

# Comparative Analysis of Deep Quantile Regression and Traditional Regression Models in Stock Price Prediction

Eddie Xu

## Table of Contents

Abstract . . . . .	1
Introduction . . . . .	1
Literature Review . . . . .	3
Multiple Regression on Short-Term Stock Price Prediction . . . . .	3
Random Forest Regression with Artificial Intelligence Techniques . . . . .	3
Deep Neural Networks on Stock Price Forecasting . . . . .	4
Deep Quantile Regression in Power and Energy Sector . . . . .	5
Research Question and Hypotheses . . . . .	6
Data Sources and Variables . . . . .	6
Statistical Methods . . . . .	7
Exploratory Data Analysis . . . . .	7
Pre-processing . . . . .	9
Discussion of Results . . . . .	10
Conclusion . . . . .	10
References . . . . .	10

## Abstract

## Introduction

Stock price prediction is a widely adopted practice in the finance industry for many years, offering valuable insights to investors and financial institutions. Accurate prediction supports informed decision-making, enhances risk management, and contributes to the development of more effective investment strategies. By anticipating future stock returns, financial professionals can allocate limited resources more efficiently, hedge against potential losses, and

capitalize on emerging opportunities that might otherwise be overlooked. Because of these benefits, it can be both a challenging and essential task due to the complexity and volatility of financial markets. To address this, various regression models have been applied to predict expected stock returns based on historical data. These include Multivariate Linear Regression, Multivariate Polynomial Regression, Random Forest Regression, and Support Vector Regression.

Each of these models offers distinct advantages in handling data complexity and capturing certain non-linear relationships or trends, making them important tools for stock price prediction. One of main strengths of traditional regression models is the simplicity and interpretability. These models provide a clear understanding on how explanatory variables influence the stock price. Another notable strength is the ease of implementation and validation. These models are supported by well-established statistical methodology, and they can be efficiently trained and validated with standard performance metric such as R-squared, Root Mean Square Error, and Mean Absolute Error. Because of that, these models often serve as the baseline for evaluating more complex machine learning models.

Despite their advantages, traditional regression models have significant disadvantages that make them less suitable for stock price prediction. One of the disadvantages is that these models typically output only the conditional mean of the target variable. This fails to capture the full distribution of possible outcomes which is crucial in a highly volatile and non-linear financial markets. Another disadvantage is the assumption of stable statistical properties over time, even though financial markets are inherently dynamic and non-stationary since the market fluctuate due to external factors. Another disadvantage is the limited interpretability on non-linear relationships. Some models oversimplify the relationships while other nonlinear models like Random Forest and Support Vector Regression provide predictions that can be hard to explain in simple terms. These issues significantly reduce the effectiveness of traditional regression models in capturing the complexities of real-world market behavior. As a result, relying on these models can lead to unreliable or even misleading predictions, posing risks and consequences for investors and financial institutions.

To address these limitations, recent advancements in machine learning have introduced modern regression models based on deep learning techniques, such as Deep Quantile Regression. Deep Quantile Regression is the extension of Quantile Regression that predict multiple conditional quantiles including the median of the target variable. It leverages deep neural networks to capture complex and non-linear relationships between the target and response variable.

In this research paper, the purpose is to explore the application of Deep Quantile Regression for stock price prediction. By leveraging the quantile loss function within a deep learning framework, the regression model would provide the entire conditional distribution of stock price returns, rather than a single expected return. This approach predicts point estimates as well as construct prediction intervals, which can offer valuable insights into the uncertainty and potential risk of price forecasts. The performance of Deep Quantile Regression on historical stock price data will be then evaluated and compare to the performance of traditional regression. In addition, the predicted values will be assessed against actual stock prices over

the 30-day forecasting horizon. This comparison aims to determine whether Deep Quantile Regression provides a meaningful improvement in predictive accuracy and uncertainty quantification, and whether its application to stock price prediction is a worthwhile advancement over conventional approaches.

## **Literature Review**

There are numerous studies that explore the stock prices prediction using various regression models. In fact, the regression models mentioned earlier have been evaluated and compared in terms of their performance and effectiveness.

