Smart Alpha Portfolio Selection Model

Ruiying Wang

August 8, 2022

1 Introduction

In this report, we study a portfolio selection model called 'Smart Alpha' that was introduced by [5]. This model introduces a new approach that bets on alpha, which makes it different from the factor investing approaches that are very popular among asset managers and large institutional investirs. It maximizes α in CAPM (eq(1)), meaning that it maximizes the expected return that is uncorrelated to various systematic risks. And in the mean time, it minimizes the portfolio's exposure to these systematic risk factors. Alpha and exposure of systematic risk factors (i.e. beta) are estimated by the Sparse Principal Component Analysis (SPCA) methodology of [20], while the optimization process is done by a convex quadratic programming optimizer called Gurobi.

$$\mu_i = \alpha_i + r_f + \beta_i(\mu_M - r_f) \tag{1}$$

The factor investing approach is popular in recent years, because it does not require a set of 'strict' assumptions such as normal returns and stable variance, and it solves the problem of poor out-of-sample performance that mean-variance portfolio and market weighting portfolio have ([3],[4],[9],[15],[19]). However, there are three concerns:

1. The number of factors reported in the academic literature has become too numerous, and the explanatory power of many of the factors are spurious. [6],[17]

2. The success of factor investing is anchored in how accurately portfolio managers can time when a given factor is going to be rewarded by the market, and this task is complicated. [11],[14],[18]

3. There is a significant probability that stocks will have alphas that arise from stochastic mispricing, corrections of earlier over-reactions to news, slow adjustments to firm-specific news, unanticipated increases in market illiquidity, and the state of sentiment. [1],[7],[8],[13],[16]

Considering the above, [5] propose the smart alpha model that could be viewed as an orthogonal take to the factor investing approach. It is shown empirically on the data of STOXX600 that the smart alpha portfolio is economically and statistically superior to many popular factor investing models, that it provides a higher average return, a higher risk-adjusted return, as well as a lower downside risk and a lower maximum drawdown. The smart alpha portfolio was inspired by the theory of [21], which documents an alpha momentum phenomenon for country and industry equity indexes in Europe, and it brings the following three advantages: [5]

1. High-beta stocks are overvalued and are thus associated with low realized alpha. This is because a significant proportion of investors are prohibited from using leverage. Smart alpha constructs a low-beta portfolio that capitalizes on this behaviour.

2. Smart alpha not only minimizes the beta of a portfolio, but also maximizes the alpha. This is superior to the BAB rule ([10]), which only selects a low-beta portfolio and allows its alpha to be realized following the negative relationship between alpha and beta. Also, smart alpha considers latent factors rather than a fixed number of empirical factors, allowing the number of factors to change over time as market conditions change.

3. Smart alpha estimates the alpha and beta using a robust approach (SPCA). The sparsity introduces stability into the loadings of the stocks, and consequently into its alphas. It cleans out as much noise as possible.

In this report, we first introduce the methodology of the smart alpha portfolio selection model, including the SPCA algorithm and the smart alpha process. Then, we test the model performance empirically on the data of STOXX600 component stocks and compare the result with that from [5], identifying the similarities and explaining the differences. And finally, we adjust the model parameters and study their impact on the model performance.

2 Methodology

This section describes the smart alpha model strategy. Let ω be a vector with N being the weights of a given portfolio, with N be the number of stocks in the investing universe. Let $\alpha = (\alpha_1, ..., \alpha_N)$ be a vector of length N representing the alpha for each stock in eq(1). Then, the smart alpha model aims to bet on α, while limiting the portfolio's exposure to the systematic sources of risk, which is represented by the $N \times N$ convariance matrix Σ_S. The optimization programme is thus written as follows:

$$\begin{cases} \tilde{\omega} = \arg\min\limits_{\omega} \omega^T \Sigma_S \omega \\ u.c. \;\; \omega^T \alpha \geq \epsilon, \omega \geq 0, \omega \leq \bar{\omega}, \omega^T e = 1, \end{cases} \tag{2}$$

where ϵ is the required α level, $\bar{\omega}$ specifies the upper bound of weights, and e is the unit vector that makes sure the weights to sum up to 1.

