f(x)$ such that the *Kullback-Leibler Relative Entropy* divergence measure from a challenger, or reference, model $y = g(x)$ is given by:

$$D(f, g) = \int \frac{g(x)}{f(x)} \log \left(\frac{g(x)}{f(x)} \right) f(x) dx \quad (9)$$

In this construct, the $f(x)$ mappings are our set of estimated CECL loss distribution models, and the $g(x)$ benchmark is some kind of alternative, such as the perfect foresight loss forecast. We can define the *likelihood ratio* $m(f, g)$ characterizing our modeling choice according to the relationship:

$$m(f, g) = \frac{g(x)}{f(x)} \quad (10)$$

It is standard in the literature to express (9) in terms as an equivalent expectation of a relative deviation in likelihood:

$$\mathbb{E}_f \left[m \log(m) \right] = D(f, g) < \delta \quad (11)$$

Where δ represents a relatively small upper bound on model risk deviations, dictated by the model risk appetite of the organization with respect to a particular model type (e.g., a model performance threshold). A key property of relative entropy is that $d(f, g) \geq 0$ and $d(f, g) = 0$ only if $f(x) = g(x)$. Given a set of reference models $g(x)$ and a relative distance measure $d(f, g) < \delta$, the solution the model error can be quantified by the following change of numeraire (Glasserman and Xu, 2013):

$$m_\theta(f, g) = \frac{\exp(\theta f(x))}{\mathbb{E}_f \left[\exp(\theta f(x)) \right]} \quad (12)$$

Where (12) is the solution or inner supremum to the optimization problem:

$$m_\theta(f, g) = \inf_{\theta > 0} \sup_{m(x)} \mathbb{E}_f \left[m(x) f(x) - \frac{1}{\theta} \left(m(x) \log(m(x)) - \delta \right) \right] \quad (13)$$

In (12) the model risk measure is parametricized by $\theta \in [0, 1]$, such that $\theta = 0$ corresponds to the best case of zero model risk, and $\theta = 1$ the worst case of maximal model risk. An important property of the change of measure (12) is that it is model free, or independent of the reference model $g(x)$.

We summarize the empirical implementation results of our model risk quantification in Table 4 and Figures 25 through 38. In Table 4 we tabulate the mean CECL loss under the model, the same in the worst case, and the *relative model risk error* (“RMRE”) measure defined as:

$$\text{RMRE}_{f, \theta}(f, g | x, l, e) = \frac{\mathbb{E}_f \left[f_\theta^{WC}(x | l, e) \right]}{\mathbb{E}_f \left[f(x | l, e) \right]} - 1 \approx \frac{\frac{1}{T} \sum_{t=1}^T f_\theta^{WC}(\mathbf{x}_{l,t} | l, e_{\mathbf{x}_{l,t}})}{\frac{1}{T} \sum_{t=1}^T \left[f(\mathbf{x}_{l,t} | l, e_{\mathbf{x}_{l,t}}) \right]} - 1 \quad l = 1, \dots, N_l; e = 1, \dots, N_e \quad (14)$$

Where $\mathbb{E} \left[f(\mathbf{x}_{l,t} | l, e_{\mathbf{x}_{l,t}}) \right]$ is a CECL loss estimate for the vector of macroeconomic variables $\mathbf{x}_{l,t}$ in loss model $l = 1, \dots, N_l$ (our 14 models VAR NCO models) and economic scenario generation model $e = 1, \dots, N_e$ (Fed, VAR and MS-VAR), and $f_\theta^{WC}(\mathbf{x}_{l,t} | l, e_{\mathbf{x}_{l,t}})$ is the worst case version at time $t = 1, \dots, T$ in the forecast period, so this represents an empirical forecast error metric where the expectation with respect to the null model is replaced by a sample average.

Table 4 shows that across credit loss and macroeconomic scenario generation models, the average RMRE of 34.01% is substantial and varies widely across both dimensions of specification, from 15.4% to 51.9%. Considering scenario generation frameworks, we observe that the MS-VAR model has a consistently lower RMSE measure as compared to the Fed or VAR models, averaging 27.8% in the former as compared to 27.5% and 36.6%, respectively, in the latter. Another pattern that we observe is that, in the majority of cases, credit loss models having either more macroeconomic, or include idiosyncratic factors in addition to a set of macroeconomic factors, have higher model risk measures. The 2-variable credit loss models with no idiosyncratic variables have RMSEs ranging in 21.6% to 25.2.4%, while those having 3 macroeconomic

Table 4: Kullback-Leibler Relative Entropy Worse Case Loss for Model Risk Quantification Measures - Vector Autoregressive CECL Models Out-of-Sample Compared (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

Model Number	Macroeconomic		Fed Base Scenario	VAR Base Scenario	MS-VAR Base Scenario	Min	Avg	Max
1	Unemployment Rate & BBB Corporate Yield	Mean Base Scenario	0.0338%	0.0448%	0.0781%	23.82%	26.82%	30.49%
		Mean Worse Case Base	0.0617%	0.0519%	0.1146%			
		Relative Model Risk Error	30.49%	26.14%	23.82%			
2	Unemployment Rate & BBB Corporate Yield	Mean Base Scenario	0.0696%	0.0344%	0.0920%	25.61%	30.17%	36.47%
		Mean Worse Case Base	0.0956%	0.0713%	0.1237%			
		Relative Model Risk Error	36.47%	28.43%	25.61%			
3	Unemployment Rate & Commercial Real Estate Price Index	Mean Base Scenario	0.1567%	0.1648%	0.1913%	19.87%	24.65%	27.88%
		Mean Worse Case Base	0.1642%	0.1777%	0.2300%			
		Relative Model Risk Error	26.20%	27.88%	19.87%			
4	Unemployment Rate & Commercial Real Estate Price Index	Mean Base Scenario	0.0868%	0.0895%	0.1366%	25.15%	31.25%	34.69%
		Mean Worse Case Base	0.1296%	0.1364%	0.2115%			
		Relative Model Risk Error	33.92%	34.69%	25.15%			
5	Unemployment Rate & BBB Corporate - 5 Year Treasury Spread	Mean Base Scenario	0.1817%	0.1845%	0.1979%	15.36%	21.64%	25.98%
		Mean Worse Case Base	0.2059%	0.2063%	0.2289%			
		Relative Model Risk Error	25.98%	23.57%	15.36%			
6	Unemployment Rate & BBB Corporate - 5 Year Treasury Spread	Mean Base Scenario	0.0661%	0.1008%	0.1177%	21.58%	26.11%	28.62%
		Mean Worse Case Base	0.1085%	0.1305%	0.1329%			
		Relative Model Risk Error	28.14%	28.62%	21.58%			
7	Commercial Real Estate Price Index & CBOE Equity Market Volatility Index	Mean Base Scenario	0.1129%	0.1166%	0.1021%	17.22%	25.18%	29.39%
		Mean Worse Case Base	0.1711%	0.1700%	0.1614%			
		Relative Model Risk Error	29.39%	28.93%	17.22%			
8	Commercial Real Estate Price Index & CBOE Equity Market Volatility Index	Mean Base Scenario	0.1458%	0.1564%	0.1187%	30.92%	39.35%	44.53%
		Mean Worse Case Base	0.1026%	0.1101%	0.0714%			
		Relative Model Risk Error	42.60%	44.53%	30.92%			
9	CBOE Equity Market Volatility Index, Unemployment Rate & BBB Corporate - 5 Year Treasury Spread	Mean Base Scenario	0.1331%	0.1381%	0.1361%	29.65%	36.78%	41.69%
		Mean Worse Case Base	0.1398%	0.1436%	0.1221%			
		Relative Model Risk Error	39.00%	41.69%	29.65%			
10	CBOE Equity Market Volatility Index, Unemployment Rate & BBB Corporate - 5 Year Treasury Spread	Mean Base Scenario	0.1658%	0.1718%	0.1741%	36.39%	43.61%	48.98%
		Mean Worse Case Base	0.2162%	0.2157%	0.2091%			
		Relative Model Risk Error	48.98%	45.47%	36.39%			
11	BBB Corporate Yield, Unemployment Rate & Commercial Real Estate Price Index	Mean Base Scenario	0.0836%	0.0598%	0.1513%	33.63%	39.64%	42.70%
		Mean Worse Case Base	0.1089%	0.0972%	0.2006%			
		Relative Model Risk Error	42.70%	42.57%	33.63%			
12	BBB Corporate Yield, Unemployment Rate & Commercial Real Estate Price Index	Mean Base Scenario	0.1796%	0.1545%	0.2326%	41.44%	47.41%	51.91%
		Mean Worse Case Base	0.2527%	0.1990%	0.3305%			
		Relative Model Risk Error	51.91%	48.89%	41.44%			
13	BBB Corporate Yield, Unemployment Rate & BBB Corporate - 5 Year Treasury Spread	Mean Base Scenario	0.2520%	0.2128%	0.0444%	26.26%	36.68%	42.36%
		Mean Worse Case Base	0.3330%	0.2400%	0.0723%			
		Relative Model Risk Error	42.36%	41.42%	26.26%			
14	BBB Corporate Yield, Unemployment Rate & BBB Corporate - 5 Year Treasury Spread	Mean Base Scenario	0.1766%	0.1981%	0.1746%	42.93%	46.81%	50.02%
		Mean Worse Case Base	0.3231%	0.3283%	0.3194%			
		Relative Model Risk Error	47.49%	50.02%	42.93%			
Relative Model Risk Error Summary Statistics - Macroeconomic Scenario Generation Models		Minimum	25.98%	23.57%	15.36%			
		Average	37.54%	36.63%	27.84%			
		Standard Deviation	8.75%	9.22%	8.48%			
		Maximum	51.91%	50.02%	42.93%			
Relative Model Risk Error Summary Statistics - Total Macroeconomic Scenario Generation & Credit Loss Models		Minimum	15.36%					
		Average	34.01%					
		Standard Deviation	9.68%					
		Maximum	51.91%					

