Asset Pricing and Deep Learning

Chen Zhang (✉ demi6d@gmail.com )
SenseTime (China)

Research Article

**Keywords:** Machine Learning, Deep Learning, Deep Neural Network, Big Data, Return Prediction, MLP, RNN, CNN, Transformer, Fintech

**Posted Date:** October 17th, 2023

**DOI:** https://doi.org/10.21203/rs.3.rs-3444130/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

**Additional Declarations:** No competing interests reported.
Deep Learning for Asset Pricing

Chen Zhang
SenseTime Research, Shanghai, China
demi6d@gmail.com

July 30, 2023

Abstract

The relentless evolution of machine learning technologies has led to profound advancements across diverse financial disciplines, with asset pricing emerging as a pivotal application. This research ventures into the intricate fusion of advanced deep learning methodologies within the domain of asset pricing, with a particular emphasis on enhancing risk premium measurement to facilitate a deeper understanding of financial phenomena.

This study evaluates an array of state-of-the-art deep learning architectures, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Transformers, and others fortified with memory mechanisms and attention features. All these models are uniformly tested on a consistent set of predictive signals, comprising firm characteristics, systematic risks, and macroeconomic indicators. Empirical evidence gleaned from this study underscores the superior predictive capabilities of these models, with special recognition given to RNNs and Transformers equipped with memory and attention mechanisms.

Concurrently, the research emphasizes the substantial economic advantages that investors can reap by deploying deep learning forecasts, thereby attesting to the instrumental role of these technologies in contemporary financial operations. Notwithstanding these technological strides, the study underscores the critical need for integrating domain-specific knowledge and financial theories into the design of these models, hence fostering a harmonious convergence between AI advancements and traditional financial acumen.

Further, the research exposes unique challenges that asset return prediction tasks present to deep learning models, particularly the issue of time-varying distribution. This invariably leads to a distribution shift problem, a crucial facet for accurate financial time series prediction.

Through rigorous experimentation, the study substantiates the potential of deep learning methodologies in significantly enhancing asset risk premium measurement. With the continual evolution of the deep learning field, such methodologies can consistently broaden our understanding of the underlying economic mechanisms that govern asset pricing.

In conclusion, this research not only reasserts the indispensable role of deep learning in the burgeoning field of financial technology but also delineates its superior potential and advantages over traditional machine learning techniques. Such insights illuminate a path for future research, aiming to harness the power of deep learning to foster a comprehensive and nuanced understanding of the complex dynamics that underpin asset pricing.

Key words: Machine Learning, Deep Learning, Deep Neural Network, Big Data, Return Prediction, MLP, RNN, CNN, Transformer, Fintech

1 Introduction

This comprehensive study serves as an extension and enhancement of the pioneering work undertaken by Gu et al. [Gu, Kelly, and Xiu(2020)], wherein it presents a rigorous comparative exploration of an assortment of state-of-the-art (SOTA) deep learning methodologies as they apply to the multifaceted and dynamic field of asset pricing. The study primarily zeroes in on the critical domain of asset risk premia measurement, a cornerstone element integral to the comprehensive understanding and effective execution of asset pricing strategies.

The study ambitiously takes on the challenging task of methodically comparing, contrasting and, ultimately, determining the relative efficiencies of various advanced deep learning models, all within the specific context of asset pricing. Each model, spanning from the versatile Convolutional Neural Networks (CNNs) to the robust Recurrent Neural Networks (RNNs) and beyond, has been meticulously studied and assessed in terms of their individual and comparative performance metrics, thereby enabling an informed, data-driven evaluation of their applicability to asset risk premia measurement.

It is to be noted that the importance of asset risk premia measurement in asset pricing cannot be overstated. It forms the backbone of financial risk management and investment strategy, fundamentally influencing the returns on assets and the decisions made by investors. Thus, by applying cutting-edge deep learning techniques to this pivotal area, the research not only breaks new ground in the potential utilization of artificial intelligence in financial technology but also serves to improve the existing methodologies and frameworks significantly.
The insights and knowledge derived from this comprehensive comparative analysis of deep learning methodologies open up avenues for future exploration and development in the rapidly advancing field of financial technology, specifically in asset pricing and its related domains. These endeavors promise to deliver innovative solutions to traditional problems, improve efficiency and, ultimately, provide substantial economic gains to stakeholders in the financial industry.

1.1 Primary Contributions

This extensive exploration into the application of deep learning methodologies in asset pricing contributes significantly to the field of financial technology. Its unique approach and pioneering findings illuminate the path forward for further research and innovation. The key contributions of this study are manifold and can be encapsulated within the following salient points:

1. The research presents a comprehensive and methodical comparative analysis of a wide array of state-of-the-art (SOTA) deep learning techniques. By casting a wide net over the available models and presenting a meticulous comparative study, this work provides substantial insights into their relative efficacies within the specific context of asset pricing.

2. This study brings a particular focus on the critical domain of asset risk premia measurement, an area fundamental to the broader asset pricing field. By applying advanced deep learning techniques to this core area, the research not only enhances our understanding of asset risk premia but also potentially revolutionizes the way it is measured and utilized in practical scenarios.

3. Another significant contribution of this study lies in the empirical validation of the substantial economic benefits that can be gained by leveraging deep learning forecasts in investment decisions. This research provides concrete evidence of the potential rewards awaiting the integration of these advanced methods into financial practices.

4. The study emphasizes the essentiality of intertwining domain knowledge and financial theory in the design of deep learning models. This assertion serves as a critical reminder of the continued relevance of traditional financial theory, even in an era of rapid technological advancement.

5. Lastly, this research identifies and illuminates some of the unique challenges that deep learning faces when employed for return prediction tasks, specifically the distribution shift problem induced by time-varying distributions. In doing so, it opens the door to future studies aimed at mitigating these challenges and optimizing the application of deep learning in financial forecasting.

By offering these extensive and critical contributions, the study stands as a landmark in the intersection of finance and artificial intelligence, pushing the boundaries of what we can achieve in asset pricing through the innovative application of deep learning techniques.

1.2 Demystifying Deep Learning: A Perspective for Financial Enthusiasts

At its core, deep learning is a subset of machine learning, inheriting and expanding upon its foundational theories. However, due to its exceptionally high performance, extensive range of applications, and consistent architectural framework based on neural networks, deep learning has carved out its distinct niche within the broader field of machine learning. It diverges considerably from traditional machine learning methods such as logistic regression, support vector machines (SVMs), tree-based models, and so on. For the sake of clarity, throughout this paper, any reference to 'machine learning' will pertain specifically to these traditional methodologies.

