DEEP LEARNING FOR FINANCIAL BANKING STRESS TEST ANALYTICS

by

Farid Razzak

A Dissertation submitted to the

Graduate School-Newark

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Management

written under the direction of

Professor Hui Xiong

and approved by

____

____

____

____

Newark, New Jersey

May 2020

ABSTRACT OF THE DISSERTATION

Deep Learning for Financial Banking Stress Test Analytics

By FARID RAZZAK

Dissertation Director: Dr. Hui Xiong

Since the recent financial crisis of late 2008, several global regulatory authorities have collaboratively mandated stress-testing exercises. These exercises evaluate the potential capital shortfalls & systemic impacts on large banks in hypothetical adverse economic scenarios, which try to simulate the macro-economic conditions similar to recent crisis'. The ability to relate dynamic economic conditions with banking performance profiles to identify meaningful relationships could provide significant insights for bank capital & loss projections.

In this dissertation, the practical challenges that face bank stress-test analytics are examined and approached using advanced analytical techniques.

Initially, (1) through a rigorous examination of an economic condition estimator (ECE), which learns joint approximation representations among exogenous factors by analyzing the complex non-linear relational combinations among the real-world economic indicators using a multi-modal conditioned variational auto-encoder (MC-VAE). Experimentation on real-world economic conditions from the U.S. regulatory stress test exercise (CCAR) over the last three decades demonstrates the model's effectiveness.

Additionally, (2) a focused study on bank capital & loss prediction (BCLP) methodology that can incorporate economic conditions as an estimated variable while also considering dynamic variability of potential crisis profiles that better provide a robust prediction of capital & loss. Demonstrations through experiments show that the BCLP model outperforms baseline & state-of-the-art methods from literature when evaluated on a sample of 1000 U.S. bank holding companies' historical consolidated financial statements (FR-9YC) from the past three decades.

Both the ECE & BCLP model frameworks together form the Integrated Multimodal Bank Stress Test Predictor (IMBSTP) framework to provide a data-driven end to end bank stress testing analytical tool.

Lastly, (3) a preliminary overview of the Transferable Knowledge for the Bank Capital Components (TKBCC) model framework is discussed. The framework assumes that banks inherently share hidden intrinsic qualities and leverages inductive transfer learning techniques to improve bank capital-components predictions for domain tasks with limited training data. The performance of preliminary experiments on the proposed model framework through consolidated financial statements from the China Stock Market Accounting Research Database (CSMAR), and the Wharton Research and Data Service's (WRDS) repositories from the last two decades demonstrate the utility of the TKBCC model framework.

ACKNOWLEDGEMENTS

First & foremost, I thank Allah for the blessed life I lead & for the privileges I possess. I want to express appreciation to everyone who has provided me with tangible & intangible support from the beginning of my doctoral studies until my graduation.

Starting with family, I thank my incredible wife, Tanzima Razzak, my daughter, Talia Inaaya Razzak & my future children, for their love, patience & dedication through this chapter in my life & beyond. I dedicate my doctoral degree to my father, Mohammed Abdur Razzaque, & mother, Fazilatun Nessa, for their spiritual & material devotion toward the success of all their children. I would also like to thank my eldest brother, Faisal Rajib, his wife, Tahmina Rajib, & their two sons, Evan Ardeen Rajib & Samil Ayzaan Rajib as well as my second eldest brother, Dr. Faruk Razzak, his wife, Dr. Devitri Razzak, their son Aiyan Kiran Razzak & daughter Anaiya Jana Razzak. I also thank my father-in-law Tarikul Islam, & mother-in-law Shewly Parveen Islam for their confidence in me through all my endeavors.

I want to express my deepest gratitude & respect to my doctoral advisor, Professor Hui Xiong, who has provided unconditional support, guidance, encouragement & wisdom, and his family. My doctoral journey was made productive & pleasant as possible under his direction. I am forever indebted to him for the nurture, kindness, & extraordinary family I have gained through being a part of his unique research group.

I also would like to thank my dissertation committee members: Professor Lin Peng of CUNY Baruch, Professor Thomas Lidbetter,& Professor Xiaodong Lin, for their engagement, constructive feedback & for having my best interests in completing my dissertation.

I acknowledge the Department of Management Science & Information Systems' (MSIS) Luz Kosar & Prof. Jaideep Vaidya, the Ph.D. in Management Program's Monnique DeSilva, Assistant Dean Goncalo Filipe,& Program Director Suresh Govindaraj for their efforts in facilitating my progress towards graduation. Humblest praise to the Center for Data Mining & Business Analytics (CDMBA) for providing the resources, platform & tools necessary to conduct research; The senior alumni members of CDMBA; Prof. Wenjun Zhou, Prof. Yong Ge, Prof. Keli Xiao, Dr. Hengshu Zhu, Dr. Zhongmou Li, Prof. Chuanren Liu, Prof. Meng Qu, Dr. Yanchi Liu, Dr. Zijun Yao, Dr. Bin Liu, Prof. Yanjie Fu, Prof. Jingyuan Yang, Prof. Junming Liu. The visiting scholars of our group; Dr. Yang Yang, Dr. Fei Yi, Nengjun Zhu, Prof. Yuanbo Xu, Prof. Fuzhen Zhuang, Prof. Guannan Liu, Prof. Ning Fu, Hao Yi. Also to current to peers, Dr. Qingxin Meng, Denghui Zhang, Mengfei Teng, Jingci Ming, Zixuan Yuan, & Xiaoru Gao, who continuously inspired me with their determined spirits.

Lastly, special thanks to my peers: Dr. Ussama Yaqub, Dr. Zamil AlZamil, Dr. Hao Zhong, Dr. Constantine Vitt, Dr. Deniz Appelbaum, Ken Chen (Can Chen), Hafiz Salman, who played essential roles during my doctoral journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Background

The recent financial crisis of 2008 and the ensuing economic recession in the United States have brought significant advances to capital adequacy requirements. Mandated regulations of financial institutions test their sustainability through hypothetical stressed macro-economic scenarios while requiring transparency of both actual and projected banking performance.

Effectively, regulators want to ensure that large banks are providing loans while carefully considering the systemic and economic risks that could impact their solvency. From the regulator's perspective, if the size of the bank deposits and revenues are more substantial than the size of the loans lent out by banks, the bank should be able to withstand the effects that potential adverse economic circumstances could inflict on their performance. However, this regulatory strategy may be counter-intuitive to the profit generation goals of the bank, which can create a unique circumstance of conflicts that warrants ongoing regulatory efforts.

Moreover, global regulatory reporting requirements of banks may be disparate due to non-adherence to accounting practice standards, such as BASEL III. The obscurity of banking performance due to restrictions that may impact regulatory transparency could circumvent the accurate depiction of national financial health. Financial systems in Europe and Asia, for example, provide limited firm-level financial statements and may experience the "invisible hand" effect of non-public interventions by their respective government authority. All of which can lead to a global financial crisis due to lack of appropriate preemptive warning signals.

1.2 Bank stress-test Analytics

Typically, regulators who administer stress-test exercises implement a dual-approach consisting of "bottom-up" & "top-down" methodologies (Covas, Rump, & Zakrajšek, 2014). In the "bottom-up" method, techniques to estimate the loan loss and revenue impacts to the bank use detailed confidential financial projections of institution-specific loan portfolios submitted directly to the regulators. Alternatively, the "top-down" approach depends on publicly released bank financial statements to estimate losses and revenues. Since this type of data is highly available, it is commonly used by regulators to benchmark aggregated projections from the "bottom-up" approach and evaluate different capital plan strategies in dynamic macro-economic conditions. Effective stress-testing can be a beneficial tool for providing insights and strategies to prevent catastrophic losses by financial institutions during severe economic conditions (Malik, 2018; Hirtle, Kovner, Vickery, & Bhanot, 2016; Reserve, 2015).

Currently, in the U.S., the Dodd-Frank Act stress-test (DFAST) (DODD-FRANK & ACT, 2010) & Comprehensive Capital Analysis and Review (CCAR) (Reserve, 2015) are regulatory exercises administered by the Federal Reserve. The results of an exercise typically indicate if a bank's projected stressed capitalization would be sufficient relative to a regulatory threshold, which could range between 3 to 8 percent depending on the bank's risk rating, capital ratio, exercise scenario & examination year (Reserve, 2015; Covas et al., 2014).

In Europe, the European Banking Authority (EBA) is responsible for ensuring proper function, integrity, and stability of financial markets and systems within the European Union. The EBA assesses market developments as well as identifies trends, potential risks, and vulnerabilities stemming from the micro-prudential level. Through the use of bottom-up based EU-wide stress-test exercise, which aims to assess the sustainability of financial institutions in adverse market developments and systemic risks within the EU financial system.

In global regions with developing or emerging markets, the International Monetary Fund (IMF) and World Bank administer the Financial System Stability Assessment Program (FSAP), an in-depth analysis of a country's financial sector due to the systemic effects the recent financial crises have illuminated both domestically and internationally. The FSAP assessment administers exercises using both top-down or bottom-up stress-testing methodologies.

Bank stress-testing exercises predominately focus on the ability of banks to maintain capital adequacy during turbulent economic conditions. Assessment of the capital adequacy of banks during expected, adverse, and severely adverse economic conditions, through financial measures about banking capital, revenues, loss rates, counterparty, industry, economy, and systemic level risks, are all considered. Banking capital is understood to mean the monetary resources firms keep aside after considering all expenses and responsibilities carried over into the next period for purposes of the firm's benefit. In the case of bank stress-tests, banking capital that firms kept on hand in proportion to their risk-weighted assets could indicate the amount of "safety" buffer. The capital buffer allows for mitigation of losses suffered from firm loans, investments, or services due to adverse economic conditions, which may provide a higher likelihood of survival through the turmoil.

Given regulatory mandates, sanctions, penalties, and market-impacts of bank stress-test exercises since the most recent financial crisis, stakeholders of all levels, such as national government, investors, economists, regulators, and consumers, all have a vested interest in understanding the relationship between bank performance and economic conditions. Mainly, if the economic conditions were to be adverse, would a bank survive the ordeal without requiring a government "bailout" or failing their stakeholders to the point of crisis? To expand on this interest and to provide a sense of security and public confidence, the development of methods to forecast simulated banking performance, economic conditions, and regulatory compliance has become of paramount importance (Reserve, 2015).

Bank stress-test analytics depicts both the quantitative & qualitative measures considered when conducting tasks related to providing accurate regulatory capital adequacy measurements. Individually, components of capital are first analyzed. Loan-portfolio loss projections, which depict the net charge off rates (loan loss rates) banks suffer from their primary loan business from previous periods. Additionally, pre-provision net revenue projects, which depict the streams of positive cash flows earned through different operations and services within the bank, and calculation of financial ratios, such as capital, common-equity, and leverage, that best capture attributes of overall capital to withstand turbulent circumstances. With the overall goal of improving economic & bank performance forecasts, providing accurate projections of both provides relevant information that may directly impact bank stress-test exercises.

As bank stress-test exercises evolve, stakeholders give particular focus on identifying preemptive safeguards against behaviors or patterns that may lead to the next financial crisis through data-driven quantitative investigation that can provide interpretive insights. Financial data standardization, availability, integrity, and quality will allow for feasible data-driven advanced analytical techniques applied to the core bank stress-test tasks. Generalizations may become more robust as banks may need to meet bespoke conditions depending on their profiles or past behaviors (Tarullo, 2016). State of the art methods in different data analysis domains may apply to the bank stress-test research area.

1.3 Preliminaries

Definition 1 (Economic Conditions)

In a top-down methodology approach to stress-testing, the belief is that exogenous economic factors influence the trajectory of a bank's performance. Therefore, considering common factors to understand the overall state of global, national, industrial, or financial health is crucial to comprehending how banks are likely to be affected by economic conditions.

The historical macro-economic variables that depict the U.S. and parts of the global financial economy, as per the Federal Reserve's Comprehensive Capital Adequacy Review (CCAR) (Malik, 2018; Reserve, 2015), consists of 16 domestic variables,

Z_{macro_D}, & 12 global variables, Z_{macro_I}, reported quarterly between 1976-2017.

Historical micro-economic variables represent specific aspects of the financial economy as it pertains to the benchmarks for representative financial instruments & asset-types. To this end, U.S. treasury, inflation, major commodity & stock indices returns, government bonds interest rates, interest rates swaps, currency swaps rates, & major commodities price are collected quarterly between 1976-2017 to depict the micro-economy, Z_{micro}.

The initial signals of the previous financial crisis emanated from the housing bubble, created due to risky loans that possessed a high probability of default to the financial institutions that held the risky assets. Intuitively, monitoring of financial and real estate sub-sectors can provide an early indication to similar scenarios. Historical sector-based indices, $M_{SP_{FIN}}$, depict the financial & real-estate market sectors since the most recent crisis impacted these areas most substantially. The S&P 500 Financial Sector Index consists of three tickers that provide illumination of conditions specific to the financial industry, while the S&P Real Estate Indices,$M_{SP_{RE}}$, consists of tickers that tracks more granular movements in the real estate industry.

Several industry index measurements that try to capture the overall financial conditions of the U.S. banking sector and financial economy, M_{FCI}, are developed by the regional Federal Reserve Boards (i.e., St.Louis, Chicago, Kansas City). Corporate research entities (Bloomberg, Goldman Sachs) also capture the directional conditions in money markets, debt and equity markets, and the traditional "shadow" banking systems to provide an overview perspective of significant banking economy aspects (Kliesen, Owyang, & Vermann, 2012).

$$ECO_{mod} = [Z_{macro}, Z_{micro}, M_{SP}, M_{FCI}] \tag{1.1}$$

Definition 2: (Banking Performance Profile)

Another aspect of the top-down stress-test methodology is the reliance on publicly released bank financial statements. The assessment of the respective bank's loan portfolio, loan loss rate, and net revenue represent significant factors related to projections of capital adequacy under different economic conditions. Loan portfolio breakdown of seven categories, $X_{loancat_{i,j}}$, of bank holding companies through their respective consolidated financial statements provide insight into banking performance. The loan categories represent a snapshot view of the bank's lending practice to different segments of borrowers who can be influenced by economic conditions and therefore affect the bank's risk exposure. Determining insight from $X_{loancat_{i,j}}$, where i is an individual bank and 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 it's loan portfolios 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, which then can be used to determine the loan loss rate, $Y_{ncoR_{i,j}}$, as seen in Eq.1.2.

$$Y_{ncoR_{i,j,t}} = 100 * \frac{X_{NCO_{i,j}}}{X_{loancat_{i,j_{t-1}}}} \tag{1.2}$$

Banks are able to generate revenue from interest earned from loans, trading income and other revenue-generating services, however they also have to consider expenses such as compensation, fixed assets, and other non interest earning operations, $X_{Cmpppnr_{i,j}}$. To determine the net revenue proportional to the bank's consolidated assets, $X_{CnsldAsts_{i,t-1}}$, the pre-provisional net revenue ratio of the bank can be derived, $Y_{ppnr_{i,j_t}}$ (Covas et al., 2014). Understanding the bank's ability to generate revenue during dynamic economic conditions, as shown in Eq.1.3, can be crucial to offset projections of losses for better forecast.

$$Y_{ppnr_{i,j_t}} = 100 * \frac{X_{Cmpppnr_{i,j}}}{X_{CnsldAsts_{i,t-1}}} \tag{1.3}$$

The bank's monetary reserves, retained net earnings, or equity capital, X_{EqCap}, are considered to be the funds it has available, however, these funds may be further reduced due to strategic decisions to issue dividends or perform stock repurchases (X_{EqPO}), payment of taxes ($\tau = 35\%$), and occurrence of regulatory capital deductions (X_{RegDt}). The relevant calculations are summarized X_{EqCap} in Eq. 1.4 and 1.5.

$$Net_{revloss_t} = (\sum_j Y_{ppnr_{i_t}} - \sum_j Y_{ncoR_{i_t}}) \tag{1.4}$$

$$X_{EqCap_t} = X_{EqCap_{t-1}} + (1-\tau) * Net_{revloss_t} - X_{EqPO_{i,t-1}} \tag{1.5}$$

Once the executive leadership of a bank make the relevant decisions, the remaining capital in proportion to the bank's previous period risk weighted assets,X_{RWA}, is considered to be the capital ratio (Y_{T1CR}), a measure that best depicts the bank's overall capital adequacy. Regulators, investors and economists are most interested in the Tier-1 common ratio, which only considers bank equity elements as part of the capital, as seen in Eq.1.6

$$Y_{T1CR} = \frac{X_{EqCap} - X_{RegDt_{t-1}}}{X_{RWA_{t-1}}} \tag{1.6}$$

Definition 3: (Transferable Banking Knowledge)

Large banks ultimately play similar roles globally with respect to their impact on consumers, national economy, and investors. Regulatory reporting practices may not be consistent throughout global nations, which may cause difficulty in attaining transparency into bank performance and overall impact on stakeholders. Access, availability, integrity, quality and national security may be reasons for the level of obscurity some nations may offer when representing firm-level financial performance characteristics.

Moreover, the accounting practices at banks of different nations may not follow global accounting standards, such as BASEL. Non-standardized practices cause further challenges when comparing banking systems as they may not collect similar performance characteristics, which can further complicate the compatibility of knowledge discovery tasks at a global level.

However, even with the aspects mentioned above, which may hinder comparisons of banks that predominately operate in international financial systems, banks ultimately serve a similar purpose to their stakeholders and conduct business with related strategic, operational initiatives.

Focus on obtaining the fundamental strategy and operation characteristics from firm-level banking performance profiles of a financial banking system may provide critical insights. A good representation of basic banking practices, reactions, and strategies could serve as a crucial factor in helping model related components in a foreign banking system that may lack necessary historical data.

Definition 1.1:(Domain) Consisting of two components, feature space, X and marginal probability distribution, $\Pr(X)$, where $X = x_1, ..., x_n$, therefore a "Domain" is equivalent to, $D = X, P(X)$ (Behbood, Lu, & Zhang, 2011)

In this dissertation, X will consist of banking profile attributes representing of characteristic features from the firm's respective consolidated financial statement, $X_{BankAttr}$, and exogenous economic modalities, ECO_{mod}. Thus, defining the domain to be specific to the financial stress-testing and banking environment.

Definition 1.2:(Task) Also consisting of two components, (1) a label space, $Y = y_1, ..., y_m$ and an objective function, $f(.)$, for the purposes of prediction of unobserved instances by learning decision boundaries from x_i, y_i pairs. Thus the "Task" is defined as $T = Y, f(.)$ (Behbood et al., 2011).

For this dissertation, Y is the target features that represent banking capital components, $[Y_{T1CR}), Y_{ncoR}, Y_{EqCap_{RetEarn}}]$ relevant to stress-test analysis.

Definition 1.3:(Transfer Learning) Provided a source domain D_s and learning task T_s, a target domain D_t and learning task T_t, transfer learning aims to help improve the learning of the target predictive function $f_t(.) \in D_t$ using the knowledge in D_s and T_s, where $D_s \neq D_t$ or $T_s \neq T_t$ (Behbood et al., 2011).

This dissertation work defines the domains, D_s, D_t to be related to the financial banking industry. At the same time, the learning tasks, T_s, T_t will focus on forecasting the bank capital components, $[Y_{T1CR}), Y_{ncoR}, Y_{EqCap_{RetEarn}}]$, related to stress-test analytics. Moreover, utilization of knowledge from, D_s, T_s or a financial banking system with regulatory transparency and plentiful firm-level instances, to then help improve the T_t by utilizing D_t or a financial banking system with limited regulatory transparency and limited access to firm-level instances as well as the knowledge available from D_s, T_s as a fundamental basis.

1.4 Research Motivation

An adverse outcome in the current post-financial-crisis era could have detrimental impacts on the bank. In the case of the CCAR, the most immediate repercussions include restrictions on a firm's capital distribution plan, effecting dividend payouts, share repurchases, and redemption of trust preferred securities (Covas et al., 2014).

Additionally, requirements to provide rectified compliance, strategy, and loss projections that, when re-assessed in the stress-test, could meet the regulatory thresholds (Reserve, 2015). However, the damage to the firm's reputation may directly influence its market value due to perception and confidence from stakeholders. This type of situation can cause pressure to change bank leadership or strategy (Wilmarth, 2014). Most importantly, the outcome could also serve as an industry-wide indicator to counter-party & systemic risk to avoid a potential crisis.

Thus, investigating the ability to robustly determine estimations or simulate adverse estimations of economic conditions for scenario-based analysis can prove to be beneficial to forecasting potential banking performance under said economic conditions. This type of scenario analysis can benefit both regulatory stakeholders as well as bank risk management stakeholders by providing foresight into potential mitigation scenarios to avoid the penalties from an adverse outcome. Additionally, assessing the scenario-based risk in international financial systems may be just as important as domestic financial systems. Due to the global interconnections that may have systemic impacts on domestic economics, global counter-party risk assessment, and for investment research purposes. Since not all nations follow a singular standardized regulatory reporting, accounting practices, or public disclosure protocols, it may be challenging to ascertain similar information from a diverse group of banking systems that are highly relate-able. Therefore, investigating approaches that can potentially circumvent the challenges of available transparent data while utilizing the knowledge from similar domains for related tasks can prove immensely beneficial. Since stakeholders who need to assess the scenario-based risks for foreign banking systems with limited available knowledge exist, demand for practical approaches for this task is evident.

1.5 Overview

In chapter 2 of this dissertation, an introduction of an economic condition estimation (ECE) model, which can incorporate multiple exogenous factors beyond what regulators typically leverage to represent the economic climate, is given. The ECE learns non-linear latent relationships among relevant macro-economic, micro-economic, and financial market indicators.

This model addresses practical challenges faced when considering economic data, such as data availability and integrity, and identifying dependent relationships among the economic conditions to understand how factors influence each other. Specifically, discussion of a bespoke multi-modal conditional & variational autoencoder (MCVAE) developed for macroeconomic & microeconomic conditions estimation.

In chapter 3 of this dissertation, elaboration on the bank capital & loss prediction (BCLP) model includes economic conditions as a feature that may be estimated, simulated, or historical with firm-level banking performance profiles. The BCLP evaluates the dimensional & temporal influences that exogenous economic factors and banking characteristics may have on banking performance components. The insights yielded from BCLP are essential to bank stress-test analytics that ultimately provides holistic projections of bank capital & loss ratios .

This model addresses the practical challenge of capturing the variability in the circumstances that depict a financial crisis and how they may impact different banking characteristics at different periods. Specifically, a customized dual-attention neural network (DA-RNN) which determines the critical dimensional and temporal aspects of the banking and economic features with the most influence on a particular target bank performance variables. The features deemed to have the most significant impact on the target at different periods are passed into a recurrent neural network (RNN) to perform time-series prediction tasks on target bank performance features.

The economic conditions estimation (ECE) model and the bank capital & loss prediction (BCLP) model work together to serve as a "top-down" bank stress-test analytical model framework as the *Integrated Multimodal Bank stress-test Predictor (IMBSTP)* model, designed to project bank capital & loan loss ratios in estimated economic scenarios. The overall framework of the proposed model is demonstrated in

Figure 3.10.

In chapter 4 of this dissertation introduces the transferable knowledge for the banking capital components (TKBCC) model framework. The TKBCC model framework (1) acquires pertinent banking knowledge from a source financial banking system that has regulatory transparency and firm-level financial data availability. (2) The knowledge is leveraged as a foundation to forecast banking capital components relevant to bank stress-test analytics on a target financial banking system that may not have as much regulatory transparency and firm-level financial data available for public disclosure.

The TKBCC model addresses the practical challenges of applying developed analysis tasks to a related domain. Capital-components prediction on U.S. financial bank system and the foreign financial bank system, which may have limited financial data available for model training due to regulatory reporting and data availability limitations, could benefit from the TBKCC. Advanced deep learning techniques determine which knowledge aspects of transferring with special consideration on performance impacts before the transfer of said knowledge across the defined source and target domains to perform a related task. Specifically, the inductive transfer learning approach seeks similarities between the source and target tasks, $T_s = T_t$. However, the domains are different, $D_s \neq D_t$ (Pan & Yang, 2009). In this dissertation, the D_s is represented by the U.S. financial banking system, while D_t as a different financial banking system or Chinese financial banks. Additionally, an initial investigation of a parameter transfer approach before extending into feature-representation & instance-transfer techniques that consider negative transfer impacts on overall model performance.

Finally, in chapter 5 of this dissertation concisely summarizes how the aforementioned deep learning techniques benefit essential aspects of bank stress-test analytics and how the approaches discussed provide practical utility.

1.6 Research Contributions

The literature in the research areas of bank stress-testing, macro-prudential supervisory regulations, economic forecasting, and financial balance sheet projections predominately and traditionally utilize statistical linear model approaches to address their respective tasks. However, contributions discussed in this dissertation leverage advanced analytical techniques that provide both robustness and sophistication to help improve the handling of research tasks.

As financial industries become more focused on data-driven solutions, analytical frameworks, data standardization & quality, and advanced technologies to help facilitate services, operations, and regulations, research areas such as bank stress-testing analytics, compliance analytics, and investment analytics will continue to demand attention from academia. Sub-industries such as FinTech (Financial Technologies), RegTech (Regulatory Technologies), & now SupTech (RegTech for Supervisors) (Center, 2017) commercially thrive in these research areas by developing production-level services or products that look to apply much of the practical elements available in the respective literature.

The focus of this dissertation aims to provide bespoke analytical frameworks from the state of the art machine learning & data mining research domains. Through investigation and experimentation, evidence indicates how the techniques benefit bank stress-test analytics tasks of economic conditions forecasting, bank capital & loss projections, and foreign banking system capital-components estimations.

The contributions of this dissertation can be summarized as follows:

- Introduce additional macroeconomic and microeconomic indicators beyond those used by global regulators to depict the economic conditions.

- Economic Conditions Estimation forecasting technique using a deep learning generative process that addresses the following:

  Incorporation of multiple microeconomic and macroeconomic conditions beyond those typically used by regulators, which can help enhance the depiction of a financial economy, and how it impacts on banking performance.

  Modification, development, and implementation of a variational auto-encoder generative model (M. Wu & Goodman, 2018) variant that leverages the following:

    The use of a conditional modality to help alleviate issues with data immediacy and availability. Requiring only a specified target modality to generate the estimation of the remaining target modalities, which can help with forecasting estimates when only one input feature is available readily and timely. Common in financial industry research is the disparate availability of regulatory, economic, or firm-level data. Conditionality will allow for the estimation process to function with one highly available, highly accessible, and inexpensive economic modality to forecast the remaining economic modalities.

The ability to learn from a multiple modality representation, which efficiently and effectively approximates the joint probability distribution of multiple exogenous factors as dependents of one and another, rather than independent. This ability would allow for the capture of the non-linear combination of relationships among the exogenous factors rather than only understanding the individualized relationships of each respective target economic modality.

Considerations for data sparsity and multi-modal co-occurrence challenge that happens when multiple modalities may not have aligned data occurrence along the same axis. Commonly found within financial and economic data since regulatory reporting may vary depending on banking characteristics, legislation, and national agency mandates. Economic conditions data may also be recorded or stored by various organizations at different periods following diverse practices causing economic modality data to have many co-occurrence observations to be missing.

- A Bank Capital & Loss prediction framework that leverages two attention-based neural networks and a recurrent neural network which offers the following:

  The ability to consider economic conditions features as a baseline, simulated or projected from the economic conditions estimation model.

  Customized attention-network adapted from (Qin et al., 2017) that can analyze banking characteristics data with economic conditions data to determine which dimensions are the most influential when predicting a target variable, in this case, a bank capital or loss ratio. Effectively this attention network performs a feature selection along the dimensional plane among the banking characteristics and economic conditions to determine which one has the most weight when forecasting the target capital & loss ratio for the respective bank.

A customized attention-network adapted from (Qin et al., 2017). The model considers temporal influences of each dimension when forecasting a target variable at different periods—performing a pseudo-feature selection to determine which dimensions have the most influence at different periods, which can best forecast the target variable. In the case of the bank stress-testing domain, being able to assess which banking characteristics, economic conditions, at which periods have the most influence when predicting bank capital & loss.

Incorporating the dimensional and temporal attention-based networks to function as a dynamic feature selection to capture the variability in banking characteristics at different periods. Allowing for effective modeling of the target banking capital & loss variables, with a recurrent neural network to complete a dual-stage recurrent neural network (DA-RNN) (Qin et al., 2017) designed to project time series target variables based on dimensional and temporal attention of input features. Effectively, an end to end model framework that analyzes the input banking characteristics & economic conditions modalities for the relevant dimensional & temporal features that best impact the prediction of the banking capital & loss target variables.

A key contribution to using this framework is the consideration of variability in crisis types. Generalizing a model based on one particular financial crisis may not be sufficient. Characteristics of the financial crisis may differ (e.g., student debt crisis, foreign credit crisis, tech bubble crisis, subprime mortgage crisis) at different periods. Therefore a model with the ability to dynamically assess profiles of banking characteristics, economic conditions profiles, and how they impact the target banking capital & loss features are crucial to generalize variability in crisis better'.

- An integrated model framework, Integrated Multi-modal Bank stress-test Prediction (IMBSTP) that consolidates economic conditions estimation and banking capital & loss prediction based on dimensional and temporal variables to conduct bank stress-testing analysis.

- The investigation, experimentation, and utility of transferable knowledge for banking capital components using transfer learning techniques to assess forecasting applicability to regulatory limited foreign banking systems.

Exploration of the assumption that financial banking systems fundamentally employ universal operations practices and strategies to help enhance forecasts of a foreign banking system with limited or restricted public regulatory transparency.

Inductive transfer learning techniques to assess the applicability and utility of knowledge available in a source domain of banks from the U.S. when leveraged in the target domain of a limited number of Chinese finance sector banks.

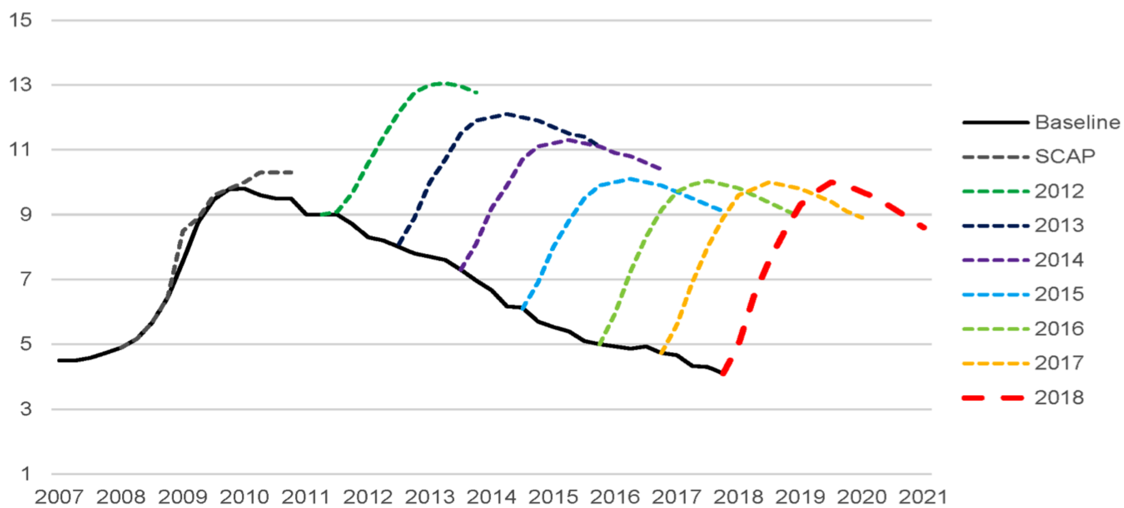
Aspects of positive, negative, and neutral transfer to evaluate the ability to improve forecasting performance by isolating knowledge that is most similar or relevant to the target domain from the source domain.

Analysis of feature-representation, instance-based, and parameter transfer approaches to determine the effectiveness of each technique when isolating financial economy and banking performance data to determine banking capital components.

- Providing contributions to state of the art for SupTech & RegTech literature. Exploring the use of advanced analytical techniques for examining bank stress-test properties (Malik, 2018; Covas et al., 2014; Hirtle et al., 2016) by utilizing deep learning techniques. Generative models, non-linear auto-regressive exogenous neural network models, and transfer learning models on a representative panel of experiments, which directly address to bank stress-test analytics tasks and challenges related to the respective problem setting.

