



Research article

A comparison of methodologies in the stress testing of credit risk – alternative scenario and dependency constructs

Michael Jacobs Jr.^{1,*} and Frank J. Sensenbrenner²

¹ Ph.D., CFA, Principal Director, Accenture Consulting, Finance and Risk Services Advisory/Models, Methodologies & Analytics, 1345 Avenue of the Americas, New York, USA

² Non-Visiting Fellow, US Commodity Futures Trading Commission, 525 West Monroe Street Chicago, IL, 60611, USA

* **Correspondence:** Email: michael.a.jacobs@accenture.com; Tel: +9173242098.

Abstract: In the aftermath of the financial crisis of the last decade, banking supervisors have sought the solution to the problem of determining the optimal capital levels that an institution should hold, in order to support their risk taking activities. The experience of this financial downturn has given rise to the conclusion that traditional approaches, such as regulatory or economic capital are inadequate to this end, leading to the prevalence of supervisory stress testing as a primary tool of prudential supervision. A critical input into this process is the set of macroeconomic scenarios, either provided by the prudential supervisors, or developed by financial institutions. Prevalent among approaches in the industry is the combination of expert opinion and an econometric methodology, for example the *Vector Autoregression* (“VAR”) model that captures the dependency structure among and between macroeconomic explanatory variables and banking loss / income target variables. Despite the prevalence of this approach, we know from the previous finance literature that Gaussian VAR models are unable to cope with the empirical fact of deviation from normality. In this paper we investigate the alternative *Markov Switching VAR* (“MS-VAR”) model, featured more commonly in the academic realm as opposed to being applied in practice. We conduct an empirical experiment using data from regulatory filings and Federal Reserve macroeconomic data released by the regulators for mandated stress testing exercises. Our finding is that the MS-VAR model performs better than the VAR model, both in terms of producing severe scenarios conservative than the VAR model, as well as showing superior predictive accuracy. Furthermore, we find that the multiple equation VAR model outperforms the single equation *autoregressive* (“AR”) models according to various metrics across all modeling segments.

Keywords: stress testing; CCAR; DFAST; credit risk; financial crisis; model risk; vector autoregression; Markov switching model; scenario generation

JEL classification numbers: C31, C53, E27, E47, E58, G01, G17, C54, G21, G28, G38

1. Introduction

Following the financial crisis of the prior decade prudential supervisors have turned to *stress testing* as a primary mechanism with which to gauge the resiliency of financial institutions' with respect to their capital and liquidity resource adequacy to withstand extreme economic scenarios (Acharya, 2009; Demirguc-Kunt et al., 2010). Prior to this, the primary means of risk measurement and management—particularly in the field of credit risk (Merton, 1974)—has been through advanced mathematical, statistical and quantitative techniques and models, which leads to *model risk*. Model risk (Board of Governors of the Federal Reserve System, 2011; “FRB-BOG”) can be defined as the potential that a model does not sufficiently capture the risks it is used to assess, and the danger that it may underestimate potential risks in the future. Stress testing has been used by supervisors to assess the reliability of credit risk models, as can be seen in the revised Basel framework (Basel Committee on Banking Supervision 2006, 2009 a, b, c, d, 2010 a, b; “BCBS”) and the Federal Reserve's *Comprehensive Capital Analysis and Review* (“CCAR”) program.

A clear pattern that we have recorded is that most of the high-profile failures of the financial crisis era included firms for which the supervisors considered the internal risk models to be robust and that they were deemed to have sufficient capital resources in order to survive a downturn (Schuermann, 2014). This set of surprise failures revealed that the question of capital adequacy was not answered, since as internal models estimate a positive default probability that is in line with the supervisor's or institution's risk appetite, the inability of such methodologies to estimate the actual potential dangers was an impetus in the search for different capital adequacy assessment methodologies, prime among these being the discipline of stress testing.

There are a number of modeling considerations that institutions must consider in estimating losses on their credit portfolios, and in the context of stress testing for CCAR purposes we can focus on some particularities relevant to the design of scenarios and the choice of risk factors. There are two broad categories of model types in use. *Bottom-up models* are loan- or obligor-level models used by banks to forecast the expected losses of retail and wholesale loans for each loan. The expected loss is calculated for each loan, and then the sum of expected losses across all loans provides an estimate of portfolio losses, through conditioning on macroeconomic or financial/obligor specific variables. The primary advantages of bottom-up models are the ease of modeling heterogeneity of underlying loans and interaction of loan-level risk factors. The primary disadvantages of loan-level models are that while there are a variety of loan-level methodologies that can be used, these models are much more complex to specify and estimate. These models generally require more sophisticated econometric and simulation techniques, and model validation standards may more stringent. In contrast, *top-down models* are pool (or segment) level models used by banks to forecast charge-off rates by retail and wholesale loan types as a function of macroeconomic and financial variables. In most cases for these models, banks use only one to four macroeconomic and financial risk drivers as explanatory variables. These variables are usually determined by interaction between model development teams and line of

business experts. The primary advantage of top-down models has been the ready availability of data and the simplicity of model estimation. The primary disadvantage of pool-level models is that borrower specific characteristics are generally not used as variables, except at the aggregate level using pool averages. Modeling challenges include determination of an appropriate loss horizon (e.g., for CCAR it is a 9-quarter duration), determination of an appropriate averaging methodology, appropriate data segmentation and loss aggregation, as well as the annualization of loss rates. In this paper we consider top-down models.

This paper shall proceed as follows. Section 2 reviews the available literature on stress testing and scenario generation. Section 3 presents the competing econometric methodologies for generating scenarios, time series *Vector Autoregressive* (“VAR”) and *Markov Switching VAR* (“MS-VAR”) models. Section 4 presents the empirical implementation, the data description, a discussion of the estimation results and their implications. Section 5 concludes the study and provides directions for future avenues of research.

2. Review of the literature

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

Stress testing first appears in the supervisory realm under the auspices of the Basel I Accord within the 1995 Amendment on Market Risks (BCBS 1988, 1996). The contemporaneous publication of RiskMetrics™ (J.P. Morgan, 1994) established the practice of market risk management as an analytical discipline in its own right and subsumed several stress testing methodologies thereto developed in this context. Jorion (1996) discusses aspects of stress testing in a book addressing *Value-at-Risk* (“VaR”). Kupiec (1999), Berkowitz et al. (1999) and the *Committee on Global Financial Systems* survey (2000; “CGFS”) analyze stress testing in a trading and treasury VaR context. Mosser et al. (2001) noted that most stress testing of the period relied upon transparent and easily identifiable historical market factors with respect to asset classes in the trading book.

However, in the case of the banking book (e.g., corporate/C&I or consumer loans), this approach of asset class shocks does not carry over as well, as to the extent these are less marketable there are more idiosyncrasies to account for. Therefore, stress testing with respect to credit risk has evolved later and as a separate discipline in the domain of credit portfolio modeling. However, even in the seminal examples of CreditMetrics™ (J.P. Morgan, 1997) and CreditRisk+™ (Wilde, 1997), stress testing was not a component of such models. The commonality of all such credit portfolio models was subsequently demonstrated (Koyluoglu and Hickman, 1998), as well as the correspondence between the state of the economy and the credit loss distribution, and therefore that this framework is naturally amenable to stress testing. In this spirit, a class of models was built upon the CreditMetrics™ framework through macroeconomic stress testing on credit portfolios using credit

migration matrices (Bangia et al., 2002). Nevertheless, prior to the financial crisis supervisory guidance for stress testing were rather unformed in the banking book as compared to other areas such as interest rate, counterparty or country risk (FRB-BOG 1996, 1999, 2002).

In the decade following the financial crisis there is a great expansion of the literature on stress testing. Foglia (2009) survey the existing credit risk stress testing literature of this era. Inanoglu and Jacobs, Jr. (2009) address the aggregation of risk types of capital models in the stress testing and sensitivity analysis of economic capital. Jacobs, Jr. (2010) extends Jacobs and Inanoglu (2009) to the validation of models for stressed capital. Schuermann (2014) analyzes the predominance of stress testing as a supervisory tool in terms of rationales for its utility, outlines for its execution, as well as guidelines and opinions on disseminating the output under various conditions. Jacobs, Jr. (2013) surveys of practices and supervisory expectations for the stress testing of credit risk in the context of a ratings migration methodology in the CreditMetricsTM framework. Rebonato (2010) proposes a Bayesian casual network model, for stress testing having the capability to cohesively incorporate expert knowledge in the model design and methodology of the stress testing process. Another recent study features the application of a Bayesian regression model for credit loss implemented using Fed Y9 data, wherein regulated financial institutions report their gains and losses in conjunction with Federal Reserve scenarios, which can formally incorporate exogenous factors such as such supervisory scenarios, and also quantify the uncertainty in model output that results from stochastic model inputs (Jacobs, Jr. et al., 2015). Jacobs (2015) presents an analysis of the impact of asset price bubbles on standard credit risk measures and provides evidence that asset price bubbles are a phenomenon that must be taken into consideration in the proper determination of economic capital for both credit risk management and measurement purposes. The author also calibrates the model to historical equity prices and in a stress testing exercise project credit losses on both baseline and stressed conditions for bubble and non-bubble parameter estimate settings. Jacobs (2017b) extends Jacobs (2015) by performing a sensitivity analysis of the models with respect to key parameters, empirically calibrates the model to a long history of equity prices, and simulates the model under normal and stressed parameter settings. While the author find statistically significant evidence that the historical S&P index exhibits only mild bubble behavior, this translates in underestimation of potential extreme credit losses according to standard measures by an order of magnitude; however, the degree of relative underestimation of risk due to asset price bubbles is significantly attenuated under stressed parameter setting in the model.

The relative merits of various risk measures and the aggregation of varying risk types, classic examples being *Value-at-Risk* (“VaR”) and related quantities, have been discussed extensively by prior research (Jorion 1997, 2006). An important result in the domain of modeling dependency structures is a general result of mathematical statistics due to Sklar (1956), allowing the combination of arbitrary marginal risk distributions into a joint distribution while preserving a non-normal correlation structure, readily found an application in finance. Among the early academics to introduce this methodology is Embrechts et al. (1999, 2002, 2003). This was applied to credit risk management and credit derivatives by Li (2000). The notion of copulas as a generalization of dependence according to linear correlations is used as a motivation for applying the technique to understanding tail events in Frey and McNeil (2001). This treatment of tail dependence contrasts to Poon et al. (2004), who instead use a data intensive multivariate extension of extreme value theory, which requires observations of joint tail events. Inanoglu and Jacobs (2009) develop a coherent approach to aggregating different risk types for diversified financial institutions. The authors model the main

risks faced—market, credit and operational—that have distinct distributional properties, that historically have been modeled in differing framework, contributing to the modeling effort by providing tools and insights to practitioners and regulators.

On the topic of scenario generation we find rather limited literature to date. Bidder and McKenna (2015) propose the use of robust forecasting analysis to estimate adverse scenarios in stress testing that are generated from a single pessimistic view with respect to a baseline predictive model, the so-called “worst case distribution”, a means of assessing weaknesses within a framework that can account for model misspecifications in a general sense. Frame et al. (2015) examine the *Office of Federal Housing Enterprise Oversight’s* (“OFHEO”) risk-based capital stress test for Fannie Mae and Freddie Mac. The authors conclude that the key driver in the model, 30-year fixed-rate mortgage performance, left the model specification and parameters settings fixed in the forecast period, and that the house price stress scenario was insufficiently severe, resulting in a significant underprediction of mortgage credit losses and associated capital requirements during the downturn. Finally, we extend Jacobs (2016) on the topic of scenario generation and stress testing employing the MS-VAR model, by in addition to this framework considering the prediction of credit loss in a multiple equation setting.

3. 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 periods¹. 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, 1976; 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

¹ 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.)

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.

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

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

Which is equivalent to:

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

Where $\Phi(B) = \mathbf{I}_r - \sum_{j=1}^p \Phi_j B^j$, $\Theta(B) = \sum_{j=0}^s \Theta_j B^j$ and $\Theta^*(B) = \mathbf{I}_r - \sum_{j=1}^q \Theta_j^* B^j$ are *autoregressive lag polynomials* of respective orders p , s and q , respectively, and B is the *back-shift operator* that satisfies $B^i X_t = X_{t-i}$ for any process $\{X_t\}$. It is common to assume that the input process \mathbf{X}_t is generated independently of the noise process $\mathbf{E}_t = (E_{1t}, \dots, E_{kt})^T$ ². 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 variables³. 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 $ARMAX(p, q, s)$

² 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 to indicate the effect of output due to unusual intervention events in the system.

³ 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 is referred to as the *reduced form model*, which has a *state space representation* (Hanan, 1988).

(“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 $\boldsymbol{\Theta}_j$ have a diagonal structure (Brockwell and Davis, 1991).

