Revisiting Fama-French Factors' Predictability with Bayesian Modelling and Copula-based Portfolio Optimization

Yang Zhao [a], Charalampos Stasinakis [b*], Georgios Sermpinis[b], Filipa Da Silva Fernandes[c]

October 2018

Abstract

This study is investigating the predictability of the five Fama-French factors and explores their optimal portfolio allocation for factor investing during 2000-2017. Firstly, we forecast each factor with a pool of linear and non-linear models. Next, the individual forecasts are combined through Dynamic Model Averaging (DMA), while their performance is benchmarked by the best performing individual predictor and other forecast combination techniques. Finally, we use the Generalized Autoregressive Score (GAS) model and the skewed *t* copula method to estimate the correlation of assets. The GAS performance is also compared with other traditional approaches such as Dynamic Conditional Correlation (DCC) model and Asymmetric Dynamic Conditional Correlation (ADCC). The performance of the constructed portfolios is assessed through traditional metrics and ratios accounting for the Conditional Value-at-Risk (CVaR) and the Conditional Diversification Benefits (CDB) approach. Our results show that combining Bayesian forecast combinations with copulas is leading to significant improvements in the portfolio optimization process, while forecasting covariance accounting for asymmetric dependence between the factors adds diversification benefits to the obtained portfolios.

Keywords: Factor Investing, Portfolio Optimization, Dynamic Model Averaging, Forecast Combinations

[a] Chinese Academy of Finance and Development, Central University of Finance and Economics, 39 South College Road, Haidian, Beijing, 100081, China
[b] Adam Smith Business School, University of Glasgow, Adam Smith Building, Glasgow, G12 8QQ, UK
[c] Aberdeen Business School, University of Aberdeen, Edward Wright Building, Aberdeen, AB24 3QY, UK

*Corresponding author. Email: Charalampos.Stasinakis@glasgow.ac.uk

1. Introduction

It is a well-known fact that reduced-form factor models are useful in asset pricing, as they provide a parsimonious summary of the cross section of asset returns (Fama and French, 1993 and 1995). The economic premise of factor-modelling is based on the fact that covariances have explanatory power over the cross-sectional expected returns and that factors are able to capture to a large extent the time-series co-movements of stock returns. Therefore, it is expected that an investor who wants to benefit from this must accept exposure to factor risk (Kozak et al., 2018). As a consequence, factor models are accepted as an econometric tool for analysing portfolio risk exposures. Decomposing risk exposure into factors not only allows for an independent vetting of managers offering investment opportunities, but also quantifies the risk exposure overlap with other funds during periods of high volatility or liquidity draughts (Luo and Mesomeris, 2015).

Factor investing has gained increased popularity over the past decades among academics and market participants (Cerniglia and Fabozzi, 2018). This is based on the fact that investors believe that portfolio returns' expectations should be evidence based and that factor-based portfolios are considered a solid example of long-term investment (Dimson et al., 2017; Briere and Szafarz, 2018). A large number of studies have identified that some style factors have historically earned attractive risk-return profiles over time (Fama and French, 1993, 2015; Carhart, 1997; Ferson et al., 2006; and Ang, 2014). There are two main types of factors that drive returns: macroeconomic factors, which capture broad risks across asset classes; and style factors, which explain returns and risk within asset classes. If an investor holds (optimized) diversified portfolios, better risk-return trade-offs can be attained in comparison to holding individual assets. This is the foundation of the traditional Mean-Variance (M-V) approach of Markowitz (1952). Assuming that this investor can invest directly in a security that replicates the return on individual factors, then it is possible to obtain diversification benefits from investing in a portfolio of stock factors. Thus, if factor investing can be implemented cheaply, it significantly raises the bar for active management.

Although practitioners have to face structural or regulatory barriers to short-selling when they construct long-short portfolios, factors still can be tradeable via different ways. Some factor premiums can be captured through long-short combinations of existing index-based instruments (Briere and Szafarz, 2018). For instance, MSCI factor indexes provide flexible access to factor investing, such as Value, Low Size, Low Volatility, High Yield, Quality and Momentum. Studies such as the works of Ferson et al. (2006), Bender et al., (2010) and Bender et al. (2013) explain how portfolios are constructed in an effort to obtain factor risk premium. This is very important, as such risk premiums are required to compensate for their underlying risk and allow risk hedging through application of different types of factors in the same portfolio. Several techniques are developed to improve upon the passive capitalization of weighted equity market portfolios through intelligent integration of factor returns.

Common factors of interest are the market, size and value factors introduced by Fama and French (1993), the momentum factor introduced by Jegadeesh and Titman (1993) and Carhart (1997), the liquidity factor identified by Pástor and Stambaugh (2003), as well as the profitability factor and investment patterns' factor found in Fama and French (2015). In general, the seminal studies of Fama and French (2015, 2016, 2018) confirm that the five-factor model is capturing adequately the returns' movements. This literature brings forward the fact that many institutions are increasingly interested in factors' congruence and how their optimal allocation can improve the risk-adjusted performance of their equity portfolios. This interest, though, goes beyond the traditional approach of Markowitz (1952).

This motivates us to explore optimal allocation methods for factor investing. Several studies postulate that portfolio optimization can yield substantial diversification benefits in terms of risk-return trade-off mainly depending on the forecasting accuracy of conditional moments of asset returns (Chan et al., 1999). Consequently, more accurate estimates can generate more successful and active investment strategies. Knowing that the expected returns and correlation (covariance matrix) of assets are the primary inputs for portfolio optimization, the aim of this study is twofold. The first target is to select superior factor return predictions. The second goal is to exploit the time-varying correlations of factor returns and their asymmetric dependence in order to maximize the diversification benefits derived from factor-based portfolios.

Miralles-Quiros and Miralles-Quiros (2017) suggest that portfolio optimization literature tends to neglect the importance of return predictability. The voluminous financial forecasting literature should be ideal for practitioners aiming at the first target mentioned above. Through that they are able to select and/or combine linear and non-linear models that apply constant or time-varying parameterization processes. Bayesian models constitute a prominent class of such techniques able to encompass the forecasting power of large number of individual predictors given powerful computational resources. Wright (2008, 2009) apply Bayesian m in forecasting exchange rates and US inflation. Feldkircher et al. (2014) utilize also the same technique in the FX markets too. The Dynamic Model Averaging (DMA) is used by Koop and Korobilis (2012) to forecasting inflation based on a set of predictors, as a recursive extension of the Bayesian approach. Another class of available forecasting tools is the Support Vector Regressions (SVRs). They are regression-based models able to explore the non-linear and data-adaptive dynamics of financial time series given a set of inputs (Vapnik, 1995). Their applications in finance are numerous (see amongst others Lu et al. (2009), Wang and Zhu (2010) and Yao et al. (2015). They exhibit, though, high sensitivity to the calibration of their parameters. For that reason, many studies in the area of heuristic and metaheuristic optimization are invested into this task, as especially the latter are able to avoid local optima trapping, over-fitting and computational costs (Parejo et al, 2012). Nature-inspired metaheuristic approaches in particular that are motivated by the evolution of species or their swarm movement behaviour have received research traction (Yang, 2010; Yang and Gandomi, 2012; Gandomi and Alavi, 2012). Mirjalili (2016) proposed the Sine Cosine (SC) algorithm that is based on mathematical objective functions rather than bio-inspired ones. This was recently adopted couple of studies, Li et al. (2018) and Fernandes et al. (2018) in a hybrid SC-SVR model. Their results indicate that SC is providing better SVR optimization compared to other robust bio-inspired algorithms.

On the other hand, portfolio researchers that focus on the univariate distributions of the individual assets and the dependence between each asset, should be able to investigate and capture time-varying correlations (Cha and Jithendranathan, 2009). The copula literature is the one that should appear attractive but also challenging. Copula modelling plays a crucial role in the portfolio optimization research of past decades (Patton, 2006; Christoffersen et al., 2012; Boubaker and Sghaier, 2013; Kakouris and Rustem, 2014; Sahamkhadam et al., 2018). Some of the Archimedean copulas, like the Clayton, the Gumbel, and the Joe-Clayton, can capture asymmetric dependence in bivariate cases, however, generalizing them to the high-dimensional case is computationally difficult (Savu and Trede, 2010). Recent studies on applications of copula-based models in finance show that the skewed *t* copula is able to incorporate the multivariate asymmetries in high-dimensional dependence modelling (Christoffersen et al., 2012, 2013; Lucas et al., 2014; Cerrato et al., 2017)[1]. In particular, Christoffersen et al. (2013) employ the skewed *t* copula to model the nonlinear dependence across four equity factors. The superiority of using skewed *t* copula is based on the fact that it not only takes into account the tail dependence, but also the multivariate asymmetry across factors. Three models are usually applied in forecasting the covariance matrix among financial assets, namely the Dynamic Conditional Correlation (DCC) model (Engle, 2002), the Asymmetric Dynamic Conditional Correlation (ADCC) model (Cappiello et al., 2006) and the Generalized Autoregressive Score (GAS) model (Creal et al. 2013). Although DCC is probably the most widely used approach, ADDC is a useful extension allowing for conditional asymmetries in correlations across assets (Fei et al., 2010). Nonetheless, GAS is able to capture the dynamic dependence of asset returns by applying the score of the conditional density function to drive the dynamics of the time-varying parameters (Lucas et al. 2014; Salvatierra and Patton 2015; Zhao et al., 2018).

Based on the above, this study is designed as follows. We investigate the return predictability of Fama-French's (2015) five factors and explore their optimal portfolio allocation for factor investing over the period of 2000-2017. This is achieved by proposing a novel three-stage optimization framework. Firstly, we perform a one-step-ahead forecast for the first moment of each factor with a pool of linear and nonlinear models. The individual forecasts are then combined through Bayesian DMA, while their performance is benchmarked by the Random Walk (RW), the best individual predictor, SVR and SC-SVR. This step will provide the most accurate factor returns, which goes towards the first aim mentioned above. Secondly, we capture the characteristics of correlation among factors and we construct a Dynamic Asymmetric Copula (DAC) model which combines the Generalized Autoregressive Score (GAS) model

[1] The skewed *t* copula is implied by the Generalized Hyperbolic (GH) distribution discussed in Demarta and McNeil (2005).

and a skewed *t* copula. The proposed model is employed to forecast the second moment of factors in portfolio optimization. Traditional multivariate models such as DCC-GARCH and ADCC-GARCH are also considered for the purpose of comparison. Finally, we obtain optimal asset allocations of the factors using the M-V approach and the mean-Conditional Value-at-Risk[2] (CVaR) optimizations. The latter approach allows us to select portfolios from an efficient frontier characterized by risk-expected shortfall trade-offs (Alexander et al., 2006; Karmakar and Paul, 2018). In order to evaluate the performance of the obtained tangency portfolios several performance metrics are applied, such as annualized returns, Sharpe and Sortino ratios, Maximum Drawdown (MDD), return over CVaR ratio and the Conditional Diversification Benefits (CDB) of Christoffersen et al. (2012). To best of our knowledge, this empirical setup is novel.

