POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Black box portfolios replication: techniques and analysis

Tesi di Laurea Magistrale in
Mathematical Engineering - Quantitative Finance

Author: **Edoardo Villa**

Student ID: 953275
Advisor: Prof. Daniele Marazzina
Co-advisors: Raffaele Zenti, Luigi Simeone
Academic Year: 2022-23

Abstract

Portfolio replication is a widespread problem in finance. Several investors, institutional or not, aim to build stable and efficient replicating strategies. The purpose of this study is to analyze the problem from a broader perspective. Two different models are proposed, then some analysis are executed and applications are shown. The models are based on the assumption, useful and widely used in finance, that the target portfolio return is a linear combination of the returns of its constituent assets. To build the replicating portfolio two separated set of factors are available, the five Fama and French factors and a collection of US Indexes. After the data presentation and analysis, the models are presented. Both models have an elastic net regression at their core, but the approaches are different: the former is based on rolling windows, where the portfolio weights are dynamically adjusted; the latter is static, with rebalancing that not allowed. Firstly, models are tested replicating 13 target portfolios, which are divided into sectorial and diversified according to their composition. Finally, a portfolio is selected to perform a more in-depth analysis and to show what can be done with the replicating portfolios built with the models of the study.

Keywords: portfolio replication, elastic net, rolling window, risk management metrics, fama and french models

Abstract in lingua italiana

La replica dei portafogli è un problema molto diffuso in ambito finanziario. Diversi investitori, istituzionali e non, mirano a costruire strategie di replica stabili ed efficienti. Lo scopo di questo studio è quello di analizzare il problema da una prospettiva più ampia. Vengono proposti due diversi modelli, quindi vengono eseguite alcune analisi e mostrate le applicazioni. I modelli si basano sull'ipotesi, utile e largamente utilizzata in finanza, che il rendimento del portafoglio target sia una combinazione lineare dei rendimenti dei suoi componenti. Per costruire il portafoglio di replica sono disponibili due insiemi separati di fattori, i cinque fattori di Fama e French e una collezione di indici US. Dopo la presentazione e l'analisi dei dati, vengono presentati i modelli. Entrambi i modelli hanno alla base una regressione elastic net, ma gli approcci sono diversi: il primo si basa su rolling windows, in cui i pesi del portafoglio vengono aggiustati dinamicamente; il secondo è statico, con il ribilanciamento che non è consentito. In primo luogo i modelli vengono testati replicando 13 portafogli target, suddivisi in settoriali e diversificati in base alla loro composizione. Alla fine, viene scelto un portafoglio per effettuare un'analisi più approfondita e per mostrare cosa si può fare con i portafogli di replica costruiti con i modelli dello studio.

Parole chiave: replica di portafogli , elastic net, rolling window, metriche di gestione del rischio, modelli di fama and french

Contents

1 | Introduction

The term portfolio replication refers to the set of strategies and techniques used to build a portfolio that mimics the performance of another one, the so called *target portfolio*, without knowing its holdings. The goal of the replica is to reproduce the risk and return characteristics of the original portfolio as closely as possible, while minimizing tracking error or deviations.

In most of the cases it is not known which are the securities contained in the portfolio, and which are their weights. The only feature observable from the outside are the returns; it can be said that this is a *Black Box* problem. A Black Box problem refers to a situation where the internal workings or mechanisms of a system are not visible or understood from the outside. This definition fits well the topic of this study, in fact, for example, a portfolio manager may use complex algorithms and trading strategies to run his portfolio, but the exact steps and parameters used in his model are not disclosed in public. This can make difficult to understand the drivers of the returns and find the best strategy to replicate them.

There are several reason for which investors, institutional or not, might want to build stable and profitable replicating strategies. The first one is to build alternative investments clones. Alternative investments are financial assets that are not part of the conventional investment categories like stocks, bonds and cash. They are also known as "non-traditional" investments and they are often riskier, more complex and less regulated. Some examples of alternative investments are: hedge funds, private equity, real estate, commodities, derivatives and cryptocurrencies. These assets are sometimes very illiquid and have really high fees, raising investor's will to build his own replica portfolio to not be affected by these problems.

Investors may be interested in the replica also for risk management purposes. The clone shows which are the factors that drive portofolio returns; consequently the investor is aware at which risks is exposed to and, if he wanted, he would have the information needed to hedge his positions. Knowing the drivers of an investment can be useful in models where there is a multiple scenario simulation: instead of lots of securities only these factors are simulated, with a less demanding operation.

Another application of portfolio replication techniques is index tracking, where the index is treated as a portfolio. In particular, it can be really useful with indexes with these two features: illiquid and big. [1] In the first case, considering the difficulty and in some situations the impossibility to buy an illiquid security, the investor is able to replicate the index buying the liquid factors that drive the returns. In the second case the strategy let an investor, who does not have under management enough asset to physically replicate the index buying all its components, to be able to replicate this big security. Related to the index topic, portfolio replication is also important for passive investment strategies, such as index funds or ETFs, [2] which aim to replicate the performance of a particular market index.

An aspect that is often ignored in academic studies but that plays an important role is taken up by the transaction costs. Replicating a portfolio can help us understand the costs associated with trading the original portfolio; this can be useful for evaluating the efficiency of trading strategies or for optimizing trade execution.

Until now it have been mentioned only possible replication strategy's usage from a "buyer" perspective: there is more. There are companies, which have customers that are not eligible by the MiFID II[3] regulation, that sell instruments that are not UCITS[4]. These companies may want to create another product with similar features tracking the previous one, by using only securities to keep this new product UCITS and sell it to retail investors.

Despite the replica is a difficult operation, there are some strategies actually available in the literature that have a good performance, that are interpretable and easy to implement. In this study replica portfolios are built using two different approaches: the first is based on *rolling windows* and rebalancing, while in the second the chosen portfolio is kept constant over time. There is no evidence that one approach is universally better than the other, but the choice depends on the scenario in which the model is used, not only considering the mere replication but also for the further risk management analysis that are exploited with the obtained portfolio.

[1] Indexes that require billions to be replicated

[2] ETFs('Exchange Traded Fund') are financial instruments that replicate the performance a particular index, sector, commodity, or other assets. ETFs advantage is that they can be purchased or sold on a stock exchange the same way that a regular stock can.

[3] Markets in Financial Instruments Directive 2014 (2014/65/EU) commonly known as MiFID 2 (Markets in financial instruments directive II),is a legal act of the European Union. Together with Regulation (EU) No 600/2014 it provides a legal framework for securities markets, investment intermediaries, and trading venues.

[4] UCITS stands for "Undertakings for Collective Investment in Transferable Securities." UCITS funds are a type of investment fund that is authorized to be sold to investors throughout the European Union. They are regulated by the European Union, and must comply with a set of standards that are designed to protect investors

During the whole development of this study a constant dialogue with R. Zenti and the Risk Management and Data Analytics team of Fondaco SGR, which has an open interest in alternative investments and portfolio replication, has taken place. They have provided, in addition to the dataset, knowledge and ideas for the success of the study.

1.1. Literature Review

The portfolio replication is inherent in a more attentioned, debated and famous problem vastly present in the literature: hedge found cloning. In fact, trying to build a clone that tracks an hedge fund is an exercise of portfolio replication. These funds are typically organized as private investment vehicles for wealthy individuals and institutional investors. Since they are not required to disclose their activities to the public, little is known about the risk of hedge fund strategies. The lack of transparency and the fear of style drift have raised the question of whether it is possible to identify and estimate the risk factors that drive hedge fund returns and to build a model to replicate them.
A first analysis has been done by Sharpe [27], who introduced the concept of style analysis; it consists in separate portfolio's returns into a set of underlying factors that explain its performance. This approach helps investors better understand the sources of risk and return in their portfolio and can help them in the construction of more efficient and diversified portfolios. Sharpe approach is used by Hasanhodzic and Lo [15] to build a linear factor model. They show that using a set of factors such as market risk, interest rate risk and credit risk it is possibile to replicate the returns of many hedge funds.
Nevertheless, looking at empirical conclusion of other studies, several limits of the factor-based hedge fund replication can be found. Agarwal and Naik [1], Fung and Hsieh [12] and Vrontos, Vrontos and Giamouridis [29] with their studies highlight that, due to the lack of transparency of hedge fund trading strategy, it is not possible fix a priori a set of factors. With the fixed set, the replication strategy allocates capital to all assets and it is not able to distinguish between important and less important factors.
A work close to this study is Giamouridis and Paterlini [13], who propose a new method for the construction of hedge funds clones. Their model minimize the tracking error volatility between the hedge fund index and the clone while imposing a constraint on the 1-norm and 2-norm of the factors. The procedure is equivalent to minimizing a penalized version of the tracking error volatility, with the penalty that allows us to reduce the sensitivity to the estimation error and, with the 2-norm, is addressed possible instability in the factor estimates due to the presence of collinearities among factors.
Recently several studies developed with new machine learning techniques came out, but most of them are more focused on the prediction rather than on the construction of an

interpretable replica portfolio.

1.2. Thesis approach

This thesis aims to show techniques of replication of the target portfolios, comparing different models used with different set of factors, and to present some possible analysis that can be done using the outcome portfolio.
The returns of a portfolio are a linear combination of the returns of the assets contained in it; for this reason, a linear model fits well the problem and it is able to obtain a clone that is close to the target. Models in this study are based on linear regression, where the coefficients of the regression are the weights of the portfolio; in particular, an Elastic Net [32] regression has been chosen to better handle multicollinearity between factors and relevant features selection. Two different approaches are exploited: in the first one the replica is built using *rolling windows* and rebalancing, the portfolio is not constant over the time, but at each time step it is built considering only the past returns contained in the rolling window; in the second one the portfolio is built using all the past returns and it is no more rebalanced, it is kept constant. Both models are evaluated twice using two different set of factors, first 14 US indexes and then the 5 Fama and French factors [7]; the two sets are never mixed.
The results of the two approaches and two different datasets are compared and it is analyzed in which circumstances one is more appropriate than the other. After that, the replica portfolio of one target portofolio is chosen. Deeper analysis are executed on it, such as adding stochasticity to the weights of the target portfolio, and some possible applications, such as the VaR calculation, are shown.
It's important to highlight that the models of this study have been thought with the goal of obtaining the exact structure of the target portfolio, its weights. Some techniques, that maybe would have led to better prediction results, have been discarded because of the fact that they are less interpretable and do not provide the weights.

1.3. Thesis structure

This study is structured as follows:

- chapter 2 : it contains the explanation of why a portfolio replication problem can be solved using a linear regression model; in particular, the Elastic Net choice is motivated. In addition, the approach used to build the models, which is the most common in a machine learning framework, is explained;

- chapter 3: it contains the description of the dataset and the explaination of Fama and French models. Collections of time series of prices and returns available for this study are two: a set of 14 US Indexes and the 5 Fama and French factors;
- chapter 4: it contains a short analysis of the dataset. Stationarity of returns is tested and some general statistics of the factors are inspected.
- chapter 5: it contains the descriptions of the two models object of these study: *Model 1*, which is dynamic and based on *rolling windows*, and *Model 2*, which is static;
- chapter 6: it contains the results of the two models in replicating 13 target portfolios;
- chapter 7: it contains some test and analysis using the replicating portfolio of one among the 13 target portfolios. Firstly, some stochasticity is added to the weights of the target portfolio. Then the result obtained with an Elastic Net inside *Model 1* is compared with the result obtained with Lasso. Gross exposure, turnover and transaction costs are analyzed later. In the end, some applications of the replicating portfolio are presented; two risk metrics are calculated (VaR and ES) and Ibbotson Cone associated to the replicating portfolio is built;
- chapter 8: it contains conclusions and future developements.

2 | Math and machine learning behind the replica

The models are based on the assumption, useful and widely used in finance, that the target portfolio price is a linear combination of the prices of its constituent assets. As well as for the prices, this relation holds for returns.
The equations that drive the values are the following:

$$z_t = q_{1,t}p_{1,t} + q_{2,t}p_{2,t} + q_{3,t}p_{3,t} + ... + q_{n,t}p_{n,t} \tag{2.1}$$

for the prices, and

$$y_t = w_{1,t}r_{1,t} + w_{2,t}r_{2,t} + w_{3,t}r_{3,t} + ... + w_{n,t}r_{n,t} \tag{2.2}$$

for the returns. The variables are:

- z_t is the price of the target portfolio at time t;
- $q_{i,t}$ is the quantity of asset i contained in the portfolio at time t;
- $p_{i,t}$ is the price of asset i contained in the portfolio at time t;
- y_t is the return of the target portfolio at time t;
- $w_{i,t}$ is the weight of asset i contained in the portfolio at time t;
- $r_{2,t}$ is the return of asset i contained in the portfolio at time t;
- the intercept is imposed to zero.

By imposing a zero intercept, the linear regression assumes that there is no value in holding cash, which is consistent with the no-arbitrage assumption in finance. This means that the portfolio replication is unbiased and there are no systematic errors in the replication process. Amenc et al. [2] have discussed in their paper the use of linear regression to replicate a target portfolio, stating that the assumption that the return on the portfolio

is zero when the returns on all underlying assets are zero implies that the regression intercept should be equal to zero.
At this point an important clarification has to be made. Considering the returns of a portfolio with this linear combination, is a good approach if the aim of the model is to understand the *structure* of the portfolio; in fact, in this way the weights can be obtained and what the portfolio is made of is clear. Instead if the ultimate goal of a model is just the *prediction* of the returns, there were some models, such as Random Forests or Neural Networks, that could perform well; nevertheless, a limit of these models is that they are black box and do not return as outcome the exact weights to construct a portfolio.
This study is centred on the structure, for this reason the outcome of both models are the weights of the replica portfolio. Obtaining the weights is useful because some strategies may be built imposing constraints on them:

- it can be imposed that they sum to 1: this would mean that all the budget available is invested on the market. Keeping the portfolio fully invested helps to optimize the returns while minimizing the risk of holding cash that is not earning any returns;
- if the sum of weights of a portfolio is greater than 1, it indicates that the portfolio is over-invested or has a leverage. This means that the total investment in the portfolio exceeds the total amount of funds available for investment, which can happen when an investor borrows money to invest in the portfolio;
- if a portfolio has some negative weights, it means that the investor is using a short selling strategy. Short selling is a technique where an investor borrows an asset and sells it in the market, hoping to buy it back at a lower price and return it to the lender, making a profit on the price difference.
- the turnover of the portfolio and transactions costs can be kept under control. The turnover of a portfolio is a measure of how frequently the assets in the portfolio are bought and sold within a specific time period; it is calculated as the total value of the securities bought or sold in the portfolio divided by the total value of the portfolio.

Differently from other studies, the set of factors used for the replica is quite large. Hasanhodzic and Lo [15], for example, did their analysis with five factors, that is, proxies for the equity, bond, currency, credit, and commodity markets. Moreover, without knowing the exact factors that drive target portfolio returns, it seems appropriate to considered a wider set of factors. Anyway, if the set is large, a procedure to pick the right factors is needed. This group of techniques, that are called *regularization* techniques, let the model pick only a restricted group of factors out of many.

Deep connected with the regularization argument is the choice of the regression technique for the models: they are both based on an Elastic Net. For the model that uses a rolling windows approach the regression is *dynamic*, the portfolio is rebalanced at each time step and the outcome are several vectors of weights. In the other model the regression is *static* and the outcome is a single vector of weights that will be kept constant over the test set.

2.1. Elastic Net choice

Regularization techniques in linear regression are methods used to avoid overfitting of the model to the training data. There is overfitting when a model becomes too complex and starts to fit the noise in the data rather than the underlying pattern. This leads to poor performance on new, unseen data. Regularization techniques are particularly useful when the number of features in the data is large relative to the number of observations, or when there is multicollinearity among the features. In this study, mainly when the models are run using the dataset of US indexes, the set of factors may be larger than necessary and applying regularization becomes crucial. Regularization makes possible to avoid a feature selection at the start of the procedure and gives to the model the capability of picking the most appropriate factors setting the others to zero.
Regularization techniques add a penalty term to the objective function, which balances the fit of the model to the data with the complexity of the model. The choice of which regularization method to use depends on the specific problem and the characteristics of the dataset. For linear regression problems several regularization techniques exist, but the most used are L^1 regularization, known as *Lasso regression* [28], and L^2 regularization, known as *Ridge regression* [17].
In the Lasso regression the penalty term added to the objective function is the sum of the absolute values of the regression coefficients. This is its equation:

$$\hat{\beta}^{Lasso} = \min_{\beta} \frac{1}{2n} \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \tag{2.3}$$

where:

- β is the vector of regression coefficients, including the intercept term β_0;
- n is the number of samples in the dataset;
- p is the number of factors in the dataset;
- x_{ij} is the value of the jth factor for the ith sample;

- y_i is the target variable value for the ith sample;
- λ is the regularization parameter.

λ is the hyperparameter of the model, which controls the strength of the penalty, that has to be calibrated. Lasso regression is a *shrinkage estimator*: it encourages the coefficients of some of the factors to become zero, performing feature selection. This technique is useful to reduce the number of irrelevant and redundant features in the model. In particular, in case of high dimensional dataset, Lasso regression can reduce the dimensionality of the problem; it identifies the most important features for the prediction task and eliminate irrelevant or redundant features, leading to simpler models. Models are not only simpler, but also more interpretable. Indeed, thanks to the shrinking procedure, Lasso regression tends to produce sparse solutions with only the more relevant factors.
In the Ridge regression the penalty term added to the objective function is proportional to the sum of the squared values of the coefficients of the model. This is its equation:

$$\hat{\beta}^{Ridge} = \min_{\beta} \frac{1}{2n} \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \alpha \sum_{j=1}^{p} \beta_j^2 \tag{2.4}$$

where α is the hyperparameter of the model, which controls the strength of the penalty, that has to be calibrated. The objective of Ridge regularization is to shrink the magnitude of the coefficients of the model, without imposing any of them to exactly zero. This can help to reduce the variance of the model and improve its ability to generalize to new data, by preventing the coefficients from taking on large values that are specific to the training data. A larger value value of alpha determines more regularization and smaller values of the coefficients.
To sum up, Lasso and Ridge regularization have the same purpose, fighting the issue of overfitting, but they use different strategies and perform well in different situations; they are consistent with the No Free Lunch Theorem of Wolpert and Macready [31] . The theorem is a fundamental result in the field of machine learning and optimization, which states that there is no single algorithm that is universally superior to all other algorithms across all possible datasets and problems. A general algorithm may perform reasonably well on many different problems, but may not be optimal for any particular problem, while a specific algorithm may perform exceptionally well on a particular problem but may not generalize well to other problems. Lasso tends to perform better than Ridge when the dataset has many predictor variables, but only a few of them are important for predicting the response variable. This is because Lasso can produce sparse solutions where some of the coefficients are exactly zero, effectively selecting a subset of the most relevant predictors. In contrast, Ridge regression may produce non-zero coefficients for all

predictors, even if some of them are irrelevant. Instead Ridge regression may be preferred over Lasso when the number of predictors is large relative to the number of observations, the predictors are highly correlated, or the goal is to improve the predictive accuracy of the model rather than selecting a subset of relevant predictors.
In this study the set of factors that can be used to build the replica portfolio is quite large: 14 securities considering only a dataset of US indexes and the 5 Fama and French factors. The exact number of factors, and their importance, is not known a priori; consequently, it is not possible to assess if it is better to use Lasso or Ridge. For this reason, Elastic Net has been chosen.
Elastic Net regularization is a technique for linear regression that combines the penalties of L^1 and L^2 to achieve a balance between the two. It was introduced as a way to overcome some of the limitations of Lasso and Ridge regularization when used individually. The equation of its objective function is:

$$\hat{\beta}^{EN} = \min_{\beta} \frac{1}{2n}||y - X\beta||_2^2 + \alpha \cdot l1_{ratio}||\beta||_1 + \frac{1}{2}\alpha \cdot (1 - l1_{ratio})||w||_2^2 \tag{2.5}$$

where:

- β is the vector of regression coefficients
- n is the number of samples in the dataset;
- X is the matrix of the j factors and the i samples;
- y is the vector of the target variables samples;
- $||.||_2$ represents the L^2 norm;
- $||.||_1$ represents the L^1 norm;
- α is the regularization strength parameter
- $l1_{ratio}$ represents the ratio of L^1 penalty term to the total penalty.

The hyperparameters α and $l1_{ratio}$ are linked in Elastic Net regularization: α controls the strenght of regularization, if it increases the amount increases; $l1_{ratio}$ controls the balance between L^1 and L^2 penalties. When $l1_{ratio}$ is 1, the penalty is pure L^1 and the regression is equivalent to a Lasso, while when it is 0 the penalty is pure L^2 and it is equivalent to a Ridge.
Compared with the Lasso, it can be noticed that that Elastic Net has a better feature selection; in fact, it handles correlated features and select groups of features together, which is not possible with Lasso. In addition, it is more stable because Lasso, in the case

of correlated predictors, can produce different set of coefficients for small variations in the data. Compared with Ridge, Elastic Net handles the multicollinearity better because it can select one feature from a group of correlated features rather than keeping all of them To conclude, Elastic Net has been chosen because it is an hybrid between the two other techniques, combining the strengths of both methods while overcoming their limitations.

2.2. Machine learning approach

The approach used in this study for the construction of the models is the most common in a machine learning framework. The steps of the procedure are:

- Problem definition: the first step is to define the problem that needs to be solved. The problem is the portfolio replication.
- Data collection: historical price time series of factors has been collected.
- Data preparation: once the data have been collected, the next step is to prepare them for analysis. In this case the returns are computed from the prices. The model uses the historical time series of returns, so all the factors are comparable in magnitude and no data pre-processing technique, such as Min-Max scaling, is needed.
- Data splitting: data are splitted in training set and test set. The training set is used to train the model, while the testing set is used to evaluate its performance. Due to the cross validation procedure that is performed, data of the validation set, used to tune the hyperparameters, are contained in the training set.
- Model selection: once the data has been prepared, the next step is to select an appropriate model to solve the problem. The model is an Elastic Net, both for the dynamic and the static approach.
- Model training: the selected model is to trained on the training set. This involves setting the hyperparameters for the algorithm and fitting it to the data to create a predictive model.
- Model tuning: using cross validation techniques the best combination of hyperparameters is chosen
- Model evaluation: the performance of the model on the test set is evaluated.