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

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

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 $\mathbf{B} = (\boldsymbol{\Phi}^T, \boldsymbol{\Theta}^T)^T \in R^{p+s}$ 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:

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

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

$$E[\mathbf{Y}_t | \boldsymbol{\Psi}_{t-1}, s_t] = \sum_{j=1}^p \boldsymbol{\Phi}_j \mathbf{Y}_{t-j}(s_t) + \sum_{j=1}^s \boldsymbol{\Theta}_j(s_t) \mathbf{X}_{t-j} \quad (5)$$

We define the innovation term as:

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

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

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

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:

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

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, 2017). 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.

4. Empirical implementation

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 scenario. 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 as well as severely adverse scenarios:

- 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")
- 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 multiple and univariate VAR and MS-VAR models⁴:

⁴ 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, 2017).

- 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:

- Real Gross Domestic Investment (“RDIG”)
- Unemployment Rate (“UNEMP”)
- Commercial Real Estate Price Index (“CREPI”)
- BBB Corporate Credit Spread (“BBBCS”)
- CBOE’s Equity Volatility Index (“VIX”)

Similarly, we identify the following loss segments (with loss measured by Gross Charge-off Rates—“GCOs”) according to the same criteria, in conjunction with the requirement that they cover the most prevalent portfolio types in typical traditional banking institutions:

- Residential Real Estate (“RESI”)
- Commercial Real Estate (“CRE”)
- Consumer Credit (“CONS”)
- Commercial and Industrial (“C&I”)

This historical data, 60 quarterly observations from 1Q01 to 4Q15⁵, are summarized in Table 1 in terms of distributional statistics and correlations, as in Figures 1 to 9 of this section. Across all series when looking at the time series dimension (in the left panels of the figures, in levels in the top and percent changes on the bottom), we observe that the credit cycle is clearly reflected, with indicators of economic or financial stress (health) and charge-off loss rates 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. Shifting focus to the smoothed histogram graphs (in the right panels of the figures, in levels in the top and percent changes on the bottom), we note that there are significant deviations from normality in terms of excess skewness and excess kurtosis

⁵ 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—refer to the Diebold-Mariano tests at the end of this section. We also choose to start our sample in 2001, as we believe that the earlier period would reflect economic conditions not relevant for the last decade and also because in the financial industry this is a standard starting point for CCAR and DFAST stress testing models.

Table 1. Summary statistics of historical Y9 credit loss rates and Federal Reserve Macroeconomic Variables.

	Output Variables - Loss Observation Segments								Input Variables - Macroeconomic Factors													
	Residential Real Estate		Commercial Real Estate		Consumer Credit		Commercial and Industrial		Real Gross Domestic Investment		Unemployment Rate		Commercial Real Estate Price Index		Baa 10 Year Corporate Bond Yield		CBOE Market Volatility Index					
	Level	Change	Level	Change	Level	Change	Level	Change	Level	Change	Level	Change	Level	Change	Level	Change	Level	Change				
	Count	Mean	Std. Dev.	Min.	25 th Prc.	Median	75 th Prc.	Max.	Coef. Var.	Skewness	Kurtosis	Count	Mean	Std. Dev.	Min.	25 th Prc.	Median	75 th Prc.	Max.	Coef. Var.	Skewness	Kurtosis
Count	60	59	60	59	60	59	60	59	60	59	60	59	60	59	60	59	60	59	60	59	60	59
Mean	0.73	6.43%	0.61	10.11	3.04	0.18%	0.87	0.07%	2.20	-1.0%	6.50	0.60%	198.9	1.13%	3.67	0.01%	26.5	3.31%				
Std. Dev.	0.794	40.7%	0.82	70.9%	1.22	11.3%	0.686	20.2%	3.65	6.0%	1.70	6.99%	37.57	4.19%	0.97	9.87%	12.2	31.9%				
Min.	0.06	-61.8%	0.01	-80.0%	1.74	-42.1%	0.18	-28.3%	-12.6	-244%	4.20	-7.6%	135.8	-14.6%	1.60	-24.2%	12.7	-49.5%				
25 th Prc.	0.14	-15%	0.06	-20%	2.30	-5.2%	0.30	-16%	1.03	-182%	5.30	-2.5%	170.6	0.00%	2.90	-6.1%	19.1	-17.4%				
Median	0.28	-1.31%	0.15	-7.18%	2.73	-0.62%	0.62	-3.95%	2.70	5.00%	5.85	-1.2%	197.5	1.27%	3.80	0.94%	21.5	-1.44%				
75 th Prc.	1.38	11.5%	0.95	17.6%	3.17	7.4%	1.34	13.6%	3.40	263%	7.85	1.07%	233.1	3.85%	4.40	5.41%	30.9	7.36%				
Max.	2.78	165%	2.95	400.0%	6.72	29.54%	2.53	60.0%	11.8	1640%	9.90	38.1%	258.9	9.81%	5.40	35.00%	80.9	111.5%				
Coef. Var.	1.08	6.33	1.36	7.02	0.40	64.23	0.78	289.97	1.65	599.21	0.26	11.68	0.19	3.73	0.27	1283.59	0.46	9.64				
Skewness	1.04	2.09	1.45	3.39	1.52	-0.50	0.92	0.80	-1.20	-1.05	0.71	3.31	-0.04	-1.65	-0.4	0.3860	2.10	1.54				
Kurtosis	-0.30	6.06	0.79	15.66	1.79	2.77	-0.27	0.21	5.50	5.41	-0.8	14.31	-1.29	4.62	-0.6	1.93	5.91	2.36				

Table 2. Correlations amongst Federal Reserve Macroeconomic Variables.

		Real Gross Domestic Investment				Unemployment Rate				Commercial Real Estate Price Index				Baa 10 Year Corporate Bond Yield				CBOE Market Volatility Index						
		Level	Change	Level	Change	Level	Change	Level	Change	Level	Change	Level	Change	Level	Change	Level	Change	Level	Change					
		Level	Change	Level	Change	Level	Change	Level	Change	Level	Change	Level	Change	Level	Change	Level	Change	Level	Change					
Real Gross Domestic Investment	Level		1																					
	Change	-41.11%		1																				
Unemployment Rate	Level	-14.78%	-1.62%		1																			
	Change	-19.37%	-7.83%	48.07%		1																		
Commercial Real Estate Price Index	Level	22.87%	0.81%	-35.99%	-12.30%		1																	
	Change	2.20%	8.95%	-24.43%	-43.00%	-42.18%		1																
Baa 10 Year Corporate Bond Rate	Level	-21.46%	-1.55%	30.03%	67.07%	-21.74%	-6.88%		1															
	Change	-1.81%	-9.31%	31.66%	11.13%	-38.29%	-34.33%	18.27%		1														
CBOE Equity Market Volatility Index	Level	-17.81%	-7.71%	36.50%	49.28%	-64.27%	-25.93%	20.57%	15.35%		1													
	Change	-14.23%	-3.34%	-7.29%	6.57%	-14.75%	-28.98%	15.29%	34.33%	44.56%		1												

Table 3. Correlations among historical Y9 credit loss rates.

		Residential Real Estate		Commercial Real Estate		Consumer Credit		Commercial and Industrial	
		Level	Change	Level	Change	Level	Change	Level	Change
		Level	1						
Residential Real Estate	Change	2.74%	1						
	Level	86.26%	3.05%	1					
Commercial Real Estate	Change	6.25%	16.51%	2.22%	1				
	Level	81.42%	1.71%	90.35%	7.41%	1			
Consumer Credit	Change	2.00%	26.15%	2.68%	16.44%	20.08%	1		
	Level	54.61%	3.92%	64.56%	-3.04%	79.79%	18.24%	1	
Commercial and Industrial	Change	-2.81%	15.49%	0.37%	19.19%	0.1822%	38.07%	19.81%	1

Table 4. Correlations amongst historical Y9 credit loss rates and Federal Reserve Macroeconomic Variables.

		Residential Real Estate		Commercial Real Estate		Consumer Credit		Commercial and Industrial	
		Level	Change	Level	Change	Level	Change	Level	Change
		Level	Change	Level	Change	Level	Change	Level	Change
Real Gross Domestic	Level	-9.39%	-10.62%	-5.73%	-9.87%	-9.63%	-14.63%	-14.03%	-1.44%
Investment	Change	-0.12%	4.73%	-2.06%	-7.71%	-3.90%	6.13%	-4.88%	-21.47%
	Level	92.76%	5.79%	89.80%	-2.27%	77.51%	11.11%	58.80%	12.57%
Unemployment Rate	Change	12.96%	33.55%	-10.18%	27.72%	7.67%	25.48%	3.26%	41.65%
Commercial Real Estate	Level	-26.31%	45.18%	-33.98%	-15.42%	-54.65%	44.23%	-79.65%	-48.00%
Price Index	Change	-4.00%	-5.50%	-0.11%	-11.47%	2.57%	-10.03%	-5.19%	-1.24%
Baa 10 Year Corporate Bond	Level	31.90%	11.26%	25.65%	5.62%	-8.32%	-15.36%	19.16%	5.36%
Rate	Change	23.67%	17.36%	-27.77%	13.18%	-28.35%	-10.77%	31.62%	35.73%
CBOE Equity Market	Level	22.16%	40.93%	26.25%	23.47%	29.05%	28.40%	33.17%	46.76%
Volatility Index	Change	9.67%	19.21%	1.05%	27.16%	3.50%	8.02%	6.82%	10.14%

relative to the Gaussian case, although the extent of these deviations exhibits significant variations across variables (e.g., in the case of the VIX, the non-normality is extreme, and obviously in the case of certain indices or loss rates the bounded domain are clear violations of normality). Furthermore, such deviations from normality are accentuated by an order of magnitude when examining these distributions of the variables in percent change form, which holds generally although with the extent of the deviations varying somewhat across variables. Finally, we note that in general the variation relative to the mean is an order of magnitude greater than looking at percent changes relative to levels.

The correlations amongst all of the independent and dependent variables, in both their level and percentage change forms, are displayed in Tables 2 through 4. First, we will describe main features of the dependency structure within the group of input macroeconomic variables, then the same for the output loss rate variables, and finally the cross-correlations between these two groups. We observe that all correlations have intuitive signs and magnitudes that suggest significant relationships, although the latter are not large enough to suggest any issues with multicollinearity. While the correlations of the percent change transformations are generally lower, they are still intuitive and of reasonable magnitudes. We also note that percent changes of variables are negatively (positively) correlated with levels when indicators are those of economic strength (weakness). The correlation matrix amongst the macroeconomic variables appears in Table 2. For example, considering some of the stronger relationships amongst the levels, the correlations between UNEMP/VIX, CREPI/UNEMP and BBBCY/RDIG are 36.5%, -36.0% and -21.5% , respectively. For example, considering some of the stronger relationships amongst the percent changes, the correlations between BBBCR/CREPI, UNEMP/RDIG and VIX/CREPI are 34.3%, -7.8% and 28.6%, respectively. The correlation matrix amongst the credit loss rate variables appear in Table 3. For example, considering some of the stronger relationships amongst the levels, the correlations between CRE/RESI, CONS/CRE and CNI/CONS are 86.3%, 90.4% and 79.8%, respectively. For example, considering some of the stronger relationships amongst the percent changes, the correlations between CONS/CRE, CNI/CRE, and CNI/CONS are 26.2%, 15.5% and 38.1%, respectively.

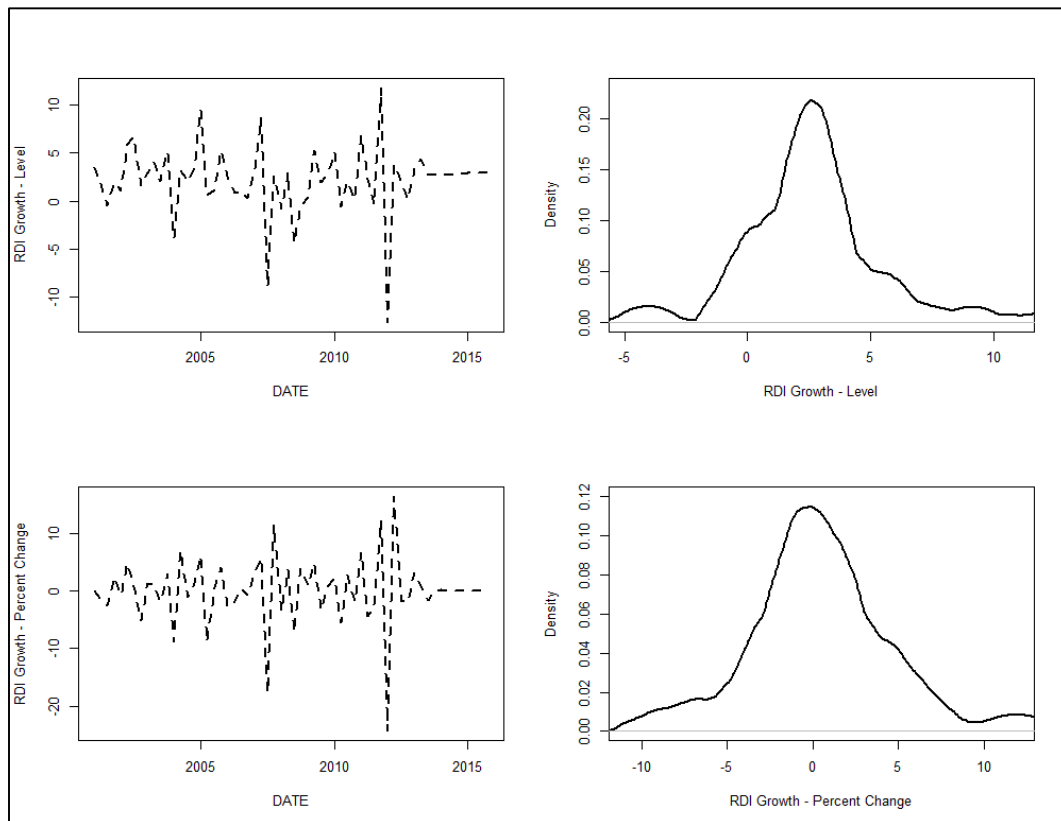


Figure 1. Time series and kernel density plot—Real Domestic Investment Growth.