### **Multiple Regression on Short-Term Stock Price Prediction**

According to Rusu, Ștefan, Boloș, M. I., & Leordeanu, M. (2024) paper, the team conducted a comparative analysis of multiple regression models to investigate the performance for short-term stock price prediction. They examined Linear Regression, Support Vector Regression, Polynomial Regression and LASSO Regression for its respective strengths in handling linearity, nonlinearity, or feature selection. The dataset was obtained from Yahoo Finance and consisted of two years of historical stock data for Apple Inc., selected due to the company's large market capitalization and market influence. They later evaluated and compared the performance of each model with standard metrics including Mean Absolute Error, R-Squared, and Mean Squared Error. While the team performed a comparative analysis on regression models, they did not include and more complex models, such as Deep Quantile Regression and compared it. Not only that, but the analysis was also based solely on data from a single company which limits the generalizability. This limitation was noted and acknowledged by the team as well.

### **Random Forest Regression with Artificial Intelligence Techniques**

Per Zheng, J., Duan Xin, Cheng, Q., Tian, M., & Yang, L. (2024) paper, the team explored the regression techniques with the application of artificial intelligences techniques in the context of smart finance. They selected Random Forest Regression as the base model because of the capability and integrated it with artificial intelligence techniques to enhance predictive performance. The modified regression was applied to predict the stock trends of three major companies (Apple, Samsung, and GE) for future time horizons of 30, 60, and 90 days based on a dataset comprising approximately 7,000 trading days. The team evaluated the performance of the model through the measurement of the Area Under the Curve (AUC) metric to assess its accuracy. Its result was compared with the performance with other regression models including Support Vector Regression, Logistic Regression, Gaussian Discriminant Analysis (GDA) and Quadratic Discriminant Analysis (QDA). This study differs from the proposed research is that Random Forest Regression was modified using artificial intelligence techniques to enhance its predictive capability. Another key difference is that the comparison did not include the Deep

Quantile Regression model, which serves as a central focus of the proposed research. Finally, the last difference is the approach of learning task. The regression models in the study were used for classification tasks like the stock trend while for the proposed research, the traditional models will be used for regression tasks and predicting the stock prices.

## **Deep Neural Networks on Stock Price Forecasting**

With advancements in machine learning, the application of deep neural network techniques for stock price prediction have been studied and evaluated in recent years. Per Omar, A. B., Huang, S., Salameh, A. A., Khurram, H., & Fareed, M. (2022) paper, the team explored the deep neural network framework and its application on stock price forecasting. They integrated the framework to three models (Autoregressive Integrated Moving Average Model, Autoregressive NN Model, and Autoaggressive Random Forest Model). Autoregressive Integrated Moving Average Model is a statical method commonly used for forecasting time series data, combining three key components of autoregression, differencing, and moving averages. Autoregressive NN Model is a forecasting approach that combines the principles of autoregression with the learning capabilities of neural networks. The model uses historical time series values as input features to the network. With that methodology, it captures both linear and non-linear temporal dependencies in the data. With those models, the team forecasted stock index prices in three different time frames: the whole period, the pre-Covid-19 period, and the Covid-19 period. For the dataset, they used the daily close price of the KSE-100 index, and it comprised a total of 5,077 observations starting from January 1, 2001, to August 20, 2021. The dataset was later divided to two subsets: pre-Covid-19 period (January 1, 2001, to February 25, 2020) and Covid-19 period (February 26, 2020, to August 20, 2021). To examine and compare the performance and effectiveness of proposed models, the team used multiple evaluation metrics including the Mean Absolute Error, Root Mean Absolute Error, Mean Absolute Percentage Error, and Correlation coefficient. While the team applied the deep learning framework to three models, they did not explore the Deep Quantile Regression model. Another difference is that the models used for the study are specifically designed for time series forecasting where past observations influence future values. This is different from the tradition regression models since they are built for general predictive modeling where observations are assumed to be independent of each other.

