CAHIER DE RECHERCHE DE LA CHAIRE FINTECH AMF – FINANCE MONTRÉAL

Equity market-neutral strategies using variable selection and regularized regression

Par Marzia A. Cremona
Charles-E Sarault
Federico Severino
Université Laval

Octobre 2024

Projet réalisé dans le cadre du 6ème appel de projets de la Chaire « Enjeux, risques et solutions : Consommateurs et investisseurs dans l'ère du numérique »

Equity market-neutral strategies using variable selection and regularized regression

Marzia A. Cremona[*] Charles-E. Sarault[†] Federico Severino[‡,§]

Working paper
October 31, 2024

Abstract

Equity market-neutral strategies are designed to feature no exposure to market risk. The implementation of such strategies relies upon the estimation of the beta coefficients. Unfortunately, traditional beta estimation methods used to implement these strategies often suffer from weak out-of-sample performance, leading to suboptimal ex-post neutrality. Therefore, we explore how machine learning techniques, particularly variable selection and regularized regression methods, can address the issues faced in the traditional development of equity market-neutral strategies. We test a range of methods, including ridge regression, lasso regression, and stepwise regressions, to construct portfolios that achieve better ex-post neutrality and lower transaction costs when compared to strategies based on traditional multiple regression models. The results demonstrate that the tested techniques enhance portfolio performance and help minimize trading costs by selecting an optimal number of risk factors to hedge. Our research contributes to the academic literature on machine learning in asset pricing and offers practical insights for portfolio management.

JEL Classification: G11, C13.
Keywords: Market-neutral portfolio · Machine learning · Penalized regression · Factor selection

[*]Université Laval, Department of Operations and Decision Systems, marzia.cremona@fsa.ulaval.ca.

[†]Université Laval, Faculty of Business Administration, charles-edouard.sarault.1@ulaval.ca.

[‡]🖂Université Laval, Department of Finance, Insurance and Real Estate, federico.severino@fsa.ulaval.ca.

[§]F. Severino and M.A Cremona acknowledge the financial support of Chaire Fintech AMF-Finance Montréal.

1 Introduction

Equity market-neutral strategies (EMNS) are built to eliminate exposure to overall market fluctuations. Their return is uncorrelated to market movements, and a suitable choice of long/short positions in some stocks can make them profitable (see, e.g. Wu et al., 2021). These strategies have always been at the core of hedge fund management due to the nature and scope of such investing vehicles. Today, they are essential for the portfolio management of retail investors as well (Feghali and Villalon, 2015).

Effectively, EMNS are designed to hedge against market risk factors, aiming to protect investors from market turmoil. A portfolio which incorporates EMNS seeks to generate a high expected return while keeping the overall volatility low (Feghali and Villalon, 2015). Moreover, paired with appropriate stock selection, these strategies theoretically allow investors to obtain pure returns independent of market directions. An appropriate selection would be to take long positions in underrated stocks and short positions in overvalued stocks. Although EMNS are interesting in theory, Patton (2009) and Muhtaseb and Colborn (2012) document that, in practice, EMNS may fail to reach their neutrality goal. The implementation of such strategies stems from the estimation of beta coefficients that quantify the stock exposure to market risk factors. Their estimation employs historical data at the time the strategy is designed, but market-neutrality is expected at the end of the investment horizon, creating a time mismatch. The weak out-of-sample performance of beta estimation through standard statistical tools can negatively affect the ex-post neutrality of the strategy, particularly when market factors suffer from multicollinearity.

Because EMNS account for an important fraction of portfolio allocations, validating their claim of neutrality is compelling to properly quantify the effectiveness of such methods. Considering the studies made in Patton (2009) and Muhtaseb and Colborn (2012), there exists evidence that the traditional methods employed to develop EMNS present some shortcomings that must be addressed to generate more robust strategies.

Consequently, we seek to contribute by developing methods that produce more stable beta estimates while addressing multicollinearity issues that prevent traditional EMNS from reaching neutrality. By testing modern machine learning techniques, we expect to uncover cost-effective solutions that achieve improved ex-post neutrality. Our project is relevant for both practitioners and academicians. The focus of the hedge fund industry on EMNS is unchanged. Moreover, it constitutes a contribution to the academic literature on machine learning for asset pricing and creates a potential pedagogical value.

The principal question we seek to answer is: can recent machine learning tools improve the ex-post neutrality of EMNS? Indeed, a panoply of approaches – such as penalized regression models and variable selection methods – have been proposed in the machine learning literature to improve the beta estimation in multiple-factor models. In general, these methods aim to obtain better out-of-sample performance by reducing the variance of the estimated beta at the price of introducing a bias in the estimation. Therefore, we set our focus on evaluating the effectiveness of these machine learning methods in improving ex-post market-neutrality. We also address the trade-off between neutrality and related strategy transaction costs. It is expected that penalized regression techniques, such as ridge regression, will produce more stable beta estimates, resulting in EMNS with ex-post neutrality closer to zero. Additionally, variable selection methods, such as stepwise regression, are expected to generate parsimonious models. These reduced models should achieve improved neutrality at a lower cost by minimizing unnecessary factors while addressing concerns of multicollinearity.

2 Literature review

Given these challenges, a closer review of the literature on the effectiveness and limitations of EMNS is pertinent at this stage. In 2009, Patton mentioned that market-neutrality involved 20% of funds under management of hedge funds, referring to (Fung and Hsieh, 1999, Table 4). Despite the popularity of such investment techniques, Patton (2009) questions the market-neutrality of hedge funds. Several definitions of risk neutrality are used, among which beta (or correlation) neutrality is used with respect to some market index or several risk factors. Astonishingly, one-quarter of the market-neutral funds under scrutiny feature significant non-neutrality. The concern of non-neutrality is further analyzed in Muhtaseb and Colborn (2012), which documents a non-null exposure to the systematic risk of some equity market-neutral hedge funds around the Great Financial Crisis of 2008. A major contribution to the previously mentioned shortcomings stems from the complexity involved in the estimation of beta coefficients. In fact, when multiple risk factors are considered, risk neutrality heavily depends on the goodness of the estimation of the beta coefficients in the model. The use of a high number of factors, as it is standard in the portfolio literature (Harvey and Liu, 2019), raises the issue of multicollinearity, which generally increases the variance of the estimated betas. Moreover, the bias-variance trade-off can negatively affect the out-of-sample performance of the beta estimators (Hastie et al., 2009, Chapter 2). This problem is not negligible in the design of EMNS because portfolio weights are chosen ex-ante. Even when the portfolio rebalancing is frequent, the out-of-sample performance of the estimated betas on future data plays an important role. These issues can make the goal of a zero-beta strategy unattainable.

Our research belongs to the booming literature on applications of machine learning in asset pricing. A general reference in the area is Gu et al. (2020), who exploit a variety of machine learning techniques to estimate security returns. Other examples include Simonian and Wu (2019), which uses ridge regression for hedge-fund replication, and Wu et al. (2021), which applies reinforcement learning to equity market-neutral portfolio management.

3 Methodology

This section outlines the methodology followed to address our goal of discovering more effective ways to generate EMNS using supervised machine learning techniques. In this section, we cover data selection, the general approach applied across all tested methods, and a detailed description of each specific machine learning technique and its differences in application.

3.1 Data

To achieve the goal of generating EMNS, two preponderant components are required. First, it is essential to select multi-factor models toward which the portfolio will be made neutral at inception. The factors selected must be sufficiently numerous so that the reported multicollinearity problems reproached to the traditional implementation of EMNS emerge. Once we select a group of factors, a stock universe of uncorrelated assets must be formed. Indeed, the selected universe should be comprised of uncorrelated stocks. This enhances diversification, reduces idiosyncratic risk and improves the effectiveness of the EMNS. Additionally, it minimizes the risk of multicollinearity in the factor model, facilitating access to improved ex-post neutrality. The second important consideration is the choice of the period to cover in our analysis. We chose a horizon that spans from January 2003 to January 2019. In doing so, we include recent data and limit the number of financial crises included. This way, our study will focus on contemporary times and will mostly be concerned with what can be seen as a normal economic period. EMNS's performances are strongly dependent on the period selected to derive the weight of their implementation. Thus, by having a sufficiently large dataset, we will be able to implement the strategies at many randomly selected dates. Then, it will be possible to trace a general portrait of which methods seem to work best across all tested periods. Therefore, even if it is known that daily data generally show more variability, it is such data that we will use.