This pipeline has been followed for both models and at the end their results are compared.

3 | Dataset

In order to replicate a Black Box portfolio a collection of financial products, each one covering a different risk factor, is needed. In this chapter dataset is presented and Fama and French models explained.
The collection of this study, provided by Fondaco SGR, is made up of two group of instruments: a set of 14 US Indexes and the 5 Fama and French factors. The Black Box portfolios used to test the model are built with a linear combination of different portfolios available in the data library on the website of Kenneth R. French. [11]
The dataset includes prices and returns on monthly basis, covering the time period from January 2001 to September 2022, with in total 261 observations. In the following sections the variables are presented and the Fama and French models explained.

3.1. US Indexes

Every product of this set is and index of the american stock market; an index is an instrument that tracks the performance of a certain group of stocks, bonds or other investments. Each one, which is built with a formula, represents only a segment of the financial market and it is used by investors to obtain quickly insights about that particular segment. For this reason having indexes which cover every area of the financial market would be really useful for the purpose of this study.
Indexes used are: RU20INTR, XNDX, LF98TRUU, I00189US, LUATTRUU, M1US000V, M1US00G, SPTR, LUACTRUU, XMI, HUI, GSCI, OEX, VIX.

- **RU20INTR**(*"Russell 2000 Small-Cap TR Index"*): ETF [1] replicating the performance of the Russel 2000 Index. It is a small-cap stock market index composed of the 2000 smallest stocks in the Russell 3000 index, which instead tracks the performance of the 3000 largest U.S.-traded stocks.
  RU20INTR is focused on small-cap companies in the U.S. market, indeed it is considered a good indicator of the U.S. economy.

[1] For simplicity ETFs and indexes in this study are considered as if they had the same properties.

- **XNDX**(*"Nasdaq 100 TR Index"*): the total return version [2] of Nasdaq 100, which is a market capitalization weighted index that includes 100 of the largest domestic and international non-financial stocks listed on the NASDAQ Stock Market based on market capitalization.
- **M1US000V** (*"MSCI US Value Net TR Index"*): index capturing large and mid cap US stocks exhibiting overall value [3] style characteristics.
- **M1US000G** (*"MSCI US Growth Net TR Index"*): index capturing large and mid cap securities exhibiting overall growth style characteristics in the US.
- **SPTR** (*"S&P 500 TR Index"*): the total return version of S&P 500, which is a market capitalization weighted index of the five hundred, largest, publicly traded companies in the United States. This index is said to represent the best picture of the large-cap equities in the US market.
- **XMI** (*"NYSE ARCA Major Market Index"*): American price-weighted stock market index made up of 20 Blue Chip [4] industrial stocks of major U.S. corporations.
- **HUI** (*"NYSE ARCA Gold Bugs Index"*): index composed of publicly-traded gold-mining companies that is useful for tracking short-term trends in gold prices.
- **GSCI** (*"S&P GSCI Commodity TR Index"*): index of commodities that measures the performance of the commodities market. The index often serves as a benchmark for commodities investments. Investing in a GSCI fund provides a broadly diversified, unleveraged long-only position in commodity futures.
- **OEX** (*"S&P 100 Index"*): index that is composed of the 100 largest Blue Chip stocks in the S&P 500 with listed stock options.
- **LF98TRUU**(*"Bloomberg Barclays US Corp High Yield TR Index"*): index provides investors with exposure to U.S. high yield bonds.
- **I00189US**(*"Bloomberg Barclays US Caa High Yield TR Index"*): index provides investors with exposure to U.S. high yield bonds of corporate with an index rating of at least Caa3.

[2] A total return index is a type of equity index that tracks both the capital gains as well as any cash distributions, such as dividends or interest, attributed to the components of the index. A look at an index's total return displays a more accurate representation of the index's performance to shareholders.

[3] There is a duality between Value and Growth stocks: value stocks are companies investors think are undervalued by the market while growth stocks are companies that investors think will deliver better-than-average returns.

[4] Blue Chip is a term that comes from poker and stands for companies that are known for being valuable, stable and established. They're typically big names in their industries, and investors count on them for their reliability.

- **LUATTRUU** (*"Bloomberg US Treasury Total Return Unhedged USD Index"*): this index track the performance of the U.S. Treasury bond market.
- **LUACTRUU** (*"Bloomberg Barclays US Investment Grade TR Index"*): index provides investors with exposure to U.S. investment grade bonds.
- **VIX** (*"Volatility Index"*): this index is a calculation designed to produce a measure of constant, 30-day expected volatility of the U.S. stock market, derived from real-time, mid-quote prices of S&P 500 Index call and put options. It is one of the most recognized measures of volatility.

3.2. Fama and French factors

Fama and French factors inherited their name from Eugene Fama and Kenneth French, two american economists that have designed a statistical model to describe stock returns. They wrote several papers questioning the validity and the precision of the "Capital Asset Pricing Model" (CAPM) [26]. The CAPM is a theoretical model that explains the relationship between the expected return of an asset and its risk. The model predicts that the expected return of an asset is equal to the risk-free rate plus a risk premium, which is proportional to the asset's beta. Beta is a measure of an asset's volatility compared to the overall market and it is the only source of risk.
However, Fama and French found that the Beta factor alone was not able to explain all the patterns of the stock market: they proposed a Three-Factor Model that was later expanded to a Five-Factor Model. The dataset of this study contains five Fama and French factors.

3.2.1. Factors Models vs CAPM

Fama and French built their factor models as a way to improve upon the CAPM, which was originally proposed as a way to explain the relationship between expected returns and systematic risk as measured by beta. However, they found that the CAPM did not explain the observed cross-sectional patterns in average stock returns. The reason why CAPM does not work is that it assumes that all stocks are perfect substitutes and that the only source of risk is market beta. Anyway, this assumption has been challenged by empirical evidence showing that factors introduced by Fama and French can explain the cross-sectional variation in average returns. As a result, CAPM has been criticized for over-simplifying the real-world dynamics of stock returns and for not being able to explain the empirical patterns in average returns.

The Fama-French Three-Factor Model wanted to explain the variation in stock returns using three factors: the market factor, the size factor, and the value factor. The model was accepted and used but some anomalies were found in the data that could not be fully explained by the three factors.
To overcome these problems, Fama and French decided to extend their model from three factors to five factors. In this paper [9] they introduced two additional factors: the profitability factor (RMW) and the investment factor (CMA). The profitability factor represents the performance of firms with high operating profitability relative to firms with low operating profitability. The investment factor represents the performance of firms with low investment relative to firms with high investment. They found that these two additional factors improved the model's ability to explain the cross-sectional variation in stock returns, particularly for small and value stocks. The Five-Factor Model has become widely accepted and used as a benchmark for explaining stock returns. The inclusion of the profitability and investment factors in the Five-Factor model has been seen as evidence that the factors play a significant role in determining stock returns and that a multi-factor approach is necessary to fully understand stock market performance.

3.2.2. Three-Factor Model

The three factors of the Fama and French Three-Factor Model are $R_{M,t}$ - $R_{f,t}$, SMB_t and HML_t. The equation of the model is the following:

$$R_{i,t} - R_{f,t} = \alpha + \beta_1(R_{M,t} - R_{f,t}) + \beta_2 SMB_t + \beta_3 HML_t + \epsilon_{i,t}, \tag{3.1}$$

where:

- $R_{i,t}$ is the excess return on asset i at time t;
- $R_{f,t}$ is the risk-free rate at time t;
- $R_{M,t}$ is the return on the market portfolio at time t;
- SMB_t is the size factor;
- HML_t is the value factor;
- α, β_1,β_2,β_3 and $\epsilon_{i,t}$ are the usual components of a linear regression.

The factor $R_{M,t}$ - $R_{f,t}$ is the market beta used in the CAPM and represents the systematic risk of the market. The SMB ("*Small Minus Big*") factor represents the difference in average returns between small-cap stocks and large-cap stocks; it is used to capture the risk premium that is associated with investing in smaller stocks relative to larger ones,

indeed there is empirical evidence of the fact that small-cap companies outperform large-cap ones considering long periods of time. The HML ("*High Minus Low*") represents the difference in average returns between high book-to-market [5] stocks and low book-to-market stocks; empirical evidence suggests that high book-to-market stocks outperforms low book-to-market stocks considering long periods of time.
To compute the everyday value of SMB and HML, firstly Fama and French sort every stock of the considered region [6] in different portfolios, dividing them using a particular feature as a criteria. They build 6 portfolios splitting stocks in two market cap and three book-to-market groups. The portfolios are:

- Small Value stocks
- Small Neutral stocks
- Small Growth stocks
- Big Value stocks
- Big Neutral stocks
- Big Growth stocks

SMB is the equal-weight average of the returns on the three small stock portfolios for the region minus the average of the returns on the three big stock portfolios:

$$\begin{aligned} SMB = &\frac{1}{3}(SmallValue + SmallNeutral + SmallGrowth) \\ &-\frac{1}{3}(BigValue + BigNeutral + BigGrowth). \end{aligned} \tag{3.2}$$

HML is the equal-weight average of the returns for the two high B/M portfolios for a region minus the average of the returns for the two low B/M portfolios:

$$HML = \frac{1}{2}(SmallValue + BigValue) - \frac{1}{2}(SmallGrowth + BigGrowth). \tag{3.3}$$

[5] The book-to-market (B/M) ratio is a financial indicator that compares the market value of a company to its book value. This ratio is obtained by dividing the book value (calculated as total assets minus liabilities) by the market capitalization.

[6] All CRSP (*"Center for Research in Security Prices"*) firms incorporated in the US and listed on the NYSE, AMEX, or NASDAQ that have a CRSP share code of 10 or 11 at the beginning of month t

3.2.3. Five-Factor Model

The five factors of the Fama and French Five-Factor Model are $R_{M,t}$ - $R_{f,t}$, SMB_t , HML_t, RMW_t and CMA_t . The equation of the model is the following:

$$\begin{aligned} R_{i,t} - R_{f,t} = \alpha + \beta_{MKT}(R_{M,t} - R_{f,t}) + \beta_{SMB}SMB_t \\ +\beta_{HML}HML_t + \beta_{RMW}RMW_t + \beta_{CMA}CMA_t + \epsilon_{i,t}, \end{aligned} \tag{3.4}$$

where:

- $R_{i,t}$ is the excess return on asset i at time t
- $R_{f,t}$ is the risk-free rate at time t
- $R_{M,t}$ is the return on the market portfolio at time t
- SMB_t is the size factor
- HML_t is the value factor
- RMW_t is the robust minus weak factor
- CMA_t is the conservative minus aggressive factor
- α, β_{MKT},β_{SMB},β_{HML}, β_{RMW}, β_{CMA} and $\epsilon_{i,t}$ are the usual components of a linear regression

It can be noticed that the market beta $R_{i,t}-R_{f,t}$, the size factor SMB_t and the value factor HML_t are in common with the Three-Factor model. The RMW (*'Robust Minus Weak'*) factor represent the difference in average returns between robust operating profitability[7] portfolios and weak operating profitability portfolios; Fama and French found that firms with high operating profitability tend to have higher returns than firms with low operating profitability, and that this effect was not captured by the market, size, and value factors. The CMA (*'Conservative Minus Aggressive'*) is the difference in average returns between conservative investment [8] portfolios and aggressive investment portfolios; it is use to account for the fact that firms with low investment tend to have higher returns than firms with high investment.

To compute the everyday value of RMW, firstly Fama and French sort every stock of the considered region in different portfolios, dividing them using a particular feature as

[7]The operating profitability of a firm is often measured by its operating income, which is calculated as revenue minus operating expenses.

[8]Investment refers to the amount of capital that a firm invests in its operations and growth, for example, in research and development, new projects, or acquiring other companies.

a criteria. They build four portfolios splitting stocks in two market cap and operating profitability groups. The portfolios are:

- Small Robust stocks;
- Small Weak stocks;
- Big Robust stocks;
- Big Weak stocks.

RMW is the equal-weight average of the returns on the two robust stock portfolios for the region minus the average of the returns on the two weak stock portfolios:

$$RMW = \frac{1}{2}(SmallRobust + BigRobust) - \frac{1}{2}(SmallWeak + BigWeak). \tag{3.5}$$

To compute the everyday value of CMA, they build two portfolios splitting stocks in two market cap and investment groups. The portfolios are:

- Small Conservative stocks;
- Small Aggressive stocks;
- Big Conservative stocks;
- Big Aggressive stocks;

CMA is the equal-weight average of the returns for the two conservative investment portfolios for a region minus the average of the returns for the two aggressive investment portfolios:

$$\begin{aligned} CMA = &\frac{1}{2}(SmallConservative + BigConservative) \\ &-\frac{1}{2}(SmallAggressive + BigAggressive). \end{aligned} \tag{3.6}$$

The Five-Factor Model present some advantages and novelties. One of them is its improved explanatory power compared to the CAPM. Additionally, the model takes into account behavioral biases such as the size and value premiums, which have been consistently observed in the stock market.
The model introduces several novelties to traditional finance models. Firstly, it introduces three new factors, which provide a more complete explanation of stock returns. Secondly, the model shifts the focus from individual stocks to common risk factors, providing a more systematic approach to portfolio construction. Finally, the model considers microeconomic variables such as profitability and investment, which are not considered in the

CAPM.

3.2.4. The role of the intercept a

The intercept in the Fama-French Factors Models is a measure of the average excess return of the stocks in the sample. It is calculated as the average return of the stocks in the sample minus the expected return predicted by the model. In financial terms, the intercept represents the average abnormal return of the stocks in the sample. Abnormal returns are the returns that are not explained by the systematic risk factors captured by the five factors. The intercept can be interpreted as the compensation investors require for taking on any residual risks that are not captured by these factors. A positive intercept indicates that the average excess return of the stocks in the sample is greater than the expected return predicted by the model, while a negative intercept indicates the opposite. If the intercept is not significantly different from zero, it suggests that the Fama-French model provides a good fit to the data.

3.2.5. The Risk Free Rate

The risk free rate R_f that is used in the models is a 1 month american Treasury Bill.
A Treasury Bill is a short-term debt security issued by the United States government, with a maturity of one year or less, which is considered to be one of the safest investments available, due to the fact that they are backed by the US government.

3.2.6. Fama and French factors available in the dataset

To exploit this study all the data library on the website of French [11] is available. In order to obtain a better replication and to keep track more sectors as possible, the *Five-Factors Model* has been chosen over the Three-Factors one. For each factor a time series of monthly returns has been downloaded; the period covered is from January 31$^{\text{th}}$, 2001 to September 1$^{\text{st}}$, 2022.

3.3. Black Box portfolios

The term *Black Box* is used to describe a portfolio whose composition in unknown. The aim of the model is to track the movements of these portfolios building a replica portfolio composed choosing instruments among the Indexes or the Fama and French factors.
As for the factors, also for the Black Box portfolios the time series of returns comes from the website of *French* [11]; the number of portfolios downloaded is 8. They are not only

used as they comes, indeed new portfolios can be built with a linear combination of them. Considering the fact that the composition of these portfolios follows the logic of *Fama and French Models*, this choice could be useful to avoid strong correlation between one of them and a particular risk factor.
The 8 portfolios are:

- **SCG**: *Small Cap Growth* contains all the stocks of the indexes NYSE, AMEX and NASDAQ which are small cap and growth [9];
- **BCM**: *Big Cap Medium BtM* contains all the stocks of the indexes NYSE, AMEX and NASDAQ which are big cap and with a medium book-to-market;
- **LCV**: *Large Cap Value* contains all the stocks of the indexes NYSE, AMEX and NASDAQ which are large cap and value;
- **Cnsmr**: contains all the stocks of the indexes NYSE, AMEX and NASDAQ which belong to the industry of consumer durables, non durables, wholesale, retail and some services (laundries, repair shops);
- **Manuf**: contains all the stocks of the indexes NYSE, AMEX and NASDAQ which belong to the industry of manufactoring, energy and utilities;
- **Hitec**: contains all the stocks of the indexes NYSE, AMEX and NASDAQ which belong to the industry of business equipment, telephone and television transmission;
- **Hlth**: contains all the stocks of the indexes NYSE, AMEX and NASDAQ which belong to the industry of healthcare, medical equipment and drugs;
- **Other**: contains all the stocks of the indexes NYSE, AMEX and NASDAQ which belong to other industries (mines, constructor, hotels, bus services, entertainment, Finance)

From now on, they are identified with the ticker written in bold. The target portfolios, some of them built with a linear combination, that are used to test the model, will be specified in the following pages.

[9] The size breakpoint is the median NYSE market equity. The book-to-market breakpoints are the 70^{th} and the 30^{th} NYSE percentiles. Stocks with book-to-market higher than the 70^{th} percentile are defined as Value, lower than the 30^{th} as Growth.

4 | Dataset Analysis

In this chapter the main features of the historical time series that compose the dataset are analyzed. Due to the fact that data are on monthly basis there may exist a choice between returns and log returns; for this reason, the first operation is to execute stationarity to decide how to continue the study, if with returns or log returns. Then, some properties of the time series are shown in tables.

4.1. Stationarity Test

When dealing with time series it's crucial to check if they are stationary or not. Stationarity is a very important concept in time serie analysis, because it guarantees that the statistical properties of the series remain constant over time. A time series is said to be *stationary* if it has constant mean, constant variance and constant covariance over time. Stationarity is an important prerequisite for many time series analysis and there are some techniques used to obtain it, such as the first differences, log transformations or detrending.

Data available in this study are the time series of monthly returns. When the returns are daily, it's a common practice in finance to use a log transformation to reach the stationarity; in this case they could already be stationary before applying any transformation.

To clarify any doubt, two different stationarity test are applied both to the monthly returns and monthly log returns. The two tests are the *Augmented Dickey-Fuller Test* (ADF) [6] and the *Phillips-Perron Test* (PP) [24].

4.1.1. Augmented Dickey-Fuller Test

The ADF test is a statistical test used to check the stationarity of a time series by regressing the it against its lagged values and testing the hypothesis that the residuals are not stationary. The p-value from the test can be used to determine if the time series is stationary or not. The null hypothesis of the test is that the time series has a *unit root*, which is a property that indicates not stationarity. A time series is said to have a unit root if it can be expressed as a first-order autoregressive model:

$$y_t = \phi y_{t-1} + \varepsilon_t, \tag{4.1}$$

where y_t is the value of the time series at time t, ϕ is the autoregressive coefficient, and ε_t is the error term. A time series with a unit root is non-stationary because the mean, variance, and autocovariance of the series are not constant over time.
These are the steps of the test more detailed:

- set the null hypothesis: the time series has a unit root and is not stationary
- build the regression equation

$$y_t = c + \phi_1 y_{t-1} + \phi_2 \Delta y_{t-1} + ... + \phi_p \Delta y_{t-p} + \varepsilon_t, \tag{4.2}$$

  where y_t is the value of the time series at time t, c is a constant, ϕ_1, ϕ_2, ..., ϕ_p are the regression coefficients, Δy_t is the first difference of the time series, and ε_t is the error term. The number of lags p is chosen based on the order of integration of the time series;
- the ADF test statistic is calculated as the t-statistic for the coefficient of the first lag of the time series, ϕ_1. If this term is significantly different from zero, then the null hypothesis is rejected, and the time series is considered to be stationary;
- the p-value for the ADF test is calculated from the test statistic and can be used to determine the level of significance for the test. If the p-value is less than 0.05, then the null hypothesis is rejected and the time series is considered to be stationary.

4.1.2. Phillips-Perron Test

To have a stronger and more robust conclusion, after the ADF test, also the Phillips-Perron test is performed on the dataset.
The Phillips-Perron (PP) is used to test for unit roots in time series data. It overcomes some of the limitations of the ADF test by using a more efficient estimation procedure. The idea behind the PP test is to estimate a time series model that includes both a linear trend and a unit root. The null hypothesis for the test is that the time series has a unit root, and the alternative hypothesis is that the time series is stationary. The estimation procedure is called Inverse Root (IR) method. After the estimation of the parameters with this procedure, a test statistic is calculated based on the estimates. If

the test statistic is greater than a critical value, then the null hypothesis of a unit root is rejected, and the time series is considered to be stationary. If the test statistic is less than the critical value, then the null hypothesis cannot be rejected, and the time series is considered to have a unit root and to be non-stationary. The critical value comes from the critical value table, which gives the critical value for a given level of alpha and the number of observations and which is based on the asymptotic properties of the test statistic and takes into account the distribution of the test statistic under the null hypothesis.

4.1.3. Stationarity Results

As it can be seen in Table 4.1 and in Table 4.2, for each factor, both for returns and log returns, the results of the two test is unanimous: p-values are really low and the null hypothesis are rejected. It can be concluded that the time series of monthly returns and the time series of monthly log returns are stationary and they can be used both inside the model.

In this study, as standard and literature practice, the choice is to use the logarithmic returns. From now on, for simplicity, in the thesis the log returns are called as returns.

