Erasmus University Rotterdam

Extreme learning: on the importance of volatility in empirical asset ranking.

Erasmus School of Economics
Master Thesis Quantitative Finance
Robeco Super Quant Internship

Author:	Reinier Vos	Student ID:	583868
Supervisor:	Dr. Prof. D.J.C. Van Dijk	Second assessor:	Dr. S.O.R. Lonn
Robeco:	M. Jansen	Robeco:	I. Honarvar Gheysary
Robeco:	K. Ūsaitė	Date:	1 October 2023

— *Abstract* —

In this research we construct and evaluate a recurrent neural network using input returns features to predict asset rankings. We investigate to what extent this complex system outperforms a simpler linear system and alternative complex system definitions. The main findings of this research can be categorized over two perspectives. From an empirical asset ranking perspective, we find that our complex system exploits the same market inefficiencies as a linear system, but also learns to allocate assets with true extreme tail rankings exclusively to the lower predicted percentiles. We identify that the use of volatility-oriented features can explain this exclusive allocation property and can lead to significantly improved Spearman correlation coefficient scores between predicted and true rankings. On the other hand, we find that ranking on volatility is detrimental on a portfolio level. We find that the inferiority of our long-short portfolio can be attributed to an imbalance of asset with extreme returns, due to the exclusive allocation of assets with extreme returns to the short portfolio. Finally, we find no substantial differences on portfolio level between our recurrent network and alternative complex system architectures.

Contents

1 | Introduction

Some of the most pervasive equity factors such as price momentum by Jegadeesh and Titman [1993] and short-term reversal by Jegadeesh [1990] rely on past returns to predict future stock returns. However, practical application of these strategies present limitations such as the existence of momentum crashes by Daniel and Moskowitz [2013] and Daniel and Moskowitz [2016].

In an attempt to overcome these limitations, two strands of research can be identified. The first strand of research focuses on reformulation of the asset's return series to which the momentum or reversal strategy is applied. One such reformulation is residual momentum as proposed by Blitz et al. [2011]. This implementation is successful in terms of absolute and risk-adjusted metrics as well as greater performance consistency with reduced portfolio crashes. Another reformulation by Chen et al. [2021] successfully proposes the use of a non-parametric method to redefine returns to ranks and signs before application of the momentum strategies. Chen et al. [2021] attribute their success to a reduced sensitivity to saliency considerations of market participants. All in all, this strand of research illustrates the promise of transformations to crude inputs such as cumulative returns for momentum-based portfolio construction strategies.

The second strand of research focuses on the application of momentum strategies to asset return forecasts. In this strand of research, the prominent global advance of machine learning presents itself through the proposal of complex system architectures. In one such work, Murray et al. [2021] apply machine learning to predict ranks from cumulative returns. They conduct this research on the US equities in the CRSP universe of assets between 1926-2022. They claim to extract signals with strong predictive power for future returns and their portfolios present an alpha factor distinct from both momentum and reversal. On the other hand, their investigation excludes a comparative analysis to simple linear parametric systems. In light of these results, we extend the research by Murray et al. [2021] by further investigating the superiority of machine learning for portfolio construction solely based on asset performance/return statistics. To this end, we first conduct a comparative analysis of a complex parametric neural network to a set of simpler linear parametric benchmarks. Specifically, we consider a system that implements both convolutional and recurrent neural networks to predict the relative ranking of an asset within a cross-section based on only its own historical performance in the form of its cumulative returns. Therefore, the first goal of this research is to determine more insight into the question,

To what extent do complex neural networks improve over simpler linear system definitions for financial portfolio construction purposes?

Through this critical question on the application of machine learning, we attempt to answer whether the predictive power in the proposed complex systems is due to their specific architecture or because they are simply data-driven and parametric. We conduct this analysis in a recent time period 1995-2022 as well as an equally-weighted portfolio setting in the CRSP and S&P500 asset universes. We do this in order to shed light on performance during this recent time period and on the practical applicability for investment respectively.

Furthermore, we consider whether alternative system configurations/architectures can improve over the investigated complex neural network architecture. In this architecture recurrent neural networks are used. However, these networks exhibit deficiencies including limited parallelizable training, exploding/vanishing gradients and difficulty with long-term dependencies as identified by Hochreiter and Schmidhuber [1997] and Goodfellow et al. [2016]. Consequently, Vaswani et al. [2017] proposed transformer system architectures levering self-attention in an attempt to

overcome the previous deficiencies. Transformers leveraging this self-attention mechanism have been successfully implemented by Zhou et al. [2019] to predict returns of commodity metals. In another setting they have been used for ranking purposes in currency ranking tasks by Poh et al. [2021]. For portfolio management, Kisiel and Gorse [2023] utilize transformers to predict positions of assets within a portfolio space. Therefore, we propose the use of these attention-based frameworks by Vaswani et al. [2017] as an alternative system architecture to recurrent neural networks. Furthermore, we investigate the use of ranked instead of absolute cumulative returns in line with Chen et al. [2021]. Therefore, the second goal of this research is to determine,

To what extent do attention-based neural network system architectures improve upon recurrent neural network system architectures for momentum-based asset ranking prediction?

and,

To what extent do ranked input feature definitions improve upon absolute input feature definitions for momentum-based asset ranking prediction?

Finally, we analyse whether the investigated machine learning systems identify other market inefficiencies besides those already identified and utilized in linear systems and in classical momentum and reversal factor approaches. Therefore, the third goal of this research is to determine,

To what extent do machine-learning based asset rank predictions differ from linear asset rank predictions and classical factor portfolios?

In order to answer these research questions, we utilize a complex recurrent neural network using past returns to predict the future rank of an asset. Alternatively, we consider a linear system and a self-attention based neural network. In our methodology, we describe these systems and address the difference between the self-attention and LSTM complex neural-network frameworks as well as the use of ranked and non-ranked inputs. In this research, we use the equally-weighted decile ten minus decile one portfolio in the CRSP and S&P500 asset universes. We exclusively estimate our systems on the CRSP universe, but analyse performance on the S&P500 as well, in order to gain insight into performance on this realistic subspace with regard to practical investments. As an initial analysis, we consider the performance of classical non-parametric benchmarks in our data. For the benchmarks, we immediately identify general limitations in profitability as observed by extensive periods of flat cumulative returns as well as severe drawdowns. This further motivates our investigation into the potential of parametric alternatives.

The main findings of this research can be categorized over two perspectives. From an empirical asset ranking perspective, we find that our complex machine-learning based system exploits the same market inefficiencies as a linear system, but also learns to allocate assets with true extreme tail rankings exclusively to lower predicted percentiles. We identify that the use of volatility-oriented features can explain this exclusive allocation property. We build this finding on the fact that neural networks present the ability to access volatility related features through their inherent complexity and that these features themselves are inherently predictable in equity returns. In addition, we connect this finding to the negative relationship between idiosyncratic volatility and returns that was established by Chen and Petkova [2012] which is clearly apparent in our regression results. Furthermore, we find that the use of volatility-based features by our complex neural networks can lead to a significantly improved Spearman correlation coefficient between predicted rank and the ranked next periods return. On the other hand, from a portfolio construction perspective we find that ranking on volatility is detrimental for portfolio performance. On a decile portfolio level, our complex system fail to achieve an increasing return pattern and present a significant volatility skew in the lower deciles. Consequently, we achieve a D10-D1

portfolio that loads strongly on the low volatility, momentum and reversal factor and is inferior to the portfolio constructed through a linear system's predictions. We find that this inferiority is attributed to complex system's prediction signal missing out on the benefits of cancelling extreme returns, due to their exclusive allocation to of assets with extreme returns to the D1 portfolio. Consequently, we observe a decreased average return and increased average volatility in the D10-D1 portfolio. Specifically, we observe that this increased volatility of D10-D1 occurs due to differences in the recovery speed between D10 and D1 around drawdowns. Finally, we find no substantial differences between our proposed complex configurations on a portfolio level or with regard to empirical asset ranking ability. Specifically, the superiority of systems utilizing LSTM's or self-attention sub-modules or ranked inputs is not clearly apparent.

This work has implications for researchers and practitioners who intend to predict the cross-section of asset ranks explicitly. Our research shows that connection between the Spearman correlation coefficient score and portfolio performance is limited. Specifically, this occurs because the Spearman correlation coefficient scores overall ranking performance and is not concerned about the loss of asymmetry/imbalance of the solution. The predictions from our complex systems serve as an example case for such an asymmetric solution, where the Spearman correlation coefficient is improved, yet the overall rank distribution becomes practically undesirable. Consequently, we do not recommend the use of this metric to measure/predict portfolio performance for predictions from systems with distinctly different empirical structures such as the 'X'-shape and 'inverted U'-shape presented in this research. Furthermore, in our research we identify that the use of a ranked target definition can amplify its relationships to other variables such as volatility. As a result of these amplified relationships, a system optimized for asset ranking might indeed attain this superior empirical asset ranking structure with inferior portfolio performance. Finally, the findings in this work have implications for future research on momentum-based strategies intending to leverage on complex system configurations. The general superiority of a linear system on portfolio level can be interpreted as evidence against an increase in the complexity of financial prediction systems. Specifically, this research can be interpreted as a critical review of the work by Murray et al. [2021], who limit the analysis of their complex parametric system to non-parametric benchmarks. More generally, this research stands in contrast to works such as Gu et al. [2020] and Kelly et al. [2021] advocating for the prominence of complex systems in finance.

This main limitations of this research are threefold. First of all, we estimate our systems on US equity data contained in the CRSP dataset and evaluate on the S&P500 universe of assets. This attempt at 'transfer learning' is shown to be inadequate, yet without a comparison to systems estimated on S&P500 specifically. Secondly, we limit our analysis on portfolio level to metrics capturing performance. However, practical considerations such as turnover, transactions costs and the impact of implementation lag are not considered. Finally, in this research we limit ourselves to use of end-of-month returns and do not consider the effects (e.g. seasonality) at this specific calendar frequency.

In what follows, chapter 2 provides the methodology of this research. Before considering the results of the proposed methodology, a description of the data and our initial analysis on the non-parametric benchmarks is provided in chapter 3. Subsequently, we consider the results of our systems in chapter 4. This analysis is performed from two main perspectives. Namely, one on empirical asset ranking ability in the cross-section and one on a portfolio level. Finally, a conclusion is presented in chapter 5.

2 | Methodology

2.1 Portfolio construction and non-parametric benchmarks

In this research we analyse portfolio performance of our systems in CRSP and S&P500 asset universes. Note that this research will utilize 'transfer learning' for the S&P500 asset universe. This means that any parametric system will solely be trained on the CRSP universe data, yet we also evaluate the out-of-sample performance of this same system on S&P500 asset universe. In these asset universes we setup equally-weighted decile portfolios according to the methodology of Jegadeesh and Titman [1993] and adhered to by Murray et al. [2021]. Thus, the breakpoints for these deciles correspond to those assets exchanged on the NYSE alone in case CRSP data is used. On the other hand, if only the high market-capitalization S&P500 universe of assets is considered, no subset of assets is used to construct the breakpoints. After allocating all assets under consideration to a decile, equally weighted portfolios are setup. The 10-1 decile portfolio (D10-D1) forms the main portfolio for analysis. Any constructed portfolio is evaluated by computation and analysis of its (cumulative) return. Unless otherwise defined we analyse the excess risk free rate return. Specific attention is paid to crashes in return quantified by downturn and risk-adjusted return quantified by Sharpe ratio's. Within the entire cross-section of assets, the difference between predicted and true ranking is quantified by the Spearman rank correlation coefficient.

In this research we will compare portfolios against a set of classical benchmarks. Explicit performance comparison is done against the momentum MOM factor Jegadeesh and Titman [1993], the short term reversal REV factor Jegadeesh [1990] and the residual momentum $RESMOM$ factor as proposed by Blitz et al. [2011]. In this research we also employ the use of factor spanning regressions. In this case we include the three factor Fama-French model (Mkt, SMB, HML) by Fama and French [1992] appended by the reversal (REV) from Jegadeesh [1990], residual momentum ($RESMOM$) by Blitz et al. [2011] and low-volatility factor ($LowVol$) by Blitz and Van Vliet [2007]. All risk factors are based on the signals defined by the previous set of authors and are constructed using equal-weighting schemes in their respective universes.

2.2 Linear parametric system benchmarks

In the research we intend to answer the question on the *the cost and gain of system complexity* by evaluating the performance of complex parametric systems versus a simple but parametric linear systems. This stands in contrast to the non-parametric benchmarks in section 2.1, which are not data-driven/parameterized. To this end, we introduce two linear system definitions whose parameters are optimized to predict ranks rather than absolute returns. As input feature set, these linear systems utilize the logarithmic, absolute excess market returns over the previous 12 months as independent variables. These corrected returns are used in an attempt to reduce market timing effect, by implying an assumed $CAPM - \beta$ coefficient equal to 1. Although we recognize that this assumption is imperfect, we also note that computation of the actual $CAPM - \beta$ coefficient would be prone to estimation error. The inputs are used to predict the normalized rank return r_{norm} as defined in Equation 2.2. Mathematically these linear systems can be represented as,

$$r_{norm_{t,i}} = \omega + \sum_{s=t-12}^{t-1} \theta_s \cdot \log\left(1 + (r_{s,i} - r_{s,m})\right) \tag{2.1}$$

where $r_{t,i}$ denotes the return of asset i at time step t. $r_{t,f}$ represents the equally-weighted market return at time step t. θ_s indicates the coefficient of the associated time step. log indicates the natural logarithm. ω indicates the intercept term which is included to account model misrepresentation. Note that ω will play no role when the predictions of all assets with a specific time step assets are ranked. Furthermore, for the target variable definition we follow Murray et al. [2021]. They construct a target variable dubbed the *normalized ranked return* and denoted for asset i at time step t as,

$$y_{t,i} = r_{norm_{t,i}} = \Phi^{-1}\left(\frac{\text{rank}(r_{t,i}; \mathbf{r_t})}{N_t + 1}\right) \tag{2.2}$$

where Φ^{-1} is the inverse cumulative Gaussian distribution function. *rank* represents the ranking operator applied to $r_{t,i}$, but logically subject to the entire set of returns; $\mathbf{r_t}$. The ranking operation labels an asset as an integer in $1, \ldots, N_t$. Although the Φ^{-1} function is unbounded, its empirical bounds are governed by the cross-section size (N_t). For our data set with a maximum $N_t \approx 6500$, the target variable is bounded by $[-3.6; 3.6]$. It is important to recognize that this r_{norm} definition maps a uniformly distributed variable (the ranked return) to the Gaussian space. Therefore, this transformation procedure is dubbed the 'Gaussian ranking' transformation procedure. This definition is considered theoretically beneficial with regard to model assumptions, since we can use normally distributed instead of uniformly distributed ranks in a regression.

For our comparative purposes we optimize this system in two ways. The first includes a conventional weighted least-squares (WLS) setting in which we opt for a look-back window of 120 months, i.e. 10 years. We opt for this size of look-back as it is inline with the $\frac{N_{months}}{N_{parameters}}$ ratio of ≈ 9 chosen for the construction of residual momentum by Blitz et al. [2011], but also to invoke stability through periods with financial crashes. Furthermore, we utilize the weighted version of least squares to account for the changes in cross-section size through time as visualized in Figure 3.1. As such, the sample weight is only governed by the assets time step t and equal to $\frac{1}{N_t}$ where N_t indicates the cross-section size at time step t. All in all, this configuration is dubbed to the *excess-market returns weighted least-squares* and abbreviated to $rtnXmkt - WLS$.

Secondly, we optimize the system using exactly the same methodology as we choose for the complex systems described next in section 2.3. This means we optimize the Spearman correlation coefficient (subsection 2.4.1) with cross-sectional sampling (subsection 2.4.2) for the linear system with an expanding window training set and a fixed validation window of 5 years. This also means that similarly to the complex systems, the performance of the linear system is evaluated on the validation set after an epoch of training and is subject to early stopping. This is configuration is included in the investigation for completeness and can be used to answer whether the optimization scheme plays a role in the results. All in all, this configuration is dubbed to the *excess-market returns Linear system* and abbreviated to $rtnXmkt - Linear$. It should be noted that we generally expect close alignment between these two linear configurations due to the equivalence between the optimization objectives as addressed in chapter B.

2.3 Complex parametric systems

In this research, we define a set of four complex system configurations. We consider one 'base' complex configuration and three extensions. Our base configuration is the convolutional and long-short term neural network designed by Murray et al. [2021], abbreviated as *CNN-LSTM*

in their work. It should be noted that compared to the methodology of Murray et al. [2021], we deviate through the use of excess market returns, explicit Spearman correlation optimization and cross-sectional batch sampling. These changes addressed in full throughout the rest of this chapter. They are implemented to better align the system's implementation with its practical use case (portfolio-construction), but ultimately are not a cause for significant changes in performance. In other words, these changes can be considered as attempts to reduce potential limitations in the original methodology, yet without substantially changing the practical performance of the systems.

2.3.1 Input and target variables

The input set as defined by Murray et al. [2021] contains 12 cumulative (excess risk-free rate) returns constructed over the period $[t-12, t-1]$. Once again, In an attempt to reduce the effects of market timing we propose the use of excess market returns instead. This set of cumulative excess market returns is denoted as $CR_{s_{t,i}}$. Mathematically, the cumulative return over m months, $CR_{m_{t,i}}$, in % can be expressed as,

$$CR_{s_{t,i}} = \left(\prod_{j=t-12}^{t-13+s} (1 + (r_{j,i} - r_{j,m})) \right) \cdot 100\% \tag{2.3}$$

where $r_{t,i}$ denotes the return of asset i at time step t. $r_{t,m}$ represents the equally-weighted market return at time step t. Thus, the set of inputs can be defined as $\{CR_{1_{t,i}}, \ldots, CR_{12_{t,i}}\}$. This will form the complete input set $x_{t,i}$, which contains an additional 'sequence' dimension (s) of size 12. As an example, $x_{t,i,1}$ denotes $CR_{1_{t,i}}$.

As an alternative to the absolute cumulative excess return, we investigate the use of the *cumulative rank* denoted as CP over the period $[t-12, t-1]$. We hypothesise that this ranked definition should improve performance on this asset ranking objective through its focus on relative performance rather than the absolute version of the cumulative return inputs. Furthermore, this definition might be more robust/stable as Chen et al. [2021] argue that rank and sign definitions of returns might overcome saliency considerations in momentum signal construction. Mathematically, the cumulative rank over s months, $CP_{s_{t,i}}$, can be expressed as,

$$CP_{s_{t,i}} = \sum_{j=t-12}^{t-13+s} 2 \cdot \left(\frac{\text{rank}(r_{j,i}; \mathbf{r_j})}{(N_s + 1)} - 0.5 \right) \tag{2.4}$$

where $r_{t,i}$ denotes the return of asset i at time step t. *rank* represents the ranking operator applied to $r_{t,i}$, but logically subject to the entire set of returns; $\mathbf{r_t}$. N_t designates the number of assets in the considered universe at time step t. The ranking operation labels an asset as an integer in $1, \ldots, N_t$, which we re-scale by $2 \cdot (x - 0.5)$ to [-1,1] centered around zero. Thus, this set of inputs can be defined as $\{CP_{1_{t,i}}, \ldots, CP_{12_{t,i}}\}$. This will form the complete alternative input set $x_{t,i}$, again consisting of the 'sequence' dimension s of size 12. Finally, it should be noted that these systems will use the *normalized ranked return* (r_{norm}) as target variable as defined in Equation 2.2.

2.3.2 System overview

In our research we consider convolutional and long-short term neural network system architecture as defined by Murray et al. [2021] (called *CNN-LSTM* in their work) and utilized recently in Eggebrecht and Lütkebohmert [2023]. In fact, this system consists of three main components

and in this research is dubbed to CNN-LSTM-FNN, consisting of a *convolutional* neural network (CNN) with max-pooling, a Long-Short term memory unit (LSTM) and a final feed-forward neural (FNN) network. Later in this section, the *self-attention* subsystem is defined as an alternative to the LSTM subsystem. This system is visualized in Figure 2.1.

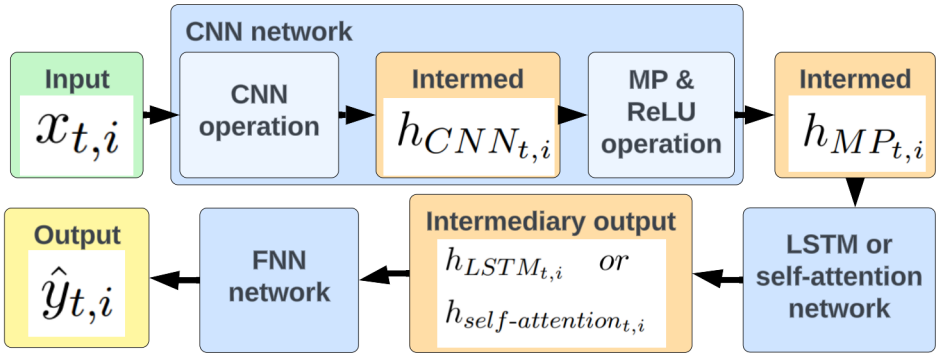


Figure 2.1: System overview for the *CNN-LSTM-FNN* or *CNN-Attention-FNN* system operating on cumulative returns (CR) or cumulative ranks (CP) to predict the *normalized* rank return (i.e. $\hat{y}_{t,i}$ or $\hat{r}_{norm_{t,i}}$). This system is inspired by the *CNN-LSTM* system in Murray et al. [2021].

Each of these three subsystems serve different purposes to obtain a rank prediction. Firstly, the one-dimensional convolutional neural networks (CNN) are considered. Similar to their multi-dimensional counterparts, these networks can be interpreted as spatial feature extractors as discussed in Goodfellow et al. [2016]. This system slides a convolutional operator known as a kernel over the input, thereby aggregating information sensitive to the parametric kernel. In an attempt to filter the information content and reduce the memory required by the system, these convolutional layers are often followed by pooling layers. Murray et al. [2021] opt for the use of max-pooling (MP) layers. The CNN and MP subsystem can be represented as,

$$
\begin{aligned}
z_{t,i} &= \text{ReLU}\left(\text{MaxPool1D}\left(\text{CNN}\left(x_{t,i}; W^{CNN}, b^{CNN}\right)\right)\right) \\
&\text{where,} \\
h_{CNN_{t,i,s}} &= \text{CNN}\left(x_{t,i}; W^{CNN}, b^{CNN}\right) = \sum_{k=0}^{Kernel^{CNN}} x_{t,i,s+k} \cdot W_s^{CNN} + b^{CNN} \\
h_{CNN_{t,i}} &= [h_{CNN_{t,i,1}}, \ldots, h_{CNN_{t,i,8}}]^T \\
&\text{and,} \\
h_{MP_{t,i,s}} &= \text{MaxPool1D}\left(x_{t,i}\right) := \text{Max}\left(x_{t,i,(s \cdot Kernel^{MP}):(s \cdot Kernel^{MP} + stride^{MP})}\right) \\
h_{MP_{t,i}} &= [h_{MP_{t,i,1}}, \ldots, h_{MP_{t,i,4}}]^T
\end{aligned}
\tag{2.5}
$$

where W^{CNN} and b^{CNN} represent the CNN's parameter weight matrix (or *kernel window*) and bias vector respectively. $z_{t,i}$, $h_{t,i}$ and $x_{t,i}$ represent the subsystem's output, intermediary and input sample respectively for asset i at time step t. Note, that the kernel window slides over the additional 'sequence' dimension s, from $s = 1 \rightarrow s = 12$. $Kernel^{CNN}$, $Kernel^{CNN}$ and $stride^{MP}$ represent the convolutional and max-pooling kernel size and the stride size respectively. It is important to note that in the CNN and MP layers the sequence dimension reduces to $s_{max} = 8$ and $s_{max} = 4$ respectively due to the aggregating operations at hand. ReLU() denotes the nonlinear 'rectified linear unit' activation function. This CNN subsystem is visualized in Figure 2.2.

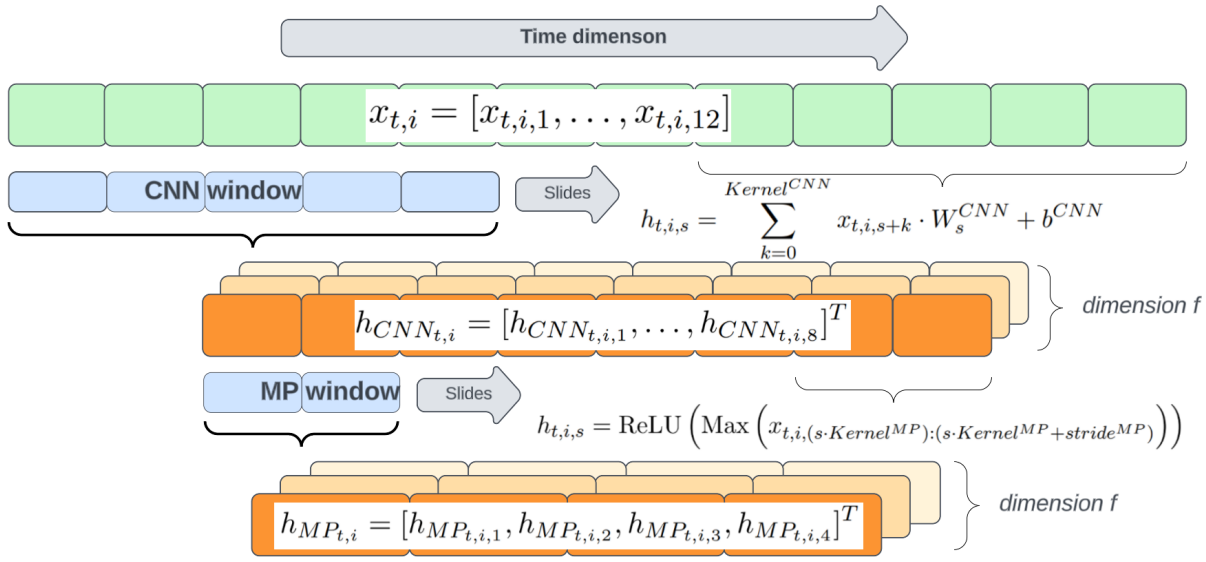


Figure 2.2: CNN network overview according to the notation in Equation 2.5. The parametric convolutional, max-pooling and ReLU operations are sequentially applied by sliding the respective windows over the time-dimension (s) of the samples. Notation according to the definitions in Equation 2.5.

The second subsystem considered is a form of the recurrent neural network (RNN), namely the long-short term memory unit (LSTM) by Hochreiter and Schmidhuber [1997]. RNN networks provide connections to previous inputs in an attempt to allow for the flow of information to previous samples as discussed in Goodfellow et al. [2016]. These previous samples are connected to the current one through a type of ordering, conventionally in the time dimension. These additional recurrent connections should allow this architecture to exploit local time series dynamics to improve its prediction. The recurrent LSTM subsystem can represented as,

$$z_{t,i,s} = \text{LSTM}(x_{t,i,s}, h_{t,i,s-1}; \psi_s) \quad \forall \quad s \in \{1, 2, 3, 4\} \tag{2.6}$$

where $z_{t,i,s}$, $h_{t,i,s}$ and $x_{t,i,s}$ represent the subsystem's output, intermediary and input sample respectively for asset i at time step t and the sequence index s. Note that the LSTM subsystem is applied to the sequence dimension with a remaining size of $s_{max} = 4$ after application of the CNN subsystem. The LSTM has trainable parameters ψ_s for each of these sequence entries, where $h_{t,i,0}$ is always initialized with zeros. In our setting, $z_{t,i,4}$ forms the final output possible and is used for any subsequent layers. For specific details on the LSTM operation we refer to the original documentation on the long-short term memory unit by Hochreiter and Schmidhuber [1997]. This LSTM subsystem is visualized in Figure 2.3.

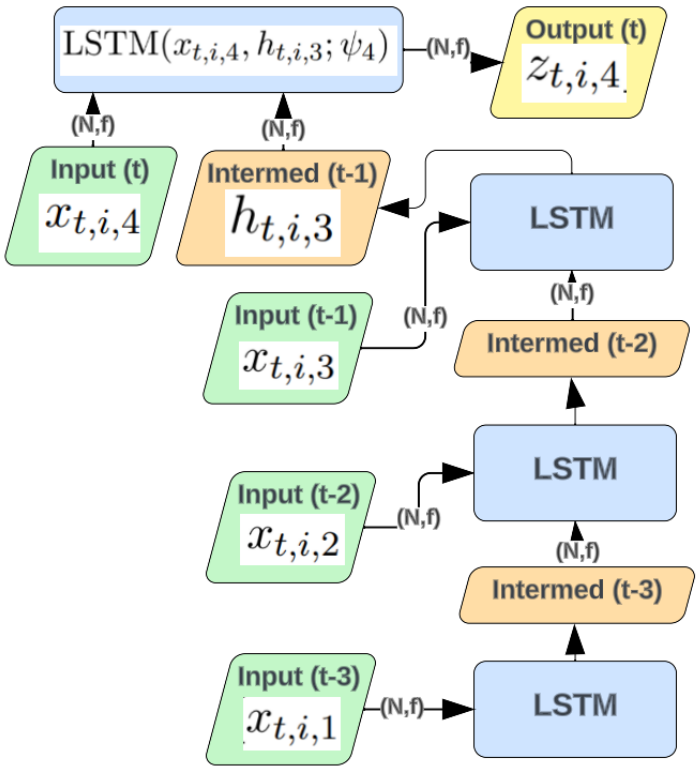


Figure 2.3: LSTM network with annotated dimensions of inputs and outputs in brackets. The blue blocks indicate a parametric operation is conducted. In this diagram the implicit temporal structure is visible as introduced in Figure 2.4. Notation according to the definition in Equation 2.6, where $z_{t,i,s}$, $h_{t,i,s}$ and $x_{t,i,s}$ represent this subsystem's output, intermediary and input sample respectively for asset i at time step t and the sequence index s.

The final subsystem considered by Murray et al. [2021] is a feed-forward neural network (FNN) outlined in Bishop et al. [1995]. This architecture consists of a layer of fully connected neural nodes with nonlinear activation's functions and a final output layer leading to a scalar output. The intention of such a network is inline with that of partial least squares (PLS) regression by Geladi and Kowalski [1986], yet without transformation of the target variable to a new hyperplane as well. Furthermore, the non-linear activation functions within these types of networks expand on PLS as they allow for subsequent transformations of the hyperplane in an attempt to further meet the optimization objective. The FNN subsystem can represented as,

$$z_{t,i} = \text{ReLU}\left(x_{t,i}W^{FNN} + b^{FNN}\right) \tag{2.7}$$

where W^{FNN} and b^{FNN} represent the FNN's parameter weight matrix and bias vector respectively. $z_{t,i}$ and $x_{t,i}$ represent the subsystem's output and input sample respectively for asset i at time step t. ReLU() denotes the nonlinear 'rectified linear unit' activation function.

All in all, this CNN-LSTM-FNN architecture can be interpreted as a feature extractor applied to the 'raw' input features through the CNN network, after which the LSTM should be able to leverage any time dependence between the extracted features. Finally, the FNN is able to transform the spatial and time extracted features once more to improve scoring on the objective.

In an attempt to limit overfitting, regularization techniques are applied. Murray et al. [2021] apply early stopping as well as dropout layers. Dropout layers are applied to other predefined

layers and deactivate nodes therein. Many interpretations of their functionality exist, such as the construction of implicit ensembles whenever none of the nodes are deactivated as discussed in Goodfellow et al. [2016].

Finally, we describe the CNN-LSTM-FNN system architecture used in this research specifically. For this system with a consistent f number of units the architecture can be described as follows with the output dimension provided in brackets with N the number of samples,

Layer. *(Output Dimension)*: *Description*

IN. $(N, 1, s = 12)$: System inputs, either the set of CP or CR.

