Portfolio Optimization Based on Barra Equity Risk Model

Group members: Cang Yimeng, Ding Yuan, Liu Yang, Meng Yan, Wang Yining, Xiao Yunhan

Abstract

The objective of a quantitative hedging strategy is to achieve robust absolute returns. The optimal portfolio is constructed in perfect equilibrium, where the investment manager fully exposes the portfolio to alpha factors while eliminating other destabilizing style factor distractions. In a multi-factor model, the equity portfolio's weighting is critical in determining the robustness of the strategy's returns. Therefore, portfolio weight optimization plays a crucial role in pursuing return stability for multi-factor models.

Regarding specific research ideas, we aim to use the Barra risk model to predict the volatility of the equity portfolio. The Barra risk model is a multi-factor model developed by Barra Inc. that measures the overall risk associated with security relative to the market. The model incorporates over 40 data metrics, including earnings growth, share turnover, and senior debt rating. It then measures risk factors associated with three main components: industry risk, exposure to different investment themes, and company-specific risk. Finally, we will calculate the weight optimization of the stock portfolio based on the predicted portfolio risk to obtain the optimal portfolio under various risk constraints.

1. Introduction

As a classic model in the field of quantitative investment research, multi-factor stock selection has been widely studied and practically applied by various investment institutions at home and abroad. In China, since the launch of the CSI 300 stock index futures in 2010, alpha hedging strategies represented by multi-factor stock selection have gradually entered the public's view. However, in the market situation of December 2014, the alpha hedging strategy encountered a major setback. The reason for this is not difficult to find. The combination with too obvious market value style characteristics is the main reason for the large fluctuation of strategy income.

This report is different from traditional multi-factor research. We did not focus on the mining of alpha factors, but focused on the impact of weight optimization of stock portfolios on the strategy style characteristics. Through the calculation of the weight optimization of the stock portfolio, we found the optimal investment portfolio under market value neutrality, industry neutrality, and style factor neutrality constraints, and obtained stable excess returns.

In the multi-factor model, the key step that determines the robustness of the strategy's returns is the weight allocation of the stock portfolio. Therefore, from the perspective of quantitative hedging strategies pursuing income stability, portfolio weight optimization plays a crucial role in multi-factor models. In terms of specific research ideas, we started from the perspective of structured multi-factor risk models, used the BARRA risk factor effectiveness verification method, and constructed a structured multi-factor risk model based on 31 industry factors and 10 style factors, laying the foundation for predicting the volatility of stock portfolios. Afterwards, we examined the strength of various factor alpha properties through the study of pure factor stock portfolios and explained the economic and financial logic behind the factors.

Finally, through the calculation of the weight optimization of the stock portfolio, we obtained the optimal investment portfolio under market value neutrality, industry neutrality, and style neutrality constraints. Empirical verification shows that the strategy's return stability is significantly improved by constraining the style characteristics of the stock portfolio. The attribution analysis results also show that the combination of style-neutral configuration has a decisive impact on the strategy's return characteristics.

2.1. Structured multi-factor risk model

The stock prices of any stock are exposed to various risk factors at the same time, and their combined effect forms the fluctuation of the stock price. In order to quantitatively study the role of various risk factors, quantitative risk models have emerged.

The significance of the risk model is to find the cause of the fluctuation of stock prices, decompose and strip the source of stock returns, and realize the prediction of future stock price fluctuations. Let's start with the structured multi-factor risk model to explore the overall risk structure of the A-share market.

Before the structured multi-factor risk model, there were three basic risk models: The first basic risk model is the Markowitz portfolio variance. In Markowitz's mean-variance theory, the risk calculation of an investment portfolio requires estimating the volatility of each asset in the portfolio and their correlation coefficients. Generally, when there are N stocks in a portfolio, the number of volatilities that need to be estimated is N, and the number of correlation coefficients that need to be estimated is $\frac{N(N-1)}{2}$. We can summarize the parameters that need to be considered into a covariance matrix V:

$$V = \begin{bmatrix} \sigma_1^2 & \sigma_{1\,2} & \cdots & \sigma_{1\,N} \\ \sigma_{1\,2} & \sigma_2^2 & \cdots & \sigma_{2\,N} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{1\,N} & \cdots & \cdots & \sigma_N^2 \end{bmatrix}$$

where σ_{ij} represents the covariance of r_i and r_j, where $\sigma_{ij} = \sigma_{ji}$. The covariance matrix contains all the elements required to calculate portfolio risk. However, the disadvantage of this method is that the covariance matrix contains too many independent parameters, making it difficult to predict the covariance matrix accurately and efficiently.

The second basic risk model requires estimation of the volatility σ_n for each stock, as well as the average correlation coefficient ρbetween stocks. This means that the covariance between any two stocks is

$$Cov(r_n, r_m) = \sigma_n \cdot \sigma_m \cdot \rho$$

The biggest advantage of this model is its simplicity. However, this model ignores subtle connections between stocks that are similar in industry or have similar attributes.

The third basic risk model relies on sample variance and sample covariance of historical data. This type of model uses T periods of samples to estimate an $N \times N$ covariance matrix, and requires $\mathrm{T} \geq \mathrm{N} + 1$, which means that it is obviously unrealistic to estimate the covariance matrix of daily returns for all A-shares in application.

To overcome and improve the shortcomings of the above three basic risk models, structured multi-factor risk models have emerged. Structured risk factor models use a set of common factors and a specific factor that is only related to the stock to explain the stock's return, and use the volatility of common factors and specific factors to explain the volatility of stock returns. The advantage of structured multi-factor risk models is that by identifying important factors, the scale of the problem can be reduced, and even if the number of stock combinations changes, the complexity of problem solving will not change as long as the number of factors remains constant.

Structured multi-factor risk models first perform a simple linear decomposition of returns, which includes four components: stock returns, factor exposure, factor returns, and specific factor returns. Then, the linear decomposition of the j-th stock is shown as follows:

$$r_j = x_1 f_1 + x_2 f_2 + x_3 f_3 + x_4 f_4 \dots x_k f_k + u_j$$

where r_j represents the return of the j-th stock; x_k represents the exposure of the j-th stock on the k-th factor (also known as factor loading); f_k represents the factor return of the k-th factor of the j-th stock (i.e., the return carried by each unit of factor exposure); u_j represents the specific factor return of the j-th stock.

For the time structure of the above equation, if we define that factor exposure is a result at time t, then stock returns, factor returns, and specific factor returns are all results at t+1. In the model, we process cross-sectional data at a daily frequency.

So for an investment portfolio containing N stocks, assuming that the weight of the portfolio is $w = (w_1, w_2, \dots, w_N)^T$, then portfolio returns can be expressed as:

$$R_P = \sum_{j=1}^{N} w_n \cdot \left(\sum_{k=1}^{K} x_{jk} f_{jk} + u_j \right)$$

Now, we assume that the specific factor returns of each stock are uncorrelated with the common factor returns and that the specific factor returns of each stock are also uncorrelated. Then, based on the above expression, we can obtain the risk structure of the portfolio as follows:

$$\sigma_P = \sqrt{w^T (XFX^T + \Delta) w}$$

where X represents the factor loading matrix ($N \times K$) of N stocks on K risk factors:

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,k} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,k} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,k} \end{bmatrix}$$

F represents the covariance matrix ($K \times K$) of factor returns for K factors:

$$F = \begin{bmatrix} Var(f_1) & Cov(f_1, f_2) & \cdots & Cov(f_1, f_k) \\ Cov(f_1, f_2) & Var(f_2) & \cdots & Cov(f_2, f_k) \\ \cdots & \cdots & \cdots & \cdots \\ Cov(f_k, f_1) & Cov(f_k, f_2) & \cdots & Var(f_k) \end{bmatrix}$$

The volatility and covariance of factor returns are estimated using daily frequency data of factor returns.

Δ represents the covariance matrix of stock-specific factor returns for N stocks ($N \times N$):

$$\Delta = \begin{bmatrix} Var(u_1) & 0 & \cdots & 0 \\ 0 & Var(u_2) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & Var(u_k) \end{bmatrix}$$

Since we assume that the correlation of stock-specific factor returns is 0, Δ is a diagonal matrix.

