An Integrated Multimodal Attention-Based Approach for Bank Stress Test Prediction

1st Farid Razzak
Rutgers University
farid.razzak@rutgers.edu

2nd Fei Yi
Northwestern Polytechnical University
yifeinwpu@gmail.com

3rd Yang Yang
Nanjing University
yangy@lamda.nju.edu.cn

4th Hui Xiong
Rutgers University
hui.xiong@business.rutgers.edu

***Abstract*—Since the financial crisis in late 2008-2009, several global regulatory authorities have mandated stress-testing exercises to evaluate the potential capital shortfalls & systemic impacts that large banks may face during adverse economic conditions. Thus, having the ability to analyze economic conditions & banking performance profiles together to determine relationships among their respective features may provide insights for stress-testing tasks.**

In this paper, we propose an Integrated Multimodal Bank Stress Test Prediction (IMBSTP) model framework consisting of a two-stages; (1) economic conditions estimator to approximate joint representation among the exogenous factors using generative models, (2) bank capital & loss forecaster to project stress-test measures based on dimensional & temporal features selected from the exogenous economic conditions & banking performance profiles using a dual-attention recurrent neural network.

Extensive experimentation is performed on historical economic conditions & consolidated financial statements of U.S. bank holdings companies to show the effectiveness of our approach when compared to state-of-the-art baseline methods.

***Keywords*-Deep Learning, Multimodal Conditional Generative Models, Recurrent Neural Networks, Bank Stress-Test**

I. Introduction

The recent financial crisis & it's ensuing economic recession in the United States have brought major advances to capital adequacy requirements, such as testing the solvency of banks through hypothetically adverse economic scenarios by examining expected performance in said dire conditions. Effectively, regulators want to ensure large banks are providing loans with considerations for economic & systemic risks that could impact bank solvency. However, this regulatory strategy is counter to the profit generation goals of the bank, which can create a unique circumstance of conflicts that warrant regulatory oversight. Effective stress-testing can provide insights & potential mitigation strategies to prevent catastrophic losses by financial institutions during severe economic conditions [1].

Currently, the Federal Reserve's Comprehensive Capital Analysis & Review (CCAR) is the most recognized regulatory exercise in the U.S [2]. An adverse outcome from the exercise may serve as an industry-wide signal for counter-party & systemic risk as well as have repercussions for the bank that may include restrictions on a firm's capital distribution plan [3] , rectification of compliance risk program, & renewal of loss projections meeting regulatory thresholds [2].

In this paper, we propose the *Integrated Multimodal Bank Stress Test Prediction* (IMBSTP) model framework, designed

TABLE I: Data Summary for X_i, Y_i (1990-2017)

Type	Variable	Mean	Std.	Med.	Max
$X_{loancat}$ (In Mn USD)	Comm. & indtrl	5.2	13.0	1.2	205.0
	Constr. & ld dev.	10.5	2.1	.43	275.5
	Consr.(excl.CredC)	13.1	114.6	.22	1710.4
	CredCrd	3.9	13.4	.007	162.4
	HELOCs	1.7	6.6	.24	121.7
	Multifam-RealEstate	.57	2.0	.14	6.9
	Non-farm-nonres-CRE	5.1	2.4	1.3	798.6
	Res-RealEstate(excl.HELOCs)	2.2	16.1	1.4	372.4
Y_{ncoR} (0-100%)	Comm. & indtrl	1.1	1.9	.44	11.0
	Constr. & ld dev.	2.2	4.7	.34	25.9
	Multifam-RealEstate	.93	2.0	.11	12.2
	Consr.(excl.CredC)	3.6	5.8	1.2	35.2
	CredC	8.3	1.0	4.1	54.8
	HELOCs	1.7	1.7	.15	12.0
	Non-farm-nonres-CRE	1.9	4.3	.15	23.1
	Res-RealEstate(excl.HELOCs)	.80	1.9	.13	12.5
	Comb-LoanLoss	2.4	2.1	1.7	14.7
Y_{ppnr} (0-100%)	Net-interest income	4.6	6.6	2.3	35.9
	Non-interest income	2.8	5.1	0.9	30.7
	Trading income	0.2	0.7	.001	4.9
	Compensation expense	2.8	4.7	1.16	26.5
	Fixed assets expense	.66	1.0	.29	6.3
	Non-interest expense	5.3	9.6	1.61	53.7
$Y_{CapRatio}$ (%)	T1 Common Equity	16.9	24.0	13.0	66.0
	T1 Risk	17.0	27.4	14.0	70
	Ttl-Risk	18.2	28.1	16.0	72.0
	T1 Leverage	5.8	20.9	9.0	37.0

to project bank capital & loan loss ratios in estimated economic conditions. Particularly, we incorporate additional exogenous factors beyond regulator standards to depict economic conditions & learn non-linear latent relationships among relevant economic & financial market indicators that can be used to understand banking performance. The contributions of this paper can be summarized as follows:

- An integrated model framework that robustly consolidates multimodal economic conditions estimation & banking capital & loss prediction based on dimensional & temporal importance for stress-test related tasks.
- Additions to the state of the art literature by exploring the use of deep learning techniques to understand bank stress-test properties [1], [3], [4] using generative & non linear auto-regressive exogenous neural network modeling.

