Anexo I. Registro del Título del Trabajo Fin de Grado (TFG)

NOMBRE DEL ALUMNO: Álvaro Olivié Molina

PROGRAMA: GITT+BA GRUPO: 5º FECHA: 27/10/2023

Director Asignado: Coronado Vaca, María

Título provisional del TFG:

Machine Learning and Big Data in Factor Investing for Bonds

Firma del estudiante:

Fecha: 27/10/2023

Table of Contents

Provisional Index

The proposed index is the following:

1. Introduction
    a. Objectives
    b. Justification
    c. Methodology
    d. Structure

2. State of the Art

3. Project Development
    a. Database
    b. Algorithm

4. Conclusions

Objectives

The project proposes to understand what factors impact bond returns and create a model that can build a profitable portfolio with the use of machine learning and big data. There are two main objectives, understand what factors most heavily affect prices and how these factors affect the price. The next objective is to then use these insights to develop a model that can understand the market and develop a portfolio with as low risk as possible and the highest possible returns.

Another secondary objective is to investigate the effect new factors like ESG (Environmental, Social, and Corporate Governance) have on the market and if there is a correlation between ESG rating and bond prices. ESG is a rating that focuses on the sustainability of the rated entity.

Methodology

The process to follow for this project is the following. The project will be divided into three stages. The first stage is dedicated to research. It is necessary for a project looking into financial topics to first research the existing knowledge as there is already extensive work done on the matter. This research creates a starting point from which to develop the project and will speed up the following stages due to a deeper understanding of the data to be analyzed.

The next stage is understanding the data and apply simple models to gain a broad idea of the inner workings. The data to be used is a database of bonds over time and includes information and price data for each bond. During this stage components such as trends, possible outliers and gaps in the data will be investigated and considered for the final stage. It is also important to use this stage to understand link anomalies in the data with real world events. This allows us to situate the data in a broader context. With this understanding we can head into the final stage with a clear idea of what the code should do with the data and what is to be expected.

The final stage is the development stage. During this program will be created and tested. The code development includes multiple sub-stages as heavy testing is required to create an accurate model. Multiple different models and iterations of the same model will be needed to produce the desired results.

Bibliography

[1] Arnott, R., Harvey, C. R., & Markowitz, H. (2019). A backtesting protocol in the era of machine learning. *The Journal of Financial Data Science*, 1(1), 64–74. https://doi.org/10.3905/jfds.2019.1.064

[2] Kaufmann, H., Messow, P., & Vogt, J. (2020). *Boosting momentum*. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3668928

[2] Simonian, J., Wu, C., Itano, D., & Narayanam, V. (2019). *A machine learning approach to risk factors: A case study using the fama–french–carhart model.* The Journal of Financial Data Science, 1(1), 32–44. https://doi.org/10.3905/jfds.2019.1.032