As is made evident in Harvey and Liu (2019), risk factors used in recent top academic journals are numerous, so numerous they call it a “factor zoo.” As of 2019, this “zoo” was reported to comprise more than 400 risk factors. Instead of getting lost in the “zoo” by developing our own multi-factor model, we selected two existing models.

The first model we chose is a five-factor model developed in Fama and French (2015). The factors considered are Mkt.Rf (Excess return on the market), SMB (Small Minus Big), HML (High Minus Low), RMW (Robust Minus Weak), CMA

(Conservative Minus Aggressive). This model is an updated version of the well-known three-factor model of the same authors. These models are interesting because they extend the traditional Capital Asset Pricing Model by including additional risk factors beyond market exposure. Starting off our analysis using a more concise factor model is appropriate because it is less computationally intensive to work with. Also, it allows us to rapidly develop the general framework to be applied to the larger factor model we test. We will refer to this model as the 5FF.

The second model we chose is the result of the work made in Jensen et al. (2023). With their work, they derived themed portfolios of capped value-weighted factors. The themed portfolios were obtained by forming clusters of factors that demonstrate a high degree of economic and statistical similarity. They started off with a total of 153 factors, which they condensed into thirteen themes. The themed factors consist of Accruals, Debt Issuance, Investment, Low Leverage, Low Risk, Momentum, Profit Growth, Profitability, Quality, Seasonality, Size, Short-Term Reversal and Value. This multi-factor model is pertinent because it presents a satisfying number of factors to give rise to multicollinearity. We expect that the presence of correlation across the risk factors will be better delt with by machine learning techniques than traditional methods. We will refer to this model as 13GF.

We show the correlations among the risk factors during the full sample of data we consider for the 5-factor model and the 13-factor model in Fig. 1 and Fig. 2 respectively.

As for the universe of stocks that will be used to generate the neutral portfolios, we chose the most uncorrelated stocks from the main US exchanges. Because the multi-factor models selected pertain to risks in US markets, it is evident that stocks from the same markets be used. The exchanges considered are AMEX, NASDAQ and NYSE. From the CRSP database, we obtained return data for every stock traded on the mentioned exchanges during the horizon of interest. The screening process begins by removing any stocks that do not have return data for the complete period. That is, any stocks having less than 4779 days of return data from January 2003 to January 2019 are taken out of the universe. While analyzing return data, we observed that many stocks had the same return for many consecutive days. An unchanged level for a stock price is possible, but such a circumstance rarely occurs. Indeed, it is almost impossible for a stock price level to remain the same, and such a phenomenon is suspicious. Thus, we removed any stock that had more than 5% of its 4779 returns equal to 0. In other words, we removed any stocks whose price remained the same for two consecutive periods more than 5% of the time. The final step, before selecting the least correlated stocks, was to remove any stocks with less than the median mean daily volume. By removing stocks with low mean daily volume, only liquid and tradable stocks are considered, ensuring that the portfolios could have been set up in practice. Then, the final step in composing the stock universe was to perform a correlation analysis, allowing to identify the least correlated assets. The correlation analysis is

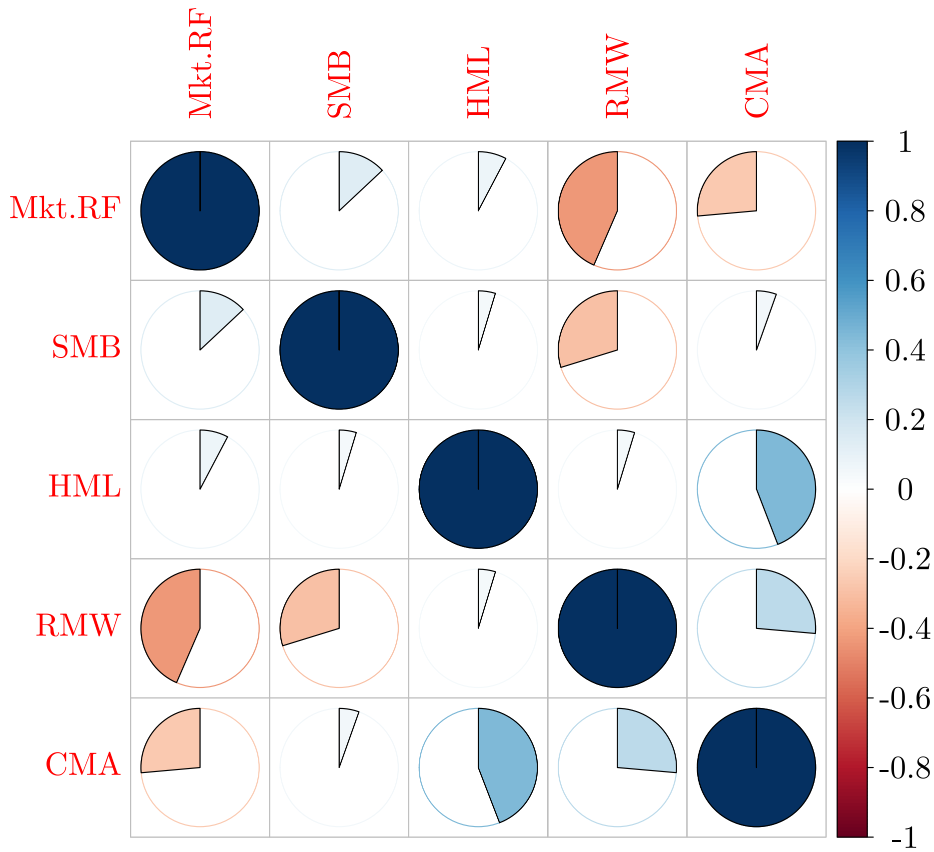

Figure 1: Correlations of 5FF during the whole sample of interest.

performed by considering a matrix of all the stocks' returns for the complete period. Each column is the full set of returns for each stock. Subsequently, the absolute values of pair-wise correlations are considered. If two stocks have a high correlation, we use a function that looks at the mean absolute correlation of each stock and removes the stock with the largest mean absolute correlation (Kuhn, 2007) The same pair-wise correlation analysis is repeated until the number of stocks needed in the final universe is reached. Respectively, six and fourteen stocks are needed for the 5FF and 13GF models. Once the pair-wise correlation analysis is done, we are left with the least correlated stocks. As is common practice in many of the consulted papers, the stock returns are transformed into excess returns using the risk-free rate included in the Fama-French Data. This way, the returns are standardized and adjusted for risk.

For both models we are using, the risk factors' returns can be obtained through each of the authors' websites.

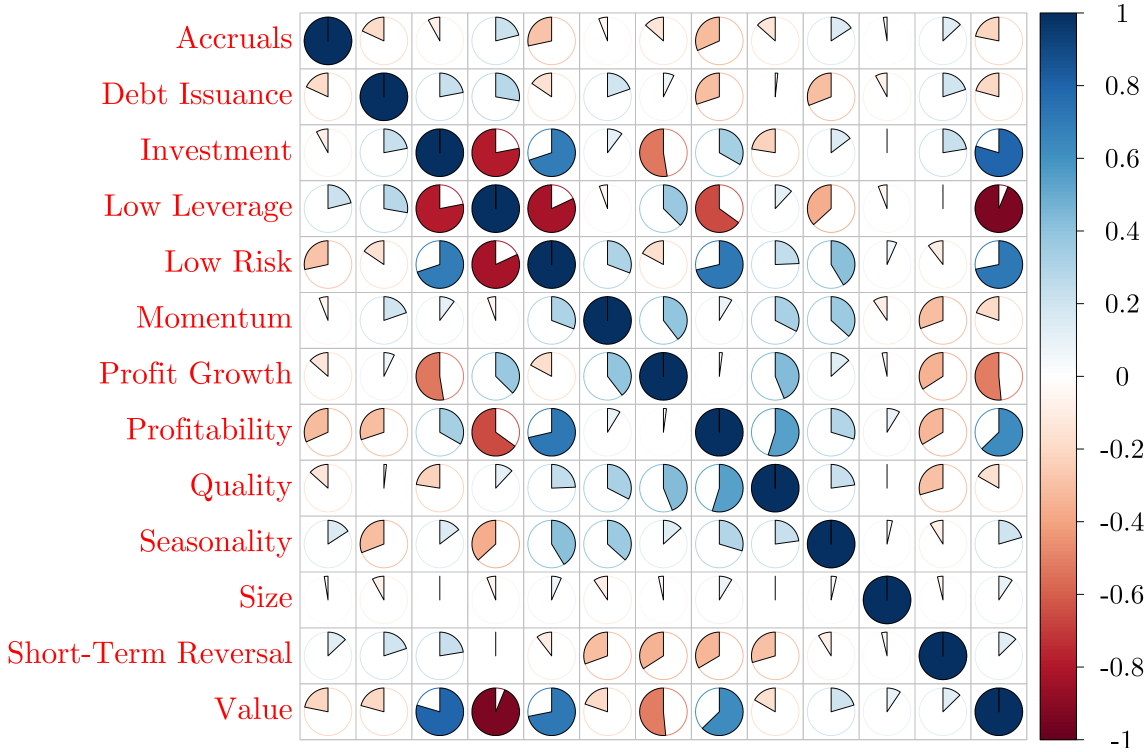

Figure 2: Correlations of 13GF during the whole sample of interest.