Per Abe, M., & Nakayama, H. (2018) paper, the team evaluated the application of deep neural networks and applied it to perform one-month ahead stock price prediction with 25 financial and market factors over a 10-year rolling timeframe. Their dataset was based on MSCI Japan Index constituents, comprised of 319 constituents of the free float-adjusted market capitalization in Japan. For the preprocessing, they ranked each financial factors across the stock at each time point and normalized it. From there, the team designed multiple Deep Neural Networks architecture starting with the model containing three hidden layers and increasing the depth to five and eight hidden layers. Its predictive performance was later compared with other regression models: Support Vector Regression and Random Forest Regression. The

models were assessed based on rank correlation, directional accuracy, and long-short portfolio returns. While the team evaluated the deep learning framework on stock price prediction, their framework was designed to generate point forecasts of expected returns using lagged financial factors. This is different from the proposal research since it focuses on applying the framework to a Quantile Regression and estimates conditional quantiles, which allows for a full distribution of stock price prediction.

## **Deep Quantile Regression in Power and Energy Sector**

While there is a growing body of research applying deep neural network techniques to stock price prediction, there remain relatively few studies utilizing Deep Quantile Regression in this context. In contrast, there are more studies that has been conducted within the power and energy sectors instead, where the model is experimented with various optimizations and evaluated for probabilistic forecasting. This highlights how Deep Quantile Regression can be adapted to different forecasting domains.

Per Yu, Y., Yang, M., Han, X., Zhang, Y., & Ye, P. (2021) paper, the team proposed a Deep Quantile Regression model designed for regional wind power forecasting purposes. The team used a combination of power historical data and numerical weather prediction data. To evaluate the performance of the proposed model, they compared it with other models as benchmarks including the Quantile Regression Neural Network model, sparse Bayesian learning model, and individual forecast power accumulation method. There are various key differences between this study and the proposed research. One of the differences is that the study focused on regional wind power forecasting using historical energy data, while the proposed research explores the application of Deep Quantile Regression in stock price prediction, where the input features are financial indicators and lag returns instead. Another key difference is the approach on comparative analysis. The team in the study compared their Deep Quantile Regression model against other probabilistic forecasting models, while the proposed research aims to compare the Deep Quantile Regression with traditional regression models to assess their relative forecasting performance.

Per Zhu, J., He, Y., Yang, X., & Yang, S. (2024) paper, the team examined the application of Deep Quantile Regression on ultra-short-term wind power probabilistic forecasting. They developed a non-crossing multi-output quantile regression deep neural network, optimized through chaotic particle swarm optimization. This modified model leveraged a multi-output deep neural network to output all quantile estimations simultaneously through a single training process. Various datasets were used for this study as they were derived from two case studies in wind power. The team utilized datasets derived from two case studies in wind power forecasting: the first consisting of wind power data from Ontario with a one-hour resolution, and the second using numerical weather prediction data from the Global Energy Forecasting Competition 2014. All datasets were split into 70% for model training, 15% for validation, and the remaining 15% for testing. Given the model's complexity, the continuous ranked probability score was used to assess the performance because it can effectively measure the

compatibility of the predicted distribution with the observation. While the study demonstrated the performance of the model on wind power prediction and provided detailed insights that can be shared with other fields like stock price prediction, the model is more complex from the model used in the proposed research. The Deep Quantile Regression used in the proposed research has a more straightforward structure and is not optimized. This simplification is intentional because the purpose is to balance model interpretability and predictive accuracy within the context of financial forecasting.