	ADF	PP
RU20INTR	2.82e-28	2.57e-28
XNDX	2.36e-28	2.30e-28
LF98TRUU	2.08e-11	5.67e-24
I00189US	2.03e-20	7.11e-20
LUATTRUU	3.65e-13	3.87e-27
M1US000V	9.32e-28	7.65e-28
M1US000G	3.22e-28	2.69e-28
SPTR	3.53e-28	2.98e-28
LUACTRUU	8.73e-27	8.67e-27
XMI	2.76e-29	2.39e-29
HUI	7.56e-30	7.45e-30
GSCI	8.16e-24	1.75e-23
OEX	2.58e-28	2.36e-28
VIX	1.12e-10	1.36e-10
Mkt-RF	6.84e-28	5.38e-28
SMB	2.54e-18	1.04e-29
HML	1.65e-09	3.47e-26
RMW	6.17e-26	3.83e-26
CMA	3.48e-26	1.47e-26
SCG	1.58e-27	1.79e-27
BCM	1.56e-27	1.58e-27
LCV	3.14e-27	3.78e-27
Cnsmr	1.15e-06	8.36e-29
Manuf	7.53e-29	7.52e-29
Hitec	1.15e-28	9.84e-29
Hlth	1.06e-29	1.10e-29
Other	6.53e-09	3.20e-27

Table 4.1: Results of Augmented Dickey-Fuller Test and Phillips-Perron Test performed on returns of the factors. P-values are low and the null hypothesis are rejected

	ADF	PP
RU20INTR	4.22e-28	4.07e-28
XNDX	5.22e-28	3.93e-28
LF98TRUU	1.63e-11	1.25e-23
I00189US	2.81e-20	1.24e-19
LUATTRUU	3.51e-13	3.86e-27
M1US000V	1.52e-27	1.13e-27
M1US000G	6.23e-28	4.87e-28
SPTR	6.02e-28	4.68e-28
LUACTRUU	9.40e-27	9.66e-27
XMI	3.78e-29	3.40e-29
HUI	7.63e-30	7.25e-30
GSCI	2.39e-23	6.12e-23
OEX	3.96e-28	3.42e-28
VIX	9.70e-16	2.34e-15
Mkt-RF	1.17e-27	8.55e-28
SMB	2.06e-18	9.22e-30
HML	1.43e-09	3.34e-26
RMW	5.67e-26	4.02e-26
CMA	2.70e-26	1.33e-26
SCG	2.29e-27	2.66e-27
BCM	2.79e-27	2.74e-27
LCV	6.00e-27	7.25e-27
Cnsmr	1.18e-06	1.21e-28
Manuf	1.32e-28	1.33e-28
Hitec	1.99e-28	1.33e-28
Hlth	1.26e-29	1.30e-29
Other	8.09e-09	4.89e-27

Table 4.2: Results of Augmented Dickey-Fuller Test and Phillips-Perron Test performed on logarithmic returns of the factors. P-values are low and the null hypothesis are rejected

4.2. Returns analysis

As decided in the previous section, the historical time series of monthly log returns is considered. Some statistics of these factors, covering the period January 2001 - September 2022, are reported in Table 4.3.

	Mean	Std Dev	Max	Min	Skewness	Kurtosis
RU20INTR	0.006253	0.058072	0.169172	-0.244961	-0.811658	2.075387
XNDX	0.007018	0.064761	0.172869	-0.306548	-0.904343	2.693173
LF98TRUU	0.005543	0.027632	0.114263	-0.173244	-1.517915	9.714089
I00189US	0.005413	0.041535	0.176501	-0.271666	-1.568950	10.667923
LUATTRUU	0.002796	0.013041	0.05171	-0.044917	-0.079753	1.164333
M1US000V	0.004304	0.044586	0.121679	-0.175236	-0.863386	2.117821
M1US000G	0.006161	0.048584	0.142352	-0.201059	-0.706960	1.647600
SPTR	0.005834	0.044276	0.120618	-0.18386	-0.743910	1.410923
LUACTRUU	0.003925	0.017301	0.065755	-0.080876	-1.001980	4.856212
XMI	0.003725	0.040932	0.12884	-0.177204	-0.721545	1.974915
HUI	0.005124	0.112642	0.362784	-0.482969	-0.098294	1.196179
GSCI	-0.000435	0.069446	0.179527	-0.348499	-1.062962	3.480165
OEX	0.003288	0.044337	0.119358	-0.157717	-0.598588	0.805051
VIX	0.001179	0.218062	0.852588	-0.614279	0.459012	1.074104

Table 4.3: Statistics of the indexes log returns

A first comment that can be done is that the means, except for the index accounting for commodities, are positive. These reflects the performance of the financial markets in the period taken into account; also considering period of decline such as the global financial crysis of 2008-2009 and the Covid pandemic, the performance has been overall positive. One thing that stands out is the line of the VIX; due to its nature, it's not surprising that it has very different statistic compared with the other securities. VIX is a measure of the volatility of the market and does not represent the ownership of a stock or of an obligation, but it is left in the set of factors because it can be traded. In fact, many investors trade VIX as a way to hedge against volatility using future contracts or options to buy and sell it. Coherently, it has a volatility that is way higher than others factors and also the max value and min value are more extreme. In Figure 4.1 it can be seen the difference between monthly log returns time series of the VIX and the one of the SPTR, the index tracking S&P 500.

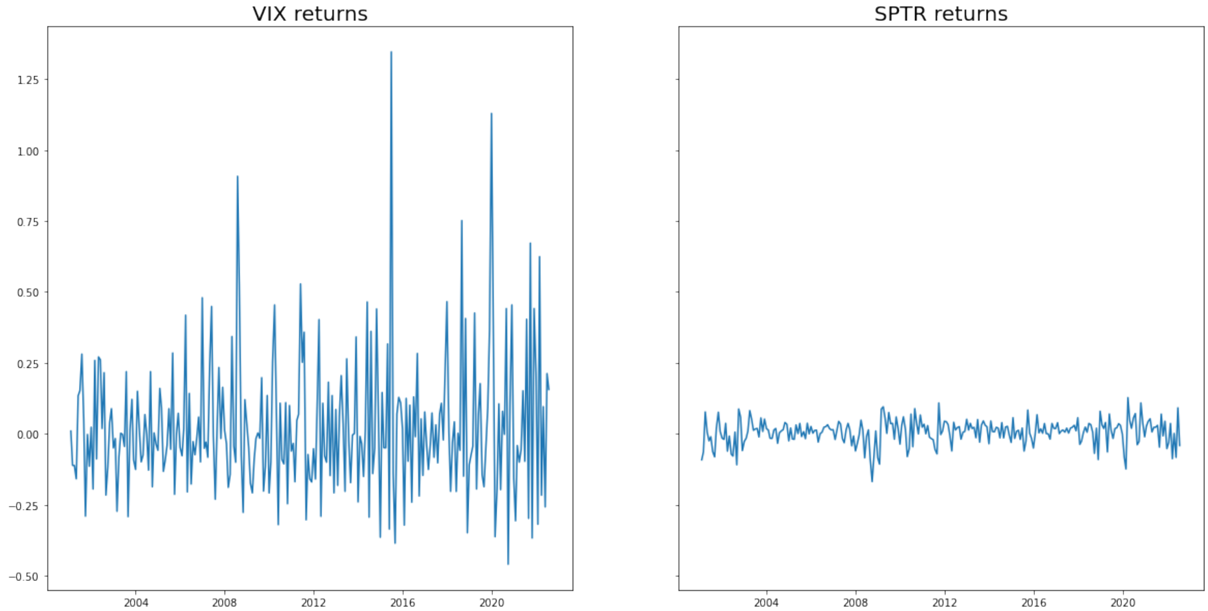


Figure 4.1: Monthly log returns of VIX and S&P 500 compared

Some consideration can be done comparing different kinds of securities: indexes tracking bonds (LF98TRUU, I00189US, LUATTRUU and LUACTRUU) are in general less volatile than equity indexes, with the ones exposed to high yield that are closer to equity than the others. As expected, the index tracking the US Treasury bond market is the less volatile and with low kurtosis. The Figure 4.2, displaying the frequency returns distribution of the index tracking U.S. high yield bonds and U.S. high yield bonds of corporate with an index rating of at least Caa3, explains their high values of kurtosis: the data are concentrated around the mean, but they have tails more dense of extreme values compared to a normal distribution.

Between equity, the most volatile and the one with more extreme values is the index composed of gold-mine companies, and this fact is not surprising because this security is exposed to a single factor that is gold's price.

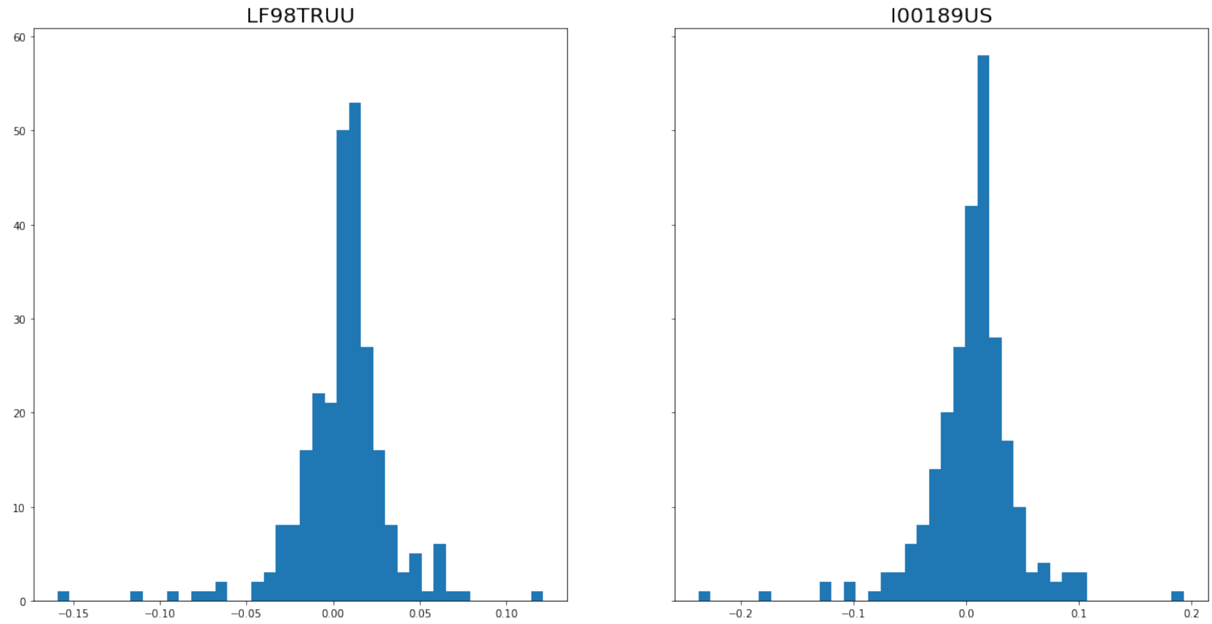


Figure 4.2: Frequency distributions of log returns of LF98TRUU and I00189US

	Mean	Std Dev	Max	Min	Skewness	Kurtosis
Mkt-RF	0.005042	0.045777	0.127953	-0.189105	-0.710379	1.408772
SMB	0.002247	0.026745	0.068873	-0.086757	0.049651	0.091710
HML	0.000239	0.031690	0.120003	-0.150474	-0.100947	3.504548
RMW	0.003526	0.023013	0.087186	-0.096621	0.104562	2.459673
CMA	0.001782	0.020161	0.086636	-0.071926	0.570115	2.126407

Table 4.4: Statistics of the Fama and French factors log returns

	Mean	Std Dev	Max	Min	Skewness	Kurtosis
SCG	0.006056	0.065927	0.20785	-0.249906	-0.539191	1.108035
BCM	0.005886	0.044750	0.134119	-0.192591	-0.957822	2.784286
LCV	0.005232	0.060876	0.167316	-0.317993	-1.251950	4.431393
Cnsmr	0.007787	0.041346	0.14583	-0.159347	-0.377811	1.479526
Manuf	0.006758	0.047186	0.126192	-0.212946	-1.040354	3.003278
Hitec	0.006216	0.059795	0.181404	-0.255925	-0.712635	1.808904
Hlth	0.005719	0.040647	0.125839	-0.115972	-0.470670	0.355578
Other	0.004768	0.053838	0.153493	-0.226148	-0.936130	2.843574

Table 4.5: Statistics of the target portfolio components log returns

Table 4.4 and Table 4.5 contain statistics of Fama and French factors and portfolio derived from them, so it is appropriate to analyze them together. Fama and French factors have standard deviations that are lower than the one of the components of the target portfolios, and also of the equity; for sure they are more similar to bond indexes. Also here all the means are positive in value, with the mean of Fama and French factors that is similar to the indexes one. It can be notice that, although they are built following Fama and French models, the returns of Table 4.5 behave similarly to the returns of the equity US indexes.

4.3. Multicollinearity

In financial problems, when dealing with multiple stocks, is always important to anaylize their correlation. A correlation coefficient measures the strength and direction of the linear relationship between two variables, and ranges from -1 to +1, where -1 indicates a perfect negative correlation, 0 indicates no correlation, and +1 indicates a perfect positive correlation. In the problem of these study study the correlation is even more important because of the model that is used: linear regression results are affected by the multicollinerity between factors. Multicollinearity is a phenomenon that occurs when two or more independent variables in a regression model are highly correlated with each other. This means that the independent variables are not independent of each other, which can cause issues in the interpretation of the regression coefficients and the predictive accuracy of the model. When multicollinearity is present, it becomes difficult to determine the individual effect of each independent variable on the dependent variable, as the effects of the variables may be overlapping. Additionally, the regression coefficients may be unstable, and small changes in the data or the model can lead to large changes in the coefficients. One common way to detect multicollinearity is to calculate the correlation matrix between the independent variables. If there are high correlations (higher than a threshold of 0.7 or 0.8 more or less) between two or more independent variables, multicollinearity may be present. To address multicollinearity, several techniques can be used. One approach is to remove one or more of the highly correlated independent variables from the model. In this study the approach is to apply regularization techniques directly on the model, which can help to shrink the coefficients of the correlated variables and improve the stability.

	RU20INTR	XNDX	I00189US	M1US000V	M1US000G	LUACTRUU	XMI	HUI	GSCI	OEX	SPTR	LUATTRUU	LF98TRUU
RU20INTR	1.000000	0.777217	0.663476	0.868828	0.837292	0.270898	0.088225	-0.055174	0.188828	-0.003694	0.884472	-0.321064	0.687114
XNDX	0.777217	1.000000	0.551451	0.771595	0.922608	0.233764	0.119995	-0.002128	0.110508	0.054809	0.880087	-0.271316	0.590195
I00189US	0.663476	0.551451	1.000000	0.661361	0.615849	0.443599	0.105836	-0.015024	0.270297	0.035683	0.658491	-0.252001	0.925783
M1US000V	0.868828	0.771595	0.661361	1.000000	0.850593	0.279631	0.120103	-0.036053	0.196640	0.011900	0.962666	-0.310751	0.671146
M1US000G	0.837292	0.922608	0.615849	0.850593	1.000000	0.335632	0.115638	-0.020254	0.160707	0.035469	0.959985	-0.233407	0.664797
LUACTRUU	0.270898	0.233764	0.443599	0.279631	0.335632	1.000000	0.092039	0.083699	0.146412	0.054632	0.315794	0.570104	0.606196
XMI	0.088225	0.119995	0.105836	0.120103	0.115638	0.092039	1.000000	0.173716	0.361964	0.815803	0.123437	0.034372	0.093767
HUI	-0.055174	-0.002128	-0.015024	-0.036053	-0.020254	0.083699	0.173716	1.000000	0.275115	0.130430	-0.030074	0.094808	-0.011634
GSCI	0.188828	0.110508	0.270297	0.196640	0.160707	0.146412	0.361964	0.275115	1.000000	0.325633	0.181279	-0.137723	0.276670
OEX	-0.003694	0.054809	0.035683	0.011900	0.035469	0.054632	0.815803	0.130430	0.325633	1.000000	0.024725	0.059013	-0.010176
SPTR	0.884472	0.880087	0.658491	0.962666	0.959985	0.315794	0.123437	-0.030074	0.181279	0.024725	1.000000	-0.281277	0.689882
LUATTRUU	-0.321064	-0.271316	-0.252001	-0.310751	-0.233407	0.570104	0.034372	0.094808	-0.137723	0.059013	-0.281277	1.000000	-0.128490
LF98TRUU	0.687114	0.590195	0.925783	0.671146	0.664797	0.606196	0.093767	-0.011634	0.276670	-0.010176	0.689882	-0.128490	1.000000

Figure 4.3: Correlation matrix of US Indexes returns from January 2001 to September 2022

	Mkt-RF	SMB	HML	RMW	CMA
Mkt-RF	1.000000	0.314528	0.044661	-0.389822	-0.201424
SMB	0.314528	1.000000	0.242843	-0.305042	0.059760
HML	0.044661	0.242843	1.000000	0.155225	0.582984
RMW	-0.389822	-0.305042	0.155225	1.000000	0.146012
CMA	-0.201424	0.059760	0.582984	0.146012	1.000000

Figure 4.4: Correlation matrix of returns of Fama and French factors from January 2001 to September 2022

Observing Figure 4.4 it can be noticed that there are no high correlation between variables and multicollinearity can be excluded. Instead, observing Figure 4.3, it can be noticed that there are group of variables with correlation between them above the threshold. This group is composed by the indexes exposed to equity and corporate bonds. The highest correlation exists between SPTR and M1US000V, with a correlation coefficient ρ of 0.96266.

5 | Models

In this chapter we present two models. One based on a rolling windows approach is defined as *Model 1*, the static one as *Model 2*. We present the explanations of concepts useful to understand how models work; concepts such as rolling windows, hyperparameters, loss function and cross validation procedures. At the end the metrics used to evaluate the performance of the models are presented.

5.1. Model 1

5.1.1. Short overview

As mentioned before, this model is based on a rolling windows approach. The first feature which has to be highlighted is that the model is dynamic: indeed, thanks to the rolling windows, portfolio weights evolve during time and are not freezed. At the core there is a linear regression, in particularly an Elastic Net that is able to perform regularization. The Elastic Net, trained using a fixed size set of past returns of the factors and of the target portofolio, finds the weights necessary to build the replica portfolio.
The paremeters of this model are three:

- WS: the size of the rolling window;
- α : the regularization strenght parameter;
- $l1_{ratio}$: the ratio of the L1 penalty term to the total penalty.

The pipeline of the model can be summed up in these steps:

- the dataset is splitted in train set and test set;
- the customized loss function that the linear regression tries to minimize is constructed. This loss is used by the cross validation procedure to select the parameters α and $l1_{ratio}$ of the Elastic Net;
- for each values of WS, an Elastic Net is trained at every month and then shifting of one step the window the procedure is repeated. At the end, the best α and $l1_{ratio}$

for all the train set (for the specific WS) are fixed by the cross validation minimizing the custom loss;

- the WS granting the best values for the metric is chosen with an hyperparameter tuning procedure;
- Chosen WS, also the linked α and $l1_{ratio}$ are recovered;
- the performance of the model is tested out-of-sample on the test set.

The model outcome from this procedure is an Elastic Net with parameters α and $l1_{ratio}$ which, at each time step t , is trained using the returns of the factors and the target of the t - WS days before; this model is able to give the best weights to build a portfolio to replicate the returns of the following month.

5.1.2. Rolling windows

A rolling windows approach is a useful and common technique in time series analysis, used in several works related to finance such as at the papers of DeMiguel et al. [5] and Frazzini et al. [10], consisting in dividing the dataset in a sequence of overlapping windows of a fixed length. A rolling window is a fixed size set of consecutive data points, part of a larger dataset; this window is used to perform a particular task, such as calculate a statistics or fitting a model like in this case. At the next time instant of the series, the rolling window remains constant in size but shift of one position and the task is repeated with this different set of data: the process is repeated until the entire dataset has been analyzed.
For example in this model, imitating what is shown in Figure 5.1, if the window size were 50 months, the situation would be the following: the set of returns of the factors and the target portfolio from the 1^{st} month to the 50^{th} would be used to train the model and obtain the weights to build the portfolio to replicate its return in the 51^{th} month. Then the window shifts, and in the 51^{th} month the model is trained with the returns from the 2^{nd} to the 51^{th} to build the best replica for the 52^{th}. This procedure until the end of the dataset. In this example the replica portfolio is rebalanced every month, but the frequency of rebalancing can be changed with an adjustment on the shift imposed to the rolling window.
Rolling windows approach have several strenghts but also some limitations. Among the strenghts, the most useful is that windows allow to analyze the data in more detailed way than looking at the overall time series: this permit to find patters or trends that would have not be found observing at the entire time series. Among the limitations, using overlapping windows may introduce correlation to data within the same window leading

to biased results. Another problem could be overfitting, with the model that becomes to specific for the data of the window and does not generalize well.
One way to tackle these limitations, besides the regularization that is already performed by the Elastic Net, is to use larger windows size, reducing the noise and the fluctuations in the data.