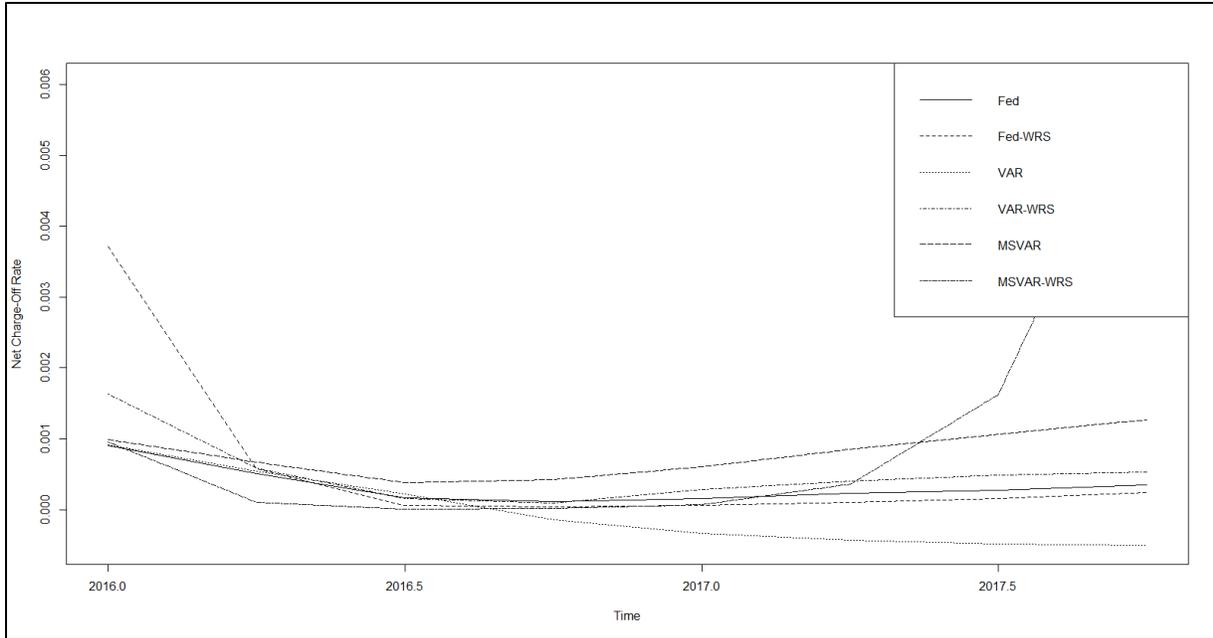


Figure 25: Kullback-Leibler Relative Entropy Worse Case Loss for Model Risk Quantification Plot – Unemployment Rate and BBB Corporate Bond Yield (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

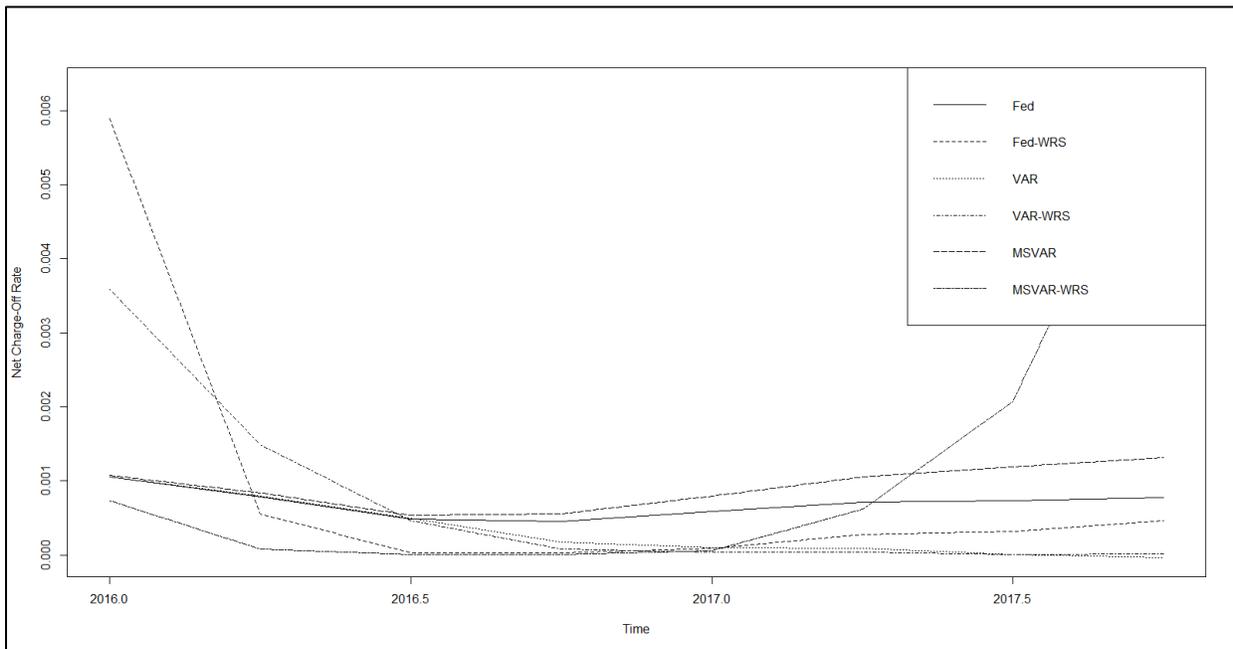


Figure 26: Kullback-Leibler Relative Entropy Worse Case Loss for Model Risk Quantification Plot – Unemployment Rate, BBB Corporate Bond Yield, Commercial Development Loan Growth and Commercial & Industrial Loans to Total Assets (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