II. Problem Statement

A. Preliminaries

Definition 1 (Economic Conditions) In a top-down approach to stress-testing, exogenous economic factors, in our case $ECO_{mod} = [Z_{macro}, Z_{micro}, M_{SP}, M_{FCI}]$, are believed to have influence on the trajectory of a bank's performance.

Currently, the historical macro-economic variables that depict the U.S. & parts of the global financial economy consists of domestic & international variables, Z_{macro_I} [2], [4].

Historical micro-economic variables represent specific aspects of the financial economy. To this end, US treasuries, inflation rates, major commodity indices, stock indices returns, government bond rates, interest rate swaps, currency swap

2374-8486/19/$31.00 ©2019 IEEE
DOI 10.1109/ICDM.2019.00161

rates, & major commodities prices are collected quarterly between 1976-2017 to depict the micro-economy, Z_{micro}.

Historical sector-based indices, $M_{SP_{FIN}}$, are collected over the same time period & frequency as previous variables to depict the financial & real-estate sectors, since initial signals of the most recent financial crisis were first noticed at mentioned sectors. The S&P 500 Financial Sector Index consists of three tickers that illuminate the conditions specific to the financial industry, while the S&P Real Estate Indices,$M_{SP_{RE}}$, consists of tickers that tracks the real estate industry.

Several industry indices that depict the overall financial conditions of U.S banking & financial economy are also collected. These indices & sub-indices, M_{FCI}, are developed by the regional Federal Reserve Boards as well as corporate research entities to capture the directional conditions in money markets, debt markets, equity markets & traditional shadow banking systems [5].

Definition 2: (Banking Performance Profile) Top-down stress-testing relies on publicly released bank financial statements to assess bank loan portfolios, loan loss rates & net revenues, to determine capital adequacy.

For the purposes of our study, the regulatory codes & calculations discussed in [3] are used to acquire corresponding features.

Loan portfolios breakdown seven major loan categories, $X_{loancat_{i,j}}$, of bank holding companies acquired from their respective consolidated financial statements through "Bank Regulatory" dataset [6], a summary of which can be seen in Table I. The loan categories represent a snapshot view of the bank's lending practice to different segments of borrowers whom could be impacted by economic conditions & therefore affect the bank's risk exposure.

Determining insight from $X_{loancat_{i,j}}$, where i is an individual bank & j is a loan category, by examining the temporal evolution from $t-1$ to t can provide details into the bank's growth & loss rates in conjunction with economic conditions, $ECO_{mod,t}$.

The net losses in loan categories are depicted by the net-charge-off amounts, $X_{NCO_{i,j}}$ which considers both losses & recoveries from each respective loan category to then determine the loan loss rates, $Y_{ncoR_{i,j,t}} = 100 * \frac{X_{NCO_{i,j}}}{X_{loancat_{i,j_{t-1}}}}$.

Banks are able to generate revenue from interest earning loans, trading income & other revenue-generating services, however they also have to consider expenses such as compensation, fixed assets, & other non interest earning operations, $X_{Cmpppnr_{i,j}}$. To determine the net revenue proportional to the bank's consolidated assets, $X_{CnsldAsts_{i,t-1}}$, we can derive the pre-provisional net revenue ratio of the bank, $Y_{ppnr_{i,j_t}}$ [3]. Understanding the bank's ability to generate revenue during dynamic economic conditions, $Y_{ppnr_{i,j_t}} = 100 * \frac{X_{Cmpppnr_{i,j}}}{X_{CnsldAsts_{i,t-1}}}$, can be crucial to offset losses for more appropriate forecasts.

The bank's monetary reserves, retained net earnings, or equity capital, X_{EqCap}, is considered to be the funds it has available after expenses & losses are deducted, however these funds may be further reduced due to capital distributions

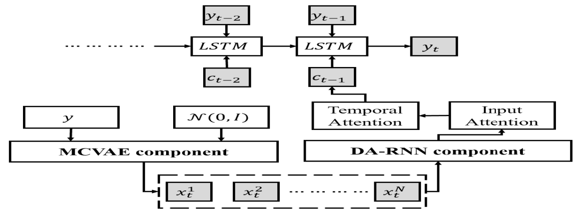

Fig. 1: The overall framework of proposed IMBSTP model.

(X_{EqPO}), payment of taxes ($\tau = 35\%$), & regulatory capital deductions (X_{RegDt}). The calculations can be summarized as $X_{EqCap_t} = X_{EqCap_{t-1}} + (1-\tau) * Net_{revloss_t} - X_{EqPO_{i,t-1}}$, where $Net_{revloss_t} = (\sum_j Y_{ppnr_{i_t}} - \sum_j Y_{ncoR_{i_t}})$.

The remaining capital in proportion to the bank's previous risk weighted assets, $X_{RWA_{t-1}}$), is considered to be the capital ratio (Y_{T1CR}), a measure that best depicts the bank's overall capital adequacy. Regulators are most interested in the Tier-1 common ratio, which only considers bank equity elements as part of the capital, $Y_{T1CR} = \frac{X_{EqCap} - X_{RegDt_{t-1}}}{X_{RWA_{t-1}}}$.