CHAPTER 2

ECONOMIC CONDITIONS ESTIMATIONS



Source: Board of Governors of the Federal Reserve

Figure 2.1. Historical CCAR Unemployment Rate in Adverse Scenarios.

2.1 Introduction

Quantitative measures that represent the overall wealth, resources, production, trade, and consumption of goods & services typically defines a respective nation's economy. Moreover, the financial economy focuses on the distribution of resources and services about financial activities that impact consumer, private, public, and government sectors of a respective nation. Economic conditions are depicted by macroeconomic & microeconomic measurements, which aggregate vital aspects of business, industry, and country to provide a summarized perspective of activities.

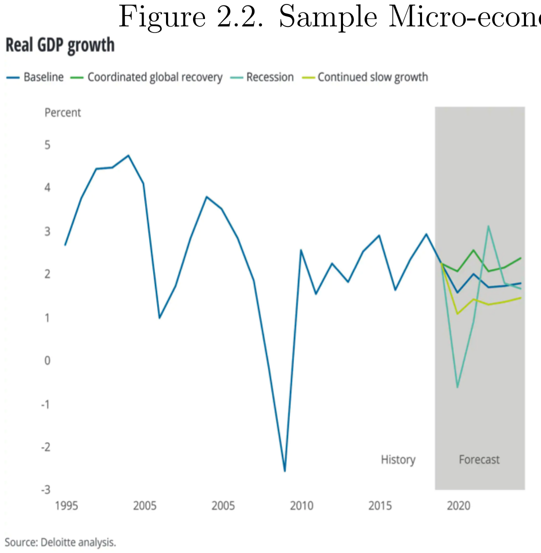
For bank stress-test analytics, economic conditions play a crucial role in determining the preparedness of a financial banking system. Economic conditions have shown to impact banking performance as turbulent circumstances directly may cause disruptions in typical operations, such as growth in loan default, loss of market value in financial assets, and decrease in demand for financial products & services. Economic adversity could impact banks implicitly or explicitly due to the systemic nature of financial banking systems. Counter-party investments & services expose banks to higher risks during economic turmoil, even when conservative strategies are employed.

Historically, financial crises' have distinguishable imprints in economic profiles. Thus, bank stress-test exercises aim to understand banking performance during simulated economic adversity. Regulators provide expected and hypothetical economic circumstances to depict adversity on banks, who must show the sufficiency of their risk management practices, leadership, and capital distribution plans to withstand potential losses. Figure 2.1 illustrates unemployment rate in different economic adversity scenarios projected in past CCAR stress-test exercises (Hughes & Poi, 2016; Reserve, 2015). Thus, enhancing the ability to forecast or generate economic conditions expected to occur or provide adversity is of great benefit to bank stress-test analytics. However, regulators typically focus on macroeconomic indicators every quarter for the last four decades. Given the complexities of how the economy may operate, investigating relationships among economic conditions co-dependently rather than independently at both macroeconomic & microeconomic levels may provide unique insights to improve estimations. Addressing practical challenges typically faced with analyzing economic data, such as sparsity, volume, availability, and immediacy, may help improve modeling efforts of economic conditions.

This chapter provides further depth on the economic condition estimation (ECE) model. The ECE incorporates multiple exogenous factors beyond what regulators typically utilize to represent the economic climate to learn non-linear latent relationships among relevant macroeconomic, microeconomic, and relevant financial market indicators. A bespoke multi-modal conditional & variational autoencoder (MCVAE) developed for robust macroeconomic & microeconomic conditions estimation. The MCVAE addresses practical challenges of data availability, integrity, and identification of dependent relationships among the economic conditions (M. Wu & Goodman, 2018).

2.2 Problem Formulation

Figure 2.2. Sample Micro-economic Forecasts by Deloitte & Touche



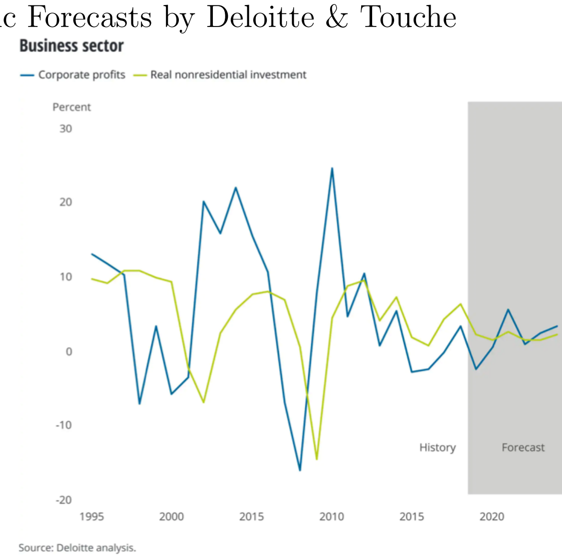


Figure 2.3. Real GDP growth rate.

Figure 2.4. Business sector growth rate.

Economic Conditions Estimation:

Bank stress-testing relies on each economic condition's baseline expectations and adverse estimations to provide scenarios. Sampled scenario circumstances resembling a financial crisis, such as increasing unemployment rate, decreasing GDP, increasing loan default rates or decreasing market value, appraise if banking strategies are capable of withstanding direct impacts to their business and stakeholders.

The hypothetical scenarios leveraged for bank stress-test exercises commonly focus on macroeconomic conditions as exogenous factors to banking performance. Thus, the task of developing a useful and practical technique to estimate economic conditions is an essential aspect of bank stress-testing, macro-prudential regulations, and risk management for stakeholders of the financial economy. Formalizing the task of economic conditions estimation may be simplified to a density estimation task, as depicted in figure 2.5 and equation 2.1.

$$\Pr(E\hat{C}O_{mod}|ECO_{mod}) \tag{2.1}$$

Where a given a set of representative economic conditions ECO_{mod} is defined in equation 1.1. The task becomes estimating the measures of potential future economic conditions by learning the probability distribution among all the input exogenous modalities.

Once probability distributions are acquired, samples can come from probability regions that befit the task, such as sampling from a probability region that may indicate a high likelihood to occur or from an area that has a very low probability to occur. 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 and sampling of economic conditions from different probability densities to attain potentially dire but possible circumstances.

2.2.1 Challenges

To accomplish the task, as described in equation 2.1, with the features, described in equation 1.1, practical challenges presented by the historical and hypothetical economic data provided by regulators or data vendors need consideration.

1. Learning independent probability distributions of economic conditions may not provide insight into how economic conditions may influence one another. Rather, they only provide an assessment of likelihood in a "vacuum." Co-linear or confounding factors among a set of economic conditions may exist. However, traditional techniques do not focus on these aspects as beneficial to forecasting.

2. When utilizing historical economic conditions data with a long time horizon, typically, the focus is on measures on the macroeconomic level as they have been vital indicators for nations to ascertain their growth. However, to consider additional economic conditions beyond those provided at macroeconomic levels and to introduce other relevant factors that may help depict the economic conditions, as illustrated in figures 2.3 & 2.4, to provide resolution, issues with data sparsity and co-occurring observations may be encountered. This challenge is familiar with financial economy data, since historical reporting of macroeconomic & microeconomic features may have different measurement practices and re-

porting intervals. Introducing techniques that consider indicators beyond those typically used to illuminate economic circumstances also need to have the ability to handle missing data throughout the time horizon of the data, which may ultimately hinder the overall performance or appropriateness of the economic conditions estimation task.

3. Economic conditions estimations for the next period typically requires the availability and access to all of the necessary input economic conditions variables to provide the best probability distribution—samples of the most likely conditions for the upcoming period. However, attaining the economic conditions input data may prove difficult. The release schedule of the data itself may differ from data vendors, economists, or regulators, hindering the utility of models that require all the inputs to be available in situations where the forecast may be needed promptly. Waiting for all the input data to be available before conducting an economic condition estimation may prove detrimental for those situations which may have limited access to the data points required on an immediate basis.

4. The overall volume of the economic conditions data available may not be sufficient for modeling purposes due to the time-horizon of which the economic conditions data may be historically available. Traditionally, macroeconomic and microeconomic data is available for the past few decades and recorded at quarterly intervals. A few hundred instances of economic conditions data may not be sufficient to generalize or represent the probability distributions of the different combinations of conditions.

Developing an approach or technique that could potentially address these challenges while achieving the task defined in equation 2.1 may help provide improvements to the effectiveness and utility of forecasts.

2.3 Related Work

Event-based market reactions in representative asset classes using the event study approach dominates the academic literature that covers the area of leveraging economic conditions to understand firm-level aspects. In works such as (Neretina, Sahin, & De Haan, 2015; Morgan, Peristiani, & Savino, 2014; Ellahie, 2013; Candelon & Sy, 2015; Petrella & Resti, 2013; Bischof & Daske, 2013; Nijskens & Wagner, 2011; Flannery, Hirtle, & Kovner, 2017) a heavy reliance on the usage of financial event-study methodology around stress-test exercise disclosures & events are analyzed to understand the impact that they may have towards financial markets in terms of abnormal returns. All of these studies rely on the economic conditions that have historically occurred rather than hypothetical scenarios administered within a bank stress-test exercise to understand the effects the exercise may have on subsets of the financial economy. Much of the bank stress-test literature that revolves around abnormal market returns on regulatory events provide insight for investment opportunities or market impact analysis.

Literature that focuses specifically on the usage of economic scenarios for stress-testing or the topic of generating feasible economic conditions for risk analysis centers around statistical methodologies that select scenarios utilizing probabilistic techniques to indicate plausibility of the conditions. Statistical approaches to economic conditions estimation or scenario generation, as seen in (Jamshidian & Zhu, 1996), where a methodology to ascertain a limited number of economic scenarios based on the discretization of multi-variate distributions of market variables while considering market and credit risks for financial risk management. Also, in the works of (M. Flood & Korenko, 2012; M. D. Flood & Korenko, 2015), techniques are explicitly discussed for financial "shock" scenarios. A grid search approach to sampling of multidimensional probability distributions about stress-test scenario selection. The severity of each scenario sampled from plausible probabilistic regions.

Research that studies the direct impacts that economic conditions may have on banking performance is also an area that relates to bank stress-tests. Sometimes referred to as reverse stress-testing, the aim is to understand if the appropriate economic conditions to inspect the impacts that economic conditions have on banks properly. In the work of (Guerrieri & Welch, 2012), testing of macro variables to assess their appropriateness. Focus on factor-based influence on loan loss rates, revenues, and capital measures by using equal-weighted average forecasting techniques that are bench-marked against a random-walk process to determine effectiveness. The work of (Glasserman, Kang, & Kang, 2015) discusses the selection of stress-test scenarios. They are obtained by first understanding the reverse effects of economic variables that cause the most significant industrial losses. Then, non-parametric empirical likelihood estimations, which ultimately help the authors model expected marginal shortfalls conditional on market stress or stress scenarios. Additionally, the work in (Grundke & Pliszka, 2018) focuses on reverse stress-testing to identify a scenario. Identifying firm-level stress points by defining the task as an inversion problem. The approach leverages principal components analysis to reduce the feature dimensions and a linear factor model to estimate the maximum likelihood for the term structure of risk-free interest rates and asset returns as an element to determine probabilities of default within corresponding selected stressed scenarios.

At the time of this writing, the work of (Malik, 2018) was the first to attempt stress-test scenario selection using machine learning techniques to help sample economic conditions scenarios using neural networks that produce robust distributions and samples for scenario generation. The work relies on machine learning research of conditional generative adversarial models (Mirza & Osindero, 2014) and generative adversarial models (Goodfellow et al., 2014). The proposed model framework learns and generates economic conditions estimation function, which can leverage a conditional modality to produce a plausible estimation of economic conditions based on historical context.

Otherwise, the research literature on generating foreseeable macroeconomic scenarios focuses on utilizing structural approaches, which relate to conceptual domain knowledge or dynamics related to relationships economic factors may have with one another. Structural methodologies "The Global Economic Model" (GEM) (Bayoumi et al., 2004; Pesaran, Schuermann, & Weiner, 2004; Lalonde & Muir, 2007) and "Global Auto-regressive Model" (GVAR) model interactive relationships between countries by using country-specific macroeconomic variables as a function of corresponding countries' indicators weighted on level of trade activity (Dees, Mauro, Pesaran, & Smith, 2007).

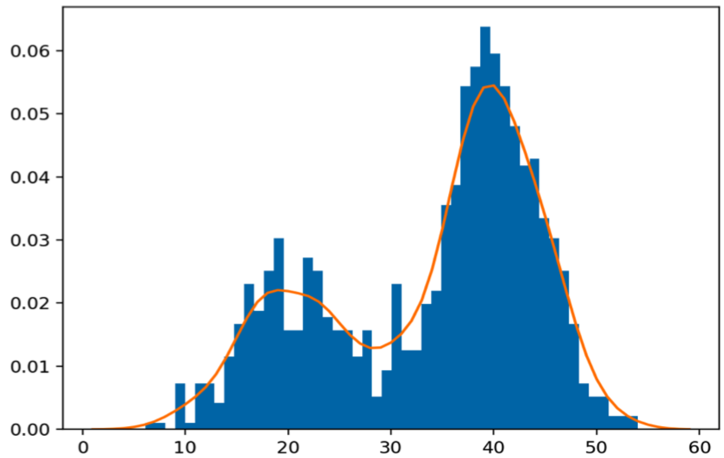


Figure 2.5. Example of Density Estimation.

2.4 Methodology

For the economic conditions estimation task, the focus is given on the generative model category of deep learning techniques to find a practical solution when considering characteristic limitations of historical economic conditions data. As depicted in figure 2.5, the task ultimately becomes a density estimation (Goodfellow, Bengio, & Courville, 2016), which needs to learn the approximated probability distributions through estimating the densities of joint distributions of the economic conditions.

To address the practical data-related challenges facing the previously defined economic conditions estimation task, depicted in equation 2.1, the methodology employed in a proposed model framework must be robust and dynamic. Advances in neural network utility and deep learning literature for various domains have shown that generative modeling processes can help learn probability distributions to then generate data samples that can potentially be indistinguishable from actual data distributions

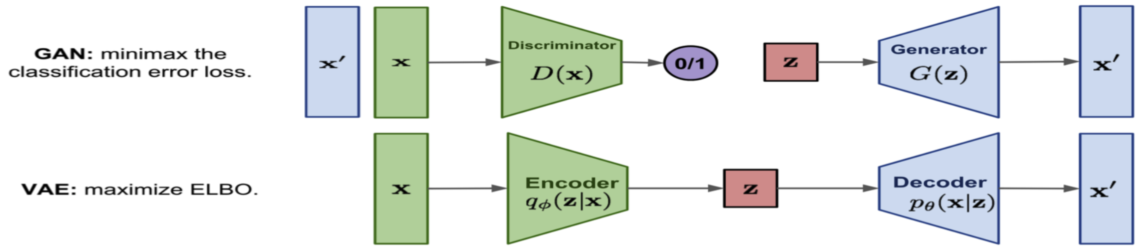


Figure 2.6. Basic Generative Model Frameworks

(Goodfellow et al., 2014; Mirza & Osindero, 2014).

In this section, further discussions about the background of generative modeling in the deep learning literature, state of the art models that may be relevant to the problem task, and unique model features that allow for model versatility, will be presented.

2.4.1 Generative Models

Generative models are a category of modeling techniques that exist under the "Unsupervised Learning" umbrella of data mining & machine learning approaches. Unsupervised learning is a particular scenario where the data for training a learner model does not have class labels. Therefore, the general objective is to ascertain a modeling function that can help describe the hidden patterns or structures that may exist within the training data (Radford, Metz, & Chintala, 2015).

Traditionally, unsupervised learning techniques centered around (1) estimating the probability density functions, $p(x)$, of a random variable, x, given a set of observations, $X_1, X_2, ...$, also known as density estimation for continuous probabilities. (2) Grouping a set of input instances, $X_1, X_2, ...$, according to distance metrics that ultimately assign group membership of a discrete set of labels to each object based on their similarity to one another, also known as clustering. (3) Transforming the raw input data into a representation of continuous vectors that can then be leveraged effectively for machine learning tasks, also referred to as representation learning. (4) Also, reducing the number of random variables for a task by acquiring variables that predominantly capture the variability of the data into a lower-dimensional representation, otherwise known as dimensional reduction (Goodfellow et al., 2016). Newer techniques have incorporated sophisticated methods to effectively and efficiently learn about the data structure of unlabeled data. (1) Introducing neural networks that can make a good representation of input data by minimizing reconstruction error as an objective function by identifying non-linear aspects in the data to enhance the representation, also known as auto-encoder (Kingma & Welling, 2013). Alternatively, (2) estimating the probability density function of a random variable by employing non-parametric approaches that consider data points with more neighbor samples as likely to be higher in density, otherwise known as kernel density estimation (Goodfellow et al., 2016).

Recently, however, with the evolution of the research literature in auto-encoders, kernel density estimation, neural networks, and deep learning, the research on generative models have emerged to be very popular. The task of generative models focuses on creating new samples from a learned probability distribution that mimic the same structural distributions that exist in a training data-set to provide the ability to sample new data with variations. However, challenges exist in acquiring the true data representation of the training set through implicit or explicit means that may not be feasible. Thus the development of a modeled distribution that is similar to the true data distribution is warranted. The abilities of neural networks are exploited by tasking their objective function to learn the ideal approximation parameters necessary to model a distribution after the true data distribution (Doersch, 2016). Common methods of generative models in the recent and relevant academic literature are Variational Autoencoders (VAE). VAEs aim to maximize the lower bound of the data log-likelihood to generate an ideal model data distribution from the training set (Doersch, 2016). Generative Adversarial Networks (GAN), which employs a generator network that tries to learn how to create potential data samples from random white noise. Additionally, a discriminator network assesses for potential validity of its understanding of the true data distribution from the training set (Goodfellow et al., 2014). Reference the general framework of both VAE and GAN in figure 2.6. The following subsections further describe the VAE and GAN generative model theory.

Variational Auto Encoder

Conceptually, auto-encoders in unsupervised learning encode input data onto a smaller dimensional representation by extracting latent factors that describe the input data distribution. Typically, the encoded input data only corresponds to its respective decoder to produce samples that resemble the input without variation. However, to generate data samples with some variability to approximate potential samples that are likely to exist in the training data, the corresponding probability distribution is required to be learned (Doersch, 2016). Variational auto-encoders learn sophisticated data distributions by using neural networks for an unsupervised learning task. Technically, Bayesian inference aspects of learning the underlying probability distribution of the training data to create a probabilistic graphical model that can sample new data points from the approximated distribution. To this end, a latent representation of the training data to be learned. The latent representation will consist of latent variables, which are implicit elements derived from a mathematical model, which is assumed to be fundamental to the training data distribution. The inferred latent variables, z, provide pertinent underlying information for model approximation through a probability distribution, $P(z)$. A Gaussian distribution as a prior, to learn, $P(z)$ to efficiently sample data during the inference period. The objective becomes modeling the data with parameter weights that maximize the likelihood of being from training data, X, under the assumption that the latent representation helps generate the approximated data, $x(x \in X)$. Therefore the non-linear mapping of the latent variables, z, to x can be achieved using a deterministic function, $f(z)$, with parameters Θ (Kingma & Welling, 2013). Thus, the maximization of the probability of each data point in X occurring in the model distribution, as shown in equation 2.2.

$$P_\theta(X) = \int P_\theta(X, z)dz = \int P_\theta(X|z)P_\theta(z)dz \tag{2.2}$$

Where $f(z) = P_\theta(X|z)$. This maximum likelihood estimation allows for the model to generate training samples from inferred latent variables with variations to produce approximated data that follows the data structure of the training set (Kingma & Welling, 2013; Sohn, Lee, & Yan, 2015). To practically accomplish this, neural networks are tasked with computing z, assuming that the derived latent representation stems from a normal distribution of z for efficient sampling during the inference period, which allows for the projection of any kind of distribution that can return to its original latent representation. Efficient methods to approximate the maximization of $P_\theta(X)$ due to the intractability of equation 2.2 over all of the dimensions of z are necessary. Thus, $P(z|X)$ is obtained through a technique from Bayesian statistics to solve an optimization task called variational inference. Essentially, $P(z|X)$ is modeled using a simple distribution, $Q(z|X)$. The minimization of the difference between the two distributions though the Kullback-Leibler (KL) divergence technique aligns the model as close as possible to the true distribution (Doersch, 2016).

The objective function of a variational auto-encoder can be denoted as:

$$\log P(X) - D_{KL}[Q(z|X)||(P(z|X)] = E[\log P(X|z] - D_{KL}[Q(z|X)||(P(z)] \quad (2.3)$$

Where $Q(z|X)$ is the encoder network, z is the encoded representation of $x(x \in X)$, $P(X|z)$ is the decoder network and the KL-divergence metric measures the difference between the true training data distribution and model distribution can be depicted as:

$$D_{KL}[N(\mu(X), \Sigma(X)||(N(0|1)] = \frac{1}{2}\sum_k (\exp(\Sigma(X))) + \mu^2 - 1 - \Sigma(X)) \quad (2.4)$$

The $P(X) - D_{KL}[Q(z|X)||(P(z|X)]$ from equation 2.3 represents the optimized objective function which will yield two terms, reconstruction loss from input data to the output sample and the KL-divergence metric. Additional optimization techniques and reparameterization tricks are applied to incorporate neural network backpropagation to maximize the lower variational bound of the reconstruction error and divergence terms to align the modeled approximation closely to the true distribution of the training data (Kingma & Welling, 2013).

Generative Adversarial Network

In contrast to VAEs, the Generative Adversarial Network (GAN) does not directly employ density estimation techniques. GANs rely on a game theory-based approach that seeks to identify the "Nash equilibrium" between a Generator neural network and a Discriminator neural network (Goodfellow et al., 2014). The technique samples from a simple random distribution, such as Gaussian, in the Generator network, which will eventually learn to generate samples that closely align with samples from the data distribution of the training data by leveraging neural networks for approximation and the Discriminator network for validation. Technically, implementing adversarial training methods for the two networks. The generator model, G, will learn the training data distribution, while the discriminator model, D, will assess the probabilistic likelihood that a sample originated from the training data distribution. Essentially, the task of G is to learn to generate data samples that would potentially exist in the training data, while the task of D is to determine if the sample could have originated from the training data based on what it knows about the training data distribution (Malik, 2018). Both networks improve over time to be better at their respective tasks of generating real-like data samples to convince the D of the genuineness and discriminating data samples from G as not actual samples from the training data distribution. Ultimately, a prior is initially defined on the input noise variables, $P(z)$, to have the G map to this data distribution by utilizing sophisticated differential functions with parameter weights, Θ_g. Also, the D first accepts data samples as an input, x, then leverages a separate differentiation capable function with parameter weights,Θ_d, to output a singular scalar value that indicates the probability that x can be from the training data distribution, $P_{data}(x)$ (Grosse, Ancha, & Roy, 2016). Thus the objective function of the GAN model architecture can be defined as:

$$\min_{G} \max_{D} V(D, G) = \mathbb{E}_{x \ p_X(x)}[\log D(x)] + \mathbb{E}_{z \ p_z(z)}[log(1 - D(G(z)))] \tag{2.5}$$

Where the $D(x)$ will output the value of 1 if the input data sample comes from the training data distribution to maximize the function in equation 2.5 concerning D. Otherwise, if the data sample was generated through $G(z)$, then $D(G(z))$ will output 1 to minimize the objective function concerning G. Ultimately G should generate realistic data samples, x, to trick D into believing that the sample x is a sample from the training data distribution. Gradient ascent solves for the maximization of the D parameters. In contrast, the minimization of the G parameters uses gradient descent. However, maximization of $E[\log(D(G(z))]$ is incorporated rather than minimization of $E[\log(1 - D(G(z))]$ for desired gradient size behaviors concerning G performance. Thus, the training process of G and D will apply stochastic gradient descent on both models simultaneously. The optimization process occurs after every k steps for D and then alternates to one step for G in a repeatable manner until D is unable to determine if input samples are from G or the training data, $D(x, \theta_d) = .5$ (Goodfellow et al., 2014).

Several variations of GANs are currently in existence, however conditional GANs (CGAN) has emerged as one of the more popular techniques due to their effectiveness in many applications. CGANs include an additional conditionality aspect with the data sampled from random noise. This additional element of information helps direct the G and D towards data structures that may be more relevant to the task at hand by statistically conditioning the models to consider the relevant conditionality vector (Mirza & Osindero, 2014).

One of the major benefits of utilizing GANs is the fact that they can operate with relatively small amounts of training data. However, the challenges of using GANs properly include the complexity in configuring their hyper-parameters for the neural network as well as the optimization convergence difficulties (Grosse, Ghahramani, & Adams, 2015).

2.4.2 Multimodal Conditional Variational Auto Encoder

Current literature in the generative models' research area, such as (Kingma & Welling, 2013; Sohn et al., 2015; M. Wu & Goodman, 2018), have proposed variant approaches to address the "modality estimation" problem. Generative models can learn an approximated model distribution based on the respective input feature's training data distribution to be able to sample highly realistic data samples with some variation for estimation of the input modalities. However, the techniques discussed in the literature do not address the "conditional multi-modality estimation" problem. This problem seeks to explore the relationship among the distributions of several target modalities conditioned on one specific modality, referred to as the conditional modality, to then be able to generate realistic data samples using the mentioned conditional modality. The benefits of economic conditions estimation applications are evident when the multi-modality and conditionality aspects are considered for modality estimation tasks since they can address practical challenges with data volume, availability,

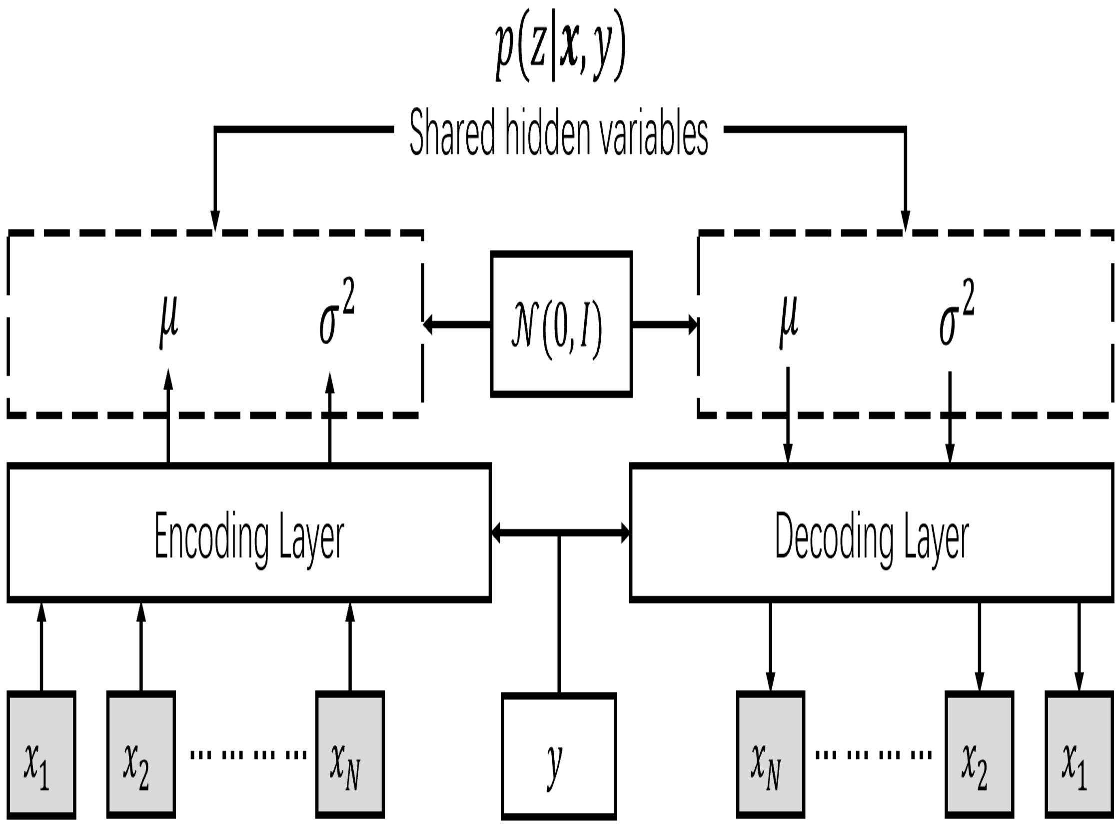


Figure 2.7. MCVAE model framework.

sparsity, and co-linearity.

Discussed in this section, the proposal of a bespoke Multi-modal and Conditional Variational Auto-encoder (MCVAE) model that addresses the ”conditional multi-modality estimation” problem form the research literature. The proposed framework also serves as a solution for the application of estimating desired economic conditions variables of different exogenous factors given one highly available economic variable.

Overview

Technically, the proposed MCVAE consists of several variational auto-encoder components for each target modality and applies a conditional mechanism by importing a conditional modality for better estimation of the target multi-aspects modality. Extended from the basic methodology of VAE (Kingma & Welling, 2013), along with inspiration from MVAE(M. Wu & Goodman, 2018) and CVAE (Sohn et al., 2015), the goal of training MCVAE is to maximize the evidence lower bound (ELBO) for generating the latent variable $\mathbf{z}$ given multi-modality set $\mathbf{x}$ and conditional modality y, which is defined via an inference network $q_\phi(\mathbf{z}|\mathbf{x}, y)$ as follows:

$$
\begin{aligned}
ELBO(\mathbf{x}, y) \triangleq & \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x},y)}[\lambda \log p_\theta(\mathbf{x}|y, \mathbf{z})] \\
& - \beta KL[q_\phi(\mathbf{z}|\mathbf{x}, y), p_\theta(\mathbf{z}|y)],
\end{aligned}
\tag{2.6}
$$

Where $p_\theta(\mathbf{z}|y)$ is the conditional network and $p_\theta(\mathbf{x}|y, \mathbf{z})$ represents the generation network. Furthermore, by extending the derivation in (Sohn et al., 2015) to a multi-modality problem, the empirical lower bound becomes:

$$\begin{aligned}\mathcal{L}(\mathbf{x}, y; \theta, \phi) = \sum_{x_i \in \mathbf{x}} &-KL(q_\phi(\mathbf{z}|x_i, y)) \parallel p_\theta(x_i|y, \mathbf{z}) \\ &+ \frac{1}{L} \sum_{l=1}^{L} \sum_{x_i \in \mathbf{x}} \log p_\theta(x_i|y, \mathbf{z}^{(l)}),\end{aligned} \tag{2.7}$$

Where x_i is the i^{th} modality in $\mathbf{x}$, and L is the number of samples that latent variable $\mathbf{z}^{(l)} = g_\phi(\mathbf{x}, y, \epsilon^{(l)}), \epsilon^{(l)} \sim \mathcal{N}(0, \mathbf{I})$.

Multi-modal Setting

When forecasting based on multiple modalities for economic conditions estimation, the modalities refer to the economic conditions to be estimated. Given that the task will likely have multiple conditions to be estimated, as they are assumed to depict the economic conditions, the assumption that N exogenous factors or modalities, $x_1, x_2, ...x_N$, are conditionally independent while recognizing the latent factors in common, z. Thus, the generative model will following the form:

$$p_\theta(x_1, x_2, ...x_N, z = p(z)p_\theta(x_1|z)p_\theta(x_2|z)...p_\theta(x_N|z) \tag{2.8}$$

Which ignores unobserved modalities when determining marginal likelihood.

Model Procedure

As shown in figure 2.7, the proposed MCVAE model consists of two components, the encoding, and decoding networks. (1) The model first passes multiple modalities and a separate conditional modality through an encoding network to learn hidden representations of shared variables across the variables. (2) Given the conditional modality and any initial variables sampled from Gaussian distribution, the decoding network can finally produce the estimation of target modalities. (3) In this problem,

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

return estimation $\hat{\mathbf{x}}$, parameters (θ, ϕ)

the target input modalities $\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$ and conditional input y are different characteristics of macroeconomic and microeconomic conditions. (4) The goal of MCVAE is to generate or simulate the overall economic conditions when given only one specific type of economic condition. (5) Algorithm 1 demonstrates the training procedure of the proposed MCVAE , which is optimized by the popular Stochastic Optimization method Adam (Kingma & Ba, 2014).