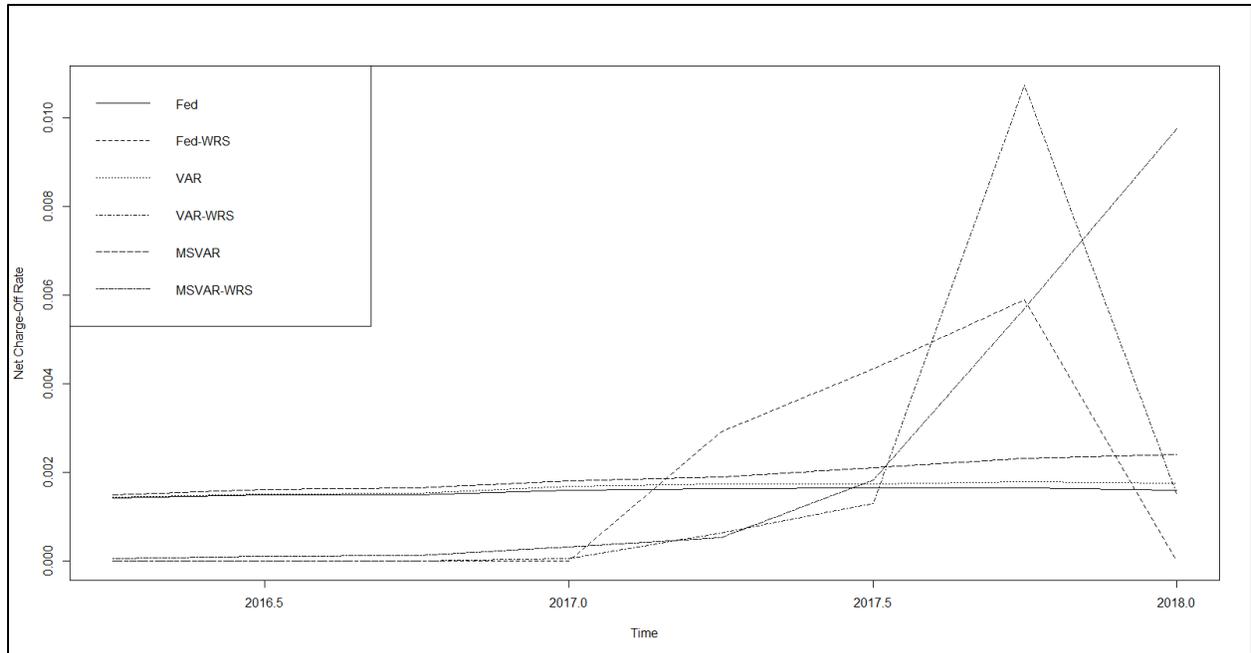


Figure 27: Kullback-Leibler Relative Entropy Worse Case Loss for Model Risk Quantification Plot – Unemployment Rate and Commercial Real Estate Price Index (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

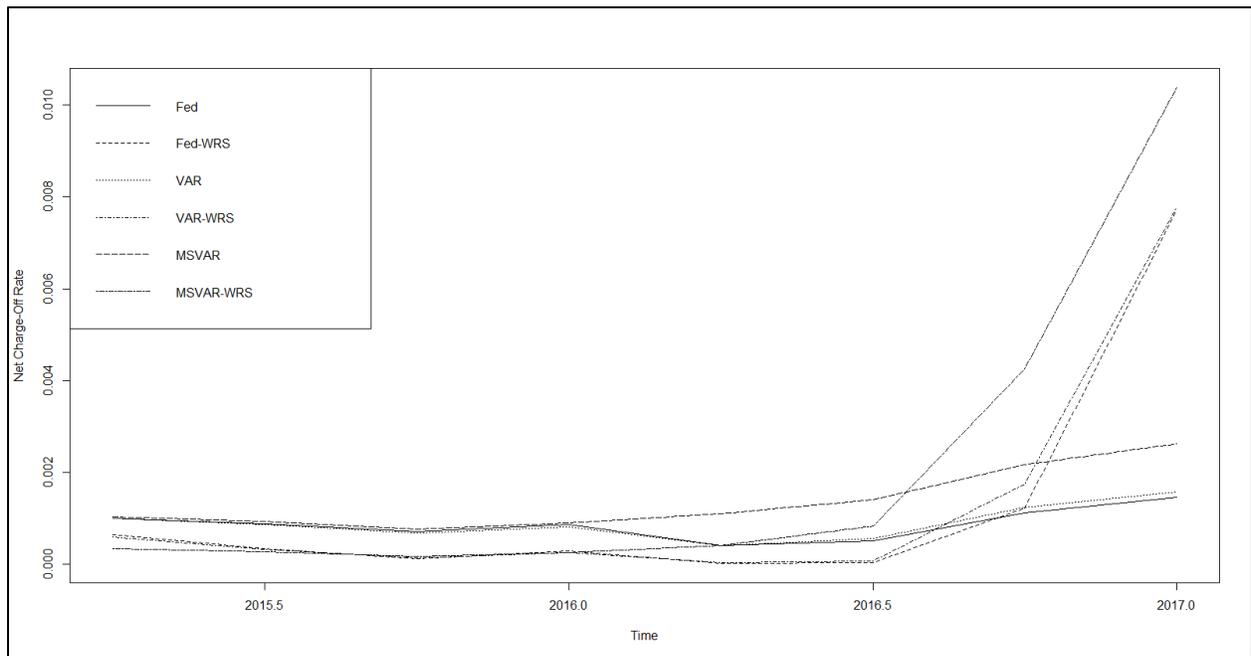


Figure 28: Kullback-Leibler Relative Entropy Worse Case Loss for Model Risk Quantification Plot – Unemployment Rate, Commercial Real Estate Price Index, Commercial Development Loan Growth and Total Trading Account Assets to Total Assets (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