Per Lu, S., Xu, Q., Jiang, C., Liu, Y., & Kusiak, A. (2022) paper, the team explored the performance of a Deep Quantile Regression model integrated with non-crossing constraints and sparse-group lasso regularization for probabilistic electric load forecasting. The non-crossing constraint was implemented to ensure that the predicted quantiles remained properly ordered, preventing unrealistic quantile crossing that can occur when each quantile is estimated independently. The sparse-group lasso regularization was introduced to enhance model simplicity by eliminating unimportant signals and emphasizing the most relevant features. The dataset used in the study consisted of daily power usage data from four million households collected between 2016 and 2018 from China’s power supply system database. To evaluate the model’s performance, the team used various metrics including Continuous Ranked Probability Score, Average Quantile Loss, and the Monotonicity of quantile outputs. The approach on Deep Quantile Regression from this study is different from the proposed research in many key aspects. For starters, the model in their study is more complex, integrating non-crossing constraints and sparse-group lasso regularization, while the proposed research utilizes a more straightforward Deep Quantile Regression model. Not only that, the objective of evaluation is also very different.

## Research Question and Hypotheses

**Research Question:** what extent Deep Quantile Regression outperforms traditional regression models (Multivariate Linear, Multivariate Polynomial, Random Forest, and Support Vector Regression) in modeling the distribution of stock returns over the 30-day forecasting horizon?

- **H1 Hypotheses:** Deep Quantile Regression provides more accurate and informative predictions of stock return distributions than traditional regression models.
- **Null Hypotheses:** Deep Quantile Regression does not perform significantly better than traditional regression models in predicting stock return distributions.

## Data Sources and Variables

For this research, the primary data source is the historical daily stock price data for a selected group of publicly traded companies and will be obtained from Yahoo Finance. Yahoo Finance is selected to ensure consistent, reliable, and reproducible data collection. Since the platform

maintains continuous historical data, the dataset contains the fundamental stock information with no missing or unknown data. The selected companies are the top 25 companies of the S&P 500 index, based on the market capitalization. This approach captures a diverse representation of different industries and company types. There are eight variables in the dataset and six of them are numeric. They are defined as below:

- ***Date***: the timeframe of the stock
- ***Ticker***: the stock symbol
- ***Open***: the opening price of the stock
- ***High***: the highest price of the stock during the period
- ***Low***: the lowest price of the stock during the period
- ***Close***: the closing price of the stock during the period
- ***Adjusted Close***: the adjusted closing price of the stock after corporate action during the period
- ***Volume***: the total number of shares traded during the period

## Statistical Methods

For statistical methodology, the approach will be a comparative analysis between traditional regression models and Deep Quantile Regression. Unlike traditional regression models where it output a single conditional mean prediction, Deep Quantile Regression applies the quantile loss or pinball function to directly estimate specific quantiles. Because of that, this allows the model to provide a better understanding of the full distribution of the target variable, which is useful in stock price prediction. By applying both traditional regression models and Deep Quantile Regression to the same dataset, this research examines and evaluates the strengths and limitations of each approach on stock price prediction.

## Exploratory Data Analysis

Once the data was successfully obtained, an exploratory data analysis was performed to assess data qualities and gain data insights before facilitating the model development and evaluation. The dataset is divided into two subsets. The first subset covers the period from March 2015 to March 2025, and it comprises of 61,489 observations. This subset is used for the training and testing purposes, providing a robust size for Deep Quantile Regression and traditional regression model comparisons. The second subset covers the period from April 2025 to May 2025, and it comprises of 3,150 observations. This subset is served as a validation period and will be compared to the predicted value in the same timeframe.

The descriptive statistic was applied on the numerical variables to understand their distributional properties. The results below show that the mean values are relatively high, which can be attributed to the presence of companies with large market capitalizations such as, Apple and Amazon, in the dataset. The statistics also reveal a wide range across all variables. This

indicates that there is a heteroscedasticity since the data appears to be highly volatile and non-stationary. This would impact the precision of traditional regression models, especially for multivariate linear and polynomial regression since the assumption of constant variance is violated.