Joint Approximation Posterior

The ability to efficiently approximate the joint representation of the multiple modalities is crucial aspect that is considered in MCVAE through the contributions in (M. Wu & Goodman, 2018). To accomplish this task, the inference networks required are 2^N or $q(z|X)\forall X \subseteq \{x_1, x_2, ..x_N\}$, for each subset of modalities. First, the need to overcome the computational issue of training the necessary inference networks for each modality is addressed by defining an optimal inference network, $q(z|x_1, x_2, ...x_N)$, which can determine the relationship between the joint and singular modalities by relying on the true posterior, $p(z|x_1, .., x_N)$, under the assumption that conditional independence the relationship can be defined as the following:

$$p(z|x_1, .., x_N) \propto \frac{\prod_{t=1}^{N} p(z|x_t)}{\prod_{t=1}^{N} p(z)} \approx \frac{\prod_{t=1}^{N} \hat{q}(z|x_t)p(z)}{\prod_{t=1}^{N} p(z)} = p(z) \prod_{t=1}^{N} p(z)\hat{q}(z|x_t) \tag{2.9}$$

Also known as the "product of experts" (PoE). The reduced quotient term and the true posteriors for each factor, $p(z|x_t)$ in the corresponding variational component, $q(z|x_t)$, produce the approximated distribution of the joint-posterior. This approach allows for an efficient and consistent solution to the task. Furthermore, the product of Gaussian experts helps to reach a tractable solution (M. Wu & Goodman, 2018).

The computational complexity becomes 2^N multi-modal inference networks, which are required for the MCVAE to effectively and efficiently learn the joint approximations of the input modalities (M. Wu & Goodman, 2018).

This feature of the model allows for multiple exogenous economic conditions with the consideration that they may be dependent on each other rather than independent. Thus addressing the need to learn the representation of their joint distribution efficiently. This formulation also accounts for potential conditional modalities that can learn the latent distributions and generate the target modality estimates without the requirement of input modalities for estimation (M. Wu & Goodman, 2018).

Sub-Sampling paradigm

To address circumstances where the training data might not have missing modalities or observations, but the testing data might while ensuring the consideration of the relationships between modalities, the transfer of a training scheme from MVAE (M. Wu & Goodman, 2018) into the MCVAE model is important. The ELBO term for both whole and partial observations is split completely into partial combinations of modality sets. Reduction of the computational complexity is possible due to consideration of all the partial combinations and sub-samples of ELBO terms at each optimization step gradient. Technically the sequence of sampling is (1) the ELBO using the product of all N Gaussians, (2) all ELBO terms using a single modality, and (3) k ELBO terms using k randomly chosen subsets, X_k (M. Wu & Goodman, 2018). The objective function of the sub-sampling scheme can be generalized as the following:

$$ELBO(x_1, ..., x_N) + \sum_{i=1}^{N} ELBO(X_i) + \sum_{j=1}^{k} ELBO(X_j) \tag{2.10}$$

This feature was originally discussed in MVAE literature (M. Wu & Goodman, 2018) and transfers into the MCVAE model. The model feature provides the ability to generalize for weakly supervised learning problem settings and samples from partial data by ignoring modalities. This modeling feature is significant for addressing the data sparsity and modality co-occurrence challenges that typically afflict economic conditions data from heterogeneous sources.

2.5 Experiment

Table 2.1. MCVAE Prediction Experiment Setup

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

This section demonstrates the effectiveness of the generative models VAE, GAN, CGAN, and the proposed MCVAE have on an economic condition estimation task related to bank stress-testing analytics. Drawing from the machine learning, computer science, and data mining literature to formulate the task as a "modality estimation problem," allowing for the use of advanced analytical techniques from the research literature. Furthermore, experimentation to better understand the practicality and utility of each technique when dealing with economic conditions data and the practical challenges that are typically faced. Particularly considering multiple exogenous

Table 2.2. $Z_{DomesticMacro}$ (1976-2017)

Type	Variable	Mean	Std.	Min	Med.	Max
$Z_{macro,D}$	Real GDP grwth	2.8	3.0	-8.4	3.0	16.4
	Nom GDP grwth	6.0	3.9	-7.2	5.4	25.5
	Real disp. income grwth	2.8	3.3	-15.1	3.0	11.5
	Nom. disp. income grwth	6.1	4.0	-13.9	5.8	16.5
	Unemp. rate	6.3	1.5	3.9	6.0	10.7
	CPI infl. rate	3.6	3.2	-8.9	3.2	16.7
	3M Tsy. rate	4.5	3.5	0	4.8	15.1
	5Y Tsy. yield	5.9	3.4	0.7	5.9	15.0
	10Yr Tsy. yield	6.5	3.1	1.6	6.3	14.6
	BBB corp. yield	8.2	3.1	3.7	7.6	17.6
	Mortgage rate	8.1	3.3	3.4	7.7	17.8
	Prime rate	7.6	3.6	3.3	7.8	20.3
	House Price Idx	100.8	50.7	23.4	82.8	196.2
	CRE Price Idx	135.4	59.3	50.9	108.5	282.6
	DJ Total Stock Mkt Idx	10811	6115	2417	10806	27673
	VIX Idx	25.9	10.7	12.7	22.7	80.9

modalities over a historical time horizon that may not all be available at the same time.

Each generative modeling technique is assessed based on their performance in approximating the actual distribution of the training data. They are effectively addressing the lack of data volume in economic conditions data typically reported at quarterly intervals. The validation of the performance compared to baseline generative models can also provide a thorough understanding of the effectiveness of the models on an application using real-life historical data.

2.5.1 Data Description

Data is from multiple data repositories, including Wharton Research Data Services (WRDS, 2019), Global Financial Data (GFD), the U.S. Federal Reserve Board of Governors (Reserve, 2015), St. Louis Federal Reserve (STLFed), and Chicago Federal Reserve (ChicagoFed). Also, several interest rates, sector indices, government bonds, and market indices spanning the past three decades from WRDS are accessed. Commodities index prices for the past thirty years from GFD. Macroeconomic and financial conditions indices from the U.S. Federal Reserve System's network of regional satellite authorities, STLFed, ChicagoFed, FRB. Tables 2.2 and 2.3 provide a data summary on the macroeconomic variables that are provided by U.S. regulators for bank-stress-test analysis purposes through their CCAR exercise (Reserve, 2015).

Tables 2.4, 2.5, and 2.6 provide data summaries on a perspective of the microeconomic conditions through measurements of more specific asset classes, currency rates, indices, and government interest rates respectively. These conditions can depict the

Table 2.3. $Z_{InternationalMacro}$ (1976-2017)

Type	**Variable**	**Mean**	**Std.**	**Min**	**Med.**	**Max**
	Euro real GDP grwth	1.9	2.3	-11.4	2.1	7.5
	Euro infl.	1.9	1.3	-1.4	2.1	5.7
	Euro bilat. ex.rate	1.2	0.7	.08	1.2	1.5
	Dev. Asia real GDP growth	7.4	2.8	-1.9	6.9	15.1
	Dev. Asia infl.	2.8	2.0	-3.00	2.7	8.2
$Z_{macro,I}$	Dev. Asia bilat. exch.rate	95.2	7.0	85.3	93.85	110.2
	Japan real GDP grwth	2.2	3.9	-17.8	2.3	12.0
	Japan infl.	1.4	2.7	-3.6	.6	10
	Japan bilat. exch.rate	2.3	57.0	77.0	119.9	299.6
	UK real GDP grwth	2.3	3.2	-8.5	2.6	18.9
	UK infl.	4.4	4.7	-1.2	3.0	32.1
	UK bilat. ex.rate	1.6	0.2	1.1	1.6	2.3

Table 2.4. Z_{micro} (1976-2017)

Type	**Variable**	**Mean**	**Std.**	**Med.**	**Max**
Commodities	$\text{Brent}_{D,Crude}$	42.7	31.8	29.4	143.9
	$\text{NYH}_{Q,Heat}$	112.5	80.2	80.9	388.9
	$\text{USG}_{Q,Heat}$	103.7	78.6	77.9	381.5
	$\text{WTC}_{Q,Crude}$	37.9	27.1	29.6	139.9
	$\text{NYGold}_{Q,\$Oz}$	560.2	384.3	393.2	1674.3
Commodities Idx, Q	$\text{S\&P}_{GSCI}$	284.5	159.9	207.7	862.8
	$\text{S\&P}_{GSCI,IndtrlM}$	196.8	104.5	157.4	479.4
	$\text{S\&P}_{GSCI,PrecM}$	756.9	520	525.3	2283.8
	Bloomberg	113.4	26.4	106.2	233.0
	Moody's	2181.2	1809.5	1211.7	7324.7
	Reuters	1875.0	434.5	1729.3	3241.8
Market Idx	$\text{S\&P1}_{D}$	489.3	250.4	555.7	1192.2
	$\text{S\&P5}_{D,Equal-Wtd}$	1434.1	893.0	1207.7	4113.4
	$\text{S\&P}_{Q,TotalMkt}$	1585.9	483.3	1397.2	2535.4
	$\text{S\&P}_{Lvl}$	10019.8	587.1	1057.0	2647.5
	$\text{S\&P}_{DRtrns}$	.009	.04	.01	.11
	$\text{S\&P}_{DRtrns,Eql-WtdwDiv.}$	.01	.04	.01	.01
	$\text{S\&P}_{DRtrns,Eql-WtdwoDiv.}$	.01	.04	.01	.18
	$\text{S\&P}_{DRtrns,Val-WtdwDiv.}$	.01	.04	.01	.12
	$\text{S\&P}_{DRtrns,Val-WtdwoDiv.}$	.009	.04	-.01	.11
US Tsy & Infl. Daily Rtrns	30D	.008	.007	.008	.03
	90D	.009	.008	.009	.04
	1Y	.014	.011	.010	.06
	2Y	.013	.016	.008	.085
	5Y	.016	.03	.017	.09
	7Y	.019	.037	.02	.11
	10Y	.019	.042	.02	.13
	20Y	.02	.058	.03	.22
	30Y	.021	.072	.039	.33
	Infl. Idx	.008	.009	.008	.04

behavior that may exist within the financial economy, which may impact or be influenced by co-linear exogenous factors or non-linear latent relationships that may exist among other economic conditions. Although there is better data resolution, to some degree, information loss during aggregation may provide hidden rudimentary insights to relationships among the modalities. Lastly, table 2.7 shows the data summary of sector-based indices for financial services, real estate, and the national financial conditions measures for credit, leverage, and non-financial leverage. Each of the modalities depicted plays a vital role in the previous financial crisis. The most recent financial crisis in 2008 stemmed from the subprime mortgage housing bubble. Therefore monitoring behaviors and patterns in the real estate industry may provide a potential signal of similar precursors. The financial sector facilitates the operations, services, and instruments for the financial economy itself. Therefore, the financial services industries' performance is paramount to understand the initial effects of a potential crisis for the national economy. Finally, the financial conditions indices are maintained by economists and regulatory authorities in response to previous crises'. They function to capture aspects of financial conditions that may be obscure to the public, such as shadow banking operations of banks (Kliesen et al., 2012; Brave & Kelly, 2017), which can help serve as a preliminary signal toward a potential economic downturn. Incorporating such modalities to determine potential non-linear patterns among them could provide for a more robust approximation of reasonable economic conditions estimates.

Table 2.5. Z_{micro} Currency Swaps (1976-2017)

Type	**VarName**	**Mean**	**Std.**	**Min**	**Med.**	**Max**
US FX	AL	1.3	.23	.80	1.3	2.0
	BZ	2.1	.77	.83	2.1	4.1
	CA	1.2	.16	.93	1.2	1.6
	CH	6.9	1.4	1.6	6.8	8.7
	DN	6.3	.91	4.6	5.7	11.2
	EC	0.84	09	0.70	0.83	1.4
	EU	.84	.13	.62	.81	1.2
	HK	7.7	.12	5.2	7.7	7.8
	IN	41.6	14.6	7.6	44	68.8
	JP	116.1	25.6	75.7	113.8	299.5
	KO	1027	206.6	667.2	1079.1	1960
	MA	3.2	.57	2.1	3.3	4.7
	MX	11.3	3.3	3.1	10.9	21.8
	NO	6.8	.99	4.7	6.6	9.5
	NZ	1.6	.30	.91	1.5	2.5
	SD	7.3	1.2	3.9	7.3	11
	SF	6.8	3.5	.66	6.7	16.8
	SI	1.5	.21	1.2	1.5	2.2
	SL	89.7	38.4	6.2	97.1	153.8
	SZ	1.3	.36	.73	1.2	2.6
	TA	30.5	2.8	24.5	30.8	40.3
	TH	32.9	6.3	20.6	32.5	56.1
	UK	1.6	.2	1.1	1.6	.88
	VZ	1625.1	13671	.17	.70	9.9
	GX	93.3	9.0	78.4	92.4	152.7
	$TWEXM$	88.3	10.3	68	88.4	139

Table 2.6. Z_{micro} Interest Rates (1976-2017)

Type	**Variable**	**Mean**	**Std.**	**Min**	**Med.**	**Max**
Govt. Bond Interest Rates	D_AH_M3	5.3	1.6	2.7	5.1	15.5
	D_AH_M6	5.4	1.6	2.8	5.1	15.6
	D_AH_Y1	5.4	1.3	3	5.3	8.7
	D_COMP_Y10P	7.4	1.3	4.8	7.2	14.2
	D_LTNOM_Y25P	5.1	.35	4.2	5.1	5.9
	D_TCMNOM_Y20	10.2	2.1	7.1	10.4	15.1
	$LTAVG\_Y10P$	1.5	.82	-.22	1.7	3.4
	TB_M3	3.1	2.6	-.02	3.	15.5
	TB_M6	3.2	2.6	.02	3.1	15.6
	TB_WK4	1.1	1.4	-.03	.29	5.1
	TB_Y1	3.4	2.8	.07	3.9	14.9
	$TCMII\_Y5$	.58	1.0	-1.6	.45	4.1
	$TCMII\_Y7$	.88	.98	-1.3	.82	4.2
	$TCMII\_Y10$	1.1	.91	-.87	1.2	3.1
	$TCMII\_Y20$	1.4	.78	-.16	1.4	3.3
	$TCMII\_Y30$	1.0	.43	.24	.98	2.2
	$TCMNOM\_M1$	1.1	1.4	0	.2	5.2
	$TCMNOM\_M3$	3	2.6	0	3	13.9
	$TCMNOM\_M6$	3.2	2.6	.02	3.2	14.2
	$TCMNOM\_Y1$	3.4	2.8	.08	3.4	16.9
	$TCMNOM\_Y2$	3.8	2.8	.16	3.9	16.7
	$TCMNOM\_Y3$	4	2.7	.28	4.2	16.3
	$TCMNOM\_Y5$	4.4	2.5	.56	4.4	16.1
	$TCMNOM\_Y7$	4.7	2.4	.91	4.6	15.7
	$TCMNOM\_Y10$	5	2.2	1.3	4.7	15.4
	$TCMNOM\_Y20$	4.7	1.5	1.6	4.8	8.3
	$TCMNOM\_Y30$	5.6	2.1	2.1	5.5	14.7

2.5.2 Dimension Normalization & Reduction

In this dissertation, normalization and principal components dimension reduction, as per equation 2.11 & 2.12, of each modality in ECO_{mod} is performed prior to estimation for the purposes of scaling numerical values and obtaining representative component variables.

Data Scaling

Min-max scaling techniques are used, as per (Pedregosa et al., 2011), to normalize the numerical values to a scale between -1 and 1. Scaling is important for the defined task, as the measurements for each modality discussed in the data description section may not use the same metrics, units, or scale. Thus, scaling or normalizing the data to its respective dimensional vector-based distribution with an upper and lower bound allows for meaningful data analysis. Typically min-max scaling can be depicted as:

$$X_{scaled} = \frac{X - Min(X)}{Max(X) - Min(X)} \tag{2.11}$$

where X represents a dimension vector within a particular modality, ECO_{mod}.

Dimension Reduction

Since several exogenous modalities depict the economic conditions, the number of dimensions to consider are significant. In the experiment setting, the consideration of the macroeconomic, microeconomic, and additional exogenous modalities creates over 130 dimensions. The need to reduce the number of dimensional variables by transforming them into the principal components of all the dimensions, which ultimately allows the experiment to capture the majority of variability among the modalities

Table 2.7. M_n (1976-2017)

Type	**Variable**	**Mean**	**Std.**	**Med.**	**Max**
$M_{FinclSectIdx}$	S&P$_{BNK.I}$	243.9	85.2	243.5	397.5
	S&P$_{DVF.I}$	430.6	165.6	453.4	759.4
	S&P$_{FIN.I}$	223.9	132.7	221.3	482.4
$M_{RelEstSectIdx}$	S&P$_{REA.I}$	132	37.7	131.2	198.5
	S&P$_{RES.I}$	108.3	20.9	103.5	144.4
	S&P$_{REI.I}$	146	37.4	151.7	200
	S&P$_{REM.I}$	78.5	28.2	81.1	122.6
$M_{FinclCondIdx}$	St.Louis Fincl Stress	-.09	.09	-.39	4.2
	Natl Fincl Conditions	-.10	.90	-.35	3.7
	Seasonally-Adjusted NFC	-.09	-1.39	-.39	4.2
	NFC-Credit	.08	.99	-.28	4
	NFC-Leverage	.06	1.0	.12	3.6
	NFC-Risk	-.12	.86	-.41	3.3
	NFC-Non-Fincl Leverage	.06	1.0	.12	2.6

Figure 2.8. PCA Correlation Heatmap

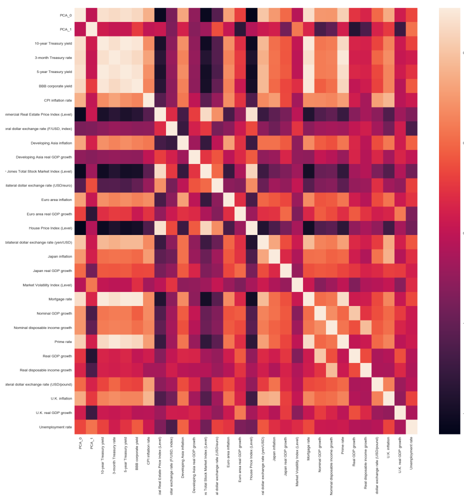


Figure 2.9. Z_{macroD}

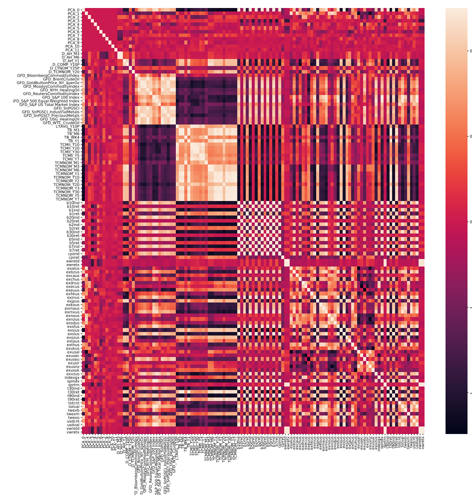


Figure 2.10. Z_{micro}

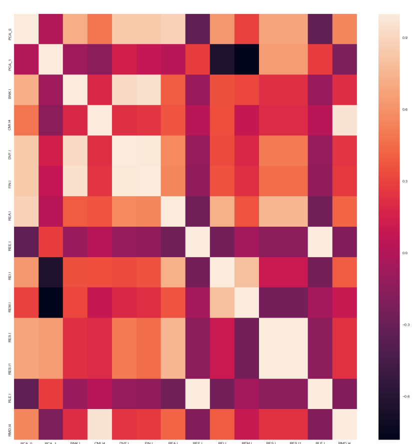


Figure 2.11. $M_{sectIDX}$

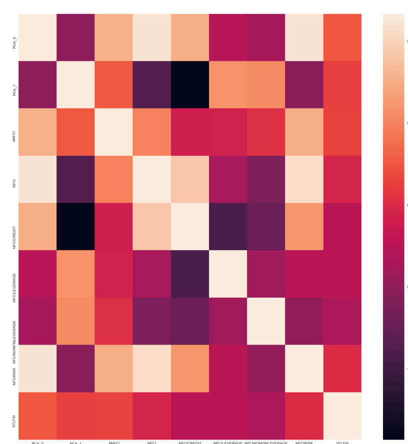


Figure 2.12. M_{NFCI}

while providing a feasible set of dimensions for consistency and efficiency.

Principle Component Analysis (PCA): Refers to an algorithmic technique that performs dimension reduction by finding the lower-dimensional projection that minimizes the reconstruction error and keeps the majority of principal information by maximizing the variability (Wold, Esbensen, & Geladi, 1987). This is achieved through the use of eigen decomposition, search for K-largest eigenvectors, and linear projection with a matrix composed of K eigenvectors (Goodfellow et al., 2016; Malik, 2018) For the purpose of this experiment, twelve principal components are acquired from the economic conditions modalities, as was done in (Malik, 2018), using the following formulation:

$$\begin{gathered}
\bar{w}_{j,n} = \frac{(w_{j,n} - \mu_i)}{\sigma_i} \quad \forall w_j \in ECO_{mod} \\
\hat{w}_{j,n} = [P_1, P_2, ...P_{12}] = [\bar{w}_{j,1}, \bar{w}_{j,2}, ...\bar{w}_{j,n}] * \underbrace{W'}_{\text{Projection Matrix}} \\
E\hat{C}O_{mod} = [\hat{w}_1, \hat{w}_2, ...\hat{w}_j] \quad \forall w_j \in ECO_{mod}
\end{gathered} \tag{2.12}$$

Upon exploratory analysis, the principal components adequately capture the linear relationship of the modality dimensions, which is apparent through a correlation heat-map of the raw modality dimensions and the principal components with particular focus on the first two principal components. Figure 2.9, 2.10, 2.11, and 2.11 and 2.12 show the first two PCA components (first two columns in heat-map) of the modalities and how they correlate to the original dimensions. Indications of strong positive or negative correlation through lighter or darker colors indicate linear relationships with the respective principal component. Overall, the principal components seem to capture the variability in the modality dimensions well enough to justify the use of the twelve components instead of the original dimensions.

Figures 2.14, 2.15, 2.16, and 2.17 show the bi-plot of the first two principal components of each modality, which technically explain the majority of the variability, as a function of time. Therefore the points in each plot present the respective year and quarter coordinates for the principal components. Thus, points that have many neighbors represent the quarter and year that economic conditions were most similar concerning the principal components for each modality. Conversely, points that have lower amounts of neighbors and are significantly distant from the other points indicate times that the economic conditions were abnormally different from previous periods, or it is an anomaly in history. The periods from late 2008 to mid-2009 were consistently outlier points on each modality, which supports the profile of crises. It is also apparent that not all crisis has the same economic condition profile, but may still be anomalous when compared to historical context. This exploratory analysis gives additional support to potential relationships among the normalized principal components that can be exploited by the generative models.

2.5.3 Economic Conditions Estimation

Experiment Setup. Modeling of $E\hat{C}O_{mod}$ consistent of experiments at both quarterly & yearly intervals that will consist of four separate experiments for each configuration, totaling twenty experiments. The purpose of each experimental configuration is to understand how well the generative modeling techniques can represent the data distribution for (1) a known crisis event (Experiment 1, e.g., recent financial crisis), (2) all of the available data (Experiment 2). Aligned economic conditions and bank-

Figure 2.13. PCA Biplot first two components as function of time

Figure 2.14. Z_{macro}

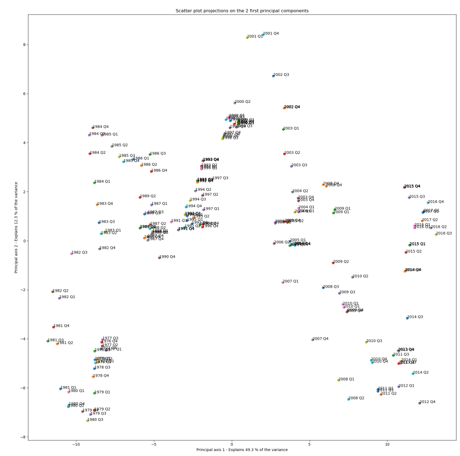


Figure 2.15. Z_{micro}

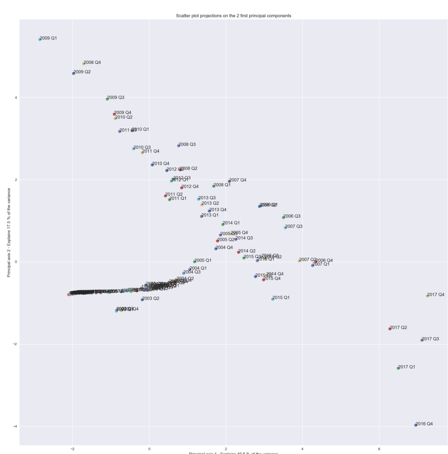


Figure 2.16. $M_{sectIDX}$

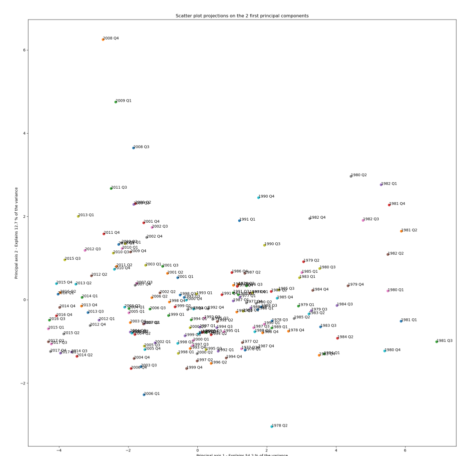


Figure 2.17. $Z_{macroDsubset}$

ing data are available (Experiment 3 & 4). Table 2.1 provides a summary of the data segmented for training and testing. For all experiments, $Z_{domestic}$ is used as the conditional modality, since the experiment seeks the relationships of the modalities as it relates to the regulator provided economic conditions, the effectiveness at discriminating the other modalities, and the high availability of historical & projected values.

Evaluation Metrics. For economic conditions estimation task, the root mean squared error (RMSE) evaluation metric is used to gauge how well the model can generate a data distribution to match the sampled data and to compare effectiveness across models. RMSE can be defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2} \tag{2.13}$$

For this task, 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. For generative models that do not support multi-modal joint approximation, the RMSE for each modality is attained by the average RMSE across all the modalities.

$$\frac{1}{n}\sum_{n=1}^{n} RMSE(e_1, e_2, ..., e_i) \forall \in E\hat{C}O_{mod}, e_i \neq e_{cond} \tag{2.14}$$

Moreover, for the generative models that do not consider conditional modalities, the $e_i \neq e_{cond}$ constraint is removed.

Additionally, the use Annealed Importance Sampling (AIS) with bidirectional Monte-Carlo, proposed in (Y. Wu, Burda, Salakhutdinov, & Grosse, 2016) and used in (Malik, 2018), for the plausibility validation of generative models by determining the

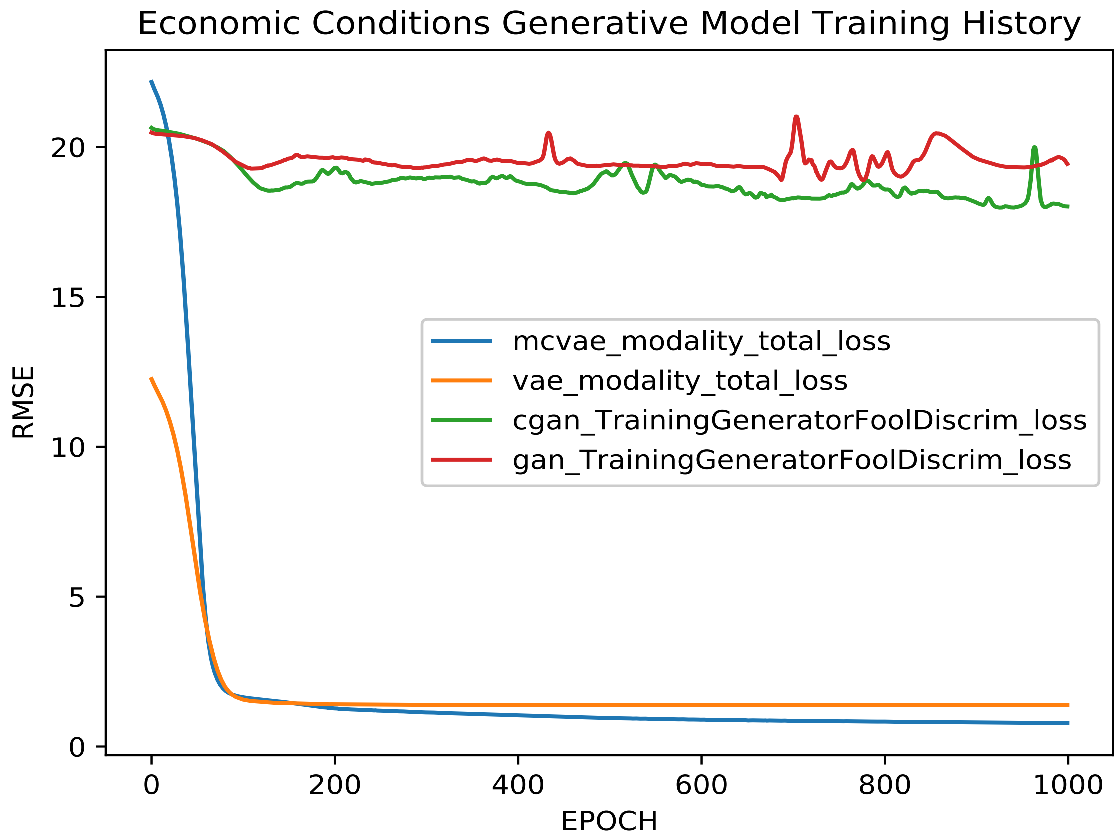


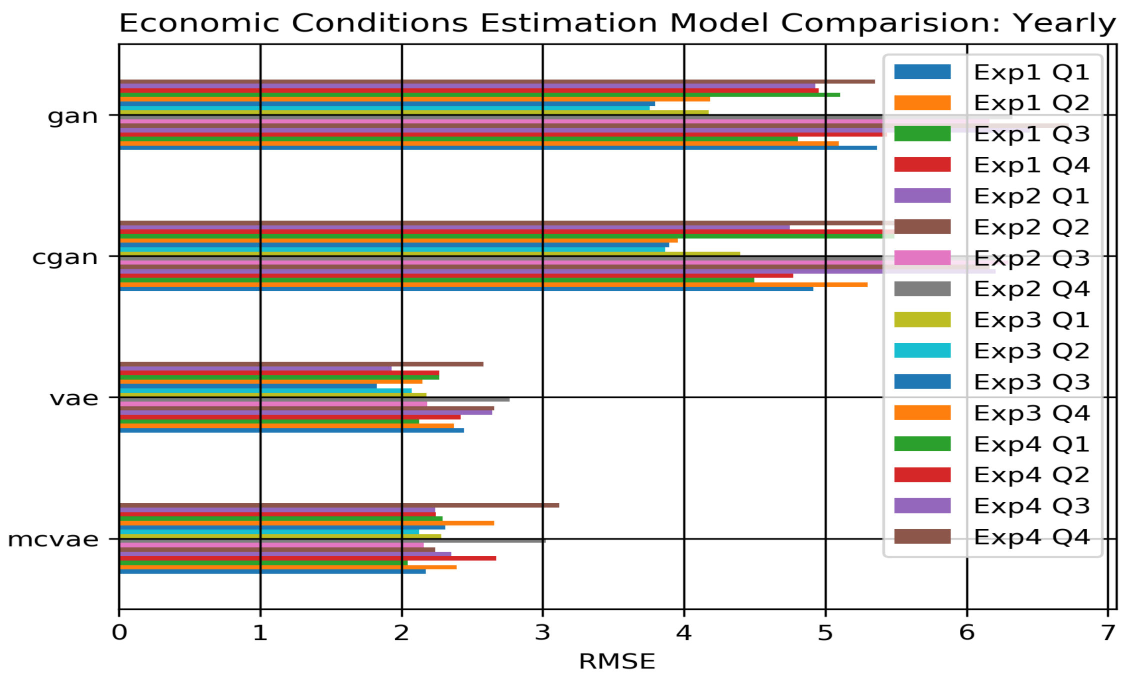
Figure 2.18. Training History of Generative Models for Exp.1

log likelihood (LLD) through calculating $P_\theta(z)$. This evaluation metric has recently gained popularity for evaluating generative models as it has shown to better depict the probabilistic potential of the model estimations.