The results show that combining Bayesian forecasts with the proposed skewed *t* copula-based GAS model leads to the best factor allocations. In particular, this study brings forward several interesting findings: (a) The portfolios based on the forecasts derived from SC-SVR and DMA offer significant improvement over the ones from RW and 1/N strategy in terms of portfolio risk reduction; (b) The optimization using the GAS model yields evidently better performance than the ones using DCC and ADCC model; (c) The portfolios that allow short-selling could offer higher diversification benefits than long-only ones. These findings have some important implications. Firstly, it is confirmed that reducing the estimation errors of the first moments of asset returns can significant improve the portfolio optimization performance. Secondly, incorporating the asymmetric dependence among asset returns in the optimization process leads to substantial increase in the diversification benefits for the investor. Finally, CVaR is a more appropriate risk measure if the investor’s utility function is characterized by minimization of downside tail risk.

The rest of the paper is organized as follows. Section 2 provides the description of the Fama-French factors’ dataset, while all forecasting models are described in section 3. Their statistical evaluation is given in section 4. The portfolio design of this study is explained in detail in section 5. The final portfolio optimization results are summarized in section 6, while some concluding remarks are given in Section 7. Finally, technical and mathematical details essential for the understanding of this study are included in the appendix.

2. Dataset

The forecasting models in this study are applied in the task of forecasting the one day ahead logarithmic returns of the five factors, namely the Market Factor (MKT), Size Factor (SMB), Value

[2] CVaR is the abbreviation of the Conditional Value-at-Risk, which is also known as the Expected Shortfall.

Factor (HML), Profitability Factor (RMW) and Investment Patterns' Factor (CMA)[3]. The descriptive statistics and correlation matrices of the return series are shown in the following table.

Insert Table 1

The MKT and RMW returns series exhibit slight negative skewness, while SMB, HML and CMA a positive one. All return series exhibit high positive kurtosis. The Jarque-Bera and Augmented Dickey-Fuller (ADF) statistic confirms that the factor return series under study are non-normal and without a unit root at the 99% confidence level respectively. Additionally, from the table we note that high negative and positive correlations are evident across the factors. The dataset design of this study is shown in table 2 below.

Insert Table 2

Figure 1 presents the factors' performance during the out-of-sample period.

Insert Figure 1

Overall, all factors present an upward trend in cumulative returns during 2000-2017. MKT seems to be performing worse compared to the others. Especially before 2003, we observe decreasing cumulative returns for MKT, while all other four factors have opposite performance. During the global financial crisis period HML's and MKT's performance are more negatively affected. After 2009, the returns' upward trend of MKT factor is vastly reinforced.

3. Forecasting Models

This section summarizes the models applied in the design of this forecasting application. Initially, the individual forecasts from a pool of traditional predictors are obtained. Then, the best predictors are selected and fed into three forecast combination techniques, namely the traditional SVR, the SC-SVR and the Bayesian DMA.

3.1 Individual Forecasts

Applying a large pool of traditional predictors should be the first step of every forecasting exercise. In our case, we employ more than three hundred linear and non-linear individual predictors to predict the five factors in-sample. The linear models belong in the classes of Simple Moving Averages (SMA), Exponential Moving Averages (EMA), Autoregressive terms (AR) and Autoregressive Moving Average (ARMA) models. There are also several non-linear models applied, such a Smooth Transition Autoregressive Model (STAR), Nearest Neighbors Algorithm (*k*-NN), a Multi-Layer Perceptron (MLP), a Recurrent Neural Network (RNN), a Higher Order Neural Network (HONN), a Psi-Sigma Neural

---

[3]All the data are publicly available on French's website at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Network (PSN), PSO Neural Network (ARBF-PSO), Genetic Programming (GP) and Gene Expression Programming (GEP).[4] The final pool size is three hundred and twenty-eight individual predictors for each forecasting exercise. After the individual forecasts are obtained it is important to screen out the best predictors. Thus, the Principal Component Analysis (PCA) is used in order to discard highly correlated variables, while accounting for the 95% of the total variance. The final principal components are presented in the following table.

Insert Table 3*

It is obvious that the PCA analysis vastly decreases the input dimensions (number of input ranges from six to nine), while non-linear models appear to be the best performing[5]. The above process is crucial, as it allow us to select robust prediction benchmarks and cope with the dimensionality issue that arises when applying forecast encompassing techniques. Only the above principal components are used as inputs for all the remaining techniques, while the RW and the best predictor of each case (bold) play the role of our benchmarks.

3.2 Forecast Combination Methods

In this study we employ three techniques to combine the best individual forecasts, namely SVR, SC-SVR and DMA. Their short descriptions are presented in this subsection.

3.2.1 Support Vector Regression (SVR)

SVR is a technique based on the principle of structural risk minimization, as proposed by Vapnik (1995). It is able to achieve good generalization in non-linear regression tasks by using only a subset of the training observations, known as the support vectors. If $\{(x_1,y_1), (x_2,y_2)\ldots, (x_n, y_n)\}$, where $x_i \in X \subseteq R, y_i \in Y \subseteq R, i=1 \ldots n$ are the training data and n the total number of training samples, the general SVR function can be specified as:

$$f(x) = w^T \varphi(x) + b \qquad (1)$$

$\varphi(x)$ is the non-linear function that maps the input data vector x into a feature space where the training data exhibit linearity. In order to obtain w and b, the following regularized risk function must be minimized:

[4] All predicting models within the pool are well known in the forecasting literature. Short specifications for the linear ones are provided in appendix A. In this paper, we do not delve into the description of the non-linear ones as this is out of the scope of this paper. Their modelling design is available upon request, while we refer the interested reader to Sermpinis et al. (2017) for more details.
[5] We have also applied a Relevance Vector Machine (RVM) model for input variable selection (Tipping, 2001). RVM is a Bayesian sparse kernel technique able to cope with large-scale data-processing and reduce the feature space to the most important vectors. This approach leads to very similar inputs' sets as with the PCA ones, while the ranking of the models and the significance of the results remain consistent to the discussion of the main text. These results are not presented here for the shake of space, but they are available upon re request.

$$\left\{ R(C) = C\frac{1}{n}\sum_{i=1}^{n} L_{\varepsilon}(y_i, f(x_i)) + \frac{1}{2}\|w\|^2 \text{ , } L_{\varepsilon}(y_i, f(x_i)) = \begin{cases} 0 \;\; if \; |y_i - f(x_i)| \le \varepsilon \\ |y_i - f(x_i)| - \varepsilon \;\; if \;\; other \end{cases}, \varepsilon \ge 0 \right\} \tag{2}$$

Parameters C and ε are calibrated by the practitioner, y_i is the actual value at time i, $f(x_i)$ is the predicted value at the same period and L_ε is ε-sensitive loss function The loss function identifies the predicted values that have at most ε deviations from the actual values y_i. The ε parameter defines the known 'tube' in the SVR literature (Vapnik, 1995; Schölkopf et al., 1999). Assuming the parameter $v \in (0,1)$, then the SVR problem transforms to a vSVR optimization problem that follows:

$$\text{Minimize } C\left(v\varepsilon + \frac{1}{n}\sum_{i=1}^{n}(\xi_i + \xi_i^*) \right) + \frac{1}{2}\|w\|^2 \text{ subject to } \begin{Bmatrix} \xi_i \ge 0 \\ \xi_i^* \ge 0 \\ C > 0 \end{Bmatrix} \text{and } \begin{Bmatrix} y_i - w^T\varphi(x_i) - b \le +\varepsilon + \xi_i \\ w^T\varphi(x_i) + b - y_i \le +\varepsilon + \xi_i^* \end{Bmatrix} \tag{3}$$

Equation (5) becomes a dual problem and its solution is based on the two Lagrange multipliers a_i, a_i^* and the kernel function $K(x_i, x)$:

$$f(x) = \sum_{i=1}^{n}(a_i - a_i^*)K(x_i, x) + b \text{, where } 0 \le a_i, a_i^* \le \frac{C}{n} \tag{4}$$

The transformation of input space is achieved with the Gaussian Radial Basis Function (RBF) for all the SVR models applied. Parameter C satisfies the need to trade model complexity for a training error and vice versa Additionally, the intuition of vSVR is that the parameter v is an approximation of the upper and lower bounds of the fraction of errors. Here it should be noted that the majority of the SVR studies suggest that v parameter should be used as it reduces computational time and increases forecasting accuracy (Schölkopf et al.,1999). In order to calibrate the parameters of the vSVR, the grid search technique is applied.[6]

3.2.2 Sine-Cosine Support Vector Regression (SC-SVR)

One recent very successful approach for SVR parameterization is the SC algorithm as proposed by Mirjalili (2016). The SC algorithm is a population-based optimization technique that is able to search different areas of the given search space by combining an exploration and exploitation phase. The modelling procedure starts with a set of random solutions and proceeds to the global optima. Once the algorithm is in the exploration phase, the probability of getting trapped in the local optima is minimized. Conversely, the higher the number of random solutions, the higher the probability of obtaining the global optima in the exploitation phase. Hence, when applying the SC the best global solution is found by updating the positions of the random candidate solutions towards the best solution. This is achieved through the use of sine and cosine functions as objective functions. The local search of different regions

[6] For more details on the mathematical solutions and SVR modelling, the interested reader should refer to Vapnik (1995) and Cherkassky and Ma (2004).

in the search space is achieved by allowing the sine and cosine functions to return to values greater than one or less than minus one. Following the work of Mirjalili (2016), the position updating equations are the following:

$$P_j'^{t+1} = \begin{cases} P_j'^{t} + r_1' \sin(r_2') \left| r_3' P_j''^{t} - P_j'^{t} \right|, r_4' < 0.5 \\ P_j'^{t} + r_1' \cos(r_2') \left| r_3' P_j''^{t} - P_j'^{t} \right|, r_4' \geq 0.5 \end{cases} \quad (5)$$

where $P_j'^{t}$ is the position of the current solution for the j^{th} dimension at t^{th} and $P_j''^{t}$ is the position of the destination point, r_1', r_2', r_3', r_4' are random variables.

The $r_1' = c'' - t(c''/T'')$ is a balancing random metric, where c'' is a constant, t is the current iteration and T'' is the maximum number of iterations. Calibrating r_1' balances the exploring and the exploitation and leads to an adaptive shift in the range of sine and cosine calculations. Consequently, r' dictates the next positions' region. This region would be either in the space between the current solution and the next destination or outside it. The random variable r_2' is bounded between [0,2π] and indicates whether the random location will be within or outside the cyclical pattern invoked by the nature of the sine and cosine functions. The third random variable r_3' is a random weight defining the emphasis of the destination position in defining the distance. Finally, r_4' is bounded as [0,1] and provides an equal switch between the sine and cosine functions. For more details on the mathematical implementation of SC algorithm, we refer the reader to Mirjalili (2016).