1. $(N, f, s = 8)$: 1 dimensional CNN network with kernel size of 5, no input padding, stride equal to 1 and intercept parameters included. ReLU activation function.
2. $(N, f, s = 4)$: 1 dimensional max-pooling with kernel size and stride equal to 2.
3. $(N, s = 4, f)$: Dimensional permutation layer, no parameters involved. This layer conducts a transpose operation on dimensions two and three.
4. (N, f): LSTM network applied over the time dimension T (N, T, f) (also referred to as sequence dimension $(N, s = 4, f)$), *many-to-one* output focussing on final timestep i.e. $(N, -1, f)$.
5. (N, f): FNN network with a fully connected layer including intercept using ReLU activation function. This layer also apply dropout with a probability of 20% inline with Murray et al. [2021].
6. $(N, 1)$: Output linear layer including intercept term transforms the layer's input vector into a scalar prediction output.

OUT. $(N, 1)$: System outputs, the target rank prediction.

It is important to note that this specific architecture and its hyperparameters were designed, motivated and selected by Murray et al. [2021]. In this research we build on their work and test the potential of this system under the changes proposed and described in this and the previous sections. Thus, this research does not intend to improve the hyperparameters of this system through intensive hyper-parameter tuning.

Self-attention framework

To extend the set of investigated systems we consider self-attention-based networks first proposed by Vaswani et al. [2017] which form the basis for the recent prominence in transformer models. This self-attention mechanism is the most crucial part of such a network as it allows for the extraction and identification of long-term dependencies between sets of spatial or temporal features.

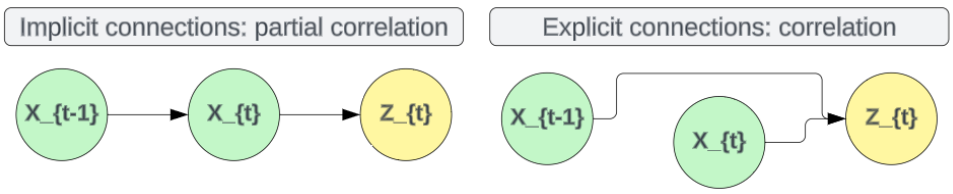


Figure 2.4: Implicit versus explicit temporal information flow overview as utilized by the LSTM and self-attention frameworks respectively.

In our research, we intend to leverage the ability of these systems to identify temporal rather than spatial dependencies. In this setting, the self-attention framework can be contrasted to that of the Long-Short Term Memory unit (LSTM) by Hochreiter and Schmidhuber [1997]; Where the LSTM provides the ability to establish implicit temporal connections between the layer's

inputs (X) of different time steps to the layer's output (Z) as visualized in Equation 2.6, the self-attention framework overcomes this characteristic by allowing for explicit versions of these temporal connections. This difference is visualized in Figure 2.4, in which we annotate that implicit connections are analogous to partial correlation, while explicit overcome this partiality. In this figure, notice the implicit/recurrent connections through the steps of X to reach the final output Z, where the explicit version is not recurrent and offers explicit/direct connections between the input steps to Z. We hypothesise that allowing for this explicit temporal information flow through self-attention networks should improve the predictive ability of the system as they allow for direct connections between all input time step to Z. As such, if there is relevant information contained in the earliest steps of X (e.g. CR_1), the self-attention framework should be superior in extracting it. On the other hand, we recognize beforehand that the potential advantage of explicit connections might be limited as the informational content is expected to diminish for inputs further from the current time step ($X_{<t}$) as given in the overview of Figure 2.4.

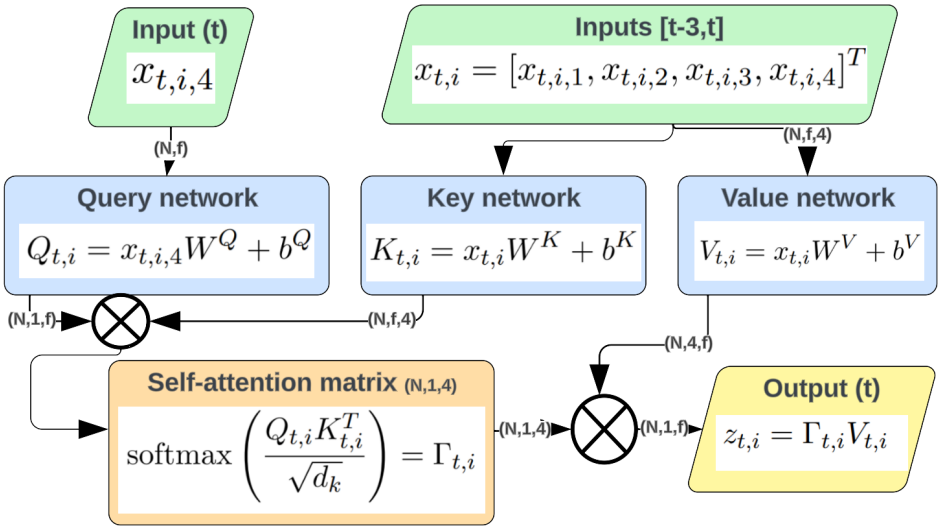


Figure 2.5: Self-attention network with annotated dimensions of inputs and outputs in brackets. The blue blocks indicate a parametric operation is conducted. In this diagram the explicit temporal structure is visible as introduced in Figure 2.4. The self-attention matrix is displayed at the bottom of the figure and holds the dot-product scalar between the different input time steps with regard to input at time step t. After applying the *softmax* operation to these dot-products the sum of these four scalars equals one. Notation according to the definition in Equation 2.8, where $z_{t,i,s}$, $h_{t,i,s}$ and $x_{t,i,s}$ represent this subsystem's output, intermediary and input sample respectively for asset i at time step t and the sequence index s.

The difference in temporal flow between the implicit LSTM and explicit self-attention architectures is visualized specifically in Figure 2.3 and Figure 2.5 respectively. Note that these diagrams are a functional higher-level schematic, for additional specific notes on motivation and implementation we refer to the original works by Hochreiter and Schmidhuber [1997] and Vaswani et al. [2017] respectively. In Figure 2.3, recognize that the flow of information occurs in an implicit and sequential fashion similar to the left-hand case of Figure 2.4, where the input at a time step is used together with the transformed information of the previous LSTM sub-module for the next LSTM sub-module. On the other hand, in Figure 2.5 we can recognize a parallel rather than sequential structure through the use of matrix multiplications, with all intermediary dimensions show in brackets. In this figure, the inputs of respective time steps are transformed through the blue parametric 'key' and 'query' sub-modules networks after which their dot product will determine how the outputs from the value network are aggregated to a single (rather than four-dimensional) output vector. Therefore, Figure 2.5 can also be used

for taxonomic purposes; in this figure the self-attention matrix allows for the network to learn parameters such that it can improve the *attention*/focus on it*self* through the establishment of connections between different time-steps.

To continue, we elaborate on the mathematics involved in the (parameterized) operations conducted in a self-attention network. Mathematically, self-attention by Vaswani et al. [2017] can be defined in our setting as,

$$\begin{aligned} z_{t,i} &= \text{softmax}\left(\frac{Q_{t,i}K_{t,i}^T}{\sqrt{d_k}}\right)V_{t,i} = \Gamma_{t,i}V_{t,i} \\ &\text{where,} \\ &Q_{t,i} = x_{t,i,4}W^Q + b^Q, \quad K_{t,i} = x_{t,i}W^K + b^K, \quad V_{t,i} = x_{t,i}W^V + b^V \\ &\text{and,} \\ &x_{t,i} = [x_{t,i,1}, x_{t,i,2}, x_{t,i,3}, x_{t,i,4}]^T \end{aligned} \tag{2.8}$$

where $Q_{t,i}$, $K_{t,i}$ and $V_{t,i}$ are the query, key and value matrices for a sample resulting from transformations of the input through parametric weight (W) and bias (b) layers. In Figure 2.5, these matrices are setup in the blue network blocks/sub-modules with identical names. d_k is the dimension of the keys. $z_{t,i}$ and $x_{t,i}$ represent the output and input sample respectively for asset i at time step t. Note that these samples still contain the additional 'sequence' dimension s of size 4, similar to definition of the LSTM given in Equation 2.6. Self-attention considers samples with 2 dimensions (s,f) which we denote as the (s) sequence and the embedding/feature (f) dimension. For a sample $x_{t,i}$, the self-attention matrix $\Gamma_{t,i}$ is setup through the dot product $Q_{t,i}$ and $K_{t,i}$ and subject to the softmax operator i.e. softmax $\left(\frac{QK^T}{\sqrt{d_k}}\right)$. Thus, the 'self-attention matrix' ($\Gamma_{t,i}$, denoted in orange in Figure 2.5) is of size (1,s) for every sample and has column sums equal to 1. The values inside these matrices, e.g. $\Gamma_{t,i,k}$ can be interpreted as the 'dependence' between sequence indices $s = k$ (i.e. any sequence entry k) and $s = 4$ (i.e. the final sequence entry). Consequently, the final matrix multiplication, denoted as ΓV applies an aggregation of the embeddings across the sequence dimension for every element in the sequence. Therefore, this final output should contain a mix of amplified and reduced features in its embedding dimension based on the importance to the optimization objective at hand. All in all, this framework should be interpreted through its parameter matrices; Q, K and V. Where Q and K parameters are optimized to amplify and reduce specific dependencies within the sequence dimension, the parameters for V are optimized to find the best transformation and aggregation of the inputs.

In this research we consider incorporating the self-attention framework into the CNN-LSTM-FNN system by replacing it with the LSTM component (previously defined as layer 4). Specifically, the replacement is incorporated as in the previously defined architecture at step 4 as follows,

4. $(N, s = 4, f)$: Self-attention network applied over the time dimension i.e. (N, T, f), such that the Γ matrix is of size (1,4). The key query and value matrices are obtained by fully connected linear layers with f number of units without intercepts and with no activation function applied as defined in Equation 2.8.

All in all, this research will consider four system configurations. First of all, we define the $CR - CLF$ and $CP - CLF$ configurations which utilize the *CNN-LSTM-FNN* architecture abbreviated to CLF. These two configurations differ in the input feature set used, namely cumulative excess returns (CR) and cumulative ranks (CP) respectively. The third and forth configuration is denoted as $CR-CAF$ and $CP-CAF$ which uses the cumulative excess returns (CR) and cumulative ranks (CP) respectively as inputs. They utilize the *CNN-Attention-FNN*

architecture abbreviated to CAF. For each of the four configurations previously mentioned we opt for $f = 32$ number of units in every intermediary layer.

2.4 Optimization

2.4.1 Objective

In their work, Murray et al. [2021] choose minimization of the weighted mean-squared error as their optimization objective. This loss function for asset i at time step t can be expressed as,

$$\mathcal{L}_{MSE_{t,i}} = w_{t,i} \cdot (y_{t,i} - \hat{y}_{t,i})^2 \tag{2.9}$$

where $w_{t,i}$ indicates the sample weight and $y_{t,i}$ and $\hat{y}_{t,i}$ denote the true and predicted normalized ranked return respectively. Murray et al. [2021] define $w_{t,i} = \frac{1}{N_t}$ in order to correct for inconsistent universe size through time as visualized in Figure 3.1. $y_{t,i}$ is the normalized ranked return as defined in subsection 2.3.1. It is important to note that the due to the structure of the applied Φ^{-1} mapping as well as the squared error scaling, a minimization of this error should inherently tends towards improved predictive ability of the tail rank compared to ranks closer to the center rank.

It is important to acknowledge that optimization of the MSE between ranks is practically equivalent to the optimization of the Spearman correlation coefficient as we illustrate in chapter B. However, this minimization of the MSE between ranks as applied by Murray et al. [2021] can be described as a *static* proxy. This is denoted as a *static* proxy because in their case, the computation of the error ($(y_{t,i} - \hat{y}_{t,i})^2$) excludes an actual ranking operation applied to the predictions (i.e. $(y_{t,i} - \text{rank}(\hat{y}_{t,i}, \hat{\mathbf{y}}_\mathbf{t}))^2$). As such, the 'predicted rank' of an asset (i.e. system's output) might change after ranking that asset's prediction to all other asset predictions at the respective time step. In other words, although the system's output ($\hat{y}$) might be interpreted as the 'predicted rank', in fact that interpretation is only correct after we rank the actual predictions i.e. $\text{rank}(\hat{y}_{t,i}, \hat{\mathbf{y}}_\mathbf{t})$. Hence, the interpretation of $\hat{y}_{t,i}$ as 'predicted rank' is denoted as a *static* proxy.

Although we expect limited divergence between using this *static* rank proxy and the actual ranked version, in this research we intend to use the actual ranked version. We do this in an attempt to achieve a further alignment between the optimization routine and the practical portfolio use-case. To this end, we utilize the ranking operation designed by Blondel et al. [2020] in order to rank our predictions during optimization and thereby maximize the Spearman correlation coefficient directly. Although a conventional ranking operation constitutes an ill-defined ranking function, the work by Blondel et al. [2020] defines an alternative well-defined ranking function with use-able gradients. As such, the use of this version allows us to theoretically achieve the improved alignment between the optimization routine and the portfolio use-case. For specific details on this alternative ranking function we refer to the work by Blondel et al. [2020], where we opt for a regularization strength (= hyper-parameter) equal to 1 as the authors propose. After applying this differentiable version of the ranking operation to the targets and predictions, the to-be-maximized Spearman correlation coefficient contribution of asset i at time step t can be represented as,

$$\begin{aligned}\mathcal{L}_{SPR_{t,i}} &= \rho_{Spearman}\left(\hat{y}_{t,i}, y_{t,i}, w_{t,i}; \hat{\mathbf{y}}_\mathbf{t}, \mathbf{y_t}, \mathbf{w_t}\right) \\ &= \rho_{Pearson}\left(\widehat{\text{rank}}(\hat{y}_{t,i}; \hat{\mathbf{y}}_\mathbf{t}), \text{rank}(y_{t,i}; \mathbf{y_t}), w_{t,i}; \hat{\mathbf{y}}_\mathbf{t}, \mathbf{y_t}, \mathbf{w_t}\right) \\ &= \frac{w_{t,i} \cdot (\widehat{\text{rank}}(\hat{y}_{t,i}; \hat{\mathbf{y}}_\mathbf{t}) - M_{\hat{\mathbf{y}}_\mathbf{t}}) \cdot (\text{rank}(y_{t,i}; \mathbf{y_t}) - M_{\mathbf{y_t}})}{\sum_j^{N_t} w_{t,j} \sqrt{S_{\hat{\mathbf{y}}_\mathbf{t}} S_{\mathbf{y_t}}}}\end{aligned} \tag{2.10}$$

where the sum over the entire sample set at time step t is bounded between $[-1, 1]$. In this equation, $y_{t,i}$, $\hat{y}_{t,i}$ and $w_{t,i}$ denote the target, prediction and weight for asset i at time step t respectively. $\widehat{\text{rank}}$ denotes the differentiable ranking operation by Blondel et al. [2020] applied to the predictions and rank denotes the conventional ranking operation applied to the targets. Note, that besides the intermediary ranking operation, this correlation coefficient definition is equivalent to that of Pearson correlation coefficient. Unless otherwise defined, in this research the asset sample weights $w_{t,i}$ are set to $\frac{1}{N_t}$, where N_t denotes the number of samples at time step t (i.e. the cross-section size of the asset universe). In that case, this scalar value can be interpreted as the conventional Spearman correlation coefficient; it measures the linear alignment between the ranks of two one-dimensional arrays. In Equation 2.10, M_v, S_v and $S_{v,u}$ represent the weighted mean, variance and covariance of vector v and u respectively. The complete set of mathematical notations required for the computation of the (weighted) Spearman rank correlation coefficient are given as follows,

$$\hat{y}_{rank_{t,i}} = \widehat{\text{rank}}\left(\hat{y}_{t,i}; \hat{\mathbf{y}}_t\right), \quad y_{rank_{t,i}} = \text{rank}\left(y_{t,i}; \mathbf{y}_t\right) \tag{2.11}$$

$$M_{\hat{\mathbf{y}}_t} = \frac{\sum_i^{N_t} w_{t,i} \cdot \hat{y}_{rank_{t,i}}}{\sum_i^{N_t} w_{t,i}}, \quad M_{\mathbf{y}_t} = \frac{\sum_i^{N_t} w_{t,i} \cdot y_{rank_{t,i}}}{\sum_i^{N_t} w_{t,i}} \tag{2.12}$$

$$S_{\hat{\mathbf{y}}_t} = \frac{\sum_i^{N_t} w_{t,i} \cdot \left(\hat{y}_{rank_{t,i}} - M_{\hat{\mathbf{y}}_t}\right)^2}{\sum_i^{N_t} w_{t,i}}, \quad S_{\mathbf{y}_t} = \frac{\sum_i^{N_t} w_{t,i} \cdot \left(y_{rank_{t,i}} - M_{\mathbf{y}_t}\right)^2}{\sum_i^{N_t} w_{t,i}} \tag{2.13}$$

which can be used together with the weighted covariance to compute the weighted Spearman correlation coefficient;

$$S_{\hat{\mathbf{y}}_t \mathbf{y}_t} = \frac{\sum_i^{N_t} w_{t,i} \cdot \left(\hat{y}_{rank_{t,i}} - M_{\hat{\mathbf{y}}_t}\right) \cdot \left(y_{rank_{t,i}} - M_{\mathbf{y}_t}\right)}{\sum_i^{N_t} w_{t,i}} \tag{2.14}$$

$$\rho_{Spearman_{\hat{\mathbf{y}}_t \mathbf{y}_t}} = \frac{S_{\hat{\mathbf{y}}_t \mathbf{y}_t}}{\sqrt{S_{\hat{\mathbf{y}}_t} S_{\mathbf{y}_t}}}. \tag{2.15}$$

In this research we opt for an expanding window approach with a validation set of five years. The first of these windows is defined as follows; 1975/01-1990/01 training set, 1990/01-1995/01 validation set and out-of-sample prediction period of 3 years between 1995/01-1998/01. After this the training window is expanded by 3 years, yet the validation window and the out-of-sample prediction window stay a consistent size of 5 and 3 years respectively. To further prevent overfitting on these windows, we optimize the systems by applying early stopping. Specifically, optimization is stopped if the validation set loss does not improve after five epochs since the last improvement. In this case, the system with the previously defined best validation loss score is returned. The system itself is trained for a maximum of 100 epochs, yet empirically early stopping is observed after [25,35] epochs.

2.4.2 Sampling

This research utilizes stochastic gradient (SGD) through the Adam optimization algorithm by Kingma and Ba [2014] as mentioned in subsection 2.4.1. Conventionally, the sampling of batches is done (pseudo-) randomly over the entire set. This is the methodology adhered to by Murray et al. [2021], who define a batch size of 2^{15}. It is important to recognize that this method of sampling places no restrictions on the origin of samples with regard to the asset space (i) or the time dimension (t). This is done in an attempt to generate batches of consistent size. Large (and consistent) batch sizes are advantageous as they invoke stability during model estimation process. This can be attributed to the main assumption of SGD that a similar distribution should be upheld across batches.

As an alternative sampling method we introduce and utilize cross-sectional sampling. In this approach we only (pseudo-) randomly sample a timestamp (t) and construct a batch by collecting all samples with that timestamp. As is visualized in Figure 3.1, this means that batches now have inconsistent sizes and are at least a factor 10^2 smaller than in the conventional sampling approach. Consequently, it is logical to expect more instability during the model estimation process as the previously mentioned SGD assumptions are weaker for this approach. To re-invoke this desired stability, we accumulate gradients across multiple batches before applying the optimization step with this accumulated gradient. We choose to update the accumulated gradients every 64 months which approximates nearly 5 years of data. On the other hand, this alternative sampling approach allows for optimization of the Spearman rank correlation coefficient ($\mathcal{L}_{SPR}$) across the asset space (i) for every time step t. Therefore, this sampling method is required as it is exactly inline with the practical use case of the system; portfolio construction based on ranked assets within a single period.

3 | Data

3.1 Filtering procedures and asset cross-section size

In this research we consider the CRSP data set as our focal asset universe, consisting of US equity assets on the NYSE, NASDAQ and AMEX exchanges. We apply the same data pre-processing rules for the inclusion of data and handling of delisted assets as Murray et al. [2021]. Besides these steps, we further reduce the dataset through the reduction of dual (or higher multiples) listings to a single listing. For this reduction we prefer listings in the NYSE, NASDAQ and AMEX order. Furthermore, we also consider portfolio performance in the S&P500 universe. For this universe we apply no specific filtering procedures. Note that return data in the CRSP and S&P500 universes is consistently sampled at the end-of-month frequency.

Moreover, in subsection 2.3.1 the methodology for the construction of input feature set is defined. To be considered a valid sample, assets must contain the complete information set (e.g. no missing months) required for the construction of these features, otherwise that sample is excluded from the data set. The result in terms of cross-section size through time from applying these data processing steps are visualized in Figure 3.1.

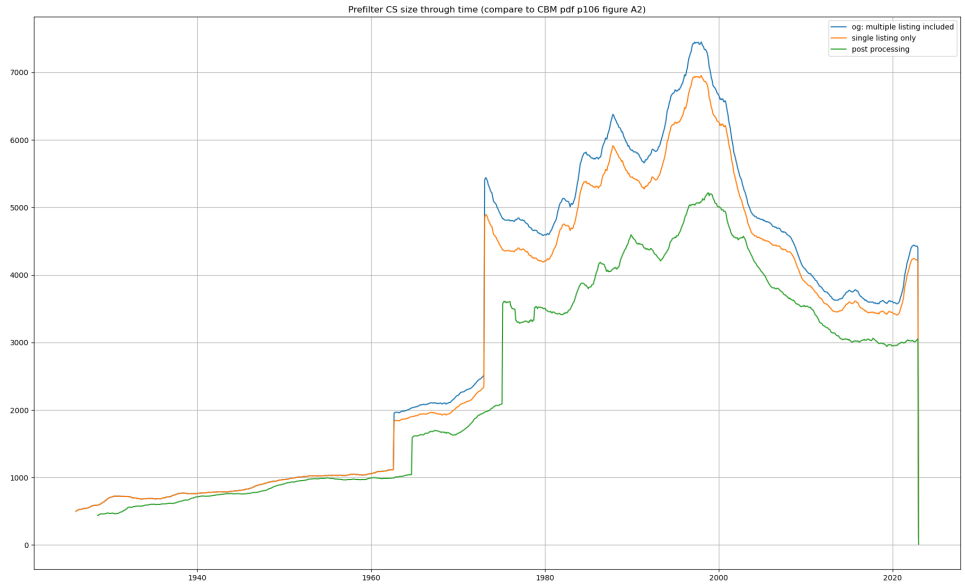


Figure 3.1: Original, single listing and final CRSP data set size through time. Final data set contains all *valid* samples with input complete feature set as required for parametric systems.

3.2 Non-parametric benchmark analysis

In this research we will consider equally-weighted portfolio performance between 1995-2022 in the CRSP and S&P500 (subset of CRSP) universes. It is important to note that while performance in S&P500 universe is considered, we always fit a system on the CRSP universe. The reason for this is an intended focus of analysis on the CRSP universe, yet evaluation of S&P500 universe to give an indication of practical investability. To give an indication of performance in this period and across these universes, we present the portfolio performance over time of the equally weighted market and classical MOM by Jegadeesh and Titman [1993], reversal REV by Jegadeesh [1990] and residual momentum $RESMOM$ by Blitz et al. [2011] benchmarks in Figure 3.2. The following results mainly focus on the performance of the MOM, $RESMOM$ and REV benchmarks as they are considered direct competitors to the proposed configurations in this research. This is because their signals definitions are based on the same (sub)set of information i.e. the previous months' returns. On the other hand, the other benchmarks e.g. HML intend to leverage on inefficiencies presented in alternative informational axes.

If we examine the portfolio performance in Figure 3.2 in the CRSP universe, we observe that the REV, MOM and $RESMOM$ benchmarks are able to generally achieve positive returns before the 2008 crash. On the other hand, this ability is diminished when considering the performance after 2008 and/or in the subset of S&P500 assets over the period 2001-2022. This observation is further supported when considering Figure 3.3 which clearly present the persistent flat/zero behaviour after 2008 of the MOM and $RESMOM$ benchmarks. Consequently, these benchmarks fail to recover from this crash of 2008. The REV benchmark shows diminished drawdowns and an improved ability for recovery in both universes. On the other hand, this signal presents limited profitability due to its lengthy periods of sustained flat performance as observed in the period between [2010,2019].

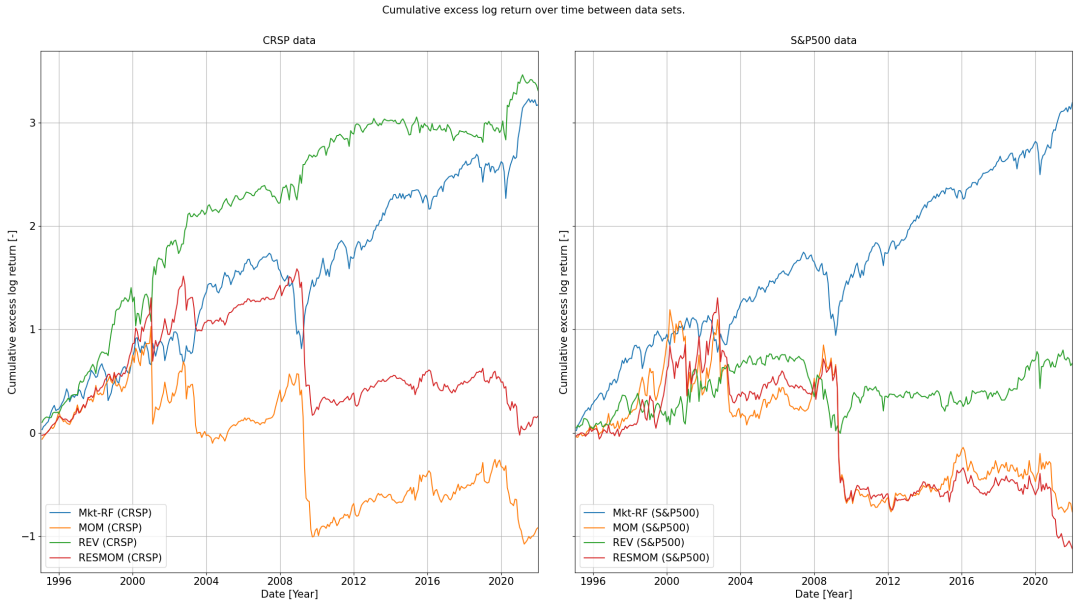


Figure 3.2: Equally-weighted portfolio performance for the Market, momentum, reversal and residual momentum factor in the period 1995-2022

Figure 3.3: Equally-weighted portfolio drawdown for the Market, momentum, reversal and residual momentum factor in the period 1995-2022

All in all, this overview of benchmark performance illustrates the unsatisfactory performance of equally-weighted portfolios in invest-able universes such as the S&P500. In their work, Murray et al. [2021] recognize this phenomenon and provide us with a parametric alternative to these classical signal definitions in the CRSP universe. As discussed in chapter 2 we further append this complex configuration with a larger set of parametric configurations. Consequently, in this work we intend to investigate to what extend these parametric definitions are superior in practically invest-able subspaces and if they can be further improved. This concludes our introductory data and non-parametric benchmark analysis.

4 | Results

In this chapter we discuss the performance and properties of our proposed system configurations. This analysis starts the results from the optimization itself in section 4.1. After this, we continue on to analyse the empirical asset ranking ability of our system configurations in section 4.2. We expand on this analysis in section 4.3, where we evaluate the predictive ability in the cross-section of assets. After these analyses we switch to an evaluation on portfolio level. This starts with the examination of decile portfolios in section 4.4. Thereafter, we evaluate our focal D10-D1 portfolio performance over time in section 4.5. Finally, we consider the focal portfolio's exposures to common risk factors in section 4.6.

4.1 Optimization analysis

In this section we provide summarizing optimization statistics for the considered configurations as well as the non-parametric MOM, $RESMOM$ and REV benchmark signals. As explained in section 2.4, optimization continues until divergence of in-sample (training set) and (semi) out-of-sample (validation set) Spearman correlation coefficient through the early stopping mechanism. The metric results are displayed in Table 4.1 which includes error metrics (MSE & MAE) and the Spearman rank correlation for the out-of-sample 1995-2022 period. The error for these error metrics (MSE & MAE) is based on the difference in *normalized ranked return* i.e. $|r_{norm} - \hat{r}_{norm}|$. Finally, to establish claims on the significant difference between prediction scores we utilize the two-sided Diebold-Mariano test by Diebold and Mariano [2002] with results given in section D.2.

Configuration	**MAE**	**MSE**	**SPR**
CR CLF32 x CRSP	1.039 (0.091)	1.825 (0.325)	0.081 (0.123)
CP CLF32 x CRSP	1.033 (0.089)	1.822 (0.352)	0.079 (0.115)
CR CAF32 x CRSP	1.294 (0.102)	2.469 (0.417)	0.071 (0.129)
CP CAF32 x CRSP	1.03 (0.09)	1.820 (0.351)	0.078 (0.126)
rtnXmkt-Linear x CRSP	1.016 (0.054)	1.801 (0.163)	0.044 (0.084)
rtnXmkt-WLS x CRSP	0.796 (0.006)	1.394 (0.015)	0.046 (0.092)
MOM (CRSP)			0.043 (0.116)
RESMOM (CRSP)			0.028 (0.109)
REV (CRSP)			0.018 (0.102)

Configuration	**MAE**	**MSE**	**SPR**
CR CLF32 x (S&P500)	1.04 (0.123)	1.929 (0.576)	0.03 (0.153)
CP CLF32 x (S&P500)	1.061 (0.121)	1.912 (0.505)	0.034 (0.152)
CR CAF32 x (S&P500)	1.469 (0.485)	2.219 (0.492)	0.016 (0.18)
CP CAF32 x (S&P500)	1.062 (0.125)	1.924 (0.53)	0.027 (0.161)
rtnXmkt-Linear x (S&P500)	1.055 (0.096)	1.927 (0.319)	0.024 (0.159)
rtnXmkt-WLS x (S&P500)	0.793 (0.009)	1.398 (0.023)	0.015 (0.136)
MOM (S&P500)			0.012 (0.201)
RESMOM (S&P500)			0.005 (0.183)
REV (S&P500)			0.023 (0.159)

Table 4.1: Optimization statistics for the investigated parametric system configurations as non-parametric benchmarks in the CRSP and S&P500 universes over the entire out-of-sample period 1995/01-2022/12. SPR denotes the Spearman correlation coefficient. Expanding window applied for the complex system configurations with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years.

To start this analysis we identify the differences between the linear and complex parametric system configurations. We first consider Table 4.1 alongside the Diebold-Mariano test statistics given in section D.2 for the MSE and MAE metrics. From these tables we conclude that the complex configurations consistently provide significantly worse MAE and MSE scores compared to the linear configurations in the CRSP universe. On the other hand, this inferiority is no longer clear and significant in the S&P500 universe. On the other hand, in Table 4.1 we observe that the complex configurations are able to achieve a higher Spearman correlation coefficient than the linear system configurations. Examining, the Diebold-Mariano test results in Table D.5 we observe that this difference is significant in the CRSP universe, yet Table D.6 also shows us that this significance does not hold up in the S&P500 universe. Specifically, we observe a {-65,-53, -77, -65}% decline in mean metric score in the S&P500 universe for the $CR-CLF$, $CP-CLF$, $CR-CAF$ and $CP-CAF$ configurations respectively and note that this decline is also visible for the linear and non-parametric signals. This general observation between these universes aligns with the works by Fama and French [2008], Fama and French [2018] and Hou et al. [2020], who remark a relatively higher frequency of theoretically exploitable market inefficiencies in lower market capitalization universes. What is more, is that Table D.5 and Table D.6 show us that the complex configurations are statistically indifferent to each other in terms of Spearman correlation coefficient in both universes, excluding the statistically significant inferiority of the $CR-CAF$ configuration. Therefore, in the remainder of this research we shall only consider the difference between $CR-CLF$ and $rtnXmkt-Linear$ when commenting on the differences between complex and linear configurations. We opt for this linear configuration because it is optimized according to the same procedure as the complex configurations. In addition, we opt for the $CR-CLF$ configurations amongst the complex ones as it is the 'non-extended' original/base configuration inline with the work *Charting by Machines* by Murray et al. [2021].