It is mentioned in [5] that eq(2) actually maximizes a risk-adjusted measure of performance defined as the ratio of alpha to systematic volatility, i.e., maximizes $\frac{\alpha_p}{\sigma_{S,p}}$. And by setting the risk free rate $r_f = 0$, one could rearrange CAPM into eq(3) and see that maximizing $\frac{\alpha_p}{\sigma_{S,p}}$ is equivalent to maximizing the Treynor Ratio $\frac{\mu_p}{\beta_p}$.

$$\frac{\alpha_p}{\sigma_{S,p}} = \frac{\mu_p - \beta_p \mu_M}{\sqrt{\beta_p^2 \sigma_M^2}} = \frac{\mu_p}{\beta_p \sigma_M} - \frac{\mu_M}{\sigma_M}. \tag{3}$$

The optimization programme specified as eq(2) requires the estimation of two inputs, α and Σ_S, and this is done using the dynamic factor model (DFM) specified as follows:.

$$r_t = \lambda_0 f_t + \lambda_1 f_{t-1} + \cdots + \lambda_s f_{t-s} + e_t, \tag{4}$$

where r_t is the return vector for the N stocks at time t, f_t is a vector of length q representing the dynamic factors, λ_j is a $N \times q$ matrix representing the exposure to the dynamic factors, and e_t is the vector of residual returns. Or, we could rewrite (4) into matrix form:

$$r_t = \Lambda F_t + e_t, \tag{5}$$

where $F_t = (f_t^T, ..., f_{t-s}^T)^T$ is a vector of static factors of length $m = q(1 + s)$, and $\Lambda = (\lambda_0, ..., \lambda_s)$ is a $N \times m$ matrix. The systematic covariance matrix

Σ_S is expressed via the covariance matrix of F:

$$\Sigma_S = \Lambda \Sigma_F \Lambda^T.$$

And α is equal to the expected residual returns e:

$$\alpha = E[e_t] = E[r_t] - \Lambda \bar{F}.$$

However, the choice of F is not obvious. One could not use the empirical factors that are identified in the literature like market, value, growth, etc., because multi-collinearity exists (see [5], table 1,2,3). With multi-collinearity, estimating the beta (Σ_S) has various pitfalls: the estimated coefficient of any one variable depends on which other predictors are included in the model; the estimated coefficients become less precise as more predictors are added; the marginal contribution of any predictor variable to reducing the error sum of squares depends on which other predictors are in the model; the hypothesis test on coefficients may yield different conclusions depending on other predictors. Considering these pitfalls, the factors are instead extracted using the dimension reduction method SPCA. The factors would be orthogonal by construction and thus do not suffer from the above drawbacks. However, since the factors are unobservable, the choice of number of factors, m, is important. In the rest of this section, we first introduce the methodology of SPCA, then present the algorithm that estimates the optimal m, and finally present the algorithm that estimates inputs α and Σ_S.

2.1 PCA & SPCA

The Principal Component Analysis (PCA) could be considered as a fundamental method of matrix decomposition, while the Sparse Principal Component Analysis (SPCA) is a sparse version of PCA. In the smart alpha model, we use the SPCA via Hard Thresholding method introduced by [20] to estimate α and Σ_S. We start by introducing the PCA method.

PCA is based on the factorization of matrix called singular value decomposition (SVD). The SVD of an $m \times n$ data matrix X is a factorization of the form as follows:

$$X = UDV^T, \tag{6}$$

where U is an $m \times m$ unitary matrix representing the set of orthonormal bases $u_1, ..., u_m$, D is an $m \times n$ rectangular diagonal matrix representing the singular values $d_i, ..., d_r, r \leq \min(m, n)$, and V is an $n \times n$ unitary matrix representing the set of orthonormal bases $v_1, ..., v_n$. Thus, one could break down X into r components:

$$X = \sum_{i=1}^{r} d_i u_i v_i^T.$$

And since u_i and v_i are unit vectors, one could ignore terms $d_i u_i v_i^T$ with very small singular value d_i. In other words, one could only select the first $k < r$ terms to represent matrix X. Thus, by reformulating eq(6), we get the principal components (PC) T:

$$T = XV = UD.$$

We select the first k PCs to represent X.

However, the derived PCs are hard to be interpreted ([12],[22]), thus [20] introduces a sparsed PCA which combines the L_1 and L_2 regularizations. It requires only one tuning parameter, the hard threshold ρ. This SPCA algorithm is computationally efficient for large n, which is exactly the case of smart alpha.