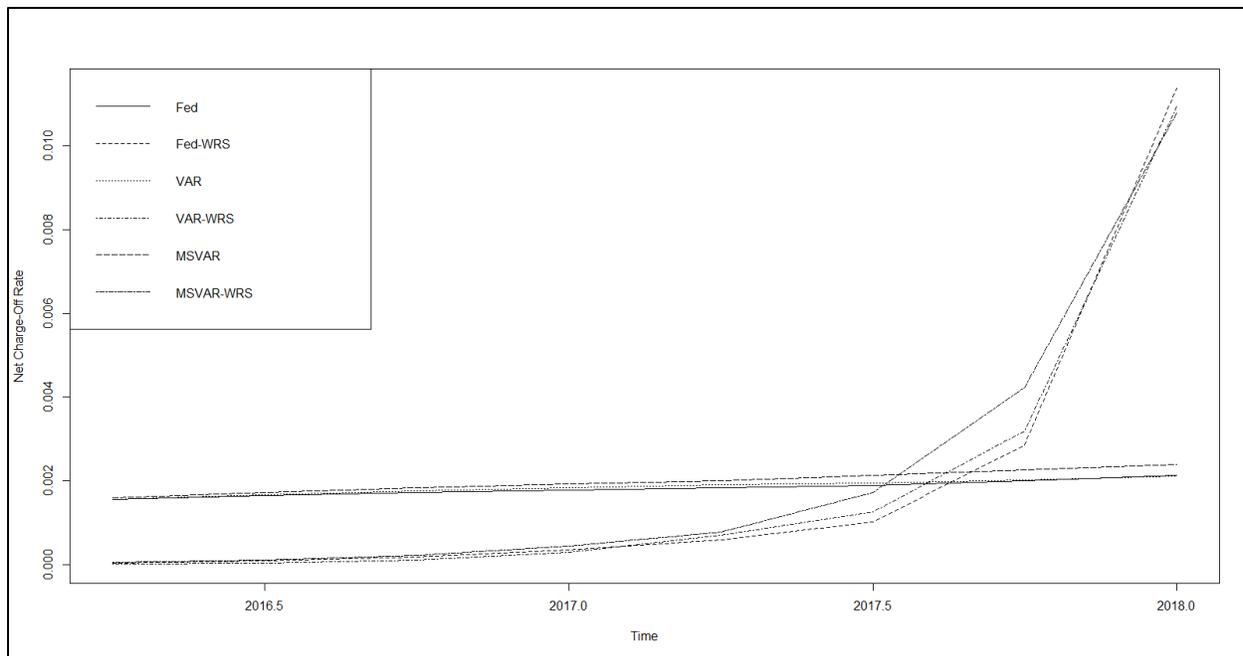


Figure 29: Kullback-Leibler Relative Entropy Worst Case Loss for Model Risk Quantification Plot – Unemployment Rate and BBB Corporate - 5 Year Treasury Bond Spread (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

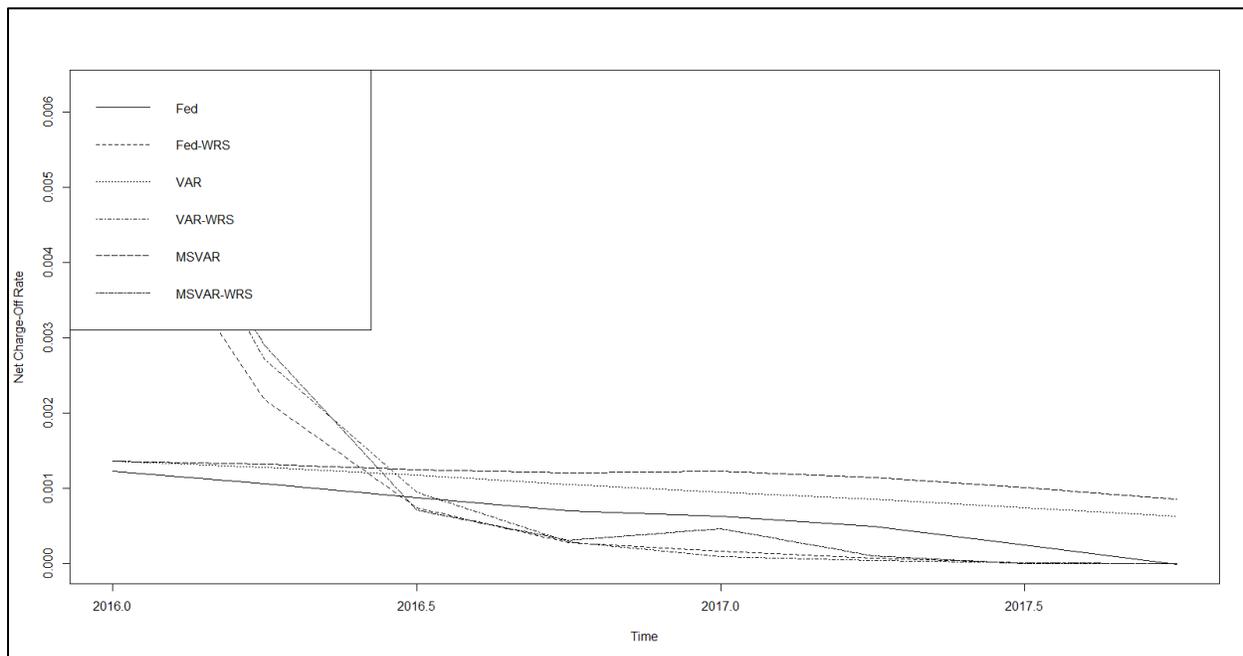


Figure 30: Kullback-Leibler Relative Entropy Worst Case Loss for Model Risk Quantification Plot – Unemployment Rate, BBB Corporate - 5 Year Treasury Bond Spread and Other Real Estate Owned to Total Assets (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

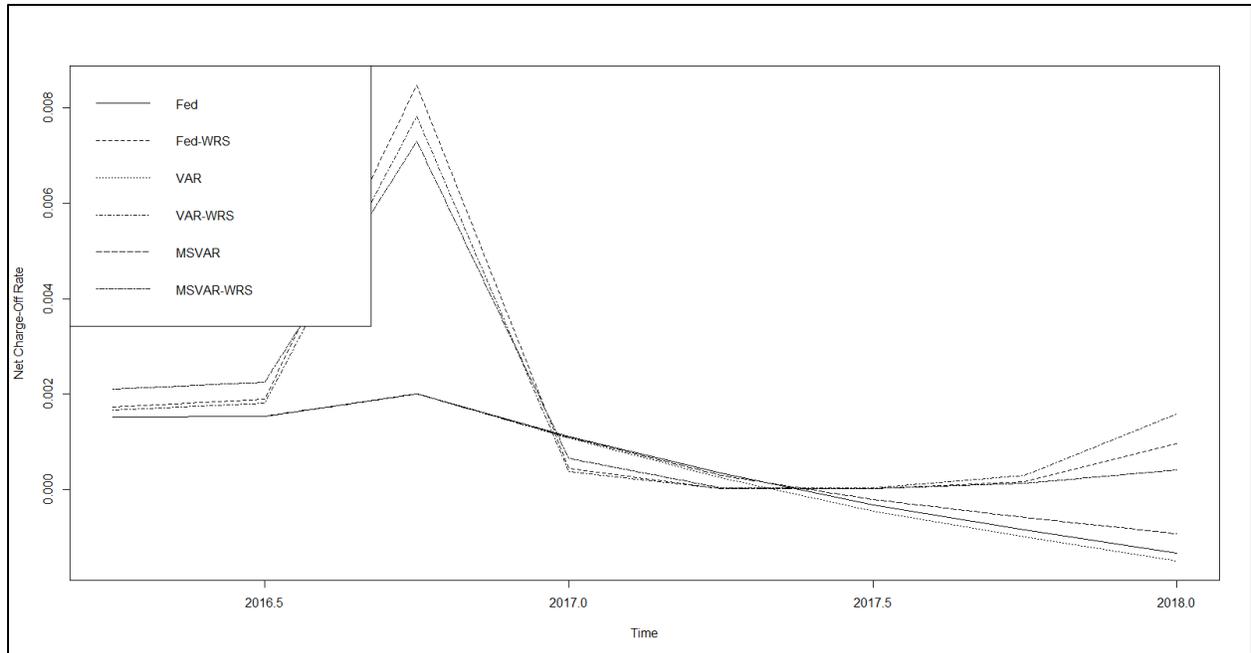


Figure 31: Kullback-Leibler Relative Entropy Worse Case Loss for Model Risk Quantification Plot – Commercial Real Estate Price Index and CBOE Equity Volatility Index (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

