

KPMG 4 First Progress Report

Names and UNI here

October 2022

1 Introduction

2 Methodology

2.1 Problem Formulation and Modeling Approach

Our final goal of this project is to build a model with good performance in multi-variate financial time series forecasting. To achieve this goal, we start from the simplest model and try to scale up the problem.

In the first stage(current stage) we try to build univariate classic ARIMA models for each individual features.

For each variable x

Feature: (x_1, x_2, \dots, x_t)

Target: $(x_{t+1}, \dots, x_{t+k})$

In the second stage, we would want to build a multivariate regression model with some classic ensemble learning models like randomforest or xgboost.

Let the multivariate feature be $\vec{x}_t = (x_{t1}, x_{t2}, \dots, x_{tm})$ at time t , for each feature x_m we need to build an individual model, that is, for the i_{th} dimension

Feature: $(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_t)$

Target: $x_{(t+1)i}$

In the final stage, we would like to explore the usage of deep learning, and see if it can be more powerful than traditional approaches.

Feature: $(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_t)$

Target: $(\vec{x}_{t+1}, \vec{x}_{t+2}, \dots, \vec{x}_{t+k})$

2.2 Related Literature

Currently there are two papers we would like to reference from.

1, A Multi-Faceted Approach to Large Scale Financial Forecasting

The paper has a very similar approach as we do, starting from a simple

baseline and gradually scale up, which can serve as a nice guideline.

2, A Transformer-based Framework for Multivariate Time Series Representation Learning

Transformer is a powerful technique which is widely used in representation learning like NLP. We would like to explore its usage in structured data regression problem.

3 Results

3.1 Feature Extraction

3.2 Uni-variate Modeling Results