The significance of a structured multi-factor risk model for investors lies not in minimizing portfolio risk, but in assigning optimal risk weights given a portfolio's expected return. We will discuss this further in our discussion on portfolio weight optimization.

2.2. Selection of risk factors

Sample size

Unlike most mature overseas markets, the stock pool with the largest market capitalization and best liquidity in the A-share market cannot fully reflect all the characteristics of the A-share market. Therefore, we select all A-share targets as the sample space for risk factor testing.

Industry factors

Now let's discuss the common factor part of the risk model. In BARRA's structured multi-factor risk model, common factors are usually divided into two parts: industry factors and style factors. Industry factors are an important part of the risk model, reflecting the unique characteristics of individual stocks by dividing all A-share stocks by industry. Our risk model includes 31 industry factors, specifically:

Food and Beverages	Agriculture, Forestry, Animal Husbandry, and Fishing	Household Appliances	Coal
Oil and Petrochemicals	Computers	Basic Chemicals	Telecommunications
Utilities	Banking	Textiles and Apparel	Non-Banking Financial Institutions
Nonferrous Metals	Media	Architectural Decoration	Environmental Protection
Power Equipment	Light Manufacturing	Building Materials	Defense and Military Industry
Transportation	Beauty and Personal Care	Automobiles	Conglomerates
Pharmaceuticals and Biotechnology	Social Services	Machinery and Equipment	Real Estate
Trade and Retail	Steel	Electronics	

Style factors

Style factors are another important part of the common factors. Style factors will play a crucial role in our discussion of pure factor combinations and stock portfolio weight optimization later. Style factors include a total of 9 major factors and 19 minor factors, including Beta, Momentum, Size, Earnings Yield, Volatility, Growth, Value, Leverage and Liquidity, specifically:

Major Factors	**Minor Factors**	**Calculation Methods**

Beta	BETA	$r_i = \alpha + \beta r_m + e_i$; Use individual stock return series and Shanghai and Shenzhen 300 index return series for univariate linear regression. The length of the yield sequence is 252 trading days. Both stock return series and Shanghai and Shenzhen 300 index return series are exponentially weighted with a half-life of 60 days.
Momentum	RSTR	$RSTR = \sum_{t=L}^{T+L} w_t[ln(1 + r_t)]$; where T = 500, L = 21, return series are exponentially weighted with a half-life of 120 days.
Size	LNCAP	$LNCAP = LN\ (total\_market\_capitalization)$; the natural logarithm on total market value of individual stocks.
Earnings Yield	EPIBS	$EPITBS = est\_eps\ /\ P$; where est_eps is consensus expected basic earnings per share for an individual stock.
	ETOP	$ETOP = earnings\_ttm\ /\ mkt\_freeshares$; historical earnings per share (EPS), calculated by dividing the net profit of an individual stock in the past 12 months by its current market value.
	CETOP	$CETOP = Cash\_earnings\ /\ P$; the individual stock's cash yield ratio to its stock price.
Volatility	DASTD	$DASTD = \sqrt{\left(\sum_{t=1}^{T} w_t \cdot \left(r_t - \mu(r)\right)^2\right)}$ The length of the yield sequence is 252 trading days and the return series are exponentially weighted with a half-life of 60 days.
	CMRA	$CMRA = ln(1 + max\, Z(T)) - ln(1 + min\, Z(T))$; where $Z(T) = \sum_{\tau=1}^{T}[ln(1 + r_\tau)]$; r_τ means the daily return of the individual stock and over the past 12 months (T) is represented.
	HSIGMA	$HSIGMA = std(e_i)$; where the residual e_i is calculated by BETA.
Growth	SGRO	Compound annual growth rate (CAGR) of total operating revenue of the company over the past 5 years.
	EGRO	Compound annual growth rate (CAGR) of net profit attributable to the parent company of the company over the past 5 years.
	EGIB_S	Consensus expected net profit growth rate of the company for the next 1 year.
Value	BTOP	$BTOP = common\_equity\ /\ current\_market\_capitalization$; calculate the total equity value of the company divided by its current market value.
Leverage	MLEV	$MLEV = (ME + LD)\ /\ ME$; where ME represents the current total market value of the company, and LD represents the long-term debt of the company.
	DTOA	$DTOA = TD\ /\ TA$; where TD represents total liabilities and TA means total assets.

	BLEV	$BLEV = (BE + LD) / BE$; where BE means book value of equity and LD represents the long-term debt of the company.
Liquidity	STOM	$\text{STOM} = \ln\left(\sum_{t=1}^{21}\{(V_t / S_t)\}\right)$; where V_t means trading volume on the day, and S_t represents the outstanding shares (i.e., shares in circulation) of the company.
	STOQ	$STOQ = ln\,(\frac{1}{T}\sum_{\tau=1}^{T} exp(STOM_\tau))$; where T = 3.
	STOA	$STOA = ln\,(\frac{1}{T}\sum_{\tau=1}^{T} exp(STOM_\tau))$; where T = 12.

The style factors of Earnings Yield, Volatility, Growth, Leverage, and Liquidity are constructed by combining several sub-factors. We obtain the weights of the sub-factors by using the method of single-factor regression within the sample to minimize the residual sum of squares.

Winsorization and Standardization of Factors

In order to standardize the regression equation, it is usually necessary to standardize the factors, including trimming outliers and normalization. Traditional methods of trimming outliers include the three times standard deviation method and the median method. In the BARRA risk model, the three times standard deviation method is widely used, but we found that this method can still produce unstable regression equations when dealing with skewed factor distributions. Therefore, we use the median method to trim outliers.

The normalization process includes methods such as the standard normal method, weighted standard normal method, Box-Cox transformation, and Johnson transformation. In the BARRA risk model, the standard normal method weighted by the square root of market capitalization is used to some extent to eliminate the influence of market capitalization. However, this method may cause the mean of the standardized factor to be unequal to zero in the cross-section, which can lead to some bias in the risk exposure setting for the optimization of the stock portfolio weights. Considering this point, we only use the simple standard normal method, that is:

$$d_{nl} = \frac{d_{nl}^{raw} - u_l}{\sigma_l}$$

where d_{nl} represents the standardized factor sequence, d_{nl}^{raw} represents the original factor sequence, u_l represents the arithmetic mean of d_{nl}^{raw}, and σ_l represents the standard deviation of d_{nl}^{raw}.

Parameters approximation

The structured risk model provides a linear decomposition of the return on any given stock as follows:

$$r_j = x_1 f_1 + x_2 f_2 + x_3 f_3 + x_4 f_4 \ldots x_k f_k + u_j$$

For a portfolio of N stocks, the vector of portfolio returns can be written as:

$$R = Xf + U$$

where *X* is the loading matrix of all factors, consisting of the industry factor dummy variable matrix and the common factor loading matrix, f is the vector of factor returns, and U is the residual sequence. For the structured risk model, given the stock returns *R* and the factor loading matrix *X*, the vector of factor returns *f* needs to be estimated.

According to the ordinary least squares (OLS) method, we need to find the vector of returns *f* that minimizes the sum of squared residuals, i.e.,

$$\begin{aligned} Min\ Q &= \sum_{i=1}^{N} \epsilon_i^2 = \sum_{i=1}^{N} (r_i - \hat{r}_i)^2 \\ &= (R - X\hat{\mathrm{f}})'(R - X\hat{\mathrm{f}}) \\ &= (R'R - R'X\hat{f} - \hat{f}'X'R' + \hat{f}'X'X\hat{f}\ ) \\ &= R'R - 2\hat{f}'X'Y + \hat{f}'X'X\hat{f} \end{aligned}$$

Let $\frac{\partial Q}{\partial \hat{f}} = 0$, we get $-X'R + X'X\hat{f} = 0$, so that $\hat{f} = (X'X)^{-1}X'R$.

The OLS method of estimation provides the optimal estimate for vector $\hat{f}$ only when the variances of the residual sequences ε_{it} from different stocks are the same. However, in most cases, financial time series data exhibit obvious heteroscedasticity, where the variances of ε_{it} for each stock are not the same. To address heteroscedasticity, the Generalized Least Squares (GLS) method is commonly used. We provide a brief derivation of the GLS method below.