B. Problem Formulation

The prediction tasks can be separated into two stages:

1) Economic Conditions Estimation: Given a set of economic conditions, ECO_{mod}, the task is to estimate future conditions by learning the true joint probability distribution among all the variables within each modality, $\Pr(E\hat{C}O_{mod})$. This representation of the overall economic conditions, $\Pr_\theta(E\hat{C}O_{mod})$, allows for both the prediction of the most likely upcoming economic conditions & sampling of economic conditions from different probability densities to acquire more dire but plausible conditions.

In our approach, normalization & principal components dimension reduction, $E\hat{C}O_{mod} = [\hat{w}_1, \hat{w}_2, ...\hat{w}_j] \quad \forall w_j \in ECO_{mod}$, of each modality in ECO_{mod} is performed prior to estimation for the purposes of scaling numerical values & obtaining representative variables.

2) Bank Capital & Loss Ratio Prediction: Given the set of banking performance profiles, $[Y_{ncoR}, Y_{ppnr}, Y_{T1CR}]$, banking loan portfolios, $X_{loancat}$, economic conditions, $E\hat{C}O_{mod}$ & economic estimations from the previous stage, $\Pr_\theta(E\hat{C}O_{mod})$, the task of predicting future bank capital, loan loss & net revenue is a function of each target variable's past with consideration for economic conditions estimations, $E\hat{C}O_{mod_t}$.

Forecasting banking capital & performance is typically derived from projected loan loss & revenue [1], [3], as:

$$Pr(Y_{ncoR_t}, Y_{ppnr_t}, Y_{T1CR_t} | Y_{ncoR_{t-1}}, Y_{ppnr_{t-1}}, X_{loancat_{t-1}}, Y_{T1CR_{t-1}}, E\hat{C}O_{mod_t}) \tag{1}$$

However, for the purposes of this paper, Y_{T1CR_t} is directly projected using loan losses, net revenue, previous loan portfolio, & estimated economic conditions.

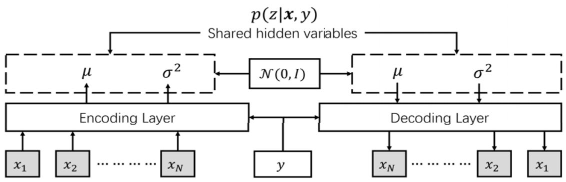

Fig. 2: Implemented MCVAE with encoder-decoder design.

Furthermore, our framework uses the estimated economic conditions, $\theta_{E\hat{C}O_{mod_t}}$, when conducting the capital, revenue & loss projections. The reasons for this are two-fold (1) all the exogenous modalities will need to be estimated together for the depiction of the future economy (2) We may not have access to economic conditions information for future time periods immediately & may need to use best estimates.

3) Challenges: To accomplish the tasks mentioned, several challenges will need to be considered in our framework.

First, our approach will need to find the non-linear latent relationships among $E\hat{C}O_{mod_t}$ to generate appropriate joint probability distributions without treating each economic modality independently. Secondly, we must address data sparsity, timeliness & limited modality observation co-occurrence, all of which are common with financial reporting data. Thirdly, we should consider the dynamic nature of dimensional influence on target variables over a temporal space since different banking & economic factors may influence banking performance at specific times.

III. Related Work

Our work relates to several research topics, including generation of economic scenarios for stress testing, prediction of bank performance in future economic scenarios & deep learning neural network techniques of machine learning literature. Works in generating foreseeable macro economic scenarios focus around utilizing structural, which models interactive relationships between countries based on economic variables with trade activity [7], [8] & statistical, which rely on linear models, methods [9], [10].

The work that aims to predict bank performance within economic scenarios are typically related to balance-sheet projections as seen in [3], where fixed effects regression within a top down approach is used for net charge offs & pre-provision net revenue predictions. Similarly in [1], several linear regression models with specific assumptions project capital gaps during adverse economic conditions.

[4] incorporates machine learning techniques to find plausible stress-tests for each bank using CGAN [11] for scenario generation & LSTM [12] for bank performance prediction.

IV. Model Design

In this section, we discuss the details of our IMBSTP model framework, which is a two-stage solution that first applies a multimodal conditioned variational autoencoder (MCVAE)

Procedure 1 MCVAE for Economic Condition Estimation

Input: target/cond modalities $\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$, y
Output: learned parameters (θ, ϕ), Estimations $\mathbf{x}$
Initialize parameters (θ, ϕ), $i \leftarrow 0$
while Convergence on (θ, ϕ) or $i = EPOCH$ **do**
  $h \leftarrow encoder.forward(\mathbf{x}, y)$ ▷ encode modalities
  $\mathbf{z} \leftarrow g_\phi(h, \epsilon^{(l)})$ ▷ learn shared variables
  $\hat{\mathbf{x}} \leftarrow decoder.forward(\mathbf{z}, y)$ ▷ modality estimation
  $Loss \leftarrow MSELoss(\mathbf{x}, \hat{\mathbf{x}})$ ▷ calculate loss
  $Loss.backward()$ ▷ propagate gradients
end while
return estimation $\hat{\mathbf{x}}$, parameters (θ, ϕ)

for economic conditions estimation, followed by an attention-based recurrent neural network for bank capital & loss prediction. The overall model framework is depicted in Figure 1.