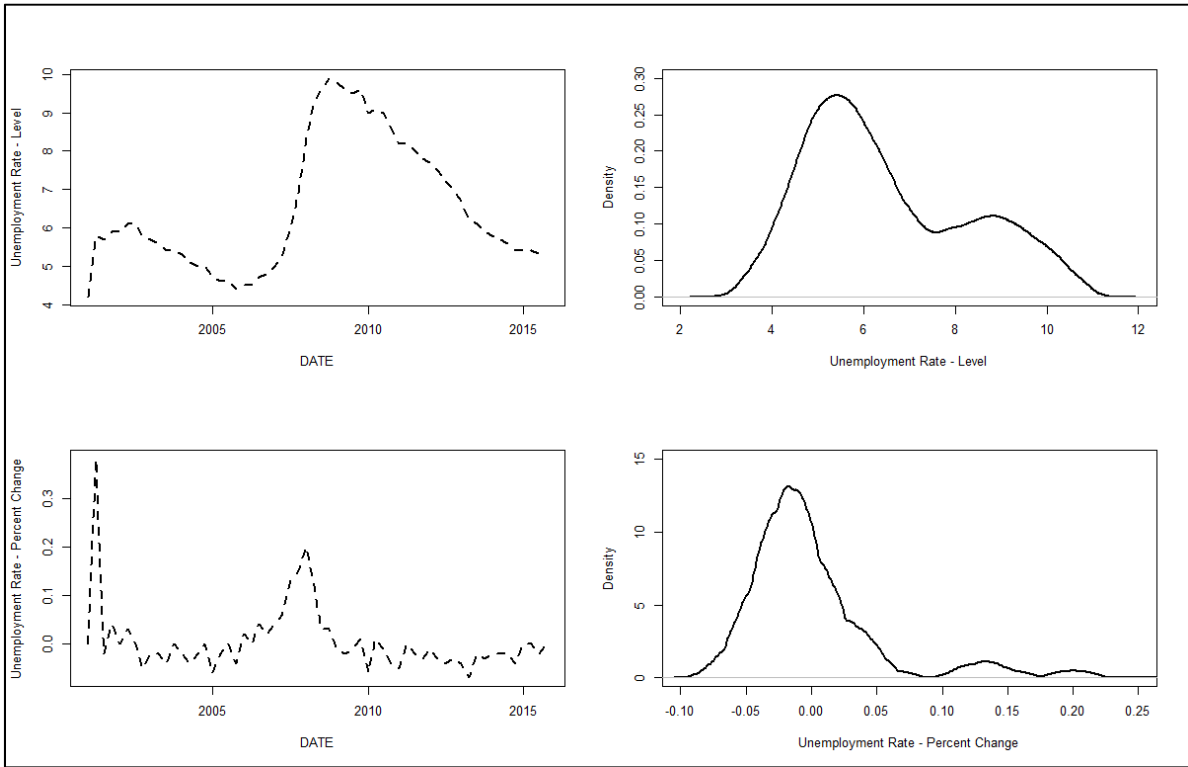


Figure 2. Time series and kernel density plot—Unemployment Rate.

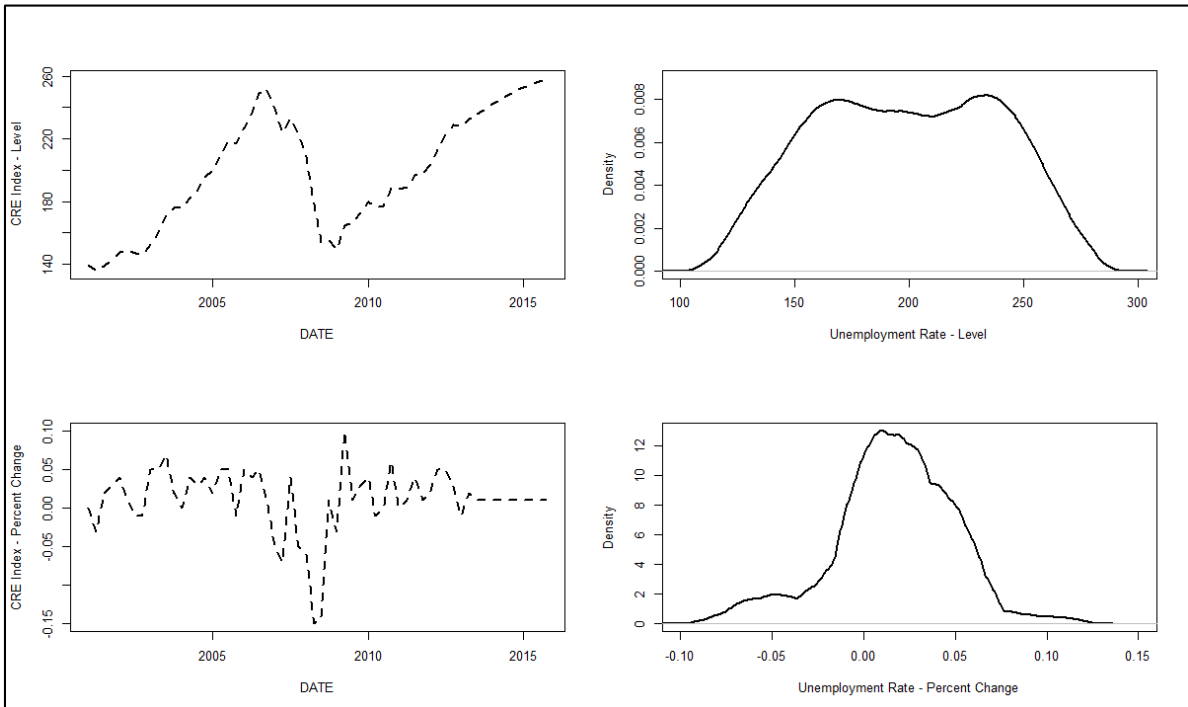


Figure 3. Time series and kernel density plot—Commercial Real Estate Index.

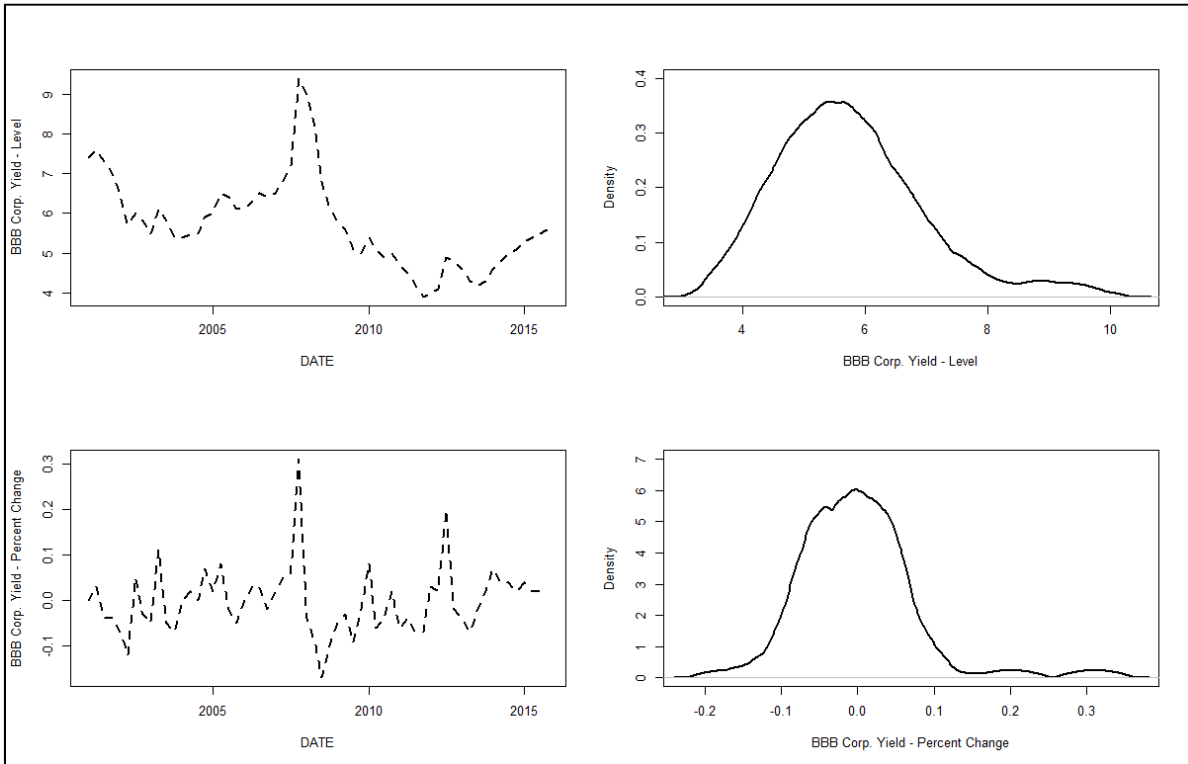


Figure 4. Time series and kernel density plot—BBB Corporate Bond Rate.

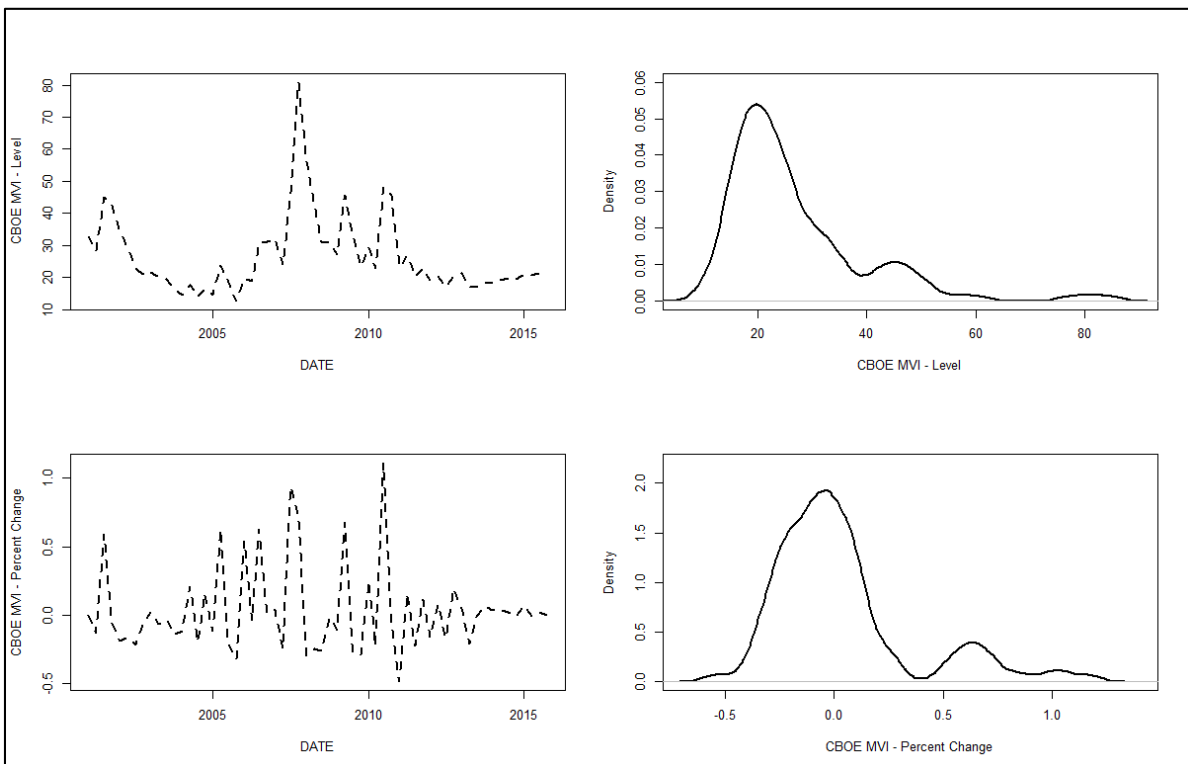


Figure 5. Time series and kernel density plot—CBOE Equity Market Volatility Index.