The GLS method assumes that the variances of ε_{it} are not the same, i.e.,

$$Var(U) = \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{n1} & \cdots & \cdots & \sigma_n^2 \end{bmatrix}$$

Since the matrix Σ is positive definite, it can be written as $\Sigma = \mathrm{K}\,\mathrm{K}'$, where *K* is a non-singular matrix.

For $R = Xf + U$, we can get

$$K^{-1}R = K^{-1}Xf + K^{-1}U$$

Setting $R^* = K^{-1}R, X^* = K^{-1}X, U^* = K^{-1}U,$ we then get

$$R^* = X^*f + U^*$$

We know that $E(U^*) = 0$, and

$$\begin{aligned} Var(U^*) &= Var(K^{-1}U) \\ &= K^{-1}Var(U)(K^{-1}) \\ &= K^{-1}KK'(K')^{-1} \\ &= I \end{aligned}$$

Hence we could conclude that $R^* = X^*f + U^*$ satisfies the homoscedasticity condition for OLS estimation, and based on the estimation result for vector f, we can ensure

$$\hat{f}_{GLS} = \left((X^*)'X^*\right)^{-1}(X^*)'Y^*$$

$$= \left((K^{-1}X)'K^{-1}X\right)^{-1}(K^{-1}X)'(K^{-1}R)$$

$$= (X'(K^{-1})'K^{-1}X)^{-1}X'(K^{-1})'(K^{-1}R)$$

Since $(K^{-1})'K^{-1} = (K')^{-1}K^{-1} = (KK')^{-1} = \Sigma^{-1}$,

$$\hat{f}_{GLS} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}R$$

Under the condition of known residual volatility matrix Σ, the Generalized Least Squares (GLS) estimation method can provide an unbiased estimator for vector $\hat{f}_{GLS}$.

In the structured risk model, where there is no correlation assumed among the residuals, the special form of GLS method, Weighted Least Squares (WLS), can be used to handle the heteroscedasticity.

WLS assumes that the variances of ε_{it} are not the same, but the covariance between the ε_{it} is zero, i.e.,

$$Var(U) = \Sigma = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \dots & \\ & & & \sigma_n^2 \end{bmatrix}$$

Rewrite Σ into

$$\Sigma = \sigma_w^2 \begin{bmatrix} 1/w_1 & & & \\ & 1/w_2 & & \\ & & \dots & \\ & & & 1/w_n \end{bmatrix}$$

Making $W = diag(w_1, w_2, \dots, w_n)$, then $\Sigma = \sigma_w^2 W^{-1}, \Sigma^{-1} = (1/\sigma_w^2)W$. From $\hat{f}_{GLS}$ in the estimation representation of *GLS*, we obtain

$$\hat{f}_{GLS} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}R$$

$$= \sigma_w^2(X'WX)^{-1}X'(1/\sigma_w^2)WR$$

$$= (X'WX)^{-1}X'WR$$

Test the effectiveness of factors

The traditional multi-factor model often uses the correlation coefficient IC between the vector of factor cross-sectional sequences and the vector of cross-sectional excess returns to test the effectiveness of the factors. It is commonly assumed that if the absolute mean value of the IC sequence exceeds a certain threshold, then the factor is considered an alpha source. However, this method has two flaws. First, the absolute mean value of the IC sequence alone cannot determine the stability of the factor. For example, a factor that has both positive and negative correlations may be identified as an alpha factor based on the absolute value of the IC sequence, but it may not have a stable directional impact on returns, leading to significant uncertainty in the model. Second, the IC method alone cannot test the issue of collinearity between factors, which leaves uncertainty in the subsequent factor weighting for dealing with collinearity.

BARRA provides a complete set of criteria for testing the effectiveness of factors, starting from the significance and stability of the factor's impact on returns and the issue of collinearity between factors. The criteria include:

1. The absolute mean value of the T-test statistic in the single-factor regression equation, which is usually considered ideal if it is greater than 2, indicating a higher level of significance of the factor's impact on returns;
2. The proportion of the absolute value of the T-test statistic in the single-factor regression equation that exceeds 2, which explains the distribution characteristics of the significance level of the factor during the testing period;
3. The annualized factor return, which indicates the degree of contribution of the factor to returns, and is annualized for easy comparison with the strategy's annualized return;
4. The annualized volatility of the factor return, which indicates the degree of volatility of the factor's contribution to returns, and is annualized for easy comparison with the volatility of the strategy's return;
5. The ratio of the factor return to the factor return volatility, which measures the stability of the factor's return after adjusting for volatility, and is an indicator of the factor's stability;
6. The correlation between the factor return and the benchmark return, which tests whether the factor return is highly correlated with the benchmark return. Ideally, the lower the correlation, the better the result (usually the return of the CSI 300 index is used as a benchmark for comparison);

Based on the above testing criteria, we conducted an effectiveness test on the industry and style factors. In the parameter estimation of the regression equation, we set $W = diag(w_1, w_2, \dots, w_n)$, where w_i is the square root of the market capitalization of the i-th stock. In other words, we used the derivative of the square root of the market capitalization of individual stocks as the weighting for *WLS* estimation.

The results of the factor effectiveness test are shown in the following table:

<u>Industry Factors Effectiveness Test</u>

Factor Name	Average Absolute t-stat	Percent Ovserv $\|t\| > 2$	Annual Factor Return	Annual Factor Volatility	Factor Return Sharp ratio
Food and Beverages	4.9293	0.3896	0.6331	0.7129	0.888
Agriculture, Forestry, Animal Husbandry, and Fishing	5.3791	0.3991	0.5504	0.5826	0.9447
Household Appliances	4.2388	0.3389	0.2239	0.4525	0.4948

Coal	3.3319	0.3148	0.3079	0.5610	0.5488
Oil and Petrochemicals	3.8014	0.3519	0.6647	0.6491	1.024
Computers	6.5858	0.4152	0.5382	0.5817	0.9252
Basic Chemicals	7.0747	0.4618	0.5147	0.5908	0.8712
Telecommunications	5.1917	0.4106	0.6832	0.5883	1.1613
Utilities	5.2414	0.3830	0.2537	0.4918	0.516
Banking	1.2192	0.1230	2.2090	1.0905	2.0256
Textiles and Apparel	5.2206	0.4011	0.4819	0.6422	0.7505
Non-Banking Financial Institutions	3.3492	0.2987	0.5023	0.5388	0.9322
Nonferrous Metals	5.2877	0.4177	0.7146	0.6998	1.0211
Media	5.4748	0.3835	0.2228	0.4481	0.4973
Architectural Decoration	5.4879	0.4202	0.6048	0.5680	1.0649
Environmental Protection	4.9210	0.4021	0.5135	0.4262	1.2048
Steel	2.8955	0.2716	0.3499	0.4906	0.7131
Power Equipment	6.7362	0.4372	0.4998	0.5623	0.8889
Light Manufacturing	5.1470	0.4081	0.5076	0.5265	0.9642
Building Materials	4.4948	0.3690	0.4757	0.4962	0.9587
Defense and Military Industry	4.7004	0.3820	0.6672	0.6585	1.0131
Transportation	5.0421	0.3966	0.4889	0.6588	0.7421
Beauty and Personal Care	1.6311	0.1611	0.5802	0.1774	3.2715
Automobiles	6.4557	0.4488	0.5889	0.5547	1.0617

Electronics	6.7032	0.4493	0.7098	0.5740	1.2365
Conglomerates	3.5459	0.3253	0.4206	0.5076	0.8286
Pharmaceuticals and Biotechnology	7.8003	0.4272	0.3089	0.4667	0.6619
Social Services	4.5639	0.3730	0.4371	0.5134	0.8514
Machinery and Equipment	7.4800	0.4428	0.3936	0.5690	0.6917
Real Estate	5.2943	0.3805	0.3402	0.6110	0.5567
Trade and Retail	5.3079	0.3780	0.2109	0.4993	0.4223
Average	**4.9849**	**0.3730**	/	/	/

