Index Tracking Strategy

Student： 1801212831 陈熙元

1801212845 龚莉

1801212897 陆金磊

1801212920 史云霞

1701213097 汪新博

Date： 2020.01.15

Group Member	Task
陈熙元	Stock selection
龚莉	Forecast Mean of Stock Return
陆金磊	Solve Index Tracking Problem
史云霞	Forecast Covariance Matrix
汪新博	Data Processing

Catalog

1 Data Processing..3

2 Forecast Mean of Stock Return..3

3 Forecast Covariance Matrix..4

3.1 Historical Covariance Method...5

3.2 Exponential Moving Average Method (EWMA)...5

4 Stock Selection..5

5 Solve Index Tracking Problem...6

Reference 10

1 Data Processing

We select the index tracking strategy as our main research topic. To keep track with market index, we get the individual stock return data, market return data and index return data from CSMAR. As the stocks included in CSI 300 keep changing during past 10 years, we also download the stock list in CSI 300 from WIND.

For the selection of data frequency, we balance the amount of data and complexity of calculation and finally use the monthly frequency data in our research framework. For the selection of data window, we select the data from 2010 to 2019, which on the one hand satisfies the need of relative ample data size, on the other hand this time period includes both stock market peak in 2015 and stock market bottom in 2013 to make sure other model fits well for different status of stock market.

2 Forecast Mean of Stock Return

To estimate the expected monthly return of individual stocks and CSI 300 index, we use both sample mean estimator and factor model estimator. The sample mean estimator at month t is calculated as the average monthly return of the previous 36 months. We also adopt the well-known Fama-French three factor model as an alternative way to estimate stock returns, and the monthly factor data before November, 2017 are available from CSMAR. Based on the linear relationship between the stock/index return and risk premium, SMB, HML factors,

$$r_{it} = \alpha_i + \beta_{i1}(r_{mt} - r_{ft}) + \beta_{i2} SMB_t + \beta_{i3} HML_t + \varepsilon_{it}$$

α_i, β_{i1}, β_{i2} and β_{i3} at the end of month t are estimated by time-series regression using data from the end of month $t - 36$ to the end of month $t - 1$, denoted as $\hat{\alpha}_i$, $\hat{\beta}_{i1}$, $\hat{\beta}_{i2}$ and $\hat{\beta}_{i3}$. Then we obtain the expected return for individual stocks or index,

$$E[r_{it}] = \hat{\alpha}_i + \hat{\beta}_{i1}(E[r_{mt}] - r_{ft}) + \hat{\beta}_{i2} E[SMB_t] + \hat{\beta}_{i3} E[HML_t]$$

where $E[r_{mt}] - r_{ft}$, $E[SMB_t]$ and $E[HML_t]$ are predicted by Autoregressive Integrated Moving Average ($ARIMA$).

At month t, $ARIMA(p, d, q)$ model is fitted to time series data of risk premium factor, SMB factor and HML factor over the past 36 months to forecast factor return at month $t + 1$. $Pmdarima$ is a statistical library designed to fill the void in Python's time series analysis capabilities, including the equivalent of R's $auto.arima$ function. The $auto\_arima$ process seeks to identify the most optimal parameters for an $ARIMA(p, d, q)$ model by conducting differencing tests to determine the order of differencing d, and then fitting models within ranges of defined $start\_p$, max_p, $start\_q$, max_q ranges. In order to find the best model, $auto\_arima$ optimizes for a given $information\_criterion$ and returns the $ARIMA$ which minimizes the value. Assuming the next-one-period return forecast at month t as the expected return for each factor, then expected return for individual stocks and CIS 300 index can be ultimately obtained.

3 Forecast Covariance Matrix

Valeriy Zakamulin's paper *A Test of covariance-matrix Forecasting Methods* (2015) evaluated the advantages and disadvantages of five different covariance matrix prediction methods from the perspectives of statistics and practical application, including rolling history covariance method, rolling shrinkage history covariance method, exponential moving average method (EWMA), DCC-GARCH method and GO-GARCH method.

The empirical results show that two prediction methods based on GARCH model perform best, EWMA method ranks the second, and rolling history covariance and rolling shrinkage history covariance method perform worst. Considering that EWMA method has a higher prediction accuracy, and it is simpler and faster than GARCH model, we use EWMA method to predict the covariance matrix and compare it with the rolling history covariance method.

3.1 Historical Covariance Method

The rolling historical method directly uses the covariance matrix of historical returns as the prediction. The formula is as follows:

$$r_t = \mu + \varepsilon_t$$

$$\Sigma_t^{Hist} = \frac{1}{L} \sum_{i=t-L}^{t-1} \varepsilon_t \varepsilon_t'$$

Where μ is the mean of monthly stock return, L is rolling window (we use 36 months).