SPCA is defined as follows:

Definition 1 (SPCA via Hard Thresholding)**.** *Given a sparse regularization matrix G. For any $\lambda > 0$, let $(\tilde{A}, \tilde{B})$ be the solution of the following:*

$$\begin{cases} \underset{A,B}{\arg\min} \ \sum_{j=1}^{k} ||Xa_j - Xb_j||^2 + \lambda ||b_j||^2 \\ u.c. \ \ A^T A = I, B_{ij} = 0 \ if \ G_{ij} = 0. \end{cases} \tag{7}$$

Then, the sparse loadings are $\tilde{v_j} = \tilde{b_j}/||\tilde{b_j}||, \forall j$.

The regularization matrix G controls the sparsity of loadings V, it is computed via Algorithm 1 using the decision Matrix H:

Definition 2 (Decision Matrix H)**.** *$H \in \{0, 1\}^{p \times p}$ is a decision matrix when $H_{ij} = 1$ if and only if hypothesis $H_{0,ij}$ is rejected. Hypotheses for identifying highly correlated variables are such as:*

$$\begin{aligned} &H_{0,ij} : \textit{variables } i, j \textit{ are not correlated} \\ &H_{1,ij} : \textit{not } H_{0,ij} \end{aligned}$$

i.e.

$$H_{0,ij} : \rho_{ij} = 0$$
$$H_{1,ij} : \rho_{ij} \neq 0$$

$H_{0,ij}$ is rejected if and only if $|\hat{\rho_{ij}}| \geq \rho$ for some decision threshold ρ, $\hat{\rho_{ij}} = \frac{\sum_{l=1}^{n}(x_{li}-\bar{x_i})(x_{lj}-\bar{x_j})}{||x_i-\bar{x_i}||||x_j-\bar{x_j}||}$ is the sample correlation coefficient of variables i and j.

Algorithm 1 Regularization Matrix G

Let α be a permutation of $\{1, ..., p\}$ such that $var(x_{\alpha_1}) \geq \cdots \geq var(x_{\alpha_p})$ and G is empty
while number of columns of $G < k$ **do**
  if $j = 1, ..., p$ and h_{α_j} has not been chosen **then**
    G has new column h_{α_j}
  end if
end while

By [20], eq(7) has an equivalent form and a closed form solution when $n \gg m$:

Theorem 1. *Let D_j be the diagonal matrix such that $[D_j]_{ii} = [G]_{ij}$, then eq(7) is equivalent to*

$$\begin{cases} \underset{A,B}{\arg\min} \ \sum_{j=1}^{k} ||Xa_j - XD_jb_j||^2 + \lambda||b_j||^2 \\ u.c. \quad A^TA = I, \end{cases} \tag{8}$$

with solution

$$b_j = D_jX^TXa_j.$$

Note that the solution of b_j depends on a_j, thus the solution for eq(8) is computed recursively. The solving algorithm is presented in detail in section 2.3.

2.2 Estimating the optimal m

We rewrite eq(5) into matrix form:

$$R = F\Lambda^T + E, \tag{9}$$

where R is the $T \times N$ matrix of returns, F is the $T \times m$ matrix of latent factors, and Λ is the $N \times m$ matrix of loadings.

Let $R = UDV^T$, we have the PCA estimations:

$$\tilde{F} = UD, \frac{1}{T}\tilde{F}^T\tilde{F} = I_m,$$

$$\tilde{\Lambda} = \frac{1}{T}R^T\tilde{F},$$

$$\tilde{E} = R - \tilde{F}\tilde{\Lambda}^T.$$

Under this framework, [2] propose an information criteria that finds a balance between residual variance and complexity:

$$IC(k) = \ln(V(k, \tilde{F}^{(k)})) + k(\frac{N+T}{NT})\ln(C_{TN}^2),$$

where $V(k, \tilde{F}^{(k)}) = \frac{1}{TN}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{E_{it}}^2$ is the average residual variance when the number of factors is set to k, and $C_{TN} = \min(\sqrt{N}, \sqrt{T})$.

The estimated value $\hat{m}$ of the optimal number of latent factors m is thus given by

$$\hat{m} = \underset{k \leq k_{max}}{\arg\min}\ IC(k), \tag{10}$$

where k_{max} is the maximum number of factors. We set k_{max} to 50 when we compute the empirical results. Note that [2] shows $Pr(\hat{m} \to m) \to 1$ as $T, N \to \infty$.