Style Factors Effectiveness Test

Factor Name	**Average Absolute t-stat**	**Percent Ovserv |t| > 2**	**Annual Factor Return**	**Annual Factor Volatility**	**Factor Return Sharp ratio**	**Correl With HS300**
BETA	10.2309	0.6014	0.6114	0.2421	2.5251	-0.0878
Momentum	3.6744	0.2410	-0.1513	0.0308	-4.9113	0.0598
Size	6.2755	0.3363	-0.2554	0.0465	-5.4952	0.2687
Earning Yield	1.9591	0.2088	0.0743	0.0497	1.4943	-0.2947
Volatility	6.4959	0.5537	0.6386	0.0892	7.1623	0.0852
Growth	1.8179	0.1707	-0.0310	0.0187	-1.6633	-0.0035
Value	3.1801	0.1727	-0.1731	0.0206	-8.403	0.0575
Leverage	2.6182	0.2445	-0.0310	0.0244	-1.2677	-0.0247
Liquidity	9.9032	0.1170	-0.7351	0.0211	-34.758	0.3168
Average	**5.1284**	**0.2940**	/	/	/	/

2.3. Portfolio Risk Forecasting

After screening the risk factors, we can use the definition of the multi-factor structured risk model in the first section to provide a linear decomposition of the expected portfolio return with respect to the risk factors:

$$R_P = \sum_{j=1}^{N} w_n \cdot \left(\sum_{k=1}^{K} x_{jk} f_{jk} + u_j \right)$$

Next, the prediction of portfolio risk is mainly based on the prediction of the covariance matrix of factor returns and the risk modeling of specific factors.

Forecasting of Factor Returns Covariance Matrix

In the first section, the portfolio risk can be expressed as

$$\sigma_P = \sqrt{w^T (XFX^T + \Delta) w}$$

In which F represents the covariance matrix ($K \times K$) of factor returns for K factors:

$$F = \begin{bmatrix} Var(f_1) & Cov(f_1, f_2) & \cdots & Cov(f_1, f_k) \\ Cov(f_1, f_2) & Var(f_2) & \cdots & Cov(f_2, f_k) \\ \cdots & \cdots & \cdots & \cdots \\ Cov(f_k, f_1) & Cov(f_k, f_2) & \cdots & Var(f_k) \end{bmatrix}$$

The structured multi-factor risk model explains the volatility of stock price returns by using the volatility of risk factor returns, while taking into account the correlation between different factor returns.

For the prediction of the covariance matrix F of factor returns, the simplest method is to calculate the historical covariance matrix of factor returns within the sample. However, this method assumes that the volatility of factor returns is stable, which may lead to significant bias in practical applications.

BARRA provides a more complete method for predicting covariance matrices in the risk models for EUE3, USE4, and CNE5:

Firstly, the daily covariance matrix $F^{(d)}$ of factor returns is calculated using the historical factor returns at a daily frequency, and the *RiskMetrics* weighted moving average (*EWMA*) method is used:

$$F_{kd}^{(d)} = cov(f_k, f_{k'})_t = \sum_{s=t-h}^{t} \lambda^{t-s} \left(f_{ks} - \overline{f_k}\right)\left(f_{k's} - \overline{f_{k'}}\right) \; / \; \sum_{s=t-h}^{t} \lambda^{t-s}$$

Where the exponential decay weight $\lambda = 0.5^{1/\tau}$ assigns higher weight to dates closer to the current date, and the half-life parameter τ indicates that the weight of the day $t - \tau$ is half of the weight of the day t, and h represents the sample time length. In actual calculation, we made h = 252, and $\tau = 90$.

Since we need to predict the risk of the next month, and the correlation coefficient matrix of the factor is calculated based on the daily return data of the factor, the impact of serial correlation between the returns of the factors must be considered.

We can perform Newey-West adjustment on the basis of $F^{(d)}$, and calculate the adjusted matrix F^{NW}, specifically:

$$F^{NW} = 21 \cdot [F^{(d)} + \sum_{\Delta=1}^{D} (1 - \frac{\Delta}{D+1})(C_{+\Delta}^{(d)} + C_{-\Delta}^{(d)})]$$

Where D represents the length of the lag time, and the calculation methods of $C_{+\Delta}^{(d)}$ and $C_{-\Delta}^{(d)}$ are as follows. The superscript d of the elements indicates that the indicator is calculated based on daily data:

$$C_{kk',+\Delta}^{d} = cov\left(f_{k,t-\Delta}, f_{k',t}\right) = \sum_{s=t-h+\Delta}^{t} \lambda^{t-s} \left(f_{k,s-\Delta} - \overline{f_k}\right)\left(f_{k',s} - \overline{f_{k'}}\right) / \sum_{s=t-h+\Delta}^{t} \lambda^{t-s}$$

$$C_{kk',-\Delta}^{d} = cov\left(f_{k,t}, f_{k',t-\Delta}\right) = \sum_{s=t-h+\Delta}^{t} \lambda^{t-s} \left(f_{k,s} - \overline{f_k}\right)\left(f_{k',s-\Delta} - \overline{f_{k'}}\right) / \sum_{s=t-h+\Delta}^{t} \lambda^{t-s}$$

It can be verified that the above two equations satisfy the following relationship:

$$C_{+\Delta}^{d} = {C_{-\Delta}^{d}}^{T}$$

In actual calculation, we set D=2 for both variance and covariance serial correlation lag time, and made h = 252, $\tau = 90$.

We used the above method to estimate the covariance matrix of factor returns at a daily frequency, with an equal-weighted portfolio. The results are shown on Github:

Barra/Barra/data/**covariance_matrix**/factor_cov.npy

The covariance matrix $F^{(d)}$ obtained above still suffers from non-stationary bias, which means that risk predictions may be persistently overestimated or underestimated. To address this issue, the factor cross-section deviation statistics B_t^F and the factor volatility multiplier λ_F need to be introduced.

Specifically, let f_{kt} be the actual factor returns of the *k*-th factor on the *t*-th day, and σ_{kt} be the predicted volatility of the factor returns on day *t* based on day *t-1*. The factor cross-section deviation statistics B_t^F can be expressed as:

$$B_t^F = \sqrt{\frac{1}{K}\sum_k \left(\frac{f_{kt}}{\sigma_{kt}}\right)^2}$$

where *K* is the number of factors. B_t^F is an intuitive method for testing whether there is bias in the predicted factor volatility, for example, if the predicted factor volatility is too low, then $B_t^F > 1$.

After obtaining the time series of B_t^F, the factor volatility multiplier λ_F can be obtained by weighting with exponential decay weights, in order to predict whether the predicted factor return volatility is unbiased, and to make corresponding adjustments, specifically:

$$\lambda_F = \sqrt{\sum_t (B_t^F)^2 w_t}$$

where w_t is the factor decay weight.

Therefore, the adjusted daily factor return covariance matrix can be expressed as:

$$\tilde{F}^{(d)} = {\lambda_F}^2 \cdot F^{(d)}$$

Modeling of Idiosyncratic Risk Factors

In the structured multi-factor risk model, the idiosyncratic risk component is the residual risk that cannot be explained by the common factors, denoted as $Var(uj)$ in the regression equation.Since the model assumes that the idiosyncratic factors are uncorrelated with the common factors and with each other across stocks, the idiosyncratic risk component Δ for the overall stock portfolio can be represented as a diagonal matrix.

The difficulty in estimating idiosyncratic risk lies in the need to calculate the risk independently for each stock, making time series methods a good choice.

Similar to the calculation of the covariance matrix *F*, we first use the exponentially weighted moving average (*EWMA*) method to calculate the daily idiosyncratic risk for each stock *j*, which can be expressed as:

$$\sigma_{u_jt}^d = (\sum_{s\,=\,t-h}^{t} \lambda^{t-s} \left(u_{js} - \overline{u_k}\right)^2 \ / \ \sum_{s\,=\,t-h}^{t} \lambda^{t-s})^{\ 1/2} \quad \lambda = 0.5^{1/\tau}$$

We use the above method to estimate the daily idiosyncratic risk matrix. The portfolio is equally weighted. The results are shown on Github:

Barra/Barra/data/covariance_matrix/single_sigma.npy

Of course, time series methods still generate estimation errors, for example:

1. For newly issued stocks, shorter historical data can result in larger errors in estimating specific factor risk;
2. For stocks with poor liquidity, the distribution of regression residuals tends to exhibit strong peak-tail behavior, making it difficult to characterize their specific risks;
3. Major events affecting listed companies can cause the residual term of the regression equation to increase significantly, and time series models cannot control the impact of this situation.