Furthermore, the mean scores in Table 4.1 and Diebold-Mariano results in Table D.5 and Table D.6 imply that the linear system configurations ($rtnXmkt-Linear$ and $rtnXmkt-WLS$) show Spearman correlation coefficients that are not statistically significant from the non-parametric MOM benchmark in both universes. For the linear configurations these results are not unexpected as they likely pick up on the same market inefficiencies as the non-parametric benchmarks and can only adjust the relative weighting of these market inefficiencies instead of picking up on other (nonlinear) ones. On the other hand, the general insignificant scores of the complex configurations compared to the linear and non-parametric benchmarks in the S&P500 universe for the complex configurations imply that these systems do not extract any additional beneficial market inefficiencies besides the linear ones. All in all, this result serves as evidence against the transfer learning capabilities of these complex systems in terms of parameter estimation in the CRSP universe and subsequent out-of-sample prediction in the S&P500 universe.

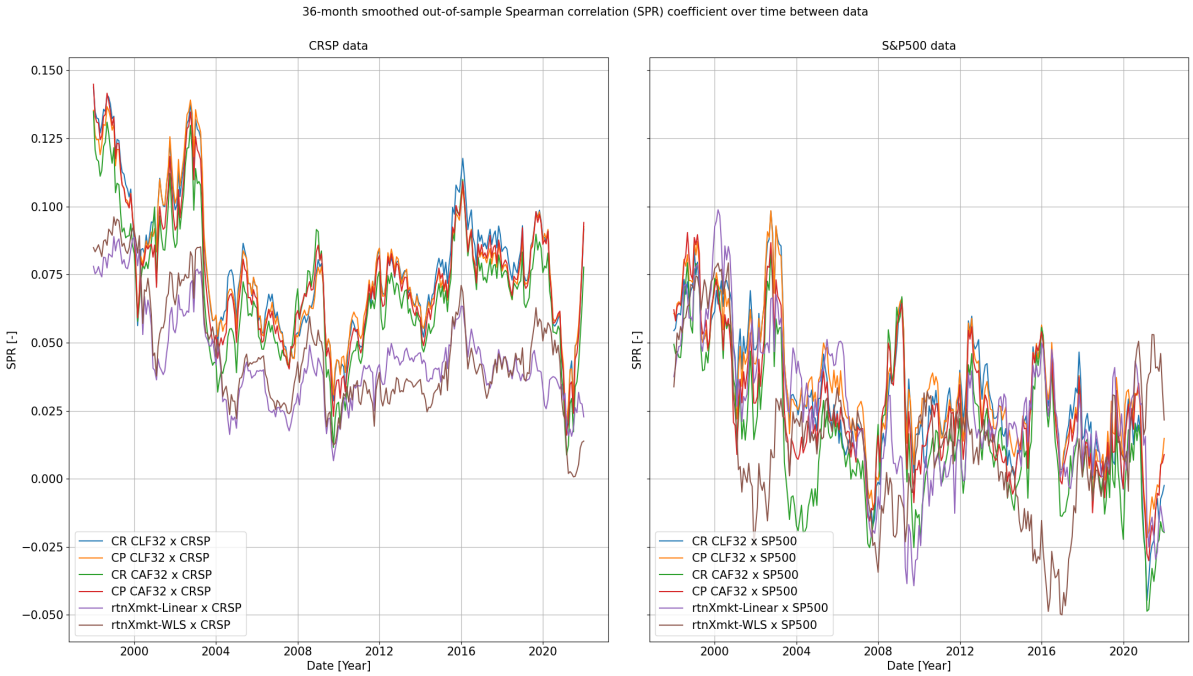


Figure 4.1: 36-month exponentially smoothed Spearman correlation coefficient over time in the CRSP and S&P500 universes over the entire out-of-sample period 1995/01-2022/12. Expanding window applied for the complex system configurations with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years.

To complete this section we consider Figure 4.1, which displays the smoothed Spearman correlation coefficient over time. The exponentially smoothed version is used to improve the intended general analysis over time as the crude scores are considered too noisy. If we examine this figure, we observe that the complex configurations are generally show larger coefficients than their linear counterparts. Furthermore, we note that between the CRSP and S&P500 universes a similar performance drop in smoothed score is visible through time and that this is most apparent when comparing the sub-periods $[1995, 2003]$ and $[2005, 2020]$ where a drop can be observed. On the other hand, there are clear differences. In the CRSP universe a rather stable score around ≈ 0.07 is apparent. In the S&P500 universe this decay does not seem limited to a drop, but a downward trend. This is most clearly observed by identifying that the extreme peaks and troughs have decreased smoothed scores over time. These observations in the S&P500 universe might indicate that this large market capitalization universe becomes increasingly efficient over time with regard to the exploited informational axes, leaving fewer market inefficiencies to be exploited.

4.2 Empirical rank distributions

In this section, we attempt to analyse the empirical asset ranking ability of the parametric configurations and the non-parametric benchmarks by examining their empirical rank distributions. We do this to extend the optimization analysis in a more granular sample-focused setting. We construct an empirical distribution map over the samples' predictions ($\hat{y}$) and target percentiles (y) in the out-of-sample period (1995-2022). For such a figure two cases can be considered, namely one where we construct this grid without considering the timestamp and the other were we do. These two cases can be visualized through the examples given in Figure 4.5 and Figure D.8 respectively. Thus, in the first case (Figure 4.5) such a map is based on the predicted

and target percentiles (i.e. grid size 100x100) of every sample regardless of its time stamp (i.e. no monthly time stamp considered). We annotate the mean predicted rank for that target decile and the standard deviation of the predicted decile for that target decile in black solid and dashed lines respectively. A similar approach can be used to attain the empirical distribution per timestamp i.e. a specific month. In this case, we achieve N_{months} of empirical distribution curves where the mean, 25% and 75% quantiles across these curves are presented in Figure D.8. Comparing Figure 4.5 and Figure D.8 or Figure 4.4 and Figure D.10, it should be clear that the general 'inverted-U' shape or 'X' shape in these figures is consistent through time. Therefore, the analysis in this section will focus on the versions of the empirical rank distribution maps without timestamp.

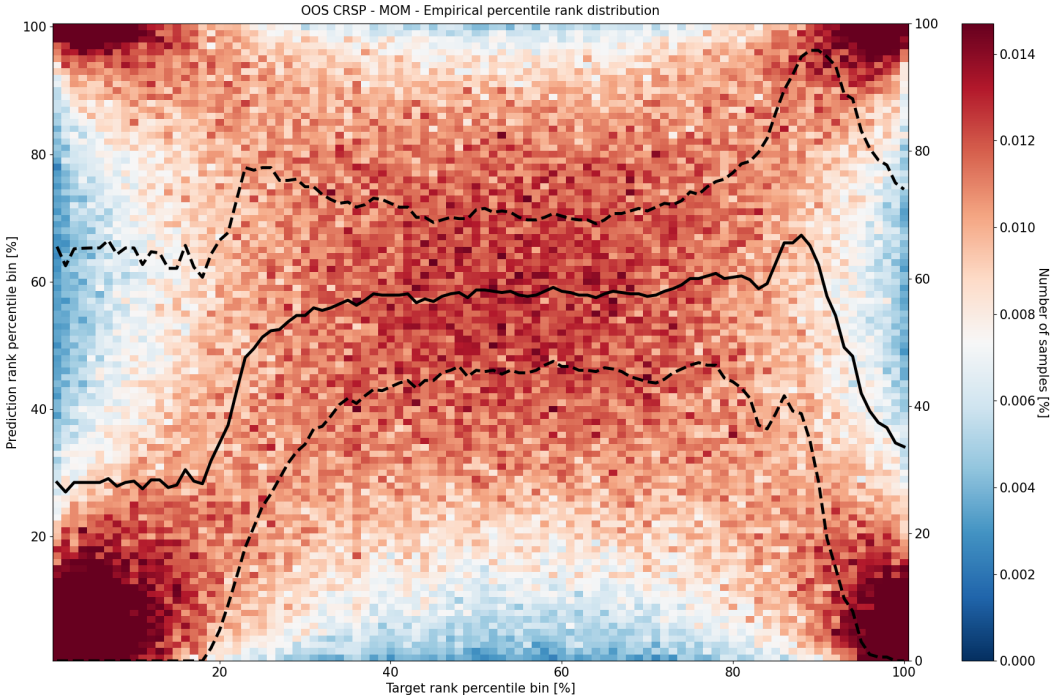


Figure 4.2: Empirical distribution grid based on target and prediction deciles for the CRSP and S&P500 universes in the out-of-sample period 1995/01-2022/12 for the MOM benchmark. Mean predicted rank for that target decile and the standard deviation of the predicted decile for that target decile in solid lines and dashed black lines respectively. Note, that a consistent colorbar numerical scale is used for all empirical distribution heatmaps across configurations.

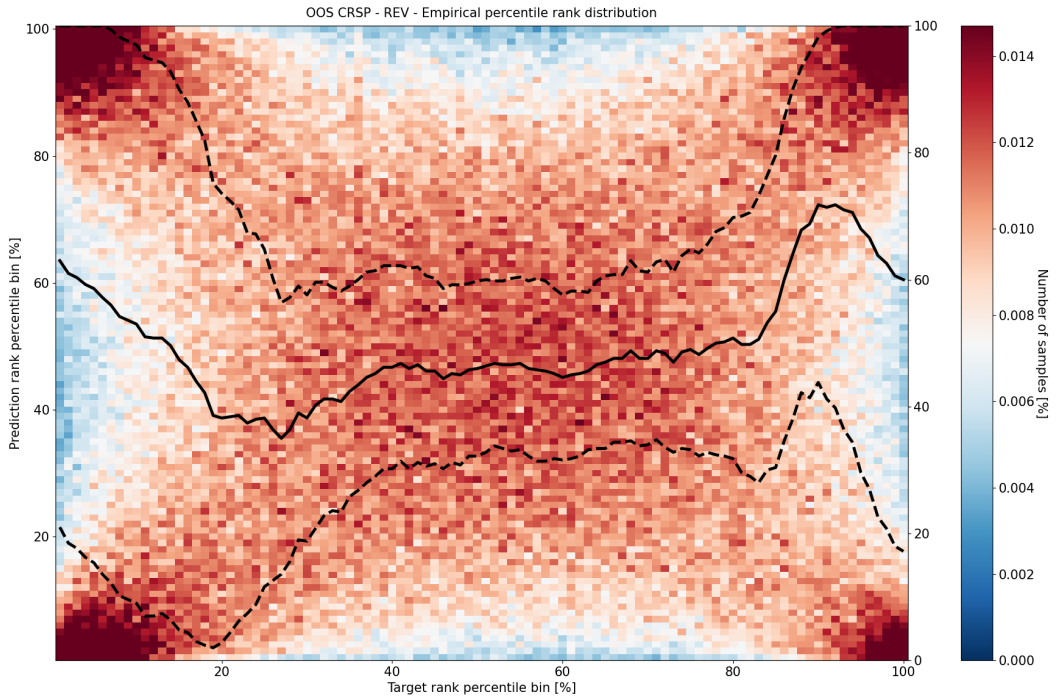


Figure 4.3: Empirical distribution grid based on target and prediction deciles for the CRSP and S&P500 universes in the out-of-sample period 1995/01-2022/12 for the *REV* benchmark. Mean predicted rank for that target decile and the standard deviation of the predicted decile for that target decile in solid lines and dashed black lines respectively. Note, that a consistent colorbar numerical scale is used for all empirical distribution heatmaps across configurations.

To start this analysis we consider the empirical distribution maps for the *MOM* and *REV* benchmark signals given in Figure 4.2 and Figure 4.3 respectively. For these benchmarks, two main clusters (i.e. distributional modes) can be identified from the predictive perspective ($\hat{y}$-axis) and an overall 'X'-like shape. For the *MOM* signal, the first predicted cluster/mode contains both samples from percentile ranges ($y = [0, 20]\%$,$\hat{y} = [90, 100]\%$) and ($y = [90, 100]\%$,$\hat{y} = [90, 100]\%$), where the coordinates are denoted as ($y_{percentile}$, $\hat{y}_{percentile}$). The second cluster presents a similar situation but mirrored, where assets from both distributional tails are ultimately predicted to belong to percentiles in the lower (y,$\hat{y} = [0, 20]\%$) range. For the *REV* signal, these clusters are similar but opposite in size; the upper cluster covers samples from the ($y = [0, 20]\%$,$\hat{y} = [80, 100]\%$) and ($y = [90, 100]\%$,$\hat{y} = [80, 100]\%$) ranges the and the lower cluster covers the ($y = [0, 10]\%$,$\hat{y} = [0, 10]\%$) and ($y = [90, 100]\%$,$\hat{y} = [0, 10]\%$) ranges. This double-cluster phenomenon is dubbed a 'bimodal' prediction structure due to the high empirical density in the tails of the predicted ($\hat{y}$) distribution. This phenomenon has already been identified for these benchmarks as well as various other factors/characteristics by Han [2022]. Practically, it is undesired signal property in terms of portfolio construction as the biggest winners and biggest losers will end up in both the predicted upper (D9 & D10) and bottom decile (D1 & D2) portfolios.

This bimodal structure should be interpreted as the signal's general inability to properly distinguish the assets with a tail (i.e. very low or very high) ranking. Notice that this distinguishing ability between tails is limited and asymmetric between the clusters' percentile ranges. Indeed, note that for the *MOM* signal we observe a relatively larger lower (i.e. low $\hat{y}$-coordinate) cluster range than for the upper cluster. This can be interpreted as this signal having a higher inclination to predict losers (with lower predicted tail ranks, $\hat{y} \approx 10$) over winners (with higher predicted tail ranks, $\hat{y} \approx 90$). For the *REV* signal we observe the opposite case, meaning this

signal has a higher inclination to predict winners over losers. On the other hand, it is important to note that the bimodal structure also means that the signal does contain the power to predict whether samples will belong to these tails (i.e. $(y \approx 10\ \%,\hat{y})$ or $(y \approx 90\%,\hat{y})$ range) or not (i.e. $(y = [20, 80]\%,\hat{y})$ range). In other words, the signal exhibits the power to distinguish assets with 'extreme' tail ranks from assets with non-tail (i.e. more center) ranks, but it cannot distinguish properly between these tails. Finally, we observe that for both benchmarks no clearly identifiable increasing trend pattern exists over the true percentile $(y = [35, 80]\%,\hat{y})$ range as indicated by the flat/horizontal solid black line. This means that these benchmark signals also do not exhibit a distinctive ability to consistently distinguish assets within this center rank range. Note that in this section we purposefully refrain from concluding on the ratio between the correct and incorrect samples within empirical clusters (and the heatmap in general) as we intend to exclusively analyse this on portfolio level in subsequent sections.

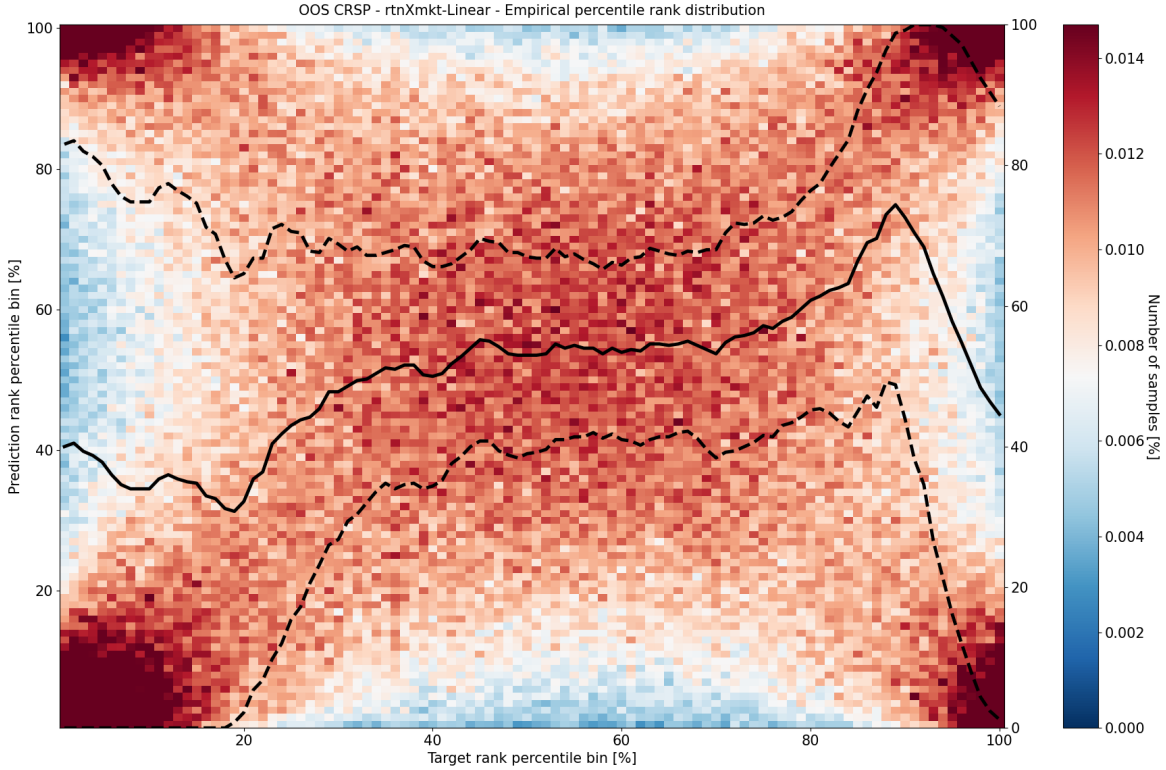


Figure 4.4: Empirical distribution for the $rtnXmkt - Linear$ system configuration over the period 1995-2022 in the CRSP universe. Expanding window applied for the linear system configuration with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years. Mean predicted rank for that target decile and the standard deviation of the predicted decile for that target decile in solid lines and dashed black lines respectively. Note, that a consistent colorbar numerical scale is used for all empirical distribution heatmaps across configurations.

We proceed with this analysis by considering the empirical distribution structure for the linear $(rtnXmkt-Linear)$ system configuration. The empirical rank distributions of this configuration is given in Figure 4.4 and aggregated over every month in Figure D.10. Examining Figure 4.4, we observe a distributional X-like pattern similar to that of the non-parametric MOM and REV benchmarks as given in Figure 4.2 and Figure 4.3 respectively. Deviation from this structure is not expected as the linear system can only adjust the relative weighting between the market inefficiencies instead of exploiting additional ones. Recall from section 4.1, that this configuration was unable to attain a significantly improved Spearman correlation coefficient compared to the MOM benchmark. In with this result, we observe that the empirical structure for this system

consists of a larger lower cluster resembles the MOM structure more than the REV signal.

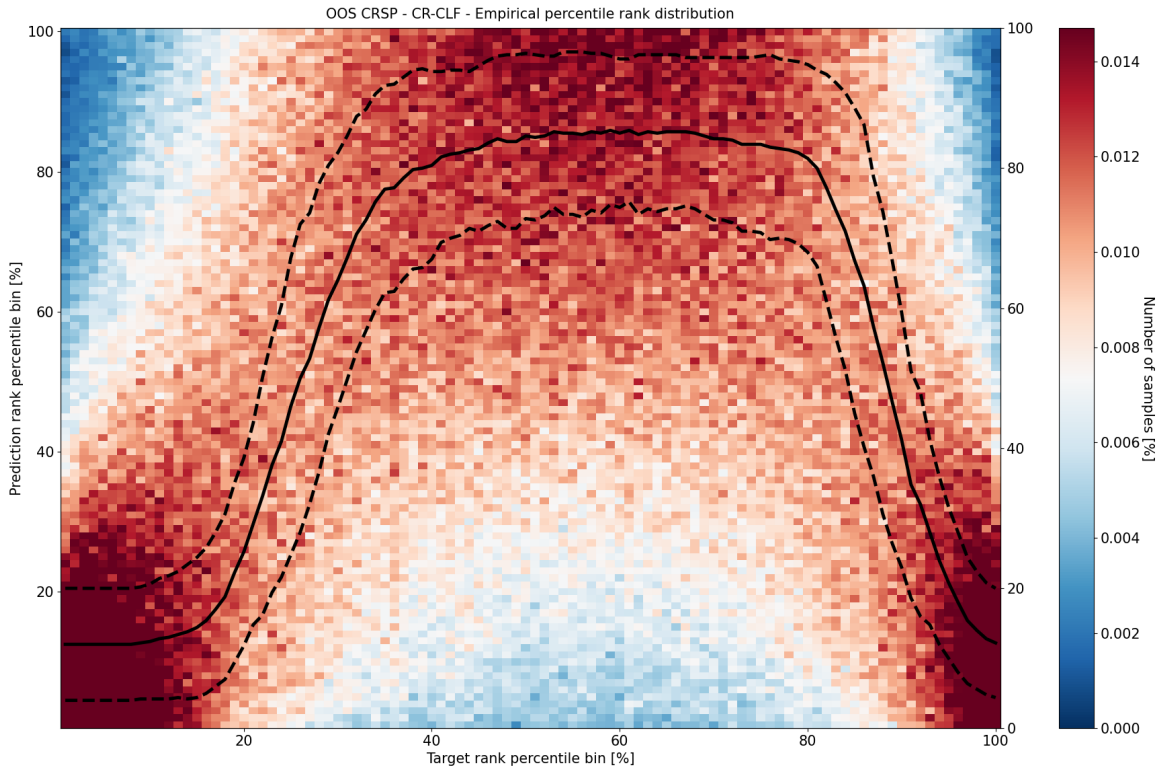


Figure 4.5: Empirical distribution grid based on target and prediction deciles for the CRSP and S&P500 universes in the out-of-sample period 1995/01-2022/12. Expanding window applied for the complex system configuration with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years. Mean predicted rank for that target decile and the standard deviation of the predicted decile for that target decile in solid lines and dashed black lines respectively. Note, that a consistent colorbar numerical scale is used for all empirical distribution heatmaps across configurations.

We continue this analysis by considering the empirical distribution structure for the complex system configurations given in Figure 4.5, Figure D.12, Figure D.13 and Figure D.14. We only present the map for the $CR-CLF$ here as their structures are indistinguishable in shape. For these complex configurations, we identify a different, but similarly bimodal prediction structure. In this case, the structure resembles an 'inverted-U' shape with two predictive modes. The first predicted mode/cluster contains samples from the $(y = [0, 20]\%, \hat{y} = [0, 20]\%)$ and $(y = [90, 100]\%, \hat{y} = [0, 20]\%)$ percentile ranges. The second cluster is less well defined, but approximately contains samples with $(y = [45, 75]\%, \hat{y} = [70, 100]\%)$ percentile range. Within this second cluster, the flat black lines indicate an inability to properly distinguish the true ranks of the assets contained in this cluster. Thus, notice that the first cluster essentially contain all samples with true extreme/tail rankings and the second cluster contains assets with true center ranks. This result clearly differs to that of the linear/non-parametric configurations in the fact that those configurations presented two clusters/predicted modes which contained a mix of assets from both true tails. On the other hand, it is not immediately apparent from these figures whether the complex configurations improve in distinguishing assets with extreme/tail rankings from those with non-extreme/center rankings. What is clear is that these complex configurations have the unique ability to allocate samples with true tail ranks exclusively to a single cluster.

All in all, the empirical structures presented Figure 4.4 and Figure 4.5 constitute three initial findings. First of all, all considered signals (parametric & non-parametric) exhibit some ability to distinguish assets with true center ranks from those with extreme/tail ranks. Secondly, the complex configurations do not show an immediate and explicit improved ability to distinguish assets with true extreme/tail rankings from each other. These complex configurations exclusively predict assets with true extreme/tail ranks as losers (i.e. low $\hat{y}$-coordinate). This means they differ from the linear/non-parametric signals which exhibit a split empirical structure with two predictive clusters containing a mix of assets from both true tails. From an econometric point of view, this means that any additional (nonlinear) market inefficiencies identified by the complex systems do not substantially improve the systems ability to distinguish extremely positive from extremely negative ranks. Rather, they are uninformative or they might further improve distinguishing between extreme/tail and non-extreme ranks. Thirdly, this exclusive 'loser' (i.e. low $\hat{y}$-coordinate) prediction of all true extreme ranks displayed by the complex configurations apparently achieves a significantly higher Spearman correlation coefficient score. This statement follows from the results in Table 4.1 and Table D.5 discussed in section 4.1 regarding the optimization statistics. Specifically, this shows that the Spearman correlation coefficient optimized in this research allows for the near-complete 'sacrifice' of one distributional tail. Specifically, this sacrifice is present in Figure 4.5 by recognizing that almost none of the assets with a higher true percentile ranks (i.e. $y > 90$) are correctly predicted as high percentile ranks (i.e. large $\hat{y}$-coordinate). Hence, the system will accumulate a substantial losses in this upper true distributional tail. It is important to recognize that this is not an erroneous outcome, because the Spearman correlation coefficient scores overall performance and is not concerned about the loss asymmetry/imbalance of a solution. However, it serves as a stark reminder that optimization of this metric means such a 'sacrificial'/asymmetric solution is a possible outcome.

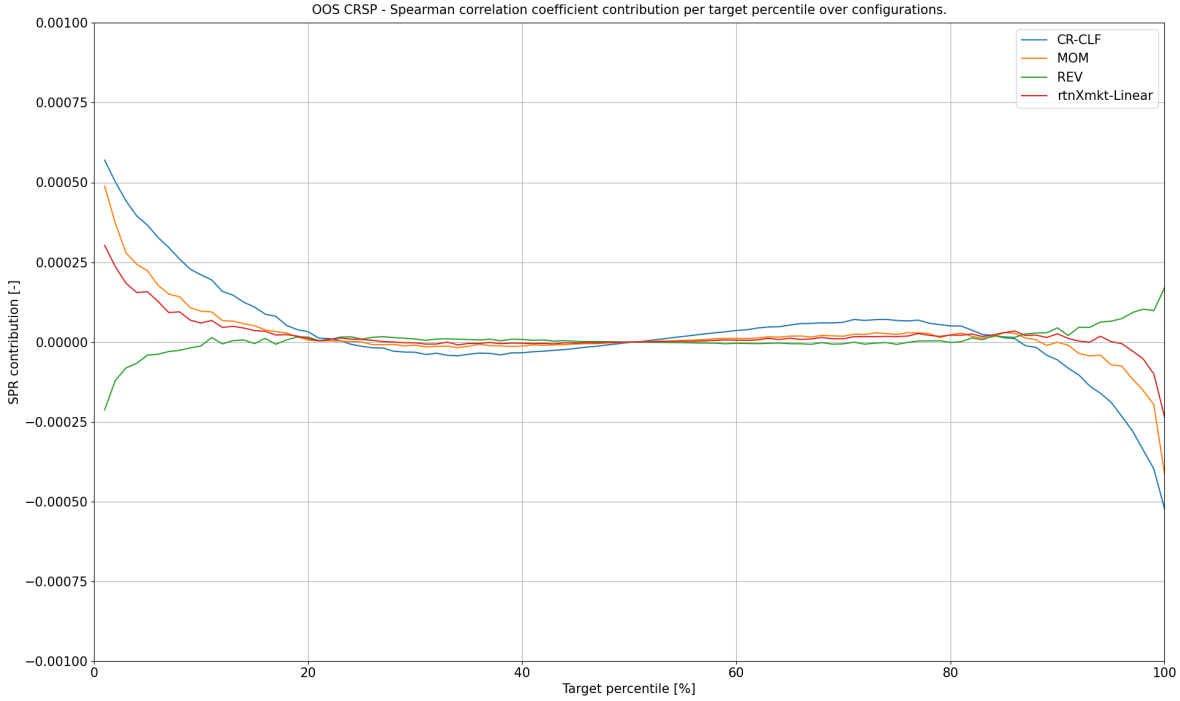


Figure 4.6: Mean Spearman correlation coefficient contribution aggregated per target percentile for the MOM, REV benchmarks and $rtnXmkt - Linear$, $CR - CLF$ system configuration over the out-of-sample period 1995-2022 in the CRSP universe.

To complete this section we attempt to the quantify the impact of the empirical predictive

structure to the Spearman correlation coefficient at a rank percentile level. To this end, we consider the *Spearman correlation coefficient contribution* per target percentile as visualized in Figure 4.6. This figure contains the aggregated mean contribution per target percentile over all months in the out-of-sample period. In other words, such a figure visualizes the average contribution of all samples with a specific true percentile to the total Spearman correlation coefficient. Mathematically, we can represent the target percentile-specific ($\mathcal{L}_{SPR_{TP}}$) and sample-specific ($\mathcal{L}_{SPR_{t,i}}$) Spearman correlation coefficient contribution as,

$$\mathcal{L}_{SPR_{TP}} = \frac{\sum_{t}^{N_{months}} \sum_{\forall j \in I_{TP,t}} \mathcal{L}_{SPR_{t,j}}}{N_{months}}$$
$$\text{where,} \tag{4.1}$$
$$\mathcal{L}_{SPR_{t,i}} = \frac{\tilde{y}_{t,i} \cdot \tilde{\hat{y}}_{t,i}}{\sigma_{\mathbf{y}_t} \sigma_{\hat{\mathbf{y}}_\mathbf{t}}}$$

where $\tilde{y}_{t,i}$ and $\tilde{\hat{y}}_{t,i}$ indicates the demeaned and ranked target and prediction for asset i at time step t respectively. $\sigma_{\mathbf{y}_t}$, and $\sigma_{\hat{\mathbf{y}}_\mathbf{t}}$ denote the standard deviation of the array of ranked targets and ranked predictions for time step t respectively. N_{months} denotes the number of months in the out-of-sample period. TP is the mathematically notation for rank percentile under consideration in the integer range $[1, 100]$. Consequently, $I_{TP,t}$ denotes the subset of asset indices i at time step t contained in target percentile TP. Besides the aggregated mean contribution, figures concerning the contribution per target percentile for every month separately are given in Figure D.17, Figure D.18, Figure D.21 and Figure D.22 for the MOM, REV, $rtnXmkt-Linear$ and $CR-CLF$ configurations respectively. Finally, for these mean contribution arrays (SPR_{TP}) we present aggregated statistics in Table 4.2.

Statistics	CR-CLF	MOM	REV	$rtnXmkt - Linear$
sum ($\approx \mathcal{L}_{SPR}$)	0.00292	0.00200	0.00054	0.00198
mean	0.00003	0.00002	0.00001	0.00002
std	0.00015	0.00010	0.00004	0.00006
max	0.00057	0.00049	0.00017	0.00030
min	-0.00052	-0.00041	-0.00021	-0.00023
$sum(x\|x > 0)$ i.e. 'sum of positives'	0.00618	0.00339	0.00129	0.00247
$sum(x\|x < 0)$ i.e. 'sum of negatives'	-0.00325	-0.00139	-0.00075	-0.00049
Custom statistics	**CR-CLF**	**MOM**	**REV**	$rtnXmkt - Linear$
$\frac{sum(x\|x>0)}{\|sum(x\|x<0)\|}$ i.e. 'ratio of sums'	1.90153	2.43885	1.72000	5.04082
Sharpe ratio	0.18975	0.20657	0.13470	0.34328
Sortino ratio	0.25444	0.26441	0.13677	0.42704

Table 4.2: Mean Spearman correlation coefficient contribution statistics aggregated per target percentile over the out-of-sample period 1995-2022 in the CRSP universe. Statistics obtained through aggregation of results presented in Figure 4.6.