3.2 General approach

To address the goal of our research, we outlined a general procedure. This involves analyzing the quality of various machine learning techniques in generating EMNS that are effectively neutral ex-post and offer reduced implementation costs. In its simplest form, generating EMNS is done by building a hedged portfolio through the selection of the right proportions of long and short positions in a matching number of assets. Although the process can appear straightforward, many topics must be covered. The objects common to all tested methods are comprised of the general process of portfolio construction, the division of the training dataset versus the test dataset, the definition of a pertinent and adequate neutrality metric and the cost model considered.

3.3 Portfolio construction

In a multiple-factor model, EMNS are portfolios of stocks with zero beta with respect to each factor. Such strategies are used, for instance, by Shyu et al. (2006) in a framework in which multicollinearity is not a major issue. If n factors are present in the model, $n + 1$ stocks are considered to build the market-neutral portfolio. Given the linearity of the portfolio betas, then zero-beta conditions with respect to each factor lead to a linear system. By solving it, the EMNS portfolio weights in closed form are obtained. If the portfolio weights are computed at time t, they are functions of the stock beta coefficients estimated from the return data until t using a multiple regression. Then, the market-neutral portfolio is held until a future time T. From the portfolio returns in the time window [t, T], we can estimate the "ex-post" portfolio betas with respect to the risk factors using another multiple regression model and measure neutrality by calculating the distance to neutrality which will be explained in Subsection 3.5.

To better understand how we obtain the strategies' weights, we will first give the definitions of the notations we use in Table 1.

Definition	Notation
Assets	$k = 1, 2, 3, ...$
Returns of assets at time t	$y_{k,t}$
Factors	$n = 1, 2, ... = k - 1$
Returns of factors at time t	$x_{n,t}$
Beta of assets	$\beta_{k,n}$
Return of portfolios as a linear combination of factor returns	$y_{k,t} = \beta_{k,0} + \beta_{k,1}x_{1,t} + \beta_{k,2}x_{2,t} + ... + \beta_{k,n}x_{n,t} + \varepsilon_{k,t}$
Weights of assets in portfolios	w_k
Neutral sum of betas	$w_1\beta_{1,n} + w_2\beta_{2,n} + w_3\beta_{3,n} = 0$

Table 1: Definition of notation.

The linear system solved to obtain the EMNS weights is shown in eq. (1). Using the matrix of coefficients and the neutrality vector on the left-hand side, we can solve for the neutral strategies' weights.

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \beta_{1,1} & \beta_{2,1} & \cdots & \beta_{p,1} \\ \beta_{1,2} & \beta_{2,2} & \cdots & \beta_{p,2} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{1,n} & \beta_{2,n} & \cdots & \beta_{p,n} \end{bmatrix} \times \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_{p-1} \\ w_p \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \tag{1}$$

Each line of coefficients in the right-hand side matrix of eq. (1) could have been individually calculated. Indeed, it would have been possible to run separate linear regressions to explain each stock by every risk factor. However, doing so would have been problematic when obtaining coefficients for models subject to variable selection. Instead, we used the fact that multiple regression can be expressed as a system of linear equations and that the coefficients can be obtained simultaneously. To do so, we used dummy variables to indicate when to include individual stocks in the strategies developed. That way, we were able to effectively measure coefficients in the case of variable selection. Working this way has implications for how lasso regression and elastic net regressions were handled. To better understand how we obtained the beta coefficients simultaneously, we now give a simple case of a portfolio comprised of two risk factors and three assets. One of our response variables (stock returns) will be chosen as the "baseline". For the "baseline", no dummy variable will be defined. Therefore, we will consider two additional predictors, which are associated with the two other response variables not chosen as "baseline". The predictors we consider are two risk factors and two dummy variables for two of the three response variables, note that we also include the intercept. The following equations detail each predictor's equation and the decomposed system of equations used to obtain the coefficients.

$$
\begin{aligned}
\text{Asset 1 (“baseline”):}\quad & y_{1,t} = \beta_{1,0} + \beta_{1,1}x_{1,t} + \beta_{1,2}x_{2,t} + \varepsilon_{1,t} \\
\text{Asset 2:}\quad & y_{2,t} = (\beta_{1,0} + \beta_{2,0}) + (\beta_{1,1} + \beta_{2,1})x_{1,t} + (\beta_{1,2} + \beta_{2,2})x_{2,t} + \varepsilon_{2,t} \\
\text{Asset 3:}\quad & y_{3,t} = (\beta_{1,0} + \beta_{3,0}) + (\beta_{1,1} + \beta_{3,1})x_{1,t} + (\beta_{1,2} + \beta_{3,2})x_{2,t} + \varepsilon_{3,t}.
\end{aligned}
$$

Collect all $y_{1,t}, y_{2,t}, y_{3,t}$ in a single vector $\mathbf{y}_t' = \begin{bmatrix} y_{1,t} & y_{2,t} & y_{3,t} \end{bmatrix}'$ and fit the regression:

$$
\begin{aligned}
y_t &= \beta_{1,0} + \beta_{1,1}x_{1,t} + \beta_{1,2}x_{2,t} \\
&\quad + (\beta_{2,0} + \beta_{2,1}x_{1,t} + \beta_{2,2}x_{2,t})\mathbb{I}_{\{y_t = y_{2,t}\}} \\
&\quad + (\beta_{3,0} + \beta_{3,1}x_{1,t} + \beta_{3,2}x_{2,t})\mathbb{I}_{\{y_t = y_{3,t}\}} + \varepsilon_t \\
&= \beta_{1,0} + \beta_{2,0}\mathbb{I}_{\{y_t = y_{2,t}\}} + \beta_{3,0}\mathbb{I}_{\{y_t = y_{3,t}\}} \\
&\quad + (\beta_{1,1} + \beta_{2,1}\mathbb{I}_{\{y_t = y_{2,t}\}} + \beta_{3,1}\mathbb{I}_{\{y_t = y_{3,t}\}})x_{1,t} \\
&\quad + (\beta_{1,2} + \beta_{2,2}\mathbb{I}_{\{y_t = y_{2,t}\}} + \beta_{3,2}\mathbb{I}_{\{y_t = y_{3,t}\}})x_{2,t} + \varepsilon_t.
\end{aligned}
$$

3.4 Dataset division

To achieve the performance comparison of the techniques we tested, we had to define how much data to use at each step. Firstly, we will describe the narrowest window of data we use. This narrow window is the basic building block we will need. Secondly, we will explain how we use the narrow window to perform cross-validation to train our models.

3.4.1 Narrow window

We chose to use 504 days (two years of trading days) of data to derive the strategies' weights. Once the strategies are in place, we then measure the distances to neutrality over the following 63 days (fiscal quarter). We use 504 days of return data to derive the weights so that the strategies are built upon sufficient historical data. The length of the period over which we measure neutrality must be consequent with the length of the period used to obtain the EMNS weights. We found that measuring neutrality over 63 days was a reasonable choice. In doing so, the portfolios do not require too frequent rebalancing, and their implementation length is reasonable when compared to the data needed to implement them. Using 504 return data to obtain the weights and then measuring neutrality over the following 63 days is the narrowest window we consider. For this case, the first 504 days of data are the training period where the models are adjusted. The following 63 measures of neutrality across 63 days can be viewed as the testing period of this narrow window. The narrow window is shown graphically in Fig. 3.

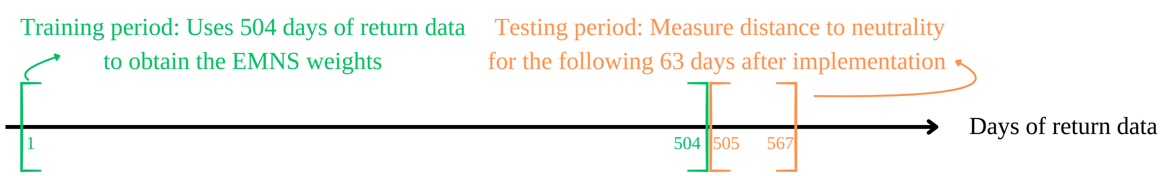

Figure 3: Narrow window.