Meaning of risk models

The significance of risk models lies in identifying the causes of stock price volatility, decomposing and separating the sources of stock returns, and predicting future stock price volatility.

In our study of risk factors, we decomposed the sources of stock returns into 31 industry factors, 10 style factors, and specific factors, and explained the reasons for stock price volatility through the fluctuation of factor effects on stock prices.

Then, through the estimation of key variables in the risk model, we achieved quantitative predictions of portfolio volatility. However, just as there is no perfect normal distribution in financial markets, there is no 100% accurate risk prediction. But in the framework of the risk model, investment managers can quantitatively evaluate risk, and help them accurately control the risk characteristics of the portfolio, eliminate the interference of risk factors with high uncertainty while obtaining alpha factor returns with high certainty.

Based on the risk model, an optimal investment portfolio with style neutrality can be constructed through an optimization method for portfolio weights. Next, we will move on to the multi-factor stock selection strategy section.

3. Style-Neutral Multi-Factor Stock Selection Strategy

Pure factors combination

In the structured risk model, we introduced a more comprehensive method for testing the effectiveness of style factors, based on which we obtained 9 major style factors. For an investment portfolio consisting of N stocks with weights $w = (w_1, w_2, \dots, w_N)^T$, the portfolio return can be expressed as:

$$R_P = \sum_{j=1}^{N} w_n \cdot \left(\sum_{k=1}^{K} x_{jk} f_{jk} + u_j \right)$$

The above expression is a simple linear decomposition form, and by calculating it, we can obtain the contribution of each factor to the overall portfolio return. In order to more accurately grasp the return characteristics of each factor, we need to construct a portfolio in which the overall return of the portfolio comes solely from the return of a particular factor, while the returns of the other factors are all zero.

Specifically, for the k-th factor, if there exists a stock investment portfolio weight $w = (w_1, w_2, \dots, w_N)^T$ such that:

$$\begin{cases} (w^T - w_{bench}^T) X_k = 1 \\ \forall k' (w^T - w_{bench}^T) X_{k'} = 0 \end{cases}$$

Here $k \neq k'$, and w_{bench} represent the portfolio weight for hedging benchmarks (for example, if the CSI 300 stock index futures and options are used as hedging benchmarks, w_{bench} represents the corresponding weights of the constituent stocks of the CSI 300 index, and the weights of non-index constituent stocks are zero). The above constraint indicates that the risk exposure of factor k to the hedging benchmark is 1, and the risk exposure of other factors k'corresponding to the hedging benchmark is completely closed. We call such portfolio weight w the unit pure factor stock portfolio of factor *k*, and we call factor *k*'s risk factor exposure 1 unit higher than the hedging benchmark.

More generally, if there exists a stock investment portfolio weight $w = (w_1, w_2, \dots, w_N)^T$ such that only:

$$\forall k'(w^T - w_{bench}^T)X_{k'} = 0$$

where $k \neq k'$. We call such portfolio weight *w* the pure factor stock portfolio of factor *k*.

The significance of a pure factor stock portfolio is that, while considering the impact of factor *k* on returns, the influence of other factors on the portfolio is completely eliminated. This allows for a more objective examination of the risk and return characteristics of a portfolio based on factor *k*.

Risk factors and alpha factors

Let's take a closer look at the definition of alpha factors. An alpha factor is a variable that significantly and consistently affects stock returns, and this effect exists independently of the influence of all other factors on returns.

In the structured multi-factor risk model in previous paragraphs, we used the selection criteria for BARRA risk factors to obtain 9 major risk factors, all of which had statistically significant effects on stock returns. However, in order to further examine whether factors have alpha properties, we must construct pure factor stock portfolios based on a particular factor to examine the robustness of its return characteristics when it acts independently on a stock portfolio.

Next, we construct pure factor stock portfolios for each of the 9 types of factors. The objective function for constructing the portfolio weights does not consider risk characteristics, but only maximizes the expected return of the portfolio. At the same time, we ensure that industry factor returns do not affect the results through industry-neutral constraints, as follows:

$$\begin{aligned} & Max\ \ w^T R_P \\ & s.t. \forall k'(w^T - w_{bench}^T)X_{k'} = 0 \\ & \quad w^T H = h^T \\ & \quad w \geq 0 \\ & \quad \sum_{i=1}^{N} w_i = 1 \end{aligned}$$

Here, $R = (r_1, r_2, \dots, r_N)^T$represents the expected return vector of the portfolio of *N* stocks; H represents the dummy variable matrix of the industries of the *N* stocks; $h = (h_1, h_2, \dots, h_{31})^T$ represents the proportional weights of the 31 industries corresponding to the CSI 300 index.

In this optimization equation, $\forall k'^{(w^T - w_{bench}^T)X_{k'}} = 0$ represents the risk exposure of any style factor *k'* other than the *k-th* factor being 0; $w^T H = h^T$represents the risk exposure of any industry factor being 0; $w^T R$ represents the objective of the weight optimization is to maximize the expected return of the portfolio. Therefore, the result $w = (w_1, w_2, \dots, w_N)^T$ of the above optimization equation is the pure factor stock portfolio of the *k-th* factor.

We separately examine the returns of the pure factor stock portfolios for the 9 major style factors in the risk model, and the cumulative return curves are as follows (in the calculations, we restrict the risk exposure of other factors to be within the range of ± 1):

Table 5: Pure Beta portfolio performance statistics

Portfolio performance		Risk factor exposure (mean)	
Combination term	60 months	Beta	0.610
Cumulative rate of return	21.8%	Momentum	0.010
Daily trade win rate	53.31%	Size	-0.010
Annualized rate of return	3.94%	Earning Yield	0.010
Annualized volatility	5.51%	Volatility	0.010
Information ratio	1.17	Growth	0.009
Maximum drawdown	4.6%	Value	-0.010
95%VaR	-0.62%	Leverage	-0.004
Daily yield T value	2.54	Liquidity	-0.010

Table 6: Pure Momentum portfolio performance statistics

Portfolio performance		Risk factor exposure (mean)	
Combination term	60 months	Beta	0.010
Cumulative rate of return	32.8%	Momentum	0.625
Daily trade win rate	56.38%	Size	-0.010
Annualized rate of return	5.72%	Earning Yield	0.010
Annualized volatility	2.83%	Volatility	0.009
Information ratio	2.076	Growth	0.009
Maximum drawdown	3.05%	Value	-0.010
95%VaR	-0.41%	Leverage	-0.009
Daily yield T value	4.64	Liquidity	-0.010

Table 7: Pure Size portfolio performance statistics

Portfolio performance		Risk factor exposure (mean)	
Combination term	60 months	Beta	-0.003
Cumulative rate of return	136%	Momentum	-0.003
Daily trade win rate	60.29%	Size	-1.292
Annualized rate of return	17.96%	Earning Yield	-0.010
Annualized volatility	6.83%	Volatility	0.001
Information ratio	2.773	Growth	-0.007
Maximum drawdown	13.22%	Value	-0.010
95%VaR	-0.86%	Leverage	-0.008

Daily yield T value	6.26	Liquidity	0.009

Table 8: Pure Earnings Yield portfolio performance statistics

Portfolio performance		Risk factor exposure (mean)	
Combination term	60 months	Beta	0.010
Cumulative rate of return	35.8%	Momentum	0.010
Daily trade win rate	53.88%	Size	-0.010
Annualized rate of return	6.18%	Earning Yield	0.641
Annualized volatility	2.66%	Volatility	-0.009
Information ratio	2.263	Growth	0.010
Maximum drawdown	2.77%	Value	-0.010
95%VaR	-0.309%	Leverage	-0.004
Daily yield T value	5.29	Liquidity	-0.010

Table 9: Pure Volatility portfolio performance statistics

Portfolio performance		Risk factor exposure (mean)	
Combination term	60 months	Beta	0.010
Cumulative rate of return	29.4%	Momentum	0.010
Daily trade win rate	53.89%	Size	-0.010
Annualized rate of return	5.19%	Earning Yield	0.010
Annualized volatility	2.67%	Volatility	-0.385
Information ratio	2.09	Growth	0.010
Maximum drawdown	2.56%	Value	-0.010
95%VaR	-2.97%	Leverage	-0.010
Daily yield T value	4.67	Liquidity	-0.010