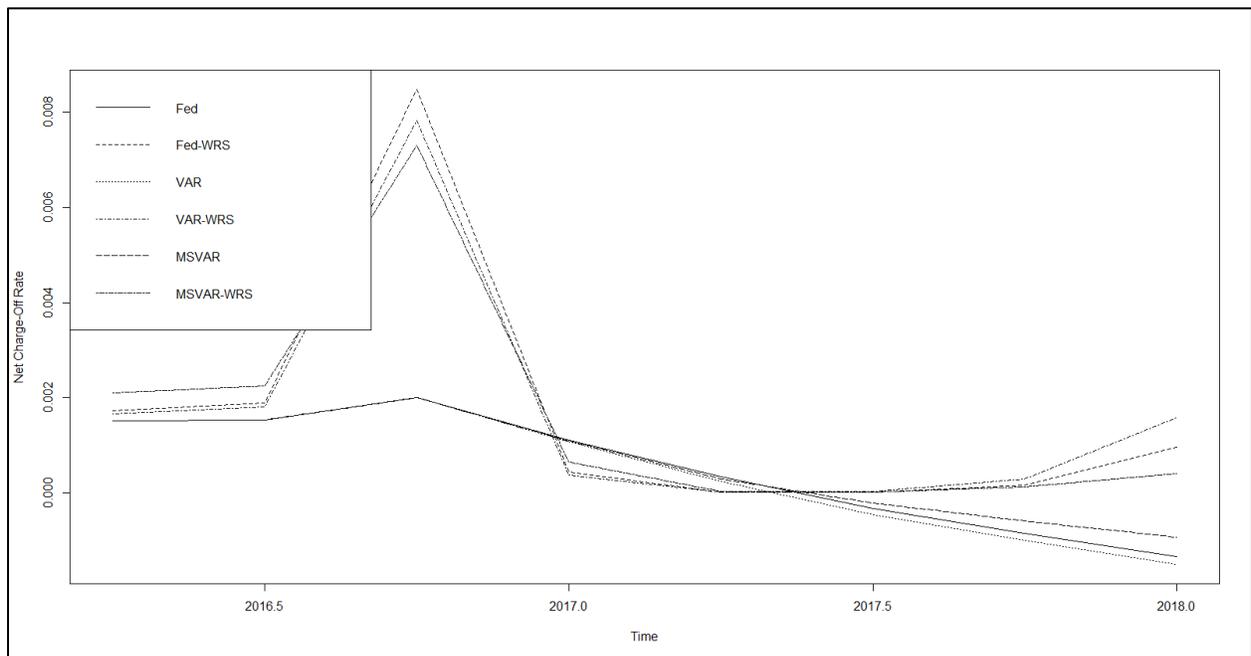


Figure 32: Kullback-Leibler Relative Entropy Worse Case Loss for Model Risk Quantification Plot – Commercial Real Estate Price Index, CBOE Equity Volatility Index, Other Real Estate Owned to Total Assets and Total Trading Account Assets to Total Assets (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

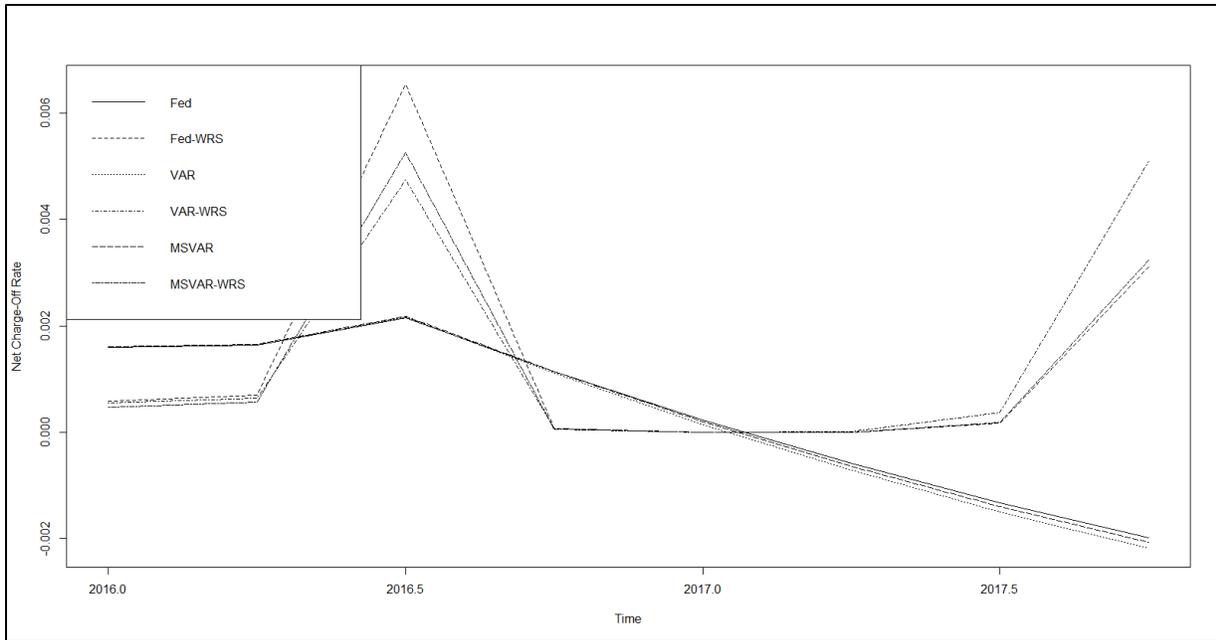


Figure 33: Kullback-Leibler Relative Entropy Worse Case Loss for Model Risk Quantification Plot – Unemployment Rate, CBOE Equity Volatility Index and BBB Corporate - 5 Year Treasury Bond Spread (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