A. MCVAE-based Modality Estimator

Existing works [13], [14] have proposed alternative approaches for dealing with modality estimation problems. However, they are not designed for the conditional multi-modality estimation problem, which seeks to explore the relationship between distributions of target modalities conditioned on one specific modality (referred to as conditional modality). In our work, we propose a Multi-modal Conditional Variational Autoencoder model (MCVAE) to address the problem & estimate the desired economic condition variables given one existing variable.

Specifically, the proposed MCVAE consists of several variational autoencoder components for each target modality & applies a conditional mechanism by importing a specified modality for better estimation of the target multi-aspect modality. According to the basic methodology of VAE [13] along with inspiration from MVAE [15] & CVAE [14], the goal of training MCVAE is to maximize the evidence lower bound (ELBO) to then generate the latent variable $\mathbf{z}$ given multi-modality set $\mathbf{x}$ & conditional modality y, which is defined via an inference network, $q_\phi(\mathbf{z}|\mathbf{x}, y)$, as:

$$\begin{aligned} ELBO(\mathbf{x}, y) \triangleq\ & \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x},y)}[\lambda \log p_\theta(\mathbf{x}|y, \mathbf{z})] \\ & -\beta KL[q_\phi(\mathbf{z}|\mathbf{x}, y), p_\theta(\mathbf{z}|y)], \end{aligned} \tag{2}$$

where $p_\theta(\mathbf{z}|y)$ is the conditional network & $p_\theta(\mathbf{x}|y, \mathbf{z})$ represents the generation network. Furthermore, by extending a similar derivation from [14] to our multi-modality problem, we ultimately obtain the following empirical lower bound:

$$\begin{aligned} \mathcal{L}(\mathbf{x}, y; \theta, \phi) = & \sum_{x_i \in \mathbf{x}} -KL(q_\phi(\mathbf{z}|x_i, y)) \parallel p_\theta(x_i|y, \mathbf{z}) \\ & + \frac{1}{L} \sum_{l=1}^{L} \sum_{x_i \in \mathbf{x}} \log p_\theta(x_i|y, \mathbf{z}^{(l)}) \end{aligned} \tag{3}$$

where x_i is the i^{th} modality in $\mathbf{x}$, & $\mathcal{L}$ is the number of samples that latent variable $\mathbf{z}^{(l)} = g_\phi(\mathbf{x}, y, \epsilon^{(l)}), \epsilon^{(l)} \sim \mathcal{N}(0, \mathbf{I})$. As shown in Figure 2, the proposed MCVAE model consists of two components, the encoding & decoding network. The model first passes multiple modalities & a separate conditional

Procedure 2 DA-RNN for bank capital & loss prediction

Input: modalities $\mathbf{x}_t$, historical records $y_{1:(t-1)}$
Output: prediction y'_t
Initialize Input Attention encoder-network layer
for $x^k_t \in \mathbf{x}_t$ **do**
  $h_t \leftarrow LSTM_{unit}.forward(x^k_t)$
  $\alpha^k_t \leftarrow Softmax.forward(h_{t-1}, x^k_t)$ ▷ weighting
  $\widehat{\mathbf{x}}_t = (\alpha^1_t x^1_t, \alpha^2_t x^2_t, \ldots, \alpha^N_t x^N_t)$
end for
Initialize Temporal Attention decoder-network layer
for $\widehat{x}_i \in \widehat{x_T}$ **do**
  $h'_i \leftarrow LSTM_{unit}.forward(\widehat{x}_i)$
  $\beta^i_t \leftarrow Softmax.forward(h'_i)$ ▷ weighting
  $c_t = \sum_{i=1}^{T} \beta^i_t h'_i$ ▷ context vector
  $d_t \leftarrow LSTM_{unit}.forward(y_{t-1}, c_{t-1})$
  $y'_t \leftarrow Linear.forward(d_t, c_t)$ ▷ prediction
end for
return prediction y'_t

TABLE II: Economic Conditions Experiments

Configuration	Training	Testing	Comment
Experiment 1	1976 Q1 - 2007 Q4	2008 Q1 - 2009 Q4	Financial Crisis
Experiment 2	1976 Q1 - 2015 Q4	2016 Q1 - 2017 Q4	All Econ Data
Experiment 3	1990 Q1 - 2007 Q4	2008 Q1 - 2009 Q4	Aligned Bank Data & Financial Crisis
Experiment 4	1990 Q1 - 2015 Q4	2016 Q1 - 2017 Q4	Aligned Bank & Econ Data

modality through an encoding network to learn hidden representations of shared variables across modalities. Afterwards, given the conditional modality & initial variables sampled from a Gaussian distribution, the decoding network can produce the estimation of target modalities. In our problem, the target input modalities $\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$ & conditional input y are different aspects of the economic conditions & the goal of MCVAE is to generate the overall economic conditions when given only one specific type of economic condition. The training procedure of the proposed MCVAE is demonstrated in Algorithm 1, which is optimized by popular stochastic optimization method Adam [16].