A key pillar of deep learning is connectionism, an approach that employs mathematical models, specifically connectionist networks or artificial neural networks (NNs), to understand human cognition. Deep learning essentially represents a synergistic convergence of machine learning principles and connectionism. Over recent years, deep learning has emerged as the mainstream approach within machine learning research, witnessing an exponential growth and achieving remarkable success in the broader field of artificial intelligence.

At a rudimentary level, machine learning involves the process of learning a prediction function that maps input data to output. Deep learning enhances this process by enabling the automatic learning of various data representations, effectively serving as an automatic feature engineering tool. This aspect sets it apart from traditional machine learning, where feature engineering is typically a manual process. Figure 1 highlights this contrast between machine learning and deep learning.

Deep learning excels at uncovering complex, high-dimensional, non-linear, and deeply-entrenched relationships within data. Domain-specific deep learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), further demonstrate exceptional proficiency in processing non-structured data.

For financial aficionados, the true power of deep learning lies in its ability to identify intricate patterns and relationships hidden within vast amounts of financial data, something that is often beyond the reach of traditional methods. This, combined with its potential for automating laborious feature engineering tasks, underscores the value of deep learning in advancing financial practices, particularly in areas such as asset pricing.
1.3 The Rationale for Applying Deep Learning to Asset Pricing

Empirical asset pricing fundamentally revolves around a dynamic prediction problem, with both the input variables (predictive signals) and output variable (asset risk premium) subject to changes over time. Given the myriad of factors influencing stock returns, the dimensionality of the input variables is typically high, and the relationship between the input information and output returns can be intricately nonlinear.

In this complex scenario, deep learning presents distinct advantages over traditional machine learning techniques. While machine learning often struggles to incorporate economic theory into the model as prior knowledge, deep learning offers the flexibility to design specific neural network architectures tailored to suit different scenarios. This flexibility enables deep learning to readily adapt to new insights and continually improve its predictive accuracy. For instance, sequence models, a type of deep learning architecture, are particularly suited to handling time series data and time-varying predictions.

At its core, deep learning, also known as deep neural network, draws inspiration from biological neuroscience. The monumental success of deep learning in the AI industry can be attributed in part to its emulation of one of nature's most intricate creations - the human brain. By incorporating mechanisms akin to the functioning of the human brain, deep learning can delve deep into data, uncovering layers of hidden patterns and relationships.

Additionally, the ’end-to-end’ nature of deep learning models simplifies the predictive process by eliminating the need for labor-intensive feature engineering, a common step in traditional machine learning methods. This not only saves time but also reduces the risk of human error and bias.

In the context of asset pricing, these advantages of deep learning present a powerful tool capable of dealing with the high dimensionality, nonlinearity, and time-variance inherent in the data. By automating feature engineering and being able to learn from the intricacies of financial data, deep learning offers the potential for more accurate and insightful asset pricing predictions, thereby redefining the landscape of financial asset management.

1.4 A Spectrum of Advanced Deep Learning Techniques: An Overview of the Methods Explored

This research embarks on a meticulous examination of a carefully selected set of deep learning methodologies. These methods, being emblematic of the most advanced and state-of-the-art (SOTA) deep learning models, have garnered substantial success across a wide variety of artificial intelligence applications.

The deep learning models under scrutiny include a comprehensive range of deep feedforward Neural Networks (NNs) such as Deep Neural Networks (DNNs) and their residual counterparts - the Residual DNNs, which incorporate an innovative skip connection functionality. This research also probes into Convolutional Neural Networks (CNNs) and their residual versions, offering a broad view of feedforward architectures and their utility.

Simultaneously, sequential models form another critical part of this study, with the inclusion of Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Gated Recurrent Unit (GRU) networks. Additionally, attention mechanism-infused RNNs also come under the lens, spotlighting the power of this unique approach within sequential models.

Lastly, this study delves into the groundbreaking Transformer model. Predicated entirely on attention mechanisms, the Transformer model has outperformed a myriad of RNNs across diverse Natural Language Processing (NLP) tasks, highlighting its potential for extrapolation into other fields, including asset pricing.

Each of these methods not only stands as a powerful tool in its own right, but when analyzed collectively, they offer invaluable insights into the nuances and dynamics of deep learning methodologies in the realm of asset pricing, pushing the boundaries of our understanding and application of artificial intelligence within the financial sphere.
1.5 Main Empirical Findings: Unraveling the Complexities of Asset Pricing Through Deep Learning

Following the foundational work of [Gu, Kelly, and Xiu(2020)], this study embarks on a rigorous empirical analysis, examining an extensive dataset of 30,000 individual stocks spanning a substantial period of 20 years, from 1980 to 2000. This temporal frame provides a rich backdrop of varied market conditions, capturing both periods of economic prosperity and downturn, thus providing a fertile ground for our analysis.

The depth of the investigation is further accentuated by the incorporation of a comprehensive set of attributes for each stock. Specifically, 81 unique characteristics, including a myriad of financial indicators and firm-specific attributes, have been considered. This provides a holistic view of each stock, enabling the deep learning models to capture intricate patterns and relationships across these features.

In addition to the firm-specific characteristics, the analysis is supplemented with 6 macroeconomic proxies that encapsulate broader economic trends and shifts, thereby providing a contextual lens through which the performance of individual stocks can be interpreted.

Finally, the study draws upon the framework proposed by [Fama and French(2015)], incorporating five key factors to represent systematic risks. These factors, which include aspects like market risk and size, play a pivotal role in determining stock returns, and their inclusion further deepens the complexity of our analysis.

In essence, this study weaves together an elaborate tapestry of data, combining a wide range of stock characteristics, macroeconomic indicators, and systematic risk factors. This expansive and diverse dataset, coupled with the substantial 20-year time frame, sets the stage for a comprehensive, in-depth exploration of deep learning methodologies in empirical asset pricing.

The findings emerging from this exhaustive empirical study are multifaceted, shedding new light on the utility of deep learning in asset pricing:

Performance Superiority of Deep Learning Models: All SOTA deep learning techniques exhibit high efficacy in stock return prediction, significantly outpacing traditional linear prediction models. This underscores the immense potential deep learning holds for advancing the field of asset pricing.

RNNs and Transformers as Top Contenders: Among various methods, RNNs with memory mechanisms and Transformers emerge as the best performers, substantiating the predictive power of historical data for asset returns. The intrinsic ability of RNNs to synthesize past and present information, aided by memory mechanisms such as attention and cell units, contributes to their effectiveness. The study also reveals the essential role of long-term memory in enhancing stock return prediction.