3.4.2 Cross-validation

Now that the narrow window is well-defined, we can move to cross-validation. For this part, using 504 return to obtain the weights and then measuring neutrality over the following 63 days is what we will call a training subperiod. Indeed, the machine learning techniques we study need to be trained to "tune" their respective parameters. Penalized regression models are trained to tune their shrinkage parameters. On the other hand, stepwise regression models must be trained over multiple periods to uncover the right factors to be included in the final model to be tested. To address this "tuning" of parameters, we perform a type of cross-validation by repeating what is done in the narrow window. We will perform our "cross-validation" in a training period, which is composed of 14 narrow windows. In each of the 14 training subperiods, an EMNS is developed and held constant over 63 trading days. The median distance to neutrality of the models is analyzed over the training period to select the parameters to be used in the test period. The test period is equivalent to any of the 14 subperiods in terms of the quantity of data used. In the test period, the strategies set up represent the best-performing strategies during training. During the training period, each machine learning techniques were set up using various parameters. Following the training period, the distances to neutrality of the various strategies were analyzed, and the specific parameters leading to the best performances were recorded. After tuning each technique using the best parameters, we test the various models in a final period. Some of the return data used in the test period are also used in the last subperiod of training. The amount of shared data is equivalent to the data shared between any of the training subperiods. Each subperiod uses 567 days of return. In subperiod 1, we use data from day 1 to day 567, in subperiod 2, we use data from day 64 to day 630, and so on. Each subperiod incorporates 63 new days of data and shares 504 values. An illustration of the dataset division we used to perform the cross-validation is shown in Fig. 4.

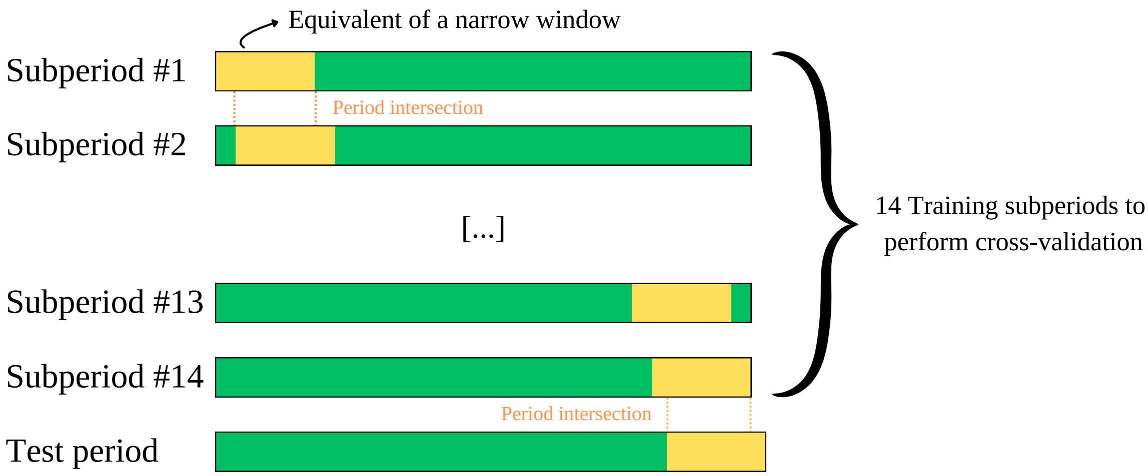

Figure 4: Cross-validation across the 14 training subperiods.

Using daily data, we are measuring neutrality over three months from the investment date. To avoid non-stationarities in the data, we choose a training set which is not excessively long. However, we do not want it to be too short because we consider up to 13 factors whose beta coefficients must be estimated, together with the tuning parameters. Since we use a cross-validation approach, we consider a training set of 882 days, divided into 14 subsamples of 63 days.

The following breakdown reiterates how we form the neutral portfolios through the dataset division. To start, we fix the starting date of the training period and use the first 504 days of stock returns to derive a matrix of beta coefficients. We use this matrix to set up the system of equations in eq. (1) to obtain the EMNS weights. For each of the following 63 days, the same process is repeated, where a matrix of coefficients is obtained using 504 days of return data. After each passing day, we obtain a new matrix of coefficients, including one new day of return and excluding the very first day of return used in the previous day's matrix. We use each of these

63 matrices alongside the initial weight vector to compute a daily neutrality vector. The process is repeated across the 14 training subperiod of 63 days. The neutrality analysis across the 14 training subperiods serves in the tuning of the final models to be used during the test period. As it was mentioned, this "cross-validation" analysis is essential because the techniques tested are dependent on the tuning of their respective parameters. The parameters we select are chosen because they produce the models with the lowest median distance to neutrality across all the 14 training subperiods. Selecting the model based on the lowest median distance ensures robustness. Indeed, it is possible that when methods are implemented, the matrix of coefficients used to derive the weights will be numerically unstable. This instability leads to weights that produce inflated distances to neutrality during the 63-day implementation of the EMNS. Such cases of numerical instability should be excluded from the tuning process. Such cases of numerical instability should be excluded from tuning decisions, as our goal is to find models that produce distances closest to 0.

3.5 Neutrality metric

At inception, we build portfolios to present no exposure toward the risk factors at play. Therefore, at first, neutrality is valued at 0. Then, after each passing day, we compute a new matrix of coefficients alongside the inception weights to measure the new neutrality of the portfolio. In eq. (1), the neutrality vector is seen on the left-hand side of the system. After inception, it won't necessarily be a null vector because the portfolio is moving away from zero neutrality. True neutrality being 0, it is then natural to measure the distance between the new neutrality vectors and the initial one. Therefore, the neutrality metric we consider will be the Euclidian distance between each passing day's new "neutrality" vector and the null vector used at inception to obtain the implementation weights.

$$\text{Euclidean distance to neutrality: } d\left(\mathbf{0}, \mathbf{N}_{t_i}\right) = \sqrt{\sum_{i=1}^{n} \left(\mathbf{0} - \mathbf{N}_{t_i}\right)^2} \tag{2}$$

We will now briefly analyze Fig. 5, a plot of the distances to neutrality of an EMNS obtained using multiple regression. We can observe on the plot that at initiation, the strategy is at a distance of 0. Since the strategies are set up to hedge against the risk factors completely, it is normal that the beginning distance be 0. After each day, the distance moves away from 0. Because the best strategy would present a distance to neutrality of 0, we believe that using the Euclidean distance as our neutrality metric was a pertinent choice.

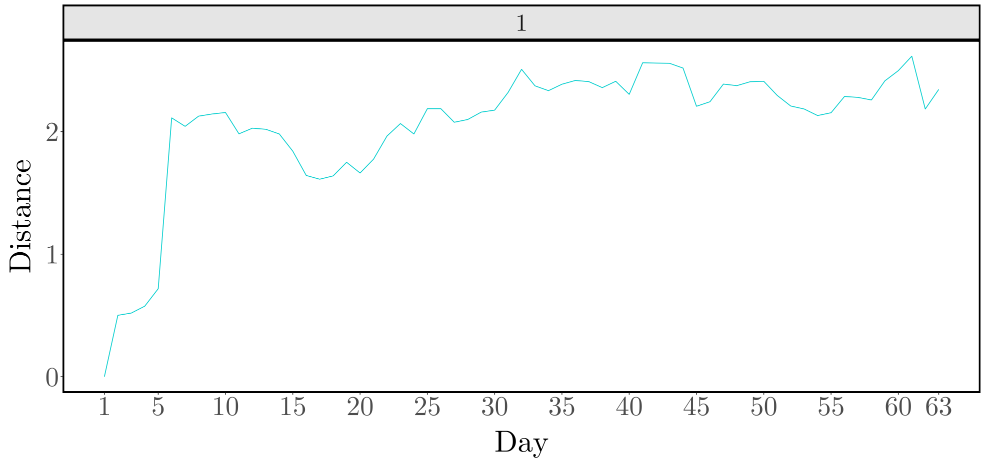

Figure 5: Distance to neutrality of an EMNS obtained using multiple regression during training subperiod 1 EMNS hedging against the 13GF.