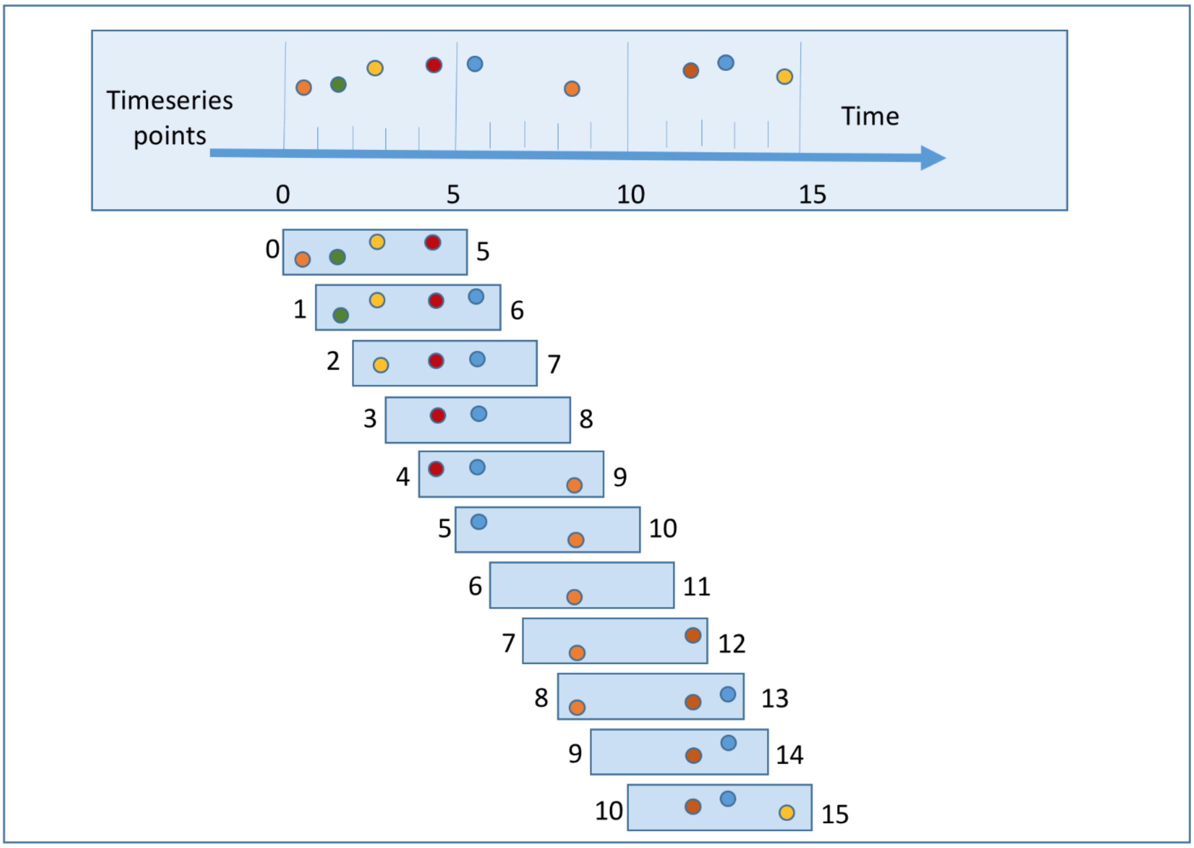


Figure 5.1: Example of a rolling window of size 5 shifting of one time step

Following these considerations, the windows sizes tested are the values in Table 5.1, where the smallest size is one year and the largest five years. The best window size is the value of WS that minimizes the tracking error volatility (TEV), a metric later explained in section 5.3.

WS		
12	27	30
33	36	39
42	44	48
52	54	60

Table 5.1: Tested values of window size for Model 1

5.1.3. Hyperparameters

In a machine learning algorithm hyperparameters are parameters of the model that are not learned from the data. They can be considered as the configuration settings the model, indeed they have to be set before training and remain constant. The procedure of setting them is called hyperparameters tuning. This procedure consist in searching the best combination of hyperparameters values that maximizes the performance of the model on a validation set. In *Model 1*, due to how cross validation is performed, the dataset is splitted only in two part and the validation set is contained in the test set. The hyperparameters are three: α , $l1_{ratio}$ and WS.
For the first two the tuning is automatized using a random search, as explained by Bergstra and Bengio [4], which is a technique that samples the hyperparameters values randomly from a defined grid or distribution. A random combination of hyperparameters is sampled from the search space, and the model is trained and evaluated using these values. The performance is then recorded, and the process continues for a fixed number of iterations or until a stopping criterion is met. This procedure can often identify good hyperparameter configurations with few evaluations. For the window size the tuning is not automatized and there is not any randomness: for each value of WS the model is trained across the training set with the window shifting of one step each iteration and at the end the WS granting the best performance is chosen. The difference in the techniques is justified by the different sizes of the grid from which the hyperparameters are searched: the best WS is chosen among twelve values, while α and $l1_{ratio}$ among respectively 10^8 and 10^7 values.

	Lower bound	Upper bound	Number of points
α	1e-6	1e-4	1e8
$l1_{ratio}$	0.1	1	1e7

Table 5.2: Grid of values for the two hyperparameters of the Elastic Net α and $l1_{ratio}$

In Table 5.2 the grid for the sampling.

5.1.4. Cross validation

Cross validation is a statistical method used to evaluate the performance of machine learning models. The training set is splitted in multiple subsets, called folds. In each iteration all the folds are used to train the model, except one that is used to test it. In the following iteration the procedure is the same, but the fold used to test changes. With this technique each data of the training is used in most of the cases to train the model, and in one case to test it. The most common type of cross-validation is k-fold cross-validation, shown by Kohavi [22], where the dataset is randomly divided into k subsets of equal size. The model is trained on $k-1$ folds and tested on the remaining fold. This process is repeated k times, with each fold being used once for testing and the other $k-1$ folds being used for training. The results from each of the k tests are then averaged to provide an overall estimate of the model's performance. The procedure permits to select hyperparameters, that are the configuration setting providing the best average result across the folds.

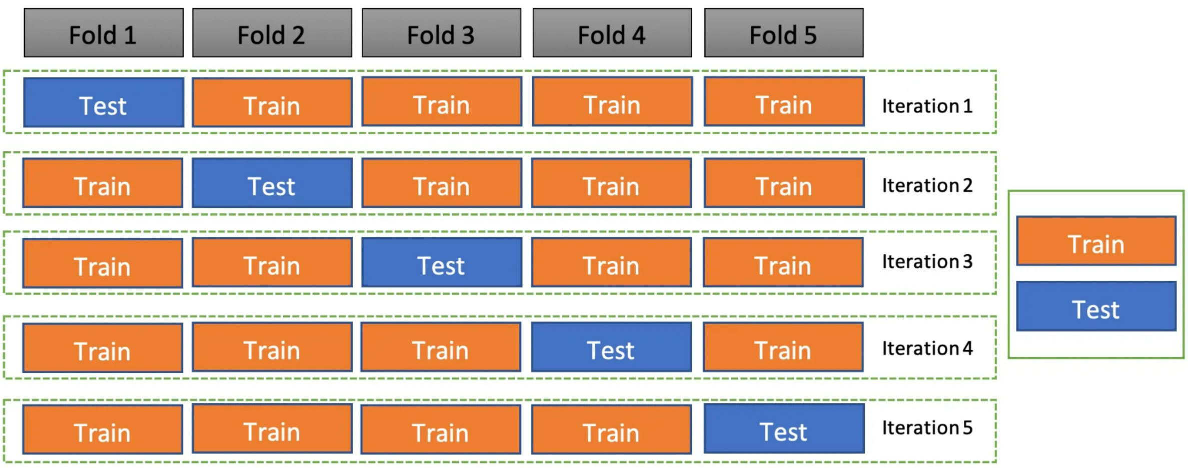


Figure 5.2: Example of k-fold cross validation procedure.

Due to the fact that this study lives in a time series framework, k-fold cross validation can not be applied; in fact, data would be shuffled while randomly divided into the k subsets. With time series this is a problem, because the data are ordered in time, and there is a relationship between them. Random shuffling of the dataset would cause the rupture of this relations causing a worse performance for the model.
To preserve the time ordering of the dataset the folds are built using a different cross validation technique, available in the paper of Pedregosa et al. [23]. In the procedure

different models are tested, keeping the WS fixed, to find the best values of α and $l1_{ratio}$ for that value of WS. This technique is based on three parameters:

- *max train size*;
- *test size*;
- *n splits.*

Max train size represents the maximum size for a single training set and it is set equal to WS.
Test size is the size of the test set internal to the cross validation and it is set equal to 1 because the model is trained to build the portfolio to replicate the next month return.
N splits is the number of splits of the dataset and it is defined as:

$$nsplits = \frac{size of the training set - window size}{test size}, \tag{5.1}$$

that is the exact number to obtain every fold of size equal to WS. With these setting of the parameters the Elastic Net is always trained with a set of dimension WS and the performance evaluated on a test set of dimension 1. The performance is determined by the value of the custom loss, explained in subsection 5.1.5, that are obtained. For each WS are chosen the α and $l1_{ratio}$ which guarantee the best average performance across the training set.

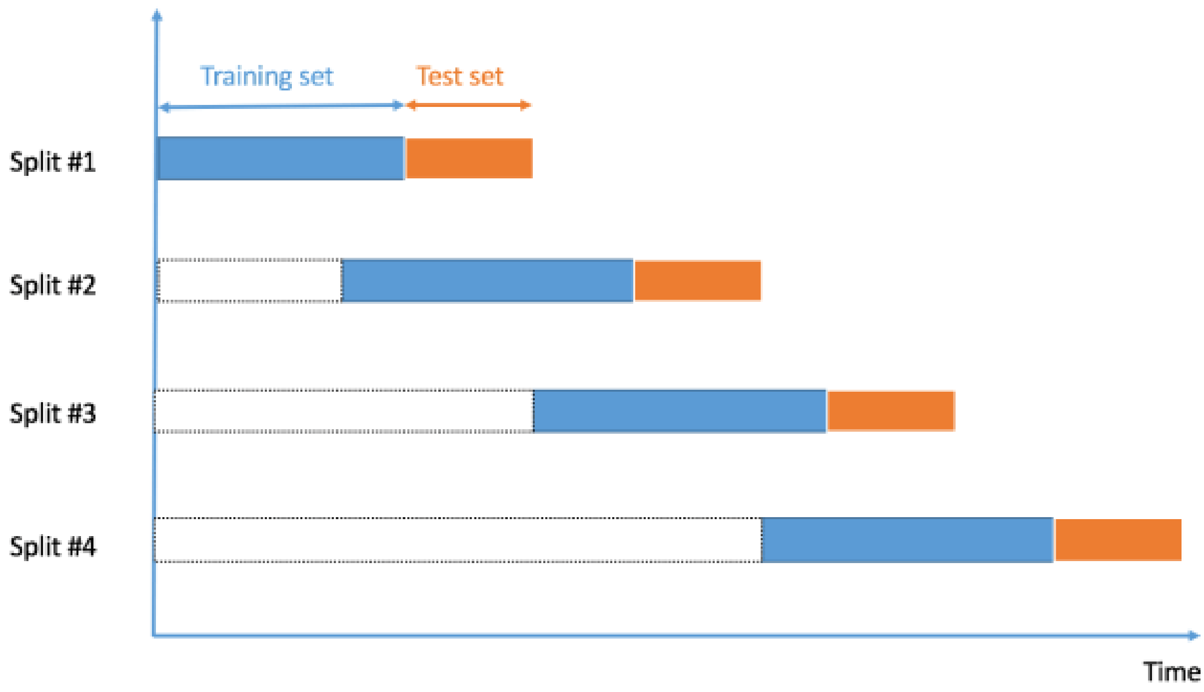


Figure 5.3: Example of cross validation procedure of *Model 1* with test size of length one, training size fixed and the windows rolling of one time step.

In the end the hyperparameter WS is tuned: at this point all the performances are available and the WS granting the best one is chosen.

5.1.5. Loss function

In machine learning algorithms the loss function is a function used to evaluate the performance during training; it calculates the distance between the predicted output and the actual output. The objective of the algorithm is to minimize the loss function by adjusting model's parameters.
The distance calculated in the loss function can be measured using different metrics, some examples are: mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and the symmetric mean absolute percentage error (SMAPE). The metric used in the loss function in this work is the SMAPE, one of the measure of forecast accuracy described by Armstrong [3]. SMAPE has been chosen because it is a symmetric measure, taking into account the relative difference between the predicted and actual values. This means that SMAPE is not affected by the direction of the prediction error and it gives equal weight to overestimation and underestimation errors. The objective of the model is to perfectly replicate the returns of the target portfolio, for this reason obtaining higher returns is equally weighted to

obtaining lower returns.
The SMAPE is defined as following:

$$\text{SMAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|)/2} \times 100 \tag{5.2}$$

where:

- n is the number of observations in the dataset;
- y_i is the actual (true) value for observation i;
- $\hat{y}_i$ is the predicted value for observation i.

As said before, the model has to minimize the loss function. The Elastic Net presented in this work uses a coordinate descendent algorithm to do it. This algorithm, explained by Zou and Hastie [32], updates the coefficients of the model one at a time while fixing the values of the other coefficients. During each cycle of the coordinate descent algorithm, the coefficients are updated in a specific order, with each coefficient being updated using a closed-form solution. The order of the updates is typically randomized at the beginning of each cycle to prevent any biases.

5.1.6. Final overview

To sum up, *Model 1* aims to solve the task of portfolio replication. It finds the best weights to build a portofolio replicating the target using a rolling windows approach. At the core of the model there is a linear regression, performed with an Elastic Net to allow regularization.
Model 1 depends on three parameters: the size of the rolling windows WS, α and $l1_{ratio}$ of the Elastic Net. The first parameter is tuned minimizing the tracking error volatility (TEV), a metric discussed in section 5.3, while the other two are found using cross validation and minimizing the SMAPE.
This model can be directly tested out-of-sample, fitting at each period the obtained Elastic Net with the set of size WS of the previous returns of the factors and of the target. This procedure may lead to a rebalancing of the portfolio between two consecutive periods.
Metrics used to evaluate the performance of *Model 1* are later discussed in section 5.3.

5.2. Model 2

5.2.1. Short overview

For *Model 2* a simpler model has been chosen to have a benchmark for *Model 1*. Its most important feature is that it is static; weights obtained after training are kept constant until the end of the test set.
The frequency of rebalancing depends on various factors, such as the investment strategy, the market conditions, and the investor's risk tolerance. Some investors may rebalance their portfolios periodically, such as monthly or quarterly, while others may prefer to rebalance based on specific triggers or thresholds. Strongly linked with the frequency of rebalancing are cost of transactions; indeed a no-rebalancing strategy leads to less costs than the strategy provided by previous model. Constant weights means that the portofolio is no more rebalanced and consequently the frequency of rebalancing is no more a parameter of interest.
Model 2 is based on a linear regression. As in *Model 1*, the regression is performed with an Elastic Net to apply a regularization and optimize the number of assets in the portfolio. The pipeline of this model is shorter and can be summed up in these steps:

- the dataset is splitted in train set and test set;
- an Elastic Net is trained using the whole training set and the best α and $l1_{ratio}$ are fixed by the cross validation minimizing the custom loss;
- the performance of the model is tested out-of-sample on the test set.

5.2.2. Hyperparameters

The hyperparameters of this model are the two parameters of the Elastic Net, α and $l1_{ratio}$. The grid of values among which they are chosen is the same of *Model 1*, shown in Table 5.2. For the tuning it is used again Bergstra and Bengio technique [4], sampling them randomly from the grid and to select the best combination, 20000 attempts are executed and the best one is chosen.

5.2.3. Cross validation

Even though it is static, the model lives in a time series framework and therefore it needs a different cross validation procedure rather than the standard k-fold.
Also the procedure use in this model, contained in Pedregosa et al. [23], takes into account the temporal ordering of the data but it is conceptually different from the one used in *Model 1*.

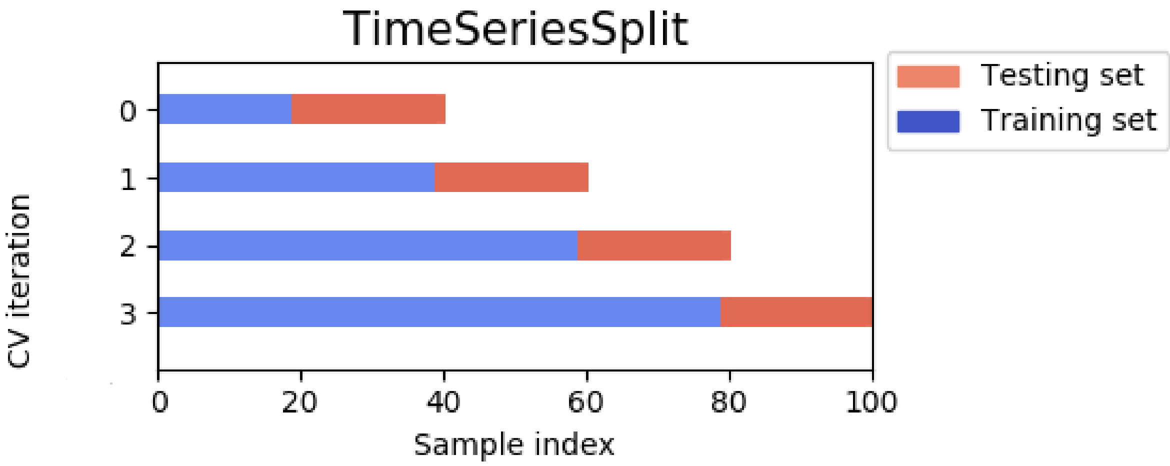


Figure 5.4: Example of cross validation procedure of *Model 2*.

The technique is a variation of k-fold cross-validation: the main idea behind this procedure is to split the time series data into several folds, where each fold consists of a contiguous subset of time points. The most important concept of the procedure is that successive training sets are supersets of those that come before them, to ensure that the model is trained on data that precedes the test data in time. For each fold i, the first i folds are used as training set and the $i + 1th$ fold is the test set. The overall performance of the model is the average of the performances across all folds.
Hyperparameters of *Model 2* are two but it could have been considered a third. Indeed, this cross validation technique requires as input parameter the number of splits of the folds. Nevertheless, a number of splits equal to five has been chosen a priori. The choice of the number of splits in cross-validation is somewhat arbitrary and can depend on the specific data and modeling problem. However, the value of five is a common default choice for the number of folds, explained by Hastie et al. [16], because it provides a good trade-off between bias and variance in the estimated performance. A larger value of k can reduce the bias of the estimated performance by reducing the dependence of the estimate on a single partition of the data; but, it can also increase the variance of the estimate by reducing the amount of data used for training the model.

5.2.4. Loss function

Model 2 does not use a customized loss function but the standard one for Elastic Net regression: the mean squared error (MSE).
MSE measures the average of the squared differences between the predicted values and the actual values of the target variable. This is its formulation:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{5.3}$$

where:

- n is the total number of observations;
- y_i is the actual value of the target variable for the i-th observation;
- $\hat{y}_i$ is the predicted value of the target variable for the i-th observation.

The MSE is non negative and it is in the same unit as the target variable, making it easy to be interpreted; a smaller value indicates better performance of the model. One possible drawback of the MSE is that it gives more weight to large errors than to small errors, as highlighted by Willmott et al. [30], due to the squaring operation; but in this case study, considering the order of magnitude of returns it has not been considered a problematic scenario.

5.2.5. Final overview

To sum up, *Model 2* aims to solve the task of portfolio replication. Differently than the rolling windows approach in which only the more recent data are used for the training, this model is trained with all the returns of the past that are available.
α and $l1_{ratio}$, the values determining the configuration setting of the Elastic Net, are the only two parameters of the model. They are selected using cross validation and random search combined, minimizing the MSE.
The weights of the replica portfolio are found using all the returns of the past and then are kept constant out-of-sample, there is no rebalancing.
Metrics used to evaluate the performance of *Model 2* are later discussed in section 5.3.

5.3. Metrics

In this section metrics used to evaluated the performance of the models are explained.
In machine learning, a metric is a quantitative measure used to to determine how well a model is able to make predictions on new, unseen data. Some common metrics for classification tasks include accuracy, precision, recall, and F1 score. For regression tasks, as the Elastic Net in this study, common metrics include mean squared error (MSE) ,

mean absolute error (MAE) , and R^2.
In portfolio replication problems, the metric used to evaluate the performance of a portfolio is typically the tracking error volatility (TEV). This metric has been used in several works; some examples are Giamouridis and Paterlini [13], the book of Grinold et al. [14] and Roncalli [25].
Tracking error volatility measures the deviation of a portfolio's returns from the returns of the target. It is calculated as the standard deviation of the difference between the portfolio returns and the target returns. A low TEV indicates that the portfolio closely matches the target, while a high TEV indicates that the portfolio is deviating significantly from the target.
In this study the tracking error volatility is chosen as metric to tune the WS hyperparameter in the *Model 1* and to evaluate performance of both models on the test set. Data are monthly, but the measure of interest is the annual TEV; it is obtained scaling the monthly tracking error volatility by the square root of the number of periods in a year. In this case, since there are 12 periods (months) in a year, it used the square root of 12 to scale the monthly tracking error volatility to an annualized tracking error volatility:

$$TEV_a = \sqrt{12} * TEV_m, \tag{5.4}$$

where:

- TEV_a is the annual tracking error volatility;
- TEV_m is the monthly tracking error volatility, standard deviation of the difference between the portfolio returns and the target returns.

To better evaluate the performance of the models on the test set and have a wider view, TEV is not the only metric chosen. The other metrics used are excess return (ER), correlation (CORR) and SMAPE, already used in the loss function.
ER is defined as the sum of the differences between the predicted returns and the target returns on a specific period:

$$ER = \sum_{i=1}^{n} (y_{rp,i} - y_{t,i}), \tag{5.5}$$

where:

- $y_{rp,i}$ is the is the return of the replica portfolio in period i;
- $y_{t,i}$ is the return of the target in period i;

- n is the total number of periods over which the excess return is calculated.

ER has been widely used in the literature: for example, by Hou et al. [19] as a metric to evaluate the performance of a portfolio replication strategy based on anomalies in stock returns; or by Holthausen et al. [18] to evaluate the prediction of stock returns obtained with financial statement information.
CORR is defined as the correlation between the returns of the replica portfolio and the target returns, measuring the strength of the linear relationship between them:

$$\rho_{rp,t} = \frac{cov(y_{rp}, y_t)}{\sigma_{rp}\sigma_t}, \tag{5.6}$$

where:

- $\rho_{rp,t}$ is the correlation coefficient between the returns of replica portfolio and target portfolio;
- $cov(y_{rp}, y_t)$ is the covariance between the returns of replica portfolio and target portfolio;
- σ_{rp} is the standard deviation of the returns of replica portfolio;
- σ_t is the standard deviation of the returns of target portfolio.