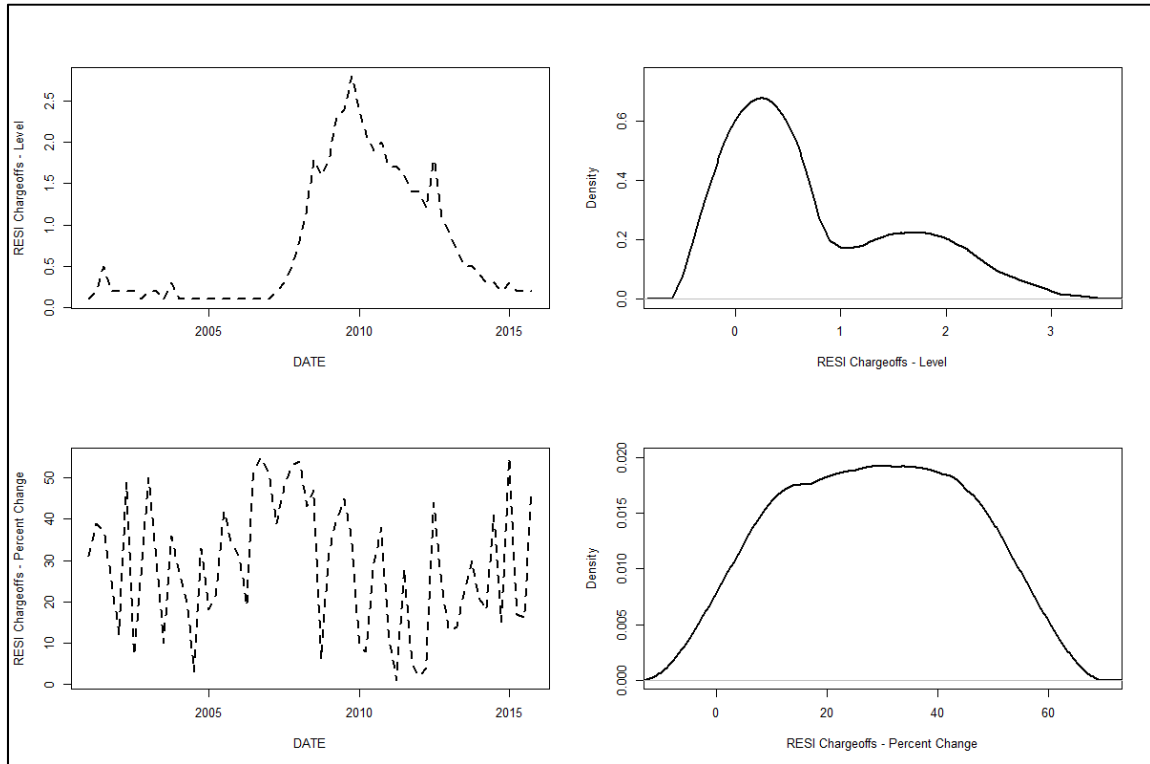


Figure 6. Time series and kernel density plot—Residential Real Estate Loan Charge-off Rates.

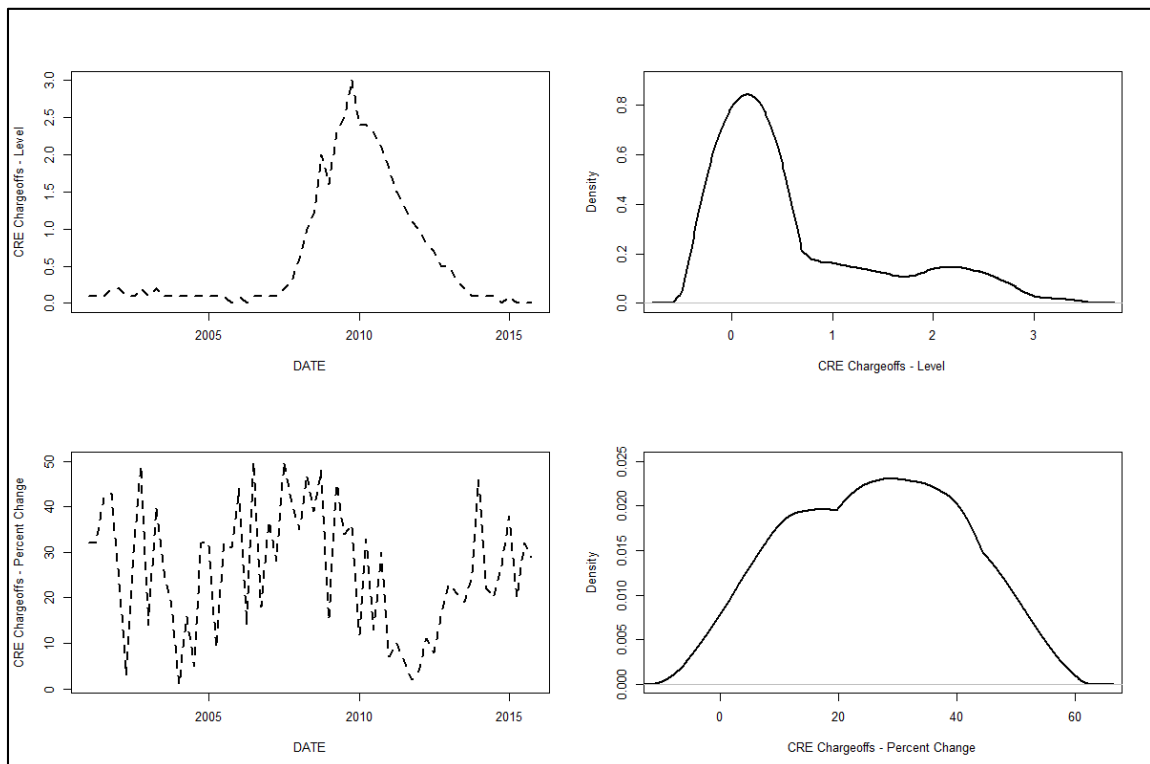


Figure 7. Time series and kernel density plot—Commercial Real Estate Loan Charge-off Rates.

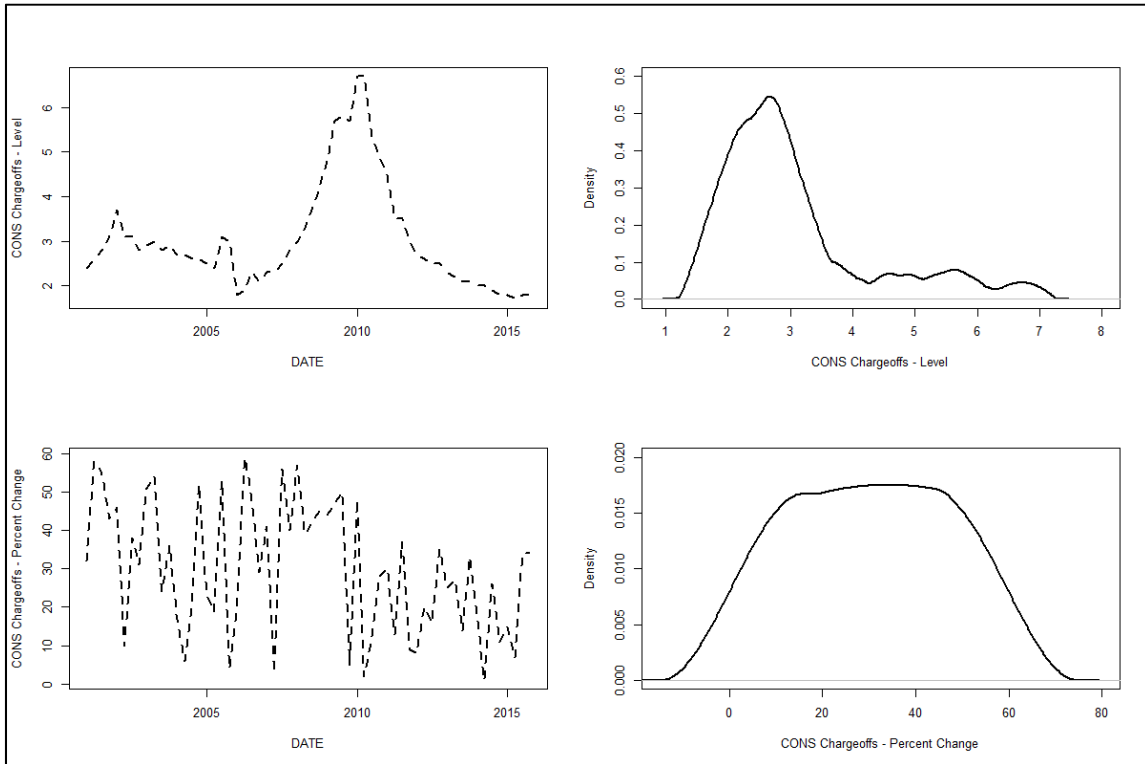


Figure 8. Time series and kernel density plot—Consumer Loan Charge-off Rates.

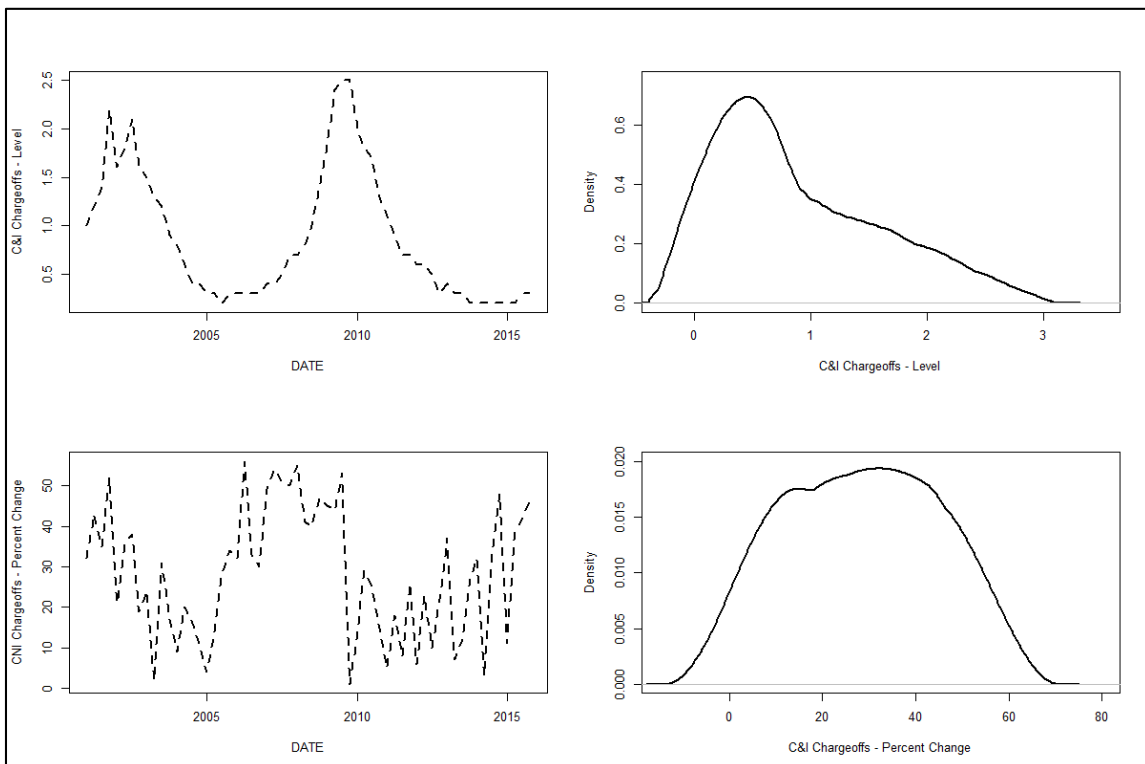


Figure 9. Time series and kernel density plot—Commercial and Industrial Loan Charge-off Rates.

Table 5. Augmented Dickey-Fuller and Kwiatkowski-Phillips-Schmidt-Shin stationarity test statistics of Credit Loss Rates and Macroeconomic Variables.