3.6 Cost model

For simplicity, in the first phase of the analysis, we consider a simple cost model. To contrast the implementation cost of all neutral portfolios stemming from the different machine learning techniques we use, we consider the absolute value sum of the strategy weights. The cost model we propose will make comparison sufficiently easy. Indeed, a strategy with a larger absolute sum of weights will be considered a costlier strategy to implement, whereas a smaller absolute sum of weights will be considered a less costly strategy to implement.

Regardless of the ease of using both the neutrality metric and cost model proposed, there is still a subjective question of what constitutes an acceptable trade-off between neutrality and cost. A return analysis of the most effective machine learning methods in achieving the lowest neutrality or cost could be pertinent to propose an acceptable trade-off.

4 Linear regression and machine learning methods

The foundation model for which we will evaluate the efficiency of implementing an EMNS is a multiple regression model. This regression-based model will serve as our benchmark. We will proceed to compare all the other strategies, obtained using the other techniques, with this benchmark. We obtain the results of this technique by following the procedure described earlier in Subsection 3.4. With this method, coefficients are obtained by solving the ordinary least squares (OLS) problem. This method does not need any parameter tuning. Although we obtained the results for this benchmark technique across the 14 training subperiods, it is only done for comparison reasons.

4.1 Subset selection

The first group of methods we test are subset selection methods. The general idea behind these methods is that only a subset of the variables is kept in the model. These methods are aimed to address two issues reproached to multiple regression based on OLS estimates (Hastie et al., 2009, Chapter 3.3). Specifically, predictive accuracy and interpretation are improved using these methods. As for other techniques we test, that also include variable selection, strategies hedging against a reduced number of factors will be hedged through an equivalently reduced number of stocks. Even if the strategy does not hedge against all factors at inception, we measure the distance to neutrality toward every factor of the full. In doing so, we can compare the reduced models with the benchmark, and the interpretation remains the same. The subset techniques we test correspond to best subset selection, forward selection and backward selection, two stepwise regression techniques.

4.1.1 Best subset selection

The first subset method we use is best subset selection. The general idea of this method is to test all possible combinations of response variables (risk factors) to include. In doing so, this process allows us to derive the best subset selection. By removing the empty set from the result of the power set formula (2^n, n is the number of risk factors), we can calculate the number of combinations of factors we will evaluate. In the 5FF case, there are 31 different combinations, and in the 13GF case, there are 8191 different combinations to be tested. It is evident that this method becomes rapidly computationally intensive as the number of factors gets larger. One of the first efficient algorithms of this computationally intensive method was proposed in Furnival and Wilson (1974). Ultimately, with this method, we test all possible combinations of risk factors and select the right combination leading to the lowest median distance to neutrality across all 14 training subperiods.

4.1.2 Stepwise regression

The following two methods we test are also part of the stepwise regression family. Namely, these methods are forward selection and backward selection. The first uses of these methods are reported in Efroymson (1960) and Hocking (1976). The main idea is to add or remove factors one at a time based on a decision criterion. For this method, we use the same decision criterion based on the median distance to neutrality across all training subperiods. To better understand how we use these methods, we will explain the process for forward selection. We start from an empty strategy only, including the regression intercept. Then, we consider all single-factor strategies that each include the individual factors of the full-factor model. Then, we measure each of the single-factor strategies' distances to neutrality across all 14 training subperiods. The strategy leading to the lowest median distance is now the "base strategy", to which, one by one, we add the remaining factors. At each step, we select the new "base strategy" according to the lowest median criterion. Once we have taken every step and the "base strategy" is the one that includes all factors, we compare the neutrality of the various strategies. We use this final comparison step to select, amongst all the strategies including a different number of factors, the one with the lowest median distance. The process works the same for backward selection but in the opposite direction, where factors are removed from the full model. Because of the order in which factors are added or removed, both techniques might select different factors in their models of equivalent sizes, leading to different "best" models.

4.2 Shrinkage methods

In contrast to subset selection methods, which use OLS to fit a linear model by selecting a subset of predictors, shrinking methods offer an alternative approach by fitting a model that applies a constraint to the coefficient estimates. This process, also known as regularization, shrinks the coefficient estimates towards zero. Although it may not seem intuitive at first, imposing such a constraint can significantly reduce the variance of the estimates, thus improving EMNS performance in many cases (James et al., 2014, Chapter 6). Thus, by minimizing the variance of the estimators in the model, shrinkage methods aim to address issues of multicollinearity when dealing with models including multiple factors (Kılıçoğlu and Yerlikaya-Özkurt, 2024). This family of methods presents advantages over OLS regressions because of how they tackle the bias-variance trade-off. Also, they are less computationally intensive than best subset selection (James et al., 2014, Chapter 6).

The shrinkage methods we consider are ridge regression, lasso regression and elastic net regression. All these techniques require some parameter tuning. Indeed, all methods need the selection of a shrinkage parameter λ. An additional parameter α is needed for elastic net regression to fix the proportion accorded to the two penalizations applied to its' coefficients.

4.2.1 Ridge regression

Ridge regression was first introduced in Hoerl and Kennard (1970). With this technique, the coefficients minimize a penalized residual sum of squares and are shrunk toward 0 and each other (Hastie et al., 2009, Chapter 3.4.1). Coefficients are obtained in a process which is almost equivalent to how they are obtained in OLS. The penalization term λ is applied to the sum of the squared coefficient. This constraint's domain can be illustrated by a circle in the case of a model with only two factors. Therefore, the constraint leads to shrunk coefficients but does not produce variable selection.

The tuning of λ will be discussed in the next subsection since the tuning process is the same for both ridge regression and lasso regression.

4.2.2 Lasso regression

Even if ridge regression presents some advantages over the benchmark method, it does not select variables to generate reduced strategies as it were the case for subset methods. Indeed, ridge regression addresses issues of multicollinearity but generally does not lead to less expansive EMNS. The lasso regression handles both variable selection and coefficient penalization. The term lasso regression was introduced in 1996 by Robert Tibshirani in Tibshirani (1996). Once again, the coefficients are obtained in a similar manner to those in OLS. The minimization problem solved to obtain the coefficient is subject to the absolute value sum of coefficients. The domain of this constraint can be illustrated by a square in the case of a model with only two factors. This domain makes it evident that some coefficients might be completely shrunk to 0, thus removing some factors completely.

The tuning process we use to determine the value of the shrinkage parameter λ goes in the same direction as what we mentioned earlier in Subsection 3.4.2. We obtain the optimal λ value the same way for both ridge and lasso regression. We select the optimal λ value by testing various values of λ across all training subperiods. We do this by training various strategies using different λ values. Then, we calculate the median distance to neutrality for each strategy. The parameter value we select is the one leading to the lowest median distance across all 14 training subperiods. To reduce computation time, we select a relatively small grid of 135 λ values. For comparison reasons, we test the same parameter values for all three shrinkage techniques. We chose the λ values while considering that when they are too large in either lasso regression or elastic net regression, the penalization leads to very few factors being selected and sometimes none. Therefore, the grid we use does not include values too large to avoid empty models. As the value of λ gets closer to 0, the coefficient becomes less and less subject to shrinking. Accordingly, for very small values of λ, both OLS and shrinking methods produce the same coefficients. For simplicity, the parameter values we use were obtained by taking the 10^{th} power of an equally spaced sequence of values starting at -2.6667 to -30.

As we mentioned earlier, we use dummy variables in the system of equations needed to obtain beta coefficients. Because we used dummy variables, we had to use lasso regression and elastic net regression variants able to handle group selection. Therefore, to remove at once the dummy variables and their associated stock returns from the simultaneous computation of coefficients, we use group lasso and group elastic net regression. Such grouping versions of these shrinkage methods were introduced in Yuan and Lin (2006).

4.2.3 Elastic net regression

This method is a hybrid of both ridge regression and lasso regression, applying both techniques' penalization constraints. Elastic net regression incorporates both advantages of the two previous techniques. Indeed, by applying both the minimization constraint of ridge regression and lasso regression, this method can select variables and shrink together the coefficients of correlated predictors (Hastie et al., 2009, Chapter 3.4.3). The penalization included in this technique was introduced in Zou and Hastie (2005). The two parameters used for this method are the shrinkage parameter λ and α. They respectively measure what proportion of penalty from ridge regression and lasso regression is applied to its coefficients.

In addition to the 135 values of λ we test in the training phase, we also test ten values of α. In other words, for each of the ten values of α, we test 135 values of λ. We perform the tuning by computing which λ value produces the lowest median distance to neutrality for each α. Then, we compare the medians of the ten remaining strategies and choose the strategy with the α value leading to the lowest median distance.