SC presents several advantages over other similar techniques used for SVR parameterization. SC generates improved sets of random solutions and benefits from high exploration and local optima avoidance, compared with individual-based algorithms, such as GAs. The algorithm is also able to divide the search space into different areas promising for exploration based on the sine and cosine. The adaptive range imposed on the two functions allows for a smooth transition between exploration and exploitation, unlike in the case of another popular algorithm, the Krill Herd. (Fernandes et al., 2018). Li et al. (2018) show how SC algorithm can be used for optimal SVR parameterization. Our approach does not differ from these guidelines. Finally, the optimal selection of the SVR parameters is achieved by minimizing the Root Mean Squared Error (RMSE) in the test-sub period. Therefore, the following fitness function needs to be maximized:

$$Fitness = 1/(1 + RMSE) \quad (5)$$

The RBF kernel is also used in SC-SVR as in traditional SVR.

3.2.3 Dynamic Model Averaging (DMA)

DMA proposed by Raftery et al. (2010) is a recursive implementation of standard Bayesian Model Averaging that allows selecting different subsets of explanatory variables over time along with variable coefficients. If we consider a candidate input set $u = 1, \dots, U$, then the state-space model at time $t = 1, \dots, T$ for the dependent variable y_t^* can be presented under observational and state equations as:

$$y_t^* = F_t^{(u)'} \zeta_t^{(u)} + \varepsilon_t^{(u)} \tag{6}$$

$$\zeta_t^{(u)} = \zeta_{t-1}^{(u)} + \eta_t^{(u)} \tag{7}$$

$$\begin{pmatrix} \varepsilon_t^{(u)} \\ \eta_t^{(u)} \end{pmatrix} \sim \mathcal{N} \begin{pmatrix} R_t^{(u)} & 0 \\ 0 & V_t^{(u)} \end{pmatrix} \tag{8}$$

where $F_t^{(u)}$ is a subset from the ν potential predictors at each time. The $\zeta_t^{(u)}$ is a $p^* \times 1$, $p^* \leq v^*$ vector of time-varying regression coefficients evolving over time. The dynamic nature of the process is that DMA allows different models to hold at each point in time.

The DMA averages the forecasts across candidate combination of models based on predictive likelihood through a recursive updating scheme. The predictive likelihood estimates the ability of model u to predict y_t^*. The models containing better predictors receive higher predictive likelihood and are associated with higher weights in the averaging process. Respectively, at each time t two vectors of weights for the model u are calculated as $\omega_{t|t-1,u}$ and $\omega_{t|t,u}$. The first quantity denotes the weight of a specific model given information available at time $t-1$, while the latter one represents the dedicated weight to the specific model after the model update at time t. Raftery *et al.* (2010) suggests the use of a forgetting factor *'δ'* where the weights for the following period are formulated as:

$$\omega_{t+1|t,u} = \frac{\omega_{t|t,u}^{\delta}}{\sum_{l=1}^{U} \omega_{t|t,l}^{\delta}}. \tag{9}$$

The δ is set to control the 'forgetting' of the entire model set and its range is $0 < \delta \leq 1$. Raftery et al. (2010) set $\delta = 0.99$ and they also introduce the second forgetting factor, λ, that is used to account for the information loss over time. This factor is used in the variance estimator as:

$$V_t^{(u)} = (1 - \lambda^{-1}) C_{t-1}^{(u)}. \tag{10}$$

where $C_t^{(u)}$ is the conditional variance. In that way, λ controls the amount of shock affecting the coefficients $\zeta_t^{(u)}$. Identical to δ, λ may also take values near to one. This determines the rate of which information loses effect on the model coefficients.

In this study, we follow the recommendations of Raftery et al. (2010) and set $\delta = \lambda = 0.99$. The computational burden for such a dynamic model is obvious, as the total number of candidate models is $U = 2^{v*}$. Unless v^* is very small, updating the parameters becomes computationally very slow using a full Bayesian approach. Although through the work of Raftery et al. (2010), DMA can become more efficient, the computational burden still increases exponentially when v^* is large. This makes DMA impractical with standard computer processing when v^* is larger than 20. Nonetheless, for our case this is not a problem as the input dimension is always significantly lower than that (see table 3).

4. Statistical Performance

The forecasting performance of our models is evaluated through four statistics, namely the RMSE, the MAE, the MAPE and the Theil-U. These traditional metrics are interpreted as the lower their output, the better the forecasting accuracy of the respective model. Table 4 presents the out-of-sample statistical performance of the models.

Insert Table 4

The above results show that the models' statistical ranking is consistent across all factor series. RW appears to be the worst model, while the best predictors are always outperforming it. This is in a sense expected, nonetheless the best individual predictors never beat the forecast combination models. This goes towards the literature that suggests that encompassing robust forecasts can boost forecasting accuracy (Diebold and Pauly, 1990). Taking a look at the forecast combination methods, we note that the Bayesian DMA is providing more accurate forecasts that the SVR counterparts. This is in line with several studies that suggest DMA can be a robust prediction tool when certain individual predictors are provided (Koop and Korobilis, 2012; Aye et al., 2015). The traditional SVR also falls short to the SC-SVR, showing once more that the SVR parameterization with more sophisticated techniques than the traditional grid search decreases forecast errors. The above results indicate a forecasting power ranking to the competing models, but further statistical validation is needed. For that reason, we perform another two tests, namely the Pesaran-Timmermann (PT) (1992) and the Diebold Mariano (DM) (1995) test[7]. The results of the two tests are provided in table 5.

Insert Table 5

The two tests support the statistical ranking presented before. The PT statistics indicate that all models are capable of capturing the directional movements of the five factors return series in the out-of-sample[8]. Moreover, DMA's statistical superiority is confirmed, as all the DM statistic realizations are negative.

In order to further validate the superiority of DMA and affirm that our results do not suffer from data-snooping bias, we resort to a multiple hypothesis testing framework. The results presented in tables 4 and 5 might be due to lack and the outperformance of the DMA in the out-of-sample insignificant. For

[7] The PT test examines whether the directional movements of the real and forecast values are in step with one another. The PT test's null hypothesis is that the model under study has no power on forecasting the relevant factor return series. The DM statistic tests the null hypothesis of equal predictive accuracy between two forecasts. In this case, the DM test is applied to couples of out-of-sample forecasts (best model vs. other model) using the MSE loss function. In our case, a negative realization of the DM value would indicate that the DMA forecast is more accurate than the competing forecast.

[8] Similar results are obtained also in the in-sample period. In-sample results are not provided within text for the sake of space and are available upon request.

this purpose, we apply the Stepwise Superior Predictive Ability test (s-SPA) of Hsu et al. (2010) and the Model Confidence Set (MCS) of Hansen et al. (2011) under the MAE criterion. The first test focuses on the comparison of the predictive abilities of multiple methods within a full set of models. Low s-SPA p-values indicate that the benchmark model is inferior to at least one of the other models, hence the null hypothesis is rejected. Here, all the models of tables 4 and 5 are used as benchmarks in turn starting with RW. The second test deduces the best models from a full set of models (in our case those of the tables before) under specified criteria and a given level of confidence. The MCS set is data-dependent, as Hansen et al. (2011) suggest that the more informative the data are, the less models are selected. Low p-values suggest that the model under study is unlikely to belong to the set of the best performing models. The results of the two tests are presented in table 6 below.

Insert Table 6

The s-SPA and MCS results confirm the superior performance of DMA. The s-SPA tests show that for each factor case the models examined are inferior to at least one alternative model. It is logical to assume that this happens because DMA achieves the best forecasting performance. The MCS results also suggest DMA is the only model that belongs to the set of the best models[9]. Thus, all the statistical findings suggest that the Bayesian forecast combinations provide the lowest forecast errors. It would be interesting to see if this superiority is translated also into successful portfolio allocations.

5. Portfolio Optimization Design

The next target of this study is to examine what (if any) diversification benefits can be obtained from the improvement of the individual forecasts. We follow two portfolio optimization approaches which are described in this section, namely the traditional Mean-Variance (M-V) optimization and the skewed *t* copula-based mean-CVaR optimization method.

5.1 The Mean-Variance Optimization

The traditional M-V optimization is by far the most widely used method to choose optimal portfolio weights, which assumes that a rational investor wishes to find portfolios that have the best expected return-risk trade-off. In this case the variance is set as the risk proxy of the portfolio. The optimal weights are obtained by minimizing the variance of the portfolio for a given level of expected return. Thus, the optimization problem in this paper can be expressed as:

$$\min_{\mathbf{w}_t} \sigma^2_{p,t}\left(\mathbf{w}_t\right) = \mathbf{w}_t^T \Sigma_t \mathbf{w}_t \text{ subject to } r_{p,t} = \mathbf{w}_t^T r_{i,t} \text{ and } \mathbf{w}_t^T \mathbf{1} = 1 \qquad (11)$$

[9] Here we should note that when DMA is dropped from the s-SPA test, the SC-SVR is found to be superior from all other models. For the case of MCS, the results remain the same when the confidence level is further relaxed or restricted with the number of replications set at 10,000.

where $\sigma_{p,t}^2$ represents portfolio variance at time t, $r_{p,t}$ represents the expected return of the portfolio, $\mathbf{w}_t$ represents the vector of portfolio weights, and Σ_t represents the covariance matrix of factor returns at time *t*. In this study, Σ_t is predicted by three different models, namely the DCC-GARCH, the ADCC-GARCH and the GAS model.

5.2 The Copula-based Mean-CVaR Optimization

Variance is commonly used as a risk measure of portfolio because of its computational advantages. However, it is not perfect from both theoretical and empirical perspectives. Variance is not a coherent risk measure since it fails to satisfy monotonicity and translation invariance, while investors tend to value downside losses and upside gains differently (Artzner et al. 1999). However, variance penalizes profits and losses in an equal way as a symmetric risk measure. Therefore, this study considers CVaR as an alternative risk measure, because it is a coherent risk measure and focuses on the tail risk of portfolio. Rockafellar and Uryasev (2000, 2002) show that the minimum portfolio CVaR and the mean-CVaR efficient frontier can be easily obtained by using programming techniques. Following their theoretical framework, we use the forecasts obtained from different models to implement a copula-based mean-CVaR portfolio optimization. This allows us to quantify the diversification benefits yielded from the improvement of the forecasts of the factor returns.

For that reason, our first step is to specify adequate models for the marginal distributions of factor returns before the copula modelling. The specification of marginal models is standard procedure provided in the appendix B. Next, selecting appropriate dependence structure is of particular importance in optimization. Christoffersen et al. (2013) show solid evidence of asymmetric tail dependence across equity factors and our empirical results in table 7 further verify the presence of asymmetry.