Stationarity Test		Levels		Percent Changes	
		Test		Test	
		Statistic	P-Value	Statistic	P-Value
Augmented Dickey-Fuller	Real Gross Domestic Investment	-3.8979	0.0202	-3.8979	0.0202
	Unemployment Rate	-2.1582	0.5108	-2.3114	0.0224
	Commercial Real Estate Price Index	-2.0886	0.5390	-3.0847	0.0455
	BBB 10 Year Corporate Bond Yield	-2.1872	0.4991	-4.4104	0.0100
	CBOE Market Volatility Index	-2.3422	0.4365	-5.1761	0.0100
	Residential Real Estate Loss Rate	-1.4057	0.8150	-1.9883	0.0483
	Commercial Real Estate Loss Rate	-1.9643	0.5892	-1.8584	0.0451
	Consumer Credit Loss Rate	-2.3685	0.4258	-3.6123	0.0398
	Commercial and Industrial Loss Rate	-3.4832	0.0512	-1.6232	0.0291
Kwiatkowski-Phillips-Schmidt-Shin	Real Gross Domestic Investment	0.6870	0.0023	0.0109	0.3166
	Unemployment Rate	0.4553	0.0091	0.1178	0.1971
	Commercial Real Estate Price Index	0.3420	0.0014	-0.2742	0.2741
	BBB 10 Year Corporate Bond Rate	0.3813	0.0079	-0.1303	0.2698
	CBOE Equity Market Volatility Index	0.4507	0.0000	0.0139	0.1389
	Residential Real Estate Loss Rate	0.4652	0.0014	0.1995	0.3581
	Commercial Real Estate Loss Rate	0.2462	0.0021	-0.1965	0.3374
	Consumer Credit Loss Rate	0.4384	0.0065	0.0659	0.1207
	Commercial and Industrial Loss Rate	0.4886	0.0098	0.0437	0.4992

The correlation matrix amongst the credit loss rate and macroeconomic variables appear in the Table 4. For example, considering some of the stronger relationships of the levels, the correlations between UNEMP/CRE, CREPI/CNI and UNEMP/RESI are 89.8%, 58.8% and 92.8%, respectively. For example, considering some of the stronger relationships amongst the percent changes, the correlations between UNEMP/CNI, UNEMP/CONS, and VIX/CRE are 41.7%, 25.5% and 27.2%, respectively.

In Table 5 we display the *Augmented Dickey-Fuller* (“ADF”; Dickey and Fuller, 1981) statistics of the macroeconomic variables under consideration. We observe that we only reject the null hypothesis of a unit root process (or of non-stationarity) in one case for the variables in level form, whereas in percent change for we are able to reject this in all cases at the 5% confidence level or better. We also show results of the *Kwiatkowski, Phillips, Schmidt and Shin* (“KPSS”; Kwiatkowski et al., 1992) test, in which the null hypothesis is a stationary time series, where we are not able to reject the null hypothesis in all cases for the percent changes, but not for the variables in level form where we do reject the null hypothesis in some cases. Taken in combination with the observations regarding the correlation analysis of Table 1, this leads to the choice of modeling the percent changes in the macroeconomic variables in order to generate base and stress scenarios. As a practice, when modeling in a time series framework, it is preferable to work with data that are jointly stationary.

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 Table 1, 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, and left skewness in variables where declines are a sign of weakening conditions such as RDIG). 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.⁶

Another important modeling consideration with respect to scenario generation is the methodology for partitioning the space of scenario paths across our 6 macroeconomic variables for the Base and Severe scenario. In the case of the Severe scenario, we choose to identify such a path in which *all six* macroeconomic variables exceed their historical 99.0th percentile in at least a single quarter, and then in that set for each variable we take an average across such paths in each quarter. It is our view that this is a reasonable definition of a Severe scenario, and in our risk advisory practice we have observed similar definitions in the industry.⁷ 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 14 where we show for each macroeconomic variable the base and severe scenarios for the VAR and MS-VAR models⁸, and also compare this to the corresponding Fed scenarios, along with the historical time series. We make the following general conclusions regarding the different scenario generation methodologies:

- In the Severe scenario, the MS-VAR model is far more conservative than the VAR model, and is always at least matching and in some cases even well exceeding historical peaks or troughs in the adverse direction.
- In terms of magnitude, the VAR model is similar to the Fed scenarios, but the trajectories of either the VAR or MS-VAR model tend to be more regular, rising at a more gradual pace into the forecast period.
- In the Base scenarios, the Fed model is rather similar to the VAR model, but in all cases the MS-VAR model produces a higher base, which is driven by the skewness of the mixture error distribution.

⁶ This is similar to the findings of Loregian and Meucci (2016) and Jacobs (2017a) 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.

⁷ We have performed a sensitivity analysis, available upon request, using the 95th and 99.9th percentiles, and the results are not greatly changed.

⁸ 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, 2017).

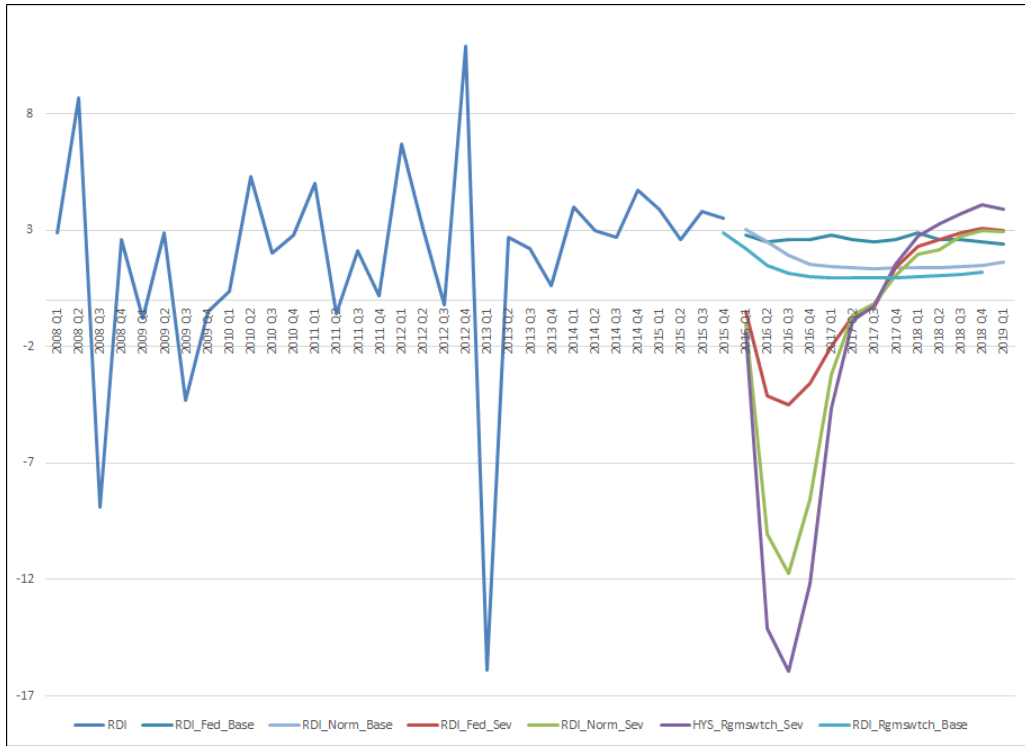


Figure 10. Historical time series, Base and Severe scenarios for the VAR, MS-VAR and Fed Models—Real Disposable Income Growth.

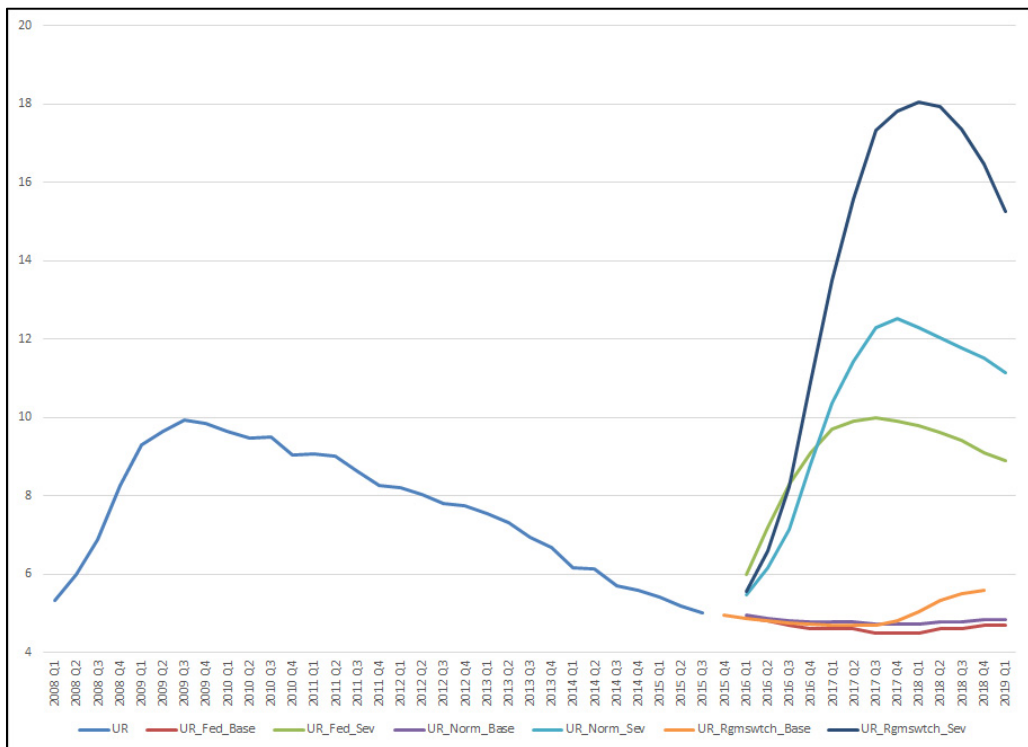


Figure 11. Historical time series, Base and Severe scenarios for the VAR, MS-VAR and Fed Models—Unemployment Rate.

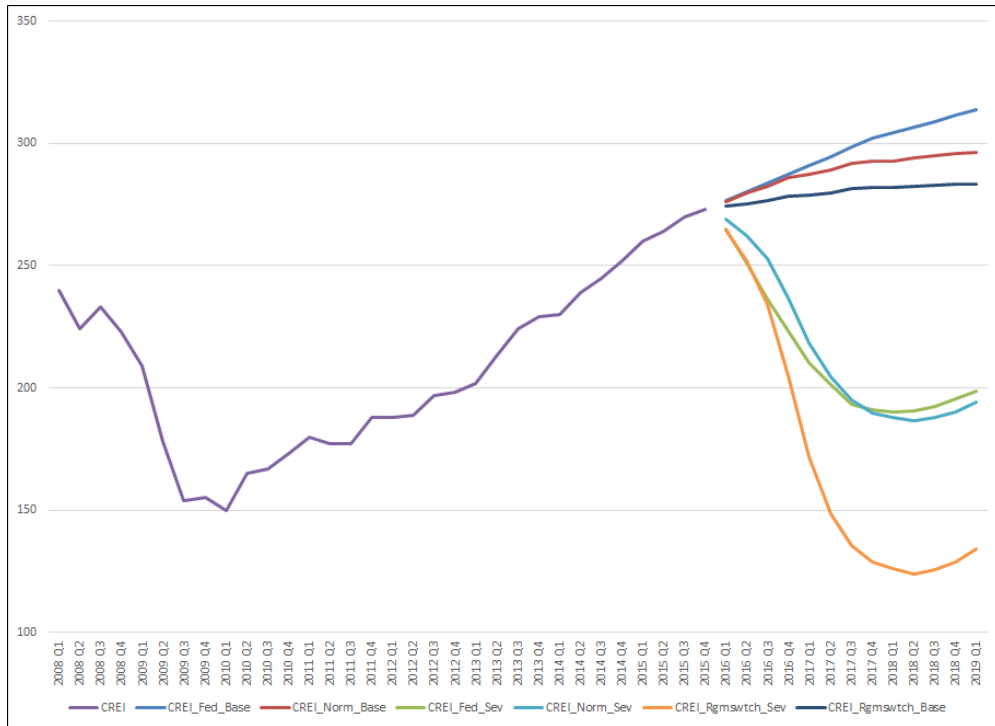


Figure 12. Historical time series, Base and Severe scenarios for the VAR, MS-VAR and Fed Models—Commercial Real Estate Price Index.

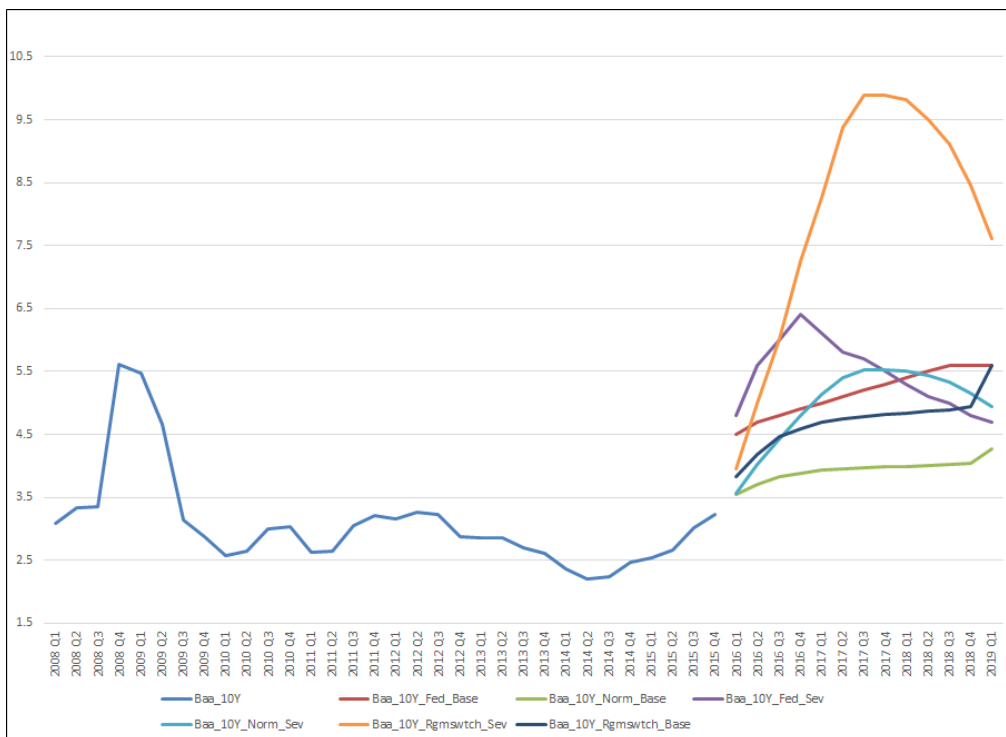


Figure 13. Historical time series, Base and Severe scenarios for the VAR, MS-VAR and Fed Models—BB Corporate Credit Spread.