Relative Performance of CNNs and Transformers: The relatively inferior performance of CNNs, along with the observation that Transformers do not universally dominate all other RNNs, emphasizes the crucial role of domain-specific knowledge and theory-guided model design. This points to the fact that specific model architecture must be tailored to individual scenarios, corroborating the “no free lunch” theorem within deep learning applications.

Minimal Impact of Skip Connections: The study finds that skip connections coupled with deep layers lead to only minor improvements, while middle or shallow networks continue to perform well. This suggests that the underlying relationship between market data and asset prices may not be as complex as in other deep learning applications. Consequently, the exploration of the economic mechanisms governing asset pricing becomes a promising avenue of study.

These findings collectively constitute a groundbreaking contribution to the existing body of knowledge in asset pricing. By navigating the intricate maze of factors that influence asset returns, this research illuminates the path forward, not only showcasing the unprecedented capabilities of deep learning but also providing a nuanced understanding that bridges theory and practice in the world of finance.

1.6 Limitations of Deep Learning: Understanding the Challenges

While deep learning boasts impressive capabilities in predicting excess returns conditioned on market information, it is essential to recognize its limitations. The field of deep learning, much like its parent field of machine learning, is characterized by an inherent complexity that makes fully elucidating the underlying mechanisms an ongoing challenge.

The field is still relatively nascent, and complete theoretical understanding of deep learning models is still under construction. Consequently, this results in the inability to offer exhaustive explanations of the internal processes and decision-making dynamics within these models. This is arguably a common drawback of many machine learning methodologies.

Deep learning models are crafted with an understanding of economic mechanisms, but they are not built entirely upon them. As such, while these models are designed to process and learn from financial data effectively, they may not fully encapsulate the intricacies of financial theory. Nonetheless, deep learning holds an edge over traditional machine learning in that it allows us to incorporate economic principles during the design of neural network architectures. This utilization of
financial theory as an induction bias helps to shape the model for corresponding financial problems, enabling a more theory-guided approach.

The field is not complacent about these limitations. There is an increasing wave of research focusing on explainable AI, which seeks to address the interpretability and transparency issues plaguing deep learning models. While there is progress to be made, the field is vibrant and actively evolving, with researchers continuously working to narrow the gap between theoretical understanding and practical application.

In conclusion, while deep learning offers transformative potential in asset pricing and broader financial applications, it is equally important to remain cognizant of its limitations and continue efforts towards surmounting them. This balance of harnessing strengths and acknowledging weaknesses ultimately fosters a more nuanced and effective approach to leveraging deep learning within the world of finance.

1.7 Literature

This study enriches the empirical asset pricing literature in several unique and meaningful ways.

To begin with, it broadens the research scope of machine learning-based empirical asset pricing. It builds upon the work of [Gu, Kelly, and Xiu(2020)], [Leippold, Wang, and Zhou(2021)], and [Bianchi, Buchner, and Tamoni(2021)], which focus on stock pricing and bond pricing. Unlike these studies, my research is not confined to a specific type of asset return prediction but strives to apply deep learning methods more broadly.

Secondly, it enhances the understanding of cross-section stock returns that employ machine learning, as explored by [Gu, Kelly, and Xiu(2021)]. This research ventures beyond factor dimension reduction and underscores the utility and performance of various deep learning methodologies in predicting stock returns.

Thirdly, it amplifies time series return prediction literature, building on the work of [Koijen and Van Nieuwerburgh(2011)] and [Rapach and Zhou(2013)]. By employing state-of-the-art deep learning techniques, this study offers fresh perspectives and more complex models to approach the prediction problem.

While numerous deep learning methodologies have been deployed in the asset pricing literature, including works like [Hutchinson, Lo, and Poggio(1994)], [Yao, Li, and Tan(2000)], [Sirignano, Sadhwani, and Giesecke(2016)], and [Heaton, Polson, and Witte(2016)], my paper takes a unique approach. It focuses on a comprehensive study of various cutting-edge deep learning methods, comparing their performance in stock return prediction tasks.

Rather than merely using deep learning as a tool for asset pricing or portfolio management, this research emphasizes understanding the inherent patterns and relationships between market data and stock returns. The goal is to conduct a profound analysis of these patterns to inform the design of more accurate and effective deep learning models for asset pricing.

In summary, this study bridges gaps and extends frontiers in the asset pricing literature, leveraging advanced deep learning methods to gain new insights and advance the current understanding of asset pricing.

2 Methodology

This section briefly outlines the deep learning methods employed in this study, taking into account the array of methodologies under analysis. For a detailed exposition of the basic mechanisms of deep learning, readers can refer to the work of [Goodfellow, Bengio, and Courville(2016)].

All models within the scope of this analysis share a common objective function - the Mean-Square Error (MSE), which is a popular choice due to its suitability for regression problems. The learning algorithms employed are fundamentally grounded in gradient descent, the cornerstone of optimization algorithms in deep learning. Notably, this study employs Adam, an adaptive learning rate optimization algorithm known for its effectiveness in training deep neural networks. Further improvements are applied to this basic optimizer, incorporating advanced techniques such as batch normalization and layer normalization to enhance learning performance.

The regularization strategy across all the models remains consistent with the application of Dropout. Dropout, a popular regularization technique, helps to mitigate the issue of overfitting by randomly ‘dropping out’ neurons during training, thereby promoting a more generalized model.

In terms of asset pricing modeling, the approach follows the method described by [Gu, Kelly, and Xiu(2020)] - an additive prediction error model for excess return. The model is expressed as:

\[ r_{i,t+1} = y_{i,t+1} = E_t(y_{i,t+1}) + \epsilon_{i,t+1} \]  

(1)

where the expected excess return at time \( t \) is defined by a function \( f \) parameterized by \( \theta^t \), and \( x_{i,t} \) represents the available predictive signals at time \( t \):
\[ E_t(y_{i,t+1}) = f(x_{i,t}; \theta) \]  

This formulation helps us systematically evaluate the efficacy of various deep learning methodologies in predicting asset prices, offering a uniform baseline for comparison across different models.