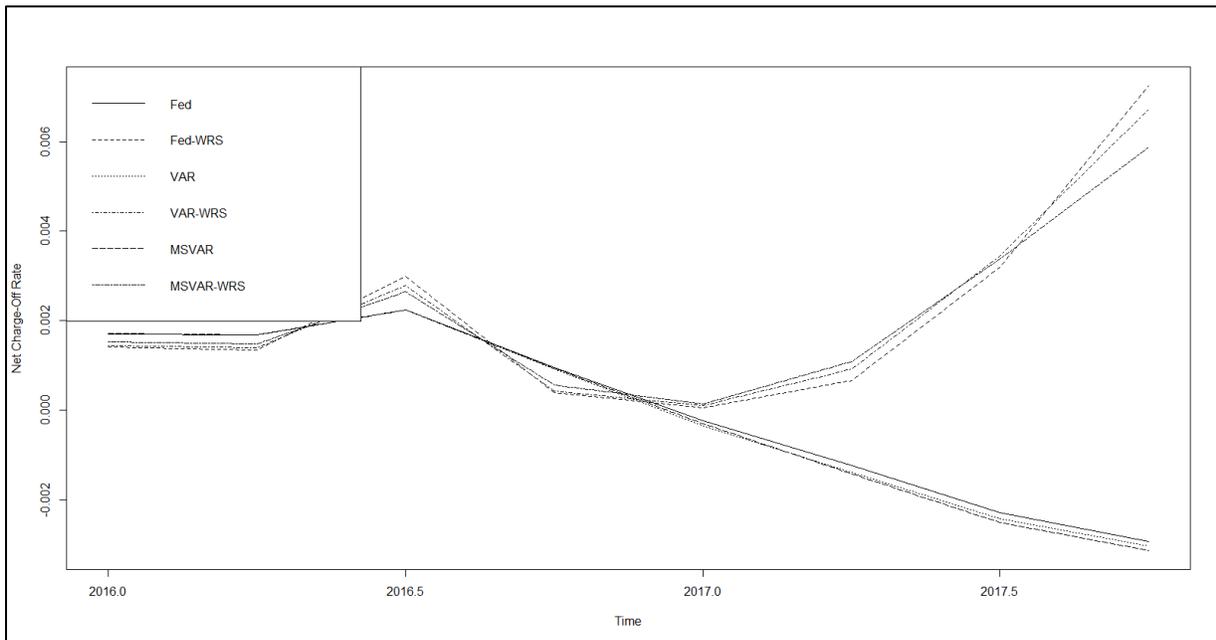


Figure 34: Kullback-Leibler Relative Entropy Worse Case Loss for Model Risk Quantification Plot – Unemployment Rate, CBOE Equity Volatility Index, BBB Corporate - 5 Year Treasury Bond Spread and Total Uncommitted Loan Growth (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

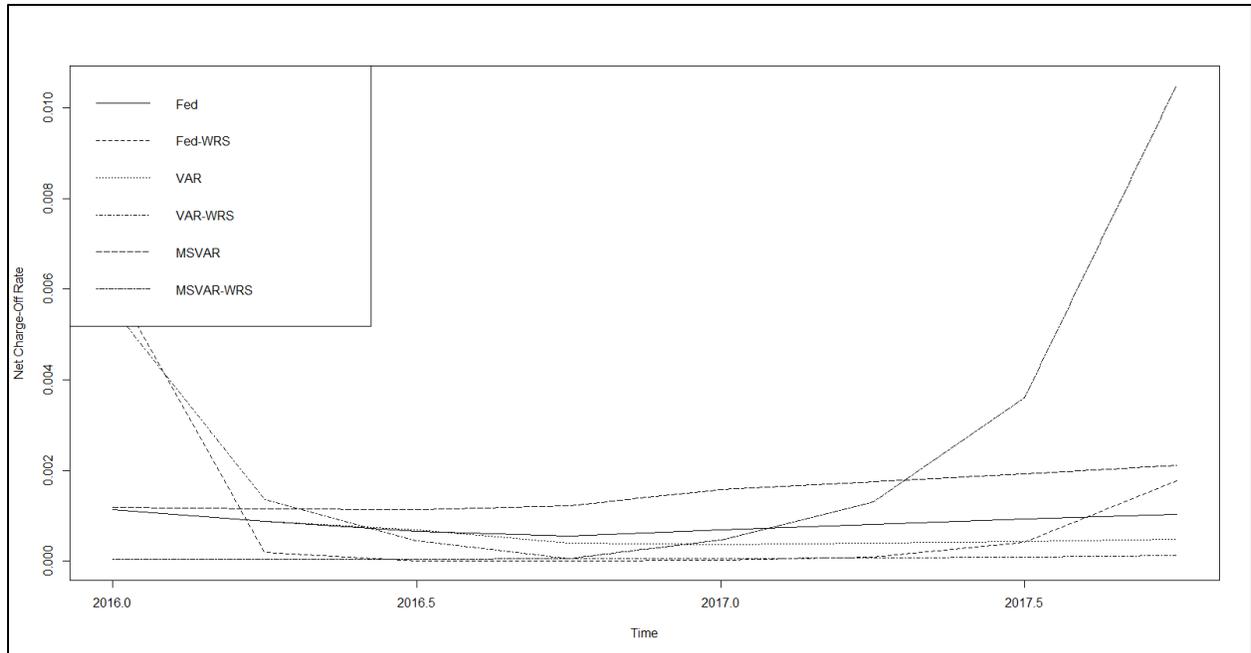


Figure 35: Kullback-Leibler Relative Entropy Worse Case Loss for Model Risk Quantification Plot – Unemployment Rate, Commercial Real Estate Price Index and BBB Corporate Bond Yield (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

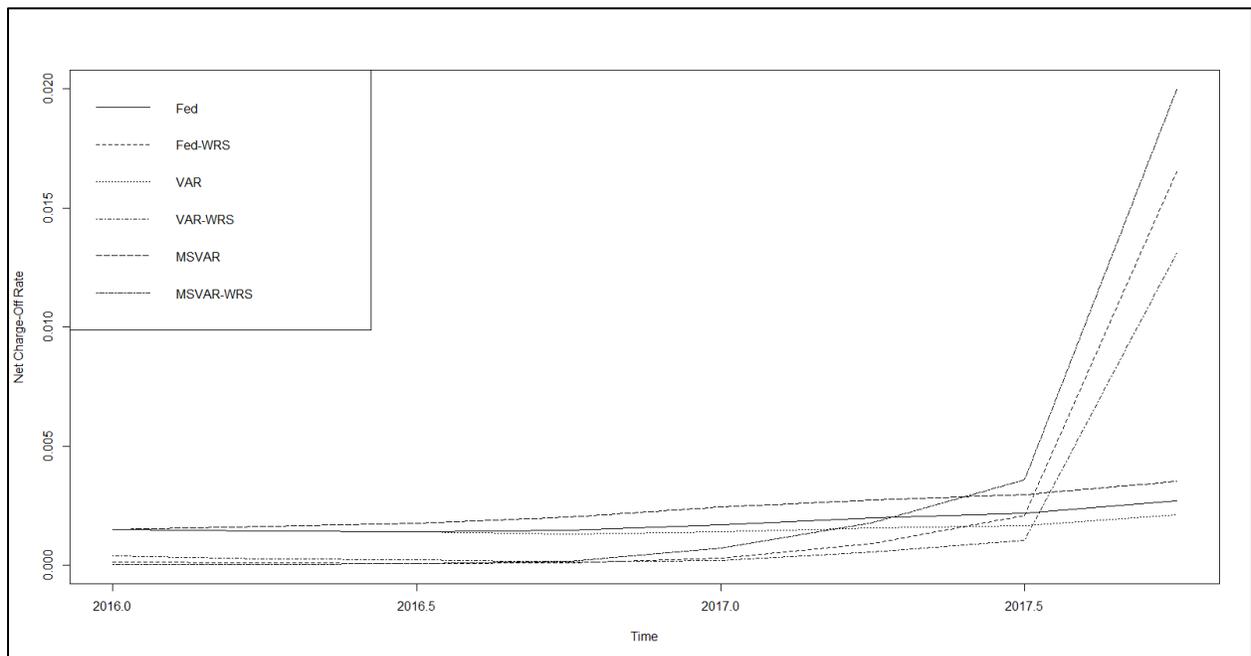


Figure 36: Kullback-Leibler Relative Entropy Worse Case Loss for Model Risk Quantification Plot – Unemployment Rate, Commercial Real Estate Price Index, BBB Corporate Bond Yield and Commercial Development Loan Growth (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