When evaluating the performance of a portfolio replication model, it is important to not consider this metric alone because an high correlation does not necessarily imply that the predicted return accurately replicates the target return; Giamouridis and Paterlini [13], to overcome this problem, have combined CORR with excess return, tracking error volatility and turnover.

6 | Results

In this chapter we present the results of *Model 1* and *Model 2*. We use the models firstly with the set of 14 US Indexes and then with the Fama and French factors, comparing the results. Models are trained with returns from January 2001 to April 2017 and are then tested out-of-sample from May 2017 to September 2022 using 13 target portfolios. Then the results are analyzed considering the fact that some of these portfolios are sectorials and others are obtained with a linear combination of different instruments.
The results of the replica of all target portfolios are presented in tables; among the columns, between the metrics, model parameter WS and N are reported. Considering that *Model 1* admits rebalancing between a time step and the following one, N is the average number of factors composing the replica portfolio in each time step. For *Model 2*, where the portfolio is fixed across the test set, N is not an average but a constant number. For the calculation of N a threshold of 5% has been chosen: if a factor is present in the portfolio with a weight less than the threshold is not added to the count.

6.0.1. Target portfolios

13 target portfolios are built using the Black Box portfolios presented in section 3.3. Depending on their composition, they are divided in two categories: *sectorial* and *mixed*. A sectorial portfolio is composed by only one security, instead a mixed portfolio is a linear combination of different portofolios.
As shown in Table 6.1, the set of sectorial is composed by P5, P6, P7, P8 and P9; the set of mixed by P1, P2, P3, P4, P10, P11, P12 and P13.
The most diversificated portfolio is P12, where 7 Black Box portfolios out of 8 are used in the composition.

Ticker	SCG	BCM	LCV	CNSMR	MANUF	HITEC	HLTH	OTHER
P1	0.50	0.50						
P2		0.50	0.50					
P3	0.50		0.50					
P4	0.35	0.30	0.35					
P5				1.00				
P6					1.00			
P7						1.00		
P8							1.00	
P9								1.00
P10				0.40	0.30	0.20	0.10	
P11						0.30	0.35	0.35
P12		0.10	0.10	0.15	0.15	0.15	0.15	0.15
P13		0.50			0.25		0.25	

Table 6.1: Weights of the target portfolios. See section 3.3 to have informations about the Black Box portfolios.

6.1. Model 1 results

Table 6.2 and Table 6.3 contain *Model 1* results, while Table 6.4 contains the difference between the values in the two tables.
Average results show that there are no important differences in the performance of *Model 1* using the two sets of factors. TEV, which is the most important among the metrics, is in average 4.37% in the US Indexes case and 4.32% for the Fama and French factors. The difference of ER is in average 5.97%: the replica with US Indexes performs better than the target portfolio more than what Fama and French replica does.
In general, the replicating portofolio is able to overperform the target; the only exception is for P7, where Fama and French replica has a very negative ER equal to -17.19%. This situation has been investigated and in Figure 6.1 weights of the replica portfolios are shown. P7 is fully composed by one portfolio, the Hitec. In the case of US Indexes, the replica portfolio is mostly composed by the XNDX, while for Fama and French by the factor Mkt-RF which across the test set period has underperformed the Hitec portfolio. With P7 the limitations of the set of factors composed by the 5 Fama and French factors emerge: it can happen that the idiosyncratic risk of a particular sector is not represented by any of the 5 factors and consequently the replica is not accurate.

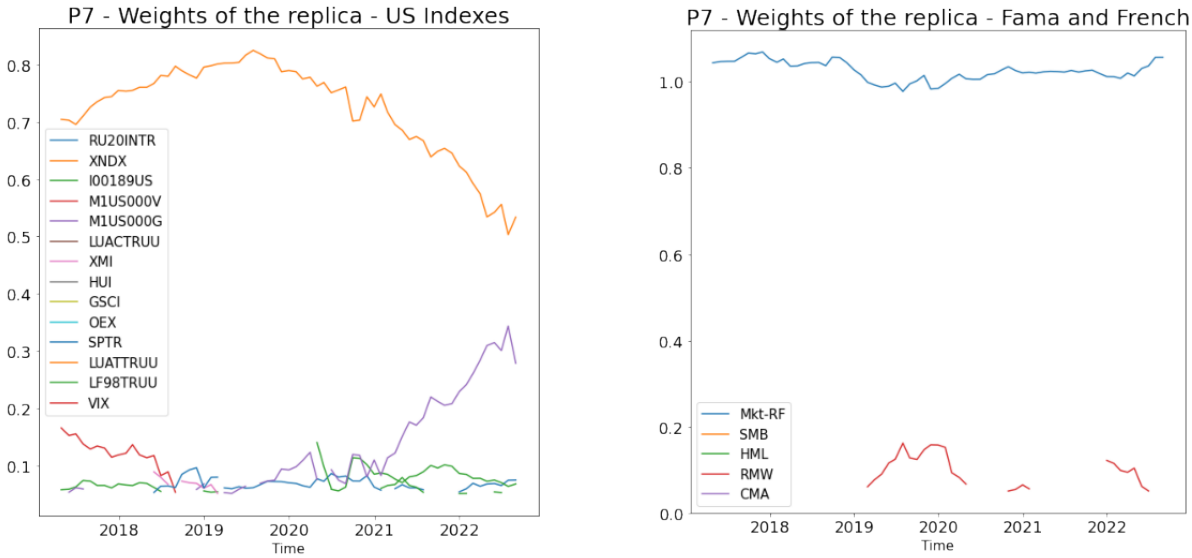


Figure 6.1: Weights of the portfolio replicating P7.

Another situation that deserves to be mentioned is the P8, fully composed by Hlth and representation of the healthcare sector. Its replica is the worst performance of *Model 1* with both set of factors. A motivation for this can be found in the fact that healthcare policies and regulations have a significant impact on the stock market; for example, a change in healthcare policies and regulations can strongly impact the profitability of healthcare companies. Among the factors, and considering how the model works, there is no possibility to take in account this kind of idiosyncratic risk.
As happened for the worst performance, also the best one is reached in the replica of the same portfolio with both set of factors; indeed, TEV of the replica portfolio of P12 with US Indexes is 1.34% and 1.80% with Fama and French factors. P12 is the most diversified portfolio, composed of 7 securities. In Figure 6.2 prices of P12 and of the two replicas are compared. It can be noticed that in the first 3 years price of the target and of the replicas are almost the same; the slight distance present after the end of 2020 is due to different returns during the first wave of Covid-19 pandemic, as it can be seen in Figure 6.3.
The smallest value for WS is 27: there are no replica portofolio in which the optimal size of the rolling window is less than 2 years. The largest value is 60: for P4 the Elastic Net of the model builds the weights of the replica portfolio using the returns of the last 5 years.
The average number of replica portofolio components with a weight larger than 5% is 4.99 for US Indexes replica and 3.86 for Fama and French. Considering that the US Indexes are 14, the Elastic Net successfully performs feature selection and it keeps in the replica

portfolio only the most appropriate factors.

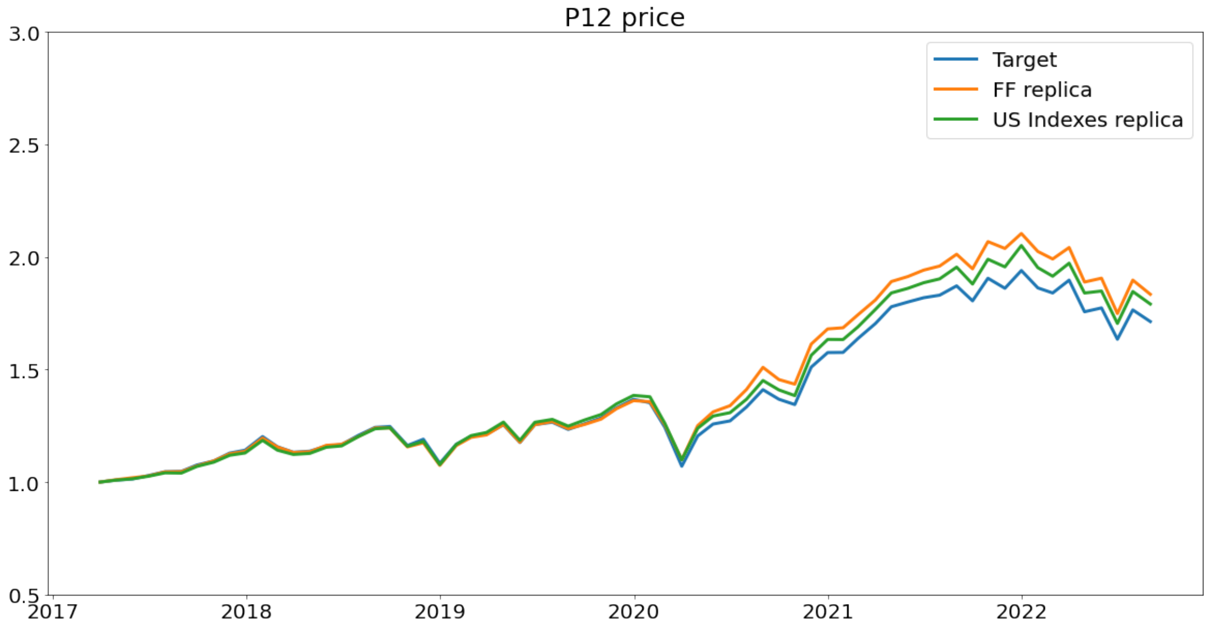


Figure 6.2: P12 price.

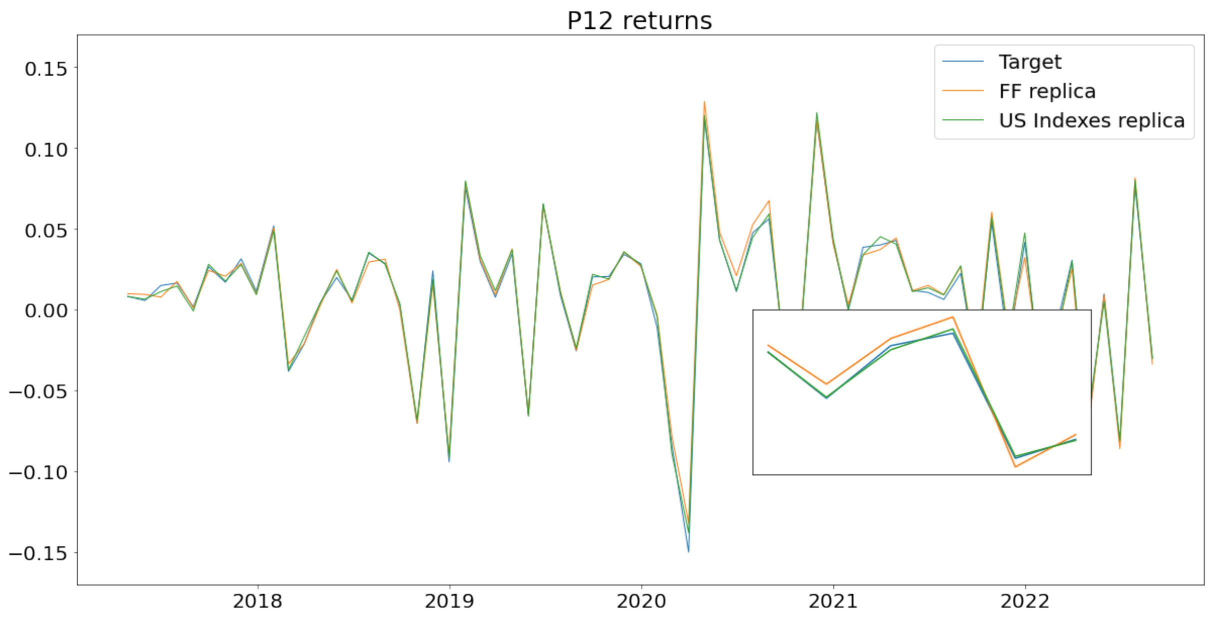


Figure 6.3: P12 returns; the zoom-in box covers from March 2020 to October 2020.

Ticker	WS	ER	CORR	SMAPE	TEV	N
P1	48	15.42%	98.72%	20.00%	3.19%	3.59
P2	27	6.06%	96.25%	23.43%	5.69%	6.89
P3	30	2.94%	97.62%	16.97%	4.95%	4.83
P4	60	2.76%	98.83%	15.99%	3.24%	4.95
P5	54	6.25%	94.95%	29.95%	5.80%	5.48
P6	39	5.36%	95.40%	36.60%	5.58%	4.03
P7	52	10.12%	98.61%	16.52%	3.16%	3.59
P8	44	18.33%	80.50%	45.85%	9.93%	6.51
P9	30	6.10%	96.05%	28.51%	5.52%	3.74
P10	52	5.11%	99.11%	17.72%	2.32%	5.06
P11	27	8.27%	98.27%	20.39%	3.07%	5.09
P12	39	5.34%	99.69%	8.21%	1.34%	4.63
P13	39	16.26%	98.38%	22.86%	2.99%	6.49
Average	41.62	8.33%	96.34%	23.31%	4.37%	4.99

Table 6.2: *Model 1*: results with US Indexes.

Ticker	WS	ER	CORR	SMAPE	TEV	N
P1	33	14.99%	98.41%	19.37%	3.55%	4.49
P2	33	13.09%	98.35%	18.44%	3.68%	3.26
P3	39	2.32%	98.19%	16.71%	4.25%	4.81
P4	42	9.99%	98.41%	14.61%	3.67%	4.58
P5	48	5.14%	95.74%	25.54%	5.51%	4.75
P6	52	15.56%	94.44%	36.60%	6.37%	2.37
P7	48	-17.19%	97.07%	21.36%	4.59%	4.23
P8	44	23.90%	85.68%	38.86%	8.12%	4.41
P9	33	5.79%	96.97%	20.51%	4.81%	3.84
P10	44	9.01%	99.09%	17.05%	2.27%	3.64
P11	33	3.47%	98.26%	19.49%	3.15%	3.11
P12	54	6.89%	99.45%	1.95%	1.80%	2.71
P13	52	23.75%	96.23%	24.78%	4.39%	3.95
Average	42.69	11.62%	96.64%	21.17%	4.32%	3.86

Table 6.3: *Model 1*: results with Fama and French factors.

Ticker	WS	ER	CORR	SMAPE	TEV	N
P1	15	0.43%	0.31%	0.63%	-0.36%	-0.90
P2	-6	-7.03%	-2.10%	4.99%	2.01%	3.63
P3	-9	0.62%	-0.57%	0.26%	0.70%	0.02
P4	18	-7.23%	-0.42%	1.38%	-0.43%	0.37
P5	6	1.11%	-0.79%	4.41%	0.29%	0.73
P6	-13	-10.20%	0.96%	0.00%	-0.79%	1.66
P7	4	27.31%	1.54%	-4.84%	-1.43%	-0.64
P8	0	-5.57%	-5.18%	6.99%	1.81%	2.10
P9	-3	0.31%	-0.92%	8.00%	0.71%	-0.10
P10	8	-3.90%	0.02%	0.67%	0.04%	1.42
P11	-6	4.80%	0.01%	0.90%	-0.08%	1.98
P12	-15	-1.55%	0.24%	6.26%	-0.46%	1.92
P13	-13	-7.49%	2.15%	-1.92%	-1.40%	2.54
Average	8.92	5.97%	-0.30%	2.13%	0.05%	1.13

Table 6.4: *Model 1*: delta of results.

6.2. Model 2 results

Table 6.5 and Table 6.6 contain *Model 2* results, while in Table 6.7 the difference between values in the two tables is reported.
As for *Model 1*, average results show that there are no important differences in the performance of *Model 2* using the two sets of factors. TEV is on average 4.44% in the US Indexes replica and 4.60% in the Fama and French one.
As for *Model 1*, the worst replica is the one of the target portfolio P8; TEV of the portfolio built using the US Indexes set of factors is equal to 9.85% and equal to 8.84% using Fama and French factors. In Figure 6.4 and Figure 6.5 the bad performance of the model can be seen.
As for *Model 1*, the best result is obtained with the replica of P12: CORR is the highest and TEV is the lowest for US Indexes replicas and also for Fama and French ones.
As for *Model 1*, due to the negative ER of the replica with Fama and French factors (-10.97%), the composition of the target portfolio P7 is investigated and can be seen in Figure 6.6: for US Indexes the XNDX is predominant and the 5 Fama and French factors lack of one or more additional factor to provide an accurate replica.
As for *Model 1*, Elastic Net successfully performs feature selection; the average number of

factors in the replica portofolio with a weight larger than 5% is 4.46 for US Indexes and 2.31 for Fama and French factors.

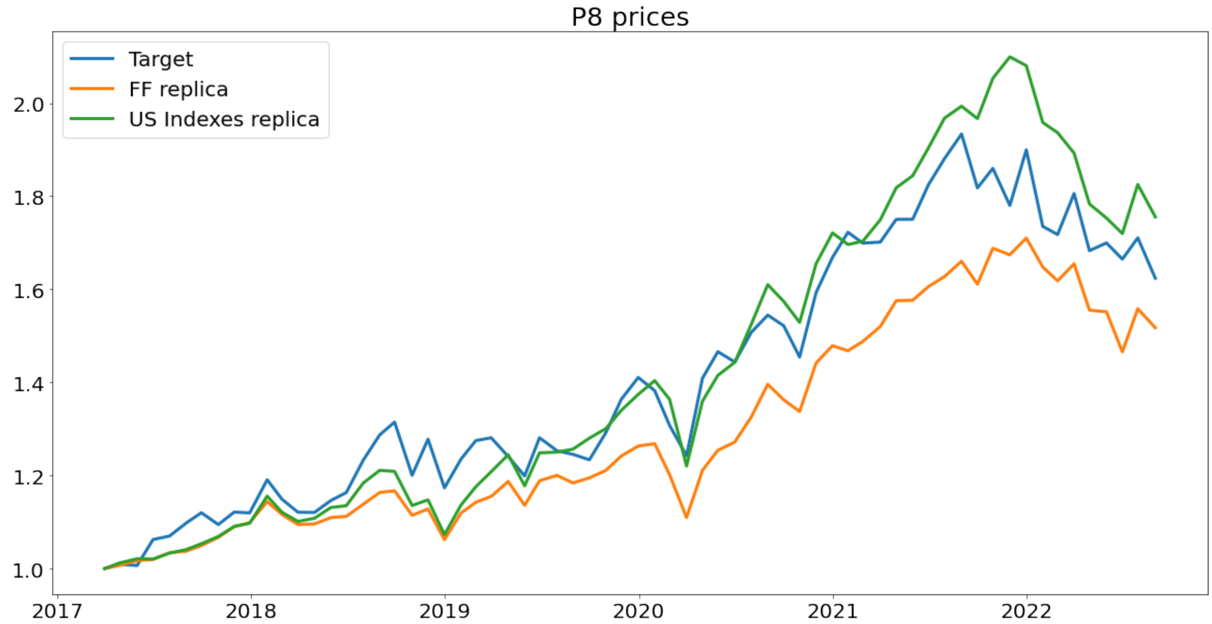


Figure 6.4: P8 price.

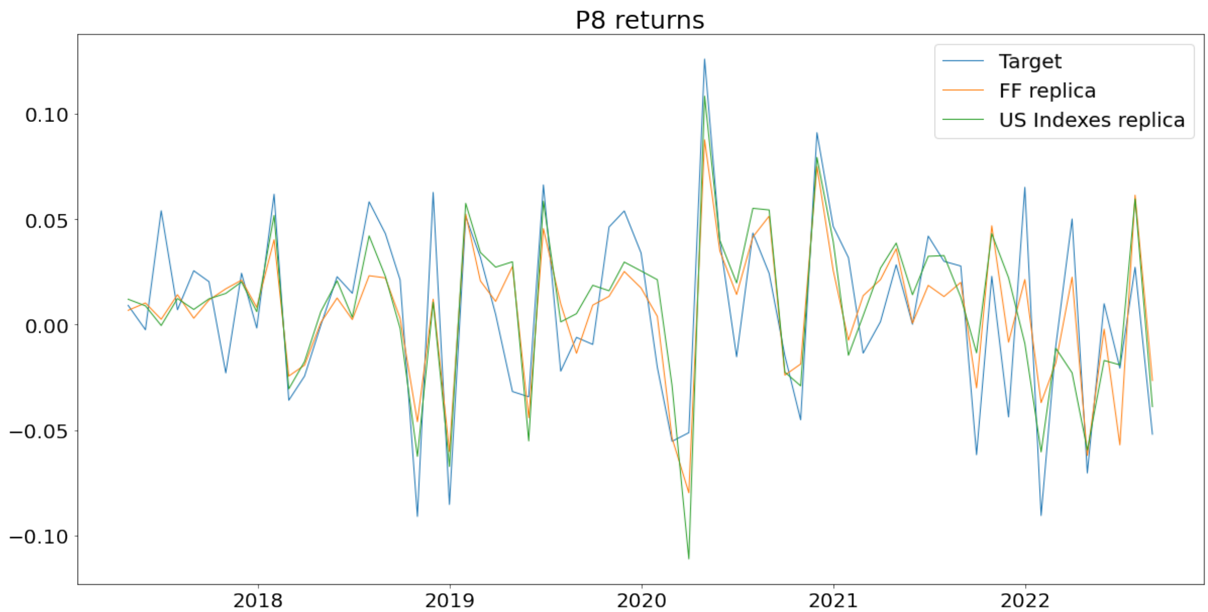
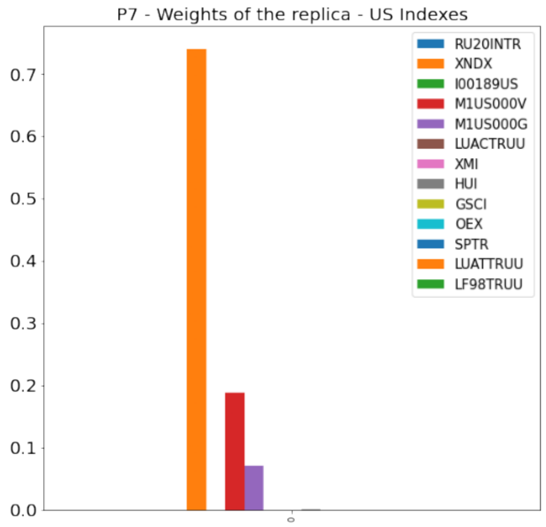


Figure 6.5: P8 returns.