### 2.1 Objective Function

For every deep learning model in this study, the parameters are estimated through the process of Maximum a Posteriori Estimation (MAPE). The core aim here is to maximize the posterior probability of the parameters \( \theta \), conditional on the observed samples of predictors \( X \) and targets \( Y \). The fundamental model is constructed as:

\[ y = f(x; \theta) + \varepsilon \]  

With an expected value of:

\[ E(y) = f(x; \theta) \]  

Two foundational assumptions underpin this model. Firstly, the noise term is assumed to follow a Gaussian distribution. Secondly, the prior distribution of the parameters \( \theta \) is assumed to be uniform. Given these assumptions, we can formulate the maximum a posteriori estimation problem as:

\[ \theta^* = \arg\max_{\theta} P(\theta | Y, X) = \arg\max_{\theta} \frac{P(Y | \theta, X) P(\theta | X)}{P(Y | X)} = \arg\max_{\theta} P(Y | \theta, X) \]  

Interestingly, under these assumptions, the maximum a posteriori estimation problem reduces to the maximum likelihood estimation (MLE) of \( \theta \) given the samples \( (X,Y) \). Additionally, given the Gaussian noise assumption, when we take the natural log of equation (5), the MLE problem becomes equivalent to minimizing the Mean Square Error (MSE).

Therefore, the objective function for the model is formulated to minimize the MSE. As a result, the loss function \( L(\theta) \) of the deep learning models under consideration is defined as the MSE:

\[ L(\theta) = \frac{1}{N \cdot T} \sum_{i=1}^{N} \sum_{t=1}^{T} (r_{i,t+1} - f(x_{i,t}; \theta))^2 \]  

This MSE-based loss function provides a simple and effective measure of the model’s ability to predict asset returns, and serves as a uniform measure for comparison across the various deep learning models.

### 2.2 Learning Algorithm

The learning algorithm employed in this study revolves around the application of the Adam algorithm, a first-order gradient-based optimization technique, introduced by [Kingma and Ba(2014)]. This algorithm incorporates both Momentum and RMSprop mechanisms, combining their advantages to improve the efficiency and reliability of the optimization process.

Additionally, to maintain consistency of the data scale and prevent the potential issue of exploding or vanishing gradients during backpropagation, normalization methods including Batch Normalization and Layer Normalization, introduced by [Ioffe and Szegedy(2015)] and [Ba, Kiros, and Hinton(2016)] respectively, are applied. These mechanisms normalize the input data and the activations of hidden layers, thereby ensuring a more stable learning process.

The issue of overfitting, which may compromise the model’s generalization ability, is mitigated through regularization. Specifically, Dropout regularization is used, which randomly deactivates a subset of neurons in each training batch. This has a dual effect. Firstly, it reduces the complexity of the network, as fewer connections between neurons result in fewer parameters to optimize. Secondly, it leads to the creation of an ensemble of various subnetworks, enhancing the generalizability of the final model.

To optimize model performance and prevent overfitting, the dataset is partitioned into three distinct subsets. This partitioning is done along the temporal dimension, with 80% of the data used for training, 10% for validation, and the remaining 10% for testing. The training set is used to fit the models, the validation set serves to fine-tune hyperparameters and prevent overfitting, while the test set is exclusively used to assess the out-of-sample predictive performance of the models.

This comprehensive application of advanced optimization, regularization, and partitioning techniques ensures that our deep learning models are robust, generalize well, and perform optimally on unseen data.

### 2.3 Model Performance Evaluation

To objectively assess the performance of the various deep learning models employed in this study, I make use of the out-of-sample \( R^2 \) statistic, a common measure of model predictive ability.

Aligning with the methodology outlined by [Gu, Kelly, and Xiu(2020)], the denominator of the \( R^2 \) metric is the sum of squared returns, calculated without the standard practice of demeaning. The out-of-sample \( R^2 \) is thus given by:
In this equation, T symbolizes the test dataset, which is reserved exclusively for the final evaluation of model performance. In this manner, I ensure that the primary focus remains squarely on the out-of-sample predictive power of each model, rather than in-sample fitting. This methodology results in a fair comparison of the deep learning methods, allowing the determination of the most effective approach for predicting stock returns.

3 Monte Carlo Simulations

To thoroughly assess and juxtapose the performance of various models, this study embraces the Monte Carlo simulation strategy. These simulations are meticulously designed, using parameters that echo those deployed in the landmark study by [Gu, Kelly, and Xiu(2020)]. Each Monte Carlo simulation carves the time series data into three sequential subsets, dedicating 80

The gallery of models being examined in this study spans an impressive range, encapsulating linear and non-linear methodologies alike. This collection includes techniques such as Random Forest (RF), Gradient Boosted Regression Trees (GBRT), LightGBM (LGBM), Multilayer Perceptron (MLP), MLP augmented with Residual connections, Convolutional Neural Networks (CNNs), Recurrent Neural Network (RNN), RNN enhanced with Attention mechanisms, Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), and the Transformer architecture. Alongside these advanced models, I also assess the performance of the more conventional Ordinary Least Squares (OLS) approach, and an oracle model is also factored in for comprehensive benchmarking.
Table 1: Comparison of Predictive $R^2$s for Machine Learning Algorithms in Simulations

| Model     | Linear | Nonlinear |
|-----------|--------|-----------|
|           | IS     | OOS       | IS     | OOS       |
| $R^2$(%)  |        |           |        |           |
| OLS       | 4.23   | 3.38      | 3.32   | 2.32      |
| RFR       | 7.37   | 1.45      | 9.06   | 3.91      |
| GBRT      | 6.69   | 4.19      | 8.52   | 5.64      |
| LGBMR     | 6.02   | 4.34      | 8.53   | 5.68      |
| MLP       | 3.61   | 2.69      | 4.92   | 3.27      |
| MLP+R     | 3.88   | 2.91      | 5.51   | 4.57      |
| CNN       | 4.12   | 3.29      | 11.84  | 2.10      |
| RNN       | 3.66   | 2.63      | 3.78   | 2.91      |
| RNN+A     | 3.78   | 2.67      | 5.73   | 4.32      |
| GRU       | 4.34   | 2.71      | 6.04   | 4.13      |
| LSTM      | 3.68   | 2.73      | 4.35   | 3.58      |
| Transformer | 3.44  | 2.67      | 5.86   | 4.14      |
| Oracle    | 5.22   | 7.23      | 7.88   | 7.71      |

For each model under scrutiny, the validation set is used to diligently fine-tune hyperparameters. The predictive prowess of each model is then gauged within the out-of-sample testing subset. To ensure a robust and reliable analysis, this sequence is replicated across 10 independent Monte Carlo samples. The average in-sample (IS) and out-of-sample (OOS) $R^2$ values for each model are tabulated in Table 1, presenting a clear picture of each model's performance.
When the underlying data generating process follows a linear trend, it is observed that OLS and advanced tree-based techniques like RF and GBRT outperform others, boasting the highest OOS $R^2$ values. However, in scenarios with an underlying non-linear data generating process, these models continue to outshine OLS, signifying their capability to capture non-linear relationships. Neural networks (NNs) also supersede OLS in these non-linear scenarios, but fall short of RF and GBRT. This suggests that while traditional machine learning methodologies thrive in handling simpler non-linear tasks, NNs excel in grappling with more complex patterns. Unsurprisingly, OLS, owing to its inherent constraint to linear relationships, trails in performance across all scenarios.