If we analyse Figure 4.6, we observe similar contribution structure for the linear configuration and the MOM benchmark. This aligns with the similar empirical structure observed in Figure 4.4 and Figure 4.2, where we noted that both signals show an asymmetric inclination to predict losers (i.e. low $\hat{y}$-coordinate). Figure 4.6 shows us that this inclination invokes a positive contribution in the left true tail (lower ranks) and a negative contribution in the right true tail (higher ranks). Table 4.2 shows us that this inclination is beneficial overall, negative contributions

($sum(x|x < 0)$) are offset by greater positive contributions ($sum(x|x > 0)$). For the REV this empirical structure is similar, but all the observations are reversed compared to the MOM signal. Note that although the contribution structure of the linear and MOM signal are similar in sign consistently, the linear configuration is lower in magnitude. This observations aligns with the expectation that the linear configuration optimizes the relative weighting between the inefficiencies captured by MOM and REV rather than identifying additional ones. On the other hand, we observe that the contribution structure for the complex configuration is consistently greater in magnitude. Recall from Figure 4.5 that this configuration sacrifices all high tail ranks for the correct prediction of all lower tail ranks. Consequently, in Figure 4.6 we observe the relatively large positive contribution in the y = [0,20] % target percentile range and large negative contribution in the y = [85,100] % range. Interestingly, the contributions in the true center rank (y =[20,85]%) range are more pronounced in magnitude. While for the linear and non-parametric signals this range carries a relatively low total contribution, for this complex system the sum of positive contribution over the y ≈[20,50]% range clearly offsets the sum of negative contribution over y ≈[50,85]% . All in all, in Table 4.2 we indeed observe that the unconditional sum of the complex configurations is superior and that the 'sacrificial' solution is more asymmetric/imbalanced than that of the other signals. This asymmetry is clearly captured in statistics such as the conditional contribution sums (sum($x|x > 0$) & sum($x|x < 0$)), standard deviation, minimum and maximum statistics.

We end this section by presenting the conditional sum ratio ('ratio of sums'), Sharpe and Sortino ratio over the Spearman contribution array in Table 4.2. These are included to provide alternative statistics to score the relationship between predicted and true asset rankings. Contrary to the Spearman correlation coefficient, these alternative statistics also score the asymmetry/imbalance of the solution. Optimization of such statistics has been attempted, yet resulted in unstable and/or vanishing gradients. Therefore, further investigation into the virtues of optimizing these metrics is left to future work. If we examine these statistics in Table 4.2, we observe that the linear system is consistently superior, suggesting that this configuration achieves a better trade-off between positive and negative contributions.

4.3 Cross-sectional predictive ability

In this section, we attempt to further analyse the empirical asset ranking ability of the complex configurations through an investigation into their predictive power in the presence of other signals. Therefore, we employ the use of Fama-Macbeth regressions. We apply this in the cross-section of assets over the entire out-of-sample period for the CRSP universe. Similar to previous sections, we focus on the $CR-CLF$ configuration due to the high resemblance across the complex configurations.

Recall from the empirical rank distribution figures in section 4.2, that the complex configurations differ from the linear/non-parametric signals in the ability to allocate samples with true tail ranks to a single lower prediction cluster. This means that these configurations achieve greater similarity in terms of the prediction $\hat{y}$ for assets with true tail rankings (i.e. low or high y). It is possible that this 'allocation' ability and the consequent empirical 'inverted U'-shape stems from a large negative contribution attributed to a feature that approximates the absolute magnitude of the target variable. Mathematically, this can be represented as,

$$\hat{y}_{t,i} \propto \xi \cdot \hat{u}_{t,i} \qquad \text{where} \quad u_{t,i} = |y|_{t,i} \tag{4.2}$$

where $y_{t,i}$ and $\hat{y}_{t,i}$ are the true and predicted rank of asset i at time step t respectively. $u_{t,i}$ and $\hat{u}_{t,i}$ denote the true and predicted absolute rank respectively. ξ denotes the coefficient attributed

to the absolute rank prediction variable. Thus, we hypothesize that these configurations acquire an accurate approximation of the absolute rank $\hat{u}_{t,i}$ and that this feature is attributed a negative contribution to the ultimate rank prediction ($\hat{y}_{t,i}$) i.e. $\xi < 0$. Given this hypothesis, we attempt to interpret this $u_{t,i}$ variable. From an econometric point of view, one can identify a close relationship between $u_{t,i}$ and the latent volatility $\sigma_{t,i}$ of an asset. If we entertain this relationship, we assume that the complex configurations can indeed acquire an accurate approximation of $u_{t,i}$, i.e. $\hat{u}_{t,i}$. The reason for this is two-fold. First of all, it is well established in econometrics that volatility of equity returns can adequately be predicted through model definitions such as the Generalized Auto-regressive Conditional Heteroskedasticity (GARCH) model proposed by Bollerslev [1986]. Secondly, Goodfellow et al. [2016] explains/demonstrates the ability of complex neural networks to learn nonlinear relationships such as the squared asset return, $r^2_{t,i}$. Thus, these complex configurations have the ability to establish a $\hat{u}_{t,i}$ feature and this feature forms an adequate approximation of $u_{t,i}$. Here, we interpret $u_{t,i}$ as the asset volatility assume it is adequately predictable.

In the remainder of this section, we intend to formally conclude on this hypothesis mathematically presented in Equation 4.2. Therefore, we introduce an additional volatility-oriented signal based on the squared returns. Similar to the $CR-CLF$ and $rtnXmkt-Linear$ signals, in this signal we also opt for the use of excess market returns to reduce market timing effects as addressed in section 2.2. The use of these corrected returns implies an assumed $CAPM-\beta$ coefficient equal to 1. Hence, we interpret this volatility signal as an approximation for the idiosyncratic rather than systemic risk of the asset. However, we recognize that this interpretation is imperfect, due to the assumed rather than estimated $CAPM-\beta$ coefficient. We define this volatility-oriented signal as the 12-month exponentially weighted volatility estimate, $\hat{\sigma}_{EWM12_{t,i}}$. This can be mathematically represented as,

$$\hat{\sigma}_{EWM12_{t,i}} = \sum_{j=t-12}^{t-1} (1-\zeta)^{t-1-j} \cdot \zeta \cdot (r_{j,i} - r_{j,m})^2 \tag{4.3}$$

where $r_{t,i}$ denotes the return of asset i at time step t. $r_{t,m}$ represents the equally-weighted market return at time step t. We opt for a ζ decay rate magnitude of 0.15, resembling a common magnitude of the α_1 innovation term coefficient in a conventional GARCH(1,1) model as proposed by Bollerslev [1986]. We note that in practice, the conclusions of this section are quite robust to this parameter. We attribute this robustness to the focus of this analysis on the rank implied by this signal instead of the absolute magnitude. In the remainder of this analysis, this additional signal is referred to as $volXmkt-EWM12$.

	r_{norm}	$CR-CLF$	$rtnXmkt-Linear$	$volXmkt-EWM12$
r_{norm}	1	0.081 (0.112)	0.044 (0.08)	-0.074 (0.144)
$CR-CLF$		1	0.465 (0.166)	-0.734 (0.078)
$rtnXmkt-Linear$			1	-0.095 (0.176)
$volXmkt-EWM12$				1

Table 4.3: Mean and standard deviation for Spearman correlation coefficients between ranked signals and target variables, computed over every month. Lower triangular section omitted. Monthly statistics obtained from CRSP data in the out-of-sample period of 1995/01-2022/12, totalling 335 months. Inconsistent sample size over months as shown in Figure 3.1. Expanding window applied for the $CR-CLF$ and $rtnXmkt-Linear$ system configurations with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years.

To continue, we consider the Spearman correlation coefficient statistics between the signals and target in Table 4.3. Note that the correlation between the target r_{norm} and the $CR-CLF$

and $rtnXmkt - Linear$ signals was already presented in Table 4.1, but is shown here again. In Table 4.3, we observe negative correlation between the target (i.e. next period rank) and $volXmkt - EWM12$ signal. Furthermore, we observe a stable and substantially negative mean correlation between the $CR - CLF$ and $volXmkt - EWM12$ signals. This observation aligns with our hypothesis mathematically presented in Equation 4.2. Indeed, from this result it is likely that the complex configurations identify a $u_{t,i}$ feature and that this feature can be interpreted as volatility. Moreover, we notice a mean positive correlation between the $CR - CLF$ and $rtnXmkt - Linear$ signals which is likely based on the comparable lower predicted cluster of assets with true tail ranks in both signals as discussed in section 4.2. It is important to acknowledge the difference in the mean correlation magnitude between the $rtnXmkt - Linear$ and $volXmkt - EWM12$ signals and the $rtnXmkt - Linear$ and $CR - CLF$ signals (i.e $-0.095(0.176)$ vs. $0.465(0.166)$). This is because this difference implies that the $CR - CLF$ signal combines both these signals instead aligning only with volatility.

To complete this section, we conduct two Fama-Macbeth style regressions. In the first regression, we consider whether the three prediction signals explain the cross-sectional excess-market returns. This approach is equal to that of a conventional Fama-Macbeth procedure, except for the use of excess-market instead of crude returns. This correction is used anyway to invoke consistency as all signals are based on these corrected returns. For the second regression, we consider whether the three prediction signals explain the cross-sectional ranks. Note that for this second regression, we apply the 'Gaussian ranking' transformation procedure to all input and target variables involved in the regression. This procedure is introduced in subsection 2.3.1 and used for the target variable r_{norm} definition in Equation 2.2. In this procedure, we rank variables to uniform distribution (bounded by $(0, 1)$) after which we transform them to standard normal through the inverse Gaussian CDF mapping (Φ^{-1}). Consequently, we omit an intercept term from this second regression as this transformation procedure removes any existing level in these variables. The results of the first and second regression are presented in Panel A and Panel B of Table 4.4 respectively.

Panel A: return target

Inputs \ Target	$rtnXmkt$	$rtnXmkt$	$rtnXmkt$	$rtnXmkt$	$rtnXmkt$	$rtnXmkt$	$rtnXmkt$
Intercept	0.003 (0.057)	0.003 (0.073)	0.003 (0.081)	0.003 (0.058)	0.003 (0.078)	0.003 (0.080)	0.003 (0.080)
$CR - CLF$	0.002 (2.018)			0.000 (1.327)	0.006 (1.837)		0.004 (0.997)
$rtnXmkt - Linear$		0.004 (1.633)		0.004 (0.787)		0.005 (1.696)	0.003 (0.720)
$volXmkt - EWM12$			-0.002 (-0.956)		0.006 (0.595)	-0.002 (-0.752)	0.005 (0.307)

Panel B: rank target

Inputs \ Target	r_{norm}	r_{norm}	r_{norm}	r_{norm}	r_{norm}	r_{norm}	r_{norm}
Intercept							
$CR - CLF$	0.081 (5.102)			0.075 (4.128)	0.057 (2.487)		0.031 (1.062)
$rtnXmkt - Linear$		0.047 (2.938)		0.013 (0.689)		0.043 (2.682)	0.031 (1.492)
$volXmkt - EWM12$			-0.074 (-4.676)		-0.032 (-1.365)	-0.071 (-4.373)	-0.049 (-1.808)

Table 4.4: Fama-Macbeth style regression mean and t-statistic results. Regression results for next periods excess-market returns (rtnXmkt) target in panel A and for 'Gaussian ranked' returns (i.e. r_{norm}, normalized rank definition) in panel B. Both panels utilize Newey-west standard errors. For panel A, conventional standardization is applied to input variables and intercept term is included. For panel B, *Gaussian ranking* procedure is applied to both target and input variables. Hence, intercept term is omitted for this case. Both panels based on CRSP data in the out-of-sample period of 1995/01-2022/12, totalling 335 months. Inconsistent sample size over months as shown in Figure 3.1. Expanding window applied for the $CR - CLF$ and $rtnXmkt - Linear$ system configurations with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years.

To start, we consider Panel A in Table 4.4 and observe that the $CR - CLF$ predictions is the only signal attaining significant t-stat at a 5% level when regressing the target on a single signal. However, we also observe that this predictive power is subsumed in the presence of the linear system's predictions. This first analysis extends the analysis conducted in Murray

et al. [2021], who show through similar Fama-Macbeth regressions that their $MLER$ system's predictions (dubbed $CR-CLF$ in this research) remain significant after including the set of cumulative returns (CR inputs as defined in subsection 2.3.1) as additional regressors. The general significance observed in Panel B, but not in Panel A can have two possible explanations. First of all, the parametric systems are optimized to predict ranks and not returns. Secondly, in the regression on returns the relationships between variables is likely more noisy due to outliers and nuances in scale. On the contrary, in the ranked regression the variables have symmetric and consistent scales which might lead to an amplification of any existing relationships.

To continue, we consider Panel B in Table 4.4. We observe that all three signals are significant in a single independent variable regression case, with coefficient magnitudes inline with Table 4.3. Moreover, the significantly negative coefficient for the $volXmkt-EWM12$ signal indicates a negative relationship between the future rank and idiosyncratic risk as estimated by exponentially weighted excess market return. This aligns with the works by Ang et al. [2006] and Chen and Petkova [2012], who identify this same negative relationship between idiosyncratic risk and returns in the cross-section. Therefore, we reflect to the hypothesis in Equation 4.2 and conclude from these insights that this hypothesized proportional relationship holds and is likely due to its beneficial effects on the predictive ability of $\hat{y}_{t,i}$ with regard to $y_{t,i}$.

We proceed with the remaining observations of Panel B in Table 4.4. When we consider both $rtnXmkt-Linear$ and $volXmkt-EWM12$ in the regression, we observe that both signals remain significant. This is to be expected as these signals contain information from different information sets, namely related to the first and second moment respectively. This statement is supported by their low correlation in Table 4.3. Furthermore, we note that the predictive power of $CR-CLF$ is maintained in the presence either the $rtnXmkt-Linear$ and $volXmkt-EWM12$ signal, but not if both are included in the regression. Therefore, we conclude that the $CR-CLF$ signal contains no additional predictive power in the cross-section of ranks over the signal already contained in the linear and volatility related aspects. In other words, if the $CR-CLF$ signal contains other nonlinear and/or interactive aspects, their predictive power in the cross-section is limited. On the other hand, the fact that in this case no signal is significant as well as the substantial correlations in Table 4.3 indicates a potential high level of multicollinearity. Attempts to overcome this potential multicollinearity such as the application of principle components analysis, where unsuccessful in further identifying and decoupling the nonlinear/interactive part of the $CR-CLF$ signal. Therefore, this analysis is omitted and left as future work.

All in all, through the analysis of this section we find a limited cross-sectional predictive ability of the $CR-CLF$ signal in the presence of a linear and volatility based signal. On the other hand, the analysis showed that predictive power of the $CR-CLF$ signal is subsumed after including both and not just either the linear- and volatility-based predictions. Furthermore, we hypothesized a volatility-oriented feature could explain the empirical 'inverted U'-shape for the complex configuration. Indeed, we observed high correlation between this complex configuration and a volatility-oriented signal. We denoted that complex neural networks present the ability to access volatility related features and that these features themselves are inherently predictable in equity returns. Finally, we confirm our hypothesis through the fact that the established negative relationship between idiosyncratic volatility and returns by Chen and Petkova [2012] is clearly apparent in our regression results.

4.4 Decile portfolio performance

Starting in this section we continue the analysis in our research on portfolio level. In this section, we start that analysis with an examination of the equally-weighted decile portfolios. The portfolios' positions are based on the rank predictions from the parametric and non-parametric

signal definitions according to the methodology discussed in section 2.1. It is important to recognize how this analysis on portfolio level is connected to the results presented in the empirical distribution heatmaps of section 4.2. The empirical heatmaps provide us with a grid of predicted and true ranks (i.e. $(y, \hat{y})$), where we recall that the 'true rank' is based on the next periods return. On the other hand, the analysis on portfolio level considers the aggregated asset performance achieved by a portfolio constructed on the predictions ($\hat{y}$) in terms of various performance metrics. In addition, notice that this analysis is less granular since it considers deciles instead of percentiles. With regard to performance metrics, we report the annualized mean return, volatility and the Sharpe ratio (all excess risk-free rate) for the equally-weighted decile portfolios for both universes in Figure 4.7, Figure 4.8 and Figure D.4 respectively. The accompanying numerical values can be found in Table D.7. To aid in a comparative analysis we also present the performance of the reversal (REV) and residual momentum ($RESMOM$) factor portfolio. The latter is included instead of its MOM counterpart, due general superiority in terms of cumulative return and drawdowns in the CRSP universe as is visible in Figure 3.2 and Figure 3.3. Note that the decile performance statistics for the non-parametric market, MOM, $RESMOM$ and REV portfolios are provided in section D.1.

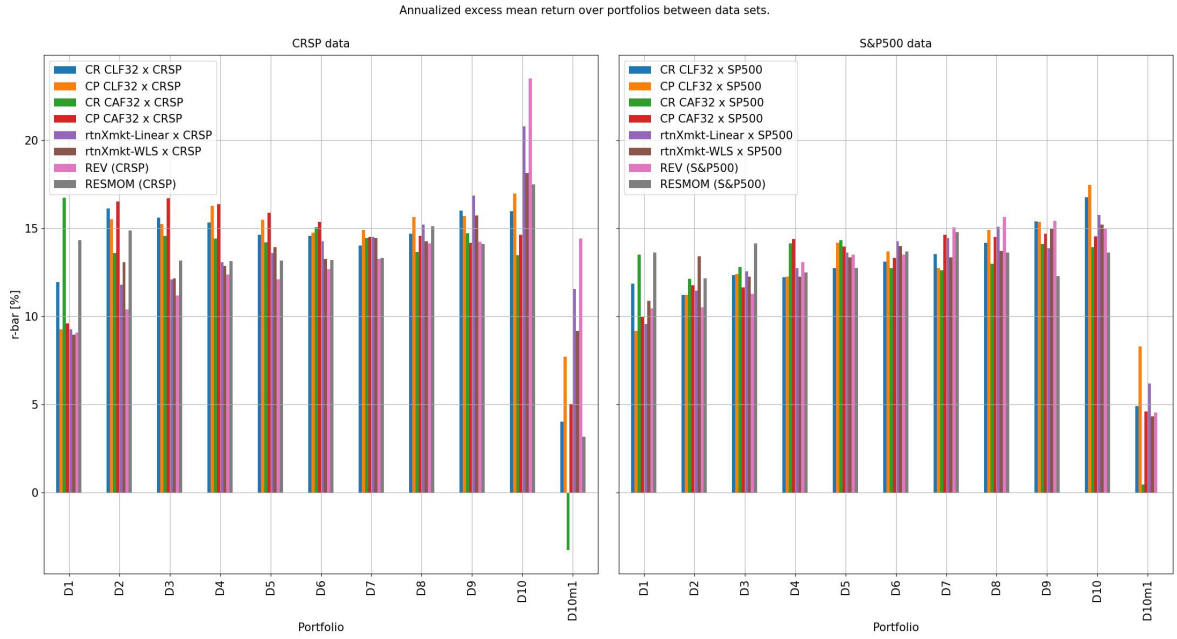


Figure 4.7: Annualized mean excess (risk free rate) return across equally-weighted deciles in the CRSP and S&P500 universes in the out-of-sample period 1995/01-2022/12. Expanding window applied for the complex system configurations with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years. Histogram values correspond to those in Table D.7.

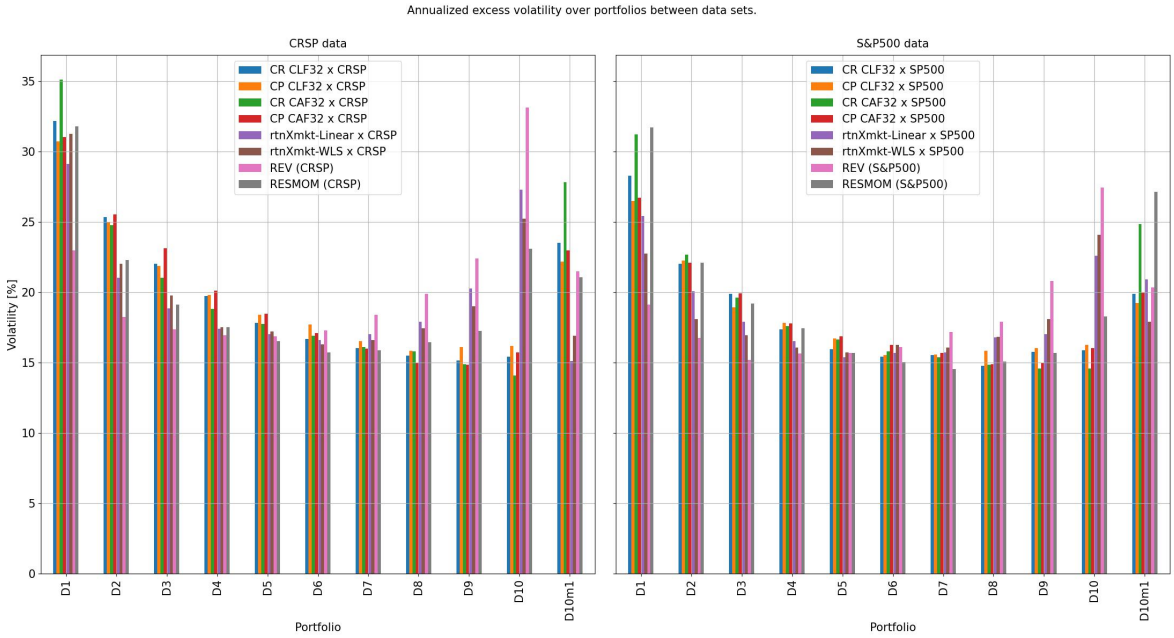


Figure 4.8: Annualized excess (risk free rate) volatility across equally-weighted deciles in the CRSP and S&P500 universes in the out-of-sample period 1995/01-2022/12. Expanding window applied for the complex system configurations with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years. Histogram values correspond to those in Table D.7.

To start this analysis, in these figures we observe that the linear parametric configurations ($rtnXmkt - WLS$ & $rtnXmkt - Linear$) resemble each other closely in terms of general performance metric structure across the deciles, yet show some numerical dissimilarities at specific deciles. This observation is consistent across universes. Hence, we infer that implied estimated parameters generally match which is expected due to the equivalence between the optimization of the MSE on ranks and Spearman correlation coefficient as discussed in subsection 2.4.1 and addressed in chapter B. Furthermore, notice from Figure 4.7 that these linear configurations as well as the non-parametric REV signal achieves the lowest and highest mean excess return in D1 and D10 respectively. This observation implies that the D10 portfolios on average correctly consist of winners with larger cumulative positive returns than the cumulative negative return of some of the losers. This statement does not hold on average in reverse for D1, because this portfolio does not attain a negative cumulative return on average. Since we observe a clearly positive mean return for the D10-D1 portfolio on average, this still means that the asymmetric mix of tail ranks in the two clusters identified in Figure 4.3 and Figure 4.4 indeed results in beneficial results on portfolio level. In addition, notice that for the linear parametric configurations an increasing trend over the deciles can be identified. Hence, we deem the relative weighting of market inefficiencies implied by the estimated parameters an adequate fit. Relating this to Figure 4.4, this means that although there is noisier mix of true ranks within the predicted center percentiles, the system shows an adequate ranking ability on decile portfolio level.

Next, we consider the four complex configurations and note that across the three portfolio performance metrics these configurations present a generally identical structure over the deciles. The clear exception is exhibited by the $CR{-}CAF$ configuration which depicts consistent inferior performance. This general alignment serves a further evidence that the solution implied by these configurations is not substantially different. Hence, in the remainder of this analysis we consider the performance of $CR - CLF$ configuration as the representative solution obtained

from our machine-learning based complex configurations. For the $CR - CLR$ configuration, we observe the lowest return on average in the D1 portfolio. Although the mean return is also not negative on average, we still deem this beneficial feature of this configurations. Similarly to the linear configuration, the imbalanced mix of true tail ranks observed in Figure 4.5 for the lower predicted percentiles apparently still constitutes a positive cumulative return. Moreover, note in this case the D10 portfolio does not clearly achieve the largest cumulative return on average. Recall from Figure 4.5, that the predicted upper percentiles contain a higher relative empirical density of center ranks than the linear/non-parametric configurations. Furthermore, the $CR - CLR$ configuration fails in obtaining an increasing excess-return pattern over the portfolios. In fact, we observe a peak in D2 and noisy pattern between from D3 to D10. Therefore, from the results in Figure 4.7 we conclude that this empirical 'inverted-U' shape exhibited by the complex configurations is undesirable at the decile portfolio level in terms of this performance metric. Specifically, recall from section 4.3 that complex configurations achieve this empirical 'inverted-U' shape through the use of volatility-oriented features. Therefore, we further conclude that the use of these volatility-oriented features for rank prediction purposes might be beneficial on cross-sectional level, yet not on the aggregated portfolio level. Generally, this serves as evidence against the general ranking ability of these configuration at an aggregate portfolio level, which seemingly stands in contrast to the significant superiority in terms of Spearman correlation coefficient observed in Table 4.1 compared to the linear/non-parametric signals. From an econometric practitioners perspective, this means that the use of this metric as a tool to predict/measure differences on aggregated portfolio level is limited. Specifically, we do not recommend the use of this metric to measure/predict portfolio performance for predictions from systems with distinctly different empirical structures such as the 'X'-shape and 'inverted U'-shape.

We extend this analysis by considering the annualized excess volatility across deciles in Figure 4.8. In this case, we observe a prominent monotonically decreasing volatility skew for the complex configurations. Thus, the predictions from these systems invoke a volatility-ranked structure across decile portfolios. On the other hand, the linear and non-parametric signals do not present this volatility ranking property. Instead, they show a non-uniform but more symmetric volatility structure with an increase in magnitude at the edge deciles. For the complex configurations, we can explain this volatility skew by considering the regression and correlation results in section 4.3. In this section, we establish that the complex configurations indeed use volatility-based features in their ranking. Specifically, we identify that these volatility features command a negative contribution to the asset rank prediction in line with the work by Chen and Petkova [2012], which results in the observed volatility skew at portfolio level. Furthermore, we can also connect this observation to the empirical heatmaps in section 4.2. For the complex case, the 'inverted U'-shape implies all extreme ranks are allocated to the lower predicted percentiles. On a portfolio level this means that the assets in the lower predicted percentiles present ranks/returns at the tails of the true distribution. In addition, in this same empirical structure we notice that for a higher predicted percentile the range of true ranks reduces (centered around the approximate true 50th rank percentile). This reduction is not indefinite, but continues until the approximate predicted 80th percentile after it stabilizes to the true percentile rank $y \approx [40, 80]\%$ range as is visible in Figure 4.5. On the other hand, the empirical 'X'-shape for the linear and non-parametric signals implies a more balanced/symmetric division of ranks/returns over the predicted percentiles. Therefore, we conclude on two findings exhibited by the complex configurations. First, the prominent volatility-skew is directly related to the asymmetry/imbalance in (extreme) rank allocation at lower predicted percentiles. Secondly, the decreasing and stabilizing trend in volatility is related to the decreasing and stabilizing range of true ranks for a higher predicted percentile.

For completeness, we compare performance between the CRSP and S&P500 universes. For the

S&P500 universe, we do observe that the complex system configurations obtain an improved increasing pattern compared to the CRSP universe. Recall from Table 4.1 that significant superiority for the complex configurations in terms of Spearman correlation coefficient is exclusively present in the CRSP universe. If we examine Figure 4.7, we observe an improved trend for the complex configurations in the S&P500 universe compared to the CRSP universe, yet the general structure across deciles for these complex configurations is not explicitly/immediately different to that of the linear and non-parametric signals. Generally, the higher resemblance across all signals suggests that the amount of exploitable market inefficiencies might be reduced in this larger market capitalization universe as theorized by Fama and French [2008], Fama and French [2018] and Hou et al. [2020]. On the other hand, Figure 4.8 shows us that the complex configurations do show differentiation in terms of volatility-skew across decile portfolios in the S&P500 universe. All in all, these result serves as further evidence against the transfer learning capabilities of these complex systems in terms of parameter estimation in the CRSP universe and subsequent out-of-sample prediction in the S&P500 universe.

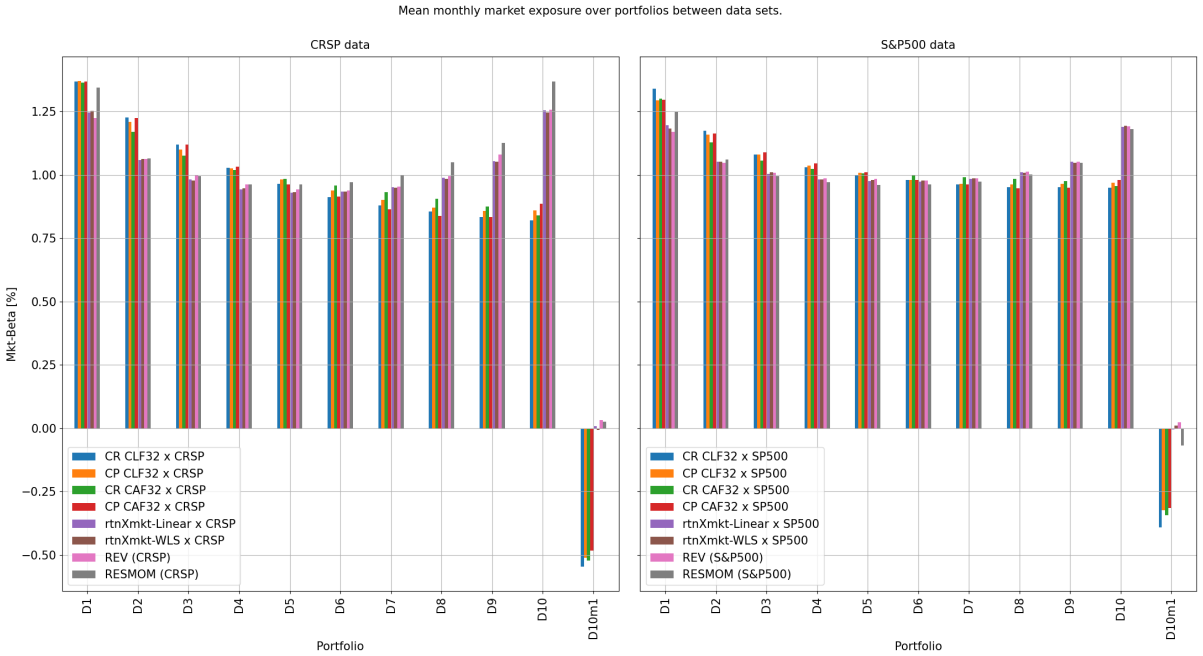


Figure 4.9: Mean monthly CAPM-β coefficient (i.e. market factor exposure) over deciles in the CRSP and S&P500 universes in the out-of-sample period 1995/01-2022/12. CAPM-β coefficient estimated over 36 months of data for every asset individually. Expanding window applied for the complex system configurations with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years. Monthly decile average CAPM-β obtained by taking the average of the CAPM-β coefficient (36-month estimation window) over all assets contained in that decile for that month t.

To complete this analysis, we consider to what extent the complex configurations can decouple risk. To this end, we consider the average $CAPM-\beta$ coefficient over all assets contained in the deciles in Figure 4.9. This coefficient is estimated over 36 months of data. This decile statistic can be interpreted as the average sensitivity to systemic risk. Recall that in this research we opt for the use of excess market returns to limit effect related to market timing. In fact, in section 4.3 we use this definition to define a volatility-based signal which should serve as an approximation of idiosyncratic risk. In the works by Ang et al. [2006] and Chen and Petkova [2012], this idiosyncratic risk is identified to hold a negative relationship to future return and not the systemic risk sensitivity. Indeed, from the cross-sectional regression results in Table 4.4, we

observe a that our idiosyncratic volatility estimate commands a negative contribution to return prediction. Now, if we consider Figure 4.9, we see that the complex configurations achieve a $CAPM - \beta$-skew similar in shape to Figure 4.8. For the linear and non-parametric signals this shape is approximately symmetric. Consequently, the linear and non-parametric signals attain a negligible net $CAPM - \beta$ exposure, while for this complex configuration this net exposure is substantially negative. This observation extends to both universes. As such, the use of excess market returns achieves the desired result for the linear parametric configuration. On the other hand, for the complex configurations this is clearly not the case. This can be explained through the increased importance of volatility-related features for the complex configurations established in section 4.3. Namely, these configurations access the ability to setup volatility-related features for prediction purposes, yet do not have the ability to decouple these into a systemic and idiosyncratic part. Hence, there is likely a 'spillover' effect of systemic risk into the idiosyncratic risk approximation. This is explanation is plausible if one considers that the complex systems only access the information contained in a single asset in their prediction. They do not contain the essential cross-sectional information required for an adequate decoupling of risk such as market portfolio volatility. To conclude, the results in Figure 4.9 entail that the intention to reduce market timing through the use of excess market returns is not successful for the complex configurations. Specifically, we attribute this to their increased focus on volatility-related features and an inability to decouple the idiosyncratic and systemic risk components.