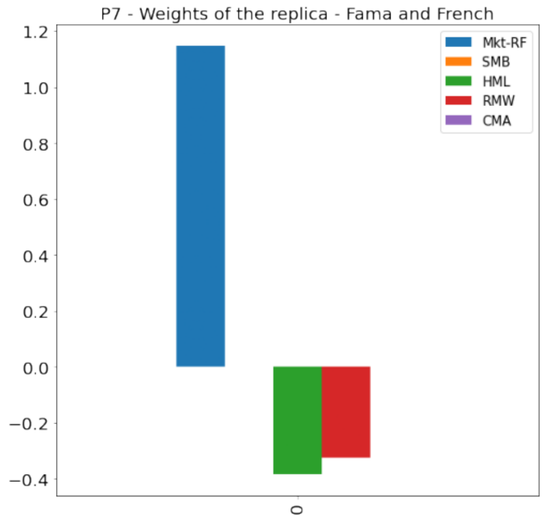


Figure 6.6: P7 weights comparison.

Ticker	WS	ER	CORR	SMAPE	TEV	N
P1	-	6.34%	98.74%	18.18%	3.31%	5
P2	-	3.08%	96.80%	26.12%	5.35%	3
P3	-	-16.70%	93.77%	35.75%	7.33%	2
P4	-	2.53%	98.80%	15.28%	3.35%	3
P5	-	8.00%	94.10%	35.94%	6.39%	3
P6	-	6.40%	98.91%	16.79%	2.80%	6
P7	-	6.50%	98.91%	16.79%	2.82%	3
P8	-	8.40%	76.54%	46.31%	9.85%	9
P9	-	8.80%	96.67%	25.49%	5.05%	5
P10	-	-1.49%	99.19%	17.65%	2.50%	3
P11	-	5.14%	98.38%	19.03%	2.98%	5
P12	-	1.30%	99.69%	10.82%	1.40%	4
P13	-	14.27%	96.20%	26.40%	4.41%	7
Average	-	6.65%	95.86%	24.25%	4.44%	4.46

Table 6.5: *Model 2*: results with US Indexes.

Ticker	WS	ER	CORR	SMAPE	TEV	N
P1	-	8.55%	98.64%	17.40%	3.30%	2
P2	-	9.80%	97.91%	18.92%	4.39%	2
P3	-	-9.51%	99.00%	14.52%	3.55%	3
P4	-	4.86%	98.52%	13.81%	3.71%	4
P5	-	-17.57%	95.17%	34.73%	6.88%	1
P6	-	19.02%	92.63%	37.22%	6.98%	4
P7	-	-10.97%	95.75%	29.04%	6.00%	1
P8	-	-6.78%	81.65%	46.24%	8.84%	1
P9	-	6.67%	96.91%	22.46%	4.97%	2
P10	-	1.40%	99.34%	16.45%	1.96%	3
P11	-	-0.03%	98.59%	18.26%	2.82%	2
P12	-	2.31%	99.36%	12.94%	1.99%	2
P13	-	18.92%	95.76%	25.23%	4.67%	3
Average	-	8.95%	96.09%	23.58%	4.60%	2.31

Table 6.6: *Model 2*: results with Fama and French factors.

Ticker	WS	ER	CORR	SMAPE	TEV	N
P1	-	-2.21%	0.10%	0.78%	0.01%	3
P2	-	-6.72%	-1.11%	7.20%	0.96%	1
P3	-	-7.19%	-5.23%	21.23%	3.78%	-1
P4	-	-2.33%	0.28%	1.47%	-0.36%	-1
P5	-	25.57%	-1.07%	1.21%	-0.49%	2
P6	-	-12.62%	6.28%	-20.43%	-4.18%	2
P7	-	17.47%	3.16%	-12.25%	-3.18%	2
P8	-	15.18%	-5.11%	0.07%	1.01%	8
P9	-	2.13%	-0.24%	3.03%	0.08%	3
P10	-	-2.89%	-0.15%	1.20%	0.54%	0
P11	-	5.17%	-0.21%	0.77%	0.16%	3
P12	-	-1.01%	0.33%	-2.12%	-0.59%	2
P13	-	-4.65%	0.44%	1.17%	-0.26%	4
Average	-	8.09%	-0.19%	0.26%	-0.19%	2.15

Table 6.7: *Model 2*: delta of results.

6.3. Overall results

In Table 6.8 the results are divided in sections; one for each combination of model number and set of factors. Results are also divided for tipology of target portfolio. Values in the rows sectorial are the average of the values of the portfolios that replicate a sectorial target portfolio; the same holds for mixed portfolios.
It can be noticed that there is not a particular combination of model and set of factors that outperforms the others. The results are similar: CORR and ER of *Model 1* are slightly higher than the correspondents of *Model 2* but this relation does not hold for SMAPE and TEV.
This result is not surprising. Indeed, the target portofolios built to test the model are constant; they never change their composition and consequently they are always exposed to the same risk factors. For this reason *Model 2*, even though it keeps the same replica portfolio across the test period, does not have worse performance than *Model 1*.

Type	**WS**	**ER**	**CORR**	**SMAPE**	**TEV**	**N**
US INDEXES *Model 1*						
SECTORIAL	43.8	9.23%	93.10%	31.49%	6.00%	4.67
MIXED	40.25	7.77%	98.36%	18.20%	3.35%	4.98
AVERAGE	42.03	8.50%	95.73%	24.84%	4.67%	4.83
F. FRENCH *Model 1*						
SECTORIAL	45	13.52%	93.98%	28.57%	5.88%	3.92
MIXED	41.25	10.44%	98.30%	16.55%	3.35%	4.11
AVERAGE	43.13	11.98%	96.14%	22.56%	4.61%	4.02
US INDEXES *Model 2*						
SECTORIAL	-	7.62%	93.03%	28.26%	5.38%	5.20
MIXED	-	6.36%	97.70%	21.15%	3.83%	4.25
AVERAGE	-	6.99%	95.36%	24.71%	4.61%	4.73
F. FRENCH *Model 2*						
SECTORIAL	-	12.20%	92.42%	33.94%	6.73%	1.80
MIXED	-	6.92%	98.39%	17.19%	3.30%	2.25
AVERAGE	-	9.56%	95.41%	25.56%	5.02%	2.03

Table 6.8: Average results of each of the four combination model / factors. Values are the mean of sectorial portfolios and mixed portfolios.

In a problem of portfolio replication, in which there is the additional information that the target portfolio is constant, using *Model 2* may be preferable to avoid transaction costs and to keep the model simpler. Some tests about this topic are shown in chapter 7.
In addition, this result underlines the importance of choosing a good set of factors to build the replica portfolio. Considering that Fama and French 5 factors is a good set for portfolio replication purposes, the fact that using US Indexes leads to the same performances indicates that this set has been well composed and no risk factors have been left out. Anyway, a crucial role is played by the regularization technique contained in the model that allows to choose only the most important factors avoiding multicollinearity.

Type	WS	ER	CORR	SMAPE	TEV	N
AVERAGE *Model 1*						
SECTORIAL	44.40	11.37%	93.54%	30.03%	5.94%	4.29
MIXED	40.75	9.10%	98.33%	17.37%	3.35%	4.55
AVERAGE *Model 2*						
SECTORIAL	-	9.91%	92.72%	31.10%	6.06%	3.50
MIXED	-	6.64%	98.04%	19.17%	3.56%	3.25

Table 6.9: Results averaging US Indexes and Fama and French Replica.

Result shown in Table 6.9 are obtained averaging the US Indexes replica and Fama and French replica.
Looking at them it can be noticed that both *Model 1* and *Model 2* provide a better performance in replicating a target portfolio which is mixed rather than sectorial; mixed portfolios are composed by multiple securities and are diversified. Coherently, as seen in section 6.1, the best performance is the replica of P12, which is the most diversified portfolio composed by 7 securities. As discussed by Grinold et al. [14], systematic factors are common to many assets in the portfolio and can be better captured by a model, while the idiosyncratic factors are specific to individual assets and can be difficult to predict. For this reason, a diversified portfolio, which has a lower level of idiosyncratic risk, may be more predictable using a linear regression model.

7 | Tests and analysis on portfolio P4

In this chapter we choose a target portfolio and we present some tests and analysis on its replicating portfolio. The portfolio chosen is P4, because it is a good compromise between sectorial and diversification: statistics are summed up in Table 7.1 and visualized in Figure 7.1, Figure 7.2, Figure 7.3 and Figure 7.4. The replica using US Indexes is slightly better than the Fama and French and the best one is obtained with *Model 1*.

Type	**WS**	**ER**	**CORR**	**SMAPE**	**TEV**	**N**
US INDEXES *Model 1*	60	2.76%	98.83%	15.99%	3.24%	4.95
F. FRENCH *Model 1*	42	9.99%	98.41%	14.61%	3.67%	4.58
US INDEXES *Model 2*	-	2.53%	98.80%	15.28%	3.35%	3
F. FRENCH *Model 2*	-	4.86%	98.52%	13.81%	3.71%	4
AVERAGE	51	5.04%	98.64%	14.92%	3.49%	4.13

Table 7.1: Results of the replication of P4.

P4 is composed by 35% of SCG, 30% of BCM and 35% of LCV and the weights of the replicating portfolio are shown in Figure 7.3 and Figure 7.4. In the US Indexes replica the RU20INTR has a large weight during all the period tested; the other two important components are M1US000V and SPTR. A correspondence exists between the replicas with the two set of factors. Indeed, in the Fama and French replica the largest components are Mkt-RF and SMB which are deeply connected with securities in US Indexes replica. RU20INTR and SMB both account for small cap stocks and SPTR and M1USOOOV are a good representation of the whole equity market.

To conclude the analysis, in the end of the chapter we show a couple of applications of the replicating portfolio.

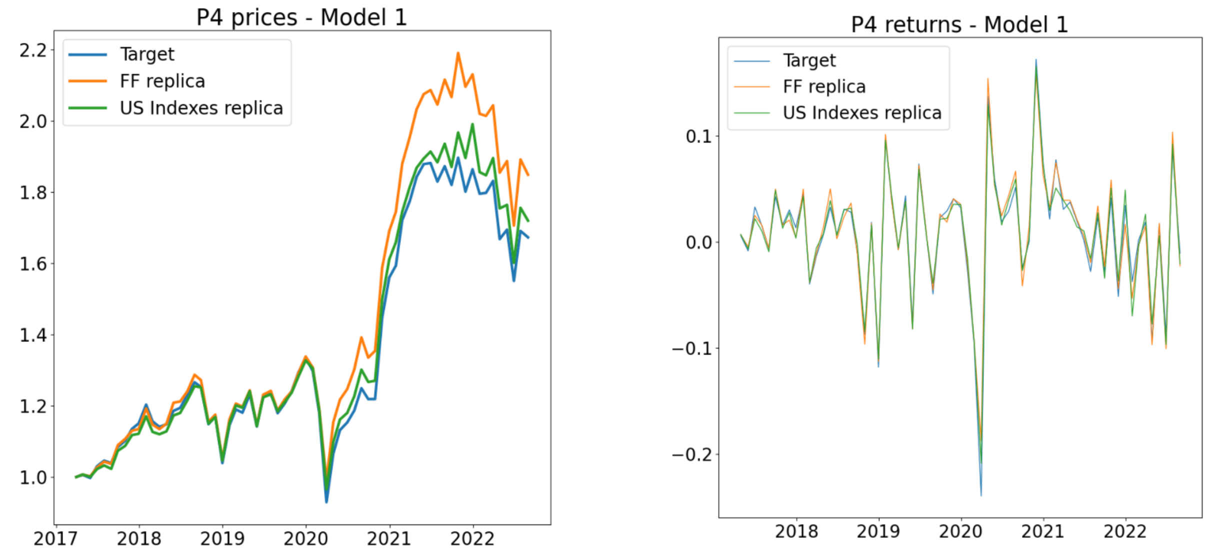


Figure 7.1: P4 analysis: price and returns calculated with *Model 1* using US Indexes and Fama and French factors.

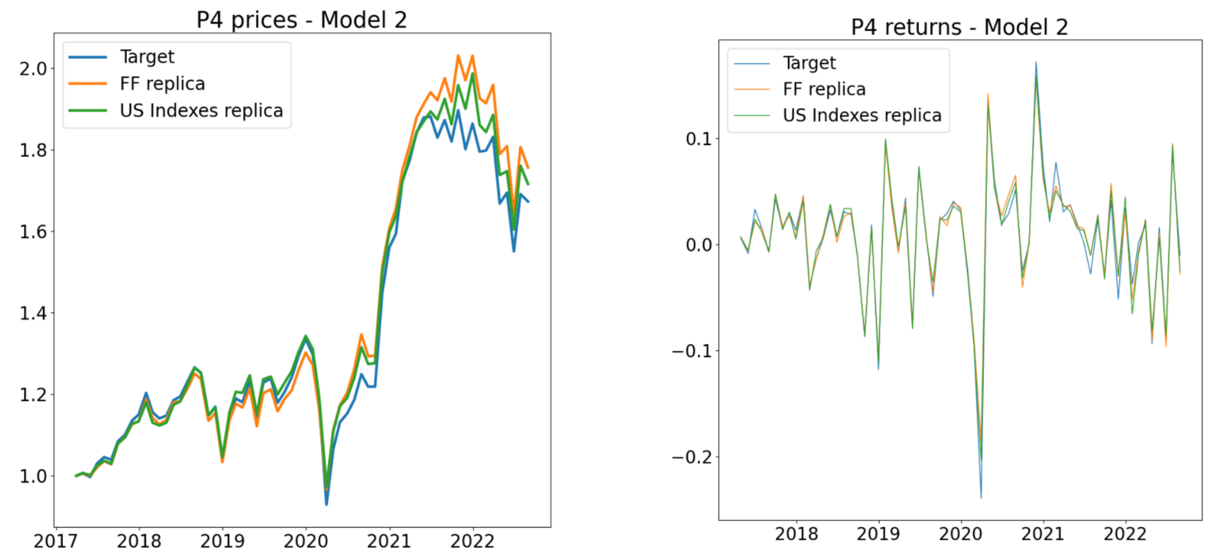


Figure 7.2: P4 analysis: price and returns calculated with *Model 2* using US Indexes and Fama and French factors.

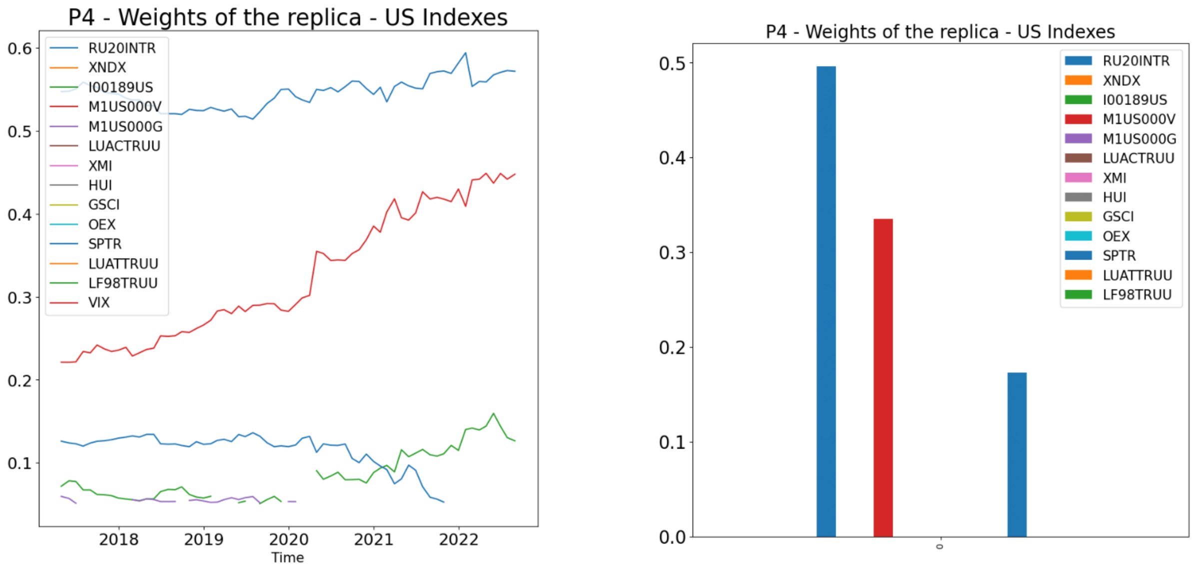


Figure 7.3: P4 analysis: weights of the portfolio built by *Model 1* and *Model 2* using US Indexes.

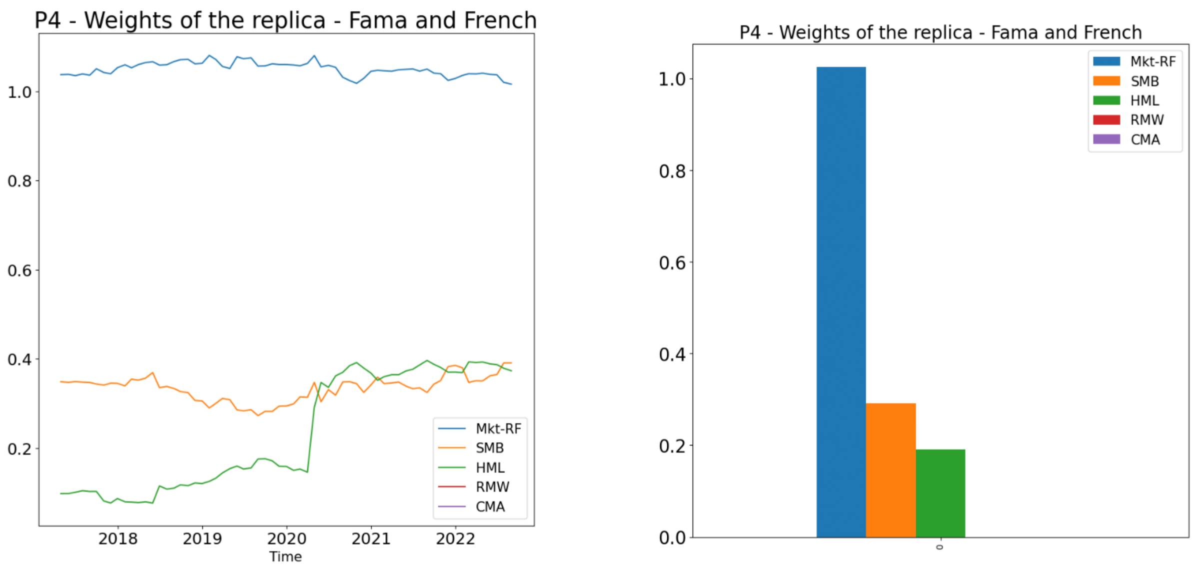


Figure 7.4: P4 analysis: weights of the portfolio built by *Model 1* and *Model 2* using Fama and French factors.

7.1. Tests and analysis

7.1.1. Stochastic target weights

In chapter 6 the target portfolios built to test the model are constant; they never change their composition and consequently they are always exposed to the same risk factors. In this section a test on the replica of P4 is performed: a stochastic component is introduced and the target weights are no more constant.
Before the explanation of the two scenario considered, the procedure needs to be clarified. The stochasticity is introduced only in the target weights of the test and models are trained with constant target weights. The situation represented by this procedure is the worst case scenario: the model has been trained to replicate a portfolio assumed to be constant, but the portfolio manager of the target has decided to change his strategy and the weights are adjusted during time. This method is a robustness test for both *Model 1* and *Model 2*, with the expectations of a better performance of the former thanks to the admitted rebalancing.
The stochasticity is introduced updating the target portfolio adding a gaussian jump to the weights:

$$w_{i,t} = w_{i,t-1} + z \tag{7.1}$$

where:

- $w_{i,t}$ is the weight of the i component at time t;
- $w_{i,t-1}$ is the weight of the i component at time $t-1$;
- z is gaussian variable with mean equal to zero and standard deviation equal to σ.

Only one simulation is considered. The frequency of the jumps depends on a variable Q, which has a Bernoulli distribution of parameter p and admits only 0 or 1 as values: if the value of Q simulated is equal to 1 the jump is applied to the weights, if the value is equal to 0 nothing happens. Two different configurational settings of the parameters are tried out: in the *Scenario 1* the value of p is equal to 0.40 and σ is equal to 0.01; in the *Scenario 2* the value of p is equal to 0.10 and σ is equal to 0.05. The evolutions of the portfolio weights of this simulation are in Figure 7.5.