	Open	High	Low	Close	Adj Close	Volume
count	61489.0	61489.0	61489.0	61489.0	61489.0	61489.0
mean	153.54670666159146	155.21145368158236	151.85603552673726	153.58026945325855	145.4639004177016	45368442.02540292
std	148.00811496307443	149.62143833153874	146.32467774870366	148.0128387225048	146.69978520243478	108850693.9958367
min	0.481249988079071	0.4882499873638153	0.4772500097751617	0.4827499985694885	0.4672989547252655	491000.0
25%	52.323333740234375	52.84999847412109	51.6875	52.27000045776367	47.18853378295898	5455700.0
50%	113.37000274658205	114.58999633789062	112.22000122070312	113.4499969482422	102.18753814697266	13083200.0
75%	189.3999938964844	191.3999938964844	187.00999450683597	189.3300018310547	182.47000122070312	35804800.0
max	1076.47998046875	1078.22998046875	1068.010009765625	1076.8599853515625	1072.45361328125	3692928000.0

Figure 1: Descriptive Statistic

In addition, the mean is larger than the median for all variables, which indicates each distribution may be right-skewed and have a long tail on the high end. This interpretation is further supported by the standard deviation being larger compared to the mean, showing that the data is very spread out. The histogram below reinforces the analysis as it shows each distribution is indeed right-skewed.

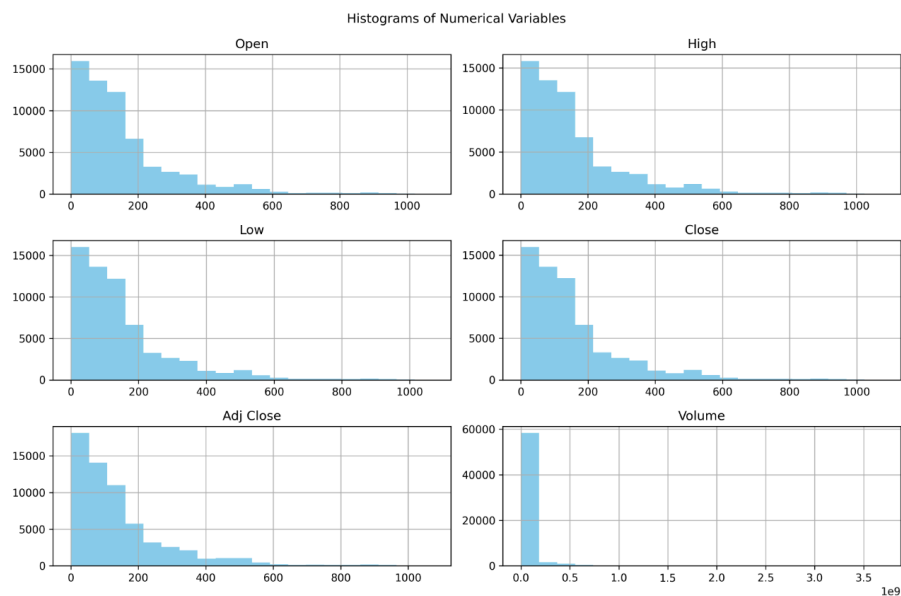


Figure 2: Histogram



## Pre-processing

Due to the findings identified during the exploratory data analysis, the original variables were not suitable as features for both traditional and Deep Quantile Regression models. Because of that, the dataset was preprocessed for model development and evaluation. Multiple features were engineered to address problems such as the skewness observed in original variables. In addition, they also capture trend behaviors and predict short-term directional movement that may influence future stock prices. This enables the models to be trained on various aspects of market complexity. The following engineered features are defined as below:

- ***Lagged Returns:*** the percentage change in the adjusted closing price of a stock from the prior day
- ***1-Day Forward Returns:*** the percentage change in value from the previous day's closing price to the current day's closing price
- ***5-Days Forward Returns:*** the percentage change in value from the closing price from 5 days ago to the current day's closing price
- ***5-Days Rolling Standard Deviation:*** the measure of the stock's volatility over the past 5 days period
- ***21-Days Rolling Standard Deviation:*** the measure of the stock's volatility over the past 21 days period
- ***Relative Strength Index (RSI):*** the momentum indicator that measure the rate and volatility of the stock's price
- ***Simple Moving Average:*** the average closing price over the timeframe of 20 days
- ***Moving Average Convergence Divergence (MACD):*** the momentum indicator that identifies a difference between two moving averages
- ***Simple Moving Average Ratio:*** the ratio of the current price to the 20-day Simple Moving Average
- ***Natural logarithm of Trading Volume*** the logarithm of trading volume

There are missing data for some of the features because they need prior values that exist outside of the dataset. Because of that, the missing value are filled with forward and backward fill methodology. All features were scaled to ensure that they contributed equally during model training.