4.5 Focal portfolio performance

In this section, we further expand our analysis on portfolio level by considering the performance of the focal D10 minus D1 (i.e. D10-D1) equally-weighted portfolios. The summarizing performance of the two linear and four complex configurations over the entire out-of-sample period $[1995, 2022]$ is provided in Table 4.5. In addition, we provide the correlations between the portfolios in this period and across the universes in Figure 4.14. Finally, we analyse D10-D1 portfolios' performance over time through analysis of the cumulative return in Figure 4.10 and Figure 4.12 and drawdown in Figure 4.13.

Configuration	**mean**	**Sharpe**	**Configuration**	**mean**	**Sharpe**
CR CLF32 x CRSP	4.021	0.171	CR CLF32 x S&P500	4.910	0.247
CP CLF32 x CRSP	7.710	0.348	CP CLF32 x S&P500	8.287	0.431
CR CAF32 x CRSP	-3.268	-0.117	CR CAF32 x S&P500	0.447	0.018
CP CAF32 x CRSP	5.034	0.219	CP CAF32 x S&P500	4.587	0.230
rtnXmkt-Linear x CRSP	11.536	0.763	rtnXmkt-Linear x S&P500	6.183	0.296
rtnXmkt-WLS x CRSP	9.178	0.543	rtnXmkt-WLS x S&P500	4.328	0.241
Mkt-RF (CRSP)	13.910	0.678	Mkt-RF (S&P500)	13.333	0.785
SMB (CRSP)	1.231	0.112	SMB (S&P500)	2.410	0.369
HML (CRSP)	4.254	0.259	HML (S&P500)	-0.732	-0.070
MOM (CRSP)	0.474	0.019	MOM (S&P500)	1.820	0.063
REV (CRSP)	14.423	0.670	REV (S&P500)	4.550	0.224
RESMOM (CRSP)	3.170	0.150	RESMOM (S&P500)	-0.012	-0.000
LowVol (CRSP)	-7.198	-0.290	LowVol (S&P500)	-1.581	-0.050

Table 4.5: Equally-weighted D10-D1 portfolio performance statistics for the investigated configurations and common risk factor portfolios; Annualized excess (risk-free rate) mean return [%] and Annualized excess (risk-free rate) Sharpe in the CRSP and S&P500 universes in the out-of-sample period 1995/01-2022/12. Expanding window applied for the complex system configurations with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years.

To start, we examine the summarizing performance of the portfolios in Table 4.5. In previous sections, we identified no immediate difference in the prediction structure between the complex

configurations in terms or empirical density or portfolio performance across deciles. For the focal portfolio, the differences are more pronounced. We observe that in both asset universes the complex configurations using rank based inputs (i.e. $CP-$) as well as the systems with a LSTM subsystem (i.e. $-CLF$) outperform their counterparts in terms of mean excess return and Sharpe ratio. On the other hand, we observe in both universes that the linear configurations' and REV benchmark portfolios show competitive or even superior performance on the two metrics. Furthermore, note that the MOM and $RESMOM$ factor portfolios score poorly on the metrics. Finally, we recognize that in the CRSP universe the low volatility factor ($LowVoL$) achieves stark negative scores on the performance metrics. One should recall that this D10-D1 portfolio acquires a net negative CAPM-β exposure position by definition. Therefore, the fact that this portfolio attains substantial losses suggests that such a strategy does exploit sizeable market inefficiencies in this universe as theorized by Blitz and Van Vliet [2007]. However, we note that the negative performance metrics suggests an inversion of portfolio positions. In addition, we observe the diminished performance among the larger market capitalization assets in the S&P500 universe, suggesting that the consistency of these inefficiencies is once again conditional on asset size. All in all, the negative performance of the $LowVol$ portfolio the CRSP universe could serve as evidence for the diminished performance of the complex configurations since their volatility ranking aspect aligns them more closely with this factor.

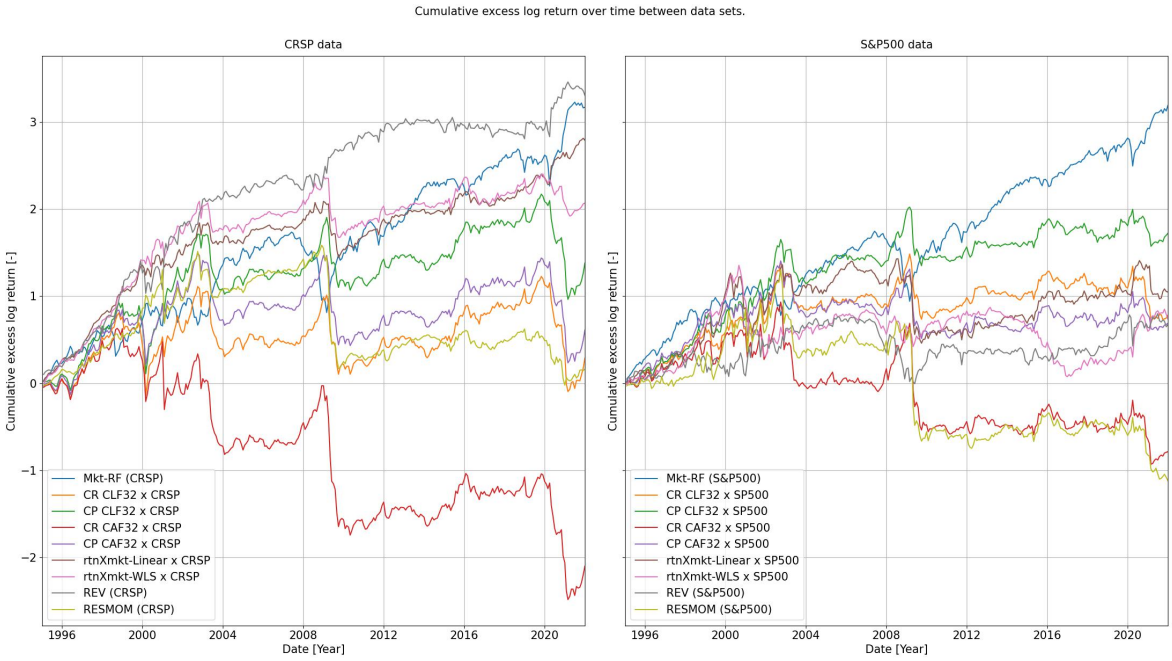


Figure 4.10: Equally-weighted D10-D1 portfolio cumulative log return in CRSP and S&P500 universes in the out-of-sample period 1995/01-2022/12. Expanding window applied for the complex system configurations with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years.

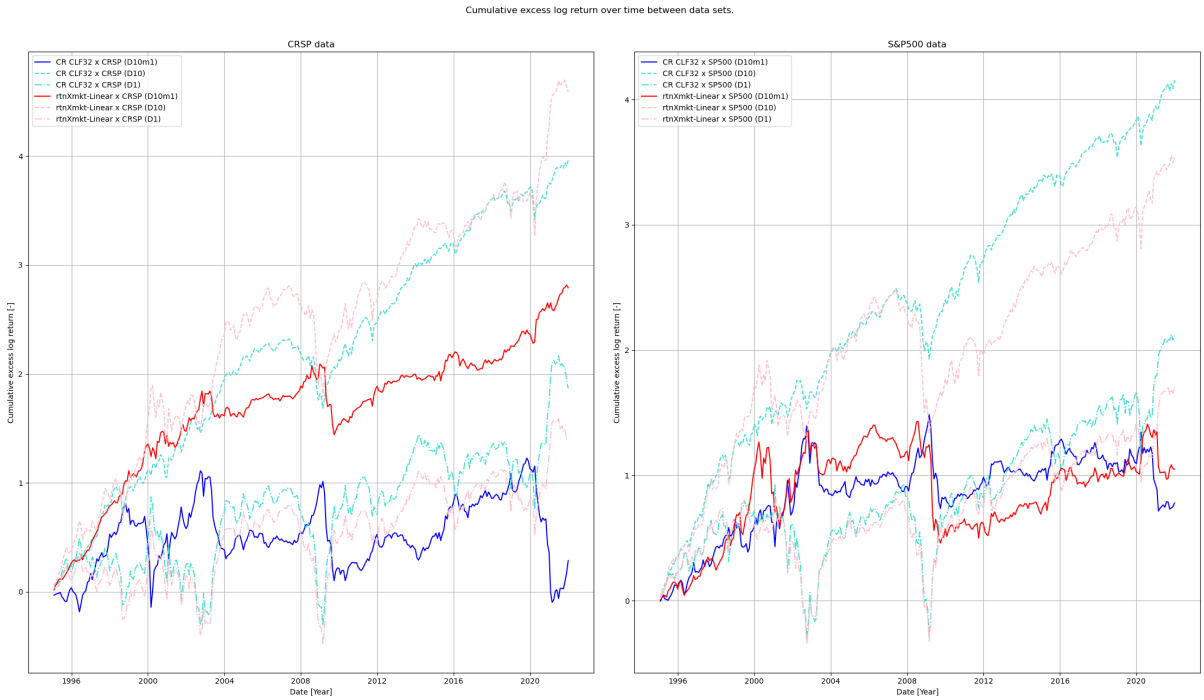


Figure 4.11: Equally-weighted D10-D1 portfolio cumulative log return in CRSP and S&P500 universes in the out-of-sample period 1995/01-2022/12 for the equally-weighted D10-D1, D10 and D1 portfolios. Expanding window applied for the complex system configurations with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years.

To continue, we consider the D10-D1 portfolios' performance over time visualized in Figure 4.10. For the complex configurations, we observe high correlation in terms of cumulative return progression in both universes. Although this serves a further evidence on the indifference of the exploited market inefficiencies between these configurations on portfolio level, it does show that the implementation of ranked inputs ($CP-$) and the LSTM subsystem ($-CLF$) can provide beneficial nuances in terms of return magnitude. On the other hand, in the CRSP universe we observe that the cumulative return of the $rtnXmkt - Linear$ configuration is consistently less volatile and competitive over time to that of the complex $CR - CLF$ configuration. On the other hand, in the S&P500 universe we observe that an increase in cumulative return is focused in the period before 2010, after which it flattens regardless of the configuration type. This result aligns with the general decline in smoothed Spearman correlation coefficient scores in this universe as shown in Figure 4.1.

In Figure 4.11, we present the cumulative return of the focal and decile portfolios separately and identify that the main difference between the configurations presents itself in the long D10 portfolio. Once again, these results can be connected to with the empirical heatmaps in section 4.2 and the aggregated decile portfolio statistics in section 4.4. Indeed, the high resemblance of the short D1 portfolios in terms of volatility can be explained by the fact that both these configurations exhibit an empirical cluster at the lower predicted percentiles containing a mix of assets with true extreme/tail rankings. For the linear configuration, this same empirical mix is present in D10 resulting in higher volatility. On the other hand, the complex configuration attains a D1 portfolio with substantially lower volatility as it contains assets with true center rankings. All in all, the portfolio statistics in Table 4.5 and the cumulative performance over time in Figure 4.11 clearly imply superiority of the linear configuration. This result suggests that a mix of extreme ranks in both the D1 and D10 portfolio instead of only D1 is beneficial on the aggregated portfolio level. This can be explained through cancellation of these extreme returns

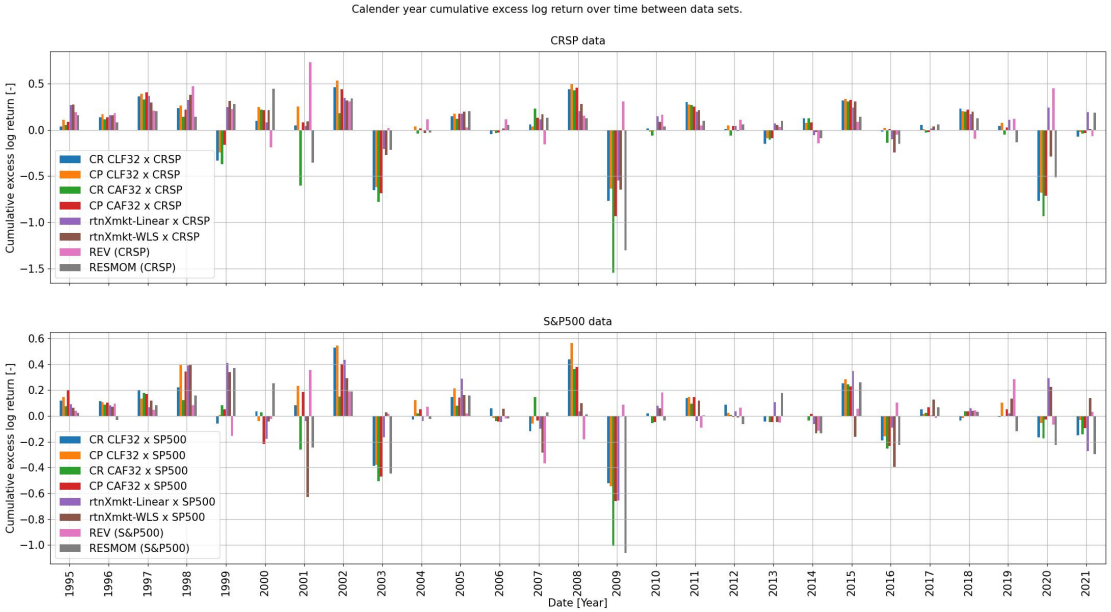


Figure 4.12: Equally-weighted D10-D1 portfolio cumulative log return of every calendar year in CRSP and S&P500 universes in the out-of-sample period 1995/01-2022/12. Expanding window applied for the complex system configurations with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years. Yearly reset occurs at 1st of January, meaning that cumulative return is that observed between 01/01/year up to and including 31/12/year.

that likely occurs due to the applied long and short positions. Figure 4.7 and Figure 4.8 imply that this cancellation causes both an increased average return and decreased average volatility in D10-D1 for the linear configuration. Finally, it is important to recognize that this cancellation phenomenon is implied by the lower volatility in D10-D1 compared to larger volatility in both D1 and D10 individually observed for the linear configuration.

We complete this section by considering Figure 4.12 and Figure 4.13. In Figure 4.12, we observe that although the calendar year cumulative return of $CR-CLF$ is superior to that of the linear configurations in a multitude of years, this performance is not consistent. Amongst the complex configurations, the $CP-CLF$ shows the highest frequency in terms of yearly out-performance, yet we note that this is diminished in the S&P500 universe and in the CRSP universe after 2010. Furthermore, Figure 4.13 indicates how the $CR-CLF$ configuration exhibits steeper and more frequent drawdowns as well as a decreased recovery time than the $rtnXmkt-Linear$ configuration in the CRSP universe. Once again, in the S&P500 universe, these differences are not as apparent. For the complex configurations, a general pattern around drawdowns can be identified over various sub-periods in Figure 4.11. To illustrate, we focus on the crash of 2008 where we observe a steeper drawdown in D1 than in D10 resulting in a steep peak initially. After the trough has been hit in both D1 and D10, the D1 portfolio recovers faster than D10. Consequently, we observe the drawdown of D10-D1 during this recovery phase. Conceptually, this means that the complex configuration experiences worse drawdowns due to a larger and more frequent mismatch in relative recovery speed between D10 and D1 compared to that of the linear configuration. Since we connected this portfolio feature to the inherent differences between D1 and D10, we can once again associate this with the decile volatility skew observed in Figure 4.8 and the empirical heatmaps in section 4.2.

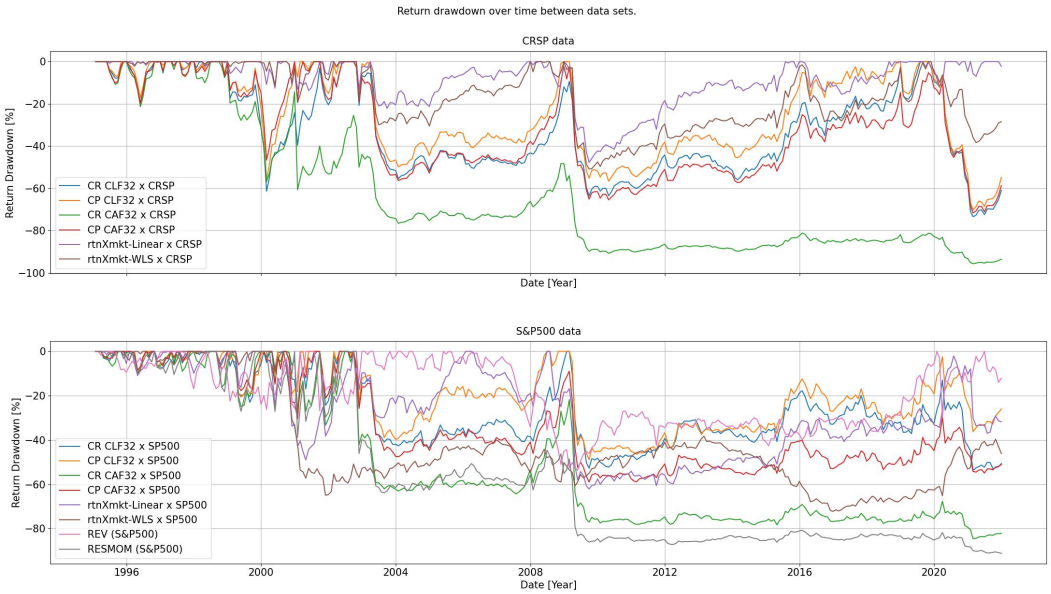


Figure 4.13: Equally-weighted D10-D1 portfolio drawdown in CRSP and S&P500 universes in the out-of-sample period 1995/01-2022/12. Expanding window applied for the complex system configurations with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years.

4.6 Factor spanning regression analysis

In this section, we complete our analysis on portfolio level by investigating the exposure of the system's portfolios to common risk factors. For this analysis, we include the three factor Fama-French model by Fama and French [1992] appended by the reversal (REV) from Jegadeesh [1990], residual momentum ($RESMOM$) by Blitz et al. [2011] and low-volatility factor ($LowVol$) by Blitz and Van Vliet [2007]. All risk factors are based on the signals defined by the previous set of authors and are constructed using equal-weighting schemes in their respective universes. In this analysis we opt for the $RESMOM$ factor instead of conventional MOM factor because of the observed reduced drawdowns, which form a characteristic property of the $RESMOM$ factor. As such, a regression on $RESMOM$ instead of its conventional counterpart should help to shed light on this aspect as well. To begin this analysis we consider factor regression results provided in Table 4.6 and the correlation matrices given in Figure 4.14.

To begin this analysis we consider factor regression results provided in Table 4.6 and the correlation matrices given in Figure 4.14. If we consider the Pearson correlation coefficients for the D10-D1 portfolios presented in Figure 4.14 we can clearly identify differences between the linear and more complex parametric systems. From these matrices, we conclude that the complex configurations are extremely correlated with a minimum coefficient of 0.97 observed between the best ($CP - CLF32$) and worst ($CR - CAF32$) performing complex configurations in the CRSP universe. Expectantly, this high correlation between the complex systems is accompanied by similar correlation coefficients to the other portfolios across the complex configurations. On the other hand, the correlation drops to $\approx \{0.6, 0.8\}$ between the linear and complex configurations in the CRSP universe. If we consider the correlation of all parametric configurations to the non-parametric factor benchmark portfolios we observe high absolute correlations to the return-based factors REV, MOM and $RESMOM$ in both universes for all (linear & complex) configurations. This is to be expected because all these portfolios originate from the same information set, albeit subsets for some of them. Furthermore, we observe high positive correlation

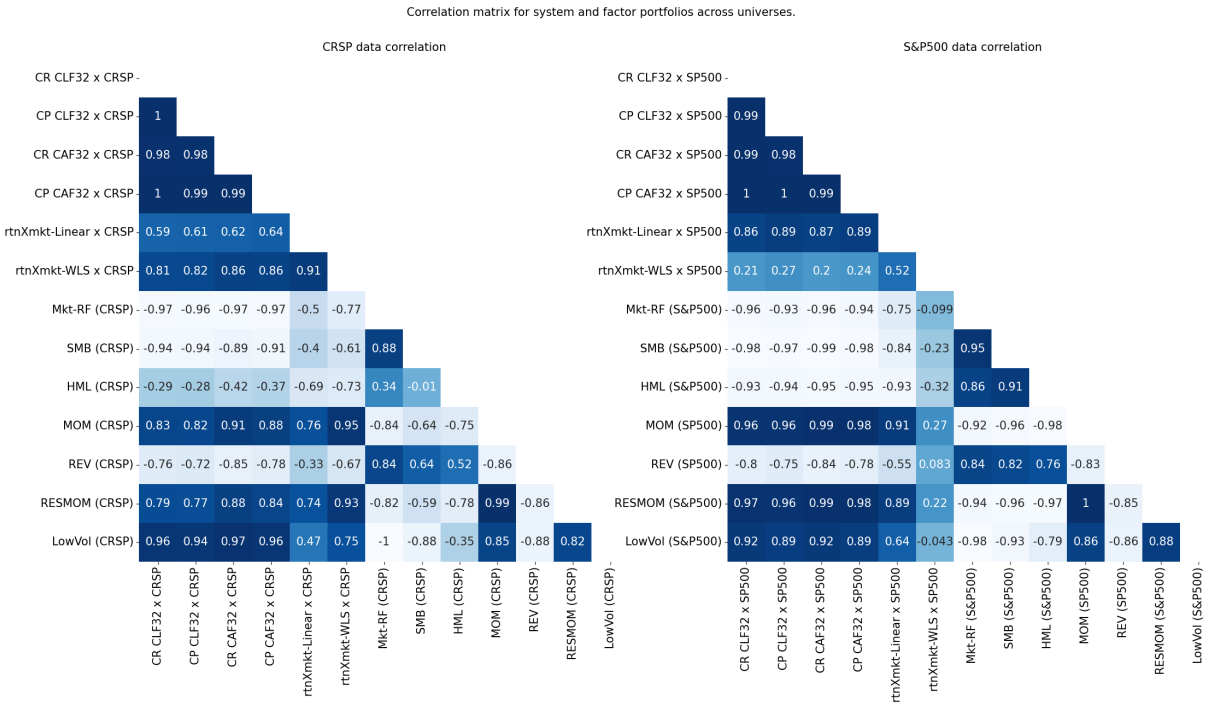


Figure 4.14: Pearson correlation coefficients between system configurations' D10-D1 portfolios and common factor portfolios in the out-of-sample period 1995/01-2022/12 in the CRSP and S&P500 universes. All portfolios setup according to an equal weighting scheme.

to the low-volatility factor for the complex configurations, but not for the linear configurations. This is in line with the results and conclusions drawn from Figure 4.8 which indicated that these configurations present a volatility skew. Recall that this resulted from the exclusive allocation of assets with true tail ranks to the lower predicted percentiles, as is visible through the empirical 'inverted U'-shape in Figure 4.5. Through the average $CAPM-\beta$ in Figure 4.9 we acknowledge that a $CAPM-\beta$ skew also exists for the complex configurations. Consequently, this allocation implies high $CAPM-\beta$ in D1 and low $CAPM-\beta$ in D10 for the complex systems, similar to the definition of the low volatility factor by Blitz and Van Vliet [2007]. Finally, we notice that the complex system configurations show substantially larger negative correlation to the market and size factor than their linear parametric counterparts. This statement extends to both asset universes. The larger correlation to the market aligns with the negative correlation to the low volatility factor and the fundamental connection between these two factors. This is because the low volatility factor obtains a net negative $CAPM-\beta$ due to the shorting of large $CAPM-\beta$ assets (larger systemic volatility sensitivity) which is not completely offset by the long position in small $CAPM-\beta$ assets (larger systemic volatility sensitivity). On the other hand, the market portfolio attains a positive $CAPM-\beta$ position by simply investing in the entire universe. Furthermore, the size factor correlation is unexpected as smaller capitalization assets should theoretically present more frequent inefficiencies.

CRSP	**CR-CLF32**	**CP-CLF32**	**CR-CAF32**	**CP-CAF32**	**rtnXmkt-Linear**	**rtnXmkt-WLS**
alpha	4.841 (2.164)	7.589 (3.205)	-0.38 (-0.182)	5.644 (2.498)	3.492 (1.942)	3.567 (1.61)
Mkt-RF (CRSP)	-0.195 (-3.364)	-0.221 (-3.608)	-0.145 (-2.687)	-0.19 (-3.255)	-0.026 (-0.556)	-0.044 (-0.952)
LowVol (CRSP)	0.375 (7.108)	0.299 (5.357)	0.328 (6.653)	0.267 (5.021)	-0.092 (-1.679)	-0.019 (-0.458)
SMB (CRSP)	-0.869 (-11.963)	-0.847 (-11.017)	-1.012 (-14.911)	-0.864 (-11.779)	-0.388 (-6.652)	-0.413 (-7.122)
HML (CRSP)	0.082 (1.352)	0.024 (0.366)	0.003 (0.052)	-0.026 (-0.42)	-0.214 (-4.382)	-0.301 (-6.214)
RESMOM (CRSP)	0.339 (7.689)	0.326 (6.987)	0.589 (14.308)	0.444 (9.993)	0.569 (16.076)	0.642 (18.256)
REV (CRSP)	0.297 (8.38)	0.365 (9.742)	0.06 (1.801)	0.257 (7.184)	0.498 (17.491)	0.378 (13.354)
S&P500	**CR-CLF32**	**CP-CLF32**	**CR-CAF32**	**CP-CAF32**	**rtnXmkt-Linear**	**rtnXmkt-WLS**
alpha	5.934 (2.676)	9.61 (4.169)	1.962 (1.053)	5.937 (2.758)	4.574 (1.812)	3.756 (1.13)
Mkt-RF (S&P500)	0.007 (0.126)	-0.063 (-1.068)	0.075 (1.577)	-0.021 (-0.38)	0.027 (0.415)	0.003 (0.031)
LowVol (S&P500)	0.218 (7.176)	0.104 (3.304)	0.196 (7.714)	0.103 (3.49)	-0.145 (-4.212)	-0.148 (-3.264)
SMB (S&P500)	-0.677 (-5.227)	-0.686 (-5.097)	-0.974 (-8.959)	-0.8 (-6.36)	-0.372 (-2.525)	-0.432 (-2.226)
HML (S&P500)	-0.114 (-1.55)	-0.121 (-1.585)	-0.277 (-4.486)	-0.231 (-3.24)	-0.361 (-4.317)	0.004 (0.04)
RESMOM (S&P500)	0.34 (10.45)	0.385 (11.359)	0.508 (18.585)	0.4 (12.644)	0.584 (15.736)	0.232 (4.742)
REV (S&P500)	0.193 (5.632)	0.298 (8.377)	0.042 (1.462)	0.232 (6.99)	0.435 (11.17)	0.296 (5.772)

Table 4.6: Annualized factor spanning regression statistics for the investigated configurations. All values are fitted coefficients with t-stat in brackets in the out-of-sample period 1995/01-2022/12. Expanding window applied for the complex system configurations with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years. Note that Newey-West 12 lags standard errors are used for standard deviation computation.

We proceed by considering the factor exposures presented in Table 4.6. To start this analysis we identify the differences between the linear and more complex parametric systems. From this regression table we conclude that the α statistics is both larger and significant for the complex configurations (excluding $CR-CAF$) compared to insignificant intercepts for the linear cases. For the linear cases an insignificant α is expected. This is because they likely pick up on the same market inefficiencies as the $RESMOM$ and REV factors and can only define a relative weighting between them. On the other hand, the significant α for the complex configurations entails that they pick up on further market inefficiencies. This is possible through their inherent access to more complex non-linear/interactive features as demonstrated by Goodfellow et al. [2016]. Note that this result seems contradictory to the cross-sectional regression results in Table 4.4, which showed us that the $CR-CLF$ configuration had limited predictive power in the cross-section beyond that contained in a linear and volatility-based signal. However, Table 4.6 shows us that on portfolio level do differentiate themselves. Therefore, we attribute the significant intercepts of these complex configurations to improved portfolio timing between time steps. This axis of improvement stands in contrast to a consistently improved predictive ability in the cross-section due to the identification and exploitation of sizeable and previously unknown market inefficiencies. Instead, portfolio timing improvement arises from nuances in the system prediction; a dynamic shift in the relative importance of features during different market conditions (e.g. crash/recovery). Furthermore, we observe that the exposure of the linear systems is different to that of the $CR-CLF$ configuration in terms of significant exposure to the value (HML) and the insignificant exposure to the low volatility factor ($LowVol$). The insignificant exposure to the $LowVol$ factor aligns with the absence of the volatility skew observed in Figure 4.8 for the linear configurations. The negative significance to the HML value factor indicates a tilt towards growth assets. Finally, the linear and complex configurations are similar in terms of significant exposure to residual momentum ($RESMOM$) and (REV) reversal factor. This is to be expected due to all these portfolios sharing the (or parts of the) same information set. This observation indicates that the complex configurations do not only rank volatility, but indeed also pick up on linear momentum- and reversal-related inefficiencies.

To continue, we consider the performance amongst the complex configurations. If we examine the exposures presented in Table 4.6, we notice that the $CR-CAF$ network obtains no significant intercept. This aligns with the consistent inferior performance; reducing the potential magnitude of the intercept term. On the other hand, in this table we also note that this reduced

intercept can be caused by an increase in exposures to the risk factors compared to the $-CLF$ configurations. The ranked based self-attention network ($CP-CAF$) does show a significant intercept alongside the identified increased exposure. What's more is that configurations with the ranked inputs (i.e. $CP-$) and the LSTM-subsystem ($-CLF$) improve over their alternative counterparts. Indeed the $CP-CLF$ acquires superiority in the intercept term to the other configurations in both universes. Considering previous results and the comments made at the start of the previous paragraph, we attribute this superiority amongst the complex configurations to a improvement portfolio timing. Inline with the decile performance, correlation and optimization results presented in previous sections, we generally do not observe substantial differences in exposures between $CR-CLF$ and $CP-CLF$ in the CRSP universe. Specifically, in the CRSP universe we observe that the complex configurations show negative exposure to the size factor indicating alignment with larger market capitalization assets such as those contained in the S&P500 universe. Here, we observe no substantial changes in the exposures of the complex configurations besides a significant exposure to the market. In this universe, the size exposure is also significantly negative indicating a tilt towards to the larger capitalization assets in this large S&P500 market capitalization universe. Similarly, the exposures to return-based factors $RESMOM$ and REV have generally remained unchanged. The exposure to the $LowVol$ factor has remained unchanged as well. This is to be expected as these complex systems have been identified to implicit rank volatility and as such will do so regardless of the universe they are operating in.

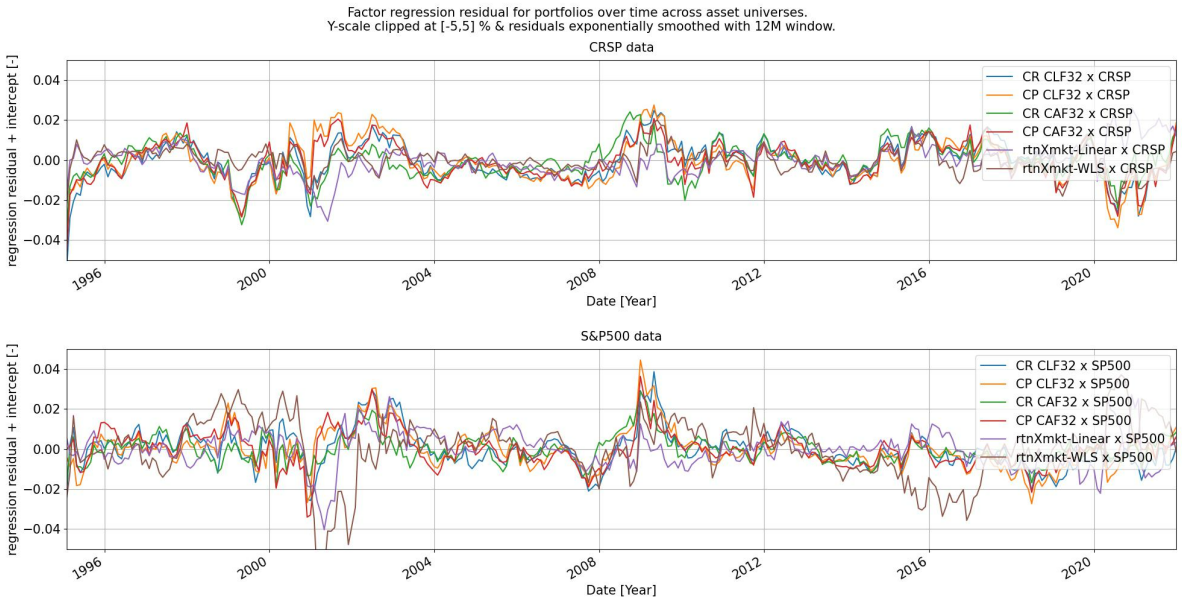


Figure 4.15: 6-factor regressions' 12-month exponentially smoothed residual for the system focal portfolios in CRSP and S&P500 universes in the out-of-sample period 1995/01-2022/12. Associated exposures given in Table 4.6. Expanding window applied for the complex system configurations with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years.