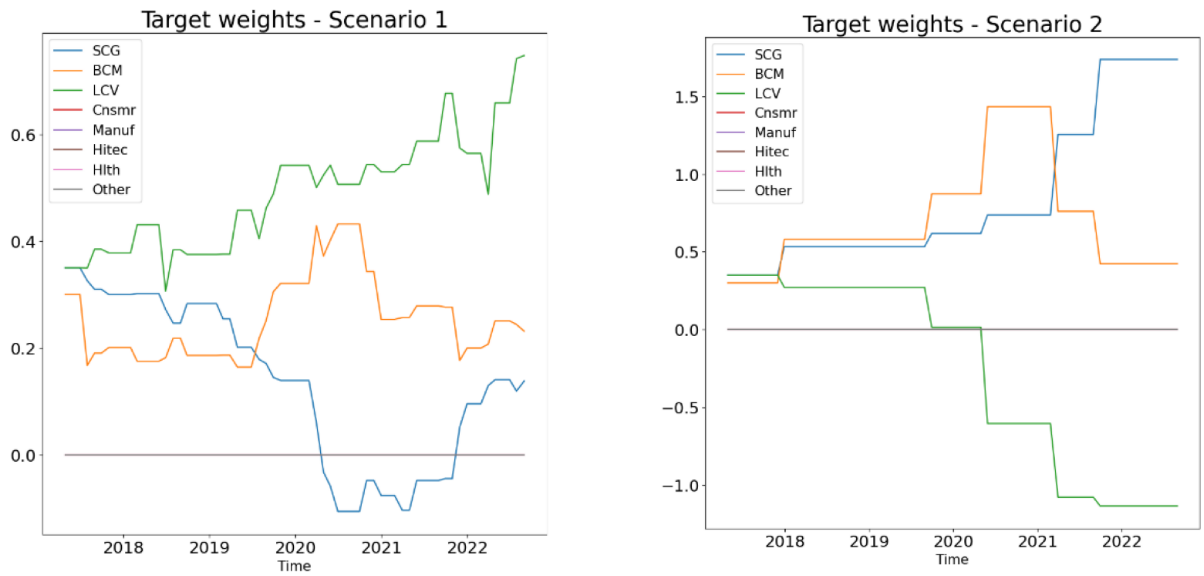


Figure 7.5: P4 stochastic weights comparison.

Scenario 1 represents the situation of a portfolio manager who decides to make small adjustments to the weights frequently. In each month there is a probability equal to 40% that a rebalance takes place; furthermore, due to the fact that its volatility is equal to 1%, the adjustment is most likely to be small. The results of the test are contained in Table 7.2 and the predicted returns are shown in Figure 7.6.
As expected, the result of each combination are worse than the results with a deterministic target portfolio. Despite this, value of the metrics are still low for each replica: TEV is lower than 5%, CORR is higher than 95% and SMAPE is lower than 20%.

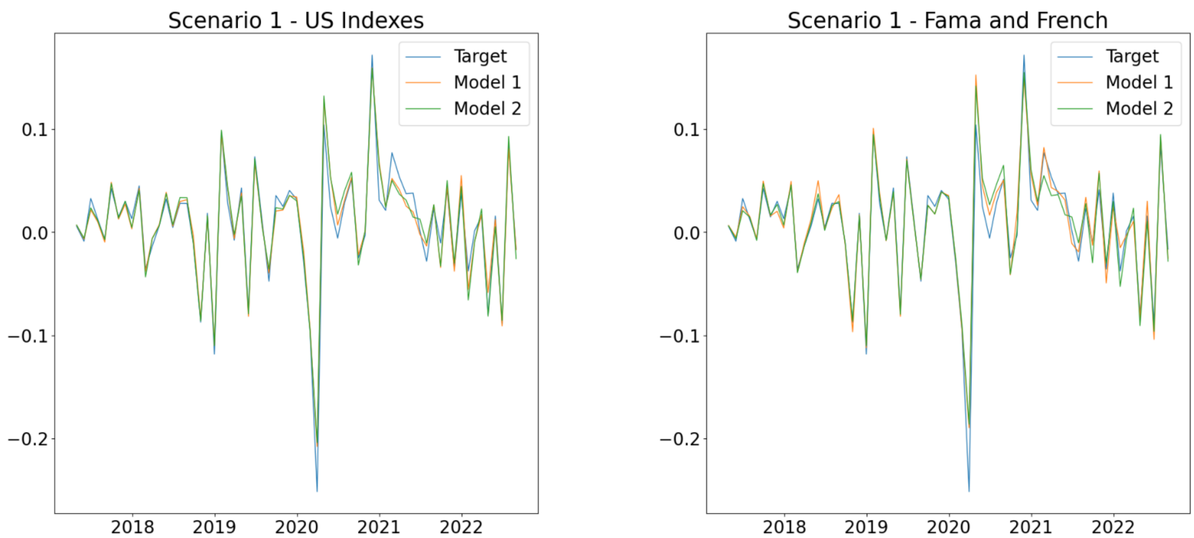


Figure 7.6: Scenario 1 replicating portfolios returns.

Type	WS	ER	CORR	SMAPE	TEV	N
US INDEXES *Model 1*	60	8.09%	97.85%	18.92%	4.21%	5.23
F. FRENCH *Model 1*	42	17.43%	97.18%	18.21%	4.75%	4.67
US INDEXES *Model 2*	-	9.50%	96.48%	19.56%	4.80%	3
F. FRENCH *Model 2*	-	11.84%	96.96%	17.84%	4.94%	4
AVERAGE	51	11.72%	97.12%	18.63%	4.68%	4.23

Table 7.2: Scenario 1 Results.

Model 1 has better performance than *Model 2* and in both cases the best set of factors for the replica is US Indexes.

Scenario 2 represents the situation of a portfolio manager that keeps the weights constant for a longer period during the year, but when he decides to rebalance the portfolio is because he wants to change his strategy and consequently the adjustment is larger in value than *Scenario 1*. In each month there is a probability equal to 10% that a rebalance takes place; furthermore, due to the fact that its volatility is equal to 5%, the adjustment is larger. The results of the test are contained in Table 7.3 and the predicted prices are shown in Figure 7.7.

In this scenario the replica is worse than *Scenario 1* and the prediction ability is considerably lower; models are not able to capture the rebalancing of the target portfolio. As it can be seen in Figure 7.7 the performance is particularly poor from mid 2020, after when some jumps are concentrated in a short time span.

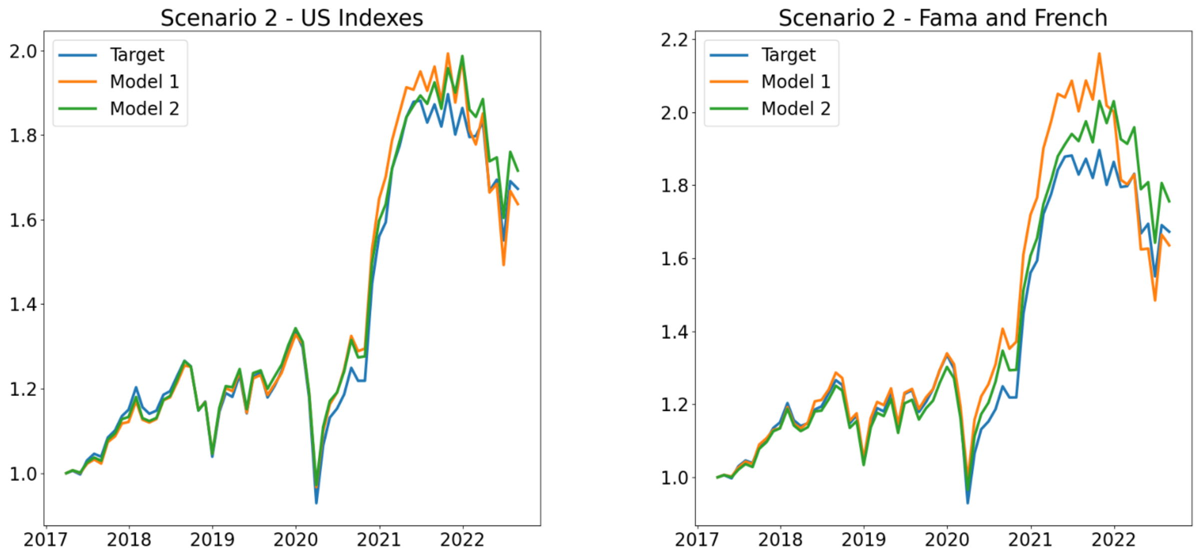


Figure 7.7: Scenario 1 replicating portfolios returns.

Type	WS	ER	CORR	SMAPE	TEV	N
US INDEXES *Model 1*	60	10.54%	94.24%	20.15%	7.27%	6.24
F. FRENCH *Model 1*	42	1.47%	92.18%	22.17%	8.12%	4.49
US INDEXES *Model 2*	-	15.26%	95.03%	18.02%	6.85%	3
F. FRENCH *Model 2*	-	17.59%	94.73%	16.75%	7.05%	4
AVERAGE	51	11.22%	94.20%	19.27%	7.32%	4.43

Table 7.3: Scenario 2 Results.

In this case holding the same portfolio better captures the stochasticity contained in the target weights than the dynamic replication: *Model 2* has better performance than *Model 1* and in both cases the best set of factors for the replica is US Indexes.

7.1.2. Lasso regression

As explained in section 2.1, Elastic Net used in the models is a regularization technique that combines Lasso regression and Ridge regression; the former tends to perform better than the latter when the dataset has many predictor variables, but only a few of them are important for predicting the response variable. *Model 1* and *Model 2*, using the set of factors of the US Indexes, build a replicating portfolio composed by on average 4.95 and 3 securities (Table 7.1): this subset is considerably lower than the complete set, composed by 14 factors. For this reason, to give more space to the capability of Lasso of performing feature selection and shrinking coefficients to zero, the Ridge component of the Elastic Net has been turned off. In Table 7.4 and Figure 7.8 the results of the replica can been seen. In the case of P4, Lasso regression performs better than the Elastic Net: comparing the corresponding model in Table 7.1 and Table 7.4 it can be noticed that each value of the metric in the Lasso replica has better values. Furthermore, the average number of portfolio components of *Model 1* has decreased.

Type	WS	ER	CORR	SMAPE	TEV	N
US INDEXES *Model 1*	54	0.10%	98.87%	14.28%	3.14%	4.2
US INDEXES *Model 2*	-	-0.74%	99.03%	14.31%	2.96%	3
AVERAGE	-	-0.32%	98.95%	14.30%	3.05%	3.6

Table 7.4: Lasso Results.

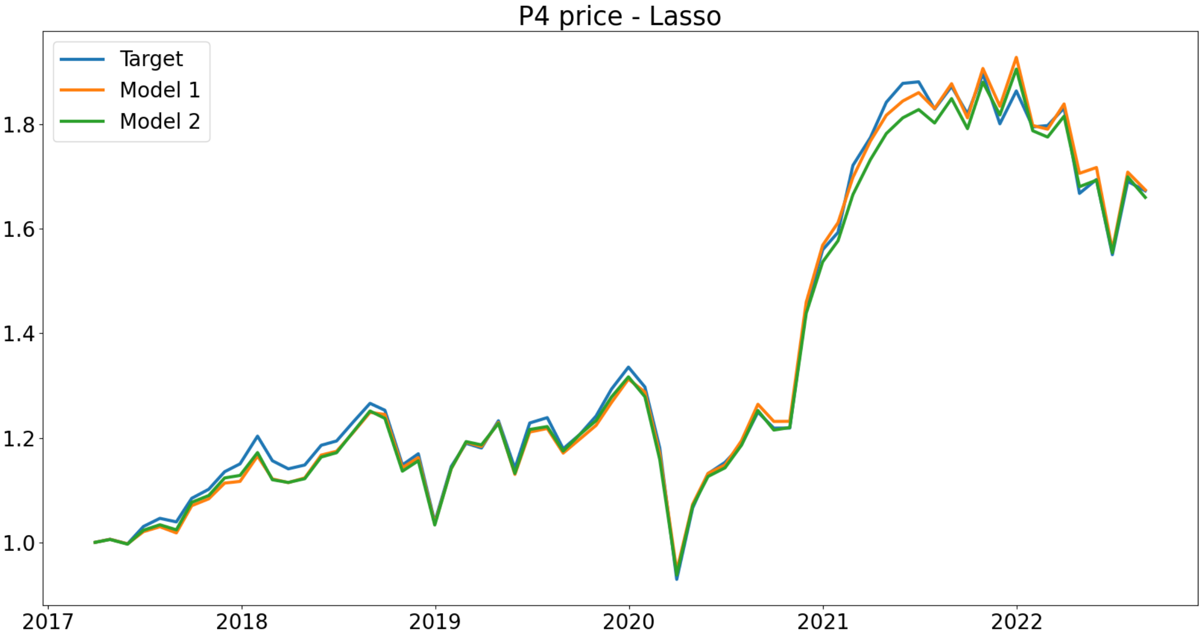


Figure 7.8: Price of the replicating portfolios built using Lasso.

This robustness test is not executed with the Fama and French set because they are only five and Lasso works better in a context of large sets of factors.

7.1.3. Gross exposure, Turnover and Transaction costs

Gross exposure, turnover and transaction costs are aspects, often ignored in academic studies, that for a practioner are really important. Portfolio managers sometimes have constraints imposed by the firm or by the regulation and they can not apply the best strategy obtained with a mathematical model. These three quantities are calculated for the replica portfolio of P4 to analyze potential problems and are shown in Figure 7.9, Table 7.5, Table 7.6 and Table 7.7.
Gross exposure is the total value of a portfolio, considering both long and short positions. In each time instant it is calculated by taking the sum of the absolute values of all positions held in a portfolio:

$$\text{Gross Exposure} = \sum_{i=1}^{n} |w_i|, \tag{7.2}$$

where:

- n is the total number of positions in the portfolio;
- w_i is the weight of the factor i.

Gross exposure is an important metric for portfolio managers because it helps them understand the total amount of risk their portfolio is exposed to.

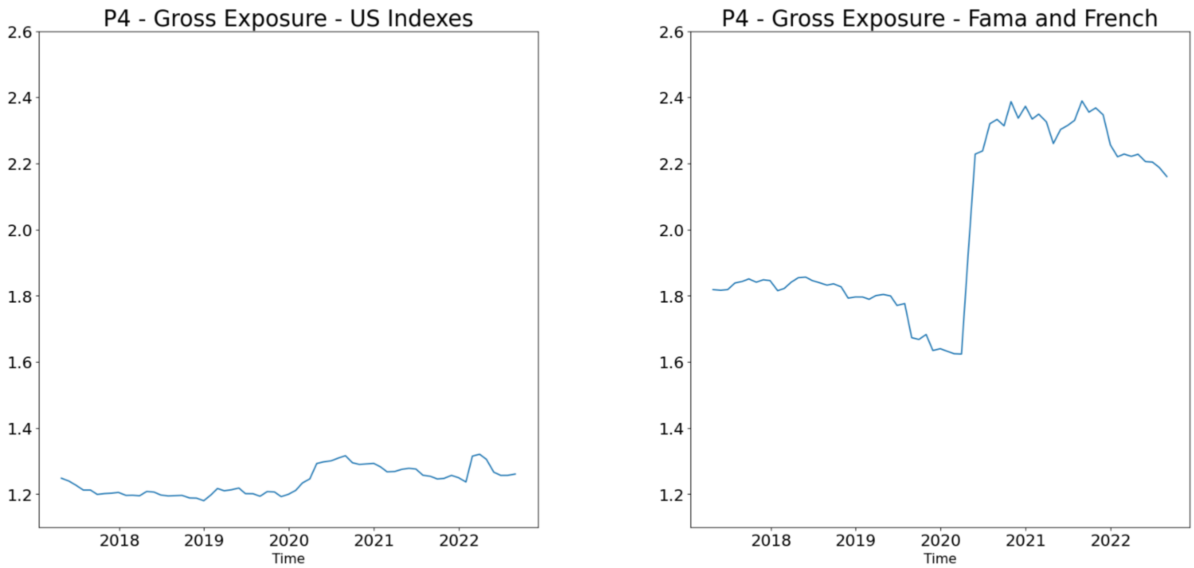


Figure 7.9: Gross Exposure of *Model 1* replica.

In Figure 7.9 the gross exposure of the portfolios obtained with *Model 1* and using the two different set of factors is shown. It can be noticed that it is always larger than 100%: this means that the replica portofolio uses *leverage*. Leverage refers to the procedure of borrowing money or using financial derivatives, such as futures or options, to gain exposure to a larger amount of assets than the amount of capital actually held by the portfolio. Leverage augments portfolio exposition: it can increase the potential returns, but also increases the potential risks. If the value of the underlying assets declines, portfolio manager may be required to provide additional collateral or liquidate positions to cover losses, which could further amplify them. The maximum leverage allowed for a portfolio depends on jurisdiction and on the type of financial institution that holds the portfolio. Gross exposure of the US Indexes replica is always under 140%; instead the Fama and French replica portfolio is highly leveraged, with the value of gross exposure that breaks the level of 200%.
Turnover is a measure of how much a factor is bought or sold, and in which quantity, during the life of a portfolio. It is calculated only in *Model 1* where the rebalancing is allowed and the results are shown in Table 7.5 and Table 7.6. For each factor the turnover is calculated with this formula:

$$\text{Turnover}_i = \sum_{t=1}^{T} |w_{i,t} - w_{i,t-1}|, \tag{7.3}$$

where:

- T is the number of periods;
- $w_{i,t}$ is the weight of the factor i at time t;
- $w_{i,t-1}$ is the weight of the factor i at time $t-1$.

Security	Turnover
RU20INTR	0.3593
XNDX	0.0017
I00189US	0.3761
M1US000V	0.5061
M1US000G	0.2110
LUACTRUU	0.0446
XMI	0.1571
HUI	0.0801
GSCI	0.1824
OEX	0.1760
SPTR	0.3537
LUATTRUU	0.6167
LF98TRUU	0.0098
VIX	0.0360
TOTAL	1.5553

Table 7.5: US Indexes turnovers. The TOTAL amount has been divided by 2 because the same percentage of turnover has been counted both in the factor bought and sold.

The total turnover of the Fama and French replica is equal to 244.64%, meaning that during the test set the factors are bought and sold for an amount of money that is more than double than the portfolio value. This percentage is higher than US Indexes case because of the portfolio adjustment that takes place on mid 2020; the jump can be observed in Figure 7.9.

Security	Turnover
Mkt - RF	0.5224
SMB	0.8408
HML	0.8961
RMW	1.2995
CMA	1.3340
TOTAL	2.4464

Table 7.6: Fama and French turnovers. The TOTAL amount has been divided by 2 because the same percentage of turnover has been counted both in the factor bought and sold.

Transactions costs can have a significant impact on portfolio replication strategies; brokerage fees, bid-ask spreads, and market impact costs, can erode the returns of a portfolio and affect the ability of the replication strategy to closely track the target. In particular, high transaction costs can make it difficult to replicate the turnover of the target portfolio, which can lead to differences in the factor exposures and returns of the replicated portfolio. In this study the transaction costs are considered both for buy and sell trades equal to 0.04% and are shown in Table 7.7.

Factors	Total Turnover	Trade Cost	Total Transaction Costs
US INDEXES	3.1106	0.0004	0.0012
FAMA AND FRENCH	4.8928	0.0004	0.0020

Table 7.7: Total transaction costs.

In this case the total turnover is double of the one showed in Table 7.5 and Table 7.6, because it is considered that when there is a change in the portfolio two trades take place, one for buy a factor and one for sell the other. The total transaction costs are the product between the total turnover and the trade cost; consequently, the replication strategy using Fama and French factors is more expensive than using US Indexes.

7.2. Applications

After the tests and analysis presented in section 7.1, we show what may be done with the replication portfolios obtained with *Model 1*. For example, they may be used to calculate the Value at Risk (VaR) and the Expected Shortfall (ES) of the position or to build the Ibbotson Cone [20] associated to the investment.

7.2.1. VaR and ES

Value at Risk (VaR) is a statistical measure that quantifies the potential loss in the value of a portfolio over a certain period of time, with a given level of confidence. VaR estimates the maximum amount of potential loss that a portfolio may suffer, under normal market conditions, within a specific time horizon and at a certain level of probability. As explained by Jorion [21], VaR is commonly used by risk managers to set risk limits and manage portfolio risk by ensuring that potential losses do not exceed a certain threshold. Several methods to compute VaR exist, for example: historical method, variance-covariance method and Monte Carlo method.
In this study the variance-covariance is used, also called *parametric* method. This method is based on the assumption that gains and losses, and so portfolio returns, are normally distributed and that are stationary. Normality of returns has been explained by Fama [8] and the stationarity has been proved in section 4.1. The formula to calculate the parametric VaR is:

$$VaR(\alpha)_t = \mu_t + \sigma_t * N^{-1}(\alpha), \tag{7.4}$$

where:

- α is the level of confidence;
- μ_t is the mean of the portfolio composed with the weights at time instant t;
- σ_t is the standard deviation of the portfolio composed with the weights at time instant t;
- $N^{-1}(\alpha)$ is the inverse of the cumulative distribution function of the standard normal distribution, representing the value of the Z-score corresponding to the desired level of confidence α.

VaR of the current replica portfolio is calculated each month of the test set and the level of confidence used are 95% and 99%; the results are shown in Table 7.8 and in Figure 7.10. For the US Indexes replica portofolio the month in which the VaR reaches its minimum is in 2018 an it is equal to 9.07%. This value is the maximum amount of potential loss for the following month in the 95% of the cases.

Factors	Min VaR 95%	Max VaR 95%	Min VaR 99%	Max VaR 99%
US INDEXES	9.07%	9.78%	12.60%	13.52%
F. FRENCH	8.55%	9.88%	11.89%	13.77%

Table 7.8: VaR of replicating portfolios built with *Model 1*.