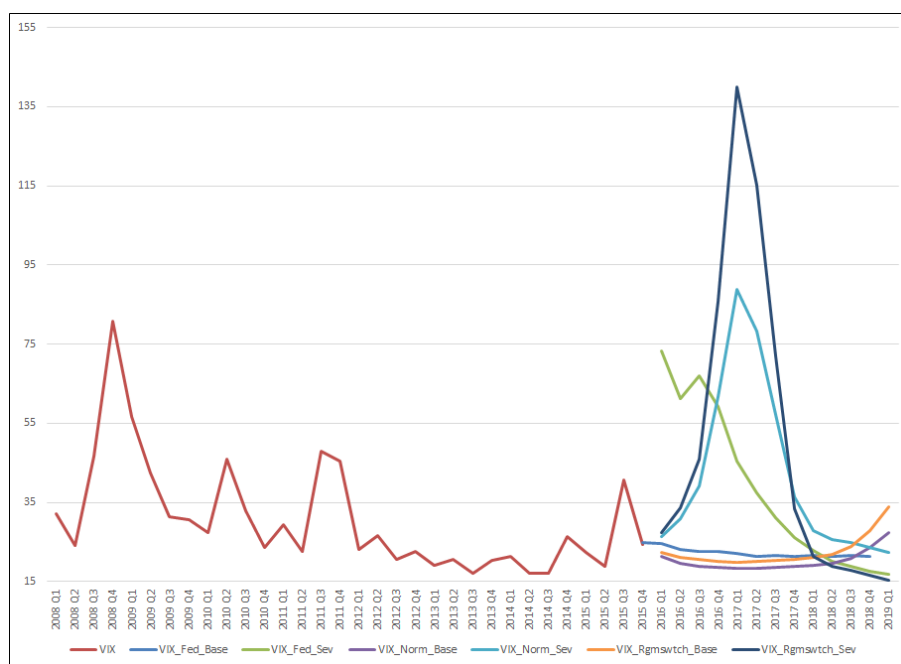


Figure 14. Historical time series, Base and Severe scenarios for the VAR, MS-VAR and Fed Models—VIX Equity Market Volatility Index.

The estimation results are summarized in Tables 6 and 7. Table 6 tabulates the results of the VAR (1) estimation of a 4-equation system, while Table 7 tabulates the results of the single equation AR (1) models for each portfolio segment separately. Below we highlight the main conclusions of this study in regard to the difference between the multiple and single equation estimations (detailed descriptions of estimation results and residual diagnostics are given in an Addendum to this paper):

- In both the VAR and AR models, all coefficient estimates are of intuitive sign, and statistically significant at conventional confidence levels, although we note that the significance levels are generally at higher levels for the VAR as compared to the AR models.
- Residual diagnostics reveal lack of serial autocorrelation and a Gaussian distribution in both VAR and AR models, although we note that the quality of residuals is somewhat better for the VAR as compared to the AR models.
- Across all 4 segments, according to the likelihood ratio statistic, we reject the hypothesis that the restrictions of the single equation AR models are justified.
- The results of the estimation are broadly consistent across the VAR and AR models, but with a few notable differences, such that the autocorrelation terms are larger in the AR models than in the VAR model.
- The VAR models show greater sensitivity to macroeconomic factors than do the AR models.
- The VAR models are generally more accurate according to standard measures of model fit with respect to each segment.
- The VAR is more conservative than the AR as by measured by cumulative 9-quarter percentage error in the sense of under-predicting (over-predicting) to a lesser degree during the downturn (recent) period.

Table 6. Vector Autoregressive Model Estimation Results Compared (Fed Macroeconomic Variables and Aggregate Y9 Bank Charge-offs).

		Coefficient Estimates				P-Values			
		Residential		Commercial		Residential		Commercial	
Statistic		Real Estate	Commercial Real Estate	Consumer Credit	and Industrial	Real Estate	Commercial Real Estate	Consumer Credit	and Industrial
Observation Segments	Output Variables - Loss								
	Residential Real Estate	0.6808	0.3765	0.0576	0.0055	0.00E+00	0.00E+00	0.00E+00	3.84E-02
	Commercial Real Estate	0.3510	0.5205	0.0950	0.0869	0.00E+00	0.00E+00	0.00E+00	1.17E-07
	Consumer Credit	0.2225	0.5346	0.3772	0.3698	2.74E-11	0.00E+00	0.00E+00	0.00E+00
Input Variables - Macroeconomic Factors	Commercial and Industrial	0.3370	0.2637	0.0054	0.7879	0.00E+00	0.00E+00	2.41E-02	0.00E+00
	Real Gross Domestic Investment	-0.0156	-0.0185	-0.0339	-0.0047	6.75E-04	1.32E-05	2.49E-02	2.14E-01
	Unemployment Rate	0.0215	0.0157	0.0598	0.0164	7.86E-10	5.19E-07	7.99E-04	7.73E-05
	Commercial Real Estate Price Index	-0.0006	-0.0227	-0.0063	-0.0053	3.52E-04	0.00E+00	7.32E-11	0.00E+00
	BBB 10 Year Corporate Bond Rate	0.0675	0.0226	0.2731	0.0676	0.00E+00	6.50E-13	0.00E+00	0.00E+00
	CBOE Equity Market Volatility Index	0.0012	0.0118	0.0158	0.0118	5.24E-02	9.83E-13	2.07E-10	1.21E-11
	Log-Likelihood					-42.90			
	Root Mean Squared Error / Mean	0.2609	0.3006	0.1016	0.2408				
	Squared Correlation	0.9397	0.9503	0.9352	0.9038				
	Cumulative Percentage Error - Entire Sample	4.126%	4.794%	3.913%	5.714%				
Model Performance Metrics	Cumulative Percentage Error - Recent Period	5.158%	5.565%	6.673%	3.995%				
	Cumulative Percentage Error - Downturn Period	-3.686%	-4.370%	-3.118%	-4.072%				

The results of the scenario analysis with respect to the credit loss segments, for both AR and VAR estimation, as well as for the three scenario generation methodologies (Fed, VAR and MS-VAR), are shown in Tables 8 and 9, as well as in Figures 15 through 18. The results across modeling segments are in line with the scenarios analysis as per macroeconomic variable as discussed in this section, but these results in terms of conservatism of the Severe forecasts are accentuated in the VAR model and dampened in the AR models. In the severe scenario, the MS-VAR model is far more conservative than the VAR model, reflecting the greater sensitivity to macroeconomic factors noted in the estimation results, and always at least matching and in some cases well exceeding historical peaks or troughs in the adverse direction.

Table 7. Single Equation Autoregressive Model Estimation Results Compared (Fed Macroeconomic Variables and Aggregate Y9 Bank Charge-offs).

Statistic		Coefficient Estimates				P-Values			
		Residential		Commercial		Residential		Commercial	
		Real Estate	Commercial Real Estate	Consumer Credit	and Industrial	Real Estate	Commercial Real Estate	Consumer Credit	and Industrial
Observation Segments	Output Variables - Loss	0.9641				0.0000			
	Commercial Real Estate		0.9525				0.0000		
	Consumer Credit			0.8774				0.0000	
	Commercial and Industrial				0.8357				0.0000
Input Variables - Macroeconomic Factors	Real Gross Domestic Investment	-0.0085	-0.0081	-0.0299	-0.0003	4.21E-02	4.57E-02	1.17E-02	4.84E-02
	Unemployment Rate	0.0135	0.0034	0.0318	0.0094	1.38E-03	2.21E-02	4.92E-04	2.82E-02
	Commercial Real Estate Price Index	-0.0003	-0.0010	-0.0028	-0.0035	7.78E-08	9.72E-08	0.00E+00	1.11E-16
	BBB 10 Year Corporate Bond Rate	0.0433	0.0130	0.1385	0.0468	0.00E+00	1.53E-09	0.00E+00	0.00E+00
	CBOE Equity Market Volatility Index	0.0012	0.0035	0.0071	0.0068	4.35E-02	4.22E-05	3.34E-02	5.09E-12
	Log-Likelihood	-11.00	-11.91	-34.81	-5.33				
	Log-Likelihood Ratio Statistic - VARMAX								
	Null	2.72	2.56	0.42	4.17	2.18E-11	9.55E-12	0.00E+00	6.69E-09
	Root Mean Squared Error / Mean	0.3293	0.3923	0.2161	0.3100				
	Squared Correlation	0.8399	0.8470	0.7858	0.8037				
Model Performance Metric	Cumulative Percentage Error - Entire Sample	8.17%	7.60%	-11.55%	-11.30%				
	Cumulative Percentage Error - Recent Period	9.25%	10.29%	-22.52%	-6.86%				
	Cumulative Percentage Error - Downturn Period	-10.33%	-8.46%	-27.76%	-8.34%				

Table 8. Vector Autoregressive Model Estimation and Gaussian Vector Autoregressive vs. Regime Switching Vector Autoregressive Scenario Generation Compared (Fed Macroeconomic Variables and Aggregate Y9 Bank Charge-offs).

Scenario	Statistic	Modeling Segment				Scenario	Modeling Segment			
		Residential Real Estate	Commercial Real Estate	Consumer Credit	Commercial and Industrial		Residential Real Estate	Commercial Real Estate	Consumer Credit	Commercial and Industrial
Fed Base	Mean	0.2587	0.0930	1.8571	0.3512	Fed Severely Adverse	0.2743	0.2173	2.1794	0.5909
	Standard Deviation	0.0384	0.0455	0.0442	0.0286		0.0406	0.1185	0.4207	0.1650
	Minimum	0.2046	0.0170	1.8067	0.3046		0.2252	-0.0217	1.2789	0.2893
	25th Percentile	0.2272	0.0595	1.8244	0.3297		0.2341	0.1685	1.8839	0.4739
	Median	0.2595	0.0967	1.8448	0.3563		0.2723	0.2281	2.2580	0.5967
	75th Percentile	0.2889	0.1291	1.8785	0.3744		0.2945	0.3253	2.5021	0.7324
	Maximum	0.3174	0.1573	1.9484	0.3888		0.3511	0.3445	2.6334	0.7953
	Cumulative Loss	3.3627	1.2089	24.1426	4.5661		3.5657	2.8243	28.3327	7.6811
	Cumulative Loss Relative to Recent Period	1.2933	2.4178	1.4286	2.2830		1.3714	5.6486	1.6765	3.8406
	Cumulative Loss Relative to Downturn Period	0.1308	0.0460	0.3755	0.2195		0.1387	0.1074	0.4406	0.3693
VAR Model Base	Mean	0.2923	0.1208	1.9015	0.3877	VAR Model Severely Adverse	0.2594	0.2218	2.9533	1.3755
	Standard Deviation	0.0574	0.0599	0.0641	0.0457		0.1006	0.1072	0.4598	0.4527
	Minimum	0.2019	0.0209	1.8115	0.3020		0.0844	0.0382	1.6095	0.4136
	25th Percentile	0.2487	0.0767	1.8499	0.3580		0.2161	0.1709	2.9414	1.1734
	Median	0.2952	0.1248	1.8943	0.3979		0.2446	0.1901	3.0590	1.4404
	75th Percentile	0.3379	0.1686	1.9440	0.4257		0.2938	0.2934	3.2333	1.6629
	Maximum	0.3754	0.2052	2.0077	0.4392		0.5067	0.4302	3.4646	2.0801
	Cumulative Loss	3.8005	1.5700	24.7199	5.0399		3.3722	2.8828	38.3931	17.8817
	Cumulative Loss Relative to Recent Period	1.4617	3.1400	1.4627	2.5200		1.2970	5.7655	2.2718	8.9408
	Cumulative Loss Relative to Downturn Period	0.1479	0.0597	0.3844	0.2423		0.1312	0.1096	0.5971	0.8597
Regime Switching Model Base	Mean	0.3360	0.1821	2.0186	0.4142	Regime Switching Model Severely Adverse	0.9249	1.2696	4.6953	1.6844
	Standard Deviation	0.0822	0.0811	0.1973	0.0559		0.5048	0.7820	2.4530	1.0219
	Minimum	0.2071	0.0382	1.8380	0.3102		0.2363	-0.0547	1.4164	0.1437
	25th Percentile	0.2742	0.1246	1.8820	0.3785		0.5456	0.6673	2.3354	0.6546
	Median	0.3417	0.1917	1.9375	0.4314		0.6909	1.4276	5.0869	2.2355
	75th Percentile	0.4019	0.2491	2.1084	0.4600		1.3450	1.9156	7.2006	2.5118
	Maximum	0.4489	0.2861	2.4608	0.4718		1.7271	2.2146	7.3885	2.6254
	Cumulative Loss	4.3675	2.3669	26.2423	5.3849		12.0235	16.5051	61.0384	21.8977
	Cumulative Loss Relative to Recent Period	1.6798	4.7338	1.5528	2.6924		4.6244	33.0102	3.6117	10.9488
	Cumulative Loss Relative to Downturn Period	0.1699	0.0900	0.4081	0.2589		0.4678	0.6276	0.9493	1.0528

Table 9. Single Equation Autoregressive Model Estimation and Gaussian Vector Autoregressive vs. Regime Switching Vector Autoregressive Scenario Generation Compared (Fed Macroeconomic Variables and Aggregate Y9 Bank Charge-offs).