In contrast to the subset methods we test, where we tune the strategies to select which risk factors to retain, in lasso regression and elastic net regression, the tuning only indicates the parameters of the final model. Thus, we cannot know in advance how many factors will be included in the strategies we test. Certainly, the best parameters will generate differently performing strategies depending on the dataset used.

5 Results

The performance of an EMNS is highly dependent on the beginning period of data used to generate it. Changing a single day of data in the set used to measure the EMNS's weights can modify the strategy and the neutrality we then measure. Therefore, it is evident that to be able to stipulate which of the techniques we test offers the best performances, we need to test randomly selected starting dates. To set the general format of the analyses we performed, we will first focus on the results we obtained by creating EMNS using the very first period of data as the starting point. To begin, we will present which method offered the best performance during the training periods regarding the lowest median distance to neutrality. In addition, we will show how each of the tuned strategies performed in the test period regarding both neutrality and implementation cost.

5.1 Training period

Following the procedure detailed in Subsection 3.4.2, we calculate the median distance to neutrality for each of the 14 training subperiods to tune our EMNS. We will now analyze the performance of the two multi-factor models across the training period.

5.1.1 5-factor model

Method	Median distance	Average number of factors
Multiple regression	0.594	5.00
Best subset selection	0.594	5.00
Forward selection	0.594	5.00
Backward selection	0.594	5.00
Ridge regression	0.588	5.00
Lasso regression	0.580	4.93
Elastic net regression	0.574	4.57

Table 2: Summary of metrics for the EMNS hedging against the 5FF during the test period.

Over the first set of training periods, the tuning of best subset selection, forward selection, and backward selection led to the same results as the benchmark. Indeed, to minimize the median distance to neutrality, all the subset methods selected each of the five factors to be included in the neutral strategies. The three best performances were obtained using the penalized regression techniques. The best performance in terms of neutrality was achieved using elastic net regression. The performance of lasso regression was better than ridge regression in this case. It comes as no surprise that the α parameter we found through training for elastic net regression leans toward more lasso regression regularization than ridge regression with $\alpha = 0.95$. Also, the training process led to the selection of $\lambda = 0.0004641589$. Such a value of λ is at a size where variable selection is observed, but still, many factors are kept in the neutral strategies on average. The average number of factors using this technique is also the lowest. As we already explained, the tuning of the EMNS in the case of either lasso regression or elastic net regression leads to the selection of parameters but not the selection of risk factors. The variable selection made by this best-performing strategy was only observed in the last four training periods. Indeed, the EMNS dropped the CMA factor in periods 11 and 12 and then additionally dropped RMW in periods 13 and 14 to finally be composed of Mkt.RF + SMB + HML. The number of factors that will be included during the test period might differ. As an example, during the test period, elastic net regression remains a three-factor strategy, but lasso regression switches from four to three factors.

The graphs in Fig. 6 illustrate the distance to neutrality for all training subperiods of 63 days following the implementation of the EMNS for the benchmark and elastic net regression. They represent the cross-validation done in the 14 training subperiods. There is little to no difference before the 11th period, where one of the factors is left out of the portfolio for the first time. When there is no selection of variable, the strategies present a distance to neutrality of 0 at inception. However, it is not the case when some variables are left out. Neutrality is measured toward each of the factors included in the full model, but the strategy was made neutral only toward a subgroup of factors. Therefore, the EMNS using a reduced number of factors does not outperform full models at inception, as it is possible to observe in subgraphs 11, 12 and 14 of Fig. 6. The reduced strategy's lower neutrality is observed after an average of seven days in this case. In the periods where a selection of variables occurs, the benchmark distance curves present more variability than the reduced model distance curves, which are mostly flat and stable.

5.1.2 13-factor model

Method	Median distance	Average number of factors
Multiple regression	3.582	13.00
Best subset selection	2.481	3.00
Forward selection	2.894	3.00
Backward selection	3.026	1.00
Ridge regression	3.582	13.00
Lasso regression	3.048	1.00
Elastic net regression	2.935	12.64

Table 3: Summary of metrics for the EMNS hedging against the 13GF during the test period.

Regarding the EMNS we developed while hedging against the 13GF, most techniques led to better neutrality performance than the benchmark. The only method resulting in an equivalent performance to the benchmark was obtained using ridge regression. All other machine learning techniques led to strategies presenting variable selection. In most cases, three factors or less were kept, and a lower neutrality was observed when compared with the benchmark. The power of variable selection is clearer here than in the case of the 5-factor model, where the improvement in neutrality was only subtle. Here, the improvement in the median distance to neutrality ranges from 0.55 for lasso regression to 1.1 for best subset selection. This improvement is more substantial than what was achieved with elastic net regression for the 5FF model, where the distance to neutrality was only 0.02 lower than the benchmark. The technique leading to the best-performing EMNS is best subset selection. The

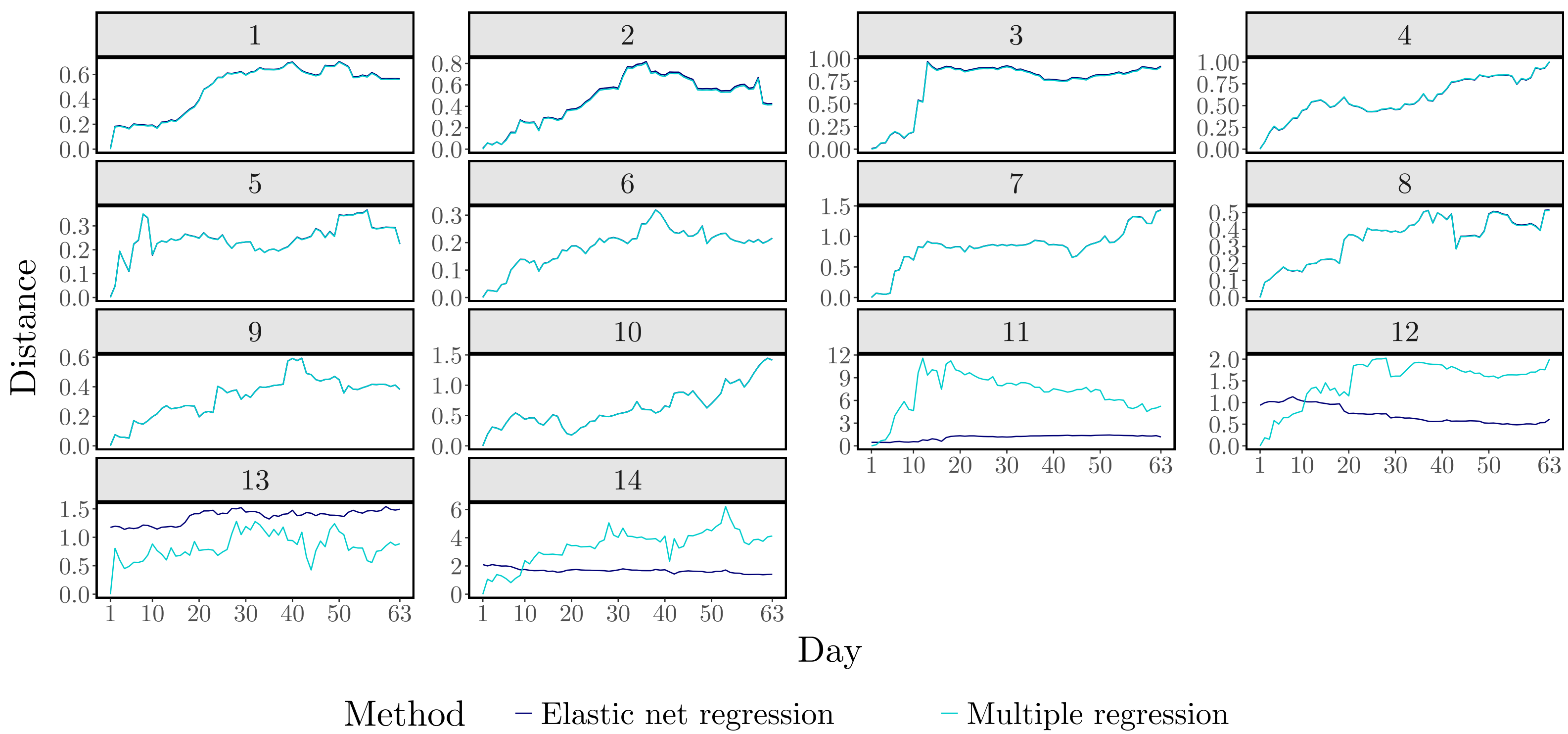

Figure 6: Comparison of the EMNS for the 5-factor model obtained using multiple regression and elastic net regression during the training period.

improvement is achieved by keeping only three factors out of the 13 factors. The strategy's portfolio consisted of investment + profitability + quality. Although lasso regression and backward selection did not lead to the lowest median distances, it is interesting to note that these strategies each hedge using only one factor during all training subperiods. The factors selected were low risk and value for lasso regression and backward selection, respectively. This result shows that an EMNS composed of only one factor offers a better performance over this period than a strategy hedging against all 13 factors at inception. To further understand this phenomenon, it is essential to analyze the results of applying the training process for different randomly selected inception dates.

