Eddie Xu

Professor Arthur O'Connor

DATA 698: Master's Research Project

September 10, 2025

Research Proposal: A Comparative Study of Deep Quantile Regression and Traditional Regression Models in stock price prediction

Introduction

Stock price prediction is a widely adopted practice in the finance industry, offering valuable insights to investors and financial institutions. Various regression models have been employed to forecast expected stock returns, including Linear Regression, Polynomial Regression, Random Forest Regression, and Support Vector Regression (SVR). Linear Regression is the most basic and commonly used model, primarily focusing on identifying the linear relationship between stock prices and time or other explanatory variables. Polynomial Regression builds upon this by incorporating polynomial terms to capture non-linear trends in historical data. Random Forest Regression introduces an ensemble learning approach that constructs multiple decision trees and aggregates their outputs to enhance accuracy and reduce the risk of overfitting. Support Vector Regression, on the other hand, attempts to find the optimal hyperplane that best fits the data within a specified margin of tolerance, making it suitable for complex, non-linear data structures.

Despite their strengths, these traditional regression models primarily output the conditional mean of the target variable. This limitation prevents them from capturing the uncertainty and asymmetry of stock return distributions, which are crucial considerations in volatile financial markets. As a result, relying solely on point estimates may lead to less effective investment decisions and risk management strategies.

To overcome these limitations, the research proposal aims to explore the application of Deep Quantile Regression for stock return prediction. Deep Quantile Regression is the standard Quantile Regression that uses deep neural networks to predict multiple quantiles or percentiles.

Unlike traditional models, Deep Quantile Regression estimates multiple conditional quantiles of stock price returns. By modeling the 5^{th}, 50^{th}, 95^{th}, and 99^{th} percentiles, it enables the construction of prediction intervals that reflect a bigger picture of potential price movements. These intervals can offer more informative and robust guidance for financial decision-making.

Literature Review

There are numerous studies that explore the stock prices prediction using machine learning models. In fact, many regression models mentioned earlier have been evaluated and compared. Per Zheng, J., Duan Xin, Cheng, Q., Tian, M., & Yang, L. (2024) paper, the team explores the machine learning techniques and decided to leverage the Random Forest Regression combined with artificial intelligence. They evaluate the performance through the measurement of the AUC and later compare the performance with SVM, logistic regression, Gaussian discriminant analysis and quadratic discriminant analysis. This is different from the proposal since the random forest is applied with artificial intelligence, and they did not compare it with the Deep Quantile Regression model.

The application of deep neural network techniques has been explored in the finance industry recently. Per Omar, A. B., Huang, S., Salameh, A. A., Khurram, H., & Fareed, M. (2022) paper, the team explores the deep neural network model and applies it to three models (Integrated Moving Average Model, NN Model, and Random Forest Model) and forecast stock index prices in three different time frames: the whole period, the pre-Covid-19 period, and the Covid-19 period. While the team applies the Deep Learning to three models, they did not explore the Deep Quantile Regression model. Per Wang, J., Wang, S., Lv, M., & Jiang, H. (2024) paper, the team explores and applies the Deep Quantile Regression in forecasting Value at Risk and Expected shortfall in a heterogeneous market. This is different from the proposal because it focuses on predicting the standard risk, not stock price. They also did not evaluate and compare the performance with traditional regression models.

Research Question and Hypotheses

__Research question__: To what extent does Deep Quantile Regression outperform traditional regression models (Linear, Polynomial, Random Forest, and SVR) in modeling the distribution of stock returns?

__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/Variables and Research Methods

In terms of research methods, the primary data source will be the historical daily stock price data for a selected group of publicly traded companies. The data will cover the period from 2015 to 2025 and will be obtained from Yahoo Finance to ensure consistent and reproducible data collection. Exploratory data analysis will be performed to assess data quality and gain data insights before the regression modeling. The key variables in this study will include time (as the temporal dimension) and stock prices, which will serve as the basis for feature engineering and predictive modeling.

The data will be preprocessed and engineer features such as lagged returns will be created. After that, it will go through all traditional regression models and the Deep Quantile Regression. For traditional regression model, they will be evaluated by the AUC as well as the Root Mean Squared Error and the Mean Absolute Error. Pinball Loss (quantile loss) and Prediction Interval Coverage Probability (PICP) are going to be used to evaluate the performance of Deep Quantile Regression.

In conclusion, the application of Deep Quantile Regression remains underutilized, particularly within the financial markets. Unlike traditional point forecasting methods, Deep Quantile Regression offers a probabilistic approach that can capture the uncertainty and distributional characteristics of stock returns. By providing not just predictions but also confidence intervals, the model presents a promising alternative framework for stock forecasting that merits further exploration and adoption in financial analytics.

Reference

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.

Wang, J., Wang, S., Lv, M., & Jiang, H. (2024). Forecasting VaR and ES by using deep quantile regression, GANs-based scenario generation, and heterogeneous market hypothesis. Financial Innovation, 10(1), Article 36. https://doi.org/10.1186/s40854-023-00564-5

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