Table 10: Pure Growth portfolio performance statistics

Portfolio performance		Risk factor exposure (mean)	
Combination term	60 months	Beta	0.010
Cumulative rate of return	55.13%	Momentum	0.010
Daily trade win rate	3.23%	Size	-0.010
Annualized rate of return	3.31%	Earning Yield	0.010
Annualized volatility	2.415	Volatility	-0.008
Information ratio	1.37	Growth	0.577
Maximum drawdown	4.11%	Value	-0.010

95%VaR	-3.03%	Leverage	-0.008
Daily yield T value	4.63	Liquidity	-0.010

Table 11: Pure Value portfolio performance statistics

Portfolio performance		Risk factor exposure (mean)	
Combination term	60 months	Beta	0.010
Cumulative rate of return	10.9%	Momentum	0.010
Daily trade win rate	53.8%	Size	-0.010
Annualized rate of return	2.07%	Earning Yield	0.010
Annualized volatility	2.94%	Volatility	0.007
Information ratio	0.712	Growth	0.009
Maximum drawdown	7.86%	Value	-0.639
95%VaR	-3.57%	Leverage	-0.010
Daily yield T value	1.58	Liquidity	-0.010

Table 12: Pure Leverage portfolio performance statistics

Portfolio performance		Risk factor exposure (mean)	
Combination term	60 months	Beta	0.010
Cumulative rate of return	16.93%	Momentum	0.010
Daily trade win rate	51.3%	Size	-0.010
Annualized rate of return	3.12%	Earning Yield	0.010
Annualized volatility	2.39%	Volatility	-0.010
Information ratio	1.29	Growth	0.009
Maximum drawdown	2.97%	Value	-0.010
95%VaR	-0.31%	Leverage	-0.495
Daily yield T value	2.88	Liquidity	-0.010

Table 13: Pure Liquidity portfolio performance statistics

Portfolio performance		Risk factor exposure (mean)	
Combination term	60 months	Beta	0.010
Cumulative rate of return	28%	Momentum	0.010
Daily trade win rate	53.74%	Size	-0.010
Annualized rate of return	5.19%	Earning Yield	0.010
Annualized volatility	2.66%	Volatility	-0.010
Information ratio	1.892	Growth	-0.001

Maximum drawdown	3.64%	Value	-0.010
95%VaR	-0.316%	Leverage	-0.003
Daily yield T value	4.04	Liquidity	-0.457

Table 14: Pure factor portfolio performance summary

Pure factor portfolio	Annualized rate of return	Annualized volatility	Information ratio	Maximum drawdown
Beta	3.94%	5.51%	1.17	4.6%
Momentum	5.72%	2.83%	2.076	3.05%
Size	17.96%	6.83%	2.773	13.22%
Earning Yield	6.18%	2.66%	2.263	2.77%
Volatility	5.19%	2.67%	2.09	2.56%
Growth	3.31%	2.415%	1.37	4.11%
Value	2.07%	2.94%	0.712	7.86%
Leverage	3.12%	2.39%	1.29	2.97%
Liquidity	5.19%	2.66%	1.892	3.64%

The examination of pure factor stock portfolios reflects the return and risk characteristics of a portfolio when the returns come solely from a particular factor.

Taking the pure Earning Yield factor portfolio as an example, when the portfolio is exposed only to the Earning Yield factor, completely eliminating industry factors and other style factors, it can achieve an annualized return of about 6.18% and an annualized volatility of 2.66%. At this point, when maximizing the expected return of the portfolio, the Earning Yield factor has a relative average risk exposure of approximately 0.641 compared to the CSI 300 benchmark.

After considering all the pure factor portfolios, we found that the performance of the pure Momentum, Earning Yield, and Volatility factor portfolios was relatively stable, with strong alpha properties, while the returns of the other pure factor portfolios varied to different extents. The results of the pure Earning Yield factor portfolio are consistent with the traditional economic view that a company's profitability has a significant impact on its stock performance. The Earning Yield factor includes a consistent expectation view on a company's profitability, indicating that market expectations have a significant impact on stock prices.

The results of the pure Momentum and Volatility factor portfolios show that market forces also have an impact on stock prices, indirectly proving the feasibility of technical analysis.

It is worth mentioning that the performance of the pure Size factor portfolio is quite special. Compared to other factors, the Size factor can be described as having enormous impact. The pure Size factor portfolio can achieve an annualized return of approximately 17.69%, the highest among all factor portfolios, while its annualized volatility of 6.83% and maximum drawdown of 13.22% are also the highest among all factor portfolios. We believe that the alpha properties of the Size factor, as a typical style factor, are worth discussing.

Industry neutrality and style neutrality

Industry neutrality refers to the industry allocation of a long portfolio being consistent with that of the hedge benchmark. The purpose of industry-neutral allocation is to eliminate the impact of industry factors on strategy returns. Unlike traditional industry allocation, which attempts to over-allocate to strong industries and under-allocate to weak ones in a future period, the characteristic of industry neutrality is to eliminate the impact of industry-level factors and only consider the excess returns of individual stocks within industries. The net value curve of an industry-neutral strategy tends to be relatively stable with small drawdowns.

In the CSI 300 index, the top 5 weighted industries are banking, mining, real estate, securities, and non-ferrous metals, while the bottom 5 are leisure services, light manufacturing, diversified finance, textile and clothing, and agriculture, forestry, and animal husbandry.

Assuming H is the dummy variable matrix of industry factors for the sample stocks, and h represents the corresponding weights of the 31 industries in the CSI 300 index, the industry-neutral weight w satisfies the following equation:

$$w^T H = h^T; w \geq 0$$

Style factor neutrality refers to the risk exposure of the long portfolio's style factors being equal to zero relative to the hedge benchmark. The significance of style factor neutrality is to completely match the style characteristics of the long portfolio with the hedge benchmark, so that the excess returns of the portfolio do not come from any specific style. Our goal is to pursue stable alpha returns, rather than returns from a specific market style. After implementing style factor neutrality, the net value curve of the strategy will be further smoothed and the maximum drawdown will be reduced, resulting in a significant improvement in the stability of the portfolio compared to only considering industry neutrality.

Assuming X_k is the loading cross-section of the k-th factor in the sample, and w_{bench} is the corresponding weight of the CSI 300 index, the style factor neutral weight w for factor k satisfies the following equation:

$$(w^T - w_{bench}^T)X_{k\prime} = 0; \quad w \geq \ 0$$

If weight w satisfies the above expression for any style factor in the portfolio, the portfolio w is considered to be style factor neutral.

In simple terms, the purpose of style factor neutrality is to closely approximate the style characteristics of the long portfolio to the hedge benchmark, in order to avoid potential volatility in the strategy caused by that specific style factor.

Portfolio weight optimization

In the introduction, we mentioned that the optimization of portfolio weights plays a crucial role in multifactor models. The purpose of portfolio weight optimization is to fully quantify the risk characteristics of the portfolio, allowing investment managers to gain a clear understanding of the sources of portfolio returns and risk exposures.

The process of portfolio weight optimization includes two factors: the objective function of weight optimization and the constraints. We have already mentioned the constraints in the previous section, which are industry neutrality and style factor neutrality of the portfolio. Regarding the objective function of weight optimization, there are several different approaches.

1. Minimizing Expected Portfolio Risk

The calculation of portfolio volatility in a structured risk model is represented by $\sigma_P = \sqrt{w^T(XFX^T + \Delta)w}$. Therefore, the expression for weight optimization that minimizes expected portfolio risk is:

$$Min\ \ w^T(XFX^T + \Delta)w$$

$$s.t. \forall k'(w^T - w_{bench}^T)X_{k'} = 0$$

$$w^T H = h^T$$

$$w \ \geq \ 0$$

$$\sum_{i=1}^{N} w_i = 1$$

where k' represents the identified style factors, H is the dummy variable matrix for industry factors of sample stocks, and h represents the corresponding weights of the 31 industries in the CSI 300 Index.