Insert Table 7

Appendix C includes the specifications of the tail dependence coefficients. The results show that an appropriate copula needs to be selected in order to incorporate multivariate asymmetry is necessary. In this study, we adopt the skewed *t* copula to model and capture the nonlinear asymmetric dependence across the five Fama-French factors. As mentioned in introduction, this type of copula is advantageous compared to others since it not only takes into account the tail dependence, but also the multivariate asymmetry across the factors. More details on how the factor dependence is modelled through this copula are given in appendix D.

Following Patton (2013) and Oh and Patton (2018), we use the skewed *t* copula with GAS dynamics to model the time-varying correlation across the factors and apply the estimated parameters to implement a Monte Carlo simulation. This allow us to estimate the portfolio CVaR for the optimization. The key idea of the mean-CVaR optimization approach is to calculate portfolio VaR and minimize CVaR simultaneously for given level of expected returns. More specifically, our optimization process contains following steps. Initially, we utilize the monthly forecasts of the RW, the best individual predictor, SVR, SC-SVR and DMA. Next, we combine the GAS dynamics with the skewed t copula to capture asymmetric dependence across the factor returns. The correlation matrix for the copula is predicted with the DCC, ADCC and GAS models. All ex ante one-step ahead forecasts are obtained with a rolling window approach. This is a sound approach as studies that employ mimicking portfolios, as ours, should avoid using fixed weights (Ferson et al., 2016). In this study, a 5-year rolling window is used to re-estimate the skewed t copula for each month[10]. Then, given the estimated time-varying copula parameters, we implement a Monte Carlo simulation to predict the VaR and CVaR for the simulated factor portfolios. Finally, the portfolio CVaR can be minimized using linear programming and an efficient frontier of optimal risk-return portfolio for each month can be obtained for a series of target returns. In this study, we choose the optimal weights of the tangency portfolio (the one with higher Sharpe ratio or Return/CVaR ratio in the frontier) to rebalance our factor portfolios every month.

The mean-CVaR optimization is pioneered by Rockafellar and Uryasev (2000, 2002). The β-VaR and β-CVaR of the equity factor portfolio at time t in integral form are given by:

$$\alpha_\beta(\mathbf{w}_t) = \min\left\{\alpha \in \mathbb{R} : \Psi(\mathbf{w}_t, \alpha) \geq \beta\right\} \tag{12}$$

$$\phi_\beta(\mathbf{w}_t) = (1-\beta)^{-1} \int_{f(\mathbf{w}_t, \mathbf{r}_t) \geq \alpha_\beta(\mathbf{w}_t)} f(\mathbf{w}_t, \mathbf{r}_t)\, p(\mathbf{r}_t)\, d\mathbf{r}_t \tag{13}$$

where Ψ is the cumulative distribution for the loss associated with $\mathbf{w}_t$, the probability that $\mathbf{r}_t$ occurs is $p(\mathbf{r}_t)$ and the loss function is presented by $f(\mathbf{w}_t, \mathbf{r}_t)$ as:

$$f(\mathbf{w}_t, \mathbf{r}_t) = -\left[w_{1,t} r_{1,t} + \cdots + w_{n,t} r_{n,t}\right] = -\mathbf{w}_t^T \mathbf{r}_t \tag{14}$$

The β-CVaR of portfolio in integral form can be well approximated with a Monte Carlo simulation (Rockafellar and Uryasev, 2000). Therefore, the following equation is a suitable approximation that can be used to minimize CVaR for a given level of portfolio return:

$$\min_{(\mathbf{w}_t, \alpha)} F_\alpha(\mathbf{w}_t, \beta) = \alpha + \frac{1}{q(1-\beta)} \sum_{m=1}^{q} \left[-\mathbf{w}_t^T \mathbf{r}_{m,t} - \alpha\right]^+ \text{ subject to } \mu(\mathbf{w}_t) = -\mathbf{w}_t^T \mathbf{r}_t \leq -R \text{ and } \mathbf{w}_t^T \mathbf{1} = 1 \tag{15}$$

where q denotes the number of samples generated by the skewed t copula-based Monte Carlo simulation, α denotes VaR at β level and $\mathbf{1}$ is a vector of ones and $\mathbf{r}_{m,t}$ is the mth vector of simulated returns. The vector of optimal weights, $\mathbf{w}_t$, can be obtained from the optimization procedure to generate the portfolio that minimizes CVaR for a given target return R.

6. Portfolio Optimization Results

[10] We use a rolling window of 60 months (5 trading years) for all the data sets. We conduct rolling forecast by moving forward a month at a time and end with the forecast for August 2017.

In this section, we compare empirically the out-of-sample performances of the different optimization strategies from various combinations of forecasting models to the benchmark 1/N strategy (an equally-weighted portfolio). The following three tables show the results for the M-V, the mean-95% CVaR and the mean-99% CVaR optimization. In each of these tables, the various strategies being examined are listed in rows, while the columns refer to the different performance measures. Except from the traditional measures of annualized return, Sharpe ratio. Sortino ratio and MDD, we incorporate also the Return/CVaR ratio and the CDB. The latter is a proposed measure from Christoffersen et al. (2012) and its description is provided in appendix E. Table 8 is providing the M-V results.

Insert Table 8

Panel A shows the performance of the factors and the 1/N strategy (equally-weighted portfolio of 5 factors). The RMW factor has the highest annualized return (5.369%) and 1/N strategy yields the highest Sharpe ratio (0.918) and Sortino ratio (1.410), as well as the lowest maximum drawdown (9.346%). In Panel B the M-V optimization results with short-selling constraints are presented. In general, we find that the DMA-based portfolios, model that provides the most accurate predictions in Section 4, yield better performances than portfolios based on the SC-SVR and RW. The poor performance of the optimization based on the RW model is expected because it yields the least accurate predictions of factor returns among all the forecasting models. Another worth noting finding is that the optimizations based on the RW significantly underperform the 1/N benchmark. This indicates that the errors in forecasting the first moment of the returns dilute all the gains from the optimization. The average Sharpe ratio, Sortino ratio and CDB of the portfolios from the DMA model (1.196, 2.175 and 0.798 respectively) are significantly higher than the average of portfolios from the SC-SVR model (0.937, 1.751 and 0.767 respectively). In addition, the average maximum drawdown of the DMA-based portfolios (13.217%) is slightly lower than the SC-SVR-based portfolios (13.393%) which suggests that more accurate return predictions can mildly reduce downside risk in optimal portfolios. It should be noted that an interesting finding is observed in terms of the second moment prediction. The optimizations based on the predictions from the GAS model yield better results than the ones based on the DCC and ADCC models. Specifically, the average Sharpe ratio, Sortino ratio and CDB of the portfolios from the GAS model (0.960, 1.679 and 0.779 respectively) are clearly higher than the average of portfolios from the ADCC model (0.902, 1.598 and 0.777 respectively). Finally, the short-selling results are presented in Panel C. The portfolios allowing short-selling (130/30 strategy)[11] yield better performance compared to the one in Panel B. In particular, the average annualized returns are above 6.6%, while the average Sharpe ratio, Sortino ratio and CDB of panel C are further improved (0.971, 1.859 and 0.821 respectively). Finally, the average maximum drawdown across all portfolios is also around 1.9% lower

[11] 130/30 strategy uses leverage by shorting poor-performing assets and purchasing well-performing ones. A 130-30 ratio means that we short assets up to 30% of the portfolio value and then use the funds to take a long position in the assets with better performances.

than the one of Panel B. This improvement implies that allowing short-selling could yield higher diversification benefits than long-only portfolios.

Next, we further investigate whether more accurate estimates of factor return moments can yield more successful trading strategies in the mean-CVaR optimization, which minimizes the tail risk instead of the variance. Table 9 presents the results of mean-CVaR optimization at 95% confidence level. We replace the Sharpe ratio with the ratio of return over CVaR because the objective of this optimization is to minimize portfolio CVaR instead of variance. The other risk-return measures remain the same.

Insert Table 9

The above results indicate that the portfolios from the DMA model still significantly outperform 1/N strategy and portfolios from the SC-SVR and RW. Specifically, without short-selling the average Return/CVaR ratio, Sortino ratio and CDB (2.859, 2.470 and 0.838 respectively) DMA portfolios are significantly higher than the average of portfolios from the second best model (SC-SVR) (2.204, 1.918 and 0.788 respectively). The average maximum drawdown of DMA portfolios is 0.9% lower than the average of SC-SVR portfolios. Similar to the findings in table 8, the portfolios from the GAS model still yields evidently better performance than the ones using DCC and ADCC model across all metrics, indicating that taking into account asymmetric dependence in the optimization can lead to substantial risk reduction. For example, GAS models, when compared with the DCC, achieve on average around 0.7% and 0.06 higher annualized returns and return/CVaR ratio respectively. Finally, we continue to observe that the portfolios with short-selling provide better performance than the ones without short-selling constraints in general. This means that short-selling can offer additional diversification benefits to the CVaR optimization.

Finally, we further implement the mean-CVaR portfolios at 99% confidence level as a robustness test. These results are shown below.

Insert Table 10

The trend of our findings remains the same. The last table confirms the empirical evidence provided by the previous tables and validates the fact that the performances reported in this study are not very sensitive to the change of risk tolerance.

7. Conclusions

This study is attempting to provide further insight to the challenging task of factor investing. More specifically, examine the predictability of the factor returns and explore whether active investment is possible to be achieved by optimally allocating them in portfolios. In order to do this, we examine the Fama and French (2015) dataset over the period of 1965-2017 on a monthly basis. The out-of-sample estimations are over the period of 2000-2017. We propose a novel three-stage optimization approach. In the first step we obtain individual forecasts for each of the five factors based on a large pool of linear and non-linear models commonly used in the forecasting literature. Then, once these forecasts are obtained, they are combined with the Bayesian DMA in order to obtain superior forecast combinations. These are then benchmarked with RW, the best performing individual predictor, SVR and SC-SVR.

The final step of the approach is to proceed with the portfolio allocation for each factor. In order to capture the characteristics of correlation among factors, a DAC model is applied combining the properties of the GAS model and skewed *t* copula. For comparison purposes, the traditional DCC and ADCC techniques are utilized. The final optimal allocations are obtained through the mean-CVaR optimization, while for comparison purposes traditional M-V results are presented. From the mean-CVaR and M-V efficient frontiers, the relevant tangency portfolios are extracted and evaluated through several performance metrics. Except from the traditional performance metrics of annualized return, Sharpe and Sortino ratios, Maximum Drawdown (MDD), we look into the return over CVaR ratio and the CDB of Christoffersen et al. (2012).

In terms of the results, combining Bayesian forecasts with the proposed skewed *t* copula-based GAS model leads to the best allocations. The factor-based portfolios obtained from the more elaborate forecast combination techniques, SC-SVR and DMA, offer significant improvements over those of RW and the equal weighted strategy 1/N strategy in terms of portfolio risk reduction. The dynamic GAS-skewed *t* copula driven optimization boosts the portfolio performance compared to the DCC and ADCC counter parts. Finally, the portfolios that allow short-selling offer higher diversification benefits than long-only ones and this is particularly verified by the CDB. In general, this study's message is that the accuracy of the first moments of returns is very beneficial in portfolio optimization when we simultaneously account for the asymmetric dependence among them. Therefore, investors and practitioners interested in portfolios mimicking factors' performance should focus seriously on both aspects if they want to maximize their utility. Finally, knowing that VaR represents a worst-case loss associated with a probability and a time horizon, investors worried about the downside tail of risk should apply CVaR minimization techniques, because CVaR quantifies the expected losses that occur beyond that VaR breakpoint.