4 An Empirical Study of US Equities

4.1 Data and Over-arching Model

The empirical study encompasses all firms listed on the NYSE, AMEX, or NASDAQ in the CRSP database. The sample includes monthly equity returns spanning a substantial 20-year period, from July 1980 to December 2000, covering approximately 14,000 stocks in total and averaging around 6,000 stocks each month.

The risk-free rate is represented by the one-month Treasury bill rate, sourced from Kenneth R. French’s data library, enabling the calculation of excess return by deducting the risk-free rate from the respective stock return.

Building upon the foundational work of [Green, Hand, and Zhang(2017)], I construct 81 stock-level characteristics, thus contributing to a thorough representation of the cross-section of stock returns literature.

In an attempt to account for systematic risk, I incorporate five factors proposed by [Fama and French(2015)]. Further enriching the dataset, I also incorporate six macroeconomic proxies, inspired by the work of [Welch and Goyal(2008)]. These proxies encompass the dividend-price ratio (dp), earnings-price ratio (ep), book-to-market ratio (bm), net equity expansion (ntis), Treasury-bill rate (tbl), and stock variance (svar).

The meticulously curated dataset, therefore, encapsulates diverse attributes of the firms and the wider economic environment, providing a solid foundation for the ensuing empirical analysis.

In alignment with the methodology of [Gu, Kelly, and Xiu(2020)], all my deep learning methods seek to emulate the empirical over-arching model $E_t(r_{i,t+1}) = f(x_t)$, as stipulated in equation (2). The baseline set of stock-level covariates, denoted as $z_{i,t}$, is defined as follows:

\[ z_{i,t} = x_t \otimes c_{i,t} \]  

(8)

The theoretical underpinning for this model is grounded in the standard beta-pricing representation of the conditional asset pricing Euler equation, as mentioned by [Gu, Kelly, and Xiu(2020)]. Here, the stock-level characteristics $c_{i,t}$ act as a stand-in for the risk exposure function $\beta_{i,t}$, while systemic risk and macroeconomic proxies emulate the role of the risk premium $\lambda_t$.

Defining $\beta_{i,t} = \theta_1 c_{i,t}$ and $\lambda_t = \theta_2 x_t$, the model transforms as:

\[ E_t(r_{i,t+1}) = \beta_{i,t} \lambda_t \]  

(9)

Further simplifying, we get:

\[ g^*(z_{i,t}) = E_t (r_{i,t} + 1) = c'_{i,t} \theta_2 x_t = (x_t \otimes c_{i,t})' \vec{\theta}_2 =: z'_{i,t} \theta \]  

(10)

Here, $\theta$ is defined as:

\[ \theta = \vec{\theta}_1 \]  

(11)

The strength of the deep learning model lies in its generality. The function $g^*(z_{i,t})$ can be any function and the input $z_{i,t}$ can be any combination of $x_t$ and $c_{i,t}$. Consequently, the factor risk premium can exhibit a nonlinear relationship with macroeconomic and systematic risk, expressed as $\lambda_t = f_1(x_t)$, while the risk exposure can embody a nonlinear relationship with firm characteristics, denoted as $\beta_{i,t} = f_2(c_{i,t})$. This provides a versatile modeling framework to capture a wide spectrum of market dynamics.

The study employs a distinct time split to examine the model’s behavior across an extended temporal scale of 20 years, offering a rich observation of the predictive mechanisms within the complexities of asset pricing.

This single time span, which spans from 1980 to 2000, provides a robust window into the economic fluctuations, market dynamics, and technological advancements of two transformative decades. This extended period captures significant events that influenced financial markets, such as regulatory changes, global market integrations, technology-driven trading, and several market cycles. The data is meticulously divided into three distinct parts:

Training Set (1980-1996): Comprising 80% of the data, this segment spans the years from 1980 to 1996. This period includes major market events, economic shifts, and varying market conditions, providing a comprehensive backdrop for the
model to learn. The long-duration training set is designed to enable the model to capture long-term patterns, cyclic behaviors, and the nuanced interplay between various predictive signals.

Validation Set (1997-1998): Encompassing 10% of the data, this subset covers the years from 1997 to 1998. The validation set is strategically selected to include transitional periods, enabling the tuning of model hyperparameters in a way that enhances its generalization capabilities across diverse market environments.

Test Set (1999-2000): Containing the remaining 10% of the data, this part spans the years from 1999 to 2000. This era represents significant market changes, including the peak of the dot-com bubble. The selection of this testing period allows for a rigorous assessment of how well the models generalize to unseen data and adapt to abrupt market changes.

By carefully crafting this 20-year time span, the study goes beyond mere predictive performance assessment. It constructs an experimental design that reflects a multifaceted understanding of the market, encapsulating varying economic conditions, regulatory landscapes, and technological trends. This approach not only bolsters the robustness of the comparative analysis but also opens up avenues for exploring how deep learning models respond to complex, long-term financial phenomena, thereby advancing the fusion of machine intelligence with traditional financial wisdom.

4.2 The Cross Section of Individual Stocks

Building on the expansive comparative studies conducted by [Gu, Kelly, and Xiu(2020)] and [Leippold, Wang, and Zhou(2021)], which firmly established basic Deep Neural Networks (DNNs) as the top-performing models among various machine learning methodologies, my research pivots the focus towards exploring state-of-the-art (SOTA) deep learning models. I begin with DNNs as the benchmark, setting the stage for a deep exploration of advanced deep learning techniques.

A total of eight distinctly potent deep learning models are rigorously tested in this study, specifically: OLS, DNNs, residual DNNs, CNNs, residual CNNs, RNNs, RNNs enhanced with attention mechanisms, GRU, LSTM, and Transformer architectures.

Table 2 offers a detailed exposition of these different deep learning models, evaluated predominantly on their out-of-sample $R^2$ performance metric. To ensure a thorough and encompassing analysis, I utilize a single, yet comprehensive, data set reflecting a long temporal span of 20 years, covering the period from 1980 to 2000.

This extended time span allows us to scrutinize how different deep learning models perform over a considerably long period. It also enables the understanding of the models’ robustness and adaptability in predicting individual stock returns across diverse market conditions. It further provides valuable insight into their ability to encode complex, long-term financial patterns and trends, thus illuminating the potential and limitations of each model in tackling the intricacies of stock return prediction.