2.3 Estimating α and Σ_S

For $R = F\Lambda^T + E$, we set up the SPCA described as eq(7):

$$\begin{cases} (\tilde{A}, \tilde{B}) = \underset{A,B}{\arg\min}\ \sum_{j=1}^{k} ||Ra_j - Rb_j||^2 + \lambda||b_j||^2 \\ u.c.\ \ A^TA = I_{\hat{m}}, B_{ij} = 0 \text{ if } G_{ij} = 0, \end{cases} \tag{11}$$

where G is a sparse regularization matrix with dimensions $N \times \hat{m}$. Then, by Theorem 1, we have the solution for eq(11):

$$\tilde{b_j} = D_j R^T R\tilde{a}_j.$$

The algorithm to find solutions for eq(11) is given as follows:

Algorithm 2 SPCA

$R = UDV^T$
Let $A = [a_1, ..., a_{\hat{m}}] = V[:, 1 : \hat{m}]$
while not converge **do**
  for $j = 1, ..., \hat{m}$ **do**
    $\tilde{b_j} = D_j R^T R \tilde{a}_j$
  end for
  Update $\tilde{B} = [\tilde{b_1}, ..., \tilde{b}_{\hat{m}}]$
  Compute SVD of $R^T R\tilde{B} = \tilde{U}\tilde{D}\tilde{V}^T$
  Update $A = \tilde{U}\tilde{V}^T$
end while
$\tilde{\Lambda}_{spca} = [\tilde{\lambda}_{1,spca}, ..., \tilde{\lambda}_{\hat{m},spca}]$, where $\tilde{\lambda}_{j,spca} = \frac{\tilde{b_j}}{||\tilde{b_j}||}$.

And we have:

$$\tilde{F}_{spca} = R\tilde{\Lambda}_{spca},$$
$$\tilde{\Sigma}_{S,spca} = \tilde{\Lambda}_{spca}\tilde{\Sigma}_{F,spca}\tilde{\Lambda}^T_{spca},$$
$$\tilde{\alpha}_{spca} = \bar{R} - \bar{\tilde{F}}_{spca}\tilde{\Lambda}^T_{spca},$$

where $\tilde{\Sigma}_{F,spca}$ denotes the covarience matrix of F, $\bar{R}$ is the sample mean of R, and $\bar{\tilde{F}}_{spca}$ is the sample mean of $\tilde{F}_{spca}$.

Substituting $\tilde{\Sigma}_{S,spca}$ and $\tilde{\alpha}_{spca}$ into eq(2), we have

$$\begin{cases} \tilde{\omega} = \arg\min\limits_{\omega} \omega^T \tilde{\Sigma}_{S,spca}\omega \\ u.c. \quad \omega^T \tilde{\alpha}_{spca} \geq \epsilon, \omega \geq 0, \omega \leq \bar{\omega}, \omega^T e = 1. \end{cases}$$

This is a convex quadratic programming problem with $\tilde{\Sigma}_{S,spca}$ positive semidefinite. However, with the constraints on the weights, there is a unique global solution [5], and it is computed via optimizer Gurobi.

3 Empirical Results

In this section, we test the smart alpha model on the data of the European stock market. Similar to [5], we use the components of STOXX600 as our investment universe, but we use a shorter time range from 2015 January 1st to 2021 December 31st. We choose a smaller dataset in this report because firstly, we accessed the data from Yahoo Finance using the tickers scraped from DividendMax. The tickers represent the current components of STOXX600, not historical ones, so choosing a more recent time range would make our investment universe more accurate to STOXX600. And secondly, using a smaller dataset is computational more friendly for a non-professional computer. However, the dataset is sufficient to obtain meaningful results, since the time range of 2015 to 2021 includes the 2018 market crash, the Covid-19 crash, and the bull market of 2020. Also, it covers the tail of 2001 to 2018 that is used in [5], so that the two datasets are comparable.

In the rest of this section, we use a rolling window approach on our dataset to test the performance of the smart alpha portfolio. We study the optimal m across time, compare the model performance with that from the original paper, and study the impact of the two paramters, ϵ and ρ. Note that there are $T = 1764$ daily returns and $N = 526$ stocks in total, we assume 21 trading days a month and 252 trading days a year. The window size is set to $n = 252$, and at each iteration, the estimation window is moving forward one month by including the data for a new month and dropping the data for the earliest month. There are $t = 84$ months in total.

3.1 Optimal m

In figure 1, we plot the estimated optimal number of latent factors, $\hat{m}$, as stated in eq(10). We observe that the optimal m takes values between 2 and 4, with an average number of 2.5. There is a rapid increase from 2 to 4 from the middle of 2017 to the beginning of 2018, which is exactly prior to the 2018 market crush. Then the optimal m immediately dropped back to 3 and stayed there untill 2020, when the market recovered. This result mostly agrees with the conclusion from [5], that the optimal m increases significantly at the begining of crisis periods, and decreases at the exit from them. However, our results suggest a smaller m in general, this might due to the different datasets we used.