Overall, the results of this paper support the notion that factor investing can be a robust asset allocation approach for institutional investors, especially when they want to mitigate exposure to risk either from market turmoil or managers' biases. With factor mimicking portfolios, investors are also offered the opportunity to create tradable funds that are engineered in a way that potential factor sensitivities of their asset are captured. Thus, the final message that this work conveys is that factor-based portfolios show another solid path to diversification benefits.

References

Aas, K. and Haff, I. H. (2006). The generalized hyperbolic skew student'st-distribution. Journal of Financial Econometrics, 4(2), pp. 275-309.
Alexander, S., Coleman, T.F. and Li, Y. (2006). Minimizing CVaR and VaR for a portfolio of derivatives. Journal of Banking & Finance, 30(2), pp. 583-605.
Ang, A. (2014). Asset Management: A Systematic Approach to Factor Investing. Oxford University Press.
Artzner, P., Delbaen, F., Eber, J.M. and Heath, D. (1999). Coherent measures of risk. Mathematical Finance, 9(3), pp. 203-228.
Aye, G., Gupta, R., Hammoudeh, S. and Kim, W.J. (2015). Forecasting the price of gold using dynamic model averaging. International Review of Financial Analysis, 41, pp. 257-266.
Boubaker, H. and Sghaier, N. (2013). Portfolio optimization in the presence of dependent financial returns with long memory: A copula based approach. Journal of Banking & Finance, 37(2), pp. 361-377.
Bender, J., Briand, R., Nielsen, F., and Stefek, D. (2010). Portfolio of risk premia: A new approach to diversification. Journal of Portfolio Management, 36(2), pp 17-25.
Bender, J., Briand, R., Melas, D. and Subramanian, R. (2013). Foundations of factor investing, MSCI Research Insight.
Briere, M. and Szafarz, A. (2018). Factor investing: The rocky road from long-only to long-short. In Jurczenko, E. (ed) Factor Investing: From Traditional to Alternative Risk Premia, Elsevier, pp. 25-45.
Cappiello, L., Engle, R.F. and Sheppard, K. (2006). Asymmetric dynamics in the correlations of global equity and bond returns. Journal of Financial econometrics, 4(4), pp. 537-572.
Carhart, M. M. (1997). On persistence in mutual fund performance. The Journal of Finance, 52(1), pp. 57-82.
Cerniglia, J. and Fabozzi, F.J. (2018). Academic, Practitioner, and Investor Perspectives on Factor Investing. The Journal of Portfolio Management, 44(4), pp.10-16.
Cerrato, M., Crosby, J., Kim, M. and Zhao, Y. (2017). Relation between higher order comoments and dependence structure of equity portfolio. Journal of Empirical Finance, 40, pp. 101-120.
Cha, H.J. and Jithendranathan, T. (2009). Time-varying correlations and optimal allocation in emerging market equities for the US investors. International Journal of Finance & Economics, 14(2), pp.172-187.
Chan, L. K., Karceski, J. and Lakonishok, J. (1999). On portfolio optimization: Forecasting covariances and choosing the risk model. The Review of Financial Studies, 12(5), pp. 937-974.
Cherkassky, V. and Ma, Y. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. Neural networks, 17(1), pp.113-126.
Christoffersen, P., Errunza, V., Jacobs, K. and Langlois, H. (2012). Is the potential for international diversification disappearing? A dynamic copula approach. The Review of Financial Studies, 25(12), pp. 3711-3751.
Christoffersen, P. and Langlois, H. (2013). The joint dynamics of equity market factors. Journal of Financial and Quantitative Analysis, 48(5), pp. 1371-1404.
Creal, D., Koopman, S.J. and Lucas, A. (2013). Generalized autoregressive score models with applications. Journal of Applied Econometrics, 28(5), pp.777-795.
Demarta, S. and McNeil, A.J. (2005). The t copula and related copulas. International Statistical Review, 73(1), pp.111-129.
DeMiguel, V., Garlappi, L. and Uppal, R. (2007). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?. The Review of Financial Studies, 22(5), pp. 1915-1953.
Diebold, F.X. and Mariano, R.S. (1995). Comparing predictive ability. Journal of Business & Economic Statistics, 13 (3), pp. 253-263.
Diebold, F.X. and Pauly, P., 1990. The use of prior information in forecast combination. International Journal of Forecasting, 6(4), pp. 503-508.
Dimson, E., Marsh, P. and Staunton, M. (2017). Factor-based investing: the long-term evidence. Journal of Portfolio Management, 43(5), pp.15-37.
Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. Journal of Business & Economic Statistics, 20(3), pp. 339-350.
Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. Journal of Financial Economics, 33(1), pp. 3-56.
Fama, E.F. and French, K.R. (1995). Size and book - to - market factors in earnings and returns. The journal of Finance, 50(1), pp.131-155.
Fama, E. F. and French, K. R. (2015). Incremental variables and the investment opportunity set. Journal of Financial Economics, 117(3), pp. 470-488.
Fama, E.F. and French, K.R. (2016). Dissecting anomalies with a five-factor model. The Review of Financial Studies, 29(1), pp. 69-103.
Fama, E.F. and French, K.R. (2018). Choosing factors. Journal of Financial Economics, 128(2), pp. 234-252.

Fei, P., Ding, L. and Deng, Y. (2010). Correlation and volatility dynamics in REIT returns: performance and portfolio considerations. Journal of Portfolio Management, 36(2), pp.113-125.
Feldkircher, M., Horvath, R. and Rusnak, M. (2014). Exchange market pressures during the financial crisis: A Bayesian model averaging evidence. Journal of International Money and Finance, 40, pp. 21-41.
Ferson, W., Siegel, A. F. and Xu, P. T. (2006). Mimicking portfolios with conditioning information. Journal of Financial and Quantitative Analysis, 41(3), 607-635.
Fernandes, F.D.S., Stasinakis, C. and Zekaite, Z. (2018). Forecasting government bond spreads with heuristic models: evidence from the Eurozone periphery. Annals of Operations Research, pp.1-32.
Gandomi, A.H. and Alavi, A. H. (2012). Krill herd: A new bio-inspired optimization algorithm. Communications in Nonlinear Science and Numerical Simulation, 17 (12), pp. 4831-4845.
Hansen, P.R., Lunde, A. and Nason, J.M., (2011). The model confidence set. Econometrica, 79(2), pp. 453-497.
Hsu, P.H., Hsu, Y.C. and Kuan, C.M., (2010). Testing the predictive ability of technical analysis using a new stepwise test without data snooping bias. Journal of Empirical Finance, 17(3), pp. 471-484.
Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. The Journal of Finance, 48(1), pp. 65-91.
Kakouris, I. and Rustem, B. (2014). Robust portfolio optimization with copulas. European Journal of Operational Research, 235(1), pp. 28-37.
Karmakar, M. and Paul, S. (2018). Intraday portfolio risk management using VaR and CVaR: A CGARCH-EVT-Copula approach. International Journal of Forecasting, 35(2), pp. 699-709
Koop, G. and Korobilis, D. (2012). Forecasting inflation using dynamic model averaging. International Economic Review, 53(3), pp. 867-886.
Kozak, S., Nagel, S. and Santosh, S. (2018) Interpreting factor models. The Journal of Finance, 73(3), pp.1183-1223.
Li, S., Fang, H., and Liu, X. (2018). Parameter optimization of support vector regression based on sine cosine algorithm. Expert Systems with Applications, 91, pp. 63-77.
Liu, C. and Maheu, J.M. (2009). Forecasting realized volatility: a Bayesian model‐averaging approach. Journal of Applied Econometrics, 24(5), pp.709-733.
Lu, C.J., Lee, T.S. and Chiu, C.C. (2009). Financial time series forecasting using independent component analysis and support vector regression. Decision Support Systems, 47 (2), pp. 115-125.
Luo, Y. and Mesomeris, S. (2015). Factor Investing and Portfolio Construction Techniques. In Risk-Based and Factor Investing (pp. 401-433). Elsevier.
Lucas, A., Schwaab, B. and Zhang, X. (2014). Conditional euro area sovereign default risk. Journal of Business & Economic Statistics, 32(2), 271-284.
Markowitz, H. (1952). Portfolio selection. The Journal of Finance, 7(1), pp. 77-91.
Pástor, Ľ. and Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. Journal of Political Economy, 111(3), pp. 642-685.
Miralles‐Quiros, M.D.M. and Miralles‐Quiros, J.L. (2017). The role of time‐varying return forecasts for improving international diversification benefits. International Journal of Finance & Economics, 22(3), pp. 201-215.
Mirjalili, S. (2016). SCA: a sine cosine algorithm for solving optimization problems. Knowledge-Based Systems, 96, pp. 120-133.
Oh, D.H. and Patton, A.J. (2018). Time-varying systemic risk: Evidence from a dynamic copula model of cds spreads. Journal of Business & Economic Statistics, 36(2), pp.181-195.
Parejo, J.A., Ruiz-Cortés, A., Lozano, S. and Fernandez, P. (2012). Metaheuristic optimization frameworks: a survey and benchmarking. Soft Computing, 16(3), pp. 527-561.
Patton, A.J. (2006). Modelling asymmetric exchange rate dependence. International economic review, 47(2), pp. 527-556.
Patton, A. (2013). Copula methods for forecasting multivariate time series. In Handbook of Economic Forecasting (Vol. 2, pp. 899-960). Elsevier.
Pesaran, M.H. and Timmerman, A.G. (1992). A simple nonparametric test of predictive performance. Journal of Business and Economic Statistics, 10 (4), pp. 461-465.
Raftery, A.E., Kárný, M. and Ettler, P. (2010). Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. Technometrics, 52(1), pp. 52-66.
Rockafellar, R.T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. Journal of risk, 2, pp. 21-42.
Rockafellar, R.T. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. Journal of banking & finance, 26(7), pp. 1443-1471.
Sahamkhadam, M., Stephan, A. and Östermark, R. (2018). Portfolio optimization based on GARCH-EVT-Copula forecasting models. International Journal of Forecasting, 34(3), pp. 497-506.