Through this approach, the study paints a detailed portrait of deep learning performance in asset pricing, exploring not just their predictive capabilities but also their adaptability and resilience across two transformative decades of market evolution. It paves the way for further exploration of advanced deep learning techniques within the realm of empirical asset pricing, heralding a new era of technological integration in financial research and practice.

|          | OLS | RFR | GBRT | LGBMR | MLP | MLP | CNN | RNN | RNN | GRU | LSTM | Transformer |
|----------|-----|-----|------|-------|-----|-----|-----|-----|-----|-----|------|-----------|
|          |     |     |      |       |     |     |     |     |     |     |      |           |
| 20 Years | -0.50 | 1.05 | 0.53 | 0.63 | 1.14 | 1.25 | 1.00 | 1.06 | 1.38 | 1.44 | 1.44 | 1.34 |
The main findings of my comparative study are as follows:

Firstly, a significant performance improvement is observed when utilizing state-of-the-art (SOTA) deep learning methods in comparison to traditional linear regression techniques for stock return prediction. This underscores the prowess of deep learning models in capturing complex patterns within financial data.

Secondly, Recurrent Neural Networks (RNNs) exhibit superior performance, indicative of the predictive power that historical data possesses in forecasting present stock returns. This underscores the inherent correlations between current returns and past data. The proficiency of RNNs in integrating past and present information for predictive tasks is validated. However, the issue of the vanishing gradient problem hampers the trainability of vanilla RNNs in long timespan datasets.

Thirdly, both attention mechanisms and cell units within RNNs prove effective, underscoring the importance of long-term memory in predicting stock returns.

Fourthly, Gated Recurrent Units (GRU), acting as a simplified version of LSTM, offer a streamlined model that doesn't compromise on performance. As per Occam's Razor principle, GRU can serve as an excellent alternative to LSTM in empirical asset pricing, highlighting the essence of simplicity in model design.

Fifthly, Convolutional Neural Networks (CNNs) show relatively weaker performance than other Neural Network architectures. This is likely attributable to the design of CNNs, which are more aligned with computer vision tasks than stock return prediction. Similarly, while Transformers generally outperform RNNs in Natural Language Processing tasks, they do not necessarily hold the same advantage in predicting stock returns. This underscores the importance of domain knowledge and the need for theory-guided model design in achieving optimal results.

Lastly, deep neural networks with skip connections show only marginal improvements, suggesting that an increase in network depth doesn't necessarily translate into improved performance. Middle to shallow networks could be sufficient and effective for the task at hand. Consequently, we should not overestimate the depth and complexity of underlying associations between market data and stock returns. This calls for an optimistic perspective towards the future of asset pricing research. One promising avenue could be the utilization of explainable AI methods to learn from data and elucidate the underlying economic mechanisms. This approach could serve as a catalyst for advancing asset pricing studies.

| Table 3: Diebold-Mariano Tests of Out-of-Sample Prediction |
|-----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
|                 | RFR | GBRT | LGBMR | MLP | MLP | CNN | RNN | RNN | GRU | LSTM | Transformer |
| OLS             | 20.64 | 13.83 | 17.4 | 21.35 | 23.83 | 28.32 | 22.08 | 26.74 | 26.86 | 27.22 | 25.9 |
| RFR             | -23.35 | -15.89 | -3.19 | 1.49 | -7.33 | -5.84 | 7.21 | 8.54 | 7.09 | 4.3 |
| GBRT            | 15.08 | 12.61 | 10.68 | 1.75 | 7.21 | 22.08 | 24.73 | 19.24 | 20.65 |
Table 3 provides a comprehensive evaluation of the statistical significance of the thirteen models through the implementation of pairwise Diebold-Mariano (DM) test statistics. These statistical measures offer an evaluation of the comparative predictive accuracy between various forecasting models. A positive statistic signifies the column model outperforms the corresponding row model, with bold numerals indicating significance at the 5% level.

The results presented in this table align harmoniously with our previous discussions, delivering robust evidence to support our initial assessments. Particularly noteworthy is the consistently high performance of Recurrent Neural Networks (RNNs) enhanced with memory units and attention mechanisms. This signifies that these network architectures not only excel in predictive accuracy but also demonstrate statistically significant improvements over other models.

Such findings enrich our understanding of the advantages of specific deep learning configurations in the realm of stock return predictions. The statistical significance of these results suggests that the superior performance of RNNs with memory and attention enhancements is not merely incidental, but rather a robust feature of these models. This provides compelling grounds for favoring these architectures in future empirical asset pricing studies.

### 4.3 Portfolio Forecasts

#### 4.4 Portfolio Predictions and Performance

To extend our understanding of how machine learning forecasts apply to real-world scenarios, I have performed portfolio experiments utilizing the one-month-ahead out-of-sample stock return predictions generated by the thirteen models under consideration. These experiments involve sorting stocks into deciles based on their predicted returns and subsequently constituting zero-net-investment portfolios each month. These portfolios are positioned long on the highest predicted decile and short on the lowest one.

Table 4 presents the performance of these portfolios over a five-year testing period. The columns labeled “Pred”, “Avg”, “Std”, and “SR” denote the predicted monthly returns, the realized average monthly returns, the standard deviations of these returns, and the corresponding Sharpe ratios, respectively. All portfolios are constructed with equal weighting for each constituent stock.

The results echo our previous findings regarding the superior forecasting performance of deep learning models. Interestingly, though, the OLS model exhibits relatively strong portfolio performance compared to its out-of-sample $R^2$. The realized returns largely align with the forecasts produced by the deep learning models, further cementing their practical utility. Among these, RNN models equipped with memory mechanisms consistently outshine both linear models and traditional machine learning approaches, as evidenced by the higher realized returns and Sharpe ratios of portfolios based on their predictions.