Training Progress. Figure 2.18 shows the convergence of each generative model from Experiment 1 under time split conditions for all of the input modalities, $EC\hat{O}_{mod}$. Interestingly, the variational auto-encoder models' training history consistently outperformed that of the adversarial network models for all experiment settings.

Baseline Algorithm. The evaluation of the effectiveness and robustness on the economic conditions estimation task of the proposed model in comparison to the

Figure 2.19. Yearly Experiments (Q1,Q2,Q3,Q4)



baseline models, includes (1) Variational Auto-encoder (VAE)(Kingma & Welling, 2013),(2) Conditional Generative Adversarial Networks (CGAN) (Mirza & Osindero, 2014; Malik, 2018) and (3) Generative Adversarial Networks(GAN)(Goodfellow et al., 2014). The training of the proposed and baseline algorithms is on the same data-set. Also, all the models' respective performance uses identical validation sets. All the algorithms are implemented by Python deep learning library PyTorch (Paszke et al., 2017).

2.6 Conclusion and Discussion

Overall Performance: The performance of the different approaches is summarized in table 2.8. The proposed generative model, MCVAE, achieves a favorable RMSE for

Table 2.8. 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	-10.6	-15.5	-10.7	-19.0	1.53	4.81	1.92	5.74
	VAE	2.34	2.56	2.05	2.26	.004	.002	.001	.002	-8.3	-8.1	-8.4	-8.7	.98	.109	.48	.48
	CGAN	4.72	6.33	3.94	5.76	1.17	1.26	.89	1.08	-10.7	-10.3	-11.0	-10.5	1.51	1.84	1.61	1.59
	GAN	5.42	6.33	4.24	5.64	.86	.89	.75	1.04	-10.5	-10.2	-11.4	-10.5	1.37	1.70	1.68	1.79
	CGAN (Malik, 2018)	-	-	-	-	-	-	-	-	-17.4	-10.5	-	-	-	-	-	-
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	-8.0	-8.0	-8.4	-8.4	1.08	.14	.68	.57
	CGAN	4.91	6.20	4.39	5.48	.99	1.46	1.23	.99	-11.0	-10.1	-11.6	-10.9	1.87	1.65	2.14	1.62
	GAN	5.36	6.44	4.17	5.11	.93	.88	.98	1.03	-10.3	-10.8	-11.5	-11.1	1.39	1.63	1.59	1.75
Q2	MCVAE	**2.37**	**2.23**	2.12	**2.24**	.06	.06	.07	.08	-10.2	-12.3	-10.4	-12.1	1.82	4.71	5.01	3.31
	VAE	2.38	2.65	2.07	2.26	.01	.006	.002	.004	-8.1	-8.0	-8.4	-8.6	78	.16	.27	.51
	CGAN	5.29	6.16	3.86	5.49	1.20	1.40	1.07	1.05	-10.8	-10.9	-10.9	-10.5	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	-10.9	-11.6	-11.1	1.57	1.64	1.49	1.51
Q3	MCVAE	**2.04**	**2.15**	2.30	2.23	.09	.07	.07	.05	-11.3	-14.2	-14.0	-16.0	3.92	4.57	6.48	6.83
	VAE	2.12	2.18	1.82	1.93	.01	.005	.002	.004	-8.2	-8.3	-8.5	-8.9	.77	.10	.45	6.84
	CGAN	4.49	6.29	3.89	4.74	1.31	1.20	1.24	.97	-10.8	-9.8	-11.0	-10.8	1.91	1.68	1.93	1.54
	GAN	4.80	6.16	3.79	4.92	.96	.65	.92	.76	-10.5	-9.6	-10.8	-11.0	1.43	1.18	1.42	1.62
Q4	MCVAE	2.67	3.03	2.65	3.11	.08	.04	.06	.04	-13.0	-15.3	-11.6	-15.4	6.51	6.73	6.72	12.0
	VAE	2.41	2.76	2.14	2.57	.008	.002	.002	.005	-8.09	-8.2	-8.5	8.7	0.72	.03	.75	.79
	CGAN	4.77	6.20	3.95	5.52	.89	.91	1.01	.76	-11.0	-10.9	-11.3	-11.6	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	-10.5	-10.4	1.43	1.58	1.38	1.34

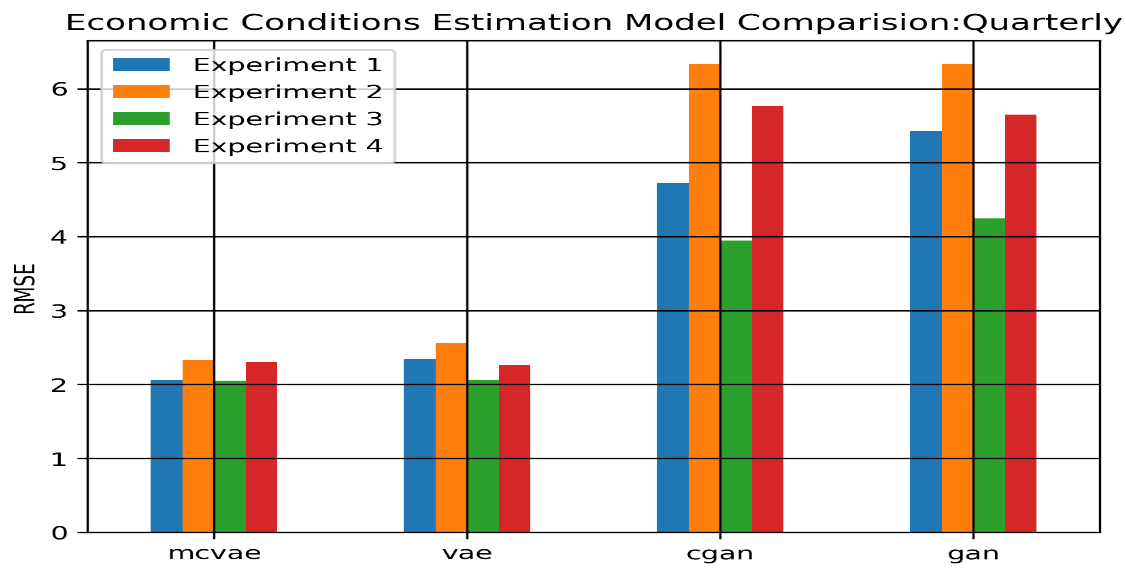
10 out of the 20 total experiments from both quarterly & yearly data configurations, outperforming the baseline generative models, seen in figures 2.20 and 2.19. Specifically, the proposed model performs the best when it can use more data, as seen in the quarterly experiment configuration. Furthermore, the proposed model can outperform baseline models in 3 of 4 experiments. The second-highest performance comes from the original version of the variational auto-encoder model. However, the evaluation metrics are close in values for cases where MCVAE does not outperform VAE.

Additionally, the mean log-likelihood results indicate that the proposed model is often close to that of the baselines. The standard deviation of the evaluation metrics through all experiment iterations indicates the robustness of each model. The variational autoencoders can achieve low standard deviations, while the generative adversarial networks have substantial variation in RMSE throughout the iterations. Lastly, the results from (Malik, 2018) are provided as context due to the partial similarity in task and experiment setup. The proposed framework can outperform the log-likelihood found in the related work.

Concluding Remarks: The proposed MCVAE model consistently outperforms the baseline models in the economic conditions estimation experiments. Moreover, the results of MCVAE are robust in terms of performance and consistent in terms of standard deviations through multiple experiment iterations. The practical utility that MCVAE may benefit the economic conditions modality estimation problem when compared to other generative models, which may not be as effective.

The MCVAE model showed through the log-likelihood metric via the Annealed

Figure 2.20. Quarterly Experiments



Importance Sampling (AIS) with bidirectional Monte-Carlo for validation of generative models, made famous by (Y. Wu et al., 2016), that the plausibility of the experiments is in line with the likelihoods of the baseline models. The model provides ”most likely to happen” estimation samples at the same probabilistic confidence range as the baseline models and existing models in literature (Malik, 2018).

Additional insights from the experimentation include the fact that the MCVAE model performs best when plentiful amounts of data, as depicted in the results of using continuous quarterly intervals against yearly intervals at quarterly periods. This finding could indicate that sequential or temporal auto-correlations may exist that could help with estimations or finding appropriate non-linear relationships to build an approximated model distribution. Alternatively, more data allows for the model to learn effectively for this setting.

Overall, the GAN based models did not perform well in this particular application domain. Although GANs typically perform better than VAEs in image learning, machine translation, computer vision tasks, they did not seem to be adequate for economic conditions estimation. Fundamental aspects of sampling from random white noise initially or the complexity of finding joint distributions in multi-modal data could explain the intuition behind the performance of the GAN based models.

The MCVAE's performance indicates that the considerations of multi-modal joint approximation, modality co-occurrence or lack thereof, and utilization of a conditional modality for data immediacy & availability play a crucial role in addressing the practical challenges typically faced with economic conditions estimation. Addressing these challenges with a novel model framework that combines multi-modal exogenous factors and conditional modality learning may provide a unique perspective in the generative models' research literature. Most importantly, the framework provides practical utility for economic conditions estimations tasks.

CHAPTER 3

BANK CAPITAL AND LOSS PREDICTION

Figure 3.1. Tier-1 Peer Group Moody's Projections

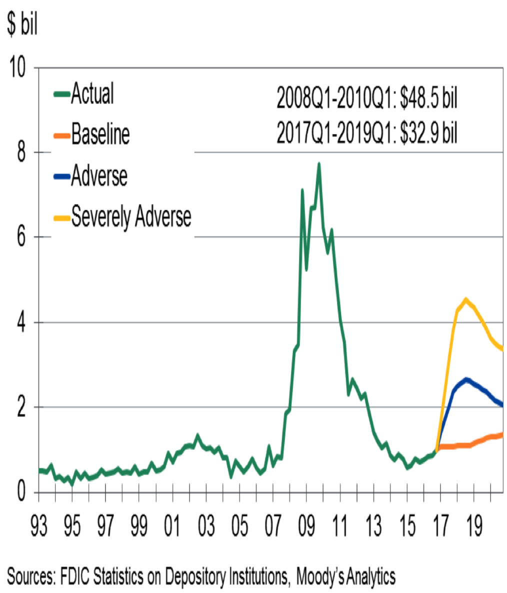


Figure 3.2. Net Charge Offs.

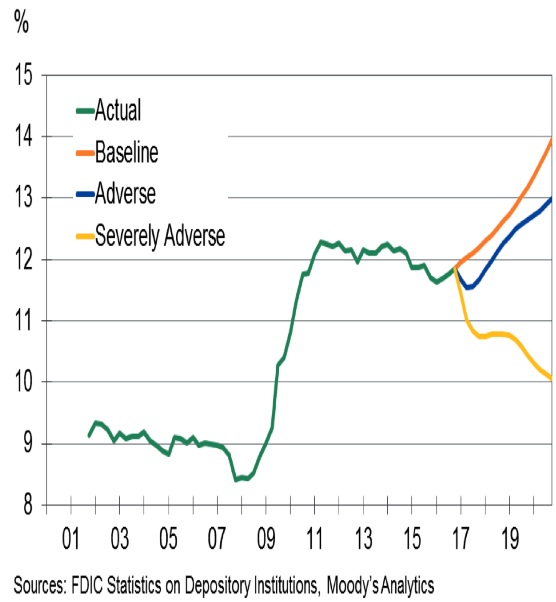


Figure 3.3. Capital Ratio.

3.1 Introduction

Traditionally, regulatory authorities refer to capital adequacy as the firm-level financial measures that represent the ability of a bank to stay solvent by having a relative abundance of resources to mitigate losses or implement strategic investments while considering for operating expenses, business profits, and investor distributions (Reserve, 2015). These ratios account for the capital funds that are kept on hand by firms in proportion to their respective risk-weighted assets. The larger the ratio, the stronger the indication that the firm would be able to operate normally. Conversely, if the capital ratio is not over a certain regulatory threshold, typically 5-8% of risk-weighted assets, the indication is a firm would not have the resources available to be sustainable if revenue generation and loan loss rates took adverse directions. Capital ratios provide stakeholders with a perspective on the firm's potential for growth and it's ability to sustain during a systemic crisis by having a buffer to withstand the effects for the adversity period.

Capital adequacy directly references regulatory mandates for firms to maintain specified thresholds of capital in proportion to their risk-weighted assets for sustainability through a financial crisis. Bank stress-test analytics perform balance sheet projections of loan portfolio loss rates, net revenues, and capital ratios during both expected and adverse economic conditions are crucial insights to determine the potential capital adequacy of a bank. Adverse economic conditions have shown to impact banking performance due to the direct impact on bank loan portfolio default and loss rates. The loan business is the primary service banks offer. Firms can expect significant losses to their core businesses caused by the inability to collect principal loans with interest because of adversity in the economy that impacts the loan consumer and risk exposures of counter-parties to investors.

The interconnections of the banking system through risky financial instruments may now incur significant losses to other banks that own said assets, potentially causing a systemic failure of the financial system. Firms without prudent risk management and conservative capital management plans may become insolvent, however, if a large portion of the banking system cannot maintain solvency, the effect on the national economy will worsen in a vicious cycle. Consequently, government-based tax-sponsored bailouts may be required to help the financial economy recover. Prevention of these types of occurrences justifies bank stress-test analytics, and exercises that seek to ensure firms sustain a satisfactory amount of monetary reserves to serve as a buffer in case of shortfalls to their business caused by economic conditions.

The goal of bank stress-test exercises is to understand banking performance during adverse economic conditions through the evaluation of risk management and conservative capital planning strategies. Regulators provide hypothetical economic conditions that vary in severity to simulate financial crisis characteristics to assess banking performance through balance sheet projections of loan portfolio loss rates, pre-provision net-revenue accumulations, and overall capital available. If the capital measures sufficiently adhere to the regulatory minimums during these adverse conditions, regulators deem the bank quantitatively satisfactory to withstand a financial crisis scenario. Therefore, the ability to expertly forecast key banking performance components of regulatory capital measures under dynamic economic conditions can be beneficial for stakeholders of bank stress-test analytics. As seen in figures 3.3 & 3.3, forecasts of net charge off rate and capital ratio in different economic adversities using historical context and commercial analytics (Hughes & Poi, 2016) provides relevant insight to stakeholders.

However, given the various characteristics depicting past financial crises' and their impact on banking performance, exploring the non-linear relationships of relevant banking characteristics and their respective periods may provide an insightful perspective, rather than generalizing all crises' to have static characteristics. Past financial crises, such as the student debt crisis, dot com bubble, housing bubble, and foreign credit debt crisis have shown varying impacts on banking characteristics at different periods. Therefore, addressing the practical challenge of modeling banking performance while considering dimensional and temporal influences that may improve overall estimations.

In this chapter, a bank capital & loss prediction (BCLP) model, which combines user-defined economic conditions modalities and firm-level banking characteristics to select relevant dimensional and temporal information for the prediction of banking capital & loss components of performance, is presented. Expanded from the Dual-Attention Recurrent Neural Network (DA-RNN) proposed in (Qin et al., 2017) as a non-auto regressive exogenous model (NARX), the proposed framework consumes economic conditions and firm-level consolidated financial statements to learn the relevant dimensions and periods. Ultimately the model predicts banking capital components based on attention-based feature selection networks that help capture the dynamic nature of economic conditions on banking performance.

3.2 Problem Formulation

Bank Capital & Loss Ratio Prediction:

Capital ratio forecasts under adverse scenarios, as administered in bank stress-tests, typically focus on the growth rates of loan portfolios through the net-charge-off, as shown in equation 1.2. Net-charge-offs are measures of each loan category that depict the banks' primary lending business, which would be directly impacted by

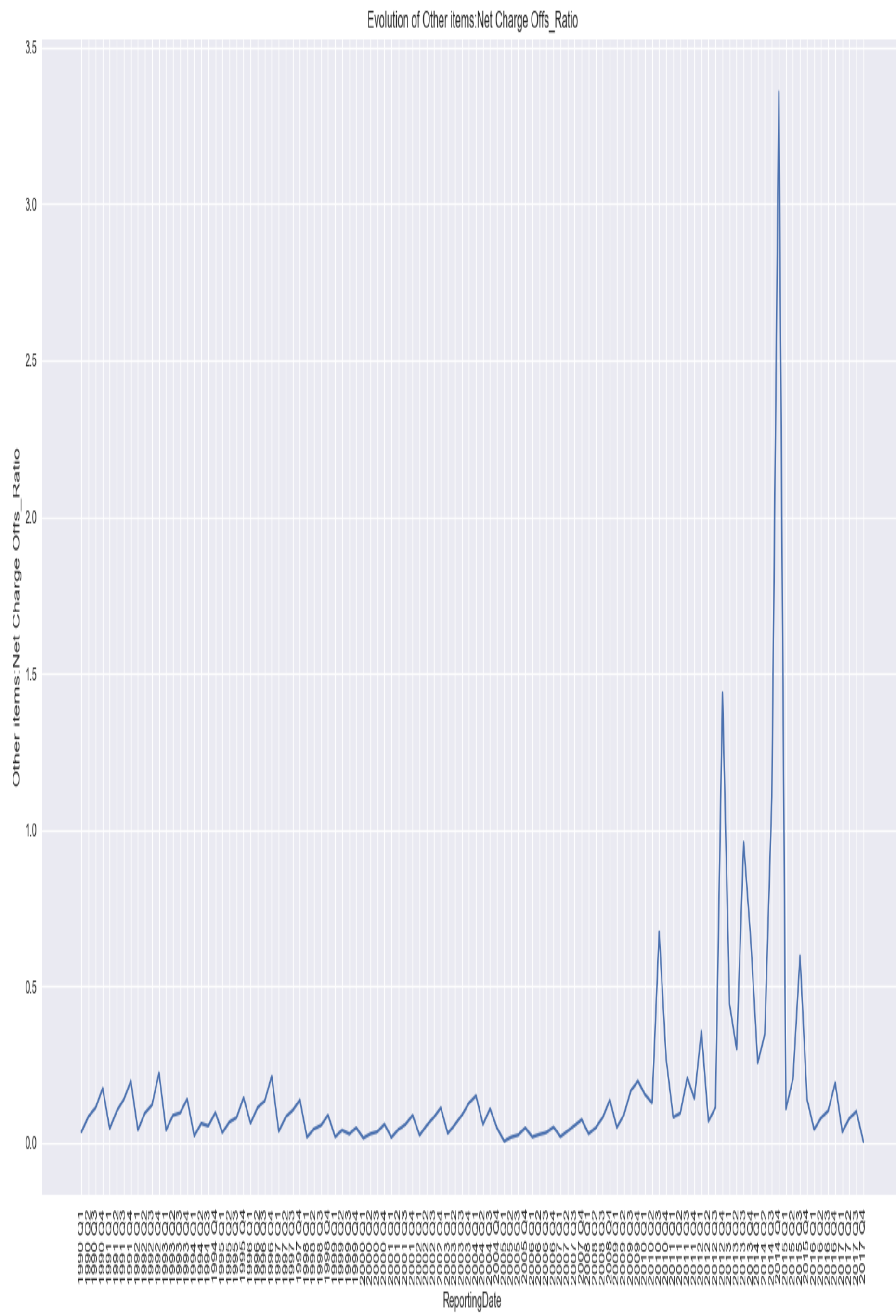


Figure 3.4. Net Charge Off Ratio of 1000 U.S. Bank Holding Companies

economic adversity. Figure 3.4 depicts the increase of the aggregate net charge off ratio of over 1000 U.S. bank holding companies during the recent financial crisis and recovery between 2008 - 2012. Additionally, aspects of pre-provision net revenue, as shown in equation 1.3, are also typically estimated to understand potential offsetting factors to said losses (Harris, Khan, & Nissim, 2018). Lastly, these estimated components of revenue and loss can subsequently derive book equity, as seen in equation 1.5, and the regulatory capital ratio measures, as seen in equation 1.6. The economic conditions, banking characteristics, and target banking performance variables are all numerical and follow a sequential quarterly reporting interval common with data from the financial domain.

Thus, the problem can be formalized as a time-series with exogenous terms prediction task, which aligns with the bank capital & loss prediction under economic conditions consideration task of bank stress-test analytics. More specifically, given the set of banking performance profiles, $[Y_{ncoR}, Y_{ppnr}, Y_{T1CR}]$, banking loan portfolios, $X_{loancat}$, economic conditions, $E\hat{C}O_{mod}$ and economic estimations from the previous chapter, $\Pr_\theta(E\hat{C}O_{mod})$, the task of predicting future bank capital, loan loss and net revenue is a function of each target variable's past in conjunction with economic conditions estimations, $E\hat{C}O_{mod_t}$, as seen in Eq.3.1.

$$\begin{gathered}\Pr(Y_{ncoR_t}, Y_{ppnr_t}, Y_{T1CR_t} | Y_{ncoR_{t-1}}, \dots \\ Y_{ppnr_{t-1}}, X_{loancat_{t-1}}, Y_{T1CR_{t-1}}, E\hat{C}O_{mod_t})\end{gathered} \tag{3.1}$$

Forecasting banking capital, $\Pr(Y_{T1CR})$ is typically derived from projected loan loss and revenue, $[Y_{ncoR_t}, Y_{ppnr_t}]$ (Covas et al., 2014; Hirtle et al., 2016), in accordance with equation 1.5 and 1.6. However for the purposes of the proposed approach,

Y_{T1CR_t} is directly projected using loan losses, net revenue, previous loan portfolio, and estimated economic conditions.

$$\Pr(Y_{T1CR_t} | Y_{ncoR_t}, Y_{ppnr_t}, X_{loancat_{t-1}}, E\hat{C}O_{mod_t}) \tag{3.2}$$

Furthermore, the proposed model framework in this chapter will use the estimated economic conditions, $\theta_{E\hat{C}O_{mod_t}}$, when conducting the capital, revenue, and loss projections. The reasons for this are two-fold (1) all the economic modalities will need to be estimated together for the future economic conditions estimations in a practical scenario. (2) Typically, access to economic conditions information for future periods may be limited and may need to use accurate estimations.

3.2.1 Challenges

The goal is to enhance the tasks described in equation 3.2 and 3.1 with features described in equation 1.1 for the bank stress-test problem. The practical challenge of generalizing the diverse nature of economic conditions profiles as they adversely impact banking performance at different periods needs addressing.

1. Identify dimensions from economic condition modalities and banking characteristics that may have the most influence in estimating target banking performance variables.

2. Identify dimensions in the temporal space combined with previously identified relevant feature-based dimensions that can provide insight as to which features are important at each period to estimate banking performance for a specific period.

3. With previously identified dimensional and temporal considerations of relevant features, robustly estimate the banking performance target variables leveraging the identification of non-linear relationships that may exist among the relevant banking characteristics, economic conditions, and banking performance.

4. Utilize the identified patterns and insights in a meaningful manner to effectively estimate banking performance that can ultimately outperform baseline models and techniques from the related literature.

The implementation of an approach that could potentially address the challenges discussed above while accomplishing the problem task defined in equation 3.2 and 3.1 may improve the effectiveness of the bank capital & loss forecasting in general and for the bank stress-test analytics domain.

3.3 Related Work

The bulk of literature related to the problem task of financial balance sheet projection of bank capital components performance with consideration of exogenous economic conditions for purposes of risk management or bank stress-test analytics is mainly surrounding the use of statistically linear approaches for time-series prediction with exogenous terms. This section focuses on related literature about forecasting banking performance with consideration for economic conditions and bank stress-test capital forecasting. The number of works in this niche area of study is currently still growing due to the challenges of requiring relevant domain knowledge and acquiring firm-level historical financial data.

The authors in (Carling, Jacobson, Lindé, & Roszbach, 2004) leverage "Vector

Auto Regression" (VAR) models to evaluate relationships between firm financial ratios and macroeconomic variables by developing a technique that allows for the balance-sheet projection of multiple target bank performance variables. The authors then conduct a factor analysis to assess the influence of particular macroeconomic and microeconomic terms have on banking performance.

In (Covas et al., 2014), the authors propose a "Fixed Effects Quantile Auto Regression" (FE-QAR) model within a top-down econometric methodology to generate predictions of net charge offs of loan portfolios and the significant components of pre-provision net revenue. The authors claim that their technique captures non-linear relationships among the feature space and yields plausible estimates. They are using real-world quarterly financial data from FR-Y9C and FFIEC 031/041 regulatory reports of 15 large banks holding companies from 1997 to the fourth quarter of 2011. The authors create 14 different models using banking characteristics, macroeconomic factors, and financial ratios closely aligned with the CCAR stress-test exercises to show the effectiveness of their model compared to the baseline model, "Ordinary Least Squares" regression. Lastly, they were able to determine the capital shortfalls by mapping estimated PPNR and net charge off ratios to capital ratios to generate a probability distribution to predict capital ratios. This work thoroughly considers the aspects of the CCAR examination, such as the hypothetical economic scenarios and the bank capital components forecasting, by creating a probability distribution that relates all the features to provide plausible capital estimations.

Similarly, in (Hirtle et al., 2016), the authors propose the "Capital and Loss Assessment under Stress Scenarios" (CLASS) model. The CLASS model employs linear regressions and specific assumptions about loan loss, asset growth, taxes, and other factors to calculate projected industry capital gaps during stressed macroeconomic conditions. The authors utilize regression models to project PPNR and loan categories while assuming growth rates and capital distributions are held constant. Twenty-two regression models are used to project the industry level capital ratios based on 200 large bank holding companies through the use of FR-Y9C and FFIEC Call reports from 1991 to 2013. They accomplish this by exploring the cumulative probability distribution of their estimations during expected and baseline economic conditions. The authors analyze the loan and revenue components of capital to understand the trends related to financial stability. With the capital projections, the authors can derive a capital buffer gap from assessing the amount of capital needed to meet regulatory thresholds. Finally, the authors compare the effectiveness of their capital gap assessment to the capital shortfall, and CDS spread measures. The authors remark that the CLASS model can serve as an early warning indicator for capital adequacy and conclude reasonable model utility for purposes of net income and capital forecasting under the economic scenarios. Overall, this work provides a very rigorous analysis of model results and relevant interpretation of the experiment findings, factor analysis, and the general perspective of the state of bank capital forecasting techniques.

The work in (Malik, 2018) incorporates scenario generation, bank performance prediction by using machine learning techniques and is the preliminary inspiration of the proposed framework in this chapter. The author's motivation is to find more plausible stress-tests for each bank using CGAN for scenario generation and LSTM for bank performance prediction to improve the current regulator methodologies. The author indicates the work will have scenarios that are tailored and comprehensive toward upcoming risks of banks or the industry. The work utilizes a generative deep learning method with competing for adversarial neural network modules to use a stochastic process that generates economic environment variables. Additionally, they have other methods that will confine the model to sample from a distribution that closely resembles plausible scenarios that are like adverse scenarios. The authors indicate their method can benefit regulators as it can be sampled from more plausible and risky distributions than the current regulatory scenarios, and can incorporate hypothetical shocks rather than relying on past scenarios. In their approach, they have two main parts, (1) Scenario Generation, where they sample from an actual joint probability distribution of all macroeconomic and microeconomic variables by leveraging a GAN. Furthermore, Performance Prediction, which relies on LSTM to cluster similar characteristic banks and historically similar performance internally. The authors model will predict the loss of each loss rate for each loan category based on the generated scenario and $t-1$ bank characteristics. The author's experiment on real-world data consisting of 16 macroeconomic variables from 1976-2017 & six microeconomic variables from 1980-2017 and FR Y-9C bank loan portfolio breakdowns for bank characteristics. Additionally, they conduct experiments based on bank and time split settings.

3.4 Methodology

For the bank capital & loss prediction task, utilization of attention-mechanism neural network, and recurrent neural network model categories of deep learning techniques to find an effective solution. Moreover, practically capturing the unique exogenous aspects that may impact banking performance concerning different periods is vital.

To address the practical challenge of enhancing the previously defined bank capital & loss prediction tasks from equations 3.2 and 3.1, the methodologies implemented into the model framework must be dynamic and robust. Popular deep learning techniques in literature have combined attention-based networks for feature selection tasks (Bahdanau, Cho, & Bengio, 2014; Qin et al., 2017). By using neural networks to place weights on input dimensions to assert the influence on target variables, and recurrent neural networks (Hochreiter & Schmidhuber, 1997; Malik, 2018), which employ sophisticated temporal structures, for time-series prediction with exogenous terms tasks.

In this section, a brief discussion on the background of attention-based neural networks, recurrent neural networks, and non-linear auto-regressive exogenous models relevant to the problem task. Additionally, elaboration into a proposed model framework that seeks to combine attention networks and neural networks to provide a useful model solution for the problem task.

3.4.1 Recurrent Neural Networks (RNN)

In deep learning literature, where sequential data that has persistent time steps, dependence on previous sequential inputs, or serves as an input, output, or both, recurrent neural network models for estimation tasks (Hochreiter & Schmidhuber, 1997). Sequential data is processed using a recurrence formulation, which shares model weights and recursively feeds the outputs back as inputs. This process infers vital aspects of the sequential data by identifying information that may be relevant for target variables. The persistence of information through looping inputs and outputs of the RNN's neural networks (Chung, Gulcehre, Cho, & Bengio, 2014). The "Recurrent" term in RNN, which refers to the same task on each sequential element. In contrast, the output elements are dependent on previous sequential components (Hochreiter & Schmidhuber, 1997).

RNN models have been popularized in literature as they have shown effectiveness in applications that may have one or more inputs or outputs. RNN usage has shown prominence in image classification, image captioning, machine translation, video classification, and sentiment analysis (Yi, Yu, Zhuang, Zhang, & Xiong, 2018). RNN components can be denoted as the hidden state, h, the input data x, the output, y, the weights, W, and the time steps, t. RNN can define the output task the following:

$$\begin{aligned} h_t &= f_\omega(h_{t-1}, x_t) \\ h_t &= \tanh\left(W_{hh}h_{t-1} + W_{h_x}X_t\right) \\ y_t &= W_{hy}h_t \end{aligned} \tag{3.3}$$

Technically, to process sequential data, a RNN will take in the input x at time t or x_t, the previous hidden state value h_{t-1} into a tanh activation function to calculate h_t and use tanh or Relu to find the non-linearity in the output, y_t. The weight matrix is shared at every step at each layer of the RNN neural network. The hidden state functions as a bridge to connect the pertinent information from the previous states to function as a memory component for the RNN model. The output at a specific time step relies on the current input along with the previous states. RNNs share the weight parameter at each time step for each layer, which differentiates it from other deep neural networks that use specified parameters at each hidden layer. Random initialization of weight matrices occurs; however, during the training of the RNN, ideal weights to accomplish the behavior needed for the RNN. By calculating a loss function, the measurement of the deviation between the ground truth output and the model predicted output. To reduce the loss, RNNs use backpropagation, which shares the weights across the layers at all time steps (Goodfellow et al., 2016). As such, the gradient of the error at each step relies on the loss at previous steps.

Computation of the gradient for h_t requires aspects of W due to the need for backpropagation of each RNN cell. However, this is inefficient computationally. Challenges about the exploding gradient refer to the most significant singular weight value being over 1. Vanishing gradient refers to the most considerable singular weight value being less than one, mitigating the challenges by gradient clipping techniques, which set gradients to a reduced acceptable value if they exceed a preset threshold, or by using variants of RNN.

3.4.2 Long Term Short Memory (LSTM)

Situations exist where an RNN may encounter the vanishing gradient problem or the difficulty in learning necessary information to predict the next time step effectively. Caused by the distance between the time steps, a variant of the RNN model, Long Short Term Memory (LSTM), is popularly applied (Yi et al., 2018).

LSTM can efficiently learn informational dependencies from the sequential data, regardless of the distance of the time step distance from the prediction. LSTMs can connect information from over one thousand time steps through the use of an optimized gradient-based algorithm, which considers consistent error flow within the internal states (Hochreiter & Schmidhuber, 1997).

Specifically, LSTMs aim to recall information that traverses time steps. Therefore it must be able to assess the relevant information to remember or forget. The cell state is an integral component of the LSTM framework, functioning as an internal memory component and using four different regulated gates to assess the previous states by determining which information to add or remove from the cell states. The gates operate in a manner that controls the amount of information that should go to the next cell state (Hochreiter & Schmidhuber, 1997).