B. Attention-based Predictor

In our work, there are two types of features that can be used to predict future bank capital & loss; 1) Exogenous factors that mainly consists of economic conditions, 2) Historical bank performance profiles' respective time-series. These two factors are integrated in our work by applying a dual-stage attention-based neural network model (DA-RNN) [17], which takes both attention mechanism [18] & long short-term temporal dependencies (LSTM) [12] into consideration for effective time-series prediction.

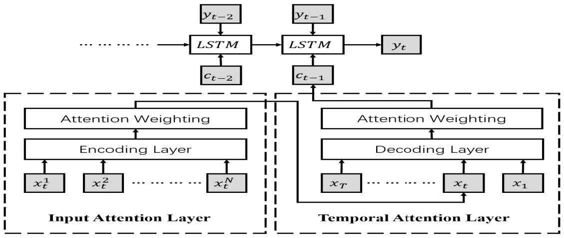

Fig. 3: Dual-Attention based RNN Architecture.

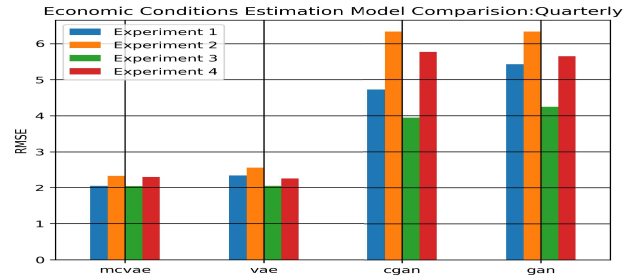

Fig. 4: Model Comparison

The design of DA-RNN is demonstrated in Figure 3, & the detailed learning procedure is illustrated in Algorithm 2. Specifically, DA-RNN consists of two LSTM networks that incorporate attention mechanism to select relevant features. The first LSTM component (input attention component) encodes the input exogenous features on different time step $\mathbf{X}_{1:T} = \{\mathbf{x}_1, \ldots, \mathbf{x}_{T-1}, \widetilde{\mathbf{x}}_T\}$, specifically $\widetilde{\mathbf{x}}_T$ is the estimation of current exogenous features (economic conditions) from our proposed MCVAE model as mentioned previously. For each $\mathbf{x}_t, 1 \leq t \leq T, t \in \mathbb{N}^+$, we have $\mathbf{x}_t = \{x^1_t, x^2_t, \ldots, x^N_t\}$ consisting of N different modalities, & each modality is given a weight factor $\alpha^k_t, 1 \leq k \leq N, k \in \mathbb{N}^+$ based on a attention-based network to construct a weighted representation of original inputs as: $\widehat{\mathbf{x}}_t = (\alpha^1_t x^1_t, \alpha^2_t x^2_t, \ldots, \alpha^N_t x^N_t)$, where α^k_t is calculated by referring to the previous hidden state h_{t-1} & the cell state s_{t-1} in the encoder LSTM unit as

$$\alpha^k_t = \frac{exp(e^k_t)}{\sum_{i=1}^{N} exp(e^i_t)}, \qquad e^k_t = v_e^\top \tanh(W_e[h_{t-1}; s_{t-1}] + U_e x^k_t) \tag{4}$$

where $v_e^\top$, W_e, & U_e are model parameters to learn.

In the temporal attention component, the LSTM units take $\widehat{\mathbf{x}}_t$ as inputs & generates it's corresponding hidden state h'_t, which is combined with the previous decoder hidden state d_{t-1} & the cell state of the LSTM unit s'_{t-1} to then generate the attention weight β^i_t of each hidden state $h'_i, 1 \leq i \leq T$ as:

$$\beta^i_t = \frac{exp(l^i_t)}{\sum_{j=1}^{T} exp(l^j_t)} \tag{5}$$

where $l^i_t = v_d^\top tanh(W_d[d_{t-1}; s'_{t-1}] + U_d h'_i)$ & $v_d^\top$, W_d, & U_d are model parameters to learn. The attention factor β^i_t indicates the importance of the i-th encoder hidden state for the prediction, & the context vector c_t is defined as:

$$c_t = \sum_{i=1}^{T} \beta^i_t h'_i \tag{6}$$

or a weighted sum of all the hidden states $\{h'_1, h'_2, \ldots, h^i_T\}$.