The estimated optimal m for each window is brought forward to compute SPCA estimation for α and Σ_S.

3.2 Performance of Smart Alpha Portfolio

In this section, we test the performance of the smart alpha model on the returns of STOXX600 component stocks. We evaluate model performance

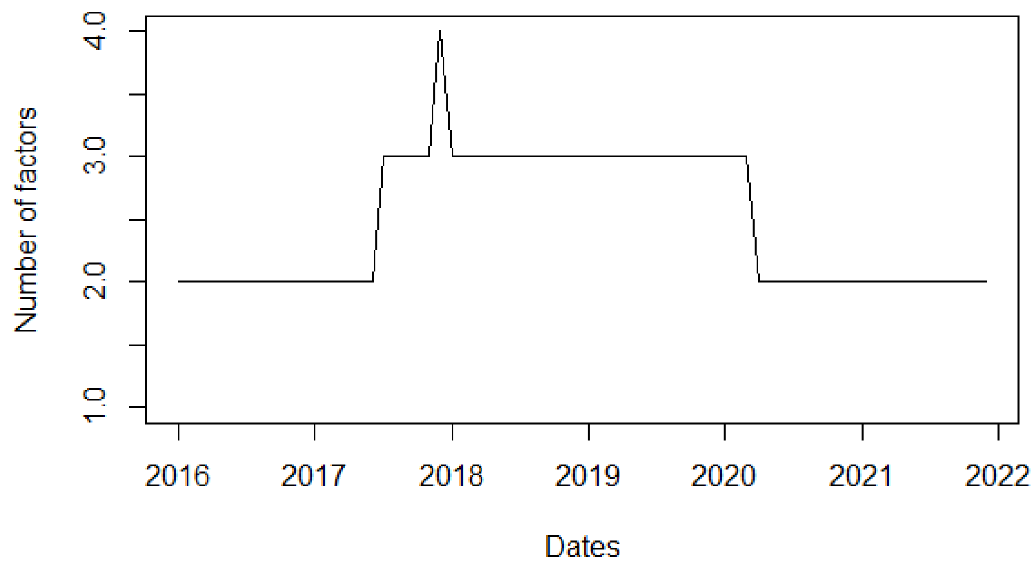

Figure 1: The estimated optimal number of latent factors m

considering its returns, its volatility, its maximum drawdown, its trading volumn, as well as its correlation to the market (STOXX600). The statstics we used are: raw return, annualized average return, annualized volatility, Sharpe ratio, maximum drawdown, monthly turnover (see eq (12), defined in [9]), beta, annualized residual risk, annualized alpha, annualized average excess return, and appraisal ratio. The last 5 statistics are obtained by fitting a linear regression of smart alpha portfolio returns to the returns of STOXX600. Note that we use log returns because log returns are additive across time, and we use a transaction cost of 25 basis points per transaction, same as the original paper. The upper bound $\bar{\omega}$ is set to 0.02, making sure that at least 50 stocks are chosen. Also, we set $\rho = 0.45$ because this gives the best performance (see in section 3.3). While for the required alpha, ϵ, we choose values 0.0005 and 0. This is because $\epsilon = 0$ gives the best model performance (see in section 3.4), but the original paper uses a higher alpha level, so we include $\epsilon = 0.0005$ (quantile 0.7) too as a comparison. Note also that we include smart alpha computed by PCA too as a comparison, and we chose $\epsilon = 0$ and $\epsilon = 0.0005$. Results are presented in table 1.

$$\text{Turnover} = \frac{1}{t}\sum_{i=1}^{t}\sum_{j=1}^{N}|w_{j,i+1} - w_{j,i}| \tag{12}$$

Table 1: Smart Alpha Portfolio Performance

	SPCA, $\epsilon = 0.0005$	SPCA, $\epsilon = 0$	PCA, $\epsilon = 0.0005$	PCA, $\epsilon = 0$
Raw Return (%)	43.92	47.51	39.43	46.70
Annualized Average Return (%)	7.32	7.92	6.57	7.78
Annualized Volatility (%)	1.56	1.57	1.79	1.68
Sharpe Ratio	2.778	3.14	1.993	2.84
Maximum Drawdown (%)	-20.02	-19.95	-23.99	-20.30
Monthly Turnover (%)	27.97	24.28	40.67	38.17
Beta	0.02	0.02	0.12	0.14
Annualized Residual Risk (%)	1.74	1.77	2.29	2.00
Annualized Alpha (%)	4.42	5.00	3.34	4.51
Annualized Average Excess Return (%)	4.48	5.07	3.39	4.57
Appraisal Ratio	1.89	3.18	0.28	0.32