The initial step in LSTM decides if the cell state needs to be remembered or forgotten through the Forget gate, depicted in equation 3.5, which uses a sigmoid activation function to give a binary output value of 0 or 1, as depicted in equation 3.4.

$$S(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{e^x+1} \tag{3.4}$$

An output of 0 from a forget get indicates that the cell state should be kept, while an output of 0 means that the cell state value should be forgotten (Hochreiter & Schmidhuber, 1997).

$$f_{(t)} = \sigma(W_{fx}X_t + W_{fh}h_{t-1} + b_f) \tag{3.5}$$

The next step assesses which new information will be contained in the cell state, denoted in equation 3.7. In this step, the task of deciding if information should be written into the cell state by leveraging a sigmoid function is addressed in the input gate. The next task to decide the amount of information to store in the cell state using a tanh activation, also known as the hyperbolic tangent function depicted in equation 3.6, is performed by the gated gate.

$$\tanh z = \frac{\sinh z}{\cosh z} = \frac{e^z - e^{-z}}{e^z + e^{-z}} = \frac{e^z - e^{-z}}{e^z + e^{-z}} \tag{3.6}$$

The output from a tanh activation function will be between -1 and 1.

$$\begin{aligned} C_{(t)} &= f_t \Theta C_{t-1} + i_t \Theta g_t \\ f_{(t)} &= \sigma(W_{fx} X_t + W_{fh} h_{t-1} + b_f) \\ i_{(t)} &= \sigma(W_{ix} X_t + W_{ih} h_{t-1} + b_i) \\ g_{(t)} &= \sigma(W_{gx} X_t + W_{gh} h_{t-1} + b_g) \\ o_{(t)} &= \sigma(W_{ox} X_t + W_{oh} h_{t-1} + b_o) \end{aligned} \tag{3.7}$$

The penultimate step in the LSTM framework creates the new cell state by concatenating the output values from the previous two steps by deriving the product between the post tanh activation for the current time step and the values yielded from the output gate.

$$h_t = O_t \Theta \tanh C_t \tag{3.8}$$

The final step in LSTM, the cell state, which functions as the internal memory of the LSTM unit, calculates the product of the previous cell states, the output of the forget gate, the newly computed hidden state, denoted in equation 3.8, from the gated gate, and the output from the input gate. The final output of the LSTM unit relies on calculations from a formalized version of the cell state.

LSTM addresses vanishing and exploding gradient challenges by using backpropagation from the current cell state to the previous cell state that only considers element-wise alignment by the forget gate without any matrix multiplication operations of the model weights, W (Hochreiter & Schmidhuber, 1997; Yi et al., 2018; Malik, 2018).

The techniques mentioned in this section in regards to LSTM are what ultimately allows for the modification of the information stored in memory at each time step by assessing what information to forget, remember, or update. This model framework allows for long term dependencies stored in memory for sequential data prediction tasks.

3.4.3 Gated Recurrent Unit (GRU)

A Gated Recurrent Unit (GRU) is a popular variant of LSTM but does not consist of internal memory features. GRU consists of a reset gate and an update gate in contrast to the three-step process in LSTM.

The reset gate in GRU determines how to combine new input information with previous time step's information. The update gate assesses the retention of information from the previous memory. When compared to LSTM components, the GRU update gate consolidates the input and forget gates (Chung et al., 2014). GRU is a version of LSTM that is less complex since it does not include features for long-term memory dependencies while addressing the vanishing gradient problem.

3.4.4 Attention-Mechanism

As input sequences get long, the ability of RNNs to adequately learn temporal dependencies from the input data becomes more complicated and of lower quality. The concept of attention strives to identify specific key input vectors of the sequence based on weights that indicate relevance to the output sequence. The attention mechanism is often used with RNNs to help improve the efficiency and performance of the model. Popular variants include additive attention mechanisms, which can perform linear combinations between encoder and decoder states to help learn common characteristics of the input sequences (Bahdanau et al., 2014). Hidden states from both the attention encoder and decoder used to create the context vector.

The context vectors provide insights about the input sequence data when predicting the output. Context vector comprises of individual weighted sum from hidden state outputs from an encoder. Each vector, h_i has some information about the entire input sequence since they are derived from encoder states, with particular insight on the i^{th} input in the sequence. The hidden state vectors are scaled by attention weights, a_{ij} which assign the relevance of the input x_j to the output, y_i, as denoted in equation 3.9.

$$c_i = \sum_{j=1}^{i} a_{ij} h_j \tag{3.9}$$

Furthermore, the attention weights are determined by a fully-connected network, fc, which leverage a softmax function. The calculation of the weights are denoted by equation 3.10.

$$\begin{aligned} a_{ij} &= \frac{\exp e_{ij}}{\sum_{k=1}^{i} \exp e_{ik}} \\ e_{ik} &= fc(s_{i-1}, h_j) \end{aligned} \tag{3.10}$$

a_{ij} represents the importance that h_j has with respect to the previous hidden state, s_{i-1} when asserting the next state, s_i, and ultimately produces the prediction output, y_i. The larger a_{ij} weight values will cause the respective RNN model to focus on input x_j when predicting output y_i (Bahdanau et al., 2014).

3.4.5 Non-linear Auto-regressive Exogenous model (NARX)

The time series analysis and modeling category which leverages nonlinear auto-regressive models with exogenous inputs is commonly known in the literature as NARX models (Lin, Horne, Tino, & Giles, 1996). Specifically, models that try to identify patterns in a current time series value from past values from the same series, current values and past values of exogenous series that may influence the target series with a error term that references knowledge that is currently unknown, but impacts the current value of the time series can be formalized in equation 3.11.

$$y_t = F(y_{t-1}, y_{t-2}, y_{t-3}, \ldots, u_t, u_{t-1}, u_{t-2}, u_{t-3}, \ldots) + \varepsilon_t \tag{3.11}$$

Where y is the target series value and u is the exogenous variables (Qin et al., 2017). In this problem task, assuming that u, and previous values of y may help predict the current value of y and ε_t is the error term that accounts for the factors not considered, but influence prediction of the target variable.

The function, F, in this model framework is a non-linear function, such as a neural network (Lin et al., 1996; Qin et al., 2017).

3.4.6 Dual-Attention-based RNN (DA-RNN)

In this dissertation 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, from microeconomic to macroeconomic perspective; and (2) Historical bank performance values of the respective time-series. These two factors are integrated in this work by applying a Dual-Stage Attention-based neural network model (DA-RNN) (Qin et al., 2017), which takes both attention mechanism (Bahdanau et al., 2014) and long short-term temporal dependencies (LSTM) (Hochreiter & Schmidhuber, 1997) into consideration for better time-series prediction.

The architecture of DA-RNN is demonstrated in figure 3.5, and the detailed learning procedure is illustrated in algorithm 2. Specifically, DA-RNN consists of two LSTM networks that incorporate attention mechanisms to select relevant features. The first LSTM component (Input attention mechanism) 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 the previously proposed MCVAE model from chapter 2. As for each $\mathbf{x}_t, 1 \leq t \leq T, t \in \mathbb{N}^+$, there is $\mathbf{x}_t = \{x_t^1, x_t^2, \ldots, x_t^N\}$ consists of N different modalities, and each modality is given a weight factor $\alpha_t^k, 1 \leq k \leq N, k \in \mathbb{N}^+$ based on a attention-based network to construct a weighted representation of original inputs (Razzak, Yi, Yang, & Xiong, 2019) as follows:

$$\widehat{\mathbf{x}}_t = (\alpha_t^1 x_t^1, \alpha_t^2 x_t^2, \ldots, \alpha_t^N x_t^N), \tag{3.12}$$

Where α_t^k is calculated by referring to the previous hidden state h_{t-1} and the cell state s_{t-1} in the encoder LSTM unit as following steps:

$$e_t^k = v_e^\top tanh(W_e[h_{t-1}; s_{t-1}] + U_e x_t^k), \tag{3.13}$$

$$\alpha_t^k = \frac{exp(e_t^k)}{\sum_{i=1}^N exp(e_t^i)}, \tag{3.14}$$

Where $v_e^\top$, W_e, and U_e are model parameters to learn.

In the temporal attention component, the LSTM units take $\widehat{\mathbf{x}}_t$ as inputs and generate its corresponding hidden state h_t', which is combined with the previous decoder

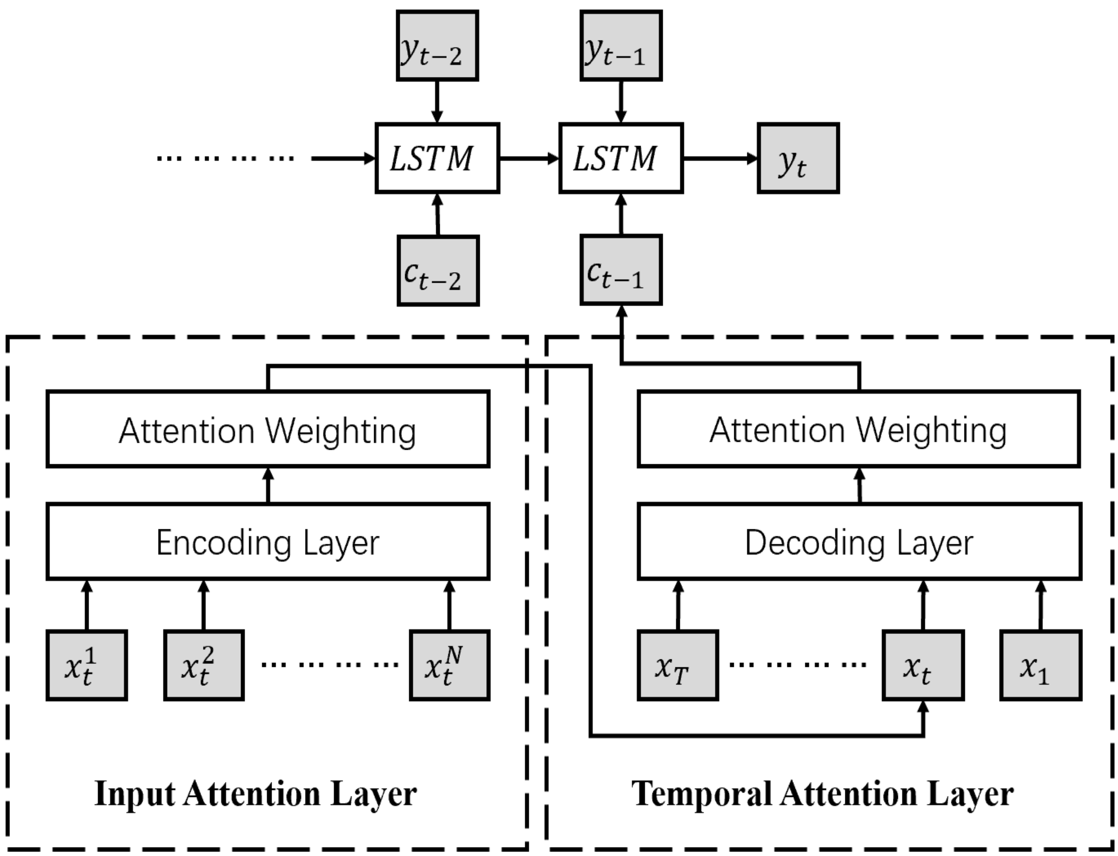


Figure 3.5. The Dual-Attention based RNN framework.

hidden state d_{t-1} and the cell state of the LSTM unit s'_{t-1} to generate the attention weight β_t^i of each hidden state $h'_i, 1 \leq i \leq T$ as follows:

$$l_t^i = v_d^\top tanh(W_d[d_{t-1}; s'_{t-1}] + U_d h'_i), \tag{3.15}$$

$$\beta_t^i = \frac{exp(l_t^i)}{\sum_{j=1}^{T} exp(l_t^j)}, \tag{3.16}$$

where $v_d^\top$, W_d, and U_d are model parameters to learn. The attention factor β_t^i indicates the importance of the i-th encoder hidden state for the prediction, and the context vector c_t is defined as a weighted sum of all the hidden states $\{h'_1, h'_2, \ldots, h_T^i\}$,

$$c_t = \sum_{i=1}^{T} \beta_t^i h'_i. \tag{3.17}$$

After all the attention-based feature learning and transformation is completed, the context information c_t for predicting bank performance at time t is obtained, along with it's corresponding historical time-series data $y = \{y_1, y_2, \ldots, y_{t-1}\}$, in which the formulation of a simple LSTM framework is established by concatenating y_{t-1} and c_{t-1} together to infer y_t as:

$$y'_{t-1} = w^\top [y_{t-1}; c_{t-1}] + b, \tag{3.18}$$

$$d_t = f(d_{t-1}, y'_{t-1}), \tag{3.19}$$

$$y'_t = v_y^\top (W_y[d_T; c_T] + b_w) + b_v, \tag{3.20}$$

where f represents a LSTM unit, w, b, v_y, W_y, b_w, and b_v are parameters to learn.

The dimensional and temporal consideration of the DA-RNN model makes it a capable candidate for the prediction task given that different banking performance

Procedure 2 DA-RNN for bank capital & loss prediction

Input: modalities $\mathbf{x}_t$, historical records $y_{1:(t-1)}$

Output: prediction y'_t

for $x_t^k \in \mathbf{x}_t$ ▷ *Initialize Input Attention encoder-network layer* **do**

  $h_t \leftarrow LSTM_{unit}.forward(x_t^k)$

  $\alpha_t^k \leftarrow Softmax.forward(h_{t-1}, x_t^k)$ ▷ weighting

  $\widehat{\mathbf{x}_t} = (\alpha_t^1 x_t^1, \alpha_t^2 x_t^2, \ldots, \alpha_t^N x_t^N)$

end

for $\widehat{x_i} \in \widehat{\mathbf{x}_T}$ ▷ *Initialize Temporal Attention decoder-network layer* **do**

  $h'_i \leftarrow LSTM_{unit}.forward(\widehat{x_i})$

  $\beta_t^i \leftarrow Softmax.forward(h'_i)$ ▷ weighting

  $c_t = \sum_{i=1}^{T} \beta_t^i h'_i$ ▷ 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

return prediction y'_t

Table 3.1. Bank Capital & Loss Prediction Experiment Configurations

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.

profile and economic condition characteristics may have influences on the capital and loss ratios at different periods (Razzak et al., 2019). This insight tends to be true for the financial industry, given that recent crises' (e.g., housing crisis, the internet dot-com bubble) stem from different banking and economic factors.

3.5 Experiment

In this section, the application of popular RNN models and performance of models from the state of the art literature, (Malik, 2018), will be compared to the proposed DA-RNN based model on a bank capital & loss prediction task that relates to bank stress-test analytics. The goal is to validate the effectiveness and relevance of the techniques on real-world banking data to ascertain model utility in bank stress-testing settings. Leveraging deep learning techniques from machine learning literature to formulate the task objective as a "non-linear auto-regressive exogenous modeling for time series analysis" promotes the use of techniques considered to be state of the art in research for a practical domain task. To further comprehend the benefits that each particular modeling technique may have on the problem task of bank capital & loss prediction and the practical challenges, comparative experimentation is required. Thus, evaluating the effectiveness of each model technique's ability to learn the relevant sequential dependencies among exogenous inputs to predict a target output series may help to address generalization complexities. These complexities are familiar with diverse banking performance impacts during different types of financial crises. Validations of performance through model comparison to baseline RNN models provide an empirical understanding of the tangible contributions and practical utility that the models offer for this specific problem setting. The use of real-world historical data relevant to bank stress-test analytics for this experiment may provide further insights unique to this domain area as well as provide a research direction for NARX models on time series data.

3.5.1 Data Description

The bank holding companies' consolidated financial statements (FR-9YC) for over 20k banks from 1986 to 2017 come from the WRDs bank regulatory data repository (WRDS, 2019). Specifically, data selection through reporting codes from (Covas et al., 2014; Hirtle et al., 2016) retrieves banking characteristics and performance attributes from the financial statements. Additionally, calculations provided in (Covas et al., 2014; Hirtle et al., 2016) are followed to derive net-charge-off rates, Y_{ncoR}, for respective loan categories, $X_{loancat}$, pre-provisioned net revenue, Y_{ppnr}, and capital ratios, $Y_{CapRatio}$. The table 3.2 summarises the data about banking characteristics and performance.

Additionally, historical economic conditions data from tables found in chapter 2 of this dissertation are estimated using the MCVAE method to provide economic condi-

Table 3.2. 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

tions vector for the experiment. Thus, this experiment predicts banking performance based on estimated economic conditions for the respective period, similar to the bank stress-test exercises (Reserve, 2015).

Outlier & Missing Data Handling

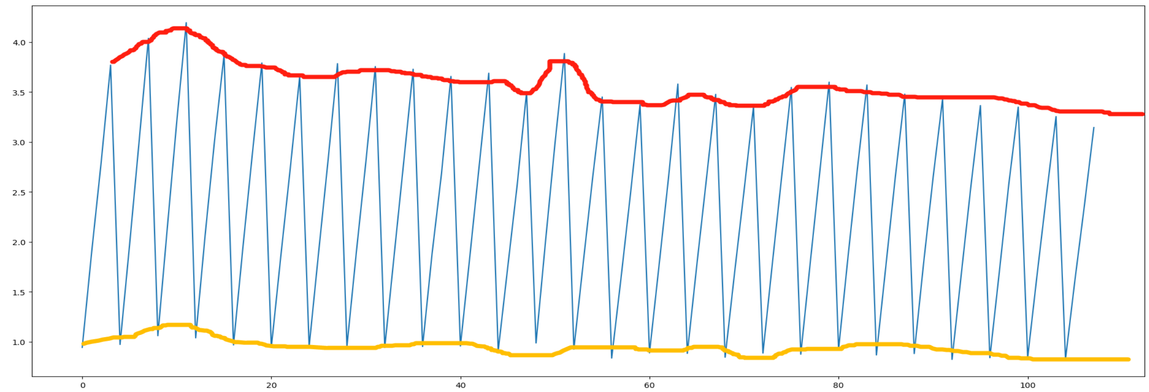
The traditional "Standard Deviation Method" (Leys, Ley, Klein, Bernard, & Licata, 2013) on the $Y_{ppnr}, Y_{nco}, Y_{T1CR}$ variables handles for data anomalies and outliers. Transformation of values that are beyond four standard deviations from the mean of the respective dimensional vector to the closest non-outlier neighbor value. Additionally, interpolation on missing data using linear methods found in the python pandas libraries (McKinney, 2010) for data quality issues.

Using the methods mentioned above allows for the usage of potentially useful data for the experiments while assuming that data quality issues are not representative in the full data set, but can be generalized for the experiment.

Data Selection

When sampling bank holding companies for this experiment, filtration of banks that have at least eight consecutive quarters since 1990 allows for meaningful analysis. The intuition is similar to the work in (Malik, 2018). Furthermore, the selection of 1000 banks based on the most considerable average total consolidated assets within the filtered data-set, are considered for experimentation. Effectively, addressing the potential for survivorship bias in the data-set while focusing on only the most representative banking entities as it relates to bank stress-tests.

Figure 3.6. Auto correlation in Bank Characteristics



3.5.2 Temporal Considerations

Upon initial inspection of the bank data and the respective target variables, it was discovered several dimensions follow a sinusoidal pattern across periods (quarters), which may cause challenges in forecasting. Specifically, yearly data has the most substantial weighted influence from $t-4$, indicating a periodic pattern after every four time periods, as shown in figure 3.6. The reasons behind this pattern could be related to dynamic regulatory reporting mandates or more banks reporting on certain quarters. Leveraging this temporal correlation on experiments for both quarterly and yearly settings can determine if the phenomena are advantageous for the problem task.

3.5.3 Bank Capital & Loss Ratios Prediction

Experiment Setup. Modeling of the target variable Y_t at both quarterly & yearly intervals, which will then consider two data split types and two experiment settings, totaling in twenty experiments for each target variable.

The purpose of the data configurations is to understand how well the prediction task can project capital and losses for a known crisis event period (Experiment 1:Train 1990-2007, Test 2008-2016) and all of the available data (Experiment 2: Train 1990-2015, Test 2016-2017).

Each experiment will have two different data splits, (a) Time Split, where the prediction of Y_t for the testing set, and (b) Bank Split, where 80% of the banks are used for training to then predict all of the Y_t for the remaining 20% of the banks. Please refer to Table 3.1 for a summary of the experiment setup.

Evaluation Metrics. For the bank capital & loss prediction task, the "root mean squared error" (RMSE) evaluation as per Eq.2.13 compares how well the predicted estimates match the testing set and to compare the effectiveness of each model.

In this experiment, Y_i is the actual loan loss or capital ratio value, $\hat{Y}_i$ is the respective estimated value, and n is the number of testing observations. Specifically, lower RMSE values indicate superior model performance.

Baseline Algorithm. Evaluation of bank capital & loss ratio prediction is performed by comparing the proposed DA-RNN (Qin et al., 2017) model with a set of baseline models, (1) Long Short Term Memory (Hochreiter & Schmidhuber, 1997); (2) Gated Recurrent Unit (GRU) (Chung et al., 2014).

Training of all baseline models are on the same data-set, and their performance evaluation uses the same validation set as the proposed model. Implementation of the experiments is done by the python deep learning library PyTorch (Paszke et al., 2017).

3.6 Conclusion and Discussion

Overall Performance.

The performance of the proposed DA-RNN model and the baseline models over all twenty experiments for two target variables, Y_{T1CR}, Y_{ncoR}, is summarized in table 3.3, figure 3.7, figure 3.8 and figure 3.9.

It can be observed that the proposed DA-RNN model outperforms the baseline models in both experiments, when using quarterly data in a time split configuration, as seen in figure 3.9. The proposed model also shows consistent performance against baseline models for a significant portion of experiments consisting of the yearly data intervals for all four quarters, as seen in figures 3.7 and 3.8.

In total, the model is able to outperform all baseline models in 24 of 40 experiments, with predominate success predicting target variable Y_{ncoR} in 15 of 20 experiments. The experiments this model did not do well on compared to the baselines are related to the bank-split configuration and target variable Y_{T1CR} for yearly Q2 and Q4 data in the time-split configuration.

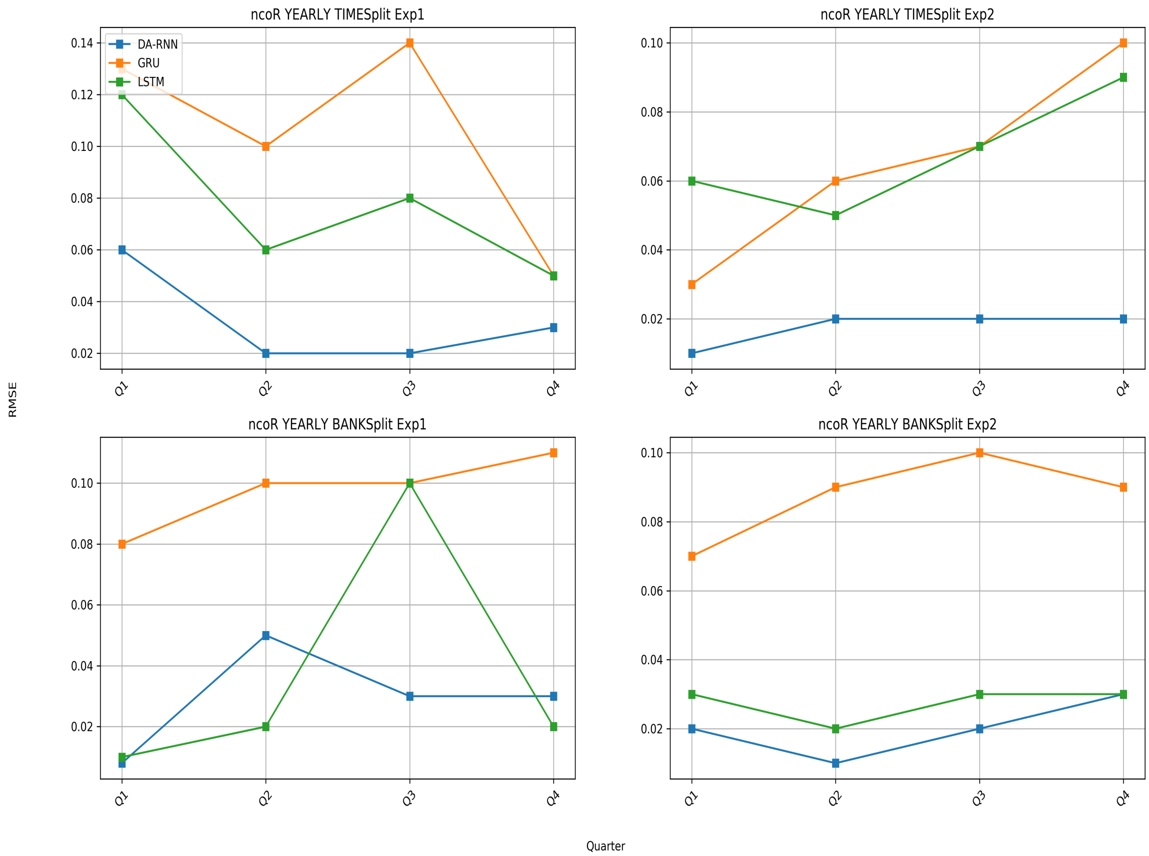
The dynamic nature and differences among bank holding companies, which may cause difficulties in generalization, could explain the banking split performance disparity. Moreover, data sparsity and quantity processed by the model may have lead to insufficient learning for generalization, which may explain the ineffectiveness of the model in the year over year experiments.

The proposed DA-RNN model shows consistent performance when compared to the baseline models in the bank capital & loss prediction experiments. Additionally,

Table 3.3. 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
	LSTM(Malik)(Malik, 2018)	.61	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	VAR(Malik)(Malik, 2018; Covas et al., 2014)	.94	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	FE-QAR(Malik)(Malik, 2018; Covas et al., 2014)	.88	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

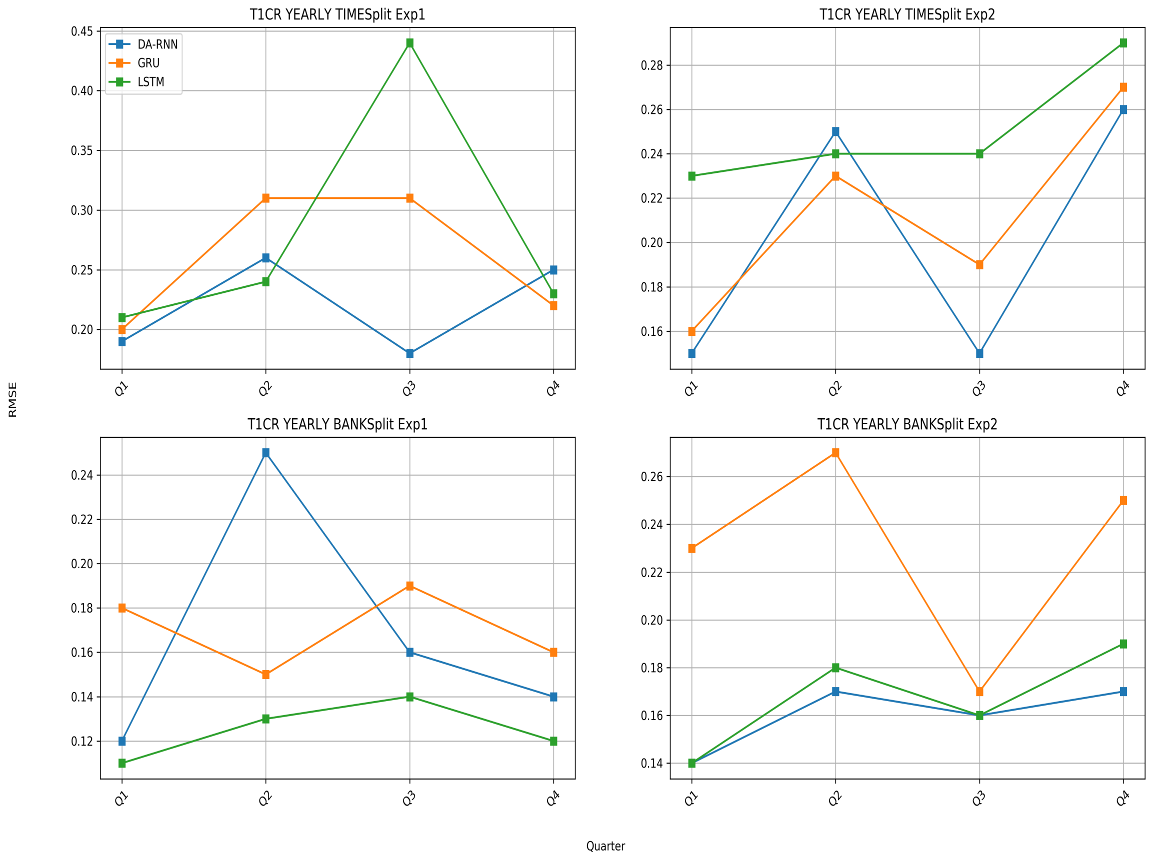
Figure 3.7. Yearly Y_{ncoR} RMSE



the experiment findings provide insights in regards to the effectiveness of predicting bank capital and loss under different data interval and splitting settings. The multi-view perspective of the task gives credence to the modeling approach and highlights the benefits it may contribute to the problem task.

It was evident that the DA-RNN model performed particularly well under time split configurations when compared to bank split configurations. Given that the DA-RNN model is a NARX model that takes sequential input data, the time-split configuration better aligns with that fundamental aspect. The bank split configuration measures the ability of the model to generalize based on a training set of banks

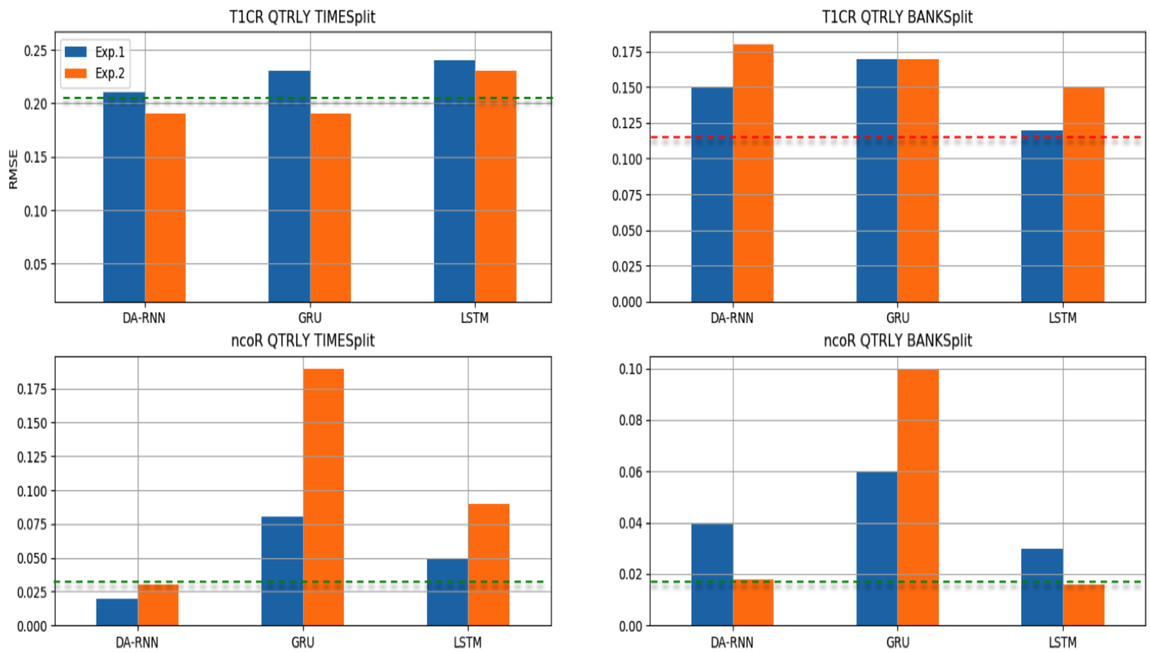
Figure 3.8. Yearly Y_{TC1R} RMSE



to predict the output of banks in the testing set fully. This setting is not appropriate for the bank stress-test domain due to the diversity of each bank's operation and strategy that may be unique to each bank, and temporal dependencies of other banks most likely would not be applicable.