After the attention-based feature learning & transformation is completed, we finally obtain the context information c_t for predicting bank performance at time t along with its corresponding historical time-series data $y = \{y_1, y_2, \ldots, y_{t-1}\}$,

TABLE III: Economic Conditions Estimation Model Comparison

Iter=30,LR=1e-3		Experiment Results (Testing Set)															
Metric		RMSE				RMSE_{std}				LLD				LLD_{std}			
Split	Model	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Qtrly	MCVAE	**2.05**	**2.33**	**2.04**	2.30	**.028**	**.03**	**.03**	**.025**	-10.64	-15.53	**-10.73**	-19.05	1.53	4.81	1.92	5.74
	VAE	2.34	2.56	2.05	**2.26**	**.004**	**.002**	**.001**	**.002**	**-8.36**	**-8.14**	**-8.44**	**-8.76**	**.98**	**.109**	**.48**	**.48**
	CGAN	4.72	6.33	3.94	5.76	1.17	1.26	.89	1.08	-10.79	-10.39	-11.00	-10.52	1.51	1.84	**1.61**	**1.59**
	GAN	5.42	6.33	4.24	5.64	.86	.89	.75	1.04	**-10.51**	**-10.21**	-11.47	**-10.51**	**1.37**	**1.70**	1.68	1.79
Q1	MCVAE	**2.17**	**2.35**	2.28	2.29	**.04**	**.04**	**.06**	**.13**	-11.78	-15.65	-13.22	-17.12	3.12	7.12	5.76	8.47
	VAE	2.44	2.64	**2.17**	**2.26**	**.01**	**.005**	**.003**	**.006**	**-8.01**	**-8.04**	**-8.47**	**-8.47**	**1.08**	**.14**	**.68**	**.57**
	CGAN	4.91	6.20	4.39	5.48	.99	1.46	1.23	.99	-11.00	**-10.13**	-11.67	**-10.99**	1.87	1.65	2.14	**1.62**
	GAN	5.36	6.44	4.17	5.11	.93	.88	.98	1.03	**-10.30**	-10.80	**-11.57**	-11.14	**1.39**	**1.63**	**1.59**	1.75
Q2	MCVAE	**2.37**	**2.23**	2.12	**2.24**	**.06**	**.06**	**.07**	**.08**	**-10.23**	-12.30	**-10.48**	-12.15	1.82	4.71	5.01	3.31
	VAE	2.38	2.65	**2.07**	2.26	**.01**	**.006**	**.002**	**.004**	**-8.10**	**-8.04**	**-8.48**	**-8.62**	**.78**	**.16**	**.27**	**.51**
	CGAN	5.29	6.16	3.86	5.49	1.20	1.40	1.07	1.05	-10.88	**-10.95**	-10.99	**-10.54**	1.73	2.01	1.80	1.68
	GAN	5.09	6.72	3.75	4.95	1.00	1.04	.81	1.05	-10.76	**-10.95**	-11.64	-11.19	**1.57**	**1.64**	**1.49**	**1.51**
Q3	MCVAE	**2.04**	**2.15**	2.30	2.23	**.09**	**.07**	**.07**	**.05**	-11.35	-14.23	-14.06	-16.05	3.92	4.57	6.48	6.83
	VAE	2.12	2.18	**1.82**	**1.93**	**.01**	**.005**	**.002**	**.004**	**-8.20**	**-8.31**	**-8.53**	**-8.96**	**.77**	**.10**	**.45**	6.84
	CGAN	4.49	6.29	3.89	4.74	1.31	1.20	1.24	.97	-10.82	-9.84	-11.07	**-10.84**	1.91	1.68	1.93	**1.54**
	GAN	4.80	6.16	3.79	4.92	.96	.65	.92	.76	**-10.59**	**-9.65**	**-10.89**	-11.04	**1.43**	**1.18**	**1.42**	1.62
Q4	MCVAE	2.67	3.03	2.65	3.11	**.08**	**.04**	**.06**	**.04**	-13.04	-15.35	-11.65	-15.44	6.51	6.73	6.72	12.05
	VAE	**2.41**	**2.76**	**2.14**	**2.57**	**.008**	**.002**	**.002**	**.005**	**-8.09**	**-8.20**	**-8.5**	**8.79**	**0.72**	**.03**	**.75**	**.79**
	CGAN	4.77	6.20	3.95	5.52	.89	.91	1.01	.76	-11.11	-10.9	-11.36	-11.65	1.58	1.83	2.23	2.20
	GAN	5.43	6.32	4.18	5.34	.58	.65	.77	.76	**-10.6**	**-10.2**	**-10.56**	**-10.47**	**1.43**	**1.58**	**1.38**	**1.34**

Config	Split	Train	Test	Desc.
Exp.1	Time	1990Q1-2007Q4	2008Q1-2016Q4	Projections
	Bank	80% of Banks in 1990-2016	Remaining 20% of Banks in 1990-2016	After Financial Crisis
Exp.2	Time	1990Q1-2015Q4	2016Q1-2017Q1	Projections
	Bank	80% of Banks in 1990-2017	Remaining 20% of Banks in 1990-2017	With All Data.

TABLE IV: Bank Capital & Loss Prediction Experiments

for which we formulate a simple LSTM framework by concatenating y_{t-1} & c_{t-1} together to infer y_t as:

$$\begin{aligned} y'_{t-1} = w^\top[y_{t-1}; c_{t-1}] + b, d_t = f(d_{t-1}, y'_{t-1}), \\ y'_t = v_y^\top(W_y[d_T; c_T] + b_w) + b_v \end{aligned} \tag{7}$$

where f represents a LSTM unit, w, b, v_y, W_y, b_w & b_v are parameters to learn. The dimensional & temporal consideration of DA-RNN model makes it an effective candidate for our prediction task given that differing banking performance profiles & economic condition characteristics may affect bank capital & loss at specific time periods. This is particularly relevant for the financial industry since the recent crisis' (e.g., housing crisis, internet dotcom bubble, etc) have shown to be predominately influenced by varying banking & economic factors.

