

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
Research Question or Hypotheses	3
Data Sources and Variables	3
Statistical Methods	3
Conclusion	3

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

Research Question or Hypotheses

Data Sources and Variables

For this research, the primary data source will be 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, it is expected that the dataset will contain the fundamental stock information and will not have any missing or unknown data. The selected companies will be the top 25 companies of the S&P 500 index, based on the market capitalization. This will capture a diverse representation of different industries and company types. To facilitate model development and evaluation, the data set will be 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 will be 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 will be served as a validation period and will be compared to the predicted value in the same timeframe.

The dataset contains 10 numerical columns, and it includes key variables that are needed for both Deep Quantile Regression and traditional regression models to forecasting future stock price. The key variables 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

Conclusion