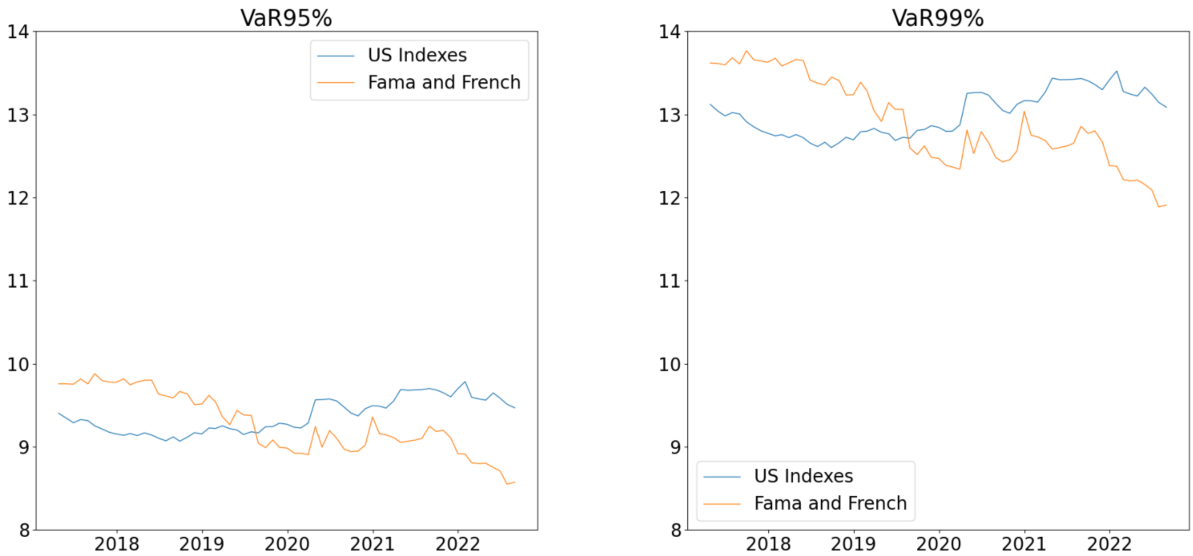


Figure 7.10: VaR of replicating portfolios built with *Model 1*.

VaR is useful because is widely used, easy to understand and provides a single number measure of risk. Nevertheless, it has some limitations. The biggest weakness of VaR is that it is not able to capture the tail risk: this means that VaR may underestimate the potential losses in extreme market conditions, with some disruptive events that could not be captured by the normal distribution. In Figure 7.11 it can be clearly noticed that the Covid pandemic, perfect example of an extreme event, breaks the threshold of both VaR95% and VaR99%.

VaR is usually accompanied by another risk measure: the Expected Shortfall (ES). ES is calculated by taking the average of the portfolio's losses that exceed the VaR threshold, weighted by the probability of those losses occurring. ES measures the average amount of loss expected if the portfolio experiences a loss greater than the VaR limit. In other words ES is a measure of the expected loss, beyond VaR, in the event of an extreme market event. The formula to calculate the parametric ES is:

$$ES(\alpha)_t = \frac{\mu_t + \sigma_t * N(z_\alpha)}{1 - \alpha} \tag{7.5}$$

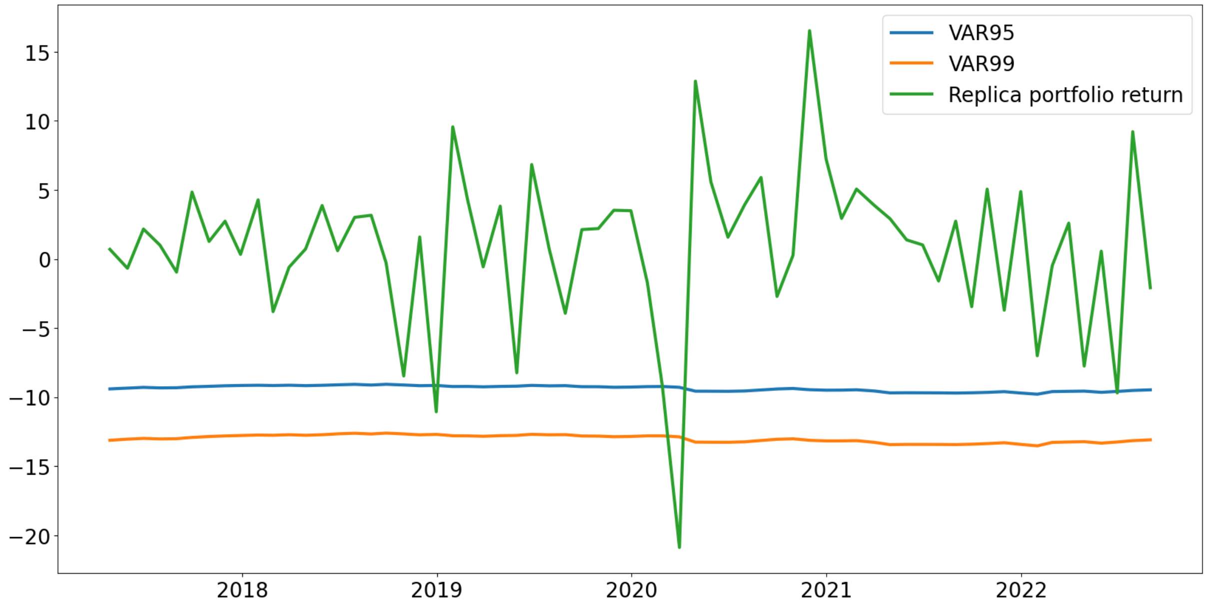


Figure 7.11: VaR test performed on replica portfolio built with *Model 1* using US Indexes.

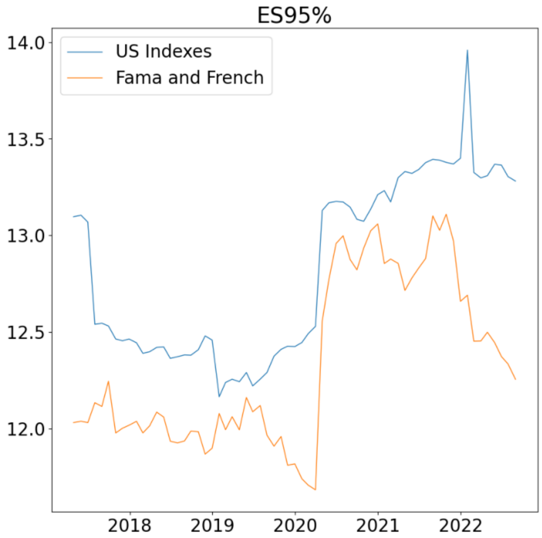
Factors	**Min ES 95%**	**Max ES 95%**	**Min ES 99%**	**Max ES 99%**
US INDEXES	12.16%	13.96%	15.64%	20.40%
F. FRENCH	11.68%	13.11%	14.43%	20.66%

Table 7.9: ES of replicating portfolios built with *Model 1*.

where:

- α is the level of confidence;
- μ_t is the mean of the portfolio composed with the weights at time instant t;
- σ_t is the standard deviation of the portfolio composed with the weights at time instant t;
- z_α is the critical value of the standard Normal distribution at the α percentile level.

The values of ES for P4 are shown in Table 7.9 and in Table 7.9.
By using both VaR and ES, a portfolio manager can better understand the potential risks of the portfolio and make more informed risk management decisions. While VaR provides a useful benchmark for measuring potential losses, ES provides a larger view of the portfolio's risk profile and helps to identify the tail risk of the portfolio.



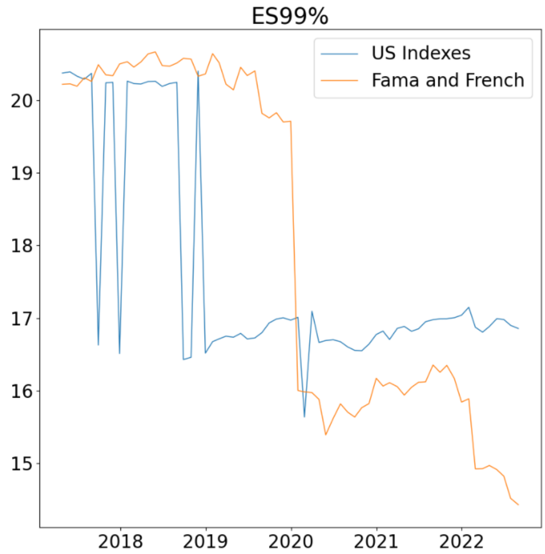


Figure 7.12: ES of replicating portfolios built with *Model 1*.

7.2.2. Ibbotson Cone

The Ibbotson Cone [20], or Volatility Cone, is a graphical representation of the expected return and risk of an investment portfolio over time. The cone makes it possible to describe, in probabilistic terms, the evolution of the portfolio by contextually representing in a single graph the evolution in a worst-case scenario, in a best-case scenario, and in the median scenario.
The Ibbotson Cone is characterized by two parameters:

- *confidence* level: it is the probability that the portfolio price evolution is under the values of the worst-case scenario;
- *protection* level: it is the probability that the portfolio price evolution is between the values of the best-case scenario and the worst-case scenario.

The cone of the replica portfolios built using *Model 1* and the two set of factors are shown in Figure 7.13 and Figure 7.14. They are built on the first month of the test set, namely May 2017; the levels chosen are 2.5% for the confidence level and 5.0% for the protection. The curves representing the best-case and worst-case scenario are calculated using the assumption of normality of the returns and the equations are the following:

$$WorstCase_t = \Delta t * \mu - \sqrt{\Delta t} * \sigma * N^{-1}(\alpha), \tag{7.6}$$

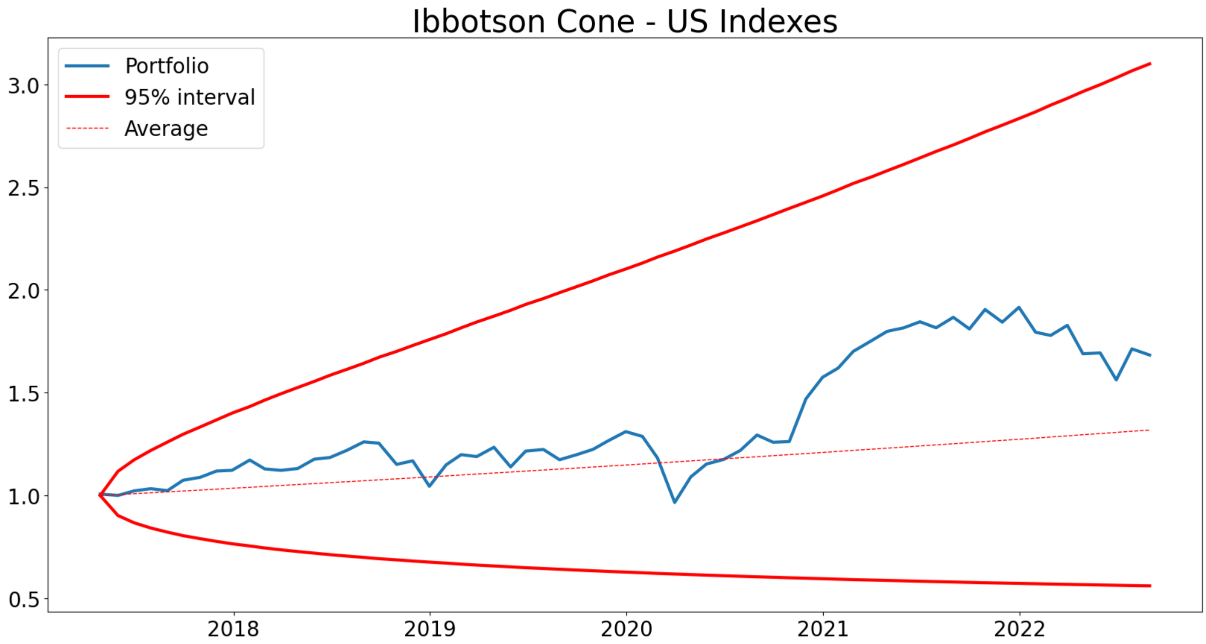


Figure 7.13: Ibbotson Cone associated to the price of the portfolio replicating P4 with US Indexes. The confidence level is equal to 2.5% and the protection level is equal to 5.0%

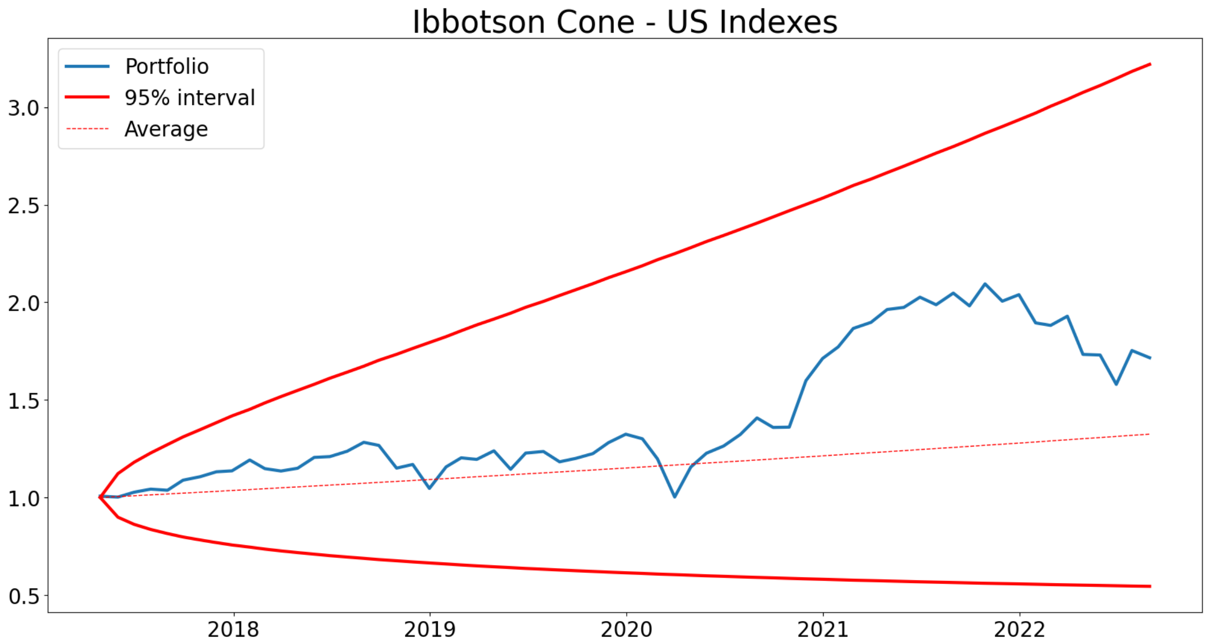


Figure 7.14: Ibbotson Cone associated to the price of the portfolio replicating P4 with Fama and French factors. The confidence level is equal to 2.5% and the protection level is equal to 5.0%

$$BestCase_t = \Delta t * \mu + \sqrt{\Delta t} * \sigma * N^{-1}(\alpha), \tag{7.7}$$

where:

- Δt is the time interval between May 2017 and t;
- μ is the mean of the portfolio, composed with the weights of May 2017, calculated considering data from January 2001 to May 2017;
- σ is the standard deviation of the portfolio, composed with the weights of May 2017, calculated considering data from January 2001 to May 2017;
- α is the percentage equal to $100\% -$ confidence level.

The curve representing the average-case is calculated considering only the first part of the equation:

$$AverageCase_t = \Delta t * \mu. \tag{7.8}$$

In the two figures the evolutions of the portfolio price, keeping the same weights of May 2017, are represented; it can be noticed that they are always contained in the 95% interval, also for example during the extreme event of the Covid 19 pandemic.

8 | Conclusions and future developments

In this study we have addressed the problem of replication of Black Box portfolios using two different models and two different set of factors. These models are based on a linear regression, developed with an Elastic Net to perform regularization. *Model 1* lives in a rolling window framework, with the replica portfolio that is rebalanced in each time period; *Model 2* is static, it does not allow rebalancing and the portfolio is kept constant. The replica is performed using two different set of factors, 14 US Indexes and the 5 Fama and French factors, and using 8 securities to build the target portfolios. In general, both models are able to build a replicating portfolio that has the same risk and return characteristics of the target portfolio using the two different datasets. They perform better with target portfolios which are diversified; indeed, they have encountered some issues in replicating the portfolio representing the Healthcare sector alone. Diversified portfolios, which has a lower level of idiosyncratic risk and are more exposed to the systematic one, are more predictable using the models; on the contrary, in the sectorial target portfolios, the idiosyncratic risk is difficult to be individuated and causes a decrease in the performance. The Elastic Net contained in the models is able to execute an efficient feature selection; it is more evident in the case of US Indexes, where the total number of securities in the replicating portfolios is very distant from the initial 14 factors. *Model 1* and *Model 2* have equivalent performance on the tested target portfolios. This does not mean that rebalancing has no value, but it is a consequence of the fact that the weights of target portfolios are constant.

Then we have chosen one target portfolio to perform a more detailed analysis and to show some possible applications with the replica. The choice is P4, because it is good compromise between diversification and sectorial. P4 is composed by three portofolios equally weighted and the result of its replica is not the best and not the worst among all the replication results. The same analysis may be performed in future with each one of the target portfolios. The first test performed is to transform his weights from deterministic to stochastic, adding a gaussian component. The simulated scenarios are two: *Scenario*

1 represents the situation of a portfolio manager who decides to make small adjustments to the weights frequently; *Scenario 2* represents the situation of a portfolio manager that keeps the weights constant for a longer period during the year, but when he decides to rebalance the portfolio is because he wants to change his strategy and consequently the adjustment is larger in value. For P4, there is evidence that *Model 1* is better than Model 2 in *Scenario 1*, instead *Model 2* is the best one in *Scenario 2*. In addition, in *Scenario 1* the overall performance of the models is slightly worse than the replica of the target portfolio with deterministic weights; on the contrary in *Scenario 2*, due to the adjustment that is larger, the performance is much worse.
Then, considering only the set of US Indexes, the Elastic Net is replaced by the Lasso regression, which performs better when the dataset has many predictor variables. The result is an increase of the performance and a reduction in the number of factors selected. The portfolio structure used in the models allows to do some analysis that for a practitioner could be really useful. After obtaining the replicating portfolio, it is not complex to calculate the gross exposure, the turnover and the transaction costs of the strategy. We have computed them for the replicating portfolios of P4, noticing that using the set of Fama and French factors in this case leads to an higher leverage and higher transaction costs. In the end, we have shown that, after that is appropriately calculated, the replicating portfolio has some useful applications. The applications considered are in the field of risk measurement and are: Value at Risk, Expected Shortfall and the Ibbotson Cone.
Possible further developments of this work are the choice of a non linear model in place of the Elastic Net or to extend the analysis, that for now have been computed only on P4, to each of the 13 target portfolios.

Bibliography

[1] V. Agarwal and N. Y. Naik. Risks and portfolio decisions involving hedge funds. *The Review of Financial Studies*, 17(1):63–98, 2004.

[2] N. Amenc, F. Goltz, and V. Le Sourd. Equity factor models. *Journal of Portfolio Management*, 39(3):67–78, 2013.

[3] J. S. Armstrong. *Principles of forecasting: a handbook for researchers and practitioners*, volume 30. Springer, 2001.

[4] J. Bergstra and Y. Bengio. Random search for hyper-parameter optimization. *Journal of machine learning research*, 13(2), 2012.

[5] V. DeMiguel, L. Garlappi, and R. Uppal. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The review of Financial studies*, 22(5): 1915–1953, 2009.

[6] D. A. Dickey and W. A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74 (366a):427–431, 1979.

[7] E. Fama and K. French. A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22, 2015.

[8] E. F. Fama. The behavior of stock-market prices. *The journal of Business*, 38(1): 34–105, 1965.

[9] E. F. Fama and K. R. French. Industry costs of equity. *Journal of financial economics*, 43(2):153–193, 1997.

[10] A. Frazzini and L. H. Pedersen. Betting against beta. *Journal of financial economics*, 111(1):1–25, 2014.

[11] K. French. Data library, 2022. URL `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

[12] W. Fung and D. A. Hsieh. The risk in hedge fund strategies: Theory and evidence from trend followers. *The review of financial studies*, 14(2):313–341, 2001.

[13] D. Giamouridis and S. Paterlini. Regular (ized) hedge fund clones. *Journal of Financial Research*, 33(3):223–247, 2010.

[14] R. C. Grinold and R. N. Kahn. Active portfolio management. 2000.

[15] J. Hasanhodzic and A. W. Lo. Can hedge-fund returns be replicated?: The linear case. *The Linear Case (August 16, 2006)*, 2006.

[16] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

[17] A. E. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

[18] R. W. Holthausen and D. F. Larcker. The prediction of stock returns using financial statement information. *Journal of accounting and economics*, 15(2-3):373–411, 1992.

[19] K. Hou, C. Xue, and L. Zhang. Replicating anomalies. *The Review of financial studies*, 33(5):2019–2133, 2020.

[20] R. G. Ibbotson and P. Chen. Long-run stock returns: Participating in the real economy. *Financial Analysts Journal*, 59(1):88–98, 2003.

[21] P. Jorion. *Value at risk: the new benchmark for managing financial risk.* The McGraw-Hill Companies, Inc., 2007.

[22] R. Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Montreal, Canada, 1995.

[23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

[24] P. C. Phillips and P. Perron. Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346, 1988.

[25] T. Roncalli. *Introduction to risk parity and budgeting.* CRC Press, 2013.

[26] W. F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442, 1964.

[27] W. F. Sharpe. Asset allocation: Management style and performance measurement. *Journal of portfolio Management*, 18(2):7–19, 1992.

[28] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

[29] S. D. Vrontos, I. D. Vrontos, and D. Giamouridis. Hedge fund pricing and model uncertainty. *Journal of Banking & Finance*, 32(5):741–753, 2008.

[30] C. J. Willmott and K. Matsuura. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate research*, 30(1):79–82, 2005.

[31] D. H. Wolpert and W. G. Macready. No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, 1(1):67–82, 1997.

[32] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2):301–320, 2005.

List of Figures

List of Tables

ACKNOWLEDGEMENTS

Acknowledgements

I would like to thank everyone I met during these years at Politecnico. Raffaele Zenti deserves a special mention: with his contribution to the Fintech course and his support in the realization of this thesis he has transmitted me a lot of knowledge. I would like to thank also Luigi Simeone for the precious advices and Professor Daniele Marazzina for reviewing the overall work.