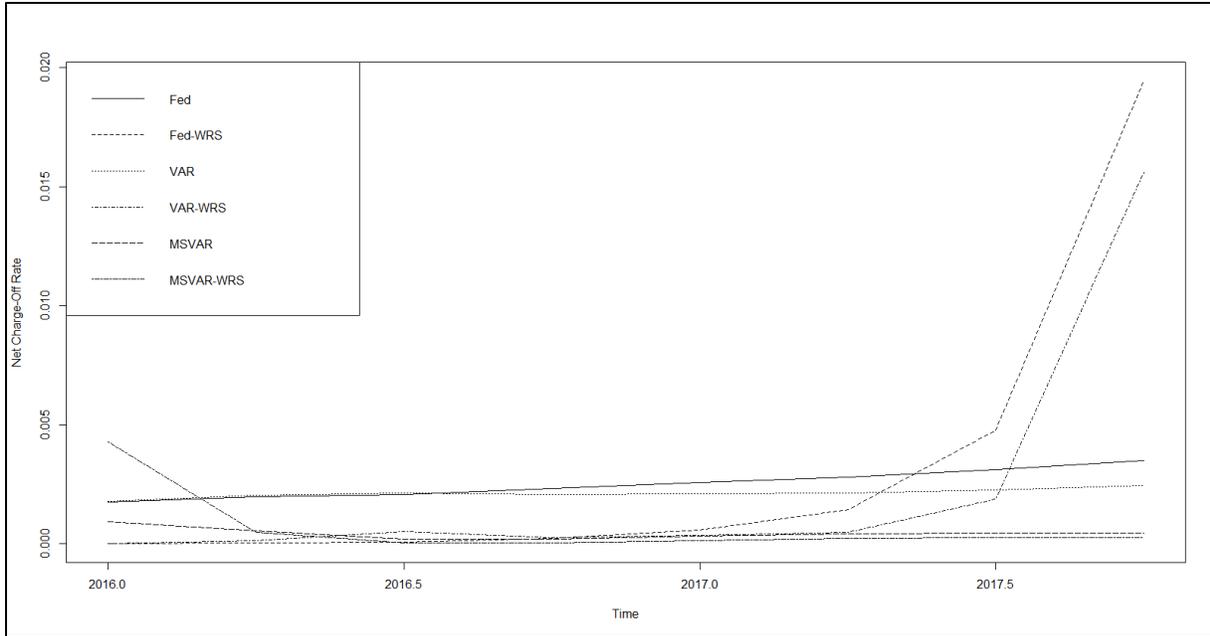


Figure 37: Kullback-Leibler Relative Entropy Worse Case Loss for Model Risk Quantification Plot – Unemployment Rate, BBB Corporate - 5 Year Treasury Bond Spread and BBB Corporate Bond Yield (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

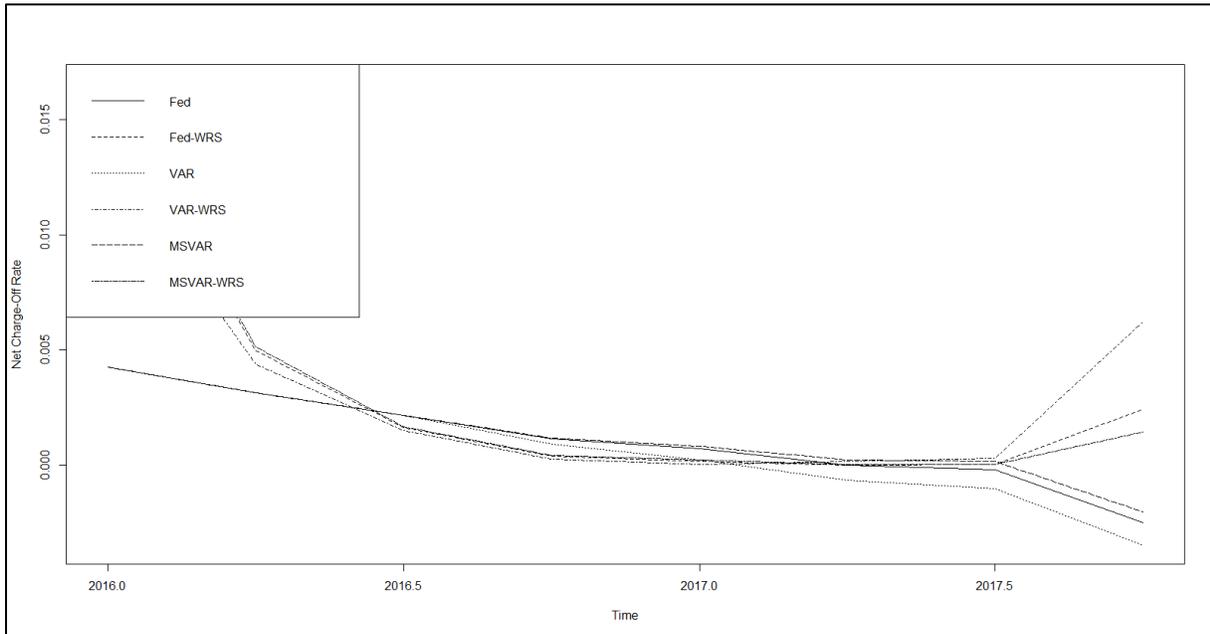


Figure 38: Kullback-Leibler Relative Entropy Worse Case Loss for Model Risk Quantification Plot – Unemployment Rate, BBB Corporate - 5 Year Treasury Bond Spread, BBB Corporate Bond Yield and Total Trading Account Assets to Total Assets (FDIC SDI Report, Federal Reserve Board 4Q91-4Q15 and Jacobs et al (2018) Models).

factors have a range of 36.7% to 39.6%. On the other hand, among the version of these models having idiosyncratic variables, the 2-variable credit loss models have RMSEs ranging in 24.7% to 39.4%, while those having 3 macroeconomic factors have a range of 43.6% to 47.4%.

There is a profound implication of this analysis that speaks to our second objective in measuring model risk, namely identifying the sources of vulnerability in assumptions that give rise to model risk. There are two sources at play herein – the joint distribution of the macroeconomic and idiosyncratic variables, and the assumptions on the error terms in the joint distribution of losses, with respect to the latter whether they are Gaussian (i.e., Fed or VAR) or follow a heavy-tailed distribution (i.e., MS-VAR). We conclude from these results that the less parsimonious are the models, the greater risk is there of model misspecification that manifests in the higher RMRE measures. On the other hand, while we may think that the additional parameter of the MS-VAR model would give rise to more model risk, the fact that this model is better capable of modeling the fat-tailed distribution of credit losses is realized in a lower MS-VAR measure, regardless of the credit loss model specification. This should be no surprise, as we saw that the Fed and VAR models had rather more egregious under-prediction than the MS-VAR model, which is at odds with the historical distribution of losses. On the other hand, the credit models with more variables fit only slightly better on an in-sample basis, but most exhibited relatively poor performance out-of-sample. The conclusion is that practitioner may wish to err on the side of more parsimonious models that can also accommodate non-normality.

7 Conclusion and Future Directions

In this study, toward the end of analyzing the impact of model specification and scenario dynamics upon expected credit loss estimates in CECL, we have implemented a highly stylized framework borrowed from the ST modeling practice. We performed a model selection of alternative CECL specifications in a top-down framework, using FDIC FR-Y9C data and constructing an aggregate or average hypothetical bank, with the target variable being NCORs and the explanatory variables constituted by Fed provided macroeconomic variables as well as bank-specific controls for idiosyncratic risk. We studied not only the impact of the ALLL estimate under CECL for alternative model specifications, but also the impact of different frameworks for scenario generation: the Fed baseline assumption, a Gaussian VAR model and a mixture-of-distributions MS-VAR model, following the study of Jacobs et al (2018a).