The dataset was then divided into three subsets: 70% for training, 15% for hyperparameter tuning, and the remaining 15% for final performance evaluation. This approach was to provide a robust and unbiased assessment for each models' predictive performance. The predictive performance of Deep Quantile Regression was evaluated and compared against traditional regression models. All models used the engineered features as input variables, and the target variable was the adjusted closing price for future time horizons of 30 days. For Deep Quantile Regression, the input variables was fed into the feedforward neural network with 2 hidden layers. Each model was trained independently to ensure consistency and enables a direct comparison of each model's learning ability on historical stock price and predictive capability on forecasting stock prices.

To assess and evaluate the predictive performance for all regression models, various metrics were used. This approach provided a comprehensive evaluation of both point prediction accuracy and uncertainty estimation. The following metrics are defined below:

- **Root Mean Squared Error:** the average of absolute difference between the predicted and actual value. It is used to measure how far the predicted values are from the observed actual values.
- **Mean Absolute Error:** the square root of averaged squared error. It measures the average magnitude of the prediction errors from both underfitting and overfitting direction
- **Pinball Loss or Quantile Loss:** the evaluation metric that is specifically designed for Deep Quantile Regression. It measures how well the predicted quantiles align with the actual outcomes
- **Prediction Interval Coverage Probability:** the evaluation metric that measures the reliability of prediction intervals

## Discussion of Results

## Conclusion

## References

- Abe, M., & Nakayama, H. (2018). Deep Learning for Forecasting Stock Returns in the Cross-Section. arXiv.Org.
- Chronopoulos, I., Raftapostolos, A., & Kapetanios, G. (n.d.). Deep Quantile Regression. Retrieved October 12, 2025, from <https://www.kcl.ac.uk/business/assets/pdf/dafm-working-papers/2021-papers/deep-quantile-regression.pdf>
- Lu, S., Xu, Q., Jiang, C., Liu, Y., & Kusiak, A. (2022). Probabilistic load forecasting with a non-crossing sparse-group Lasso-quantile regression deep neural network. *Energy (Oxford)*, 242, Article 122955. <https://doi.org/10.1016/j.energy.2021.122955>
- Omar, A. B., Huang, S., Salameh, A. A., Khurram, H., & Fareed, M. (2022). Stock Market Forecasting Using the Random Forest and Deep Neural Network Models Before and During the COVID-19 Period. *Frontiers in Environmental Science*, 10. <https://doi.org/10.3389/fenvs.2022.917047>
- Rusu, Ștefan, Boloș, M. I., & Leordeanu, M. (2024). COMPARATIVE ANALYSIS OF REGRESSION MODELS FOR STOCK PRICE PREDICTION: LINEAR, SUPPORT VECTOR, POLYNOMIAL, AND LASSO. *Revista de Studii Financiare*, 9(17), 143–156. <https://doi.org/10.55654/JFS.2024.9.17.09>
- Yu, Y., Yang, M., Han, X., Zhang, Y., & Ye, P. (2021). A Regional Wind Power Probabilistic Forecast Method Based on Deep Quantile Regression. *IEEE Transactions on Industry Applications*, 57(5), 4420–4427. <https://doi.org/10.1109/TIA.2021.3086077>

- Zheng, J., Duan Xin, Cheng, Q., Tian, M., & Yang, L. (2024). The Random Forest Model for Analyzing and Forecasting the US Stock Market in the Context of Smart Finance. arXiv.Org.
- Zhu, J., He, Y., Yang, X., & Yang, S. (2024). Ultra-short-term wind power probabilistic forecasting based on an evolutionary non-crossing multi-output quantile regression deep neural network. *Energy Conversion and Management*, 301, Article 118062. <https://doi.org/10.1016/j.enconman.2024.118062>