These findings underscore the tangible benefits of incorporating advanced deep learning techniques in portfolio management. They showcase the efficacy of RNNs with memory mechanisms in not just forecasting stock returns, but also driving investment strategies that yield superior risk-adjusted returns.
Table 4: Performance of Machine Learning Portfolios

|            | OLS       | RFR       | GBRT      |
|------------|-----------|-----------|-----------|
|            | Pred     | Avg       | Std       | SR        | Pred     | Avg       | Std       | SR        | Pred     | Avg       | Std       | SR        |
| Low(L)     | -3.78 0.3 | 7.31 0.14 | -3.38 1.13 | 7.58 0.52 | -0.08 1.11 | 5.89 0.65 |
| 2          | -2.23 0.41 | 5.63 0.25 | -2.38 1.33 | 6.75 0.68 | 0.05 1.08 | 5.72 0.65 |
| 3          | -1.7 0.69 | 5.26 0.45 | -1.91 1.5  | 6.05 0.86 | 0.15 1.36 | 5.94 0.79 |
| 4          | -1.32 0.74 | 4.88 0.52 | -1.47 1.46 | 5.82 0.87 | 0.22 1.13 | 5.93 0.66 |
| 5          | -1.02 1.23 | 4.66 0.91 | -1.07 0.98 | 5.19 0.65 | 0.22 0.84 | 4.82 0.6 |
| 6          | -0.73 1.2 | 4.74 0.88 | -0.72 0.9 | 4.79 0.65 | 0.22 1.17 | 4.84 0.84 |
| 7          | -0.44 1.48 | 4.85 1.05 | -0.39 1.33 | 4.98 0.93 | 0.23 0.81 | 4.55 0.62 |
| 8          | -0.1 1.58 | 4.97 1.1 | -0.04 1.38 | 4.86 0.98 | 0.29 2.05 | 5.95 1.19 |
| 9          | 0.37 2.4 | 5.78 1.44 | 0.44 1.48 | 5.69 0.9 | 0.4 1.7 | 5.78 1.02 |
| High(H)    | 1.55 4.54 | 7.64 2.06 | 2.3 3.06 | 7.48 1.42 | 0.83 3.31 | 7.5 1.53 |
| H-L        | 5.32 4.24 | 4.43 3.31 | 5.68 1.93 | 7.25 0.92 | 0.91 2.21 | 4.14 1.84 |

|            | LGBMR     | MLP       | MLP+R     |
|------------|-----------|-----------|-----------|
|            | Pred     | Avg       | Std       | SR        | Pred     | Avg       | Std       | SR        | Pred     | Avg       | Std       | SR        |
| Low(L)     | 0.39 1.26 | 5.9 0.74  | 0.33 0.2  | 7.16 0.1  | -0.62 -0.04 | 7.94 0.02 |
| 2          | 0.56 1.08 | 5.8 0.64  | 0.79 0.34 | 5.36 0.22 | 0.02 0.15 | 6.39 0.08 |
| 3          | 0.62 1.42 | 6.02 0.82 | 1.04 0.76 | 6 0.44  | 0.3 0.53  | 5.57 0.33 |
| 4          | 0.67 1.31 | 6.42 0.71 | 1.22 0.62 | 5.25 0.41 | 0.51 0.79 | 5.12 0.53 |
| 5          | 0.67 1.25 | 5.58 0.78 | 1.39 0.9 | 4.92 0.63 | 0.71 0.9 | 4.86 0.64 |
| 6          | 0.67 1.1  | 4.76 0.8 | 1.55 1.33 | 4.65 0.99 | 0.92 1.19 | 4.85 0.85 |
| 7          | 0.68 1.13 | 4.84 0.81 | 1.7 1.53 | 5.14 1.03 | 1.15 1.69 | 4.79 1.22 |
| 8          | 0.7 1.98 | 6.32 1.08 | 1.86 1.77 | 5.09 1.2  | 1.43 1.78 | 4.89 1.26 |
| 9          | 0.75 1.78 | 5.72 1.08 | 2.03 2.9 | 5.83 1.72 | 1.85 2.49 | 5.24 1.65 |
| High(H)    | 1.01 2.25 | 5.66 1.38 | 2.4 4.22 | 7.34 1.99 | 2.96 5.09 | 7.9 2.23 |
| H-L        | 0.62 0.99 | 3.54 0.97 | 2.07 4.02 | 4.63 3.01 | 3.57 5.12 | 5.38 3.3 |

|            | CNN       | RNN       | RNN+A     |
|------------|-----------|-----------|-----------|
|            | Pred     | Avg       | Std       | SR        | Pred     | Avg       | Std       | SR        | Pred     | Avg       | Std       | SR        |
| Low(L)     | -1.47 -0.45 | 7.16 -0.22 | -0.61 -0.07 | 8.24 -0.03 | -0.78 0.02 | 7.99 0.01 |
| 2          | -0.47 0.18 | 5.84 0.11 | 0.16 -0.03 | 6.17 -0.02 | -0.06 0.37 | 6.23 0.2 |
| 3          | -0.11 0.43 | 5.5 0.27  | 0.5 0.31  | 5.66 0.19 | 0.31 0.48 | 5.59 0.3 |
| 4          | 0.15 0.79 | 4.95 0.55 | 0.74 0.4 | 5 0.28  | 0.58 0.42 | 4.89 0.3 |
| 5          | 0.38 1.11 | 4.99 0.77 | 0.94 0.81 | 4.62 0.61 | 0.8 0.75 | 4.58 0.57 |
| 6          | 0.59 1.26 | 4.65 0.94 | 1.12 1.4 | 4.5 1.08 | 1 1.22 | 4.57 0.92 |
| 7          | 0.82 1.61 | 4.79 1.17 | 1.31 1.76 | 4.37 1.39 | 1.22 1.69 | 4.38 1.34 |
| 8          | 1.1 2.1 | 5.11 1.42 | 1.54 2.09 | 5.05 1.43 | 1.48 2.05 | 4.94 1.44 |
| 9          | 1.47 2.58 | 5.6 1.59 | 1.86 2.88 | 5.67 1.76 | 1.86 2.72 | 6.09 1.55 |
| High(H)    | 2.48 4.95 | 7.79 2.2 | 2.6 5 8.09 2.14 | 2.76 4.83 | 8.12 2.06 |
| H-L        | 3.95 5.41 | 5.09 3.68 | 3.21 5.07 | 4.34 4.05 | 3.54 4.8 | 5.06 3.29 |

|            | GRU       | LSTM      | Transformer |
|------------|-----------|-----------|-------------|
|            | Pred     | Avg       | Std       | SR        | Pred     | Avg       | Std       | SR        | Pred     | Avg       | Std       | SR        |
| Low(L)     | -1.04 -0.23 | 7.87 -0.1 | -0.63 -0.26 | 7.88 -0.11 | -0.55 -0.07 | 7.88 -0.03 |
To offer a more intuitive understanding of the portfolio performance, Figure 2 visually presents the cumulative log returns of the long and short positions of the portfolios organized by the forecast predictions of our models. The bold black baseline represents the cumulative market excess return for reference.

The solid lines illustrate the performance of the long positions (comprising stocks in the top decile of predicted returns), while the dashed lines depict the trajectory of the short positions (formed from the bottom decile of forecasted returns).