The DA-RNN model's ability to outperform baseline models in the majority of net charge off rate, Y_{ncoR}, predictions is a crucial contribution of this work. Net-charge-off rates are a vital component to calculating banking capital, thus being able to estimate them well with given economic conditions provides a strong basis supporting the model's utility.

Figure 3.9. Quarterly Y_{TC1R} and Y_{ncoR} RMSE



Concluding Remarks:

Interestingly, capital ratio predictions for the quarterly, yearly first quarter and yearly third-quarter data interval showed robust and consistent performance. For the quarterly intervals, this indicates many reporting dates and data volume available help the model learn appropriate aspects to provide accurate predictions. The level of granularity and temporal dependencies that the model can find in the input sequential data to learn better how to produce relevant outputs, similar to most models that perform better with more relevant data, may explain this experimental finding. However, the performance of the model during the year first quarter and yearly third-quarter predictions may indicate that data quality, integrity, and volume may be more reliable during Q1 and Q3 of the year. Regulatory reporting deadlines,

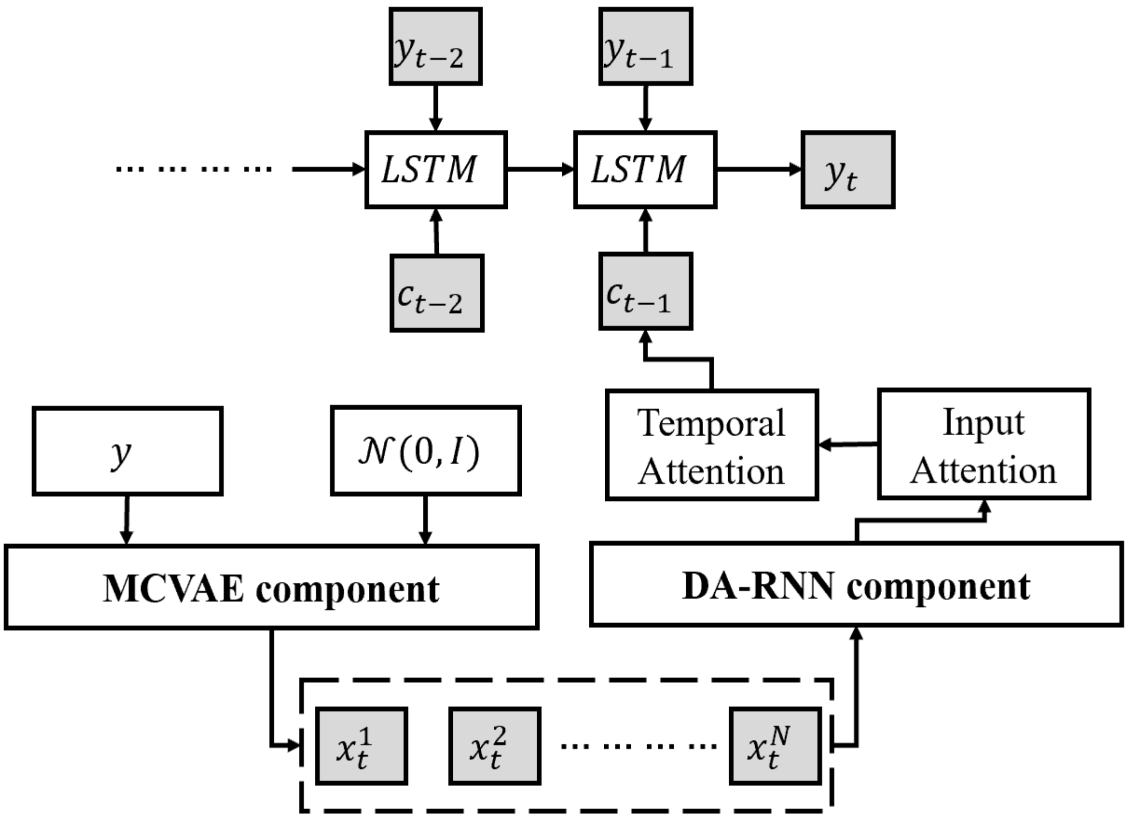


Figure 3.10. The Integrated Multi-modal Bank Stress-Test Prediction framework.

stakeholder reports, or other banking aspects that seem to make the data of these quarters more compatible with the model may explain this finding.

Ultimately, the experiment findings support that the model considerations help improve effectiveness in performing the problem task. Mainly for this experiment setting, dimensional and temporal attention mechanisms help to model banking performance during different economic conditions in a dynamic manner rather than dependence on a static generalization. These model considerations are highly applicable to the bank stress-testing exercise setting due to the emphasis on firm-level reaction to economic conditions. The model can analyze which bank characteristics are impacted by specific economic conditions at certain periods to make an estimation on banking performance for that specific time-step. Much like the real-world, banks may be impacted differently at diverse periods during varying economic conditions. This contribution indicates that a new perspective may be necessary for balance-sheet projections that consider exogenous terms in which both dimensional and temporal features are important factors.

Lastly, the utility of the DA-RNN model may be highly relevant to domain level stakeholders of banking stress-test analytics. Compatibility of the ECE and BCLP to work together allows for the Integrated Multi-modal Bank Stress-Test Prediction (IMBSTP) model framework as an end to end analytical tool, seen in figure 3.10. Furthermore, the model contributes relevant real-world experiments to the academic literature on NARX models to support domain-level considerations that may help improve the general time-series prediction task.

CHAPTER 4

TRANSFER-ABILITY OF BANKING KNOWLEDGE

4.1 Introduction

Figure 4.1. Potential for Financial Crisis in China

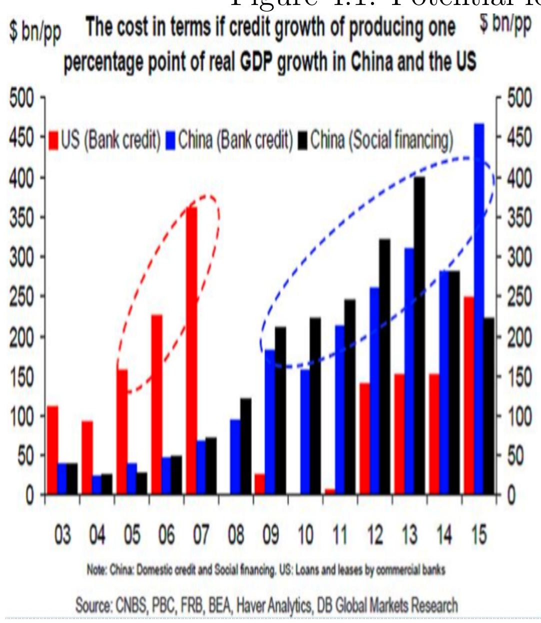


Figure 4.2. Cost per real GDP growth.

Figure 4.3. Housing area sales growth.

Although the recent financial crisis in late 2008 stemmed from the fragility of the subprime housing mortgage bubble in the U.S. banking system (Reserve, 2015), the impacts reverberated across global economies due to the systemic nature of the financial system (Ellahie, 2013). Not only did the U.S. have regulatory mandates to conduct bank stress-testing, but foreign banking authorities also enacted regulatory supervision. Ensuring that each respective global banking system would also be able to withstand crisis scenarios that were similar to recent experiences. External stress-testing methodologies and techniques are often similar to those applied in the U.S. However, the regulatory transparency at the firm-level is often obscure or aggregated for public consumption. Further complicating the identification of similarities among nations, the data collection and reporting for purposes of bank stress-test exercises may be disparate due to non-standardized accounting practices employed by global regulators.

Though differences may exist among international financial banking systems concerning regulatory reporting, accounting practices, and public disclosure of information, fundamental operations and strategies of a bank ultimately gravitate towards the same principles. Obtaining these fundamental aspects of strategy and operation may provide universal insights on firm-level banking performance. The insights could potentially serve as an integral proxy for forecasting related components from a foreign banking system, which may lack regulatory reporting transparency and data availability.

Global banking systems generally follow similar accounting practices, albeit non-standardized, which measure and collect information in regards to assets, liabilities, and shareholder's equity. Lending practices that accrue interest, investment divisions that seek to grow value through financial instruments and operations, and stakeholder capital distributions are common characteristics of banks that represent their primary methods of attaining revenue, expense, and capital. Financial statements are typically filed quarterly to their respective regulatory authority for purposes of market integrity and to assess capital adequacy in a potential economic downturn. Countries, such as the U.S., provide the public with a firm-level top-down perspective into banking characteristics and overall performance. However, banking systems, such as the European Union, provide only aggregated views to similar information, reserving the fine-level data for regulators to preserve competitive integrity. Countries, such as China, also have rigorous regulatory oversight. However, the firm-level transparency of regulatory intervention, the number of large banks, and the overall capital adequacy of these banks may be challenging to acquire for academic modeling for stress-test exercises.

The performance of banks in China and their impact on the overall economy play a significant role in the systemic effects of the global economy, because of the general global synergies that exist among larger nations. However, applying sophisticated analytical techniques may not be feasible due to the limited number of banking instances that are available to the public for research. Modeling efforts may not correctly capture potential adversity or dynamic events in the economy to generalize the variability in banking capital components. If this is the case, forecasts of capital for these banks in China may be at risk of failing to have enough capital to function through an economic crisis (Zhao, Zhan, Jiang, & Pan, 2017) properly.

Increased concern for regulatory capital adequacy were recently arose through the recent government bailouts of the Bank of Jinzhou and Baoshang Bank. Additional concerns from the IMF's FSAP on shadow banking activities & looming credit bubble, and the delay of annual reports from 20 Chinese banks are indications of potential turmoil in the financial systems. The culmination of these concerns amounts to an expected financial crisis in China, according to many analysts (Shen, Lee, Wu, & Guo, 2016). Figures 4.3 & 4.2 depict concerning macroeconomic and microeconomic signals of a potential credit crisis.

A looming demand for enhanced forecasting techniques on Chinese banking capital components for both expected economic scenarios and adverse scenarios seems to have materialized with the growing importance of the role China plays in the overall global economy. China serves as a critical counter-party, whose systemic risk exposures through capital adequacy impacts trade-partners, global investors, and foreign economies. Thus, improving risk management and forecasting of capital components in Chinese banks can help prepare stakeholders to prepare for appropriate reactionary strategies. However, practical challenges of attaining sufficient data volume, integrity, and quality to model using advanced analytical techniques exist.

This chapter introduces the transferable knowledge for the bank capital components (TKBCC) model. The TKBCC (1) acquires the pertinent banking knowledge from a source financial banking system that has regulatory transparency and firm-level financial data availability. The TKBCC then (2) leverages this knowledge as a foundation to forecast banking capital components relevant to bank stress-test analytics on a target financial banking system that may not have sufficient firm-level financial data available for public consumption.

4.2 Problem Formulation

As established previously in chapter 3, firm-level balance sheet projections of components relevant to capital adequacy calculation while considering economic conditions is a crucial aspect of bank stress-testing exercises and analytics. However, not all foreign banking systems provide the same amount of regulatory transparency in terms of public disclosures, firm-level banking characteristics and performance statements, or details into regulatory supervision. Some foreign nations may not have a sufficient amount of large-bank data or instances to provide reliable generalizations for modeling purposes. Regardless of the limitations in data, systemic risks from foreign banking systems and counter-parties play a crucial role in mitigating potential global financial crises. Thus, the importance of understanding banking performance under dynamic economic conditions is still an important task to achieve in foreign banking systems that may have regulatory obscurity or limited banking instances for modeling.

Assuming that banking strategy and operation ultimately follow some underlying core principles, finding ways to utilize the fundamental information is vital for particular scenarios. For example, a banking system that provides an abundant amount of publicly available information through regulatory transparency may allow for training examples sufficient for modeling. These banking systems can help similar systems that have an insufficient amount of training data available to complete an estimation task.

Techniques in deep learning to utilize potentially pertinent information from a source domain, D_s, on a related domain, D_t, to address a related task, T, are classified under the transfer learning category (L. Y. Pratt, 1993). For this dissertation work, the D_s is the banking system that offers firm-level regulatory transparency sufficient to model and forecast capital components that are crucial for calculating capital adequacy, or the U.S. bank holding companies. At the same time, the D_t is the Chinese banks representing a central counter-party banking system that is foreign to the U.S. banking system, which may not have sufficient data volume or transparency to facilitate robust modeling. The T is a time-series estimation with exogenous inputs of the bank capital components which are relevant to calculating regulatory capital adequacy measures (He, Pang, & Si, 2019).

Furthermore, a domain, D, is defined by two elements, the feature space, X, and the marginal distribution, $P(X)$, where X is a sample data point. Formally a domain can be denoted as shown in equation 4.1 (Behbood et al., 2011).

$$\begin{gathered} D = \{X, P(X)\} \\ \chi = \{x_1, x_2, ..., x_i\}, x_i \in X \end{gathered} \tag{4.1}$$

Where x_i is a specified vector within the X variable, for this work, X represents the banking characteristics, banking performance, and economic conditions for the respective domain.

The task T can be formally defined as two elements consisting of the target variable label space, Y, and the objective function,η, which is tasked with learning the non-linear modeling weights to capture the variability that can then help predict Y with the information provided in the D or can be denoted as $P(Y|X)$. Formally, the T is defined in equation 4.2.

$$\begin{gathered} T = \{Y, P(Y|X)\} = \{Y, \eta\} \\ Y = \{y_1, y_2, ..., y_n\}, y_i \in Y \end{gathered} \tag{4.2}$$

Where η learns from the feature vector and label pairs, $(x_i, y_i), x_i \in X, y_i \in Y$, to then predict the corresponding labels using the domain feature vectors, $\eta(x_i) = y_i$.

In this dissertation work, the label space, Y, reflects the target capital components dependent variable modeled using the feature vector, X. Moreover, the task will be trying to learn and predict numerical values as financial data in this domain is quantitatively measured. The T for this section is similar to the problem statement described in the equation 3.2 of chapter 3 concerning the bank capital & loss prediction task.

Given the defined domains, $[D_s, D_t]$ and their respective tasks, denoted as $[T_s, T_t]$, the problem objective is to learn the conditional probability distribution of T_t from D_t with information gleaned from the related D_s and T_s. Formally, the transfer learning objective is denoted in equation 4.3

$$\begin{gathered} P(Y_T|X_T \in D_t, P(Y_S|X_S) \in D_s) \\ D_s \neq D_t, T_s \neq T_t \end{gathered} \tag{4.3}$$

The underlying assumption in this problem definition is that the labeled target domain observations, Y_t, is significantly limited when compared to the corresponding source domain's label observations, Y_s (Behbood et al., 2011). This stipulation aligns with the problem in this dissertation work as the U.S. banking system has a large number of relevant observations for the task publicly available. In contrast, the Chinese banking system has significantly lower amounts of instances for a similar task.

4.2.1 Challenges

To improve the performance of estimation tasks, similar to equation 3.2 and 3.1, combined with features described in equation 1.1 for the bank stress-test problem on a foreign financial economy, the consideration of practical challenges is necessary. Interpretation of the feature space available to standard banking characteristics, modeling with an insufficient or limited amount of labeled training observations to obtain a robust generalization that captures the variability in target banking capital components are challenges that face this problem scenario.

1. Identify features from bank characteristics and performance dimensional vectors that have a similar contextual basis in the banking domain to help understand underlying strategy and operation patterns regardless of domestic, national, or international aspects of the domain space.

2. Determine the appropriateness of applying transfer learning techniques to a target domain with limited labeled observation instances. Effectively, understand if the transfer of knowledge improves or hinders the forecasting task. Furthermore, assess if the source & target domain and task are related enough to proceed with the approach.

3. Consider and interpret aspects of the potential negative, positive, and neutral transfer, which may have direct impacts on the overall performance of forecasting. Moreover, interpret which model parameters may have beneficial effects when modeling to provide insight into transferable components among banking systems potentially.

4. Assert the overall model utility using transferable knowledge from a related banking system domain for a forecasting task for bank stress-test analytics.

Implementation of an approach that addresses the challenges discussed in equation 4.3 while performing the problem task similar to the equation 3.2 and 3.1 may improve the performance of bank capital components forecasting. Banking systems with limited relevant regulatory data for the bank stress-test analytics may benefit from models that incorporate techniques that consider the challenges mentioned above.

4.3 Related Work

Research literature in transfer based learning with neural networks for deep learning applications stemmed from the work in (L. Y. Pratt, 1993), which formulated the discriminability-based transfer algorithm that showed indications of faster learning when compared to learning for randomly initialized neural networks. Soon after, machine learning literature dedicated a special issue, (L. Pratt & Thrun, 1997), which focused on the works dealing with inductive transfer learning. The field of transfer learning had become an increasingly popular research area and grew to include multi-task learning, which learns to solve tasks simultaneously (Caruana, 1997). Theoretical foundations were formally defined and researched (Baxter, 1998) to help the progression of transfer learning from state of the art technique to an area of science that has categories, structure, and methodology that can be empirically and theoretically proven. Overlap of literature from research areas of domain adaptation and multi-task optimization has also helped the overall prominence of transfer learning.

In deep learning literature, transfer learning techniques are for simulation learning, image learning, computer vision, speech recognition, machine translation, and natural language processing tasks (Pan & Yang, 2009) due to the high availability of academic data-sets for domain related tasks for each respective research area (Yosinski, Clune, Bengio, & Lipson, 2014).

However, transfer learning on sequential data that may have auto-correlated aspects, such as time-series data, is a research area that is not as rich but continues to emerge. In the work (Fawaz, Forestier, Weber, Idoumghar, & Muller, 2018), the authors investigate using transfer learning techniques on the publicly available time-series data sets from the University of California Riverside's Time Series Classification Archive (Chen et al., 2015). The authors use dynamic time warping measures to find the highest similarity between the source and target domains to match the most related domains for best performance.

In the work of (Laptev, Yu, & Rajagopal, 2018), the authors explore the use of convolutional neural networks, typically used in image learning tasks, to learn the line charts of time series data. The authors assess the usefulness of the knowledge on target domains for time series forecasting and classification tasks.

The work of (Qureshi, Khan, Zameer, & Usman, 2017), the authors develop an ensemble deep learning technique leveraging auto-encoders and deep belief networks. They utilize transferred model weight initialization from wind power stations to then predict the wind power prediction at a target station based on a similar feature space.

In the healthcare domain, transfer learning techniques to predict metrics important to hospital effectiveness is shown (Gupta, Malhotra, Vig, & Shroff, 2018). The authors leverage Recurrent Neural Networks trained on an academic critical care database (Johnson et al., 2016) to learn generic features about patients to use as an initialization point to robustly predict in-hospital mortality rates.

Works in the area of transfer learning for financial services or banking-related tasks are very few due to the relatively recent emergence of deep learning for financial applications literature. In the work of (Behbood et al., 2011), the authors propose fuzzy refinement, which leverages transductive transfer learning techniques to improve the performance of bank failure prediction of U.S. firms. The recent work of (He et al., 2019) utilizes two source domain data-sets to make useful financial time-series predictions in the target domain and shows that their approach outperforms traditional techniques.

Opportunities to improve the effectiveness of transfer learning techniques on relevant tasks are related to the negative transfer, or the transfer of irrelevant knowledge from sources, and imbalanced distributions, where the training samples may not be well represented, thus causing a bias (Pan & Yang, 2009). In the work of (Ge, Gao, Ngo, Li, & Zhang, 2014), the authors discuss a two-phase approach that first employs a sophisticated supervised weighting scheme on source domain observations to determine effects on predicting local regions on the target domain. Next, the approach performs an optimization that simultaneously minimizes training error while maintaining useful weighted prediction characteristics achieved from the first phase. The authors can show that the performance of their approach is effective because of how it can efficiently handle additional source domains and how it addresses potential negative transfer.

Furthermore, the work of (Rosenstein, Marx, Kaelbling, & Dietterich, 2005) focuses on both detecting and avoiding potential negative transfer scenarios that could hinder the desired performance of transfer learning by using hierarchical Bayesian methods. They were able to determine that the source and target domain task dissimilarity plays a vital role in the negative transfer and is an important aspect when applying transfer learning techniques. The authors conclude that clustering techniques on source data could potentially identify the most relevant source domain data that could help enhance the performance of the target domain's task.

The technical aspects of transfer learning beyond the consideration of relevant source domain knowledge include the actual source model weights in the target domain. In (Raina, Battle, Lee, Packer, & Ng, 2007), the authors develop high-level features from unlabelled data using sparse coding. The features ultimately provide an input representation that can be well distinguishable for transfer learning or self-taught learning settings. Additionally, in the work of (Mesnil et al., 2011), discussion into the feature representation known as the "triangle code," which relies on K-means and specialized pre-processing methods. The authors also indicate that the use of principal components in the transductive transfer learning setting performs best as both the first and last layers. The work also thoroughly reviewed other methodologies that try to leverage feature representations for a transfer learning task.

Generally, the parameters approach is the basic methodology of transferring knowledge from the source domain to the target domain to perform a related task (Pan & Yang, 2009). As seen in works like (Al-Sahaf, Zhang, & Johnston, 2016), which uses the parameters approach in their transfer learning framework for a binary image classification task. The recent work of (Alawad et al., 2019) also leverages the parameters approach on cancer registries to assert the improvements to performance that are possible.

4.4 Methodology

For bank capital components prediction of a foreign banking system with limited regulatory reporting data, methodologies from transfer learning techniques help provide a potential modeling framework. Specifically, an investigation into the effectiveness of identifying relevant knowledge from a related banking system with sufficient training examples that can benefit a robust forecasting model for a related task with limited training examples.

To address the practical challenge of improving the bank capital-components prediction task, similar to those mentioned in equations 3.2 and 3.1, the methodologies to be implemented into the model framework must consider the contextual similarities and differences in the feature space. Additionally, the framework must address the lack of relevant training examples in the domain space to provide a robust model that can capture the variability of potentially unforeseen circumstances. Limited training examples in this domain setting can cause the model to not correctly learn potential banking performance reactions to dynamic economic conditions appropriately. Deep learning literature has continued to foster transfer learning techniques that can help provide appropriate solutions for circumstances where training examples are insufficient. However, related domains may have relevant insight that could help in performing a problem task (Pan & Yang, 2009) by using specific methodologies, which may amount to deep learning optimizations, to improve overall model utility.

The domain setting in question forecasts bank capital-components using knowledge from a related domain using relevant financial data from the U.S. banking system as a source domain and Chinese banking system as the target domain. Ultimately, this equates to a time-series forecasting task using transfer learning techniques that are compatible with available labeled data in both the source and target domains but have limited training examples in the target domain (Pan & Yang, 2009).

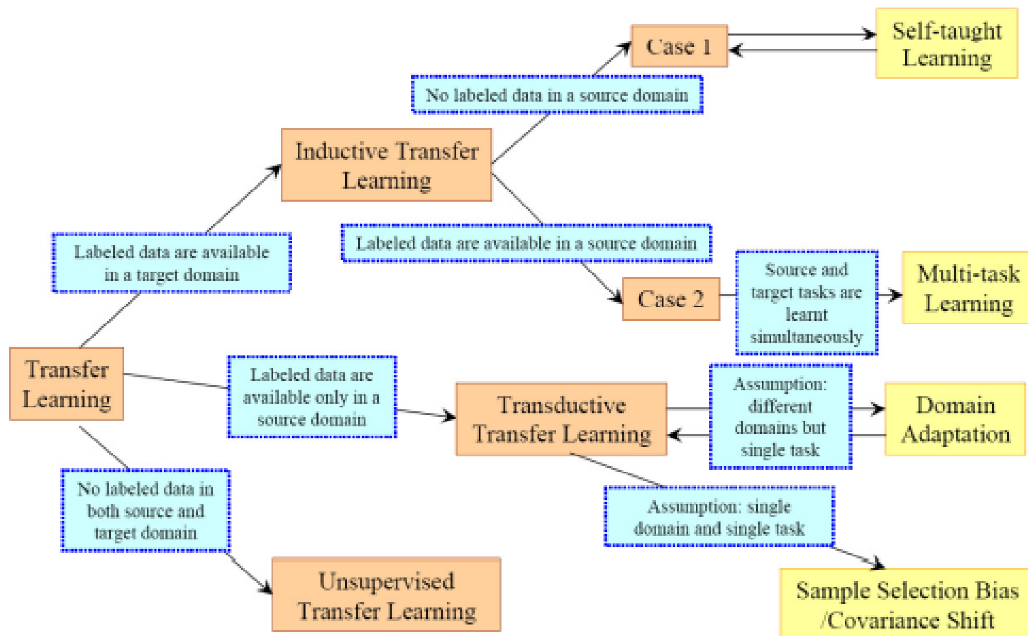
The general methodology of transfer learning techniques, which will obtain existing knowledge from a source learning model for application for a target domain task requires the holistic understanding of the following criteria (Sarkar, Bali, & Ghosh, 2018):

1. Identifying relevant aspects of knowledge elements from the source domain that can help enhance the performance to accomplish the target domain task. Moreover, isolating the source knowledge aspects that may be most similar or useful between both domains.

2. The applicability of transferring knowledge for this problem needs assessment. If the performance does not improve due to latent dissimilarities between the source and target domain, it may not be prudent to use transfer learning for this application. Negative transfer refers to obtaining irrelevant information from the source domain to solve the target domain's task, which may degrade the overall performance when compared to efforts to approach the task without transfer learning.

3. If the previous two criteria are applicable and the setting is deemed appropriate for transfer learning, the consideration of technical methodologies to accomplish

Figure 4.4. Transfer Learning Strategies



the transferring of the relevant knowledge from the source domain to the target domain learner is necessary. Typically, advanced modifications to existing algorithms or model frameworks are necessary by using specific techniques that allow for transference of latent information, which deep learning models can contextualize as fundamental knowledge to build on top of at the target task.

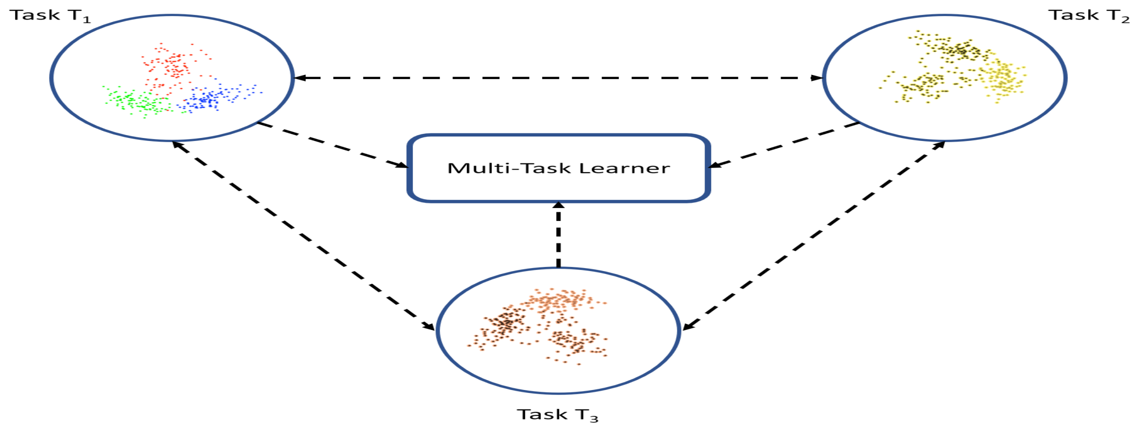
In this section, further elaboration in regards to the background of inductive transfer learning, multi-task learning, feature-representation transfer, parameter transfer and negative transfer will be illuminated to highlight key aspects that are crucial to using transfer learning to determine meaningful knowledge for bank capital components prediction. Additionally, an introduction of a proposed model framework, which will consider the mentioned aspects to provide a practical solution for the problem task.

4.4.1 Inductive Transfer Learning

Depending on the availability of data, the relationship between the source and target domain, and the similarity between the source domain task and the target domain task, specific transfer learning strategies may be the most applicable (Pan & Yang, 2009). The strategies, such as inductive transfer learning, unsupervised transfer learning, and transductive transfer learning, are generally categorized according to the traditional machine learning algorithm they may leverage (Sarkar et al., 2018). The category of methods that align with the problem task, as defined for this dissertation work's domain setting, is inductive transfer learning.

Inductive transfer learning refers to the circumstances when the source and target domains are mostly the same. In this case, the relation of both domains to the banking system, and their characteristics, which may help explain variability in the respective bank's performance, are different between the domains, due to the geopolitical aspects that may impact the banking industry for each respective financial system. In the inductive transfer learning strategy, design algorithms to leverage inductive biases from the source domain to enhance the ability to perform the target domain task (Pan & Yang, 2009). Moreover, this strategy can be sub-categorized if labeled data is present in the source domain. If labels are present, the sub-category of multi-task learning befits the problem. Otherwise, if labels are not present in this scenario, the approach belongs to the sub-category of self-taught learning (Sarkar et al., 2018). Additional details of transfer learning strategies are found in figure 4.4 from the literature (Pan & Yang, 2009). For the work in this dissertation, the focus is

Figure 4.5. Basic Multi-Task Learning Overview



predominately on the problem in which labeled data exists for both source and target domains through data selection. However, the potential of using unlabelled examples in both the source and target domains may help provide relevant information for the problem task.

Multi-task Learning

In contrast to traditional transfer learning, multi-task learning aims to simultaneously learn multiple tasks without defining source or target domains (Caruana, 1997). Thus, this learning algorithm acquires relevant information on how to achieve multiple tasks at once, as shown in figure 4.5, which contrasts to the sequential nature of transfer learning where the source domain task is initially learned before the target task is considered (Sarkar et al., 2018).

Instance transfer

Ideally, the use of knowledge gained from the source domain for the target domain task is the overall goal of transfer learning. Often, however, the source domain data cannot be directly used for the target task because of structural or dimensional incompatibilities. Interestingly, there may be specific instances from the source domain that can potentially be used with the target domain's training data to yield better performance. The instance transfer approach identifies and uses potentially meaningful instances from the source domain with the target domain's training data to accomplish favorable target task performance (Sarkar et al., 2018). Recently, in the area of inductive transfer learning, instance transfer techniques are popularly used with modified AdaBoost techniques, as first seen in (Dai, Yang, Xue, & Yu, 2007).

Feature-Representation transfer

Minimization of divergence and error rates among the feature space data distributions by obtaining feature-representations that can best align the source and target domains are the key goals of feature-representation transfer (Sarkar et al., 2018). Effectively, the method seeks to reduce negative transfer, irrelevant knowledge elements, from the source domain to predict the target domain task by finding the similarity in the feature-space data distribution structures and compatibility in training. This transference approach works for both supervised and unsupervised strategies of transfer learning (Pan & Yang, 2009).

Parameter transfer

Assuming that designed models for related tasks share latent parameters and prior distributions of deep learning model hyper-parameters. Transferring model parameter weights from a source domain model designed for a task related to the target domain task could serve as an important initialization point for the target model, rather than initializing from random (Sarkar et al., 2018). Additionally, parameter weighting can be modified to optimize the loss functions of the target domain task, which enhances performance by effectively favoring positive knowledge transfer elements.