From table 1, we conclude that in general, SPCA models outperform PCA models (this agreess with the original paper), and models with a more 'greedy' choice of alpha fail to bring a high realized alpha and underperform models with required alpha set to 0. The best model is SPCA with $\epsilon = 0$, followed by PCA with $\epsilon = 0$, while PCA with $\epsilon = 0.0005$ gives the worst performance. Note that using a low required alpha level could greatly improve the PCA model performance and make it outperform SPCA with a high ϵ. This means that PCA might be a good choice when the budget is low, because PCA has a much shorter running time compared to SPCA due the optimzing process (SPCA finds a solution recursively while PCA uses a single SVD).

Comparing our best-performed model with the results from [5], we observe that our return is slightly lower but our volatility is also lower, making our Sharpe ratio much higher than that from the original paper, also the maximum drawdown is much lower. So, we can conclude that our portfolio is more stable than that from the original paper, the return is not as high but there is enough compensation for the risk. This result could also been seen from its correlation to the market portfolio: SPCA model has an extremely low correlation (0.02) with the market portfolio compared to 0.45 in the paper, so even though its alpha is low, it has a high appraisal ratio. This result is mainly due to the difference in the optimal m we used. The original paper uses a higher m (average 4) while we use a lower m (average 2.5). Using fewer factors creates a more diversified portfolio (the highest weight is 0.008 and the lowest is 0.0001) and thus the portfolio is more stable and is affected less by the market. Another reason is that the original paper considers a much longer time interval including several huge economic crises, while our dataset includes only the 2018 market crash. Another difference is that our portfolios have a higher monthly turnover than those from the original paper, meaning that our portfolios trade more often. And the difference between PCA and SPCA portfolios are much larger. This might also due to the diversity of our portfolios: if one considers more stocks while rebalancing, the total amount of weight change is larger.

In general, our conclusions agree with those from the original paper: the smart alpha portfolio provides a high return with a low risk, resulting in a high Sharpe ratio. It is lowly correlated with the market thus generates a high alpha and high appraisal ratio and is not affected much by economic crises (low maximum drawdown). Also, sparsity is necessary since it makes the portfolio more stable and it decreases the trading volumn thus decreases transaction costs. Although we could not compare our results with the smart beta portfolios or BAB as the original paper does, since we do not have access to Bloomberg, the conclusion is not affected, because our results are an enhanced version of the original paper and they provide more evidence for the main conclusion.

3.3 Impact of sparsity

In this section, we study the impact of sparsity on model performance and running time. [5] compares SPCA model with PCA model and concludes that sparisty does improve model performance. We dig more into this by comparing SPCA models with different levels of sparsity.

Sparsity is controlled by the regularization matrix G stated as in Algorithm 1. By varying the threshold ρ, we get different levels of sparse G, and in table 2, we list the number of non-zero rows in G, i.e., the total number of variables used in the SPCA process. From it, we conclude that at $\rho \geq 0.5$, G is too sparse: all variables are considered as independent. As we decrease the threshold ρ, more variables are included, and when ρ drop to 0.2, almost every variable is used in the SPCA process. Note also that m in table 2 represents the number of columns in G. Comparing $m = 2$ and $m = 3$, we conclude that the variable with smaller variance (3rd variable) brings fewer correlated variables into the information pool.

Table 2: Impact of ρ on the Sparisty level

ρ	0.5	0.45	0.4	0.35	0.3	0.25	0.2
$m = 2$	2	73	73	328	429	458	515
$m = 3$	3	73	73	450	503	520	521

We then fix the required alpha $\epsilon = 0.0005$ to study the impact of sparsity. We vary ρ from 0.3 to 0.5 to each get a smart alpha portfolio, then we compare their perfomance and running time. The results are shown in table 3. Note that we include the portfolio computed by traditional PCA as a comparison.