We further expand this analysis by considering the stability of the residual/error distribution of the previous 6-factor regressions as presented in Figure 4.15. In this figure we consider the 12-month exponentially smoothed residual time series instead of their original counterpart in an attempt to shed light on the stability of the residual series' mean. In general, we observe stability across all configurations in the CRSP universe except for the crashing periods [2000,2003], [2008,2010] and [2020,2022]. In there periods, we generally observe relatively larger smoothed positive residuals for the complex configurations compared to the linear configurations. This

observation can serve as evidence in favour of an improved portfolio timing aspect of the complex configurations. This is because a larger positive residual can be interpreted as out-performance of the portfolio over its estimated (linear) portfolio based on the factor portfolios. In the S&P500 universe, this same conclusion can be drawn except for the linear $rtnXmkt-WLS$ configuration which shows a more volatile pattern even outside the previously identified periods. This is likely due to the increased exposures of this configuration which might cause an improved fit generally, but a more divergent one during specific sub-periods when factor portfolios might show increased correlation amongst themselves.

CR-CLF	D10-D1	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
alpha (CRSP)	4.841 (2.164)	-2.273 (-1.579)	1.613 (2.003)	0.747 (0.96)	1.381 (1.724)	1.006 (1.352)	1.08 (1.496)	0.76 (0.941)	1.379 (1.712)	2.526 (2.862)	2.568 (2.376)
Mkt-RF (CRSP)	-0.195 (-3.364)	0.065 (1.75)	0.038 (1.816)	0.06 (2.985)	-0.004 (-0.178)	-0.037 (-1.931)	-0.064 (-3.438)	-0.069 (-3.31)	-0.091 (-4.387)	-0.096 (-4.186)	-0.13 (-4.633)
LowVol (CRSP)	0.375 (7.108)	-0.23 (-6.765)	-0.1 (-5.243)	0.004 (0.192)	0.024 (1.258)	0.057 (3.235)	0.072 (4.225)	0.109 (5.726)	0.122 (6.441)	0.146 (7.027)	0.145 (5.703)
SMB (CRSP)	-0.869 (-11.963)	0.509 (10.887)	0.187 (7.17)	0.026 (1.034)	-0.06 (-2.318)	-0.131 (-5.445)	-0.21 (-8.953)	-0.248 (-9.463)	-0.318 (-12.174)	-0.329 (-11.488)	-0.36 (-10.262)
HML (CRSP)	0.082 (1.352)	-0.082 (-2.091)	-0.019 (-0.873)	0.047 (2.215)	0.061 (2.806)	0.076 (3.751)	0.09 (4.597)	0.062 (2.815)	0.062 (2.834)	0.035 (1.475)	0.0 (0.015)
RESMOM (CRSP)	0.339 (7.689)	-0.248 (-8.742)	-0.09 (-5.657)	0.016 (1.012)	0.041 (2.62)	0.089 (6.057)	0.116 (8.161)	0.107 (6.75)	0.108 (6.785)	0.128 (7.378)	0.091 (4.272)
REV (CRSP)	0.297 (8.38)	-0.125 (-5.494)	-0.037 (-2.869)	-0.008 (-0.613)	-0.001 (-0.044)	0.017 (1.472)	0.038 (3.329)	0.058 (4.518)	0.095 (7.477)	0.125 (8.93)	0.172 (10.031)

CR-CLF	D10-D1	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
alpha (S&P500)	5.934 (2.676)	-1.928 (-1.204)	-2.149 (-1.831)	-1.998 (-1.775)	-1.188 (-1.297)	-0.232 (-0.299)	-0.155 (-0.2)	-0.019 (-0.024)	1.232 (1.557)	2.4 (2.397)	4.006 (3.754)
Mkt-RF (S&P500)	0.007 (0.126)	-0.039 (-0.948)	-0.036 (-1.218)	0.066 (2.29)	0.019 (0.809)	0.009 (0.432)	0.015 (0.772)	0.025 (1.241)	-0.005 (-0.27)	-0.02 (-0.786)	-0.032 (-1.161)
LowVol (S&P500)	0.218 (7.176)	-0.206 (-9.396)	-0.091 (-5.653)	-0.007 (-0.48)	0.028 (2.22)	0.042 (4.015)	0.043 (4.071)	0.067 (6.168)	0.063 (5.797)	0.052 (3.826)	0.012 (0.806)
SMB (S&P500)	-0.677 (-5.227)	0.411 (4.394)	0.415 (6.06)	0.187 (2.838)	0.012 (0.226)	-0.139 (-3.086)	-0.132 (-2.918)	-0.109 (-2.361)	-0.187 (-4.05)	-0.196 (-3.351)	-0.266 (-4.264)
HML (S&P500)	-0.114 (-1.55)	0.04 (0.755)	0.039 (0.996)	0.063 (1.69)	0.096 (3.17)	0.054 (2.12)	-0.063 (-2.44)	-0.063 (-2.416)	-0.058 (-2.21)	-0.036 (-1.082)	-0.074 (-2.088)
RESMOM (S&P500)	0.34 (10.45)	-0.255 (-10.829)	-0.118 (-6.862)	-0.029 (-1.741)	0.01 (0.75)	0.028 (2.435)	0.073 (6.451)	0.07 (6.042)	0.08 (6.923)	0.057 (3.898)	0.086 (5.457)
REV (S&P500)	0.193 (5.632)	-0.081 (-3.27)	-0.137 (-7.587)	-0.078 (-4.484)	-0.04 (-2.853)	-0.017 (-1.389)	0.027 (2.28)	0.058 (4.729)	0.052 (4.226)	0.105 (6.815)	0.112 (6.793)

Table 4.7: Annualized alpha statistics for the $CR-CLF$ configuration. All deciles portfolios are excess market returns, except for the D10-D1 portfolio. All values are fitted coefficients with t-stat in brackets in the OOS period 1995/01-2022/12. Expanding window applied with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years. Note that Newey-West 12 lags standard errors are used for standard deviation computation.

rtnXmkt-Linear	D10-D1	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
alpha (CRSP)	3.492 (1.942)	-2.461 (-1.861)	-0.335 (-0.413)	0.142 (0.177)	1.236 (1.686)	1.148 (1.57)	1.51 (2.107)	0.815 (1.292)	0.889 (1.04)	0.585 (0.62)	1.031 (0.96)
Mkt-RF (CRSP)	-0.026 (-0.556)	0.07 (2.059)	0.025 (1.202)	-0.027 (-1.313)	-0.065 (-3.435)	-0.057 (-3.006)	-0.065 (-3.532)	-0.038 (-2.322)	0.004 (0.159)	0.022 (0.909)	0.045 (1.604)
LowVol (CRSP)	-0.092 (-1.679)	-0.084 (-2.688)	0.025 (1.303)	0.066 (3.52)	0.072 (4.179)	0.087 (5.066)	0.08 (4.752)	0.074 (5.006)	0.098 (4.862)	0.018 (0.805)	-0.176 (-6.955)
SMB (CRSP)	-0.388 (-6.652)	0.428 (9.977)	-0.026 (-0.986)	-0.085 (-3.268)	-0.125 (-5.245)	-0.179 (-7.534)	-0.201 (-8.658)	-0.204 (-9.979)	-0.212 (-7.655)	-0.174 (-5.671)	0.04 (1.147)
HML (CRSP)	-0.214 (-4.382)	-0.039 (-1.083)	0.11 (5.006)	0.129 (5.927)	0.137 (6.865)	0.112 (5.655)	0.116 (5.957)	0.08 (4.683)	0.02 (0.854)	-0.021 (-0.818)	-0.253 (-8.666)
RESMOM (CRSP)	0.569 (16.076)	-0.363 (-13.927)	-0.12 (-7.499)	-0.074 (-4.673)	-0.021 (-1.425)	0.018 (1.23)	0.069 (4.862)	0.11 (8.838)	0.095 (5.666)	0.155 (8.35)	0.206 (9.76)
REV (CRSP)	0.498 (17.491)	-0.209 (-9.997)	-0.135 (-10.464)	-0.084 (-6.655)	-0.064 (-5.549)	-0.019 (-1.671)	-0.003 (-0.291)	0.032 (3.165)	0.067 (4.912)	0.136 (9.101)	0.289 (16.966)

rtnXmkt-Linear	D10-D1	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
alpha (S&P500)	4.574 (1.812)	-3.164 (-2.195)	-1.358 (-1.215)	-0.88 (-0.976)	-0.196 (-0.211)	0.993 (1.421)	1.221 (1.678)	0.969 (1.125)	1.031 (1.163)	-0.022 (-0.023)	1.409 (0.84)
Mkt-RF (S&P500)	0.027 (0.415)	-0.044 (-1.188)	-0.026 (-0.926)	0.025 (1.11)	-0.003 (-0.119)	-0.019 (-1.088)	0.006 (0.313)	0.019 (0.849)	0.047 (2.087)	0.012 (0.508)	-0.017 (-0.396)
LowVol (S&P500)	-0.145 (-4.212)	-0.098 (-4.948)	-0.004 (-0.242)	0.087 (7.026)	0.077 (6.051)	0.06 (6.241)	0.055 (5.529)	0.075 (6.34)	0.029 (2.401)	-0.034 (-2.627)	-0.243 (-10.586)
SMB (S&P500)	-0.372 (-2.525)	0.232 (2.759)	0.211 (3.238)	0.056 (1.062)	-0.005 (-0.097)	-0.063 (-1.55)	-0.118 (-2.771)	-0.036 (-0.708)	-0.089 (-1.716)	-0.053 (-0.953)	-0.14 (-1.429)
HML (S&P500)	-0.361 (-4.317)	0.141 (2.945)	0.068 (1.828)	0.065 (2.178)	0.072 (2.35)	0.015 (0.655)	-0.014 (-0.578)	0.024 (0.83)	-0.072 (-2.448)	-0.08 (-2.54)	-0.22 (-3.964)
RESMOM (S&P500)	0.584 (15.736)	-0.303 (-14.32)	-0.165 (-10.051)	-0.111 (-8.366)	-0.045 (-3.306)	-0.009 (-0.899)	0.013 (1.181)	0.067 (5.334)	0.088 (6.723)	0.187 (13.454)	0.28 (11.369)
REV (S&P500)	0.435 (11.17)	-0.167 (-7.49)	-0.155 (-9.009)	-0.057 (-4.079)	-0.051 (-3.549)	-0.044 (-4.044)	0.002 (0.2)	0.02 (1.538)	0.082 (5.966)	0.102 (6.959)	0.268 (10.368)

Table 4.8: Annualized alpha statistics for the $rtnXmkt-Linear$ configuration. All deciles portfolios are excess market returns, except for the D10-D1 portfolio. All values are fitted coefficients with t-stat in brackets in the OOS period 1995/01-2022/12. Expanding window applied with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years. Note that Newey-West 12 lags standard errors are used for standard deviation computation.

Finally, we expand this analysis by considering the exposures of all decile portfolios for the complex $CR-CLF$ and linear $rtnXmkt-Linear$ configurations as presented in Table D.8 and Table D.9. For the $LowVol$, SMB, $RESMOM$ and REV factors these results are also visualized in Figure 4.16 and Figure 4.17 for the CRSP and S&P500 universe respectively. Due to the high correlation between the complex configurations presented in Figure 4.14, we only conduct this analysis for the $CR-CLF$ configurations. Note that all decile portfolios are regressed with on excess market portfolio returns.

We start this analysis by considering the intercept (α) for both configurations. We observe that neither configuration is able to achieve a significant intercept for D1. On the other hand, for the $CR-CLF$ configuration D9 & D10 are significant in the CRSP universe, yet show the noisy/flat pattern across deciles previously observed in Figure 4.7. The linear configuration shows fewer deciles with significant intercepts, which aligns with remarks made in the previous paragraphs;

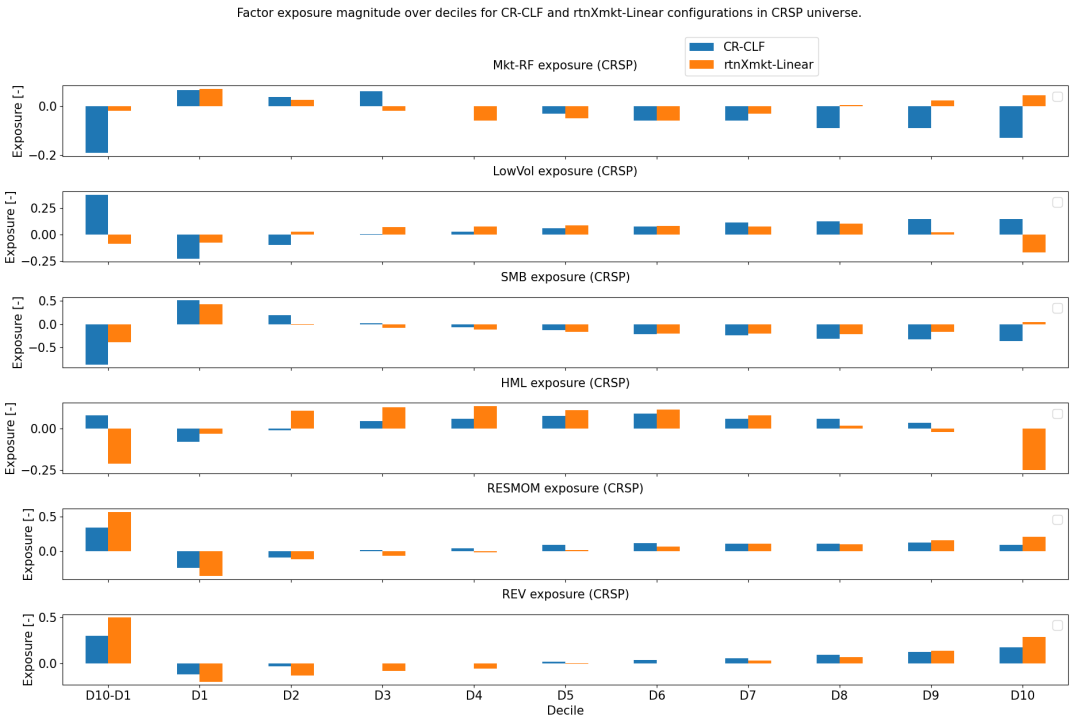


Figure 4.16: Factor exposure magnitude across deciles for the $Mkt - RF$, $LowVol$, SMB, HML, $RESMOM$ and REV equally-weighted factor portfolios. Exposures presented for the CR-CLF and rtnXmkt-Linear system configurations in the out-of-sample period 1995/01-2022/12 for the CRSP universe. Note that factor exposure for deciles (excluding D10-D1) is computed by using the portfolio returns excess market returns.

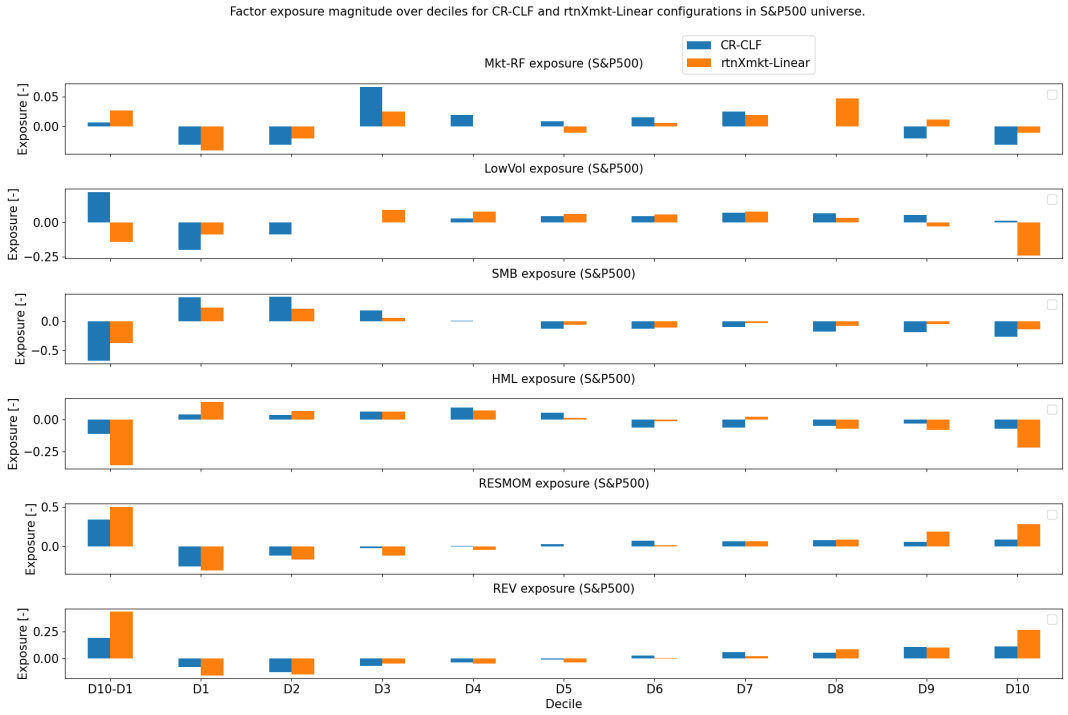


Figure 4.17: Factor exposure magnitude across deciles for the $Mkt - RF$, $LowVol$, SMB, $RESMOM$ and REV equally-weighted factor portfolios. Exposures presented for the CR-CLF and rtnXmkt-Linear system configurations in the out-of-sample period 1995/01-2022/12 for the S&P500 universe. Note that factor exposure for deciles (excluding D10-D1) is computed by using the portfolio returns excess market returns.

this configuration likely picks up on the same linear effects as the $RESMOM$ and REV factors.

To continue, we observe that the complex configuration attains significant and decreasing exposure to the market in the CRSP universe in D6 to D10. Since the decile portfolios present the results for excess market returns, this indicates that these portfolios likely increasingly contain assets with lower CAPM-beta assets (i.e. $\beta_{CAPM} < 1$). This can be expected as the previous results in Figure 4.8 indicated that these portfolios show lower volatility which is indirectly related to this CAPM relationship. Moreover, this significant market exposure is not observed in the S&P500 universe further indicating that these portfolios are neutral towards this factor in the higher market capitalization asset space. Contrary to the complex configuration, the linear configuration shows significant negative exposure to the market in D4 to D7. This significant exposure similarly fades in the S&P500 universe.

If we proceed by analysing the exposure to the $LowVol$ factor, we observe clear differences between the linear and complex configurations. We observe that this exposure to this factor increases and changes in sign in D5 and D4 in the CRSP and S&P500 universe respectively for the complex configuration. While this exposure clearly increases in the CRSP universe over this subset, this pattern is not as pronounced in the higher deciles for the S&P500 universe. From the difference in exposure between universes we conclude that the complex system's volatility ranking ability seems stronger in the CRSP universe than the larger capitalization S&P500 universe. On the other hand, this statement does not seem to hold for the high volatility assets; The significant negative exposure is observed for D1 and D2 regardless of the universe and is similar in magnitude. This indicates that selection of high volatility assets is less related to the assets' market capitalization size. Finally, for the linear configuration we do not observe the same pattern as for the complex configuration. The exposure structure is a bit more symmetric in both universes; with negative exposure in the tails and more positive exposures near the center deciles. This indicates that the tails (D1 and D10) both contain high volatility assets which aligns with the results shown in Figure 4.11 and Figure 4.8.

With regard to the value factor HML we obtain no net exposure for complex configurations but do so for the linear configurations. Analysing the decile exposures we can observe that this is caused by a steeper decile in D10 exposure specifically for the linear configurations. Consequently, the D10-D1 portfolio neutralizes the HML-axis for the complex configuration, while this does not occur for the linear configuration. This means that the linear configurations attain a general tilt towards value firms.

Next, we consider the decile exposures to the $RESMOM$ and REV factors. Due to the input feature set of the configurations aligning with information contained in these factors, we expectantly observe high significant throughout the deciles. Over the deciles we note an increasing exposure pattern between D1 to D10 for these factors across configurations and universes. This can be interpreted as the linear and complex configurations consistently picking up on the same linear inefficiencies captured in these factors. However, we observe that while this trend is clearly present for the linear case, the complex configuration shows a diminished trend after D5.

Finally, we consider the exposure to the size factor, SMB. We notice a decreasing exposure pattern from D1 to D6 for both configurations across universes, with positive exposures for the lowest deciles D1 and D2. For the complex configurations this pattern is almost strictly decreasing in both universes while the linear configuration presents a noisier pattern near the outer deciles (i.e. D1,D2, D9 and D10) in the CRSP universe. For the linear configuration we observe a less strict decreasing pattern, especially in the S&P500 universe.

All in all, from this analysis we conclude that the complex configuration mainly differs from its linear counterpart in its exposure to the $LowVol$ and HML factor. However, we noted that the difference with regard to the HML factor is not as distinctly present over the deciles. We

observe that both configurations take a short position on small capitalization assets through the exposure to the SMB factor. For the complex configurations we can combine this observation with the remarks on the $LowVol$, $RESMOM$ and REV factor exposures; it is apparent that this configuration tends to allocate assets with higher volatility ($-LowVol$), lower previous performance/returns ($-RESMOM$ & $-REV$) and relatively smaller market capitalization ($+SMB$) to the lowest deciles. This ranking focus is clearly visible across universes in the lower deciles (D1 to D4) which present positive trending patterns for the $LowVol$, $RESMOM$ and REV factors. On the other hand, for the higher deciles (D6-D10) with negative size exposures we observe that these factors show a much noisier pattern with a diminished or non-existent trend. The exposures to the return-based $RESMOM$ and REV factors do show us that the complex system still resembles the linear system in terms of exposures, yet this is mainly present in the lower deciles. On the other hand, we observe that it differentiates from the linear system in terms of its trending $LowVol$ exposure in these same lower deciles. Hence, we conclude that this complex system is still clearly resembles the classical $RESMOM$ and REV benchmark similar to the linear case, yet expands on these by leveraging on volatility for its ranking purposes in the lower deciles. Generally, the upper deciles show noisier patterns with regard to these aspects. Hence, we conclude that for the complex configurations the lower deciles generally contain smaller market capitalization assets. Smaller market capitalization theoretically contain more exploitable market inefficiencies and as such the ranking ability of complex systems improves for these assets with regard to the set of market inefficiencies captured by the $LowVol$, $RESMOM$ and REV factors.

5 | Conclusion

In this research we considered whether complex neural-network systems are able to improve over simpler linear systems as well as non-parametric methods in terms of financial portfolio construction. Specifically, we attempt to determine whether self-attention and ranked input definition could improve over non-ranked definitions and LSTM-based networks as well as linear systems when provided only the monthly returns as input set. To this end, we set up sets of inputs, system definitions, optimization procedures and portfolio construction frameworks in chapter 2. To start our analysis we considered the performance of the non-parametric benchmarks in chapter 3. For these benchmarks, we identify general limitations in profitability as observed by excessive drawdowns and extensive periods of flat cumulative return.

After implementing these parametric systems we analysed the optimization results in section 4.1. Here, we observe that the complex system configurations are able to achieve significantly higher Spearman correlation coefficients than the linear configurations in the CRSP universe. On the other hand, the significant superiority is not exhibited in the S&P500 universe. Furthermore, we show that complex system configurations are not significantly different to each other. In addition, we observed that the Spearman correlation coefficient metric score declines over time, with a more severe and constant decile in the S&P500 universe.

In section 4.2, we analyse the empirical asset ranking ability of signals by examining their empirical rank distributions. From this analysis we constitute three findings. First of all, all considered signals (parametric non-parametric) exhibit some ability to distinguish assets with

true center ranks from those with extreme/tail ranks. Secondly, these complex configurations exclusively predict assets with true extreme/tail ranks as losers. In practice, this means that we obtain an 'inverted U'-shape in the empirical distribution. The linear/non-parametric signals do not show this exclusivity property and present an empirical 'X'-shape. Thirdly, this exclusive loser prediction of all true extreme ranks displayed by the complex configurations achieves a significantly higher Spearman correlation coefficient score.

In section 4.3, we further analyse the empirical asset ranking ability of the complex configurations through an investigation into their predictive power in the presence of other signals through Fama-Macbeth style regressions. From this analysis we find a limited cross-sectional predictive ability of the complex system's signal in the presence of a linear and volatility based signal. On the other hand, the analysis showed that predictive power of the complex system's signal is subsumed after including both and not just either the linear- and volatility-based predictions. Furthermore, we hypothesized a volatility-oriented feature could explain the empirical 'inverted U'-shape for the complex configuration. Indeed, we observed high correlation between this complex configuration and a volatility-oriented signal. We denoted that complex neural networks present the ability to access volatility related features and that these features themselves are inherently predictable in equity returns. Finally, we confirm our hypothesis through the fact that the established negative relationship between idiosyncratic volatility and returns by Chen and Petkova [2012] is clearly apparent in our regression results.

After these analyses into the empirical ranking ability of our systems, we consider performance on portfolio level. In section 4.4, we analysed the decile portfolios' performance statistics. We observed that the complex systems generally showed a non-monotonically increasing pattern in excess return from D1 to D10. In addition, we observe a substantial volatility skew over the deciles peaking in the predicted lower deciles. Hence, for the complex configurations We conclude that the use of these volatility-oriented features for rank prediction purposes might be beneficial on cross-sectional level, yet not on the aggregated portfolio level. This stands in contrast to the significant superiority in terms of Spearman correlation coefficient of these configurations. Hence, we do not recommend the use of this metric to measure/predict portfolio performance for predictions from systems with distinctly different empirical structures such as the 'X'-shape and 'inverted U'-shape.

In section 4.5 continue the analysis on portfolio level, by considering the focal portfolio D10-D1. This analysis shows that the complex systems are inferior to the linear counterparts in common performance metrics. In addition, the complex configurations show great similarity on portfolio level amongst themselves. This result further suggests that ranking on volatility is detrimental on the aggregated portfolio level. This is attributed to these configurations missing out on the benefits of cancelling extreme returns, due to their exclusive allocation to of assets with extreme returns to the D1 portfolio. Consequently, we observe a decreased average return and increased average volatility in the D10-D1 portfolio. Specifically, we observe that this increased volatility of D10-D1 occurs due to differences in the recovery speed between D10 and D1 around drawdowns.

In section 4.6 we completed our analysis on portfolio level by investigating the exposure of the system's portfolios to common risk factors. We first noted that a high correlation between the complex configurations further indicated that these systems are not distinctly different from eachother. To continue, the D10-D1 factor regression analysis shows a significant negative tilt towards size and, expectantly, positive exposures to the return-based factors REV and $RESMOM$ across configurations and universes. Furthermore, for the complex configurations a significantly positive tilt towards the $LowVol$ factor is observed. We expanded this analysis by considering the exposures of every decile portfolio. In the exposures we observe a clear trending exposure patterns exists between D1 and D5 with regard to the SMB, REV, $RESMOM$

and $LowVol$ factor for the complex configurations. Ultimately, we concluded that the complex system still clearly resembles the classical $RESMOM$ and REV benchmark similar to the linear configuration, yet expands on these by leveraging on volatility for its ranking purposes in the lower deciles. Generally, the upper deciles show noisier patterns with regard to these aspects.

All in all, in this research we implemented and analysed a set of complex machine-learning based configurations. From our analysis we concluded that a linear parametric system is able to attain competitive or superior performance on portfolio level to these more complex signal implementations. Furthermore, we conclude that nuances between the complex configurations might exist, but their general performance in not significantly distinguishable. In an attempt to determine the difference between a linear and more complex parametric system's prediction, we identified that the complex implementations utilize volatility related features in their predictions. While the use of these features is beneficial on an empirical asset ranking level in the cross-section, improved performance was not visible on portfolio level resulting in portfolios with lower average returns and higher volatility.

Our research presents three main limitations. First of all, we estimate our systems on US equity data contained in the CRSP dataset and evaluate on the S&P500 universe of assets. This attempt at 'transfer learning' is shown to be inadequate, yet without a comparison to systems estimated on S&P500 specifically. Secondly, we limit our analysis on portfolio level to metrics capturing performance. However, practical considerations such as turnover, transactions costs and the impact of implementation lag are not considered. Finally, in this research we limit ourselves to use of end-of-month returns and do not consider the effects (e.g. seasonality) at this specific calendar frequency.

Finally, we provide recommendations for future work. First of all, in section 4.2 we provide examples of alternative statistics to score the relationship between predicted and true asset rankings. Contrary to the Spearman correlation coefficient, these alternative statistics also score the asymmetry/imbalance of the solution. We do not further investigate the optimization of such metrics which remain unexplored. Secondly, in section 4.4 we recognize that the complex systems fail at decoupling risk into idiosyncratic and systemic parts. Since we denote a significant importance of risk proxied by volatility in our complex systems, we recommend further work into an improved reduction of market effects in asset returns and risk. Third of all, we observe significant intercepts in our spanning regression results in section 4.6, yet acknowledge that our attempts are limited in terms of identification of their specific sources.

A | Additional analysis: complex network comparison over time

Temporal information flow analysis of self-attention and recurrent networks

In this section, we investigate whether the differences in implicit and explicit temporal information flow structures as discussed in subsection 2.3.2 are apparent between the CAF and CLF configurations. To this end, consider again high-level subsystem overview of the LSTM and self-attention framework given in Figure 2.3 and Figure 2.5 respectively. Both systems utilize 4 time steps of inputs denoted $\{X_{t-4}, \ldots, X_{t-1}\}$ to obtain a single output denoted as Z_{t-1}. All of these vectors are of dimension f, which in our configurations is set equal to 32. In order to analyse the difference in temporal information flow, we setup the correlation matrices between one of the time-step inputs (i.e. X_{t-j}) and the output signal Z_{t-1} for every sample. In this correlation matrix of dimension $(64, 64)$ we focus on the non-symmetric and non-diagonal $(32, 32)$ matrix sub-block. If we rank the Pearson correlation coefficients observed within this against the matrix sub-block for the other time steps in the $[0, 1]$ and aggregate the result a mean statistic over this entire sub-block we obtain 4 sets of scalar mean ranked Pearson correlation coefficients. These statistics can be analyzed over time in an attempt to determine whether the implicit or explicit network structures exhibit differences in temporal informational flow. A clear limitation of this analysis is the fact that both of these networks are designed to allow for nonlinear and interactive effects which will not captured by the linear Pearson correlation coefficient. However, we continue the analysis anyway in an attempt to shed light on the linear aspect of the temporal information flow. The procedure to setup these aggregated statistcis is visualized in Figure A.1.

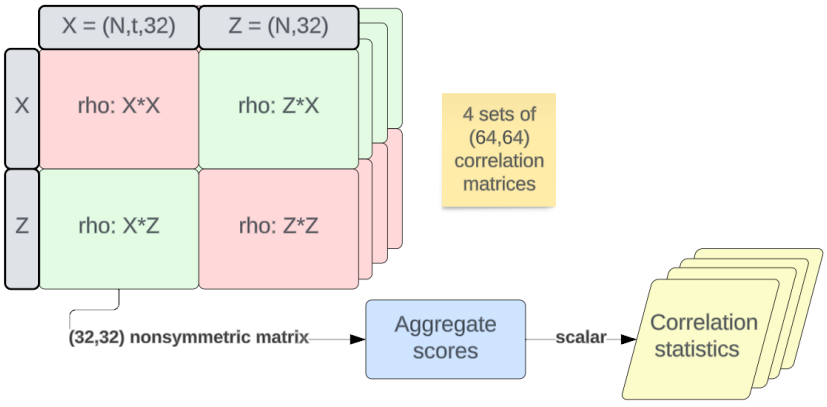


Figure A.1: Procedure for subsystem input and output to obtain a set of 4 scalar aggregated correlation statistics.