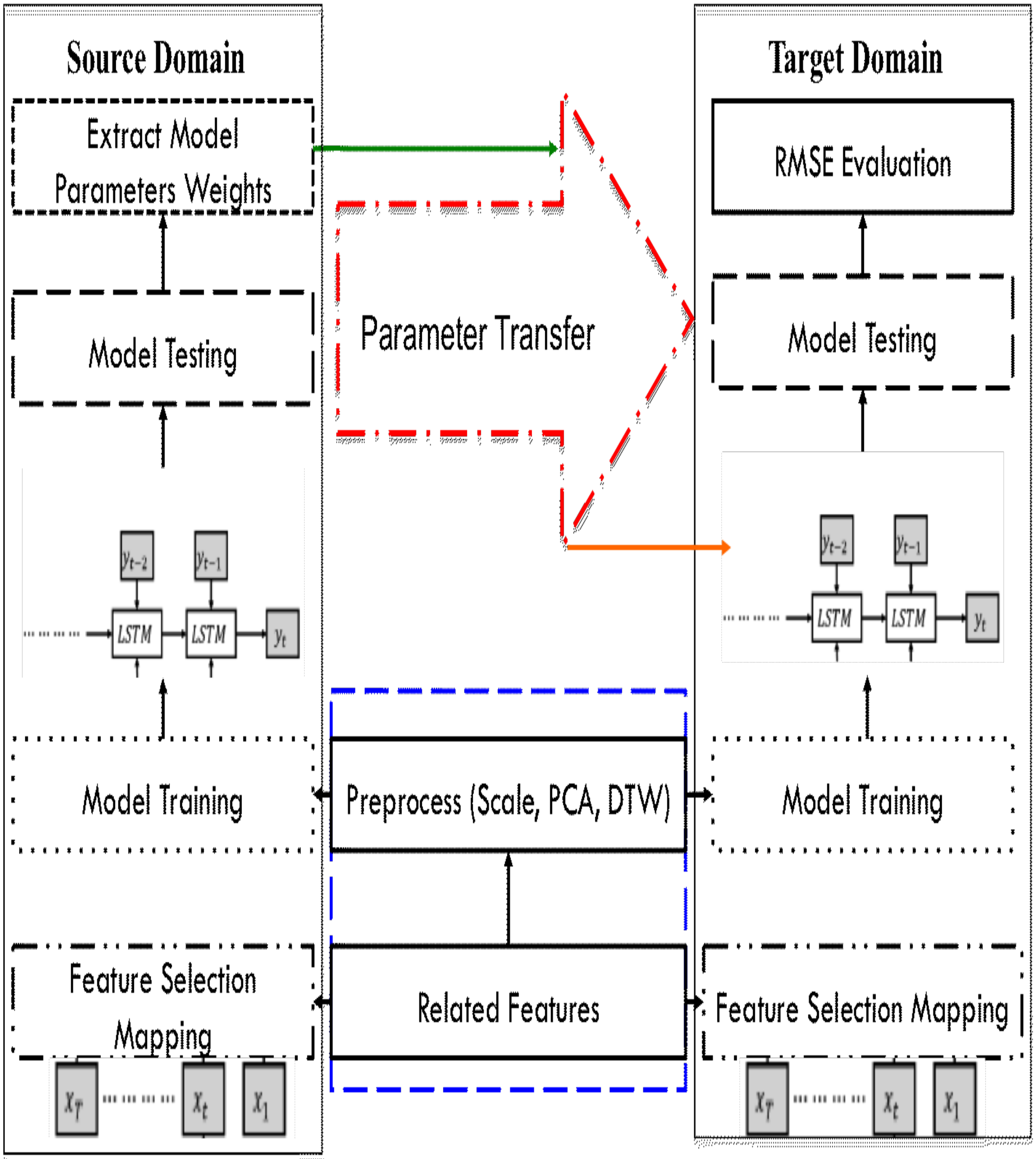
Relational-knowledge transfer

Consideration of non-independent and identically distributed data, which may have co-linearity or dimensional dependence with one another, is an important aspect to consider when transferring knowledge from a source domain which may have data instances that have relationships (Sarkar et al., 2018). Social network data typically would benefit from using relational-knowledge-transfer techniques due to the highly related and dependent characteristics of the data that may exist in the training data.

4.4.2 Transferable Knowledge for Banking Capital Components (TKBCC)

In this dissertation work, exploration of inductive transfer learning techniques to address the problem of using a model tasked with predicting banking capital-components for the source banking domain to transfer the relevant knowledge elements on to a model tasked with predicting similar in a target banking system. The proposed transferable knowledge for the banking capital components (TKBCC) model framework is a multi-stage approach. (1) Use of similarity string matching based on Levenshtein distance ratios (Levenshtein, 1966) and Euclidean distance metrics (Goodfellow et al., 2016) to find the similarity among the feature space between the source and target domain to isolate the most relevant financial characteristics of the banks. (2) Identification and acquisition of transferable components while accounting for potential for negative transfer (Sarkar et al., 2018). Figure 4.6 depicts the proposed TKBCC model framework. This model framework will analyze financial time-series data con-

Figure 4.6. TKBCC Framework



sisting of economic conditions, banking characteristics, and banking performance to make robust time-series forecasts of banking capital components.

4.4.3 Related Domain Feature Mapping

As previously mentioned, identifying relevant aspects of knowledge elements from the source domain can help enhance the performance abilities to accomplish the target domain task. Moreover, isolating the source knowledge aspects that may be most similar or useful between both domains. Given the diverse nature of standardized accounting practices across foreign nations, it is necessary to identify the most similar financial statement dimensions based on contextual information from variable descriptions. Fortunately, accounting syntax and semantics are generally universal, therefore using sophisticated distance metrics to find similarities among variable descriptions across domains will help to focus on the most applicable features for the transfer learning task.

Variable Description Similarity based Feature Selection

Simple string similarity is measured across all of the source and target domain variable descriptions using spatial distance programming libraries found in (Pedregosa et al., 2011) by computing the distance between each pair of the the dimensional variable descriptions as inputs. The Euclidean distance metric is used to determine the similarity between the variable descriptions in the two domains using the the commonly used definition found in equation 4.4.

$$\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{4.4}$$

Table 4.1. U.S. & Chinese Bank Sector Similarity Based Feature Mapping

China (CSMAR) Description	CSMAR Code	US (FRB Description	FRB Code
Total Assets	A001000000	Total Assets	BHC02170
Net Account Receivable	A001111000	Accounts receivable	BHSPA024
Net Goodwill	A001220000	Goodwill	BHCK3163
Short-Term Borrowings	A002101000	Short-term borrowings	BHSP3150
Accounts Payable	A002108000	Accounts Payable	BHSP3066
Deferred Tax Liabilities	A002208000	Net deferred tax liabilities	BHCK3049
Other Comprehensive Income	A003111000	Other Comprehensive Income	BHCKB511
Net Interest Receivable	A001119000	Accrued interest receivable	BHCKB556
Deferred Tax Assets	A001222000	Net deferred tax assets	BHCK2148
Trading Financial Liabilities	A002105000	Trading liabilities, total	BHCK3548
Interests Payable	A002114000	Accrued interest payable	BHSP3166
Taxes Payable	A002113000	Income taxes payable	BHSPC257
Notes Payable	A002107000	Subordinated & special-purpose subsidaries notes payable...	BHCKC699 + BHCPC255
Dividends Payable	A002115000	Dividends declared but not yet payable	BHSP2932
Minority Interests	A003200000	Minority Interests	BHCK4484
Consolidated Income Attributable to Minority Shareholders	B006000102	Net income (loss) attributable to bank...	BHCKG104
Assets Purchased under Agreements to Resell	A0F1122000	Securities purchased under agreements to resell	BHCKB989
Other Assets	A0F1300000	Other assets	BHCK2160
Assets Sold under Agreements to Repurchase	A0F2110000	Securities sold under agreements to repurchase	BHCK0279
Other Liabilities	A0F2300000	Other liabilities, total	BHCP2930
Net Reinsurance Receivable	A0I1114000	Reinsurance recoverables	BHCKC247
Net Fees And Commissions Income	B0D1104000	Investment banking.. fees and commissions	BHCKB490 + BHCKC886 + BHCKC887
Total Operating Revenue	B001100000	Parent total operating income from associated banks...	BHCP0520
Other Operating Expenses	B0F1213000	All other operating expenses	BHCP0522
Other Operating Revenue	B0F1105000	All other operating income	BHCP0447
Retained Earnings	A003105000	Undivided Profits And Capital Reserves (Retained Earnings)	BHCT3247

Essentially, Euclidean distance sums up the cumulative squared distances between corresponding data points and gets the square root of the result, effectively solving for the more direct distance between the point x and y.

Additionally, the Levenshtein similarity ratio (Levenshtein, 1966) is the spatial distance calculation used to provide a variance vector to standardize the Euclidean distance calculation. The Levenshtein distance is a specialized metric that measures the divergence between two sequences of words. Specifically, it measures the amount of edits, such as insertions, deletions or substitutions, necessary to change one sequence to match the other (Miller, Vandome, & McBrewster, 2009). The Levenshtein distance can be depicted in equation 4.5 and the respective similarity ratio in equation 4.6.

$$lev_{(a,b)}(i,j) = \begin{cases} \max(i,j), & \text{if } \min(i,j) = 0, \\ \min \begin{cases} lev_{(a,b)}(i-1,j)+1 \\ lev_{(a,b)}(i,j-1)+1 \\ lev_{(a,b)}(i-1,j-1)+1_{a_i \neq b_j} \end{cases} & \text{otherwise.} \end{cases} \tag{4.5}$$

Where $1_{a_i \neq b_j}$ is 0 when $a = b$, otherwise it is 1. The conditions in the min case refer to deletions, insertions or substitutions of the string sequence.

$$\frac{(|a|+|b|) - lev_{(a,b)}(i,j)}{|a|+|b|} \tag{4.6}$$

Where $|a|$ and $|b|$ are the lengths of string sequence a and string sequence b respectively.

Procedure 3 Training for TKBCC Model Framework

Input: $\mathbf{T_s}([\mathbf{X_s}, \mathbf{Y_s}])$ where $\mathbf{x_s} \in \mathbf{D_s}$, $\mathbf{X_t}$, Y_s where $Y_s \in D_t$

Output: prediction y_t'

if *TTC.type() = 'Parameter'* **then**

  $T_t.parameters() = T_s.TTC().parameters()$

  for $j_t \in$ $\mathbf{T_t}$**.parameters()** ▷ *Initialize Transfer* **do**

    $j_t.RequiresGradient \leftarrow False$ ▷ Freeze parameters

  end

  $T_t(x_t).OutputSize = dim(y_s)$ ▷ Modify Input, Output Layers to target

  $X \leftarrow Linear.forward(\mathbf{X_t}, T_s(x_s).InputSize)$

end

else if *TTC.type() = 'Instance'* **then**

  if $(x_s, x_t).similarity \geq \alpha$ **then**

    $X = X_t$.append(x_s) ▷ Add instance to target domain

  end

end

else if *TTC.type() = 'Feature-Representation'* **then**

  $X_s = X_s.PCA(n), X_t = X_t.PCA(n)$ ▷ Get n principle elements

  if $X_s \approx X_t$ **then**

    $X = X_t$.append(X_s) ▷ Add Features to domain

  end

end

while *Convergence on* (θ, ϕ) *or* $i = EPOCH$ **do**

  $\hat{y} \leftarrow T_t(x_t).forward(X, y)$ ▷ prediction

  $Loss \leftarrow MSELoss(y, \hat{y})$ ▷ calculate loss

  $Loss.backward()$ ▷ propagate gradients

end

return prediction y_t'

4.4.4 Transference of Knowledge

Using inductive transfer learning strategies on the problem task by setting up a model framework that can investigate the types of transferable components that may be most effective in improving the overall performance is the aim of the TKBCC model framework. As previously mentioned, understanding the applicability of transfer learning for the problem setting needs to be determined by assessing the improvement in performance attained when applying techniques that may mitigate negative transfer. To determine the feasibility of transfer learning, the evaluation of different types of knowledge components to transfer is necessary. Instance-transfer, feature-representation transfer, parameter transfer, and relational-knowledge transfer need to be understood to determine if a more in-depth study is prudent.

The development of a model framework that incorporates source domain and task knowledge, $[D_s, T_s]$, for the target domain and task learning model, $[D_t, T_t]$. (1) The model framework will use features determined to be similar using the related domain feature mapping methods described previously. (2) The designed model framework is compatible with different types of transfer components to assess the effectiveness of each for the problem task, thus instance transfer, feature-representation transfer, parameter transfer, and relational-knowledge transfer techniques, $T_s.TTC()$ or the function that obtains the specified transfer type of transferable components from the source domain task. (3) The $T_s.TTC()$ is incorporated into the target domain task as a model initialization to be fine-tuned further with the application of a stochastic gradient loss function that propagates gradients specific to the target domain to make predictions based on transferred knowledge. The algorithm procedure 3 depicts the high-level training sequence of the TKBCC model framework.

The feature-representation aspects are developed as a model variant that will ultimately utilize parameter transfer of the model parameter weights. However, the model weights will be trained from a feature space that leverages principal components analysis (PCA). Briefly, PCA is an unsupervised learning technique that acquires information regarding the variance from a higher dimension feature space. PCA leverages orthogonal projection or transformation on the underlying feature space to yield a lower dimensional or latent subspace, which will have maximized the variance of the projected data (Wold et al., 1987). The ability of learning the feature-representation through PCA allows for visualizations due to reduced dimensions, processing efficiency by machine learning algorithms due to data quality improvement through noise removal, better generalization through fewer dimensions, and reduced computation resources required due to fewer dimensions. For this study, PCA will be used to reduce dimensions in both the source, D_s and target, D_t domain feature spaces, $[X_s, X_t]$, to twelve principal components, similar to chapter 2 equation 2.12 and depicted in equation 4.7, that will be then modeled in the source domain to generate the appropriate parameter weights to be transferred using the parameter transfer approach.

$$\begin{gathered}
\bar{w}_{j,n} = \frac{(w_{j,n} - \mu_i)}{\sigma_i} \quad \forall w_j \in X_{[s,t]} \\
\hat{w}_{j,n} = [P_1, P_2, ...P_{12}] = [\bar{w}_{j,1}, \bar{w}_{j,2}, ...\bar{w}_{j,n}] * \underbrace{W'}_{\text{Projection Matrix}} \\
\hat{X}_{[s,t]} = [\hat{w}_1, \hat{w}_2, ...\hat{w}_j] \quad \forall w_j \in X_{[s,t]}
\end{gathered} \tag{4.7}$$

The instance-transfer aspects in a model variant will also leverage parameter transfer of model weights. However, the trained model weights are from a subset of source domain instances, $\hat{D}_s$, which will be determined based on the similarity of the observations between the source and target domains, $\hat{D}_s \mid\mid\mid D_t$. Effectively, training data in the source domain will be limited to instances that resemble the target observations, which will potentially allow for mitigation of negative transfer while providing insight that may be most applicable for the sequence-based data for the target task. To determine the similarity of the sequences-data in the two domains, Dynamic Time Warping (DTW) (Müller, 2007) techniques will be used to determine which instances qualify to be in $\hat{D}_s$, similar to the work in (Fawaz et al., 2018). Distance measurements of entries of one sequence compared to the entry of another sequence will be determined using the simple numeric metric shown in equation 4.8

$$d(x, y) = |x - y| \tag{4.8}$$

Furthermore, given two sequences a_i, b_j, comparison of their local distance for a specially chosen set of indices, m_k for a_i and n_k for b_j. Thus, the dynamic time warping distance can be depicted as the quantity yielded from equation 4.9.

$$C(a_i, b_j) = \sum_{k=0}^{M} d(a_{m_k}, b_{n_k}) \tag{4.9}$$

However, constraints are required on the indices m_k, n_k to allow for interpret-able results. Under the assumptions, a_i has length M, b_j has length N, the constraints can be defined as the following:

Definition: A warping path for a_i, b_j is a pair of sequences (m_k, n_k), both with a length L, with following conditions:(1) $\leq m_k \leq M and 1 \leq n_k \leq N \forall k$. Without this condition the m_k, n_k is not able to index a_i, b_j (2) Sequence endpoints are $(m_1, n_1) = (1,1), (m_L, n_L) = (M, N)$. This condition makes sure that all the information in both sequences are compared. (3) Monotonic increase of m_k, n_k. This condition indicates that backward temporal steps are not allowed when comparing local sequence entries. (4) $m_k - m_{k-1}, n_k - n_{k-1}) \in [(1,0),(0,1),(1,1)]$. Lastly, this condition permits the indexing of one sequence while the other may be paused, ultimately creating the "time warping" effect" that allows for certain sections of a sequence to be consolidated or expanded when compared to the other sequence.

Additionally, the optimal "warping path" will be selected from the several available paths between the two sequences. To accomplish this, minimization of equation 4.9 is required, denoted as $DTW(a_i, b_j)$. A dynamic programming equation is required to solve the sub-problems and compute the optimal path, as depicted in equation 4.10.

$$DTW(a_i, b_j) = (a_{1\ldots M-1}, b_{1\ldots N}), (a_{1\ldots M}, b_{1\ldots N-1}), \text{ and } (a_{1\ldots M-1}, b_{1\ldots N-1}) \tag{4.10}$$

The derivation of transfer-learning based parameter-transfer for experimentation is from techniques found in (Sarkar et al., 2018; Al-Sahaf et al., 2016; Alawad et al., 2019). In this dissertation work, the source domain will be trained and tested on a model using the Pytorch python library (Paszke et al., 2017), which provides the experimenter the ability to export the model parameter weights as a dictionary that can serve as an initialization point for a different Pytorch model. Thus, the trained source model will have the model parameter weights transferred into the target model as initialization parameter-weights. The parameters in the target model will be frozen, requiring the target model to fine-tune model weights atop the initial parameters, ultimately utilizing both the source domain model weights as well as the finely tuned target domain trained model weights.

The model framework investigates the effectiveness of different inductive transfer learning techniques, including parameter transfer, instance transfer, and feature representation transfer. Currently, the model framework is not considering relational-knowledge due to the domain expertise and technical development required; however, future research directions may incorporate the technique.

In this dissertation work, the problem setting necessitates a method to acquire additional training samples that are relevant enough to help enhance the prediction performance of the target domain task. Given the current state of available financial data across global nations, there is a need to leverage rich information available in globally similar industries for forecasting local tasks. This model framework and methodology implement sophisticated deep learning techniques to help utilize available related data to improve the performance on tasks that have insufficient amounts of training examples. This designed framework also evaluates the relevance of transfer knowledge by providing insight into which technique may be more favorable for the problem as it relates to bank capital components prediction.

4.5 Experiment

Table 4.2. Transferable Knowledge for Capital Components Experiment Setup

Experiment	Train	Test	Target Variable
Exp. 1	2003Q1-2015Q4	2016Q1-2017Q4	$Y_{RetEarnings_t}$

In this section, the implementation of compatible popular RNN models, such as LSTM, GRU, and Linear Regression into the proposed transfer learning model framework, TKBCC, will be explored. Assessment of the effectiveness of different transfer techniques on the problem task of predicting bank capital components of a foreign banking system that has insufficient training data. Experimentation on real-world financial and economic data representing the U.S. and China to validate the benefits of the model framework.

Evaluation of each model technique to obtain relevant knowledge elements and address negative transfer challenges for time-series with exogenous terms prediction tasks for a financial problem setting can provide meaningful insights for both transfer learning and financial forecasting literature.

Discussion on the validation of performance through model comparison among transfer learning techniques, deep learning techniques without transfer learning, and linear techniques take place in this section. Measuring the performance of the effectiveness at the problem task as well as understanding the improvement achieved when using a specific transfer learning technique are relevant to provide illumination to the utility of each model.

4.5.1 Data Description

Acquired data consists of U.S. Bank holding companies' consolidated financial statements (FR-9YC) for over 20k banks since 1986 - 2017 from the WRDs bank regulatory data repository (WRDS, 2019). Also obtained are Chinese bank financial balance sheets and income statements from the CSMAR data repository offered by WRDS

(WRDS, 2019). Additionally, data retrieval of historical economic conditions data for the U.S. from the federal reserve's (Reserve, 2015) stress -testing exercises. In contrast, the data for the historical economic conditions of China is from the National Bureau of Economic Research's data repository on China (Higgins & Zha, 2015).

Data Selection

The features used from each respective banking system, China and the United States, are selected using the domain-related feature mapping, referenced in the methodology section of this chapter, and equations 4.6 & 4.4. The feature space obtained from the domains, which were related, is summarized in table 4.1. The banking capital component to be predicted, Y, will be represented by firm-level retained earnings, which is a critical component to calculating banking capital, as shown in equation 1.6 and 1.5.

All numerical data is normalized using a min-max scaling technique discussed in (Pedregosa et al., 2011) and denoted in equation 2.11.

Furthermore, the economic conditions, depicted by macroeconomic indicators, are summarized in table 4.3, which have been scaled and reduced in dimensions using the principal components analysis method, as previously discussed in equation 2.12 to 60 quarters between 2003 and end of 2017. The original 28 macroeconomic condition values of $US_{macro,PCA}$ prior to normalization and dimension reduction can be found in tables 2.2 and 2.3. The original 55 values of $CN_{macro,PCA}$ before scaling and dimension reduction is from the National Bureau of Economic Research's data repository on China's macro-economy (Chang, Chen, Waggoner, & Zha, 2016; Higgins

& Zha, 2015).

Lastly, data anomalies and outliers are handled using Standard Deviation Method (Leys et al., 2013) on the domain banking variables by transforming values that are beyond four standard deviations from the mean of the respective dimensional vector to the closest non-outlier neighbor value. Additionally, interpolation of missing data using linear methods found in the python pandas libraries (McKinney, 2010). Using the methods mentioned above allows for the usage of potentially useful data for the experiments while assuming that data quality issues are not representative in the full data set, but can be generalized for the experiment setting.

Similar to the selection of banks in chapter 3, U.S. banks will be selected based on their consolidated total assets and existence of at least eight consecutive quarters to address potential bias, as per (Malik, 2018). Over 2000 banks are sampled and filtered for training in the source domain.

Chinese banks will be selected based on their sector, "Finance," and industry, "Bank," which will net 90 training samples for the target domain.

Lastly, economic and banking dimensions missing significant amounts of data, more than 100 missing rows, will be dropped, and missing data otherwise will be linearly interpolated using methods in (McKinney, 2010).

Temporal Considerations

Given the auto-correlative nature of financial data, features will be derived, consisting of up to four time-lags. The time-lags provides the feature space relevant context, which can be useful in RNN modeling of sequential input data. Four time-lags rep-

Table 4.3. US & China Principal Components for Macro-Economy (2003-2018)

Economic Domain	**Principal Component**	**Mean**	**Std.**	**Min**	**Med.**	**Max**
$US_{macro,PCA}$	1	-3.36	1.17	-6.40	-3.05	-1.72
	2	0.29	2.08	-2.99	-0.10	8.20
	3	-0.91	1.47	-3.40	-1.11	4.97
	4	0.06	1.19	-2.25	0.45	2.21
	5	-0.15	1.41	-3.84	-0.12	4.45
	6	0.30	1.26	-3.27	0.21	3.79
	7	0.09	1.19	-2.23	0.09	2.95
	8	-0.04	0.84	-1.74	-0.13	2.24
	9	-0.22	0.91	-2.04	-0.35	1.59
	10	0.06	0.78	-2.19	0.08	2.39
	11	0.02	0.81	-2.01	0.11	2.50
	12	0.00	0.86	-1.60	-0.07	2.19
$CN_{macro,PCA}$	1	3.68	4.43	-3.20	3.39	10.91
	2	0.38	1.73	-2.25	0.09	3.51
	3	0.53	1.41	-2.29	0.56	5.01
	4	-0.65	0.91	-2.37	-0.71	1.25
	5	0.19	0.91	-1.47	0.24	3.88
	6	0.01	0.72	-1.24	-0.16	1.79
	7	-0.07	0.84	-2.05	-0.12	2.81
	8	0.12	0.62	-1.21	0.16	1.27
	9	-0.04	0.46	-1.16	-0.08	1.44
	10	-0.05	0.40	-0.78	-0.13	0.94
	11	0.01	0.43	-0.93	-0.04	0.98
	12	0.03	0.41	-1.20	0.01	1.33

resent the previous four quarters of each economic condition, banking characteristic, and banking performance feature space, which can capture recent trends or provide appropriate context to detect patterns of dynamic behavior.

4.5.2 Transferable Knowledge for Capital Components Prediction

Discussions on the experimentation and investigation of the utility of the parameter transfer approaches are prevalent in this dissertation work. Other transfer approaches, such as instance-based transfer, feature-representation transfer, and relational knowledge transfer, will be examined shortly. However, variants of the parameter transfer approach will try and capture instance transfer and feature-representation through dynamic time warping similarity (Müller, 2007), Euclidean distance, as seen in equation 4.4, and principal components analysis (Wold et al., 1987).

Experiment Setup. The target variable for this experiment, Y_{T_t}, will be modeled at continuous quarterly intervals and be represented by bank retained earnings. Furthermore, to maximize the utilization of the source training data for target testing tasks, the experiment training data for both domains will be the following: (Train 2003 Q1-2015 Q4, Test 2016 Q1 -2017 Q4). This training and testing split to the economic conditions, banking characteristics, and banking performance dimensions, is summarized in table 4.2. Finally, the learning rate for the stochastic gradient descent is 1e-2, however, during training, if the difference in the loss from the previous iteration is not higher than the threshold 1e-3, the learning rate will be divided by a factor of 10. It will continue with the same threshold rule. Once the learning rate reaches the lower learning limit of 1e-5, the training process will stop. The implemented procedure trains the model effectively without missing any potential opportunity to fine-tune it accordingly.

Evaluation Metrics. For the bank capital-components prediction task, root mean squared error (RMSE) as per equation 2.13 is used to compare how well the predicted estimates match that of the testing set and compare the effectiveness of each model across both domains. In this experiment, Y_i is the bank retained earnings, $\hat{Y}_i$ is the respective estimated value, and n is the number of testing observations. Specifically, lower RMSE values indicate superior model performance.

Additionally, to measure the effect transferred knowledge and the transfer learning technique used when compared to a baseline algorithm that does not use transferred elements, the traditional "percent error" formulation will be used. Essentially, the metric will measure the improvement in terms of relative proportionality the transfer learning based model's performance had when compared to it's non-transfer learning counterpart. The formula is summarized in equation 4.11

$$Improve\% = \frac{RMSE_{noTransfer} - RMSE_{withTransfer}}{RMSE_{noTransfer}} * 100 \tag{4.11}$$

Where $RMSE_{noTransfer}$ refers to an RNN model that does not utilize any transfer learning components to accomplish the prediction task, while $RMSE_{withTransfer}$ refers to a RNN model that does leverage transfer learning as an initialization point that is then fine tuned to accomplish the prediction task.

Baseline Algorithm. The evaluation performed on bank capital-component prediction, and transfer-learning effectiveness consists of comparing the proposed RNN based TKBCC model with a set of baseline models. (1) Long Short Term Mem-

Table 4.4. Chinese Bank Feature Space (2003-2018)

Variable Code	Mean	Std.	Min	Med.	Max
A001000000	8.90E+11	1.88E+12	5.00E+08	2.64E+10	6.91E+12
A001109000	2.37E+07	8.02E+07	0.00E+00	0.00E+00	4.04E+08
A001119000	4.41E+09	9.43E+09	0.00E+00	7.71E+07	3.54E+10
A001220000	5.94E+08	1.51E+09	0.00E+00	5.82E+06	6.98E+09
A001222000	2.24E+09	5.28E+09	0.00E+00	5.38E+07	2.09E+10
A002101000	4.27E+08	9.89E+08	0.00E+00	1.44E+07	4.16E+09
A002105000	3.37E+09	8.90E+09	0.00E+00	0.00E+00	3.80E+10
A002107000	1.64E+07	4.44E+07	0.00E+00	0.00E+00	1.87E+08
A002208000	1.91E+08	4.06E+08	0.00E+00	1.10E+07	1.63E+09
A003102000	1.73E+10	2.78E+10	1.42E+07	2.74E+09	9.88E+10
A003105000	1.84E+10	3.76E+10	-6.75E+08	1.39E+09	1.42E+11
A003111000	-3.41E+06	1.30E+09	-3.54E+09	1.01E+07	4.29E+09
A003200000	1.11E+09	2.25E+09	0.00E+00	8.01E+07	8.94E+09
A0F1122000	4.10E+10	8.91E+10	0.00E+00	1.17E+09	3.45E+11
A0F1300000	1.98E+10	3.77E+10	0.00E+00	1.94E+09	1.57E+11
A0F2110000	1.87E+10	3.14E+10	0.00E+00	3.00E+09	1.17E+11
A0F2300000	1.67E+10	3.86E+10	0.00E+00	9.37E+08	1.70E+11
A0I1114000	7.96E+07	3.54E+08	0.00E+00	0.00E+00	2.02E+09
B001000000	8.53E+09	1.78E+10	-3.76E+07	4.72E+08	6.59E+10
B001100000	2.26E+10	4.89E+10	5.15E+06	1.16E+09	1.81E+11
B002000201	5.13E+07	1.17E+08	-2.56E+06	5.89E+04	4.54E+08
B006000102	6.68E+07	1.36E+08	-2.43E+06	1.21E+06	5.24E+08
B0D1104000	4.20E+09	9.27E+09	0.00E+00	3.51E+08	3.79E+10
B0D1104401	4.24E+09	1.12E+10	0.00E+00	0.00E+00	4.93E+10
B0F1105000	2.60E+08	7.73E+08	0.00E+00	6.99E+06	3.62E+09
B0F1213000	3.71E+08	1.12E+09	0.00E+00	1.28E+06	5.33E+09

ory (Hochreiter & Schmidhuber, 1997). (2) Gated Recurrent Unit (GRU) (Chung et al., 2014). (3) Auto-regressive Integrated Moving Average (ARIMA) (Asteriou & Hall, 2011), which represents a traditional univariate linear statistical time series forecasting model.

Each RNN model will have variants that incorporate the only model based parameter transfer, suffixed with "_pt," incorporate instance similarity-based parameter transfer, suffixed with "_it" and feature-representation based parameter transfer, suffixed with "_fr."

All trained baseline models use the same data-set, and their respective performance for comparison uses the same validation set as the proposed model. The python deep learning libraries PyTorch (Paszke et al., 2017) implements all baselines models.

4.6 Conclusion and Discussion

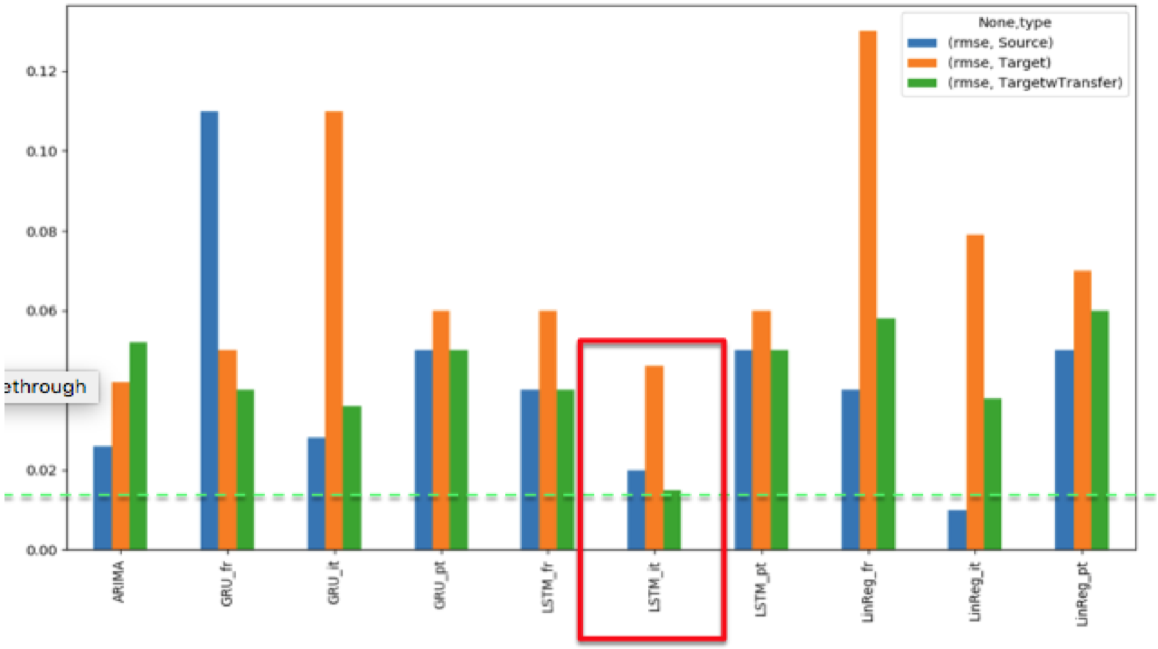
Overall Performance.