2. Maximizing Risk-Adjusted Returns

Maximizing risk-adjusted returns is the objective function, which considers the role of expected returns and expected risk, and introduces the risk aversion coefficient λ within the framework of Markowitz's mean-variance theory. The specific expression for weight optimization is:

$$Max\ R_P - \lambda\sigma_P^2 - TC(w)$$

$$s.t. \forall k'(w^T - w_{bench}^T)X_{k'} = 0$$

$$w^T H = h^T$$

$$w \ \geq \ 0$$

$$\sum_{i=1}^{N} w_i = 1$$

where $TC(w)$ represents the transaction cost of constructing the portfolio using weights *w*. $R_P = \sum_{j=1}^{N} w_n \cdot \left(\sum_{k=1}^{K} x_{jk} f_{jk} + u_j\right)$, $\sigma_P = \sqrt{w^T(XFX^T + \Delta)w}$.

3. Maximizing Portfolio Information Ratio

Maximizing portfolio information ratio is the objective function, which uses the ratio of expected portfolio returns to expected portfolio risk as the target function. The specific expression for weight optimization is:

$$Max\ \frac{R_P - TC(w)}{\sigma_P}$$

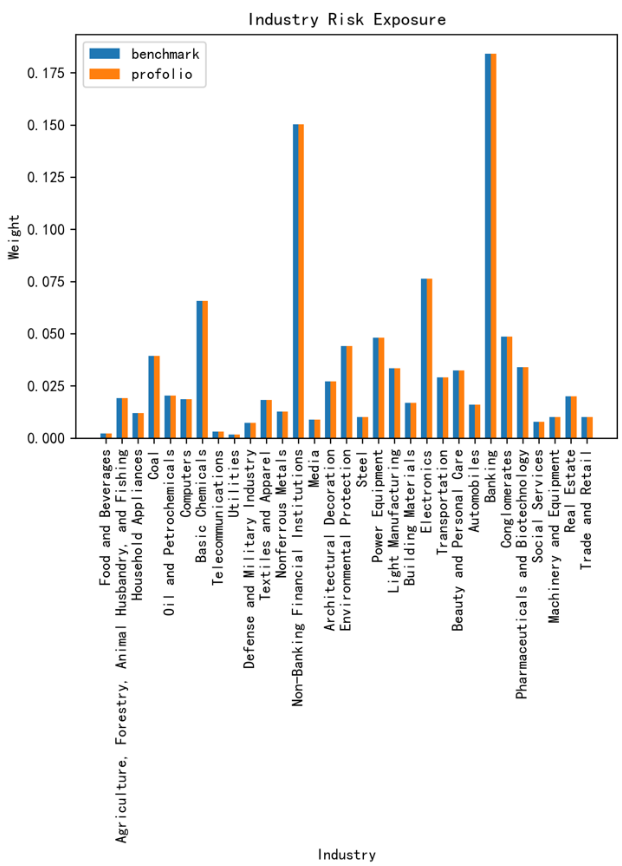
$$s.t. \forall k'(w^T - w_{bench}^T)X_{k'} = 0$$

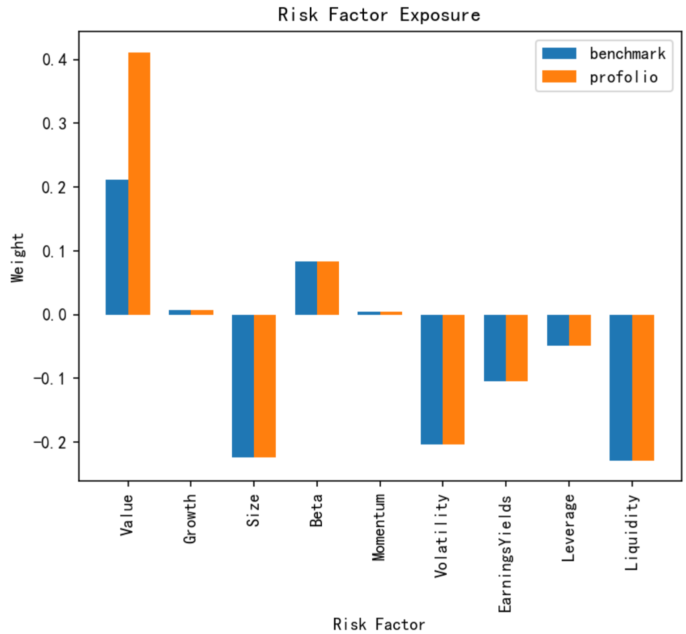
$$w^T H = h^T$$

$$w \geq 0$$

$$\sum_{i=1}^{N} w_i = 1$$

Among the three optimization objective functions mentioned above, the first and third methods rely entirely on the data results provided by the risk model for calculation, while the second method, which maximizes risk-adjusted returns as the objective function, introduces the risk aversion coefficient λ to enhance the flexibility of weight calculation. This enables investment managers to make differentiated choices based on their own risk preferences.





Empirical verification

So far, we have completed all the basic work for portfolio construction, including factor validity testing, regression equation parameter estimation, portfolio risk modeling, pure factor portfolio testing, and portfolio weight optimization based on industry neutrality and style factor neutrality. Next, empirical analysis will be conducted to examine the strategy effects of market-neutral portfolios, market and industry-neutral portfolios, and market, industry, and style-neutral portfolios, respectively, and to compare the differences among the three.

We first present the relevant hypothesis parameters for the empirical analysis:

The backtesting period is from January 2015 to December 2019, with January 2015 to January 2016 as the in-sample period for extracting factor combination parameters;

1. The stock pool consists of all A-share non-ST stocks;
2. The transaction cost is set at 0.1% per side, and stamp duty is 0.1%;
3. The weight upper limit for individual stocks in the portfolio is set at 1% (the banking, securities, and insurance industries have relatively high weights, and the weight upper limits are set at 3%, 2%, and 2%, respectively, to achieve industry-neutral allocation);
4. We use the form $Max\ R_P - \lambda\sigma_P^2 - TC(w)$ for the optimization objective function, where $\lambda = \frac{1}{2}$;
5. For the industry neutrality constraint, the factor exposure is set at $\pm 5\%$;
6. For the style neutrality constraint, the factor exposure for the nine style factors is set as follows:

Factor type	Risk exposure constraint upper bounds	Risk exposure constraint lower Bounds
Beta	-0.01	+0.01
Momentum	∞	+0.3
Size	∞	-0.3
Earning Yield	∞	+0.5
Volatility	-0.3	∞
Growth	-0.01	+0.01
Value	-0.01	+0.01
Leverage	-0.01	+0.01
Liquidity	-0.01	+0.01

If the weight optimization equation has no solution under the set factor exposure constraints, the exposure of the Momentum and Earning Yield factors is successively reduced by 0.1 until the optimal solution is found.

The results of the strategy testing are presented below:

1. Constructing Market-Neutral Hedged Portfolio.

$$Max\ R_P - \lambda\sigma_P^2 - TC(w)$$

$$s.t.w \ \geq \ 0$$

$$\sum_{i=1}^{N} w_i = 1$$

Table 15: Market capitalization neutral portfolio performance statistics

Strategy equity	4.743	Maximum drawdown	37.35%
Trading win rate	60.77%	Maximum drawdown start time	2018/11/3
Annualized rate of return	32.45%	Maximum drawdown end time	2018/12/29
Annualized volatility	15.89%	Daily return distribution skewness	-0.92
Information ratio	2.01	Daily return distribution kurtosis	8.47
Profit and loss ratio	0.864	95% VaR	-1.793%
Portfolio average annual turnover rate	360%	Portfolio stocks (average)	100

2. Constructing Market and Industry-Neutral Hedged Portfolio.

$$Max\ R_P - \lambda\sigma_P^2 - TC(w)$$

$$s.t.\ w^T H = h^T$$

$$w \geq 0$$

$$\sum_{i=1}^{N} w_i = 1$$

Table 16: Market capitalization and industry neutral portfolio performance statistics

Strategy equity	3.304	Maximum drawdown	21.67%
Trading win rate	62.76%	Maximum drawdown start time	2018/11/20
Annualized rate of return	23.69%	Maximum drawdown end time	2018/12/29
Annualized volatility	9.98%	Daily return distribution skewness	-1.03
Information ratio	2.433	Daily return distribution kurtosis	7.27
Profit and loss ratio	0.879	95% VaR	-1.061%
Portfolio average annual turnover rate	340%	Portfolio stocks (average)	95