Savu, C. and Trede, M. (2010). Hierarchies of Archimedean copulas. Quantitative Finance, 10(3), pp. 295-304.
Salvatierra, I.D.L. and Patton, A.J. (2015). Dynamic copula models and high frequency data. Journal of Empirical Finance, 30, pp. 120-135.
Sermpinis, G., Stasinakis, C. and Hassanniakalager, A. (2017). Reverse adaptive krill herd locally weighted support vector regression for forecasting and trading exchange traded funds. European Journal of Operational Research, 263(2), pp. 540-558.
Schölkopf, B., Bartlett, P. L., Smola, A. J. and Williamson, R. (1999), Shrinking the tube: a new support vector regression algorithm, In Kearns, M.J. (ed.), Advances in neural information processing systems 11. Cambridge, Mass, MIT Press, pp. 330-336.
Tipping, M.E. (2001). Sparse Bayesian learning and the relevance vector machine. Journal of machine learning research, 1(Jun), pp. 211-244.
Vapnik, V. (1995), The nature of statistic learning theory. Springer-Verlag, New York.
Yang, X. S. (2010). Firefly algorithm nature-inspired metaheuristic algorithms. United Kingdom: Luniver Press.
Yang, X. S. and Gandomi, A. H. (2012). Bat algorithm: a novel approach for global engineering optimization. Engineering Computations, 29 (5), pp. 464–483.
Yao, X., Crook, J. and Andreeva, G. (2015). Support vector regression for loss given default modelling. European Journal of Operational Research, 240(2), pp. 528-538.
Wang, L. and Zhu, J. (2010). Financial market forecasting using a two-step kernel learning method for the support vector regression. Annals of Operations Research, 174(1), pp. 103-120.
Wright, J.H. (2008). Bayesian model averaging and exchange rate forecasts. Journal of Econometrics, 146(2), pp. 329-341.
Wright, J.H. (2009). Forecasting US inflation by Bayesian model averaging. Journal of Forecasting, 28(2), pp.131-144.
Zhao, Y., Stasinakis, C., Sermpinis, G. and Shi, Y. (2018). Neural network copula portfolio optimization for exchange traded funds. Quantitative Finance, 18(5), pp.761-775.

Appendix

A. Linear Individual Predictors

Table A.1 provides short descriptions of the linear individual predictors.

Insert Table A.1

B. Univariate Modelling

The results of the ACF test indicate that the HML, RMW and CMA factors exhibit some degree of autocorrelation and all factors exhibit significant heteroscedasticity. To compensate for autocorrelation, the conditional mean is modelled with a simple AR model:

$$r_{i,t} = c + \sum_{j=1}^{p} \varphi_j r_{i,t-j} + \varepsilon_{i,t} \tag{B.1}$$

where $\varepsilon_{i,t} = \sigma_{i,t} z_{i,t}$. The optimal order is selected using Bayesian Information Criterion (BIC). To capture the heteroscedasticity and asymmetric volatility clustering, the conditional variance of factor returns are modelled using the GJR-GARCH dynamics:

$$\sigma_{i,t}^2 = \omega + \sum_{j=1}^{p} \alpha_j \varepsilon_{i,t-j}^2 + \sum_{k=1}^{q} \beta_k \sigma_{i,t-k}^2 + \sum_{k=1}^{q} \gamma_k \varepsilon_{i,t-k}^2 I\left[\varepsilon_{i,t-k} < 0\right] \tag{B.2}$$

The indicator function $I[\varepsilon_{i,t-k} < 0]$ equals 1 if $\varepsilon_{i,t-k} < 0$, and 0 otherwise. It allows us to capture the so-called "leverage effect", which implies a negative shock has a stronger impact on the conditional variance than a positive one. Using these models, we obtain the estimated standardized residuals as:

$$z_{i,t} = r_{i,t} - c - \sum_{j=1}^{p} \varphi_j r_{i,t-j} / \sigma_{i,t} \tag{B.3}$$

The descriptive statistics indicate that all factors exhibit significant skewness and the hypothesis of normality is rejected by the Jarque–Bera test. In order to capture the skewness, we use the univariate skewed *t* distribution of Hansen (1994) to model the standardized residuals of each factor. Assuming $z_{i,t} \sim F_{skt}\left(\eta_i, \lambda_i\right)$, then

$$u_{i,t} = F_{skt}\left(z_{i,t}; \eta_i, \lambda_i\right),\ \eta_i \in (2, \infty),\ \lambda_i \in [-1, 1] \tag{B.4}$$

where $u_{i,t}$ is the probability integral transform of $z_{i,t}$, λ_i is the skewness parameter and η_i is the degrees of freedom.

C. Tail Dependence Coefficients

The multivariate fat tails between the factor returns can be measured by the tail dependence coefficients (Christoffersen et al., 2013). The lower tail dependence (LTD) and upper tail dependence (UTD) coefficients are defined as:

$$\text{LTD}\left(R_i, R_j\right) = \lim_{q \to 0^+} \Pr\left[R_j \le F_j^{-1}(q) \middle| R_i \le F_i^{-1}(q)\right] = \lim_{q \to 0^+} \frac{C(q,q)}{q} \tag{C.1}$$

$$\text{UTD}\left(R_i, R_j\right) = \lim_{q \to 1^-} \Pr\left[R_j \ge F_j^{-1}(q) \middle| R_i \ge F_i^{-1}(q)\right] = \lim_{q \to 1^-} \frac{1 - 2q + C(q,q)}{1-q} \tag{C.2}$$

If the copula *C* has an analytic solution, the coefficients can be easily calculated. The copula *C* has lower tail dependence if $\text{LTD} \in (0,1]$ and no lower tail dependence if LTD = 0. A similar conclusion holds for the upper tail dependence coefficients. In our application, the *t* copula is applied to compute tail dependence coefficients.

D. Modeling Factor Dependence: A Skewed *t* Copula Approach

This study employs the skewed *t* copula proposed by Demarta and McNeil (2005). The cumulative distribution function of this skewed *t* copula defined from the skewed t distribution is given by:

$$\mathbf{C}_{skt}\left(u_{1,t}, \ldots, u_{n,t}; \Sigma, \lambda, \upsilon\right) = \mathbf{F}_{skt}\left(F_1^{-1}\left(u_{1,t}\right), \ldots, F_n^{-1}\left(u_{n,t}\right)\right). \tag{D.1}$$

where λ is the parameter of skewness, υ is the parameter of degree of freedom, $\mathbf{F}_{skt}$ is the cumulative distribution function of the multivariate skewed *t* density with correlation matrix Σ, and F_i^{-1} is the inverse cumulative distribution function of the univariate skewed *t* distribution.
From Patton (2006), if the joint distribution function $\mathbf{F}_{skt}$ is n-times differentiable, the following equation is obtained by taking the n[th] cross-partial derivative:

$$\begin{aligned}\mathbf{f}_{skt}\left(u_{1,t},\ldots,u_{n,t}\right) &= \frac{\partial^n}{\partial x_1\cdots\partial x_n}\mathbf{F}_{skt}\left(z_{1,t},\ldots,z_{n,t}\right) = \prod_{i=1}^{n} f_i\left(z_{i,t}\right)\cdot\frac{\partial^n}{\partial x_1\cdots\partial x_n}\mathbf{C}_{skt}\left(F_1\left(z_{1,t}\right),\ldots,F_n\left(z_{n,t}\right)\right) \\ &= \prod_{i=1}^{n} f_i\left(z_{i,t}\right)\cdot\mathbf{c}_{skt}\left(F_1\left(z_{1,t}\right),\ldots,F_n\left(z_{n,t}\right)\right).\end{aligned} \quad .2)$$

The equation suggests that the joint density $\mathbf{f}_{skt}$ is equal to the product of the marginal densities and the skewed *t* copula density $\mathbf{c}_{skt}$. Thus, the joint log-likelihood is equal to the sum of univariate log-likelihood and the skewed *t* copula log-likelihood:

$$\log\mathbf{f}_{skt}\left(z_{1,t},\ldots,z_{n,t}\right) = \sum_{t=1}^{T}\sum_{i=1}^{n}\log f_{i,t}\left(z_{i,t}\right) + \sum_{t=1}^{T}\log\mathbf{c}_{skt}\left(F_1\left(z_{1,t}\right),\ldots,F_n\left(z_{n,t}\right)\right). \quad \text{(D.3)}$$

More details on the implementation of the skewed *t* copula can be found in Demarta and McNeil (2005) and Christoffersen et al., (2012). Our choice to use the skewed *t* copula is supported by the literature (See Christoffersen et al., 2012, 2013; Patton 2013; and Lucas et al., 2014, among others).

E. Conditional Diversification Benefit (CDB)

The CDB is a dynamic measure of portfolio diversification benefits proposed by Christoffersen et al. (2012). One advantage of CDB is that it takes into account higher order moments and nonlinear dependence of assets. This measure is based on the CVaR:

$$\text{CVaR}_t^q\left(R_{i,t}\right) = -\text{E}\left[R_{i,t}\,\middle|\,R_{i,t}\le F_{i,t}^{-1}(q)\right] \quad \text{(E.1)}$$

where $F_{i,t}^{-1}(q)$ is the inverse cumulative distribution function of factor *i* at time *t*, and *q* is a probability normally set to 5% or 1%[12]. The upper bound on the portfolio CVaR can be defined as the case of no diversification benefits:

$$\overline{\text{CVaR}_t^q} = \sum_{i=1}^{N} w_{i,t}\text{CVaR}_t^q\left(R_{i,t}\right) \quad \text{(E.2)}$$

where $w_{i,t}$ denotes the portfolio weight on factor *i* at time *t*. The lower bound of portfolio CVaR is defined as the extreme case that the portfolio never loses more than its (1-*q*) [th] VaR:

$$\underline{\text{CVaR}_t^q} = -F_{p,t}^{-1}(q) \quad \text{(E.3)}$$

[12] In our study, we use 1% (namely 99% confidence level) to compute CDB and also compute the 5% CDB as the robustness check. The results of these CDBs are qualitatively identical.

Then, the diversification benefit is measured by:

$$\text{CDB}_t\left(w_t, q\right) \equiv \frac{\overline{\text{CVaR}_t^q} - \text{CVaR}_t^q\left(w_t\right)}{\overline{\text{CVaR}_t^q} - \underline{\text{CVaR}_t^q}}, \quad \text{CDB}_t\left(w_t, q\right) \in [0,1] \tag{E.4}$$

where the $\text{CVaR}_t^q\left(w_t\right)$ represents the CVaR of portfolio at time *t*. Higher CDB indicates higher level of diversification benefits for the portfolio.