The graphs in Fig. 7 illustrate the distance to neutrality for all training subperiods of 63 days following the implementation of the EMNS for the benchmark approach and best subset selection. By observing these graphs, we can more easily justify why the tuning of the parameters is based on the lowest median distance to neutrality. Although it was also visible in the graphs of the 5-factor model using elastic net regression, it is even clearer in this case. Indeed, periods 11 and 13 show very large distances for the benchmark. This is explained by the numerically unstable matrix of coefficients used in obtaining the implementation weights. If we had used an alternate selection criterion, such as the average distance to neutrality, very large distances would have greatly impacted the averages. Therefore, by using the median distance, the cases presenting large distances do not hold any weight in the tuning of parameters. By doing so, only the strategy giving rise to the lowest neutrality 50% of the time is selected. Even if the best subset selection strategy's initial matrix of coefficients presents numerical instability in period 14, the median distance remains the lowest of all the tested methods. Here again, the distance curves are more stable for the reduced model than the benchmark.

5.2 Test period

After tuning the models during the training phase and identifying the best parameters for each method, we proceeded to test and analyze their performance. In the following section, we explore the performance of all the techniques in the test period and discuss the implementation cost of the different strategies.

5.2.1 5-factor model

Curves of the distance to neutrality for all techniques during the test period for the 5-factor model are shown in Fig. 8. By observing the distance curves shown in Fig. 8, the best-performing technique is once again obtained through elastic net regression. As for the neutral portfolios of the last two training periods, this strategy is composed of Mkt.RF + SMB + HML. Not only does this model offer improvements in terms of

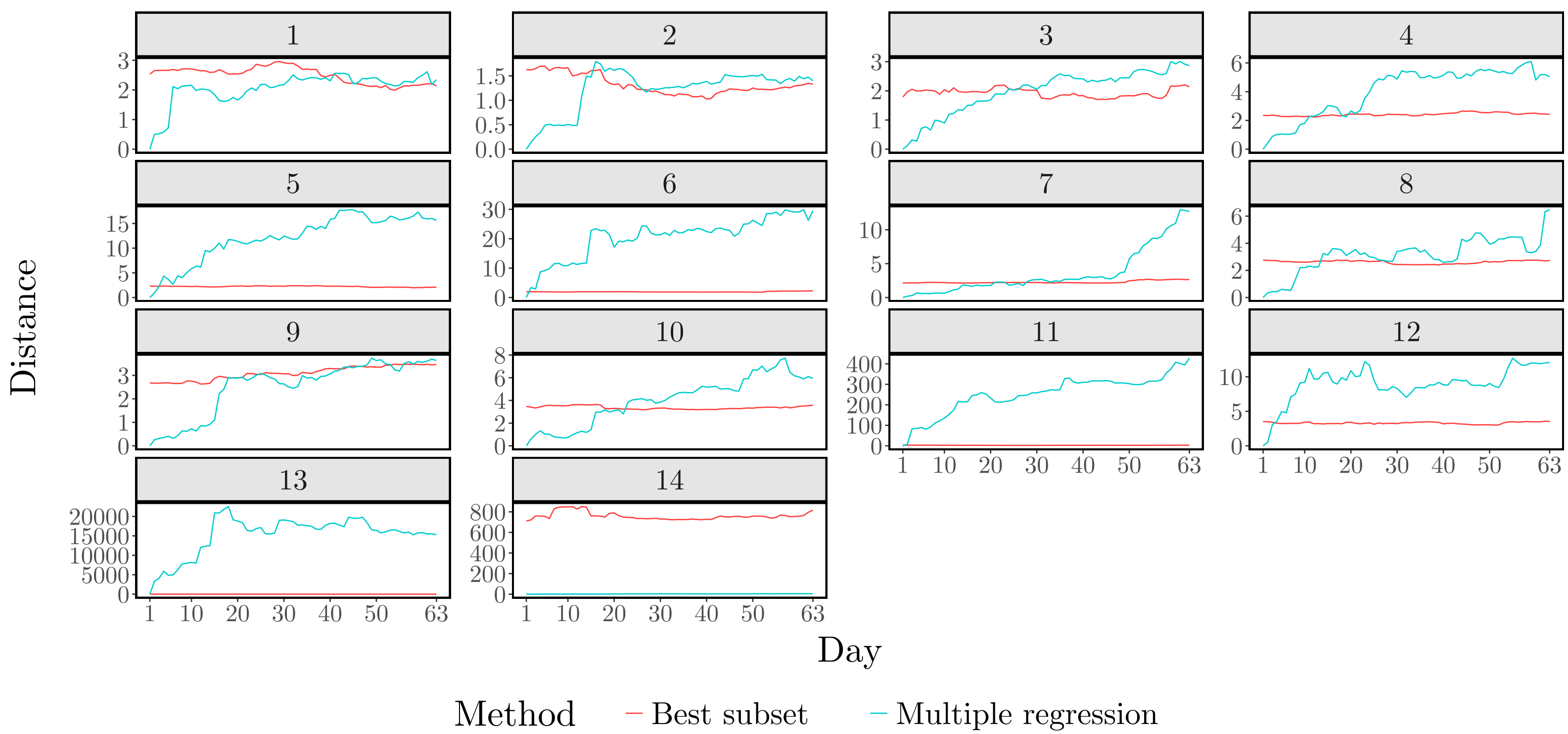

Figure 7: Comparison of the EMNS for the 13-factor model obtained using multiple regression and best subset selection during the training period.

distance to neutrality, but it also improves in terms of implementation cost. Indeed, using our cost measure, the benchmark is implemented through an absolute sum of weights of 28.42, whereas elastic net regression has a "cost" of only 2.76.

Even though ridge regression uses all five factors, it presents a greater distance to neutrality than the benchmark (linear regression). This is explained by the fact that through cross-validation, we selected $\lambda = 0.001668101$ as the parameter generating the lowest median distance. For EMNS developed using ridge regression to offer the same performance as linear regression, a sufficiently small value of λ must be used. Although the λ offered the lowest median distance during training, there is no assurance that it will still outperform in the testing period.

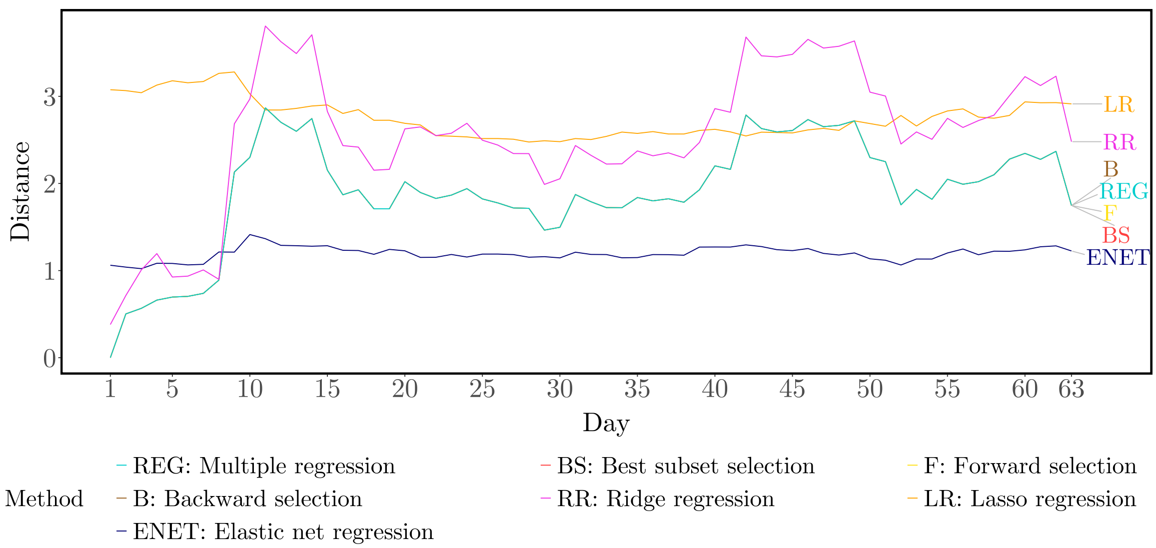

Figure 8: Comparison of the EMNS for the 5-factor model using each technique during the test period.