V. Experiment

A. Data Description

Data is acquired from Wharton Research Data Services, Global Financial Data, & the U.S. Federal Reserve Board of Governors. We sample 1000 of the largest banks based on consolidated assets that have at least eight consecutive quarters reported.

B. Economic Conditions Estimation

Experiment Setup. $E\hat{C}O_{mod}$ will be modeled with both quarterly & yearly intervals consisting of four separate experiments for each configuration. The purpose of each experiment configuration is to understand how well generative modeling can represent the data distribution for a known crisis event, all of the available data & time aligned data (Ref. Table II). For all experiments, we use $Z_{domestic}$ as our conditional modality.

Evaluation Metrics. Root mean squared error (RMSE) evaluation metric, $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}$, is used to measure model generated data compared to our ground truth testing data across models. Y_i is the actual value of each dimension in the economic condition represented in a particular modality, $e_i \neq e_{cond} \forall \in E\hat{C}O_{mod}$, while $\hat{Y}_i$ is the estimated value based on the generative model's approximation of e_i data distribution. A lower RMSE value indicates better model performance. We also use the popular Annealed Importance Sampling (AIS) with bidirectional Monte-Carlo for validation of generative models [19] to determine log likelihood upper bound (LLD). The higher the value of this metric indicates more plausibility of model output.

Baseline Algorithm. We compare our model to set of state-of-the-art baselines, including (1) Variational Auto-encoder(VAE) [13], similar to our model but does not consider multimodal or conditional aspects, (2) Conditional Generative Adversarial Networks(CGAN) [11], learns to sample data from random white noise until generated data is indistinguishable from the true distribution with conditional aspects, & Generative Adversarial Networks(GAN) [20], same as CGAN but without the conditional aspects.

Overall Performance. The performance of the models is summarized in Table III. Our proposed generative model, MCVAE, performs well in 10 out of the 20 total experiments (Fig. 4). The MCVAE & VAE were able to achieve low RMSE standard deviations indicating robustness, while the GAN & CGAN had large variation. Also, the mean LLD of MCVAE shows that it is often close & provides plausible estimates that are in range with baseline models. However, VAE & GAN dominate this category.

C. Bank Capital & Loss Ratios Prediction

Experiment Setup. Y_t will be modeled at both quarterly & yearly intervals while considering two data split types, (a) Time, where we predict the Y_t for the testing set (b) Bank , where 80% of the banks are used for training to then predict all of the Y_t for the 20% of the banks in testing, & two experiment settings (Ref. Table IV).

Evaluation Metrics. The previously mentioned RMSE is used to compare model estimates to the testing set values. Note, Y_i is the actual loan loss or capital ratio, $\hat{Y}_i$ is the respective estimated value & n is the number of testing observations. Lower RMSE indicate better performance.

Baseline Algorithm. Baseline models used to compare our proposed method to evaluate performance include, (1) Long Short Term Memory [12]; (2) Gated Recurrent Unit (GRU)

TABLE V: Bank Capital & Loss Ratio Prediction Model Comparison

Performance (RMSE); Epoch=5,LR=1e-2																					
Setup	Interval	Qrtly				Q1				Q2				Q3				Q4			
Experiment	Config	Time		Bank		Time		Bank		Time		Bank		Time		Bank		Time		Bank	
Target	Exp.#	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
Y_{T1CR}	DA-RNN*	**.21**	**.19**	.15	.18	**.19**	**.15**	.12	**.14**	.26	.25	.25	**.17**	**.18**	**.15**	.16	**.16**	.25	**.26**	.14	**.17**
	LSTM	.24	.23	**.12**	**.15**	.21	.23	**.11**	**.14**	**.24**	.24	**.13**	.18	.44	.24	**.14**	**.16**	.23	.29	**.12**	.19
	GRU	.23	**.19**	.17	.17	.20	.16	.18	.23	.31	**.23**	.15	.27	.31	.19	.19	.17	**.22**	.27	.16	.25
Y_{ncoR}	DA-RNN*	**.02**	**.03**	.04	.018	**.06**	**.01**	**.008**	**.02**	**.02**	**.02**	.05	**.01**	**.02**	**.02**	**.03**	**.02**	**.03**	**.02**	.03	**.03**
	LSTM	.05	.09	**.03**	.016	.12	.06	.01	.03	.06	.05	**.02**	.02	.08	.07	.10	.03	.05	.09	**.02**	**.03**
	GRU	.08	.19	.06	**.10**	.13	.03	.08	.07	.10	.06	.10	.09	.14	.07	.10	.10	.05	.10	.11	.09

[21]. LSTM & GRU are popularly used in the natural language processing domain, but have also enjoyed success in time-series prediction tasks due to their dynamic temporal structures.