We have established in general the CECL methodology is at risk of not achieving the stated objective of reducing the pro-cyclicality of provisions relative to the legacy incurred loss standard, as across models we observe chronic underprediction of losses in the last 2-year out-of-sample period, which arguably is a period that is late in the economic cycle. Furthermore, we have illustrated that the amount of such pro-cyclicality exhibits significant variation across model specifications and scenario generation frameworks. In general, we have found that the MS-VAR scenario generation framework produces the best performance in terms of fit and lack of under-prediction relative to the perfect foresight benchmark, which is in line with the common industry practice of giving weight to adverse but probable scenarios, which the MS-VAR regime switching model can produce naturally and coherently as part of the estimation methodology that places greater weight on the economic downturn. We have also found that for any scenario generation model, across specification the more lightly parameterized models tend to have better out of sample performance. Furthermore, relative to the perfect foresight benchmark, the MS-VAR model was found to have produced a lower level of variation in the model performance statistics across loss predictive model specifications.

As a second and related exercise, we quantified the level of model risk in this hypothetical CECL exercise, using the principle of *relative entropy*. We found that more elaborate modeling choices, such as more highly parametrized models in terms of more macroeconomic or idiosyncratic covariates, tend to introduce more measured model risk. However, the more highly parametrized macroeconomic scenario

generation framework that can accommodate heavy tails in the credit loss distribution (the MS-VAR model, which has separate parameters for normal and stressed conditions), tend to introduce less measured model risk than Gaussian approaches (Fed or VAR models). The implication is that while banks may wish to err on the side of more parsimonious approaches in order to manage the increase model risk that the introduction of the CECL standard gives rise to, they are advised to balance this against the need to model the non-normality in the credit loss distribution.

The implication of this analysis is that the volume of lending and the amount of regulatory capital held may vary greatly across banks, even when it is the case that the respective loan portfolios have very similar risk profiles. Another consequence of this divergence of expected loan loss estimates under the CECL standard is that supervisors and other market participant stakeholders may face challenges in comparing banks at a point of time or over time. There are also implications for the influence of modeling choices in specification and scenario projections on the degree of model risk introduced by the CECL standard.

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 score-card/regression
- Alternative data-sets, for example bank or loan level data
- More general classes of regression model, such as logistic, semi-parametric or machine learning / artificial intelligence techniques (Jacobs, 2018b)
- Applications related to stress testing, such as regulatory or economic capital

References

- Basel Committee on Banking Supervision (2006) 'International Convergence of Capital Measurement and Capital Standards: A Revised Framework', The Bank for International Settlements, Basel, Switzerland.
- Basel Committee on Banking Supervision (2009) 'Principles for Sound Stress Testing Practices and Supervision - Consultative Paper No. 155', The Bank for International Settlements, Basel, Switzerland.
- Batchelor, R. A., & Dua, P. (1990). Product differentiation in the economic forecasting industry. *International Journal of Forecasting* 6(3), 311-316.
- Berrospide, J., & Edge. R. (2010). The effects of bank capital on lending: What do we know, what does it mean. *International Journal of Central banking* 6(4), 5-55.
- Bernanke, B. S., & Lown C. S. (1991). The credit crunch. *Brookings Papers on Economic Activity* 2, 205-247.
- Board of Governors of the Federal Reserve System. (2011) 'Supervisory Guidance on Model Risk Management', Supervisory Letter 11-7, Washington, D.C., April 4th.
- Box, G., & Jenkins, G. (1970). *Times series analysis: forecasting and control*. San Francisco, C.A.: Holden-Day.
- Brockwell P.J., & Davis R.A. (1991). *Time series: theory and methods*. New York, N.Y.: Springer-Verlag.
- Carlson, M., H. Shan, & Warusawitharana, M. (2013). Capital ratios and bank lending: A matched bank approach. *Journal of Financial Intermediation* 22, 663-687.
- Chae, S. Sarama, R. Vocjtech, C., & Wang, J. (2017). The impact of the current expected credit loss standard (CECL) on the timing and comparability of reserves, SSRN Working Paper (October).
- Commandeur, J. J. F., & Koopman, S.J. (2007). *Introduction to state space time series analysis*. New York, N.Y.: Oxford University Press.

- Cornett, M., McNutt, J., Strahan, P., & Tehranian, H. (2011). Liquidity risk management and credit supply in the financial crisis. *Journal of Financial Economics* 101, 297-312.
- Financial Accounting Standards Board (2012) "Accounting Standards Update No. 2012-260, Financial Instruments—Credit Losses (Subtopic 825-15): Measurement of Credit Losses on Financial Instruments," December.
- Financial Accounting Standards Board (2016) "Accounting Standards Update No. 2016-13, Financial Instruments—Credit Losses (Topic 326): Measurement of Credit Losses on Financial Instruments," June.
- Francis, W.B., & Osborne, M. (2009). Bank regulation, capital and credit supply: measuring the impact prudential standards, *Occasional Paper 36*, Financial Services Authority.
- Glasserman, P., & Xu, X. (2013). Robust risk measurement and model risk. *Quantitative Finance*, 14(1), 29-58.
- Hanan, E.J. (1971). The identification problem for equation systems with moving average errors. *Econometrica*, 39, 751-766.
- Hanan, E.J. (1988). The statistical theory of linear systems. New York, N.Y.: John Wiley.
- Hansen, L.P., & Sargent, T.J. (2007). Robustness. Princeton, N.J.: Princeton University Press.
- Hirtle, B. A., Kovner, A., Vickery, J. & 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)*.
- International Accounting Standards Board (2014) "International Reporting for Financial Statement Number 9," July.
- Jacobs, Jr., M. (2015). The quantification and aggregation of model risk: perspectives on potential approaches. *The Journal of Financial Engineering and Risk Management*, 2(2), 124–154.
- Jacobs, Jr., M., Klein L. and Merchant, A., 2015a (September), "Emerging Trends in Model Risk Management", Accenture Consulting.
- Jacobs, Jr., M., Karagozoglu, A.K , & Sensenbrenner, F.J. (2015b). Stress testing and model validation: application of the Bayesian approach to a credit risk portfolio. *The Journal of Risk Model Validation*, 9(3), 41–70.
- Jacobs, Jr., M.. (2017). A mixture of distributions model for the term structure of interest rates with an application to risk management. *American Research Journal of Business and Management*, 3(1), 1-17.
- Jacobs, Jr., M., & Sensenbrenner, F.J. (2018a). A comparison of methodologies in the stress testing of credit risk – alternative scenario and dependency constructs. *Quantitative Finance and Economics*, 2(2), 294
- Jacobs, Jr., M. (2018b). The validation of machine learning models for the stress testing of credit risk. *The Journal of Risk Management in Financial Institutions* 11(3), 1-26.–324.
- Kishan, R., and Opiela, T (2000). Bank size, bank capital, and the bank lending channel. *Journal of Money, Credit, and Banking* 32, 121-141.
- R Development Core Team, 2019: "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
- Sims, C.A. (1980). Macroeconomics and reality. *Econometrica*, 48, 1-48.
- Skoglund, J., 2018 (April), "Quantification of Model Risk in Stress Testing and Scenario Analysis", SAS Institute.
- Stock, J.H., & Watson, M.W. (2001). Vector autoregressions. *Journal of Economic Perspectives* 15(4), 101-115.