Scenario	Statistic	Modeling Segment				Scenario	Modeling Segment			
		Residential	Commercial	Consumer	Commercial		Residential	Commercial	Consumer	Commercial
		Real Estate	Real Estate	Credit	and Industrial		Real Estate	Real Estate	Credit	and Industrial
Fed Base	Mean	0.2062	0.0588	0.8989	0.2422	Fed Severely Adverse	0.1884	0.1406	1.9620	0.4649
	Standard Deviation	0.0704	0.0310	0.3462	0.0758		0.0573	0.0515	0.1279	0.0943
	Minimum	0.0907	0.0092	0.5308	0.1250		0.0633	0.0260	1.7759	0.2962
	25th Percentile	0.1663	0.0351	0.6120	0.1933		0.1709	0.1305	1.8680	0.4179
	Median	0.2009	0.0604	0.7984	0.2511		0.2110	0.1562	1.9857	0.4927
	75th Percentile	0.2529	0.0835	1.1134	0.3111		0.2311	0.1743	2.0854	0.5514
	Maximum	0.3275	0.1036	1.5795	0.3512		0.2417	0.1979	2.1230	0.5674
	Cumulative Loss	2.6805	0.7639	11.6863	3.1491		2.4491	1.8279	25.5062	6.0443
	Cumulative Loss Relative to Recent Period	0.4622	1.5277	0.6183	1.5746		0.3401	0.9140	1.5092	3.0221
Cumulative Loss Relative to Downturn Period	0.1136	0.0290	0.1817	0.1514	0.0878	0.0695	0.2997	0.2906		
VAR Model Base	Mean	0.2105	0.0076	0.7330	0.2422	VAR Model Severely Adverse	0.2396	0.2323	2.6082	1.1489
	Standard Deviation	0.1315	0.0037	0.5570	0.0758		0.1727	0.0837	0.3808	0.2650
	Minimum	0.0095	0.0013	0.1246	0.1250		0.0220	0.0433	1.7796	0.5654
	25th Percentile	0.1083	0.0049	0.2747	0.1933		0.1111	0.2176	2.4853	1.0446
	Median	0.2177	0.0079	0.6006	0.2511		0.2158	0.2603	2.7857	1.2318
	75th Percentile	0.3172	0.0106	1.0737	0.3111		0.3643	0.2904	2.8854	1.3786
	Maximum	0.3983	0.0124	1.8310	0.3512		0.5431	0.3030	2.9230	1.4185
	Cumulative Loss	2.7360	0.0983	9.5293	3.1491		3.1149	3.0197	33.9062	14.9357
	Cumulative Loss Relative to Recent Period	0.4717	0.1966	0.5042	1.5746		0.5371	6.0393	2.0063	7.4678
Cumulative Loss Relative to Downturn Period	0.1159	0.0037	0.1482	0.1514	0.1303	0.1148	0.3984	0.7181		
Regime Switching Model Base	Mean	0.3126	0.1380	1.2556	0.2422	Regime Switching Model Severely Adverse	0.8319	1.1002	3.7624	1.1576
	Standard Deviation	0.1912	0.0598	0.2199	0.0758		0.3556	0.6910	1.4307	0.6079
	Minimum	0.0259	0.0289	1.0623	0.1250		0.0861	0.0322	1.8287	0.2345
	25th Percentile	0.1641	0.0963	1.0996	0.1933		0.6731	0.4644	2.2674	0.5267
	Median	0.3258	0.1489	1.1895	0.2511		0.7692	1.4092	4.1468	1.5085
	75th Percentile	0.4684	0.1906	1.3244	0.3111		1.0180	1.6879	5.1048	1.6317
	Maximum	0.5754	0.2063	1.7838	0.3512		1.4150	1.7707	5.3215	1.6693
	Cumulative Loss	4.0640	1.7942	16.3224	3.1491		10.8147	14.3029	48.9108	15.0490
	Cumulative Loss Relative to Recent Period	0.6395	3.5885	0.8636	1.5746		1.8646	28.6058	2.8941	7.5245
Cumulative Loss Relative to Downturn Period	0.1034	0.0682	0.2538	0.1514	0.4583	0.5438	0.5747	0.7235		

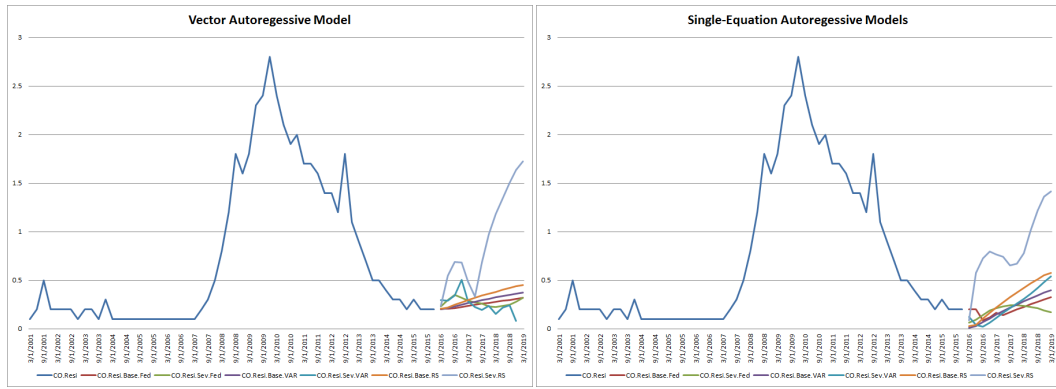


Figure 15. Vector Autoregressive vs. Single Equation Autoregressive Model Estimation and Gaussian Vector Autoregressive vs. Regime Switching Vector Autoregressive Scenario Generation Compared—Residential Real Estate.

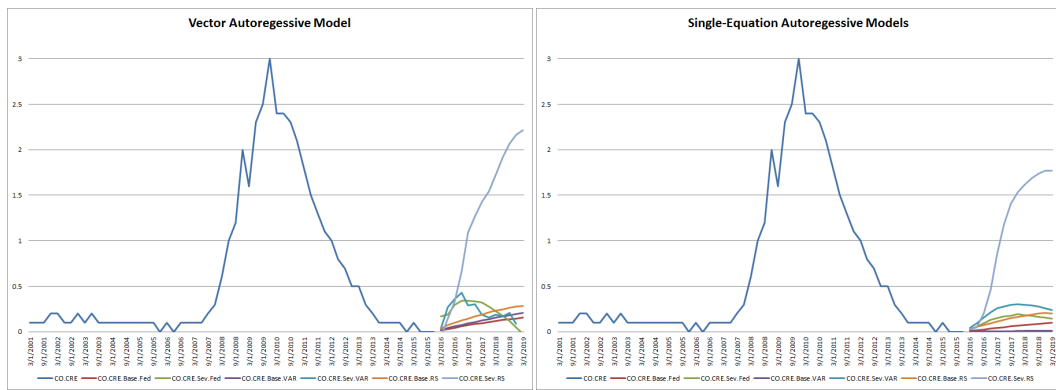


Figure 16. Vector Autoregressive vs. Single Equation Autoregressive Model Estimation and Gaussian Vector Autoregressive vs. Regime Switching Vector Autoregressive Scenario Generation Compared—Commercial Real Estate.

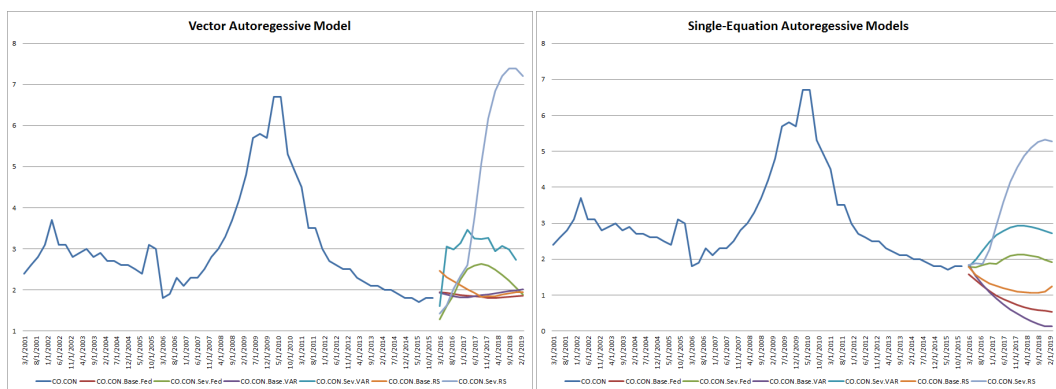


Figure 17. Vector Autoregressive vs. Single Equation Autoregressive Model Estimation and Gaussian Vector Autoregressive vs. Regime Switching Vector Autoregressive Scenario Generation Compared—Consumer Loans.

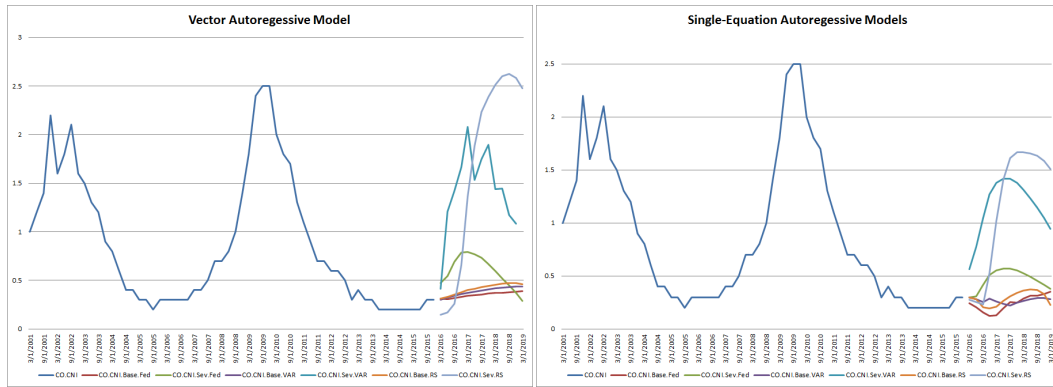


Figure 18. Vector Autoregressive vs. Single Equation Autoregressive Model Estimation and Gaussian Vector Autoregressive vs. Regime Switching Vector Autoregressive Scenario Generation Compared—Commercial and Industrial Loans.

Table 10. Diebold-Mariano Accuracy Tests—Gaussian Vector Autoregressive vs. Regime Switching Vector Autoregressive Scenario Generation Compared (Fed Macroeconomic Variables and Aggregate Y9 Bank Charge-offs).

Forecast Horizon	1-Step Ahead		8 Quarter Out-of-Sample/Time	
Modeling Segment	Test Statistic	P-Value	Test Statistic	P-Value
Residential Real Estate	2.5675829	0.0106	0.7492074	0.2838
Commercial Real Estate	2.8342814	0.0096	0.5472589	0.2522
Consumer Credit	3.8153666	0.0028	0.3536867	0.1341
Commercial and Industrial	2.102496	0.0071	1.9498362	0.1565

As an example, in the case of the C&I segment in the VAR estimation and as measured by the cumulative loss relative to that in the downturn period in VAR estimation, in the C&I segment this multiple is 1.05 in the MS-VAR model but only 0.85 (0.36) in the VAR (Fed) scenario generation models. However, in the corresponding multiple is 0.75 in the MS-VAR model but only 0.71 (0.29) in the VAR (Fed) scenario generation models.

In Table 10 we implement the Mariano-Diebold (Diebold and Mariano, 2002) accuracy tests of the hypotheses that the multiple equation VAR model outperforms the RS-VAR model. This is an important exercise, as the literature notes that often regime switching time series models are prone to the problem of over-fitting (Dacco and Satchell, 1999; Engel, 1994). We are able to reject the null hypothesis that the MS-VAR model is outperformed by the VAR model, both on a 1-step ahead and on an out-of-sample basis⁹. In the latter, we recalibrate the model leaving out the last 8 quarters of data, and predict credit losses over this period.