Tables

Table 1: Descriptive statistics and correlation matrices

Panel A: Descriptive Statistics					
Ticker	**MKT**	**SMB**	**HML**	**RMW**	**CMA**
Mean	0.399	0.363	0.321	0.447	0.341
Median	0.920	0.265	-0.075	0.375	0.020
Standard deviation	4.378	3.131	3.206	3.124	2.097
Skewness	-0.591	0.535	0.236	-0.403	1.006
Kurtosis	3.934	9.558	5.660	11.697	5.912
Jarque-Bera (p value)	0.000***	0.000***	0.000***	0.000***	0.000***
ADF (p value)	0.000***	0.000***	0.000***	0.000***	0.000***
Panel B: Linear Correlation Matrix					
Ticker	**MKT**	**SMB**	**HML**	**RMW**	**CMA**
MKT	1				
SMB	0.253	1			
HML	-0.057	-0.071	1		
RMW	-0.505	-0.525	0.424	1	
CMA	-0.238	0.039	0.616	0.278	1
Panel C: Rank Correlation Matrix					
Ticker	**MKT**	**SMB**	**HML**	**RMW**	**CMA**
MKT	1				
SMB	0.309	1			
HML	-0.018	0.103	1		
RMW	-0.580	-0.385	0.137	1	
CMA	-0.089	0.080	0.519	0.088	1

*Note: Panel A presents the descriptive statistics per factor. *** denotes that the null hypothesis is rejected at 1% significance level. Panel B and C presents the linear and rank correlations among all factors.*

Table 2: The total dataset

Datasets	Start Date	End Date	Trading Days
Total Dataset	01/01/1965	01/08/2017	632
In-sample Dataset	01/01/1965	01/12/1999	420
Training Dataset	01/01/1965	01/12/1983	228
Test Dataset	01/01/1984	01/12/1999	192
Out-of-sample Dataset	01/01/2000	01/08/2017	212

Note: The in-sample period is the sum of the training and test datasets.

Table 3: Best predictors' set

MKT	SMB	HML	RMW	CMA
AR(5), AR(6), SMA(6), SMA(8), EMA(2) MLP, RNN, **PSN,** ARBF-PSO	AR(1) AR(6) SMA(4), ARMA(1, 6), MLP, **HONN**, GP	SMA(3), ARMA(1, 2), RNN, HONN, PSN, **ARBF-PSO**	MLP, RNN, HONN, **PSN**, ARBF-PSO, *k*-NN, GEP, GP	ARMA(1,4), ARMA(2, 2), PSN, GP, GEP, **ARBF-PSO**

Note: The table presents the final input set per factor used for each forecast combination technique. The model in bold is the best predictor among all the individual predictors. For example, in the case of HML the forecasts of SMA(3), ARMA(1, 2), RNN, HONN, PSN and ARBF-PSO are used as inputs to the forecast combination methods, while the performance of ARBF-PSO acts as a benchmark to the forecast combination results.

Table 4: Out-of-sample statistical performance

Factor	Statistic	RW	Best	SVR	SC-SVR	DMA
MKT	MAE	0.0063	0.0056	0.0055	0.0052	**0.0049**
	MAPE	169.45%	167.44%	164.52%	141.21%	**130.31%**
	RMSE	0.0079	0.0078	0.0076	0.0075	**0.0072**
	THEIL-U	0.9129	0.9125	0.8286	0.7598	**0.6995**
SMB	MAE	0.0075	0.0054	0.0055	0.0051	**0.0048**
	MAPE	184.32%	161.07%	155.88%	128.81%	**118.29%**
	RMSE	0.0073	0.0073	0.0072	0.0069	**0.0065**
	THEIL-U	0.9043	0.9256	0.7635	0.7086	**0.6744**
HML	MAE	0.0075	0.0073	0.0068	0.0065	**0.0055**
	MAPE	134.54%	133.52%	128.77%	119.22%	**117.56%**
	RMSE	0.0092	0.0088	0.0088	0.0083	**0.0065**
	THEIL-U	0.9102	0.9077	0.8322	0.7980	**0.7884**
RMW	MAE	0.0065	0.058	0.0053	0.0050	**0.0049**
	MAPE	170.45%	144.03%	130.99%	121.86%	**118.45%**
	RMSE	0.0092	0.0084	0.0071	0.0065	**0.0064**
	THEIL-U	0.8440	0.8023	0.7348	0.7049	**0.6448**
CMA	MAE	0.0080	0.0074	0.0067	0.0062	**0.0058**
	MAPE	132.65%	130.74%	125.04%	123.13%	**120.45%**
	RMSE	0.0090	0.0085	0.0087	0.0080	**0.0076**
	THEIL-U	0.8689	0.8457	0.8279	0.8110	**0.7659**

Note: The third and fourth column refer to individual predictors/benchmarks. The forth column refers to the statistical performance of the best predictor as denoted in bold in Table 3. For example, in the case of MKT, the best predictor is PSN. The last three columns present the statistical accuracy of the forecast combination models that use as inputs the individual forecasts of the models presented in Table 3.

Table 5: PT and DM statistics.

Statistic	Factor	RW	Best	SVR	SC-SVR	DMA
PT	MKT	(5.18)***	(7.02)***	(8.34)***	(8.55)***	(8.78)***
	SMB	(6.03)***	(6.56)***	(7.05)***	(7.78)***	(8.54)***
	HML	(6.56)***	(6.93)***	(7.44)***	(7.93)***	(7.98)***
	RMW	(7.18)***	(7.25)***	(7.78)***	(8.46)***	(8.82)***
	CMA	(7.09)***	(7.95)***	(8.05)***	(8.66)***	(9.15)***
DM	MKT	(−9.55)***	(−8.16)***	(−7.80)***	(−5.55)***	-
	SMB	(−9.11)***	(−9.02)***	(−8.45)***	(−6.72)***	-
	HML	(−10.06)***	(−9.88)***	(−9.18)***	(−7.32)***	-
	RMW	(−10.18)***	(−9.75)***	(−9.48)***	(−8.40)***	-
	CMA	(−8.06)***	(−8.02)***	(−7.82)***	(−4.94)***	-

*Note: The values in the parentheses are the calculated PT and DM statistics. *** denotes that the null hypothesis is rejected at 1% significance level.*

Table 6: s-SPA and MCS tests

Test	Factor	RW	Best	SVR	SC-SVR	DMA
s-SPA	**MKT**	0.0000	0.0000	0.0001	0.0004	0.7586
	SMB	0.0000	0.0000	0.0003	0.0011	0.8227
	HML	0.0000	0.0000	0.0000	0.0001	0.6153
	RMW	0.0000	0.0000	0.0008	0.0015	0.5940
	CMA	0.0000	0.0000	0.0005	0.0005	0.6854
MCS	**MKT**	0.0000	0.0000	0.0009	0.0013	1.0000*
	SMB	0.0000	0.0000	0.0002	0.0016	1.0000*
	HML	0.0000	0.0000	0.0001	0.0023	1.0000*
	RMW	0.0000	0.0000	0.0012	0.0037	1.0000*
	CMA	0.0000	0.0000	0.0005	0.0008	1.0000*

*Note: The table reports the p-values for the s-SPA and MCS tests in terms of the MAE criterion. Low s-SPA values indicate that the benchmark is inferior to at least one of the other models, while low MCS values indicate that the model is not likely to belong to the set of the best models. * denotes that the model under study belongs to the set of best models at the 95% confidence level.*

Table 7: Estimates of Tail Dependence and Asymmetric Test

Factor Pairs	UTD	LTD	UTD-LTD	p-value
MKT-SMB	0.007	0.048	-0.041	0.000***
MKT-HML	0.003	0.001	0.002	0.701
MKT-RMW	0.046	0.051	-0.005	0.009***
MKT-CMA	0.035	0.024	0.010	0.036**
SMB-HML	0.013	0.027	-0.014	0.008***
SMB-RMW	0.002	0.004	-0.002	0.530
SMB-CMA	0.011	0.005	0.005	0.315
HML-RMW	0.075	0.057	0.018	0.005***
HML-CMA	0.103	0.123	-0.020	0.009***
RMW-CMA	0.069	0.017	0.051	0.007***

Note: This table reports the estimates of tail dependence coefficients and the test of asymmetric dependence. The column "UTD" and "LTD" report the estimates of upper tail dependence and lower tail dependence implied by the t copula, respectively. We follow the approach of Patton (2013) to test whether the tail dependence coefficients is equal, namely $H_0 : \lambda^L = \lambda^U$ vs. $H_a : \lambda^L \neq \lambda^U$ *. The last column shows the corresponding p-values for each factor pairs. *** denotes that the null hypothesis is rejected at 1% significance level.*

Table 8. Performances of different trading strategies (Mean-Variance)

Panel A: Factors and 1/N portfolio					
	Annualized return (%)	**Sharpe ratio**	**Sortino ratio**	**MDD (%)**	**CDB**
MKT	4.789	0.316	0.582	25.450	-
SMB	4.352	0.401	0.529	33.120	-
HML	3.848	0.347	0.468	24.000	-
RMW	5.369	0.496	0.653	23.010	-
CMA	4.091	0.563	0.497	15.050	-
1/N	4.490	0.918	1.410	9.346	0.825
Panel B: Mean–Variance optimization without short-selling					
	Annualized return (%)	**Sharpe ratio**	**Sortino ratio**	**MDD (%)**	**CDB**
RW-DCC	3.631	0.583	0.944	14.441	0.733
RW-ADCC	3.848	0.598	0.949	14.440	0.761
RW-GAS	4.675	0.615	0.951	14.671	0.758
Average	4.051	0.599	0.948	14.517	0.751
SC-SVR-DCC	4.744	0.899	1.747	13.416	0.751
SC-SVR-ADCC	4.871	0.945	1.746	13.416	0.773
SC-SVR-GAS	5.519	0.967	1.759	13.346	0.777
Average	5.045	0.937	1.751	13.393	0.767
DMA-DCC	6.810	1.127	2.099	14.690	0.798
DMA-ADCC	6.975	1.162	2.099	12.928	0.796
DMA-GAS	8.012	1.298	2.326	12.032	0.801
Average	7.266	1.196	2.175	13.217	0.798
Total Average	5.454	0.910	1.624	13.709	0.772
DCC Average	5.062	0.870	1.597	14.182	0.761
ADCC Average	5.231	0.902	1.598	13.595	0.777
GAS Average	6.069	0.960	1.679	13.350	0.779
Panel C: Mean–Variance optimization with short-selling (130/30 portfolios)					
	Annualized return (%)	**Sharpe ratio**	**Sortino ratio**	**MDD (%)**	**CDB**
RW-DCC-S	3.895	0.618	0.983	13.267	0.775
RW-ADCC-S	4.026	0.652	1.031	13.267	0.782
RW-GAS-S	5.849	0.701	1.153	13.267	0.780
Average	4.590	0.657	1.056	13.267	0.779
SC-SVR-DCC-S	5.845	0.908	1.763	12.424	0.805
SC-SVR-ADCC-S	6.088	0.947	1.830	12.354	0.806
SC-SVR-GAS-S	7.677	0.971	1.883	12.549	0.811
Average	6.537	0.942	1.825	12.442	0.807
DMA-DCC-S	8.529	1.289	2.636	9.724	0.871
DMA-ADCC-S	8.727	1.290	2.654	9.724	0.872
DMA-GAS-S	9.309	1.361	2.800	9.634	0.887
Average	8.855	1.313	2.697	9.694	0.877
Total Average - S	6.661	0.971	1.859	11.801	0.821
DCC-S Average	6.090	0.938	1.794	11.805	0.817
ADCC-S Average	6.280	0.963	1.838	11.782	0.820
GAS-S Average	7.612	1.011	1.945	11.817	0.826

Note: The table reports the out-of-sample performances of the mean-variance optimization over the period January 2000 to August 2017 (212 monthly observations). Panel A reports performances of the factors and the 1/N portfolio (equally weighted buy-and-hold portfolio). Panel B reports performances of different mean-variance portfolios without short-selling. All the portfolios are monthly rebalanced tangency portfolios obtained by the different mean-variance optimization based on various model combinations. For example, DMA-DCC refers to the performance of the tangency portfolio of the efficient frontier of the factors, where the expected returns are obtained through DMA forecasts, while the variance-covariance matrix is predicted by DCC. Panel C reports performances of different mean-variance portfolios with short-selling (130/30 portfolios). '-S' denotes optimizations allowing short-selling.