Table 3: Impact of sparsity on model performance

ρ	0.5	0.45	0.4	0.35	0.3	0.25	0.2	PCA
Raw Return (%)	43.22	43.92	42.98	41.36	39.42	40.17	41.35	39.43
Annualized Average Return (%)	7.20	7.32	7.16	6.89	6.57	6.70	6.89	6.57
Annualized Volatility (%)	1.57	1.56	1.57	1.63	1.61	1.61	1.62	1.79
Sharpe Ratio	2.676	2.778	2.645	2.392	2.212	2.289	2.395	1.993
Maximum Drawdown (%)	-20.53	-20.02	-20.46	-22.09	-21.52	-21.75	-21.90	-23.99
Beta	0.025	0.023	0.020	0.021	0.025	0.018	0.014	0.119
Annualized Residual Risk (%)	1.78	1.74	1.78	1.91	1.88	1.88	1.90	2.29
Annualized Alpha (%)	4.30	4.42	4.25	3.99	3.67	3.78	3.97	3.34
Annualized Average Excess Return (%)	4.36	4.48	4.31	4.04	3.72	3.83	4.02	3.39
Appraisal Ratio	1.748	1.888	2.162	1.892	1.480	2.106	2.941	0.281
Running Time (minutes)	5.404	6.445	6.481	7.333	7.548	8.043	8.131	0.486

We conclude that sparsity greatly affects model performance and running time. In general, a more sparse model ($\rho \geq 0.4$) outperforms less sparse models ($\rho \leq 0.35$ and PCA): it has higher returns, lower volatility, higher alpha, and smaller maximun drawdown. The reason is explained by [5]: sparsity adds a de-noising process which can produce stable exposures and alphas for the stocks, also it requires less portfolio rebalancing and lower transaction costs. Another advantage of high sparsity is that it takes less running time because it is easier to converge when solving SPCA. However, the level of sparsity should be chosen carefully. If choosing the 'wrong' threshold ρ, for example, 0.3, one could result in a portfolio that is nearly as bad as non-sparse PCA portfolio, but PCA has a great advantage in running time due to its optimization process.

Note that an interesting discovery is that although non-sparse PCA model is more correlated to the market, an extremely sparse SPCA model could be more correlated to the market compared to a less sparse one. The reason might be that an extremely sparse portfolio is more concentrated, thus is more affected by the market.

So, in conclusion, we choose $\rho = 0.45$ as our best performing portfolio on the dataset of STOXX600. It greatly outperforms all other portfolios except for appraisal ratio.

3.4 Impact of level of required alpha

In this section, we study the impact of the level of required alpha to model performance. [5] choose a required alpha level which equals to a high quantile of the stocks' alphas. This is because instead of relying on the negative relation between alphas and betas, smart alpha model also maximizes alpha by setting a high level of required alpha. So, we list the different quantiles of the stocks' alpha in table 4 and conclude that a 'high quantile' of alpha could go from 0.0005 to 0.001. We then set $\rho = 0.45$ and choose ϵ from 0, to 0.001 to study its impact on model performance. The results are presented in table 5.

Table 4: Quantile of stocks' alpha

Quantile	Alpha
0.95	0.00106
0.9	0.000935
0.85	0.000743
0.7	0.00054
0.5	0.000355

Table 5: Impact of Required Alpha on model performance

ϵ	0.001	0.00054	0.00035	0.0001	0	PCA
Raw Return (%)	41.84	44.01	46.06	46.98	47.51	46.70
Annualized Average Return (%)	6.97	7.33	7.68	7.83	7.92	7.78
Annualized Volatility (%)	1.58	1.56	1.56	1.57	1.57	1.68
Sharpe Ratio	2.51	2.79	3.01	3.08	3.14	2.84
Maximum Drawdown	-20.56	-20.04	-19.98	-19.96	-19.95	-20.30
Beta	0.02	0.02	0.02	0.02	0.02	0.14
Annualized Residual Risk (%)	1.80	1.74	1.74	1.77	1.77	2.00
Annualized Alpha (%)	4.06	4.43	4.77	4.92	5.00	4.51
Annualized Average Excess Return (%)	4.12	4.50	4.84	4.99	5.07	4.57
Appraisal Ratio	2.00	1.84	2.24	2.65	3.18	0.32

We observe that the model performance is improved as ϵ decreases, with a higher return, higher alpha, lower risk and lower maximum drawdown. Also, we put the result of a PCA portfolio using $\epsilon = 0$ here as a comparison, and one could conclude that the PCA's problem of unstableness could be greatly solved by leaving alpha unbounded, since the PCA portfolio totally outperforms the $\epsilon = 0.001$ SPCA portfolio.