5.2.2 13-factor model

Curves of the distance to neutrality for all techniques during the test period for the 13-factor model are shown in Fig. 9. In the test period, the best-performing technique in terms of neutrality is not best subset selection as during training, but instead, it is backward selection. Although best subset selection does not continue to offer better neutrality performance in the test period, it does offer improvement in terms of implementation cost. The benchmark has a "cost" of 7.45, backward selection has a "cost" of 1.37, and best subset selection has a "cost" of 1.35, which is a slight improvement.

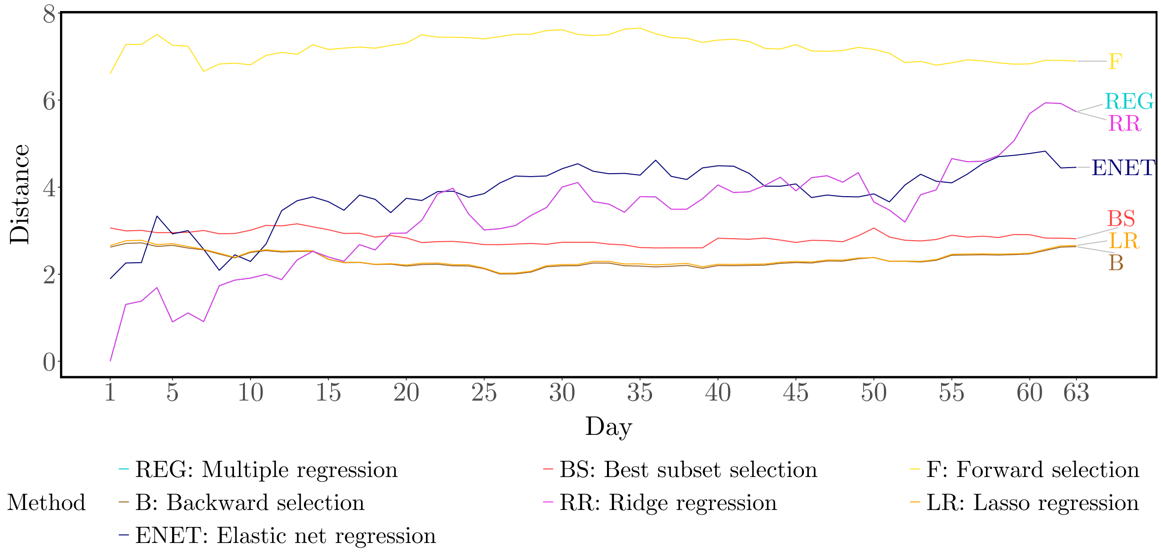

Figure 9: Comparison of the EMNS for the 13-factor model using each technique during the test period.

6 Discussion and limitations

The results we detailed up to now are limited to a single period. We must obtain results by developing portfolios at various starting dates to state which machine learning methods generate the best EMNS.

Generally, any problem we faced was solved as they appeared. Although we have been able to attend to most issues, some limitations are still impacting our work. Discussing these limitations is of central interest. By identifying how our work can be deepened, it will be easier to generate EMNS that are even better. The limitations we will address concern the tuning of parameters, problems of unstable matrices of coefficients and defining an acceptable trade-off between cost and neutrality.

6.1 Parameter tuning

The number of parameters tested was a limitation in the analysis of penalized regression. To reduce computation time, we selected a non-exhaustive list of λ and α values to be tested. Because our goal is to repeat the analysis detailed through Subection 3.4.2 using multiple starting dates, we hope to identify the best parameters for each technique. By testing a more exhaustive list of parameters, we could possibly reach a finer tuning. Testing more parameter values is even more important for lasso regression and elastic net regression. If the grid of parameters is too shallow, models of every length won't necessarily be obtained by these techniques that perform variable selection. To offer a more robust answer to the question of transaction cost, testing a wider range of parameters would be ideal.

6.2 Starting date

When studying Fig. 9, it is possible to see that the linear regression reached excessive distance to neutrality in periods 11 and 13. The simple reason behind these large distances is explained by numerical instability. At inception, the portfolios are made neutral by solving a system of equations similar to eq. (1). If the matrix of coefficients used in this calculation presents numerical instability, the weights obtained will be erroneous and lead to absurd distances to neutrality after inception. The presence of numerical instability is at the heart of our decision to use median distance as our decision criterion. By repeating the work detailed in Subsection 3.4.2, we will be able to select training periods where numerical instability is not a major problem. Uniquely considering these periods will offer better ground for the comparison of machine learning techniques.

6.3 Subjective trade-off between cost and neutrality

To better address the question of acceptable trade-off between cost and neutrality, it would be interesting to consider various investor profiles. For investor wanting to preserve their wealth and protect their assets from any fluctuation, the focus would be set toward neutrality over cost. As for diversified funds, attaining neutrality might require more active management, and a balance between cost and neutrality will be set to mitigate the cost of such management.

7 Conclusion

To conclude, with regard to the results obtained using the very first period of 504 days to derive the various strategies, we can already affirm that improvement over traditional EMNS is possible. Indeed, in both multi-factor models we tested, the training phase led to the development of EMNS that outperformed our benchmark, which represented the traditional method of implementation for EMNS. In the case of the 5-factor model, the best-performing strategy in terms of neutrality remained the same during the test period and even presented significantly lower implementation costs versus the benchmark. In the case of the 13-factor model, although the best-performing strategy during the training periods did not offer the best performance during the test period, it still offered the cheapest implementation cost. To crown any techniques as the best method to generate EMNS, we would have to study which techniques lead to better performances across varying starting dates.

References

Efroymson, M. A. (1960). Multiple regression analysis. *Mathematical Methods for Digital Computers*, 1.

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

Feghali, G. and Villalon, D. (2015). Building a better equity market neutral strategy [white paper]. AQR. https://www.aqr.com/Research-Archive/Research/White-Papers/Building-a-Better-Equity-Market-Neutral-Strategy.

Fung, W. and Hsieh, D. A. (1999). A primer on hedge funds. *Journal of Empirical Finance*, 6(3):309–331.

Furnival, G. M. and Wilson, R. W. (1974). Regressions by leaps and bounds. *Technometrics*, 16(4):499–511.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273.

Harvey, C. R. and Liu, Y. (2019). A census of the factor zoo. *Capital Markets: Asset Pricing & Valuation eJournal*.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.

Hocking, R. R. (1976). A biometrics invited paper: The analysis and selection of variables in linear regression. *Biometrics*, 32(1):1–49.

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

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2014). *An Introduction to Statistical Learning: with Applications in R*. Springer.

Jensen, T. I., Kelly, B., and Pedersen, L. H. (2023). Is there a replication crisis in finance? *The Journal of Finance*, 78(5):2465–2518.

Kılıçoğlu, Ş. and Yerlikaya-Özkurt, F. (2024). A novel comparison of shrinkage methods based on multi criteria decision making in case of multicollinearity. *Journal of Industrial and Management Optimization*, 20.

Kuhn, M. (2007). Caret: Classification and regression training. https://CRAN.R-project.org/package=caret, Accessed: 31-10-2024.

Muhtaseb, M. and Colborn, R. (2012). Neutrality of equity market neutral strategy to period and index. *Journal of Derivatives and Hedge Funds*, 18:301–332.

Patton, A. (2009). Are "market neutral" hedge funds really market neutral? *The Review of Financial Studies*, 22(7):2295–2330.

Shyu, S.-d., Jeng, Y., Ton, W., Lee, K.-j., and Chuang, H. (2006). Taiwan multi-factor model construction: Equity market neutral strategies application. *Managerial Finance*, 32(11):915–947.

Simonian, J. and Wu, C. (2019). Factors in time: Fine-tuning hedge fund replication. *The Journal of Portfolio Management*, 45(3):159–164.

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

Wu, M.-E., Syu, J.-H., Lin, J. C.-W., and Ho, J.-M. (2021). Portfolio management system in equity market neutral using reinforcement learning. *Applied Intelligence*, 51(11):8119–8131.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2):301–320.