The performance of the RNN models in the TKBCC model framework and traditional linear models are summarized in table 4.6. The measured task performance of predicting bank retained earnings at the source domain, target domain, and the target domain with transferred components in terms of RMSE. %Improve is the improvement percentage of performance in the target domain with and without transferred components. Lastly, for the deep learning RNN models, the %Δ depicts the improvement in performance for predicting the target variable using transferred components when compared to the previous model variant. That is to say, the comparison of naive

Table 4.5. U.S. Bank Feature Space (2003-2018)

Variable Code	**Mean**	**Std.**	**Min**	**Med.**	**Max**
BHC02170	1.12E+05	2.01E+05	2.58E+03	3.07E+04	8.34E+05
BHCK0279	2.20E+04	6.25E+04	0.00E+00	3.56E+03	4.42E+05
BHCK2148	6.78E+03	1.37E+04	0.00E+00	1.22E+03	5.66E+04
BHCK2160	8.26E+04	1.51E+05	2.34E+03	2.32E+04	6.07E+05
BHCK3049	5.91E+02	1.94E+03	0.00E+00	0.00E+00	8.86E+03
BHCK3163	2.83E+04	6.50E+04	0.00E+00	1.38E+03	2.43E+05
BHCK3368	2.47E+06	4.36E+06	1.82E+05	7.80E+05	1.82E+07
BHCK3548	4.44E+02	1.71E+03	0.00E+00	0.00E+00	7.71E+03
BHCK4484	2.24E+01	7.09E+01	0.00E+00	0.00E+00	3.35E+02
BHCKB490	6.50E+02	1.49E+03	0.00E+00	8.55E+01	6.60E+03
BHCKB511	-5.53E+01	3.39E+03	-9.37E+03	-4.00E+01	8.95E+03
BHCKB556	1.04E+04	1.85E+04	1.12E+03	3.48E+03	8.03E+04
BHCKB989	1.07E+03	4.23E+03	0.00E+00	0.00E+00	1.97E+04
BHCKC699	1.95E+04	3.54E+04	0.00E+00	6.00E+03	1.49E+05
BHCKG104	1.44E+04	3.34E+04	-8.32E+03	3.56E+03	1.73E+05
BHCP0447	1.78E+02	5.05E+02	0.00E+00	0.00E+00	2.24E+03
BHCP0520	5.84E+03	1.27E+04	0.00E+00	9.95E+02	5.43E+04
BHCP0522	1.43E+03	2.72E+03	0.00E+00	3.43E+02	1.08E+04
BHCP2930	3.60E+03	8.61E+03	0.00E+00	3.03E+02	3.53E+04
BHCT3247	3.54E+05	6.71E+05	-2.27E+04	1.36E+05	3.94E+06
BHSP2932	9.80E-02	2.58E-01	0.00E+00	0.00E+00	1.00E+00
BHSP3150	9.13E+01	2.77E+02	0.00E+00	0.00E+00	1.72E+03
BHSP3166	4.20E-01	9.17E-01	0.00E+00	0.00E+00	4.00E+00

Figure 4.7. TKBCC Model Comparison Results



parameter-transfer to feature representation-based parameter transfer, which is then finally compared to instance-based parameter transfer. Ultimately, the experiments try to identify the benefits of each approach.

Traditionally, the linear ARIMA model is effective at predicting the target variables at both source and target domains. Moreover, when an ARIMA model fitted for the source domain predicts the target domain, it achieves consistent results. Although ARIMA models typically do not take the exogenous terms into account without additional modification, it can provide a meaningful time-series forecast of the target variable. The experimental finding aligns with much econometric and balance-sheet projection literature that leverages linear models for their forecasting methods, adding to the support of linear based models being the staple of the financial time-series analysis domain.

The linear regression RNN model using naive parameter transfer, $LinReg\_pt$, was not able to outperform the baseline ARIMA model in any domain category for this problem task. The experiment-based insight could be explained by the considerations that ARIMA makes from seasonality, trend, and stationary. Although the linear regression model makes certain assumptions, it may not be ideal for financial time series data of this problem setting as it may not consider aspects of the variability in the data. Also, the linear regression model performance improved slightly, by %3.33, when we look at the feature representation-based parameter-transfer variant on the source and target domains. However, the relative percentage of improvement, when using this model with transferred components, increased the performance by %56.31, indicating that the $LinReg\_fr$ RNN model may be an improvement from the $LinReg\_pt$ and ARIMA models for this problem. The instance-based parameter transfer, $LinReg\_it$, showed the best performance in the model group by an improvement of %34.48 when compared to $LinReg\_fr$ on the target domain with transferred components, and %51.29 when compared to $LinReg\_it$ without transferred components.

The Gated Recurrent Unit RNN model using naive parameter transfer, GRU_{pt} showed some improvement when compared to ARIMA and $LinReg_{pt}$ when performing the prediction task for source, target, and target with transferred knowledge. GRU_{pt} also showed a %18.86 improvement when using transferred knowledge from the source domain. However, when using feature-representation based parameter transfer, GRU_{fr}, performance improves by %20, when compared to GRU_{pt}, and showed an increase of %21.31 in performance when using transferred knowledge. The instance-based parameter transfer, GRU_it showed the best performance in the model group by an improvement of %48.57 when compared to GRU_fr on the target domain with transferred components, and %67.64, when compared to GRU_it without transferred components.

The Long Short Term Memory RNN model using parameter transfer, $LSTM\_pt$, showed high effectiveness when performing the task with transferred components when compared to the other models, but performed slightly below the ARIMA performance at the source and target domains. The experiment finding could be due to the non-linear relationships that $LSTM\_pt$ seeks to find in the sequential input data with considerations for long term dependencies, which may have created some level of complexity when considering the feature space. ARIMA is not considering any exogenous factors when making it is a uni-variate forecast, which may not be practical for the problem setting even though the performance is desirable. As initially mentioned, $LSTM\_pt$ does show a %18.86 improvement in performance when using transferable components, which may indicate that pertinent fundamental aspects learned in the source domain could help as a foundation for the target modeling. This insight is further evident in the performance of the feature-representation based parameter transfer of LSTM, $LSTM\_fr$, which showed a %20 improvement from $LSTM\_pt$ and ARIMA when using transferred components.

Additionally, $LSTM\_fr$ showed a %22.28 improvement when using transferred components on the target domain problem task, indicating that feature-representation based parameter transfer has meaningful benefits for this problem. The instance-based parameter transfer, $LSTM\_it$, showed the best performance out of all the models by an improvement of %62.5 when compared to $LSTM\_fr$ on the target domain with transferred components, and %67.53 when compared to $LSTM\_it$ without transferred components. Figure 4.7 illustrates the dominance of the instance-based transfer technique.

Existing works in transfer learning (Raina et al., 2007; Mesnil et al., 2011; Rosenstein et al., 2005) have shown PCA based feature-representation for transfer learning tasks have been effective due to the ability to simplify and capture variability in the dimensions to foster better model learning. Additionally, the use of similarity measures, such as Dynamic Time Warping (Müller, 2007), to determine the similarity amount time series to help attain instances that may share approximate marginal probabilities have shown to mitigate negative transfer, as seen in the (Fawaz et al., 2018). The consideration of long-term temporal dependencies in the sequential input data combined with the principal components represented feature space has shown a strong impact in the literature. It may also be beneficial for this problem. Additionally, the consideration of source data instances that share similarity with target domain training data may provide a subset of source observations that could potentially align with the latent aspects of the fine-tuning portion of the model training.

Concluding Remarks:

The proposed TKBCC model framework provides meaningful insights into the applicability of transfer learning for the bank capital-components prediction task when bank training data may be limited. The experiments conducted provide a general overview of the effectiveness of predicting capital components when considering knowledge from a source domain through parameter-based, feature-representation

Table 4.6. Bank Capital Component Prediction Model Comparison

Performance (RMSE); Epoch=100,LR= 1e-2 to 1e-5 ; Iterations = 10						
Target	Model	Source	Target	Transfer	% Improve	% ΔModelImprove
$Y_{RetEarnings_t}$	ARIMA	0.026	0.042	0.052	%-23.78	N/A
	LinReg_pt	0.05	0.07	0.06	%8.31	N/A
	LinReg_fr	0.04	0.13	0.058	%56.31	%3.33
	LinReg_it	0.01	0.079	0.038	%51.29	%34.48
	GRU_pt	0.05	0.06	0.05	%18.86	N/A
	GRU_fr	0.11	0.05	0.04	%21.31	%20
	GRU_it	0.028	0.11	0.036	%67.64	%48.57
	LSTM_pt	0.05	0.06	0.05	%18.86	N/A
	LSTM_fr	0.04	0.06	0.04	%22.28	%20
	LSTM_it	0.02	0.046	**0.015**	%67.53	%62.5

based, and instance-based transfer methods. The perspective that the experiments provide inspires the motivation to investigate methodologies and model development that could potentially contribute benefits in performing the problem task.

The effectiveness of RNN models, particularly LSTM, when considering the problem using different transfer methods, provides meaningful insight that could further guide this research direction. LSTM's inherent design provides abilities to capture long-term sequential dependencies and addresses gradient phenomenons, which seem to be aligned with the financial data and the problem task's latent forecasting requirements, as evidenced by the performance of the LSTM based models. Additionally, the feature-representation seems to make a difference when modeling with LSTM. Possibly due to the reduction in complexity and consideration of variability within the feature space. The related research literature has been able to show the effectiveness of feature-representation transfer scenarios, to which these experiments seem to support.

Analysis of additional target variables may further justify the findings from the mentioned experiments. However, reporting of banking capital components may not use similar intervals or accounting practices, which may bring forth difficulties in properly using additional target variables. Moreover, feature mapping between the source and target domains may play a vital role in the overall effectiveness of the experiments. Once again, challenges in accounting practices at each domain may cause inaccurate mappings that should depict similar aspects of the bank but may not do so, which ultimately will impact the intentions of the experiment.

The development of a novel deep learning model that can help address challenges causing hindrance to performance may be fruitful based on the results of these experiments. Intuitively, implementing the DA-RNN model from chapter 3 may provide relevant dimensional and temporal feature selection, similar to a feature-representation proxy to the transfer method. However, the compatibility to transfer the DA-RNN model parameters is complicated due to the number of attention network and prediction network parameters that need to be consolidated and transferred appropriately. Also, the investigation of transfer methods that are not ultimately parameter transfer is necessary to evaluate the most worthwhile methods holistically. The challenge in incorporating each transfer method, other than parameter transfer, into the experiment design, data structure, and overall model development requires additional effort. Each transfer method may warrant a separate research work due to the considerations and alternations required among them (Pan & Yang, 2009). Lastly, the utility of the TKBCC model framework is highly dependent on the considerations of negative transfer. Each global banking systems adhere to a diverse set of latent geopolitical regulatory idiosyncratic settings, which directly impacts the notion of universal similarity among banks. Transparency and availability of firm-level banking data further enhance the implicit differences among banking systems. Further study into negative transfer analysis may help isolate the fundamental commonalities among banks that may help identify viable training examples for transfer learning. This work, however, shows that there is a preliminary promise to using transfer learning methods to enhance this problem. The practicality of the model framework warrants further investigation and research to find relevant insights about related banking systems and forecasting tasks. Ultimately, regulators, investors, economists, and national economies may benefit from being able to forecast key capital components of foreign banking systems with limited training observations through the use of transferable knowledge. The application of transfer learning methods is transcending the traditional problem areas to more complicated ones due to the amount of data available that may be useful in solving tasks that were once infeasible due to the insufficiency of data. The current state of information availability, regulatory transparency, and national security, all of which may hinder straight forward modeling efforts. For the bank stress -testing domain, the notion of potentially leveraging related domains of data to help execute tasks in domains that lack pertinent modeling data aligns appropriately.

CHAPTER 5

CONCLUDING REMARKS

5.1 Conclusion

The popularization of advanced analytics, machine learning, big data, artificial intelligence, and data mining has helped several industries enhance, improve, or develop robust tools to help address tasks through data-driven analytical insights. The financial industry is no different. However, a particular focus on regulatory exercises since the recent financial crisis has necessitated the research and development of advanced analytical techniques. Specifically, bank stress-testing exercises have become an important part of maintaining global financial stability through bank stress-test analytics. This dissertation focuses on discussing the state of the art and popular deep learning techniques from the machine learning literature that are relevant to bank stress-test tasks, such as economic forecasting, scenario analysis, and bank performance in adversity. This dissertation also considers the systemic aspects of global financial stability and investigates deep learning methods to assess foreign banking systems with limited regulatory transparency.

The dissertation first discusses a modified conditional multi-modal generative approach, MCVAE, proposed in chapter 2 to model the joint probability distribution among multiple economic condition modalities based on domain-specific macroeconomic & microeconomic variables and point forecast values. Through rigorous experimentation on real-world historical economic conditions data, insights relevant to the problem domain about economic forecasting for bank stress-test exercises. The MCVAE's performance indicates that the consideration of multi-modal joint approximation, modality co-occurrence, and utilization of a conditional modality for data immediacy and availability concerns play a crucial role in addressing the practical challenges typically faced with economic conditions estimation. Addressing these challenges with a novel model framework that combines multi-modal exogenous factors and conditional modality learning may provide a unique perspective in the generative models' research literature, but more importantly, provide model utility for economic conditions estimations tasks.

This dissertation work also introduced a bespoke dual-attention recurrent network model, DA-RNN, in chapter 3 for banking capital & loss prediction with consideration for economic conditions estimation and dimensional influence over a temporal space. Experimentation on real-world bank performance and economic conditions data indicated that the model considerations help improve effectiveness in performing the problem task of bank capital & loss prediction. Mainly, dimensional and temporal attention mechanisms help to model banking performance during different economic conditions in a dynamic manner, rather than dependence on a static generalization. These considerations are highly applicable to the bank stress-test exercise setting due to the emphasis on firm-level reaction to economic conditions. The model can analyze bank characteristics that are impacted by specific economic conditions at certain periods to make an estimation on banking performance for that specific time-step. Much like the real-world, banks may be impacted differently at varying periods during dynamic economic conditions. This contribution indicates that a new perspective may be necessary for balance-sheet projections that consider exogenous terms, in which both dimensional and temporal features are essential. The utility of the DA-RNN model may be highly relevant to domain level stakeholders of banking stress-test analytics. DA-RNN's considerations on modeling economic crisis' and banking characteristics in a dynamic manner, which incorporates dimensional and temporal feature space to learn target output variability better. Furthermore, the model contributes relevant real-world experiments to the academic literature on NARX models to support domain-level considerations that may help improve the overall time-series prediction task.

This dissertation proposes the integration of the two proposed models from chapters 2 and 3 to develop a deep learning-based framework for bank stress-test prediction named IMBSTP. Uniquely, IMBSTP can provide a holistic perspective of economic conditions estimation by considering non-linear joint latent relationships among conditioned multi-modal exogenous factors through MCVAE and dynamic feature selection based on dimensional and temporal relevance through DA-RNN.

Lastly, this dissertation presents a series of experiments using an inductive transfer learning model framework, TKBCC, in chapter 4 that provides meaningful insights to the applicability of transfer learning for the bank capital-components prediction task when bank training data may be limited. The experiments conducted in this chapter provide a general overview of the effectiveness of predicting capital components when considering knowledge from a source domain through parameter-based, feature-representation based, and instance-based transfer methods. The perspective that the experiments provide inspires the motivation to investigate methodologies and model development that could potentially contribute benefits in performing the relevant problem task. Regulators, investors, economists, and national economies may benefit from being able to forecast key capital components of foreign banking systems with limited training observations through the use of transferable knowledge. The application of transfer learning methods is transcending the traditional approaches due to the ability to leverage the related data available that may be useful in solving tasks that were once infeasible due to the insufficiency of training data. Further study into negative transfer analysis may help isolate the fundamental commonalities among banks that may help identify viable training examples for transfer learning.

5.2 Future Work

This dissertation presents research directions for economic forecasting, bank performance prediction, and transfer-ability of knowledge for banking capital components. The proposed models and frameworks to address each defined research problem in the previous chapters of this dissertation brings forth interesting insight that may warrant further investigation that could contribute meaningful knowledge to the bank stress-test literature.

In chapter 2 of this dissertation, economic conditions estimation tasks are explored and addressed using the MCVAE model. Insights gained from the experiments present the applicability of VAE based models for the generative modeling process. Additionally, the multi-modal and conditional aspects of MCVAE showed benefits to the approximation of the actual economic conditions data distributions. For future work, additional investigation into assessing the most critical exogenous terms to the economic conditions to help reduce the dimensional variance that may hinder modeling. Also, better isolating the exogenous factors that have co-linearity or dynamic relationships may help identify unique patterns that exist in the economy on both macroeconomic & microeconomic levels. By leveraging the model probability distributions and reconstruction error to sample adverse but plausible circumstances, much like (Malik, 2018), to justify utility for scenario selection tasks. Research into effectively handling the modality co-occurrence task is also essential to handle practical challenges presented in the domain. Lastly, continuing to investigate the relevant conditional modalities that can learn input modalities to provide robust representative estimations for related density estimation tasks.

In chapter 3 of this dissertation, banking capital & loss prediction tasks were discussed and addressed by using a DA-RNN model, which considered dimensional and temporal aspects of the target prediction. Experimental findings indicate that the considerations of the DA-RNN model improved the performance of the prediction task in most scenarios. Future work on analyzing the dimensional and temporal features that influence the target variables can be useful to understand better the impacts on bank performance from banking characteristics and economic conditions. Furthermore, investigations into the temporal effects of continuous quarterly financial data versus yearly financial data could provide relevant insight to help design experiments in related literature. Finally, using the DA-RNN model in tandem with MCVAE model as IMBSTP to perform case study based validation on regulatory disclosures to understand the ability of the model to align with supervisory results.

In chapter 4 of this dissertation, the assessment of transferable banking capital components was discussed and performed using a TKBCC model framework, which considered parameter-based, instance-based, and feature-representation based transfer methods. Experimentation found that RNN models showed promise when transferring knowledge from a related domain with a related feature space. Future work in negative transfer analysis may help isolate the fundamental commonalities among banks that may help identify viable training examples for transfer learning. An investigation into the use of multi-task learning could provide additional understanding and insight to assess the effect that deep learning techniques can have in solving the problem task.

Bank stress-test analytics through the use of deep learning techniques is entirely new. Expected growth in the related research literature due to the burgeoning availability of financial data, demands from the RegTech and FinTech stakeholders, and the impact on the stability of the global economy necessitates continued contributions. This dissertation presented the utilization of deep learning techniques for model development, framework design, and utility using real-world financial data related to bank stress-test tasks. Ultimately, the dissertation aimed to contribute practical insights into the emerging lines of research work. With ongoing aspirations of continuing the additions of meaningful knowledge through investigations of novel research related to bank stress-test analytics.

BIBLIOGRAPHY

Alawad, M., Gao, S., Qiu, J., Schaefferkoetter, N., Hinkle, J. D., Yoon, H.-J., ... Jeong, J. C. (2019). Deep transfer learning across cancer registries for information extraction from pathology reports. In *2019 ieee embs international conference on biomedical & health informatics (bhi)* (pp. 1–4).

Al-Sahaf, H., Zhang, M., & Johnston, M. (2016). Binary image classification: A genetic programming approach to the problem of limited training instances. *Evolutionary computation*, *24*(1), 143–182.

Asteriou, D., & Hall, S. G. (2011). Arima models and the box–jenkins methodology. *Applied Econometrics*, *2*(2), 265–286.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Baxter, J. (1998). Theoretical models of learning to learn. In *Learning to learn* (pp. 71–94). Springer.

Bayoumi, M. T., Faruqee, M. H., Laxton, M. D., Karam, M. P. D., Rebucci, M. A., Lee, M. J., ... Tchakarov, I. (2004). *Gem: A new international macroeconomic model* (No. 239). International Monetary Fund.

Behbood, V., Lu, J., & Zhang, G. (2011). Long term bank failure prediction using fuzzy refinement-based transductive transfer learning. In *2011 ieee international conference on fuzzy systems (fuzz-ieee 2011)* (pp. 2676–2683).

Bischof, J., & Daske, H. (2013). Mandatory disclosure, voluntary disclosure, and stock market liquidity: Evidence from the eu bank stress tests. *Journal of accounting research*, *51*(5), 997–1029.

Brave, S. A., & Kelly, D. L. (2017). Introducing the chicago fed's new adjusted national financial conditions index. *Chicago Fed Letter*, *386*.

Candelon, B., & Sy, A. N. (2015). *How did markets react to stress tests?* International Monetary Fund.

Carling, K., Jacobson, T., Lindé, J., & Roszbach, K. (2004). Exploring relationships between firms balance sheets and the macro economy. *Research Department, Sveriges Riksbank*.

Caruana, R. (1997). Multitask learning. *Machine learning*, *28*(1), 41–75.

Center, T. (2017). *Fintech, regtech, and suptech: What they mean for financial supervision.* August.

Chang, C., Chen, K., Waggoner, D. F., & Zha, T. (2016). Trends and cycles in china's macroeconomy. *NBER Macroeconomics Annual*, *30*(1), 1–84.

Chen, Y., Keogh, E., Hu, B., Begum, N., Bagnall, A., Mueen, A., & Batista, G. (2015, July). *The ucr time series classification archive.* (`www.cs.ucr.edu/ eamonn/time_series_data/`)

Choi, B. (2012). *Arma model identification.* Springer Science & Business Media.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.

Covas, F. B., Rump, B., & Zakrajšek, E. (2014). Stress-testing us bank holding companies: A dynamic panel quantile regression approach. *International Journal of Forecasting*, *30*(3), 691–713.

Dai, W., Yang, Q., Xue, G.-R., & Yu, Y. (2007). Boosting for transfer learning. In *Proceedings of the 24th international conference on machine learning* (pp. 193–200).

Dees, S., Mauro, F. d., Pesaran, M. H., & Smith, L. V. (2007). Exploring the international linkages of the euro area: a global var analysis. *Journal of applied econometrics*, *22*(1), 1–38.

DODD-FRANK, W. S. R., & ACT, C. P. (2010). public law 111-203. In *Approved by the house-senate conference on hr* (Vol. 4173).

Doersch, C. (2016). Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*.

Ellahie, A. (2013). Capital market consequences of eu bank stress tests. *Available at SSRN 2157715*.

Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L., & Muller, P.-A. (2018). Transfer learning for time series classification. In *2018 ieee international conference on big data (big data)* (pp. 1367–1376).

Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *The Journal of Machine Learning Research*, *15*(1), 3133–3181.

Flannery, M., Hirtle, B., & Kovner, A. (2017). Evaluating the information in the federal reserve stress tests. *Journal of Financial Intermediation*, *29*, 1–18.

Flood, M., & Korenko, G. (2012). *Systemic scenario selection* (Tech. Rep.). Washington, DC, USA: Working Paper 0006, Office of Financial Research, US Treasury Department ....

Flood, M. D., & Korenko, G. G. (2015). Systematic scenario selection: stress testing and the nature of uncertainty. *Quantitative Finance*, *15*(1), 43–59.

Ge, L., Gao, J., Ngo, H., Li, K., & Zhang, A. (2014). On handling negative transfer and imbalanced distributions in multiple source transfer learning. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *7*(4), 254–271.

Glasserman, P., Kang, C., & Kang, W. (2015). Stress scenario selection by empirical likelihood. *Quantitative Finance*, *15*(1), 25–41.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* MIT press.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672–2680).

Grosse, R. B., Ancha, S., & Roy, D. M. (2016). Measuring the reliability of mcmc inference with bidirectional monte carlo. In *Advances in neural information processing systems* (pp. 2451–2459).

Grosse, R. B., Ghahramani, Z., & Adams, R. P. (2015). Sandwiching the marginal likelihood using bidirectional monte carlo. *arXiv preprint arXiv:1511.02543*.

Grundke, P., & Pliszka, K. (2018). A macroeconomic reverse stress test. *Review of Quantitatdeive Finance and Accounting*, *50*(4), 1093–1130.

Guerrieri, L., & Welch, M. (2012). Can macro variables used in stress testing forecast the performance of banks? *FEDS*.

Gupta, P., Malhotra, P., Vig, L., & Shroff, G. (2018). Transfer learning for clinical time series analysis using recurrent neural networks. *arXiv preprint arXiv:1807.01705*.

Harris, T. S., Khan, U., & Nissim, D. (2018). The expected rate of credit losses on banks' loan portfolios. *The Accounting Review*, *93*(5), 245–271.

Hassani, B., & Hassani, B. K. (2016). *Scenario analysis in risk management.* Springer.

Hassani, B. K. (2016). The consensus approach. In *Scenario analysis in risk management* (pp. 39–50). Springer.

He, Q.-Q., Pang, P. C.-I., & Si, Y.-W. (2019). Transfer learning for financial time series forecasting. In *Pacific rim international conference on artificial intelligence* (pp. 24–36).

Higgins, P. C., & Zha, T. (2015). China's macroeconomic time series: Methods and implications. *Unpublished Manuscript, Federal Reserve Bank of Atlanta*.

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

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Holz, C. A. (2005, January). *OECD — China Governance Project: The Institutional Arrangements for the Production of Statistics* (OECD Statistics Working Papers No. 2005/1). OECD Publishing. Retrieved from `https://ideas.repec.org/p/oec/stdaaa/2005-1-en.html` doi: 10.1787/640443218516

Hughes, T., & Poi, B. (2016). Improved deposit modeling: using moody's analytics forecasts of bank financial statements to augment internal data. *Moody's Analytics, working paper*.

Jamshidian, F., & Zhu, Y. (1996). Scenario simulation: Theory and methodology. *Finance and stochastics*, *1*(1), 43–67.

Johnson, A. E., Pollard, T. J., Shen, L., Li-wei, H. L., Feng, M., Ghassemi, M., ... Mark, R. G. (2016). Mimic-iii, a freely accessible critical care database. *Scientific data*, *3*, 160035.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kingma, D. P., Rezende, D. J., Mohamed, S., & Welling, M. (2014). Semi-supervised learning with deep generative models. *CoRR*, *abs/1406.5298*. Retrieved from http://arxiv.org/abs/1406.5298

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Kliesen, K. L., Owyang, M. T., & Vermann, E. K. (2012). Disentangling diverse measures: A survey of financial stress indexes. *Federal Reserve Bank of St. Louis Review*, *94*(5), 369–397.

Lalonde, R., & Muir, D. (2007). *The bank of canada's version of the global economy model (boc-gem).* Bank of Canada.

Laptev, N., Yu, J., & Rajagopal, R. (2018). *Applied timeseries transfer learning.* Retrieved from https://openreview.net/forum?id=BklhkI1wz

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady* (Vol. 10, pp. 707–710).

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766.

Liaw, A., & Wiener, M. (2002). Classification and regression by randomforest. *R news*, *2*(3), 18–22.

Lin, T., Horne, B. G., Tino, P., & Giles, C. L. (1996). Learning long-term dependencies in narx recurrent neural networks. *IEEE Transactions on Neural Networks*, *7*(6), 1329–1338.

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

McKinney, W. (2010). Data structures for statistical computing in python. In S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th python in science conference* (p. 51 - 56).

Mesnil, G., Dauphin, Y., Glorot, X., Rifai, S., Bengio, Y., Goodfellow, I., ... Warde-Farley, D. (2011). Unsupervised and transfer learning challenge: a deep learning approach. In *Proceedings of the 2011 international conference on unsupervised and transfer learning workshop-volume 27* (pp. 97–111).

Miller, F. P., Vandome, A. F., & McBrewster, J. (2009). *Levenshtein distance: Information theory, computer science, string (computer science), string metric, damerau?levenshtein distance, spell checker, hamming distance*. Alpha Press.

Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (Vol. 821). John Wiley & Sons.

Morgan, D. P., Peristiani, S., & Savino, V. (2014). The information value of the stress test. *Journal of Money, Credit and Banking*, *46*(7), 1479–1500.

Müller, M. (2007). Dynamic time warping. *Information retrieval for music and motion*, 69–84.

Neretina, E., Sahin, C., & De Haan, J. (2015). Banking stress test effects on returns and risks. *Working Paper Available at SSRN: https://ssrn.com/abstract=2423720*.

Nijskens, R., & Wagner, W. (2011). Credit risk transfer activities and systemic risk: How banks became less risky individually but posed greater risks to the financial system at the same time. *Journal of Banking & Finance*, *35*(6), 1391–1398.

Outlook, K. E. (2017). Joining the dots, deloitte and touché publication. *Retrieved November*, *22*, 2017.

Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, *22*(10), 1345–1359.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., ... Lerer, A. (2017). Automatic differentiation in pytorch. *NIPS 2017 Workshop Autodiff Submission*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Dubourg, V. (2011). Scikit-learn: Machine learning in python. *Journal of machine learning research*, *12*(Oct), 2825–2830.

Pesaran, M. H., Schuermann, T., & Weiner, S. M. (2004). Modeling regional interdependencies using a global error-correcting macroeconometric model. *Journal of Business & Economic Statistics*, *22*(2), 129–162.

Petrella, G., & Resti, A. (2013). Supervisors as information producers: do stress tests reduce bank opaqueness? *Journal of Banking & Finance*, *37*(12), 5406–5420.

Polyak, B. T., & Juditsky, A. B. (1992, July). Acceleration of stochastic approximation by averaging. *SIAM J. Control Optim.*, *30*(4), 838–855. Retrieved from `http://dx.doi.org/10.1137/0330046` doi: 10.1137/0330046

Pratt, L., & Thrun, S. (1997). Machine learning-special issue on inductive transfer. *Machine Learning-Special Issue on Inductive Transfer*.

Pratt, L. Y. (1993). Discriminability-based transfer between neural networks. In *Advances in neural information processing systems* (pp. 204–211).

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

Qureshi, A. S., Khan, A., Zameer, A., & Usman, A. (2017). Wind power prediction using deep neural network based meta regression and transfer learning. *Applied Soft Computing*, *58*, 742–755.

Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.

Raina, R., Battle, A., Lee, H., Packer, B., & Ng, A. Y. (2007). Self-taught learning: transfer learning from unlabeled data. In *Proceedings of the 24th international conference on machine learning* (pp. 759–766).

Rasmussen, C. E. (2003). Gaussian processes in machine learning. In *Summer school on machine learning* (pp. 63–71).

Razzak, F., Yi, F., Yang, Y., & Xiong, H. (2019). An integrated multimodal attention-based approach for bank stress test prediction. In *2019 ieee international conference on data mining (icdm)* (pp. 1282–1287).

Reserve, F. (2015). Comprehensive capital analysis and review 2015: Assessment framework and results. *US Federal Reserve, Washington, DC, March*, *11*.

Rosenstein, M. T., Marx, Z., Kaelbling, L. P., & Dietterich, T. G. (2005). To transfer or not to transfer. In *Nips 2005 workshop on transfer learning* (Vol. 898, pp. 1–4).

Sarkar, D., Bali, R., & Ghosh, T. (2018). *Hands-on transfer learning with python: Implement advanced deep learning and neural network models using tensorflow and keras*. Packt Publishing Ltd.

Schimmer, M., Levchenko, A., & Müller, S. (2012). Eventstudytools (research apps). *St. Gallen. http://www. eventstudytools. com. Accessedon: 15.01.2019*.

Shen, C.-H., Lee, Y. H., Wu, M.-W., & Guo, N. (2016). Does housing boom lead to credit boom or is it the other way around? the case of china. *International Review of Economics & Finance*, *42*, 349–367.

Sohn, K., Lee, H., & Yan, X. (2015). Learning structured output representation using deep conditional generative models. In *Advances in neural information processing systems* (pp. 3483–3491).

Tarullo, D. K. (2016). Next steps in the evolution of stress testing. *New Haven, CT: sn*.

Wilmarth, A. E. (2014). Citigroup's unfortunate history of managerial and regulatory failures. *Journal of Banking Regulation*, *15*(3-4), 235–265.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, *2*(1-3), 37–52.

WRDS. (2019). Bank regulatory. *wrds.wharton.upenn.edu*.

Wu, M., & Goodman, N. (2018). Multimodal generative models for scalable weakly-supervised learning. In *Advances in neural information processing systems* (pp. 5580–5590).

Wu, Y., Burda, Y., Salakhutdinov, R., & Grosse, R. B. (2016). On the quantitative analysis of decoder-based generative models. *CoRR*, *abs/1611.04273*. Retrieved from http://arxiv.org/abs/1611.04273

Yi, F., Yu, Z., Zhuang, F., Zhang, X., & Xiong, H. (2018). An integrated model for crime prediction using temporal and spatial factors. In *2018 ieee international conference on data mining (icdm)* (pp. 1386–1391).

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? In *Advances in neural information processing systems* (pp. 3320–3328).

Zhao, S. X., Zhan, H., Jiang, Y., & Pan, W. (2017). How big is china's real estate bubble and why hasn't it burst yet? *Land Use Policy*, *64*, 153–162.