We hypothesize that the LSTM-based networks should generally focus on the closest time step (i.e. $t-1$) as every subsequent time step is 'deeper' in the system and harder to reach. As such, we expect that the inputs associated with this time step $t-1$ should obtain a consistent higher rank than the other time steps. Furthermore, we might also be logical because of the expected higher relevancy of more recent time steps to the prediction objective. On the other hand, the self-attention framework allows for the explicit temporal flow. As such we hypothesise that in

contrast to the LSTM-based networks (i.e. CLF) the self-attention networks (i.e. CAF) might exhibit less consistency in correlation ranks, obtaining a noisier pattern.

The results of this analysis are displayed in Figure A.2. In general, we do not observe that there is a distinct structure of correlation ranks for the CLF configurations. In addition, the hypothesized less strict importance of time steps for the CLF configurations is also not apparent. In fact, main contributor to the change in input time step importance seems to be the refitting period as annotated by the red dashed vertical lines. All in all, this analysis shows us that the importance of different time steps of inputs to the output is indifferent between the respective subsystems. Note that this conclusion only hold with regard to their linear contribution as was captured by the Pearson correlation coefficient.

Figure A.2: Mean ranks of correlation statistics for the complex configurations defined in chapter 2. Results obtained in CRSP universe and in the out-of-sample period 1995/01-2022/12. Expanding window applied for the complex system configurations with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years. New refitted out-of-sample subperiod start indicated by the red dashed vertical lines.

B | On the relationship between MSE and SPR optimization

In this section we address the equivalence between optimization of the Spearman rank correlation coefficient and optimization of the MSE between true and predicted rank. We define the true rank and predicted rank as $y_{t,i}$ and $\hat{y}_{t,i}$ respectively. To achieve this equality we prove the equivalence between MSE minimization and Pearson correlation coefficient maximization. This proof can then analogously be extended to MSE rank minimization and Spearman correlation coefficient.

First, we consider MSE minimization and show that the optimization is focused on maximization of the product between vectors y and $\hat{y}$. Mathematically, this is represented through,

$$\begin{aligned}
\text{argmin}_\theta \mathcal{L} &= (y - M(x;\theta))^2 \\
&= y^2 - 2 \cdot y \cdot M(x;\theta) + M(x;\theta)^2 \\
&= y^2 - 2 \cdot y \cdot \hat{y} + \hat{y}^2, \quad \text{where} \quad \hat{y} = M(x;\theta) \\
\text{argmax}_\theta \mathcal{L} &= \underbrace{-y^2}_{\text{Not optimized}} + 2 \cdot y \cdot \hat{y} - \underbrace{\hat{y}^2}_{*} \\
&\propto y \cdot \hat{y} - \underbrace{\hat{y}^2}_{*}
\end{aligned} \tag{B.1}$$

where $M(x;\theta)$ denotes the prediction model $M()$ subject to parameters θ and input data x. For the Pearson correlation coefficient we show that the optimization is focused on maximization of the product between vectors y and $\hat{y}$ as well. Mathematically, this is represented as,

$$\begin{aligned}
\text{argmax}_\theta \mathcal{L} &= \frac{(y - \mu_y)(\hat{y} - \mu_{\hat{y}})}{\sigma_y \sigma_{\hat{y}}} \\
&= \frac{y \cdot \hat{y} - \mu_y \cdot \hat{y} - \mu_{\hat{y}} \cdot y + \mu_y \cdot \mu_{\hat{y}}}{\sigma_y \sigma_{\hat{y}}} \\
&\propto \underbrace{\frac{y \cdot \hat{y}}{\sigma_{\hat{y}}}}_{*}
\end{aligned} \tag{B.2}$$

where μ and σ are the mean and standard deviation respectively. It is important to note that the variance term (annotated by $*$) of the prediction vector is proportional to $\hat{y}^2$ or $\sigma_{\hat{y}}$. All in all, one should recognize from both these derivations that a parameterized system intent on maximizing either the Pearson correlation coefficient or minimizing the mean-squared error is similar. The system will maximize the 'product' term $(y \cdot \hat{y})$ and minimize the 'variance' term ($\sigma_{\hat{y}}$ or $\hat{y}^2$).

C | Charting By Machines Reproduction results

In this appendix we present the result for our reproduction attempt of *Charting by Machines* by Murray et al. [2021]. In their research, Murray et al. [2021] dub their system to the *CNN-CLF* and its predictions to $MLER$. In our research, we replicate this system in chapter 2 which we call *CR-CLF* in our research. In this appendix we adhere to their signal definition and also call our signal predictions $MLER$. Compared to their system, the $CR-CLF$ configuration differs in the fact that it uses cumulative returns based on excess market return. In their case, they opt for the use of risk free rate returns. In practice, no difference is encountered from this setting. All in all, the methodology and data as described by Murray et al. [2021] are followed in an attempt to reproduce two main results. First, the Spearman rank correlation for a variety of system configurations is provided in Table C.1. Secondly, the descriptive statistics associated with the decile portfolios for the first out-of-sample period ($MLER_{i,t}^{1927/01,1963/06}$ for $1963/07 \leq t \leq 1974/12$) are provided in Table C.2. Finally, the cumulative log excess (risk-free rate) return for the $MLER_{10-1}$ portfolio of this out-of-sample period are provided in Figure C.1 and Figure C.2.

Table C.1: Spearman rank correlation in percent for system configurations described by Murray et al. [2021] in *Charting by Machines* Table 1. The *CBM* row indicates the original results and *Thesis* row indicates the reproduction attempt results.

	Dependent variable	Weighting methodology	FNN MAE	CNN MAE	LSTM MAE	CNN-LSTM MAE
CBM	r_{norm}	EWPM	6.9	9.1	10.1	10.6
Thesis	r_{norm}	EWPM	6.8	6.6	10.1	10.3
		$n_{parameters}$	9217	16897	17217	37889

Table C.2: Mean excess return ($\bar{r}$) in % and annualized Sharpe ratio for the CNN-LSTM system architecture described by Murray et al. [2021] in *Charting by Machines* Table 5. Results are shown for the decile portfolios in the $1963/07 \leq t \leq 1974/12$ out-of-sample period based on the $MLER_{i,t}^{1927/01,1963/06}$ predictions. Decile breakpoints are computed using only NYSE stocks. The *CBM* row indicates the original results and *Thesis* row indicates the reproduction attempt results.

	Portfolio	$MLER_1$	$MLER_2$	$MLER_3$	$MLER_4$	$MLER_5$	$MLER_6$	$MLER_7$	$MLER_8$	$MLER_9$	$MLER_{10}$	$MLER_{10-1}$
CBM	$\bar{r}$	-0.90	-0.47	-0.26	-0.14	-0.20	-0.03	0.15	0.07	-0.03	0.25	1.15
	Sharpe	-0.56	-0.32	-0.19	0.10	-0.15	-0.03	0.12	0.06	-0.02	0.17	1.16
Thesis	$\bar{r}$	-0.86	-0.46	-0.15	-0.29	-0.15	0.05	0.01	0.15	0.05	0.31	1.17
	Sharpe	-0.57	-0.33	-0.11	-0.22	-0.11	0.04	0.01	0.12	0.04	0.21	1.29

Figure C.1: Cumulative log return for $MLER_{10-1}$ portfolio as presented in *Charting by Machines* Figure 2 by Murray et al. [2021].

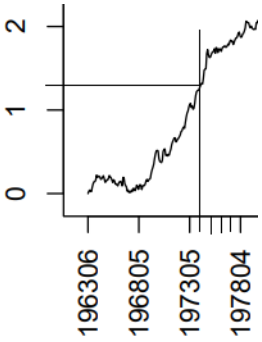
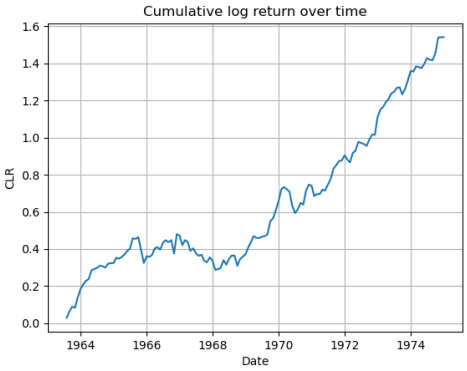


Figure C.2: Cumulative log returns for $MLER_{10-1}$ portfolio from the reproduction attempt.



By comparing all of the results from the reproduction attempt to those in the original *Charting by Machines* paper, we conclude general and consistent close alignment between results. Finally, we present the decile portfolio statistics in Table C.3 to conclude this overview of results.

Table C.3: Annualized Sharpe ratio and mean monthly in [%] excess return for $MLER$, MOM and REV signal portfolios in the CRSP and S&P500 universes in the period 1995/01-2014/12. *CRSP1926x* denotes portfolios constructed using the CNN-LSTM model trained on all available CRSP data up to 1994 while *CRSP1975x* denotes only data between 1975-1995 is used. *x<Dataset>* denotes the out-of-sample data set.

		D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D10m1
CRSP1926xCRSP	$\bar{r}$	0.272	0.656	0.643	0.641	0.587	0.94	0.794	0.706	0.944	0.969	0.697
	Sharpe	0.136	0.36	0.408	0.45	0.459	0.747	0.647	0.573	0.719	0.571	0.462
CRSP1975xCRSP	$\bar{r}$	0.333	0.61	0.71	0.743	0.62	0.877	0.666	0.768	0.911	1.057	0.724
	Sharpe	0.136	0.308	0.413	0.476	0.435	0.645	0.55	0.639	0.783	0.799	0.381
MOM (CRSP)	$\bar{r}$	0.316	0.619	0.713	0.792	0.735	0.667	0.737	0.853	0.721	1.011	0.695
	Sharpe	0.112	0.304	0.428	0.549	0.563	0.539	0.615	0.721	0.556	0.559	0.279
REV (CRSP)	$\bar{r}$	0.318	0.388	0.728	0.821	1.098	0.902	0.962	0.705	0.743	0.386	0.067
	Sharpe	0.251	0.283	0.492	0.566	0.722	0.622	0.632	0.5	0.567	0.297	0.113

D | Complete tables and figures

D.1 Non-parametric benchmark decile excess return, volatility and Sharpe

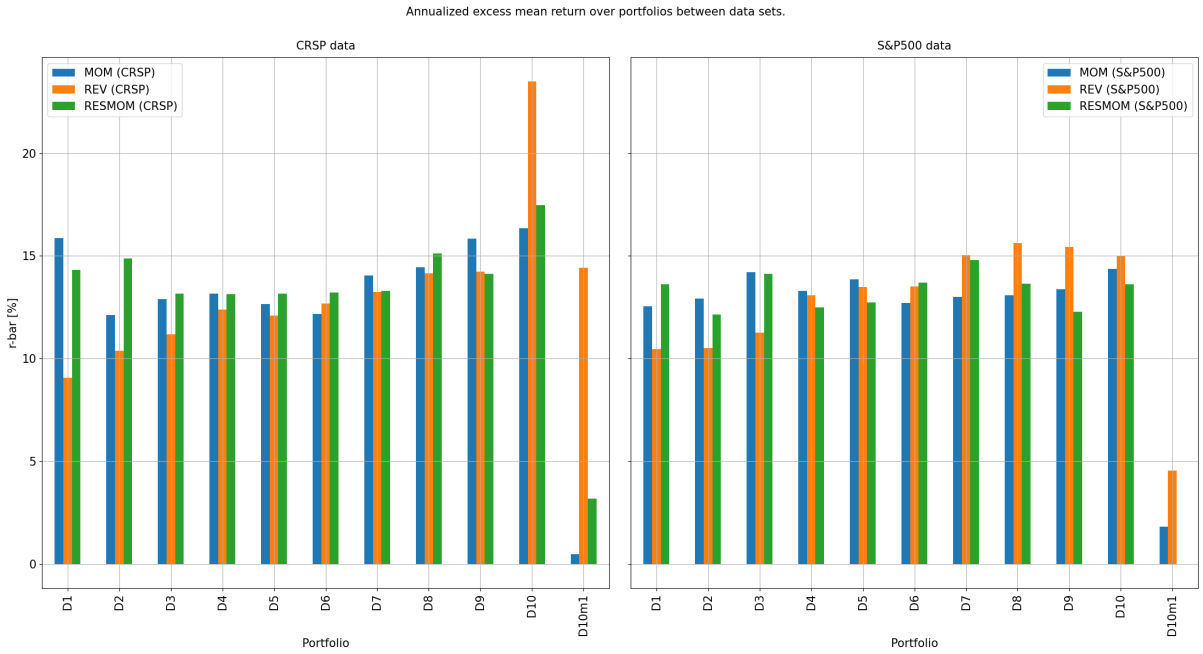


Figure D.1: Equally-weighted decile portfolio annualized mean excess return in the CRSP and S&P500 universes in the OOS period 1995/01-2022/12.

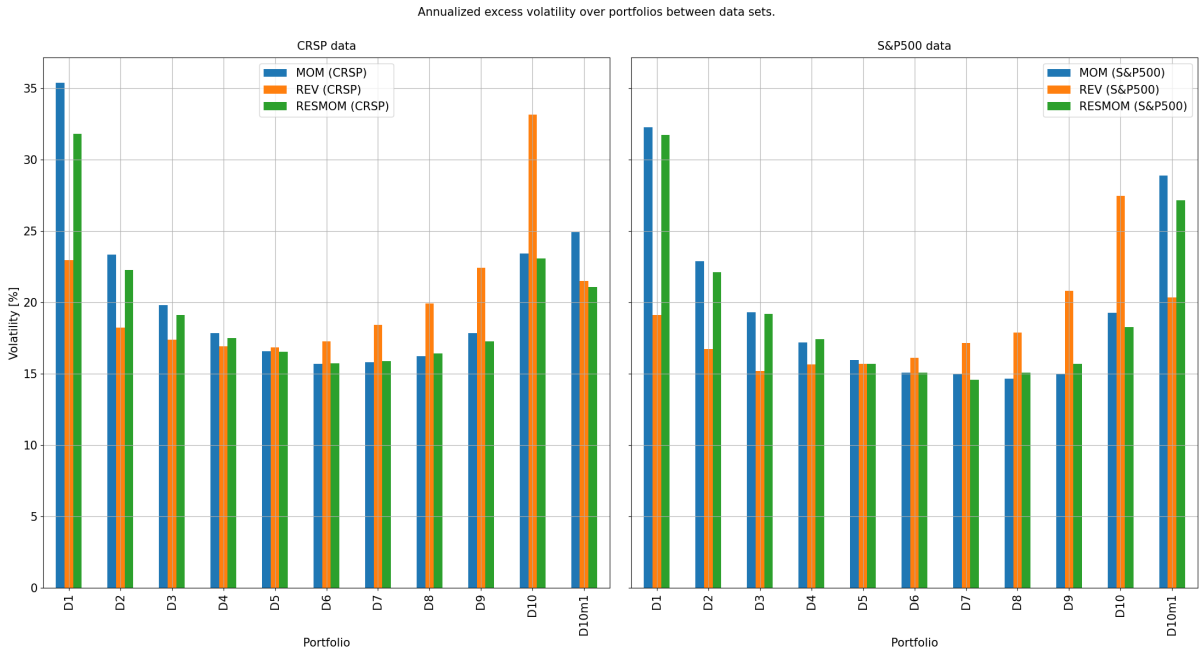


Figure D.2: Equally-weighted decile portfolio annualized return volatility in the CRSP and S&P500 universes in the OOS period 1995/01-2022/12.

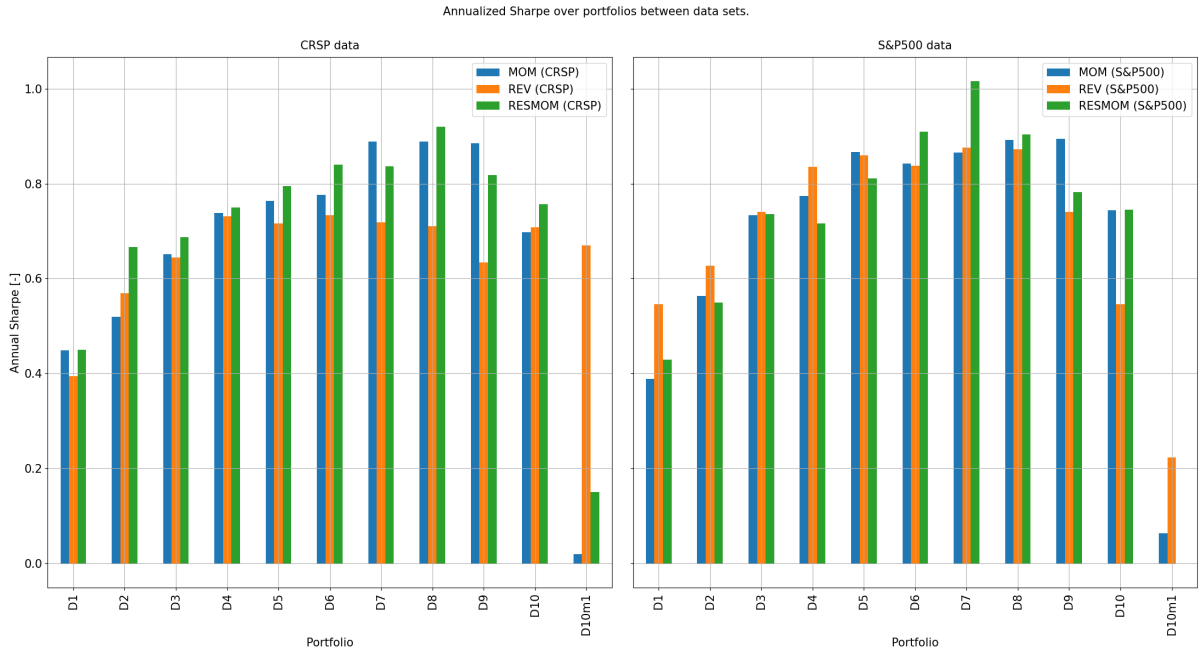


Figure D.3: Equally-weighted decile portfolio annualized Sharpe in the CRSP and S&P500 universes in the OOS period 1995/01-2022/12.

D.2 Optimization score related Diebold-Mariano test statistics

MAE x CRSP	CR CLF32	CP CLF32	CR CAF32	CP CAF32	rtnXmkt-Linear
CR CLF32					
CP CLF32	0.0001				
CR CAF32	0.0000	0.0000			
CP CAF32	0.0000	0.0645	0.0		
rtnXmkt-Linear	0.0000	0.0013	0.0	0.0059	
rtnXmkt-WLS	0.0000	0.0000	0.0	0.0000	0.0

Table D.1: p-values for Diebold-Mariano 2-sided tests applied to monthly MAE statistics for system configurations over the period 1995-2022 in the CRSP universe. Expanding window applied for the complex system configurations with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years.

MAE x S&P500	CR CLF32	CP CLF32	CR CAF32	CP CAF32	rtnXmkt-Linear
CR CLF32					
CP CLF32	0.0000				
CR CAF32	0.0000	0.0000			
CP CAF32	0.0000	0.5007	0.0		
rtnXmkt-Linear	0.0297	0.3953	0.0	0.3104	
rtnXmkt-WLS	0.0000	0.0000	0.0	0.0000	0.0

Table D.2: p-values for Diebold-Mariano 2-sided tests applied to monthly MAE statistics for system configurations over the period 1995-2022 in the S&P500 universe. Expanding window applied for the complex system configurations with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years.

MSE x CRSP	CR CLF32	CP CLF32	CR CAF32	CP CAF32	rtnXmkt-Linear
CR CLF32					
CP CLF32	0.5987				
CR CAF32	0.0000	0.0000			
CP CAF32	0.4394	0.7071	0.0		
rtnXmkt-Linear	0.0003	0.0004	0.0	0.0003	
rtnXmkt-WLS	0.0000	0.0000	0.0	0.0000	0.0

Table D.3: p-values for Diebold-Mariano 2-sided tests applied to monthly MSE statistics for system configurations over the period 1995-2022 in the CRSP universe. Expanding window applied for the complex system configurations with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years.

MSE x S&P500	CR CLF32	CP CLF32	CR CAF32	CP CAF32	rtnXmkt-Linear
CR CLF32					
CP CLF32	0.2200				
CR CAF32	0.0000	0.0000			
CP CAF32	0.7062	0.1764	0.0		
rtnXmkt-Linear	0.9484	0.5632	0.0	0.9104	
rtnXmkt-WLS	0.0000	0.0000	0.0	0.0000	0.0

Table D.4: p-values for Diebold-Mariano 2-sided tests applied to monthly MSE statistics for system configurations over the period 1995-2022 in the S&P500 universe. Expanding window applied for the complex system configurations with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years.

SPR x CRSP	CR CLF32	CP CLF32	CR CAF32	CP CAF32	rtnXmkt-Linear	rtnXmkt-WLS	MOM	REV
CR CLF32								
CP CLF32	0.4057							
CR CAF32	0.0001	0.0017						
CP CAF32	0.0608	0.2441	0.0007					
rtnXmkt-Linear	0.0000	0.0000	0.0001	0.0				
rtnXmkt-WLS	0.0000	0.0000	0.0000	0.0	0.6273			
MOM	0.0000	0.0000	0.0000	0.0	0.7709	0.4260		
REV	0.0000	0.0000	0.0000	0.0	0.0000	0.0002	0.0139	
RESMOM	0.0000	0.0000	0.0000	0.0	0.0006	0.0000	0.0000	0.3253

Table D.5: p-values for Diebold-Mariano 2-sided tests applied to monthly SPR (Spearman correlation coefficient) statistics for system configurations and benchmarks over the period 1995-2022 in the CRSP universe. Expanding window applied for the complex system configurations with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years.

SPR x S&P500	CR CLF32	CP CLF32	CR CAF32	CP CAF32	rtnXmkt-Linear	rtnXmkt-WLS	MOM	REV
CR CLF32								
CP CLF32	0.1604							
CR CAF32	0.0054	0.0001						
CP CAF32	0.4673	0.0094	0.0013					
rtnXmkt-Linear	0.4731	0.1565	0.3350	0.6696				
rtnXmkt-WLS	0.1326	0.0446	0.9151	0.2252	0.3215			
MOM	0.0358	0.0044	0.4293	0.0267	0.1452	0.8119		
REV	0.5675	0.3481	0.6621	0.7314	0.9074	0.4443	0.5092	
RESMOM	0.0030	0.0001	0.0660	0.0017	0.0145	0.3767	0.0795	0.2304

Table D.6: p-values for Diebold-Mariano 2-sided tests applied to monthly SPR (Spearman correlation coefficient) statistics for system configurations and benchmarks over the period 1995-2022 in the S&P500 universe. Expanding window applied for the complex system configurations with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years.

D.3 Parametric configurations decile portfolio statistics and factor exposures

CRSP	**D1**	**D2**	**D3**	**D4**	**D5**	**D6**	**D7**	**D8**	**D9**	**D10**	**D10-D1**
CR CLF32 (mean)	11.93	16.11	15.61	15.34	14.63	14.56	14.02	14.69	15.98	15.96	4.02
CR CLF32 (volatility)	32.17	25.34	22.02	19.74	17.82	16.68	16.04	15.51	15.17	15.45	23.51
CR CLF32 (Sharpe)	0.37	0.64	0.71	0.78	0.82	0.87	0.87	0.95	1.05	1.03	0.17
CP CLF32 (mean)	9.25	15.51	15.25	16.29	15.48	14.75	14.89	15.63	15.71	16.96	7.71
CP CLF32 (volatility)	30.73	24.98	21.90	19.81	18.38	17.71	16.54	15.83	16.09	16.17	22.16
CP CLF32 (Sharpe)	0.30	0.62	0.70	0.82	0.84	0.83	0.90	0.99	0.98	1.05	0.35
CR CAF32 (mean)	16.74	13.61	14.57	14.41	14.19	15.06	14.43	13.65	14.72	13.47	-3.27
CR CAF32 (volatility)	35.09	24.78	21.03	18.84	17.76	16.92	16.12	15.82	14.89	14.08	27.93
CR CAF32 (Sharpe)	0.48	0.55	0.69	0.77	0.80	0.89	0.90	0.86	0.99	0.96	-0.12
CP CAF32 (mean)	9.60	16.53	16.70	16.37	15.87	15.37	14.50	14.56	14.17	14.64	5.03
CP CAF32 (volatility)	31.07	25.54	23.13	20.14	18.50	17.11	16.01	14.98	14.84	15.72	22.99
CP CAF32 (Sharpe)	0.31	0.65	0.72	0.81	0.86	0.90	0.91	0.97	0.96	0.93	0.22
rtnXmkt-Linear (mean)	9.25	11.79	12.11	13.07	13.60	14.27	14.51	15.22	16.84	20.78	11.54
rtnXmkt-Linear (volatility)	29.17	21.02	18.89	17.42	17.02	16.60	17.04	17.92	20.29	27.31	15.12
rtnXmkt-Linear (Sharpe)	0.32	0.56	0.64	0.75	0.80	0.86	0.85	0.85	0.83	0.76	0.76
rtnXmkt-WLS (mean)	8.94	13.07	12.15	12.87	13.91	13.26	14.45	14.25	15.71	18.12	9.18
rtnXmkt-WLS (volatility)	31.27	22.04	19.76	17.51	17.22	16.29	16.63	17.44	19.00	25.24	16.90
rtnXmkt-WLS (Sharpe)	0.29	0.59	0.62	0.74	0.81	0.81	0.87	0.82	0.83	0.72	0.54
MOM (mean)	15.87	12.12	12.90	13.17	12.66	12.17	14.05	14.44	15.83	16.34	0.47
MOM (volatility)	35.42	23.36	19.79	17.85	16.57	15.69	15.80	16.24	17.86	23.41	24.95
MOM (Sharpe)	0.45	0.52	0.65	0.74	0.76	0.78	0.89	0.89	0.89	0.70	0.02
REV (mean)	9.07	10.38	11.19	12.38	12.09	12.68	13.24	14.15	14.23	23.49	14.42
REV (volatility)	22.95	18.24	17.38	16.94	16.86	17.28	18.41	19.90	22.44	33.13	21.53
REV (Sharpe)	0.40	0.57	0.64	0.73	0.72	0.73	0.72	0.71	0.63	0.71	0.67
RESMOM (mean)	14.31	14.87	13.15	13.14	13.16	13.21	13.30	15.13	14.12	17.48	3.17
RESMOM (volatility)	31.81	22.29	19.11	17.52	16.55	15.72	15.89	16.44	17.26	23.10	21.13
RESMOM (Sharpe)	0.45	0.67	0.69	0.75	0.80	0.84	0.84	0.92	0.82	0.76	0.15

S&P500	**D1**	**D2**	**D3**	**D4**	**D5**	**D6**	**D7**	**D8**	**D9**	**D10**	**D10-D1**
CR CLF32 (mean)	11.85	11.23	12.33	12.21	12.74	13.11	13.53	14.17	15.38	16.76	4.91
CR CLF32 (volatility)	28.27	22.01	19.88	17.37	15.95	15.44	15.55	14.77	15.78	15.88	19.88
CR CLF32 (Sharpe)	0.42	0.51	0.62	0.70	0.80	0.85	0.87	0.96	0.98	1.06	0.25
CP CLF32 (mean)	9.17	11.21	12.41	12.24	14.17	13.69	12.74	14.89	15.35	17.46	8.29
CP CLF32 (volatility)	26.50	22.24	18.95	17.85	16.73	15.52	15.57	15.86	16.03	16.26	19.23
CP CLF32 (Sharpe)	0.35	0.50	0.66	0.69	0.85	0.88	0.82	0.94	0.96	1.07	0.43
CR CAF32 (mean)	13.49	12.13	12.79	14.15	14.31	12.75	12.62	13.00	14.10	13.94	0.45
CR CAF32 (volatility)	31.23	22.67	19.65	17.60	16.64	15.82	15.37	14.87	14.58	14.58	24.83
CR CAF32 (Sharpe)	0.43	0.54	0.65	0.80	0.86	0.81	0.82	0.87	0.97	0.96	0.02
CP CAF32 (mean)	9.96	11.75	11.63	14.38	13.94	13.32	14.63	14.50	14.68	14.55	4.59
CP CAF32 (volatility)	26.70	22.09	19.94	17.78	16.88	16.25	15.69	14.89	14.97	16.06	19.94
CP CAF32 (Sharpe)	0.37	0.53	0.58	0.81	0.83	0.82	0.93	0.97	0.98	0.91	0.23
rtnXmkt-Linear (mean)	9.57	11.44	12.54	12.75	13.63	14.27	14.44	15.09	13.85	15.75	6.18
rtnXmkt-Linear (volatility)	25.44	20.08	17.92	16.55	15.40	15.70	15.71	16.81	17.04	22.60	20.89
rtnXmkt-Linear (Sharpe)	0.38	0.57	0.70	0.77	0.89	0.91	0.92	0.90	0.81	0.70	0.30
rtnXmkt-WLS (mean)	10.88	13.40	12.25	12.25	13.33	13.99	13.33	13.70	14.97	15.21	4.33
rtnXmkt-WLS (volatility)	22.72	18.11	16.97	16.06	15.72	16.26	16.08	16.83	18.10	24.07	17.96
rtnXmkt-WLS (Sharpe)	0.48	0.74	0.72	0.76	0.85	0.86	0.83	0.81	0.83	0.63	0.24
MOM (mean)	12.54	12.91	14.19	13.30	13.85	12.71	13.00	13.08	13.38	14.36	1.82
MOM (volatility)	32.23	22.90	19.34	17.18	15.97	15.07	15.02	14.65	14.95	19.27	28.89
MOM (Sharpe)	0.39	0.56	0.73	0.77	0.87	0.84	0.87	0.89	0.90	0.75	0.06
REV (mean)	10.45	10.50	11.26	13.07	13.49	13.51	15.04	15.63	15.43	15.00	4.55
REV (volatility)	19.13	16.75	15.22	15.64	15.68	16.12	17.17	17.90	20.82	27.47	20.31
REV (Sharpe)	0.55	0.63	0.74	0.84	0.86	0.84	0.88	0.87	0.74	0.55	0.22
RESMOM (mean)	13.63	12.15	14.13	12.49	12.74	13.69	14.79	13.63	12.29	13.62	-0.01
RESMOM (volatility)	31.69	22.09	19.20	17.44	15.71	15.06	14.56	15.08	15.71	18.28	28.29
RESMOM (Sharpe)	0.43	0.55	0.74	0.72	0.81	0.91	1.02	0.90	0.78	0.75	0.00

Table D.7: Decile equally-weighted portfolio statistics for the investigated configurations. Annualized excess return [%], excess volatility [%] and Sharpe ratio [-] for complex and linear system configuration, MOM, $RESMOM$ and REV signal portfolios in the CRSP and S&P500 universes in the period 1995/01-2022/12. Expanding window applied with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every year.