⁹ The holdout sample covers the last 2 years, the 1st quarter of 2016 through the 4th quarter of 2017.

5. Conclusion and future directions

This paper has considered analyzing the estimation methodologies and the macroeconomic scenarios provided, key ingredient of the stress testing process, such as the Federal Reserve's CCAR program. We have analyzed the estimation methodology implications around the supervisory requirements that banks develop their own macroeconomic scenarios. A standard approach such as the VAR statistical model, that exploits the dependency structure between both macroeconomic drivers and modeling segments, has been examined in the context of the well-known phenomenon of fat-tailed distributions that deviate from a Gaussian error structure. We have investigated the implications the MS-VAR challenger model, commonly seen in academics yet not prevalent in practice. These competing models have been empirically tested with Federal Reserve macroeconomic data released for CCAR purposes and Y9 regulatory filings. Our main finding is that the MS-VAR model produces more conservative Severe loss projections as compared to the VAR model, as well as greater forecast greater accuracy according to the KPSS testing, which we explain in the ability of the regime switching paradigm to better accommodate extreme events observed in history that deviate from normality. The MS-VAR model is capable in the Severe scenario of at least matching and sometimes well exceeding historical extremes in the direction of augmented losses, as compared to the VAR model. The VAR model bears similarities to the Fed model in terms of the magnitude of scenarios, but we observe that the trajectories of either the VAR or MS-VAR models tend to be more regular. The Fed model in the Base scenarios is rather close to that of the VAR model, but the MS-VAR model projects an augmented base in all cases, which attribute to the skewness error distribution in the regime-switching or mixture setting.

As a second main conclusion, we have considered the case of banks that model the risk of their portfolios using top-of-the-house modeling techniques, and have addressed an issue of how to incorporate the correlation of risks amongst the different segments. An approach to incorporate this consideration of a dependency structure was proposed, and the bias that results from ignoring this aspect is quantified, through estimating a VAR time series models for credit loss using Fed Y9 data. We found that the multiple equation VAR model outperforms the single equation AR models according to various metrics across all modeling segments. The results of the estimation are broadly consistent across the VAR and AR models, but with a few notable differences (e.g., most segments exhibit significant but mild autocorrelation, and different subsets of the macroeconomic variables are significant across different segments). Across all 4 segments, according to the likelihood ratio statistic, we reject the hypothesis that the restrictions of the single equation AR models are justified. Furthermore, while the VAR models are generally more accurate according to standard measures of model fit with respect to each segment, it is inconclusive whether the VAR or AR models are more or less conservative as measured by cumulative 9-quarter losses.

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

- More granular classes of credit risk models, such as ratings migration or PD/LGD scorecard/regression
- Alternative data-sets, for example bank or loan level data
- More general classes of regression model, such as logistic, semi-parametric or machine learning / artificial intelligence techniques (Jacobs, 2018)
- Applications related to stress testing, such as regulatory or economic capital

Conflict of interest

There is no conflict of interest with respect to either author. Regarding Michael Jacobs, Jr. the views and conclusions expressed in this paper do not reflect a position or opinion held by Accenture LLP or any affiliated firms. The research for this article was co-authored by Frank J. Sensenbrenner, in collaboration with Michael Jacobs, Jr., in Mr. Sensenbrenner's personal capacity and not in his official capacity as an employee of the Commodity Futures Trading Commission (CFTC). The analyses and conclusions expressed in this research paper are those of Mr. Sensenbrenner and Mr. Jacobs, Jr., and do not reflect the views of other employees of the CFTC Division of Clearing and Risk, other CFTC staff, the Commission itself, or the United States Government.

References

- Acharya VV, Schnabl P, Richardson M (2010) How Banks Played the Leverage Game. *John Wiley* 18: 144–145.
- Bangia A, Diebold FX, Kronimus A, et al. (2002) Ratings migration and the business cycle, with application to credit portfolio stress testing. *J Bank Financ* 26: 445–474.
- Bank for International Settlements, Basel Committee on Banking Supervision (BCBS) (1988) Internal convergence of capital measurement and capital standards: A revised framework.
- Basel Committee on Banking Supervision (BCBS) (1996) Amendment to the capital accord to incorporate market risks.
- Basel Committee on Banking Supervision (BCBS) (2006) International convergence of capital measurement and capital standards: A revised framework.
- Basel Committee on Banking Supervision (BCBS) (2009a) Principles for sound stress testing practices and supervision.
- Basel Committee on Banking Supervision (BCBS) (2009b) Guidelines for computing capital for incremental risk in the trading book.
- Basel Committee on Banking Supervision (BCBS) (2009c) Revisions to the Basel II market risk framework.
- Basel Committee on Banking Supervision (BCBS) (2009d) Analysis of the trading book quantitative impact study (October).
- Basel Committee on Banking Supervision (BCBS) (2010a) Strengthening the resilience of the banking sector—consultative document (December).
- Basel Committee on Banking Supervision (BCBS) (2010b) An assessment of the long-term economic impact of stronger capital and liquidity requirements (August).
- Basel Committee on Banking Supervision (BCBS) (2010c) Basel III: A global regulatory framework for more resilient banks and banking systems (December).
- Baum LE, Petrie T (1966) Statistical inference for probabilistic functions of finite state Markov chains. *Ann Math Stat* 37: 1554–1563.
- Baum LE, Petrie T, Soules G, et al. (1970) A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann Math Stat* 41: 164–171.
- Berkowitz J (1999) A coherent framework for stress-testing. *FEDS Work Pap*.
- Bidder R, McKenna A (2015) Robust stress testing. *Work Pap*.
- Blackwell D, Koopmans L (1959) On the identifiable problem for functions of finite Markov chains. *Ann Math Stat* 28: 1011–1015.

- Board of Governors of the Federal Reserve System (1996) Joint Policy Statement on Interest Rate Risk, Supervisory Letter 96-13, May 23rd.
- Board of Governors of the Federal Reserve System (1999) Supervisory Guidance Regarding Counterparty Credit Risk Management, Supervisory Letter 99-03, February 1st.
- Board of Governors of the Federal Reserve System (2002) Interagency Guidance on Country Risk Management, Supervisory Letter 02-05, March 8th.
- Board of Governors of the Federal Reserve System (2011) Supervisory Guidance on Model Risk Management, Supervisory Letter 11-7, April 4th.
- Box G, Jenkins G (1976) *Time Series Analysis: Forecasting and Control*, New York, Wiley, 2nd Edition.
- Brockwell PJ, Davis RA (1991) Time series: theory and methods. *Technometrics* 31: 121.
- Commandeur JJF, Koopman SJ (2007) Introduction to state space time Series analysis. *RePEc* 36: 3–25.
- Dacco R, Satchell S (1999) Why do regime-switching models forecast so badly? *J Forecasting* 18: 1–16.
- Demirguc-Kunt A, Detragiache E, Merrouche O (2013) Bank capital: lessons from the financial crisis. *J Money Credit Bank* 45: 1147–1164.
- Dickey DA, Fuller WA (1981) Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* 49: 1057–1072.
- Diebold FX, Mariano RS (2002) Comparing predictive accuracy. *J Bus Econ Stat* 20: 134–144.
- Embrechts P, Straumann EMD (1999) Correlation: pitfalls and alternatives. *Risk* 12: 69–71.
- Embrechts P, McNeil AJ, Straumann D (2002) Correlation and dependence in risk management: properties and pitfalls, In: Dempster, M.A.H., ed., *Risk Management: Value at Risk and Beyond*, Cambridge University Press, Cambridge, UK, 176–223.
- Embrechts P, Lindskog F, McNeil AJ (2003) Modeling dependence with copulas and applications to risk management, In: Rachev, S., ed., *Handbook of Heavy Tailed Distributions in Finance*, Elsevier, Rotterdam, 329–384.
- Engel C (1994) Can the Markov switching model forecast exchange rates? *J Int Econ* 36: 151–165.
- Frame RS, Fuster A, Tracy J, et al. (2015) The rescue of Fannie Mae and Freddie Mac. *J Econ Perspect* 29: 25–52.
- Foglia A (2009) Stress testing credit risk: a survey of authorities' approaches. *Soc Sci Electron Publ* 5: 9–45.
- Frey R, McNeil AJ (2001) Modeling dependent defaults, *Work Pap*.
- Goldfeld SM, Quandt RE (1973) A Markov model for switching regressions. *Journal Econ* 1: 3–15.
- Hamilton JD (1988) Rational expectations econometric analysis of changes in regime: an investigation of the term structure of interest rates. *J Econ Dyn Control* 12: 385–423.
- Hamilton JD (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57: 357–384.
- Hanan EJ (1971) The identification problem for multiple equation systems with moving average errors. *Econometrica* 39: 751–766.
- Hanan EJ, Deistler M (1988) The statistical theory of linear systems. *J Wiley* 32: 292–293.
- Heller A (1965) On stochastic processes derived from Markov chains. *Ann Math Stats* 36: 1286–1291.
- Inanoglu H, Jacobs, Jr. M (2009) Models for risk aggregation and sensitivity analysis: an application to bank economic capital. *J Risk Financ Manage* 2: 118–189.
- J.P. Morgan (1994) *RiskMetricsTM*, Second Edition, New York.
- J.P. Morgan (1997) *CreditMetricsTM*, First Edition, New York
- Jacobs, Jr. M (2010) Validation of economic capital models: state of the Practice, supervisory expectations and results from a bank study. *J Risk Manage Finance Inst* 3: 334–365.

- Jacobs, Jr. M (2013) Stress Testing Credit Risk Portfolios. *J Financ Transform* 37: 53–75.
- Jacobs, Jr. M, Karagozoglu A, Sensenbrenner F (2015) Stress testing and model validation: application of the Bayesian approach to a credit risk portfolio. *J Risk Model Validation* 9: 1–70.
- Jacobs, Jr. M (2015) The impact of asset price bubbles on credit risk measures. *J Financ Risk Manage* 4: 251–266.
- Jacobs, Jr. M (2016) Stress testing and a comparison of alternative methodologies for scenario generation. *J Appl Finan Bank* 6: 123–156.
- Jacobs, Jr. M (2017a) A mixture of distributions model for the term structure of interest rates with an application to risk management. *Am Res J Bus Manage* 3: 1–17.
- Jacobs, Jr. M (2017b) Asset price bubbles and the quantification of credit risk capital with sensitivity analysis, empirical implementation and an application to stress testing. *J Risk Model Validation* 11: 1–35.
- Jacobs, Jr. M (2018) The validation of machine learning models for the stress testing of credit risk. *Int J Econ Manage Sci* 7: 1–21.
- Jorion P (1996) Risk2: measuring the Value at Risk. *Financ Analyst J* 52: 47–56.
- Jorion P (1997) Value at Risk: The New Benchmark for Controlling Market Risk, volume 2.
- Jorion P (2006) Value at Risk: The Benchmark for Managing Financial Risk, Third Ed.
- Kwiatkowski D, Phillips PCB, Schmidt P, et al. (1992) Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? *J Econometrics* 54: 159–178.
- Kohn R (1979) Asymptotic results for ARMAX structures. *Econometrica* 47: 1295–1304.
- Koyluoglu H, Hickman A (1998) Reconcilable differences. *Risk* 56–62.
- Kupiec PH (2009) Risk capital and VaR. *J Deriv* 7: 41–52.
- Li DX (1999) On default correlation: A copula function approach. *J Fix Inc* 9: 43–54.
- Loregian A, Meucci A (2013) Neither “Normal” nor “Lognormal”: modeling interest rates across all regimes. *Financ Analyst J* 72: 68–82.
- Merton R (1974) On the pricing of corporate debt: the risk structure of interest rates. *J Financ* 29: 449–470.
- Mosser PC, Fender I, Gibson MS (2001) An international survey of stress tests. *Curr Issues Econ Finan* 7: 1–6.
- Pearson K (1895) Contribution to the mathematical theory of evolution. *Philos Trans Royal Soc* 185: 71–110.
- Poon SH, Rockinger M, Tawn J (2004) Extreme value dependence in financial markets: diagnostics, models and financial implications. *Rev Financ Stud* 17: 581–610.
- R Development Core Team (2017) R: A Language and Environment for Statistical Computing.
- Rebonato R (2010) Coherent Stress Testing: A Bayesian Approach. Wiley, New York.
- Schuermann T (2014) Stress testing banks. *Int J Forecasting* 30: 717–728.
- Sims CA (1980) Macroeconomics and reality. *Econometrica* 48: 1–48.
- Stock JH, Watson MW (2001) Vector autoregressions. *J Econ Perspect* 15: 101–115.
- Tjøstheim D (1986) Estimation in nonlinear time series models. *Stoch Process their Appl* 21: 251–273.
- Wilde T (1997) CreditRisk+ A Credit Risk Management Framework. *Credit Suisse First Boston*.

