

Project Progress Report

This project aims to conduct a comprehensive comparison between traditional financial models and machine learning models for equity valuation. The primary objective is to assess the predictive accuracy of both approaches when using the same set of variables, ensuring a controlled evaluation of their performance. By applying these models to identical datasets, the analysis will isolate the impact of the modeling techniques themselves rather than differences in data inputs. Additionally, the study will be segmented by industry sectors to account for variations in market structure, volatility, and factor sensitivities. This sectoral analysis will help uncover how model performance differs based on industry-specific characteristics, providing insights into the robustness, adaptability, and limitations of both traditional and ML models in various contexts. Ultimately, the goal is to identify which modeling approach offers superior predictive capabilities and to understand the conditions under which each model excels.

This Week

1. **Academic Paper Review**

I continued reviewing academic literature to deepen my understanding of traditional asset pricing models and advanced ML techniques relevant to the project. The papers reviewed provide theoretical foundations and practical applications of traditional valuation models, and various ML model techniques using financial data. This literature review has helped refine the model selection criteria and provided insights into common challenges in financial predictive modeling.

2. **Data Preparation Progress**

I finalized the data frequency and time span for the analysis to ensure alignment with the structure of financial returns and factor loadings. This step was critical to maintaining consistency in time-series data modeling. Additionally, my WRDS account has been successfully unfrozen granting access to essential datasets required for data collection.

3. **Technical Planning**

I developed a structured approach for the analysis, focusing on parallel model runs. For each traditional model, I will run multiple ML models (OLS, Support Vector Regression, Random Forest Regression, Quantile Regression, Quantile Regression Forests, Gradient Boosting Machines and Neural Networks) using the same set of factors. This will control for variable selection bias and allow for a direct performance comparison. The evaluation framework will rely on error metrics. Additionally, hypothesis testing will be used to statistically assess whether observed differences in model performance are significant.

Next Week

1. **Determine Sectors and Equities**

I will identify specific sectors and equities to be included in the dataset. This will involve defining inclusion criteria based on market capitalization, trading volume, data completeness, etc.

2. **Data Collection and Preparation**

Once the sectors and equities are determined, I will begin downloading data from WRDS. The initial phase of data preparation will include data cleaning, factor creation, etc.

3. **Continued Literature Review**

I will continue reviewing academic papers to deepen my understanding of advanced ML techniques, particularly focusing on hyperparameter tuning for Random Forests, kernel optimization in Support Vector Regression, and architecture design in Neural Networks.