Overall Performance. The performance of the models over all twenty experiments for each of the target variables, $[Y_{T1CR}, Y_{ncoR}]$, is summarized in Figure V. DA-RNN shows consistent performance against baseline models for 24 of 40 experiments (Ref. Table V) with particular success in predicting Y_{ncoR}(15 of 20 experiments), for yearly data (Fig. 5). Poor performance in the bank split setting could be due to the differences among bank holding companies which may affect generalization. Moreover, challenges with yearly experiments could be explained by data quality, volume & variability among the banks.

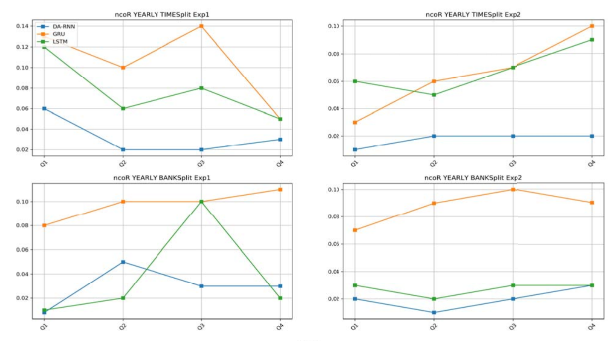

Fig. 5: Yearly Y_{ncoR} RMSE

VI. Conclusion

In this paper, we first proposed a modified multimodal generative model with conditional aspects to derive a joint probability distribution among multiple exogenous modalities based on domain specific economic variables. We then proposed a bespoke dual-attention recurrent neural network model specific to bank stress-test tasks with consideration for dimensional significance over a temporal space. To the best of our knowledge, this is the first attempt to incorporate a MC-VAE & DA-RNN for the stated application tasks. Furthermore, extensive experiments were conducted on real-world economic & banking data to validate on baseline models to determine the effectiveness of our proposed framework.

VII. Acknowledgement

Partial funding by Postgraduate Research & Practice Innovation Program of Jiangsu province (KYCX18-0045).

References

[1] B. Hirtle, A. Kovner, J. Vickery, and M. Bhanot, "Assessing financial stability: the capital and loss assessment under stress scenarios (class) model," *Journal of Banking & Finance*, vol. 69, pp. S35–S55, 2016.

[2] F. Reserve, "Comprehensive capital analysis and review 2015: Assessment framework and results," *US Federal Reserve, Washington, DC, March*, vol. 11, 2015.

[3] F. B. Covas, B. Rump, and E. Zakrajšek, "Stress-testing us bank holding companies: A dynamic panel quantile regression approach," *International Journal of Forecasting*, vol. 30, no. 3, pp. 691–713, 2014.

[4] N. Malik, "Can banks survive the next financial crisis? an adversarial deep learning model for bank stress testing," *SRN: https://ssrn.com/abstract=3208177*, 2018.

[5] K. L. Kliesen, M. T. Owyang, E. K. Vermann *et al.*, "Disentangling diverse measures: A survey of financial stress indexes," *Federal Reserve Bank of St. Louis Review*, vol. 94, no. 5, pp. 369–397, 2012.

[6] WRDS, "Bank regulatory," *wrds.wharton.upenn.edu*, 2019.

[7] M. T. Bayoumi, M. H. Faruqee, M. D. Laxton, M. P. D. Karam, M. A. Rebucci, M. J. Lee, M. B. Hunt, and I. Tchakarov, *GEM: A new international macroeconomic model*. International Monetary Fund, 2004, no. 239.

[8] M. H. Pesaran, T. Schuermann, and S. M. Weiner, "Modeling regional interdependencies using a global error-correcting macroeconometric model," *Journal of Business & Economic Statistics*, vol. 22, no. 2, pp. 129–162, 2004.

[9] M. Flood and G. Korenko, "Systemic scenario selection," Working Paper 0006, Office of Financial Research, US Treasury Department , Tech. Rep., 2012.

[10] P. Glasserman, C. Kang, and W. Kang, "Stress scenario selection by empirical likelihood," *Quantitative Finance*, vol. 15, no. 1, pp. 25–41, 2015.

[11] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[13] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[14] K. Sohn, H. Lee, and X. Yan, "Learning structured output representation using deep conditional generative models," in *Advances in neural information processing systems*, 2015, pp. 3483–3491.

[15] M. Wu and N. Goodman, "Multimodal generative models for scalable weakly-supervised learning," in *Advances in Neural Information Processing Systems*, 2018, pp. 5580–5590.

[16] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[17] Y. Qin, D. Song, H. Chen, W. Cheng, G. Jiang, and G. Cottrell, "A dual-stage attention-based recurrent neural network for time series prediction," *arXiv preprint arXiv:1704.02971*, 2017.

[18] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[19] Y. Wu, Y. Burda, R. Salakhutdinov, and R. B. Grosse, "On the quantitative analysis of decoder-based generative models," *CoRR*, vol. abs/1611.04273, 2016. [Online]. Available: http://arxiv.org/abs/1611.04273

[20] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[21] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.