Table 9. Performances of different trading strategies (Mean-95% CVaR)

	Annualized return (%)	Return/CVaR	Sortino ratio	MDD (%)	CDB
Panel A: Factors and 1/N portfolio					
	Annualized return (%)	**Return/CVaR**	**Sortino ratio**	**MDD (%)**	**CDB**
MKT	4.789	0.343	0.582	25.450	-
SMB	4.352	0.413	0.529	33.120	-
HML	3.848	0.366	0.468	24.000	-
RMW	5.369	0.381	0.653	23.010	-
CMA	4.091	0.735	0.497	15.050	-
1/N	4.490	1.119	1.410	9.346	0.825
Panel B: Mean–CVaR optimization without short-selling					
	Annualized return (%)	**Return/CVaR**	**Sortino ratio**	**MDD (%)**	**CDB**
RW-DCC-SKT	3.913	0.759	0.652	14.990	0.739
RW-ADCC-SKT	4.780	0.806	0.696	14.990	0.741
RW-GAS-SKT	4.540	0.825	0.704	14.971	0.740
Average	4.411	0.797	0.684	14.984	0.740
SC-SVR-DCC-SKT	6.197	2.171	1.885	11.265	0.777
SC-SVR-ADCC-SKT	6.314	2.209	1.916	11.266	0.777
SC-SVR-GAS-SKT	6.624	2.231	1.952	11.330	0.779
Average	6.378	2.204	1.918	11.287	0.778
DMA-DCC-SKT	7.003	2.827	2.453	10.630	0.834
DMA-ADCC-SKT	7.061	2.862	2.477	10.630	0.834
DMA-GAS-SKT	8.035	2.887	2.480	9.894	0.846
Average	7.366	2.859	2.470	10.385	0.838
Total Average	6.052	1.953	1.691	12.218	0.785
DCC Average	5.704	1.919	1.663	12.295	0.783
ADCC Average	6.052	1.959	1.696	12.295	0.784
GAS Average	6.400	1.981	1.712	12.065	0.788
Panel C: Mean–Variance optimization with short-selling (130/30 portfolios)					
	Annualized return (%)	**Sharpe ratio**	**Sortino ratio**	**MDD (%)**	**CDB**
RW-DCC-SKT-S	4.041	0.795	0.661	14.911	0.766
RW-ADCC-SKT-S	4.185	0.811	0.725	14.911	0.768
RW-GAS-SKT-S	4.525	1.191	1.066	15.036	0.772
Average	4.250	0.932	0.817	14.953	0.769
SC-SVR-DCC-SKT-S	7.480	2.190	1.918	11.734	0.806
SC-SVR-ADCC-SKT-S	7.537	2.263	1.998	11.796	0.808
SC-SVR-GAS-SKT-S	7.687	2.314	2.047	11.022	0.813
Average	7.568	2.256	1.988	11.517	0.809
DMA-DCC-SKT-S	9.321	2.904	2.606	9.544	0.855
DMA-ADCC-SKT-S	9.431	2.924	2.614	9.544	0.856
DMA-GAS-SKT-S	9.735	3.043	2.682	9.110	0.870
Average	9.496	2.957	2.634	9.399	0.860
Total Average - S	7.105	2.048	1.813	11.956	0.813
DCC-S Average	6.947	1.963	1.728	12.063	0.809
ADCC-S Average	7.051	1.999	1.779	12.084	0.811
GAS-S Average	7.316	2.183	1.932	11.723	0.818

Note: The table reports the out-of-sample performances of the mean-95%CVaR optimization over the period January 2000 to August 2017 (212 monthly observations). Panel A reports performances of the factors and the 1/N portfolio (equally weighted buy-and-hold portfolio). Panel B reports performances of different mean-CVaR portfolios without short-selling. All the portfolios are monthly rebalanced tangency portfolios obtained by the different mean-CVaR optimization based on various model combinations. For example, DMA-DCC refers to the performance of the tangency portfolio of the efficient frontier of the factors, where the expected returns are obtained through DMA forecasts, while the variance-covariance matrix is predicted by DCC. Panel C reports performances of different mean-CVaR portfolios with short-selling (130/30 portfolios). 'SKT' represents that the 95% CVaR is predicted using a Monte-Carlo simulation with the skewed t copulas to allow for asymmetric tail dependence '-S' denotes optimizations allowing short-selling.

Table 10. Performances of different trading strategies (Mean-99% CVaR)

Panel A: Factors and 1/N portfolio					
	Annualized return (%)	**Return/CVaR**	**Sortino ratio**	**MDD (%)**	**CDB**
MKT	4.789	0.343	0.582	25.450	-
SMB	4.352	0.413	0.529	33.120	-
HML	3.848	0.366	0.468	24.000	-
RMW	5.369	0.381	0.653	23.010	-
CMA	4.091	0.735	0.497	15.050	-
1/N	4.490	1.119	1.410	9.346	0.825
Panel B: Mean–CVaR optimization without short-selling					
	Annualized return (%)	**Return/CVaR**	**Sortino ratio**	**MDD (%)**	**CDB**
RW-DCC-SKT	4.430	1.035	0.854	14.158	0.753
RW-ADCC-SKT	4.493	1.046	0.858	14.158	0.756
RW-GAS-SKT	4.627	1.120	0.869	14.139	0.755
Average	4.517	1.067	0.860	14.152	0.755
SC-SVR-DCC-SKT	7.719	1.983	2.137	10.659	0.769
SC-SVR-ADCC-SKT	7.727	1.988	2.151	10.659	0.771
SC-SVR-GAS-SKT	7.708	2.041	2.277	10.634	0.777
Average	7.718	2.004	2.188	10.651	0.772
DMA-DCC-SKT	8.044	2.156	2.990	9.350	0.861
DMA-ADCC-SKT	8.015	2.160	2.994	9.285	0.872
DMA-GAS-SKT	8.836	2.184	3.257	9.201	0.886
Average	8.298	2.167	3.080	9.279	0.873
Total Average	6.844	1.746	2.043	11.360	0.800
DCC Average	6.731	1.725	1.994	11.389	0.794
ADCC Average	6.745	1.731	2.001	11.367	0.800
GAS Average	7.057	1.782	2.134	11.325	0.806
Panel C: Mean–Variance optimization with short-selling (130/30 portfolios)					
	Annualized return (%)	**Sharpe ratio**	**Sortino ratio**	**MDD (%)**	**CDB**
RW-DCC-SKT-S	4.920	1.174	1.030	14.034	0.773
RW-ADCC-SKT-S	4.924	1.197	1.139	14.034	0.775
RW-GAS-SKT-S	5.238	1.211	1.158	14.152	0.780
Average	5.027	1.194	1.109	14.073	0.776
SC-SVR-DCC-SKT-S	7.722	1.985	2.680	10.179	0.825
SC-SVR-ADCC-SKT-S	7.727	1.985	2.688	10.179	0.829
SC-SVR-GAS-SKT-S	8.298	2.060	2.695	10.109	0.829
Average	7.916	2.010	2.688	10.156	0.828
DMA-DCC-SKT-S	9.089	2.275	2.641	9.435	0.866
DMA-ADCC-SKT-S	9.122	2.349	3.067	9.435	0.878
DMA-GAS-SKT-S	9.773	2.366	3.172	9.352	0.887
Average	9.328	2.330	2.960	9.407	0.877
Total Average - S	7.424	1.845	2.252	11.212	0.827
DCC-S Average	7.244	1.811	2.117	11.216	0.821
ADCC-S Average	7.258	1.844	2.298	11.216	0.827
GAS-S Average	7.770	1.879	2.342	11.204	0.832

Note: The table reports the out-of-sample performances of the mean-99%CVaR optimization over the period January 2000 to August 2017 (212 monthly observations). Panel A reports performances of the factors and the 1/N portfolio (equally weighted buy-and-hold portfolio). Panel B reports performances of different mean-CVaR portfolios without short-selling. All the portfolios are monthly rebalanced tangency portfolios obtained by the different mean-CVaR optimization based on various model combinations. For example, DMA-DCC refers to the performance of the tangency portfolio of the efficient frontier of the factors, where the expected returns are obtained through DMA forecasts, while the variance-covariance matrix is predicted by DCC. Panel C reports performances of different mean-CVaR portfolios with short-selling (130/30 portfolios). 'SKT' represents that the 99% CVaR is predicted using a Monte-Carlo simulation with the skewed t copulas to allow for asymmetric tail dependence '-S' denotes optimizations allowing short-selling.

Table A.1: The specification of the linear models

LINEAR MODELS	DESCRIPTION	TOTAL INDIVIDUAL FORECASTS
SMA (q)	$E(R_t) = (R_{t-1} + ... + R_{t-q}) / q, \quad q = 3...30$	28
EMA (q')	$E(R_t) = \frac{R_{t-1} + (1-a')R_{t-2} + ... + (1-a')^{q'-1} R_{t-q'}}{a' + (1-a') + ... + (1-a')^{q'-1}},$ $q' = 3...30, \quad a' = 2/(1+N_{days}),\ N_{days}$ *is the number trading days*	28
AR (q'')	$E(R_t) = \beta_0 + \sum_{i'=1}^{q''} \beta_{i'} R_{t-i'}, \quad q'' = 1, \ldots, 24, \quad \beta_0, \beta_{i'}$ *the regression coefficients*	24
ARMA (m', n')	$E(R_t) = \overline{\varphi}_0 + \sum_{j'=1}^{m'} \overline{\varphi}_{j'} R_{t-j'} + \overline{a}_0 + \sum_{k'=1}^{n'} \overline{w}_{k'} \overline{a}_{t-k'}, \quad m',\ n' = 1,..,15, \overline{\varphi}_0, \overline{\varphi}_{j'}$ *the regression coefficients,* $\overline{a}_0, \overline{a}_{t-k'}$ *the residual terms,* $\overline{w}_{k'}$ *the weights of the residual terms*	210

Note: The total number of individual inputs calculated is 290. In all the specifications above, R_t *is the factor return at time t.*

Figures

Figure 1: Cumulative Returns of Fama-French's Factors