3. Constructing Market, Industry, and Style-Neutral Hedged Portfolio.

$$Max\ R_P - \lambda\sigma_P^2 - TC(w)$$

$$s.t. \forall k' (w^T - w_{bench}^T) X_{k'} = 0$$

$$w^T H = h^T$$

$$w \geq 0$$

$$\sum_{i=1}^{N} w_i = 1$$

Table 17: Market capitalization, industry, style neutral portfolio performance statistics

Strategy equity	1.874	Maximum drawdown	2.065%
Trading win rate	60.78%	Maximum drawdown start time	2018/11/26
Annualized rate of return	12.61%	Maximum drawdown end time	2018/12/07
Annualized volatility	3.25%	Daily return distribution skewness	-0.07

Information ratio	4.15	Daily return distribution kurtosis	3.76
Profit and loss ratio	1.18	95% VaR	-0.318
Portfolio average annual turnover rate	350%	Portfolio stocks (average)	100

Table 18: Segmental Performance Statistics of Style-Neutral Portfolios

Date	Annualized rate of return	Maximum drawdown	Information ratio
2015	15.7%	2.01%	4.62
2016	8.2%	1.77%	3.01
2017	11.34%	0.88%	4.85
2018	18.1%	1.13%	5.79
2019	10.5%	2.052%	3.46

Table 19: Performance comparison of the three combinations

Combination method	Annualized rate of return	Annualized volatility	Maximum drawdown	Information ratio
Market capitalization neutral	32.45%	15.89%	37.35%	2.01
Market capitalization and industry neutral	23.69%	9.98%	21.67%	2.433
Market capitalization, industry, style neutral	12.61%	3.25%	2.065%	4.15

From the results of the empirical analysis, it can be seen that after achieving market, industry, and style neutrality in the portfolio, the net value volatility of the portfolio significantly decreased, with a maximum drawdown reduced to 2.01% and an information ratio increased to 4.15.

Although style neutrality may lower the annualized return of the strategy, the stability and certainty of the return source have greatly improved. This is an ideal outcome for those pursuing stable absolute returns or alpha.

Attribution analysis

Attribution analysis can be divided into two categories: stock attribution analysis and factor attribution analysis. From the perspective of overall risk control of the strategy, factor attribution analysis enables investment managers to quantitatively study which factors contributed to the strategy's returns and risks after the fact.

- **Factor return attribution**

In a structured risk model, the expected portfolio return is linearly expressed using industry factors, style factors, and idiosyncratic residual factors:

$$R_P = \sum_{j=1}^{N} w_n \cdot \left(\sum_{k=1}^{K} x_{jk} f_{jk} + u_j \right)$$

Factor return attribution is the product of the portfolio's risk factor exposure and the factor return rate in the current period, with the remaining return sources attributed to idiosyncratic residual factors. Specifically, at the beginning of period *t*, on the portfolio construction day, the portfolio's exposure to the style factor k is obtained by optimizing the portfolio weights $w^T X_k - w_{bench}^T X_k$. Then, on the portfolio rebalancing day at the end of period *t*, a regression analysis is performed using the actual stock returns and common factors to obtain the factor return rate *f* for period *t*, where the return rate of the *k-th* factor is f_k. Therefore, the factor return contribution of the *k-th* factor in period *t* is:

$$(w^T - w_{bench}^T) X_k \cdot f_k$$

We conducted factor return attribution analysis on the results of the three empirical tests mentioned above, and the results are presented below:

From the factor return results, it can be observed that both the market-neutral and industry-neutral portfolios have significant exposure to the Size factor, exhibiting a clear small-cap, high volatility, and low liquidity characteristic. The majority of the portfolio's return contribution also depends on the Size factor, and the strategy as a whole has a clear style rotation feature between small and large caps.

On the other hand, the style-neutral portfolio has relatively small deviations in its factor exposure due to the constraints on the style factor exposure. Most of the exposure is concentrated in the Earning Yield and Momentum factors, and the majority of the portfolio's return contribution comes from these two factors. The Size factor's return contribution is relatively low. The strategy as a whole does not exhibit significant style characteristics and meets the expected style neutrality requirements.

- **Factor risk attribution**

In a structured risk model, the expression of portfolio risk is:

$$\sigma_P = \sqrt{w^T (X F X^T + \Delta) w}$$

For each type of factor, its factor risk contribution comes from two parts: factor risk exposure and factor volatility. The greater the factor exposure and the greater the factor volatility, the higher the factor risk contribution.

Specifically, for a combination of *N* stocks and *K* factors, the risk contribution of the *k-th* factor is:

$$(w^T - w_{bench}^T) X F ((w^T - w_{bench}^T) X_k)$$

The risk attribution results also indicate that compared to the previous two portfolio configurations, the style-neutral portfolio has relatively evenly distributed risk contributions across all factors, without a significant concentration of Size factor risk exposure, and the portfolio does not exhibit any significant style characteristics.

Therefore, both the strategy backtesting and attribution modeling demonstrate the effectiveness of the style-neutral portfolio optimization method in controlling strategy risk and making alpha returns more robust.

4. Research summary and prospects

In contrast to previous multi-factor stock selection research reports, this report does not focus on the exploration of effective factors, but instead investigates the impact of portfolio optimization on strategy performance and introduces the method of style-neutral portfolio allocation.

Starting with a structured risk model, we decompose stock returns into 31 industry factors, 10 style factors, and idiosyncratic factors and predict the covariance matrix of factor returns and the risk matrix of idiosyncratic factors to quantitatively forecast portfolio volatility. Through the risk model, we quantitatively grasp the risk characteristics of the portfolio, laying a foundation for constructing the optimal investment portfolio.

To test the independent impact of each factor on returns, we constructed a pure factor portfolio by constraining the optimization equation. The test showed that Earning Yield, Momentum, and Volatility factors have strong alpha characteristics, while Size factors embody high returns and volatility, and significant style features.

In the study of portfolio optimization, we introduced three different weight optimization objective functions and the constraints of industry neutrality and style neutrality. The purpose of portfolio weight optimization is to expose the sources of the portfolio's returns more purely to factors with strong alpha characteristics and neutralize all other risk factors. Under this concept, we constructed the optimal investment portfolio with market neutrality, industry neutrality, and style neutrality.

In the empirical test section, we examined the three methods of portfolio construction. The test results showed that considering style neutrality constraints significantly improves the stability of strategy returns. The portfolio's annualized return is 12.61%, the maximum drawdown is 2.065%, and the information ratio is 4.15.

The attribution analysis results showed that the sources of factor returns for the first two portfolio construction methods depended largely on the Size factor, and the strategy had a strong market style feature. The factor returns of the style-neutral portfolio were relatively evenly distributed, and the strategy did not have obvious factor features. The results of risk attribution analysis also confirmed this.

In future multi-factor model research, we will consider the following relevant issues: 1) exploration of new alpha factors; 2) construction methods of portfolios hedged against the CSI 500 index; 3) relevant research on style factor exposure adjustment.

Github link for all the data files and python codes

https://github.com/XiaoYunhan/Barra

Reference

Han, Zhenguo. "20180303-Fangzheng Securities-'Spark' Multi-factor Series (II): Advanced Barra Model, Multi-factor Model Risk Prediction." 方正证券研究所, 03 Mar. 2018.

"Get Factor Covariance." RiceQuant. Accessed 13 Mar. 2023. https://www.ricequant.com/doc/rqdata/python/risk-factors-mod.html#get-factor-covariance-%E8%8E%B7%E5%8F%96%E5%9B%A0%E5%AD%90%E5%8D%8F%E6%96%B9%E5%B7%AE

Song, Yang, and Shiliang Zhang. "Barra 风险模型（CNE6）之纯因子构建与因子合成: Pure Factor Construction and Factor Synthesis in Barra Risk Model (CNE6) - Multi-factor Model Research Series No. 9." 渤海证券研究所, 20 June 2019.

USE4 Methodology Notes August 2011." Top 1000 Funds, Aug. 2011. https://www.top1000funds.com/wp-content/uploads/2011/09/USE4_Methodology_Notes_August_2011.pdf