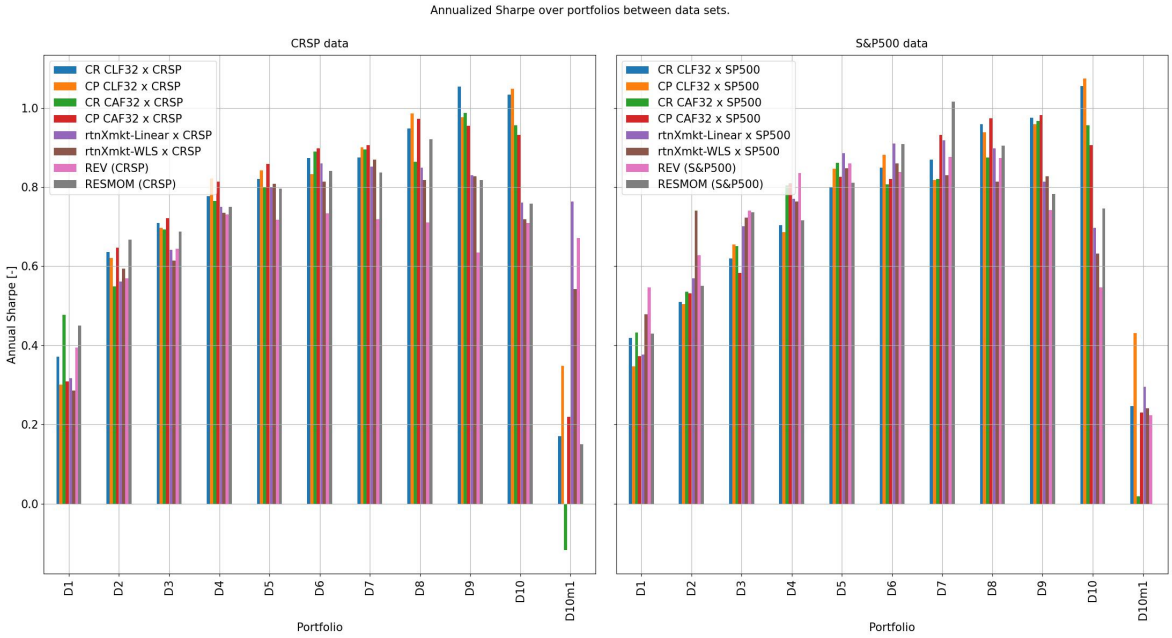


Figure D.4: Annualized excess (risk free rate) Sharpe ratio across equally-weighted deciles in the CRSP and S&P500 universes in the OOS period 1995/01-2022/12. Expanding window applied for the complex system configurations with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years. Histogram values correspond to those in Table D.7.

CR-CLF	D10-D1	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
alpha (CRSP)	4.841 (2.164)	-2.273 (-1.579)	1.613 (2.003)	0.747 (0.96)	1.381 (1.724)	1.006 (1.352)	1.08 (1.496)	0.76 (0.941)	1.379 (1.712)	2.526 (2.862)	2.568 (2.376)
Mkt-RF (CRSP)	-0.195 (-3.364)	0.065 (1.75)	0.038 (1.816)	0.06 (2.985)	-0.004 (-0.178)	-0.037 (-1.931)	-0.064 (-3.438)	-0.069 (-3.31)	-0.091 (-4.387)	-0.096 (-4.186)	-0.13 (-4.633)
LowVol (CRSP)	0.375 (7.108)	-0.23 (-6.765)	-0.1 (-5.243)	0.004 (0.192)	0.024 (1.258)	0.057 (3.235)	0.072 (4.225)	0.109 (5.726)	0.122 (6.441)	0.146 (7.027)	0.145 (5.703)
SMB (CRSP)	-0.869 (-11.963)	0.509 (10.887)	0.187 (7.17)	0.026 (1.034)	-0.06 (-2.318)	-0.131 (-5.445)	-0.21 (-8.953)	-0.248 (-9.463)	-0.318 (-12.174)	-0.329 (-11.488)	-0.36 (-10.262)
HML (CRSP)	0.082 (1.352)	-0.082 (-2.091)	-0.019 (-0.873)	0.047 (2.215)	0.061 (2.806)	0.076 (3.751)	0.09 (4.597)	0.062 (2.815)	0.062 (2.834)	0.035 (1.475)	0.0 (0.015)
RESMOM (CRSP)	0.339 (7.689)	-0.248 (-8.742)	-0.09 (-5.657)	0.016 (1.012)	0.041 (2.62)	0.089 (6.057)	0.116 (8.161)	0.107 (6.75)	0.108 (6.785)	0.128 (7.378)	0.091 (4.272)
REV (CRSP)	0.297 (8.38)	-0.125 (-5.494)	-0.037 (-2.869)	-0.008 (-0.613)	-0.001 (-0.044)	0.017 (1.472)	0.038 (3.329)	0.058 (4.518)	0.095 (7.477)	0.125 (8.93)	0.172 (10.031)

CR-CLF	D10-D1	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
alpha (S&P500)	5.934 (2.676)	-1.928 (-1.204)	-2.149 (-1.831)	-1.998 (-1.775)	-1.188 (-1.297)	-0.232 (-0.299)	-0.155 (-0.2)	-0.019 (-0.024)	1.232 (1.557)	2.4 (2.397)	4.006 (3.754)
Mkt-RF (S&P500)	0.007 (0.126)	-0.039 (-0.948)	-0.036 (-1.218)	0.066 (2.29)	0.019 (0.809)	0.009 (0.432)	0.015 (0.772)	0.025 (1.241)	-0.005 (-0.27)	-0.02 (-0.786)	-0.032 (-1.161)
LowVol (S&P500)	0.218 (7.176)	-0.206 (-9.396)	-0.091 (-5.653)	-0.007 (-0.48)	0.028 (2.22)	0.042 (4.015)	0.043 (4.071)	0.067 (6.168)	0.063 (5.797)	0.052 (3.826)	0.012 (0.806)
SMB (S&P500)	-0.677 (-5.227)	0.411 (4.394)	0.415 (6.06)	0.187 (2.838)	0.012 (0.226)	-0.139 (-3.086)	-0.132 (-2.918)	-0.109 (-2.361)	-0.187 (-4.05)	-0.196 (-3.351)	-0.266 (-4.264)
HML (S&P500)	-0.114 (-1.55)	0.04 (0.755)	0.039 (0.996)	0.063 (1.69)	0.096 (3.17)	0.054 (2.12)	-0.063 (-2.44)	-0.063 (-2.416)	-0.058 (-2.21)	-0.036 (-1.082)	-0.074 (-2.088)
RESMOM (S&P500)	0.34 (10.45)	-0.255 (-10.829)	-0.118 (-6.862)	-0.029 (-1.741)	0.01 (0.75)	0.028 (2.435)	0.073 (6.451)	0.07 (6.042)	0.08 (6.923)	0.057 (3.898)	0.086 (5.457)
REV (S&P500)	0.193 (5.632)	-0.081 (-3.27)	-0.137 (-7.587)	-0.078 (-4.484)	-0.04 (-2.853)	-0.017 (-1.389)	0.027 (2.28)	0.058 (4.729)	0.052 (4.226)	0.105 (6.815)	0.112 (6.793)

Table D.8: Annualized alpha statistics for the $CR - CLF$ configuration. All deciles portfolios are excess market returns, except for the D10-D1 portfolio. All values are fitted coefficients with t-stat in brackets in the OOS period 1995/01-2022/12. Expanding window applied with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years. Note that Newey-West 12 lags standard errors are used for SE computation.

rtnXmkt-Linear	D10-D1	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
alpha (CRSP)	3.492 (1.942)	-2.461 (-1.861)	-0.335 (-0.413)	0.142 (0.177)	1.236 (1.686)	1.148 (1.57)	1.51 (2.107)	0.815 (1.292)	0.889 (1.04)	0.585 (0.62)	1.031 (0.96)
Mkt-RF (CRSP)	-0.026 (-0.556)	0.07 (2.059)	0.025 (1.202)	-0.027 (-1.313)	-0.065 (-3.435)	-0.057 (-3.006)	-0.065 (-3.532)	-0.038 (-2.322)	0.004 (0.159)	0.022 (0.909)	0.045 (1.604)
LowVol (CRSP)	-0.092 (-1.679)	-0.084 (-2.688)	0.025 (1.303)	0.066 (3.52)	0.072 (4.179)	0.087 (5.066)	0.08 (4.752)	0.074 (5.006)	0.098 (4.862)	0.018 (0.805)	-0.176 (-6.955)
SMB (CRSP)	-0.388 (-6.652)	0.428 (9.977)	-0.026 (-0.986)	-0.085 (-3.268)	-0.125 (-5.245)	-0.179 (-7.534)	-0.201 (-8.658)	-0.204 (-9.979)	-0.212 (-7.655)	-0.174 (-5.671)	0.04 (1.147)
HML (CRSP)	-0.214 (-4.382)	-0.039 (-1.083)	0.11 (5.006)	0.129 (5.927)	0.137 (6.865)	0.112 (5.655)	0.116 (5.957)	0.08 (4.683)	0.02 (0.854)	-0.021 (-0.818)	-0.253 (-8.666)
RESMOM (CRSP)	0.569 (16.076)	-0.363 (-13.927)	-0.12 (-7.499)	-0.074 (-4.673)	-0.021 (-1.425)	0.018 (1.23)	0.069 (4.862)	0.11 (8.838)	0.095 (5.666)	0.155 (8.35)	0.206 (9.76)
REV (CRSP)	0.498 (17.491)	-0.209 (-9.997)	-0.135 (-10.464)	-0.084 (-6.655)	-0.064 (-5.549)	-0.019 (-1.671)	-0.003 (-0.291)	0.032 (3.165)	0.067 (4.912)	0.136 (9.101)	0.289 (16.966)

rtnXmkt-Linear	D10-D1	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
alpha (S&P500)	4.574 (1.812)	-3.164 (-2.195)	-1.358 (-1.215)	-0.88 (-0.976)	-0.196 (-0.211)	0.993 (1.421)	1.221 (1.678)	0.969 (1.125)	1.031 (1.163)	-0.022 (-0.023)	1.409 (0.84)
Mkt-RF (S&P500)	0.027 (0.415)	-0.044 (-1.188)	-0.026 (-0.926)	0.025 (1.11)	-0.003 (-0.119)	-0.019 (-1.088)	0.006 (0.313)	0.019 (0.849)	0.047 (2.087)	0.012 (0.508)	-0.017 (-0.396)
LowVol (S&P500)	-0.145 (-4.212)	-0.098 (-4.948)	-0.004 (-0.242)	0.087 (7.026)	0.077 (6.051)	0.06 (6.241)	0.055 (5.529)	0.075 (6.34)	0.029 (2.401)	-0.034 (-2.627)	-0.243 (-10.586)
SMB (S&P500)	-0.372 (-2.525)	0.232 (2.759)	0.211 (3.238)	0.056 (1.062)	-0.005 (-0.097)	-0.063 (-1.55)	-0.118 (-2.771)	-0.036 (-0.708)	-0.089 (-1.716)	-0.053 (-0.953)	-0.14 (-1.429)
HML (S&P500)	-0.361 (-4.317)	0.141 (2.945)	0.068 (1.828)	0.065 (2.178)	0.072 (2.35)	0.015 (0.655)	-0.014 (-0.578)	0.024 (0.83)	-0.072 (-2.448)	-0.08 (-2.54)	-0.22 (-3.964)
RESMOM (S&P500)	0.584 (15.736)	-0.303 (-14.32)	-0.165 (-10.051)	-0.111 (-8.366)	-0.045 (-3.306)	-0.009 (-0.899)	0.013 (1.181)	0.067 (5.334)	0.088 (6.723)	0.187 (13.454)	0.28 (11.369)
REV (S&P500)	0.435 (11.17)	-0.167 (-7.49)	-0.155 (-9.009)	-0.057 (-4.079)	-0.051 (-3.549)	-0.044 (-4.044)	0.002 (0.2)	0.02 (1.538)	0.082 (5.966)	0.102 (6.959)	0.268 (10.368)

Table D.9: Annualized alpha statistics for the $rtnXmkt - Linear$ configuration. All deciles portfolios are excess market returns, except for the D10-D1 portfolio. All values are fitted coefficients with t-stat in brackets in the OOS period 1995/01-2022/12. Expanding window applied with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years. Note that Newey-West 12 lags standard errors are used for SE computation.

D.4 Empirical distribution maps

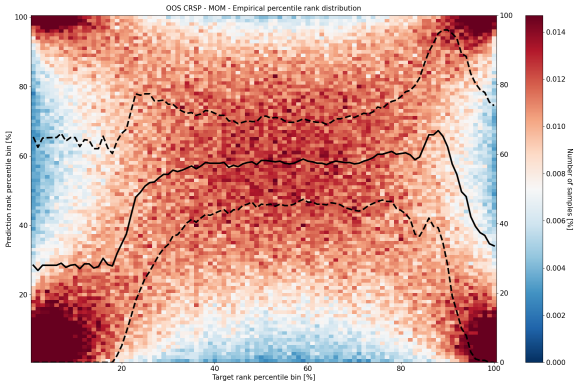


Figure D.5: Empirical distribution grid based on target and prediction deciles for the CRSP universe in the OOS period 1995/01-2022/12 for the MOM benchmark.

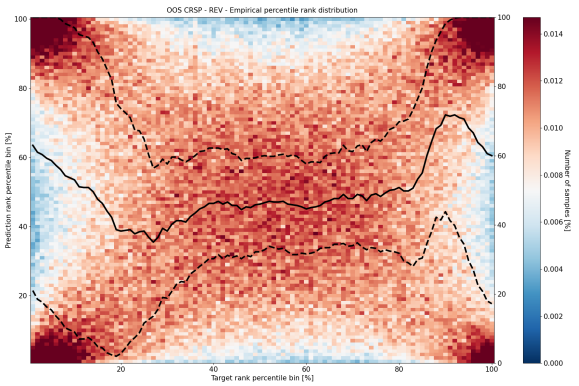


Figure D.6: Empirical distribution grid based on target and prediction deciles for the CRSP universe in the OOS period 1995/01-2022/12 for the REV benchmark.

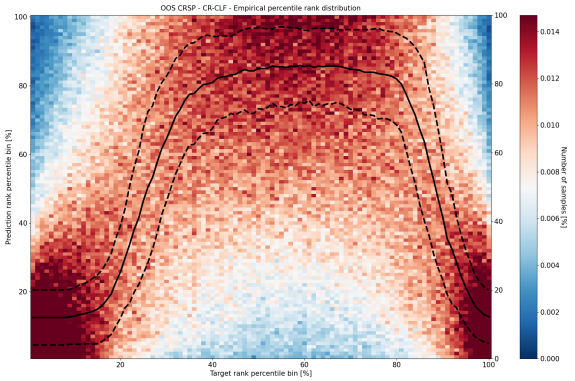


Figure D.7: Empirical distribution for the $CR - CLF$ system configuration over the period 1995-2022 in the CRSP universe. Expanding window applied for the complex system configuration with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years.

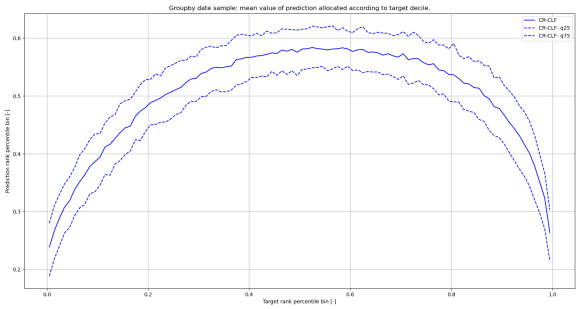


Figure D.8: Aggregated results of the empirical distribution over time for the $CR - CLF$ system configuration over the period 1995-2022 in the CRSP universe. Expanding window applied for the complex system configuration with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years.

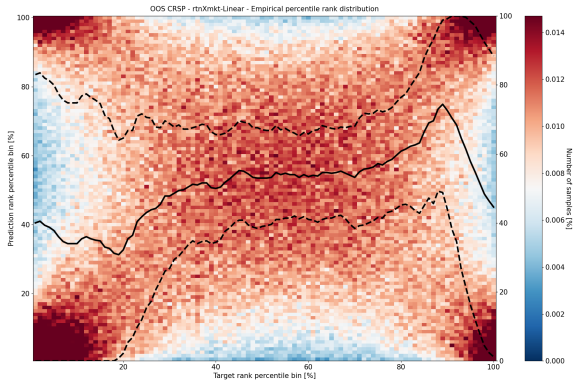


Figure D.9: Empirical distribution for the $rtnXmkt - Linear$ system configuration over the period 1995-2022 in the CRSP universe. Expanding window applied for the linear system configuration with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years.

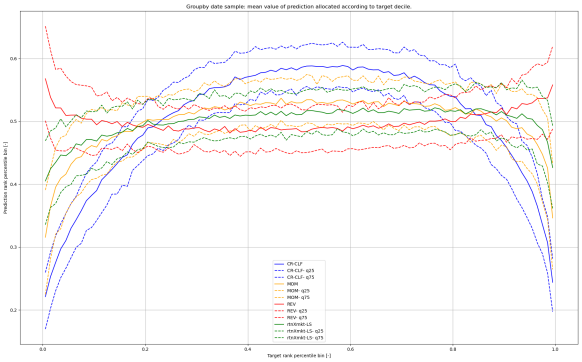


Figure D.10: Aggregated results of the empirical distribution over time for the $rtnXmkt - Linear$ system configuration as well as the MOM and REV benchmarks over the period 1995-2022 in the CRSP universe. Expanding window applied for the system configuration with training period starting at 15 years in 1975-1990, validation period of constant 5 years 1990-1995, refit every three years

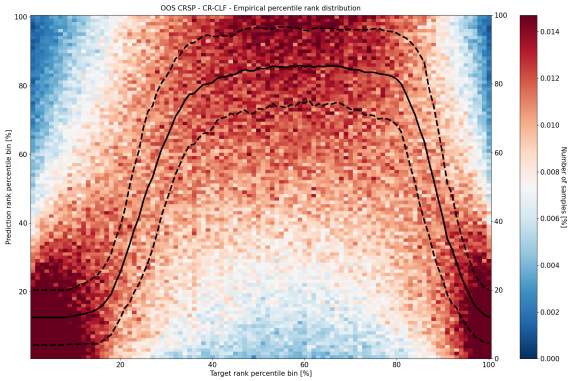


Figure D.11: Empirical distribution grid based on target and prediction deciles for the CRSP universe in the OOS period 1995/01-2022/12 for the $CR - CLF$ configuration.

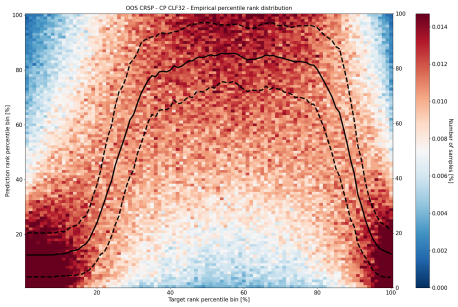


Figure D.12: Empirical distribution grid based on target and prediction deciles for the CRSP universe in the OOS period 1995/01-2022/12 for the $CP - CLF$ configuration.

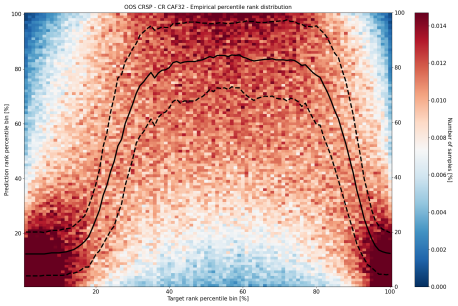


Figure D.13: Empirical distribution grid based on target and prediction deciles for the CRSP universe in the OOS period 1995/01-2022/12 for the $CR - CAF$ configuration.

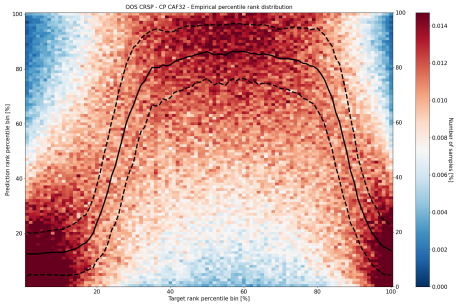


Figure D.14: Empirical distribution grid based on target and prediction deciles for the CRSP universe in the OOS period 1995/01-2022/12 for the $CP - CAF$ benchmark.

D.5 Spearman contribution plots

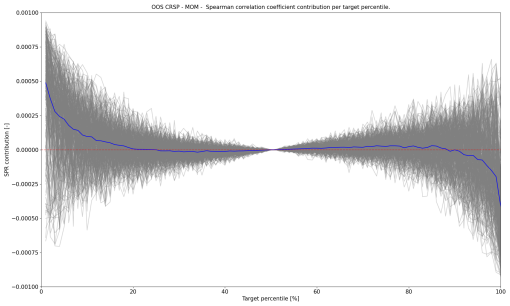


Figure D.15: Spearman correlation coefficient contribution aggregated per target percentile for the MOM benchmark over the out-of-sample period 1995-2022 in the CRSP universe. Mean contribution over time dimension per target percentile given in blue.

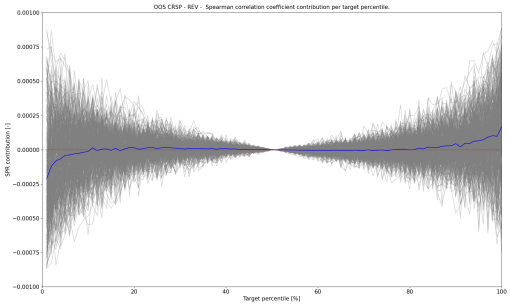


Figure D.16: Mean spearman correlation coefficient contribution aggregated per target percentile the REV benchmark over the out-of-sample period 1995-2022 in the CRSP universe. Mean contribution over time dimension per target percentile given in blue.

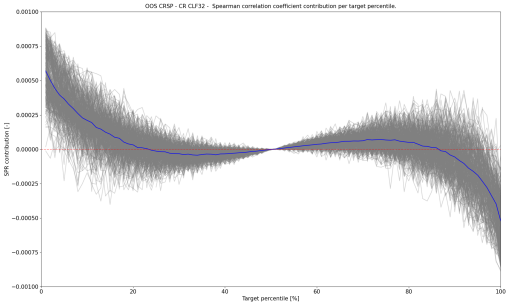


Figure D.17: Spearman correlation coefficient contribution aggregated per target percentile for the $CR - CLF$ complex system configuration over the out-of-sample period 1995-2022 in the CRSP universe. Mean contribution over time dimension per target percentile given in blue.

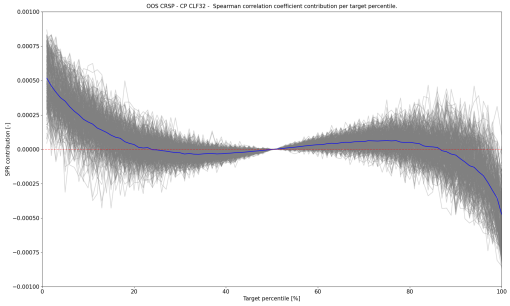


Figure D.18: Spearman correlation coefficient contribution aggregated per target percentile for the $CP - CLF$ system configuration over the out-of-sample period 1995-2022 in the CRSP universe. Mean contribution over time dimension per target percentile given in blue.

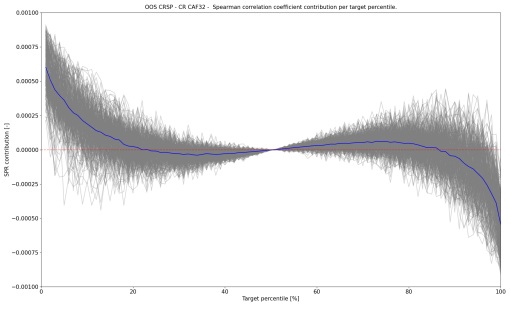


Figure D.19: Spearman correlation coefficient contribution aggregated per target percentile for the $CR - CAF$ system configuration over the out-of-sample period 1995-2022 in the CRSP universe. Mean contribution over time dimension per target percentile given in blue.

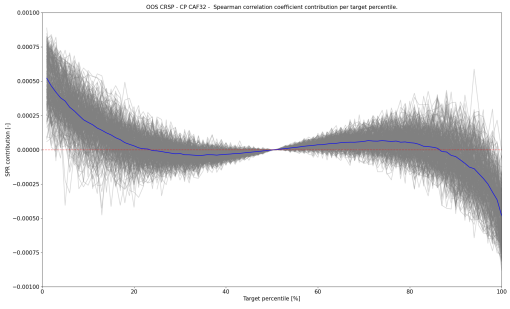


Figure D.20: Spearman correlation coefficient contribution aggregated per target percentile for the $CP - CAF$ system configuration over the out-of-sample period 1995-2022 in the CRSP universe. Mean contribution over time dimension per target percentile given in blue.

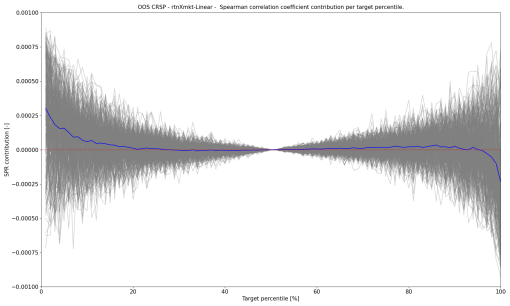


Figure D.21: Spearman correlation coefficient contribution aggregated per target percentile for the $rtnXmkt - Linear$ system configuration over the out-of-sample period 1995-2022 in the CRSP universe. Mean contribution over time dimension per target percentile given in blue.

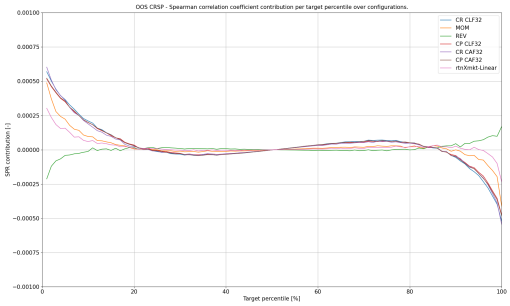


Figure D.22: Mean Spearman correlation coefficient contribution aggregated per target percentile for the complex system configurations ($CR-CLF$, $CP-CLF$, $CR-CAF$ & $CP-CAF$) as well as $rtnXmkt-Linear$ and the MOM and REV benchmarks over the out-of-sample period 1995-2022 in the CRSP universe. Mean contribution over time dimension per target percentile given in blue.

E | Python script description

This results in this research were obtained by the python scripts contained in the zipfile *ExtremeLearning MScThesisQF RWVOS 583868*. It consists of one main jupyter notebook and multiple python scripts. In this appendix we describe these script. The files are given in the following list,

NAME.FILETYPE: DESCRIPTION

- *requirements.txt*: Lists all environment package requirements.
- *Extreme learning.ipynb*: Main python notebook that allows for data pre- and post-processing and system implementation and estimation. Note that this file uses pytorch for its system implementation and estimation. This notebook calls functions provided in the .py script files.
- *utility EL.py*: This script contains all 'utility' definitions including all definitions required for the data pipelines (i.e. dataprocessor and backtester). Note that model and optimization related functions are not included in this script.
- *models EL.py*: This script contains all functions/classes which are used to setup models and fit them and properly store any results. Note that the models/functionality in this script is built to work with pytorch. Note that this file uses pytorch for its system implementation and estimation.
- *runner EL.py*: This script forms the runner file. It mainly consists of functions which calls all functions and sets up all class instances required to properly train a single estimation window out of the multiple estimation windows involved in the experiment.

Bibliography

Stanislav Anatolyev and Alexander Gerko. A trading approach to testing for predictability. *Journal of Business & Economic Statistics*, 23(4):455–461, 2005.

Andrew Ang, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang. The cross-section of volatility and expected returns. *The journal of finance*, 61(1):259–299, 2006.

Doron Avramov, Guy Kaplanski, and Avanidhar Subrahmanyam. Moving average distance as a predictor of equity returns. *Review of Financial Economics*, 39(2):127–145, 2021.

Christopher M Bishop et al. *Neural networks for pattern recognition*. Oxford university press, 1995.

David Blitz and Pim Van Vliet. The volatility effect: Lower risk without lower return. *Journal of Portfolio Management*, pages 102–113, 2007.

David Blitz, Joop Huij, and Martin Martens. Residual momentum. *Journal of Empirical Finance*, 18(3):506–521, 2011.

Mathieu Blondel, Olivier Teboul, Quentin Berthet, and Josip Djolonga. Fast differentiable sorting and ranking. In *International Conference on Machine Learning*, pages 950–959. PMLR, 2020.

Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

Mark M Carhart. On persistence in mutual fund performance. *The Journal of finance*, 52(1): 57–82, 1997.

Tsung-Yu Chen, Pin-Huang Chou, Kuan-Cheng Ko, and S Ghon Rhee. Non-parametric momentum based on ranks and signs. *Journal of Empirical Finance*, 60:94–109, 2021.

Zhanhui Chen and Ralitsa Petkova. Does idiosyncratic volatility proxy for risk exposure? *The Review of Financial Studies*, 25(9):2745–2787, 2012.

Kent Daniel and Tobias J Moskowitz. Momentum crashes. *Journal of Financial economics*, 122 (2):221–247, 2016.

Kent D Daniel and Tobias J Moskowitz. Momentum crashes. *Swiss Finance Institute Research Paper*, (13-61):14–6, 2013.

Francis X Diebold and Robert S Mariano. Comparing predictive accuracy. *Journal of Business & economic statistics*, 20(1):134–144, 2002.

Patrik Eggebrecht and Eva Lütkebohmert. A deep trend-following trading strategy for equity markets. *The Journal of Financial Data Science*, 2023.

Lerby Ergun, Alexander Molchanov, and Philip A Stork. Technical trading rules, loss avoidance, and the business cycle. *Loss Avoidance, and the Business Cycle*.

Eugene F Fama and Kenneth R French. The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465, 1992.

Eugene F Fama and Kenneth R French. Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56, 1993.

Eugene F Fama and Kenneth R French. Dissecting anomalies. *The journal of finance*, 63(4): 1653–1678, 2008.

Eugene F Fama and Kenneth R French. A five-factor asset pricing model. *Journal of financial economics*, 116(1):1–22, 2015.

Eugene F Fama and Kenneth R French. Dissecting anomalies with a five-factor model. *The Review of Financial Studies*, 29(1):69–103, 2016.

Eugene F Fama and Kenneth R French. Choosing factors. *Journal of financial economics*, 128 (2):234–252, 2018.

Eugene F Fama and James D MacBeth. Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3):607–636, 1973.

Paul Geladi and Bruce R Kowalski. Partial least-squares regression: a tutorial. *Analytica chimica acta*, 185:1–17, 1986.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

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

Chulwoo Han. Bimodal characteristic returns and predictability enhancement via machine learning. *Management Science*, 68(10):7701–7741, 2022.

Yufeng Han, Guofu Zhou, and Yingzi Zhu. A trend factor: Any economic gains from using information over investment horizons? *Journal of Financial Economics*, 122(2):352–375, 2016.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8): 1735–1780, 1997.

Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

Kewei Hou, Chen Xue, and Lu Zhang. Replicating anomalies. *The Review of financial studies*, 33(5):2019–2133, 2020.

Narasimhan Jegadeesh. Evidence of predictable behavior of security returns. *The Journal of finance*, 45(3):881–898, 1990.

Narasimhan Jegadeesh and Sheridan Titman. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, 48(1):65–91, 1993.

Jingwen Jiang, Bryan T Kelly, and Dacheng Xiu. (re-) imag (in) ing price trends. *Chicago Booth Research Paper*, (21-01), 2020.

Bryan T Kelly, Semyon Malamud, Kangying Zhou, et al. The virtue of complexity in machine learning portfolios. Technical report, Swiss Finance Institute, 2021.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Damian Kisiel and Denise Gorse. Portfolio transformer for attention-based asset allocation. In *Artificial Intelligence and Soft Computing: 21st International Conference, ICAISC 2022, Zakopane, Poland, June 19–23, 2022, Proceedings, Part I*, pages 61–71. Springer, 2023.

Bryan Lim, Sercan Ö Arık, Nicolas Loeff, and Tomas Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37 (4):1748–1764, 2021.

Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

Scott Murray, Houping Xiao, and Yusen Xia. Charting by machines. *Available at SSRN 3853436*, 2021.

Daniel Poh, Bryan Lim, Stefan Zohren, and Stephen Roberts. Enhancing cross-sectional currency strategies by context-aware learning to rank with self-attention. *arXiv preprint arXiv:2105.10019*, 2021.

David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Kieran Wood, Sven Giegerich, Stephen Roberts, and Stefan Zohren. Trading with the momentum transformer: An intelligent and interpretable architecture. *arXiv preprint arXiv:2112.08534*, 2021.

Junhao Zhou, Zhanhong He, Ya Nan Song, Hao Wang, Xiaoping Yang, Wenjuan Lian, and Hong-Ning Dai. Precious metal price prediction based on deep regularization self-attention regression. *IEEE Access*, 8:2178–2187, 2019.