This visualization provides compelling evidence of the dominance of RNN models with memory mechanisms and Residual MLP over other models. Their corresponding solid and dashed lines consistently surpass the returns from both long and short positions of the portfolios created based on other models’ predictions.

This superior performance further reinforces the potential utility of advanced deep learning methods in financial market analysis. In particular, it underscores the value of RNNs with memory mechanisms and Residual MLP in effectively guiding investment strategies for both long and short positions, enabling superior returns relative to the market excess return.

To offer a more intuitive understanding of the portfolio performance, Figure 2 visually presents the cumulative log returns of the long and short positions of the portfolios organized by the forecast predictions of our models. The bold black baseline represents the cumulative market excess return for reference.

The solid lines illustrate the performance of the long positions (comprising stocks in the top decile of predicted returns), while the dashed lines depict the trajectory of the short positions (formed from the bottom decile of forecasted returns).

This visualization provides compelling evidence of the dominance of RNN models with memory mechanisms and Residual MLP over other models. Their corresponding solid and dashed lines consistently surpass the returns from both long and short positions of the portfolios created based on other models’ predictions.

This superior performance further reinforces the potential utility of advanced deep learning methods in financial market analysis. In particular, it underscores the value of RNNs with memory mechanisms and Residual MLP in effectively guiding investment strategies for both long and short positions, enabling superior returns relative to the market excess return.

5 Conclusion

This study presents an extensive comparative analysis of diverse deep learning methodologies within the realm of empirical asset pricing. The results unequivocally illustrate the superior predictive abilities of Recurrent Neural Networks (RNNs) with
memory mechanisms and Transformers. More importantly, it unveils the potential of these methodologies to unlock significant economic benefits for investors leveraging deep learning forecasts.

A crucial insight emerging from this study is the indispensable role of domain-specific knowledge and financial theory in the architecture and design of effective deep learning models. This underlines an exciting pathway for future research in empirical asset pricing, blending the power of deep learning with traditional financial theories.

Nevertheless, implementing predictive models in asset pricing introduces novel challenges to the domain of deep learning. These stem from the violations of the Independent and Identically Distributed (IID) assumption, which underpins most deep learning applications. This violation brings forth issues such as distribution shifts over time and the low signal-to-noise ratio endemic to financial data. These urgent problems warrant further exploration and innovative solutions.

Another noteworthy finding is that high performing networks are not always deep, thereby opening new avenues for asset pricing theory exploration. The potential to extract valuable insights from the patterns revealed by these models suggests an intriguing direction for future research. Deploying explainable AI methodologies can illuminate the economic mechanisms underpinning asset pricing, enriching our understanding of this complex domain.

The successful application of deep learning methods to predict stock returns bodes well for the future development of novel economic models. This achievement underscores the valuable role deep learning can play in fostering empirical understanding of asset prices. Furthermore, the continuous advancements in the rapidly evolving field of deep learning, including explainable AI theory, innovative architectures, mechanisms, and models, promise to catalyze further breakthroughs in asset pricing studies.

Improved risk premium prediction through enhanced measurement capabilities of deep learning methods can simplify the analysis of the intricate economic mechanisms governing asset pricing. This potential signifies a substantial contribution to the study of asset pricing.

In conclusion, the findings of this study not only endorse the applicability of deep learning methods in financial innovation but also highlight their future prospects and superiority over conventional machine learning techniques. As such, they open up a plethora of research opportunities for scholars and practitioners aiming to push the boundaries of our understanding of asset pricing using the power of deep learning.

References

[Ba, Kiros, and Hinton(2016)] Ba, J. L., J. R. Kiros, and G. E. Hinton. 2016. Layer normalization. arXiv preprint arXiv:1607.06450 .

[Bianchi, Büchner, and Tamoni(2021)] Bianchi, D., M. Büchner, and A. Tamoni. 2021. Bond risk premiums with machine learning. The Review of Financial Studies 34:1046–89.

[Fama and French(2015)] Fama, E. F., and K. R. French. 2015. A five-factor asset pricing model. Journal of Financial Economics 116:1–22.

[Goodfellow, Bengio, and Courville(2016)] Goodfellow, I., Y. Bengio, and A. Courville. 2016. Deep learning. MIT press.

[Green, Hand, and Zhang(2017)] Green, J., J. R. Hand, and X. F. Zhang. 2017. The characteristics that provide independent information about average us monthly stock returns. The Review of Financial Studies 30:4389–436.

[Gu, Kelly, and Xiu(2020)] Gu, S., B. Kelly, and D. Xiu. 2020. Empirical asset pricing via machine learning. The Review of Financial Studies 33:2223–73.

[Gu, Kelly, and Xiu(2021)] ———. 2021. Autoencoder asset pricing models. Journal of Econometrics 222:429–50.

[Heaton, Polson, and Witte(2016)] Heaton, J., N. G. Polson, and J. H. Witte. 2016. Deep learning in finance. arXiv preprint arXiv:1602.06561 .

[Hutchinson, Lo, and Poggio(1994)] Hutchinson, J. M., A. W. Lo, and T. Poggio. 1994. A nonparametric approach to pricing and hedging derivative securities via learning networks. The Journal of Finance 49:851–89.

[Ioffe and Szegedy(2015)] Ioffe, S., and C. Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International conference on machine learning, 448–56. PMLR.

[Kingma and Ba(2014)] Kingma, D. P., and J. Ba. 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 .
[Koijen and Van Nieuwerburgh(2011)] Koijen, R. S., and S. Van Nieuwerburgh. 2011. Predictability of returns and cash flows. *Annu. Rev. Financ. Econ.* 3:467–91.

[Leippold, Wang, and Zhou(2021)] Leippold, M., Q. Wang, and W. Zhou. 2021. Machine learning in the chinese stock market. *Journal of Financial Economics* ISSN 0304-405X. doi:https://doi.org/10.1016/j.jfineco.2021.08.017.

[Rapach and Zhou(2013)] Rapach, D., and G. Zhou. 2013. Forecasting stock returns. In *Handbook of economic forecasting*, vol. 2, 328–83. Elsevier.

[Sirignano, Sadhwani, and Giesecke(2016)] Sirignano, J., A. Sadhwani, and K. Giesecke. 2016. Deep learning for mortgage risk. *arXiv preprint arXiv:1607.02470*.

[Welch and Goyal(2008)] Welch, I., and A. Goyal. 2008. A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies* 21:1455–508.

[Yao, Li, and Tan(2000)] Yao, J., Y. Li, and C. L. Tan. 2000. Option price forecasting using neural networks. *Omega* 28:455–66.