This is an interesting discovery that partially violates the conclusions of [5]. The results in table 5 suggest that maybe the best way to maximize alpha is to minimize beta and rely on the negative relationship between alpha and beta.

4 Conclusion

In conclusion, this report studies the smart alpha portfolio selection model introduced by [5]. It presents the methodology, tests model performance empirically on the data of STOXX600, and studies the impact of model parameters. The conclusion generated by this report in general agrees with that from [5], that the smart alpha portfolio generates high stable returns with a low risk and small extreme losses, but it does question the model's effect of maximizing alpha. However, the smart alpha portfolio still outperforms most of the portfolios in the market, and it shows a strong competitiveness by provding an excellent Sharpe ratio and appraisal ratio.

References

[1] Amihud, Y., Illiquidity and stock returns: Cross-section and time-series effects. J. Financ. Mark., 2002, 5(1), 31–56.

[2] Bai, J. and Ng, S., Determining the number of factors in approximate factor models. Econometrica, 2002, 70(1), 191–221.

[3] Behr, P., Guettler, A. and Truebenbach, F., Using industry momentum to improve portfolio performance. J. Bank. Financ., 2012, 36(5), 1414–1423.

[4] Bloomfield, T., Leftwich, R. and Long, J.B., Portfolio strategies and performance. J. Financ. Con., 1977, 5(2), 201–218.

[5] Boucher, C. and Jasinski, A. and Kouontchou, P. and Tokpavi, S., Smart alpha: Active management with unstable and latent factors. Quantitative Finance 2021; 893–909.

[6] Cochrane, J,H., Presidential address: Discount rates. J. Finance, 2011, 66(4), 1047-1108.

[7] de Bondt, W.F.M. and Thaler, R.H., Does the stock market overreact?. J. Finance, 1985, 40(3), 793–805.

[8] de Bondt, W.F.M. and Thaler, R.H., Further evidence on investor overreaction and stock market seasonality. J. Finance, 1987, 42(3), 557–581

[9] DeMiguel, V., Garlappi, L. and Uppal, R., Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? Rev. Financ. Stud., 2009, 22, 1915–1953.

[10] Frazzini, A. and Pedersen, L.H., Betting against beta. J. Financ. Econ., 2014, 111(1), 1–25.

[11] Gennaioli, N. and Shleifer, A., A Crisis of Beliefs: Investor Psychology and Financial Fragility, 2018 (Princeton University Press: Princeton, NJ).

[12] Jolliffe, I. T., Trendafilov, N. T., and Uddin, M., "A modified principal component technique based on the lasso," Journal of Computational and Graphical Statistics, vol. 12, no. 3, pp. 531–547, 2003.

[13] Jegadeesh, N. and Titman, S., Returns to buying winners and selling losers: Implications for stock market efficiency. J. Finance, 1993, 48(1), 65–91.

[14] Kelly, T.B., Pruitt, S. and Su, Y., Characteristics are covariances: A unified model of risk and return. J. Financ. Econ., 2019, 134(3), 501–524.

[15] Kourtis, A., Dotsis, G. and Markellos, R.N., Parameter uncertainty in portfolio selection: Shrinking the inverse covariance matrix. J. Banking Finance, 2012, 36(9), 2522–2531.

[16] Lee, C.M.C., Shleifer, A. and Thaler, R.H., Investor sentiment and the closed end fund puzzle. J. Finance, 1991, 46(1), 75–109.

[17] Lewellen, J., Nagel, S. and Shanken, J., A skeptical appraisal of asset pricing tests. J. Finance. Econ,. 2010, 96(2), 175-194.

[18] Shiller, R.J., Narrative economics. Am. Econ. Rev., 2017, 107(4),967–1004.

[19] Tu, J. and Zhou, G., Markowitz meets talmud: A combination of sophisticated and naïve diversification strategies. J. Financ. Econ., 2011, 99(1), 204–215.

[20] Wu, M.-C. and Chen, K.-C., Sparse PCA via hard thresholding for blind source separation. IEEE International Conference on Acoustics, Speech and Signal Processing, 2016.

[21] Zaremba, A., Umutlu, M. and Karathanasopoulos, A., Alpha momentum and alpha reversal in country and industry equity indexes. J. Empir. Financ., 2019, 53(C), 144–161.

[22] Zou, H., Hastie, T., Tibshirani, R., "Sparse principal component analysis," Journal of Computational and Graphical Statistics, vol. 15, no. 2, pp. 265–286, 2006.