3.2 Exponential Moving Average Method (EWMA)

The calculation method can be expressed in the following recursive form:

$$\Sigma_t^{EWMA} = (1 - \lambda)\varepsilon_{t-1}\varepsilon_{t-1}' + \lambda \Sigma_{t-1}^{Hist}$$

Where $0 < \lambda < 1$ is a constant decay parameter (we use $\lambda = 0.97$).

4 Stock Selection

We use the model based on similarity following

$$\begin{aligned} &\max_{y_j, x_{ij}} \sum_{i=1}^{N} V_i \sum_{j=1}^{N} \rho_{ij} x_{ij} \\ &s.t. \sum_{j=1}^{N} y_j = n \\ &\sum_{j=1}^{N} x_{ij} = 1 \\ &x_{ij} \le y_j, i = 1, ..., N; j = 1, ..., N \\ &x_{ij}, y_j = 0 or 1, i = 1, ...N; j = 1, ...N \end{aligned}$$

where V_i is the market value of stock i, ρ_{ij} is the level of similarity of stock i and stock j, x_{ij} represents whether stock i will be replaced by stock j in the portfolio, y_j represents whether stock j will enter the portfolio, and n is the number of stocks in portfolio. And we set the value of n as 50.

Here two methods are employed to estimate similarity. The first is the simply correlation coefficient which is derived from covariance forecasted in the former part. The second is the distance between factors of two stocks. We use the most common three factors: market risk coefficient (Beta), market value of equity (ME), the book to market value of equity (BM). In every period, values of three sets are normalized by

$$\mathrm{z}_{\mathrm{ft}} = \frac{x_{ft} - mean(x_{ft})}{std(x_{ft})}$$

where f is factor type, and t is the period. And then we calculate the distance between two stocks as

$$Dist_{ijt} = \sqrt{(Beta_{it} - Beta_{jt})^2 + (ME_{it} - ME_{jt})^2 + (BM_{it} - BM_{jt})^2}$$

This distance is used to measure the level of similarity.

5 Solve Index Tracking Problem

First, for each month in the back-testing periods, we input the most current CSI300 constituent stock list at that time into the “stock selection” subroutine. Based on the selected stock list, we use the methods introduced earlier to forecast the next-month’s mean vector and covariance matrix of individual stocks’ returns, as well as the next-month’s mean return and volatility of benchmark index (i.e. CSI300).

Having all these inputs available, we use the formulation of index tracking problem introduced in the class. That is

$$\min_x x'\Sigma x - 2\sigma_m^2 \beta' x\,, s.t.\, \mu' x = \mu_m, 1'x = 1$$

Moreover, in order to consider transaction cost, we also introduce the transaction cost constraint, i.e. $x_i - x_i^0 \leq y_i, y_i \geq 0, x_i^0 - x_i \leq z_i, z_i \geq 0 \ (i = 1, \dots n)$, $\sum_{i=1}^{n}(\mu_i\, x_i - t_i y_i - t_i' z_i) \geq \mu_m$. For simplicity, we assume that the transaction costs for buying and selling stocks are both 3‰. Finally, to avoid the problem that the solution of optimization problem can be sensitive to the possible estimation error of inputs, we add the extra diversification constraint, i.e. $x_i \leq 0.05$ as well as the no-short sales constraint, i.e. $x_i \geq 0$ $(i = 1,2, \dots, n)$.

In order to evaluate the tracking performance, we first calculate our constructed index fund's monthly return as well as benchmark index's monthly return. The plots show that the overall tracking performance of our constructed index fund is quite good, no matter we use factor model and EWMA to forecast mean vector and covariance matrix, or just use simple historical sample mean and covariance matrix. However, we can see that in terms of extreme market environment, such as when benchmark index return increases or drops more than 10% in a month, our constructed index fund tends to show more volatility. The possible explanation is that the index fund holds far less stocks than the benchmark index (figure 3 & 4). Finally, we also calculate our constructed index fund's daily return as well as benchmark index's daily return. The plots show that when using factor model and EWMA, the cumulative daily return of our constructed index fund tracks the benchmark index level quite well in the first half of our back-testing period while not so well in the second half; when using simple historical sample mean and covariance, the cumulative daily return of our constructed index fund tracks the benchmark index level quite well when the market experienced large sharp increase in late 2014 followed by a large drop in 2015 (known as 2015 A-share market plunge), while in less volatile periods its tracking performance deteriorate to some extent.

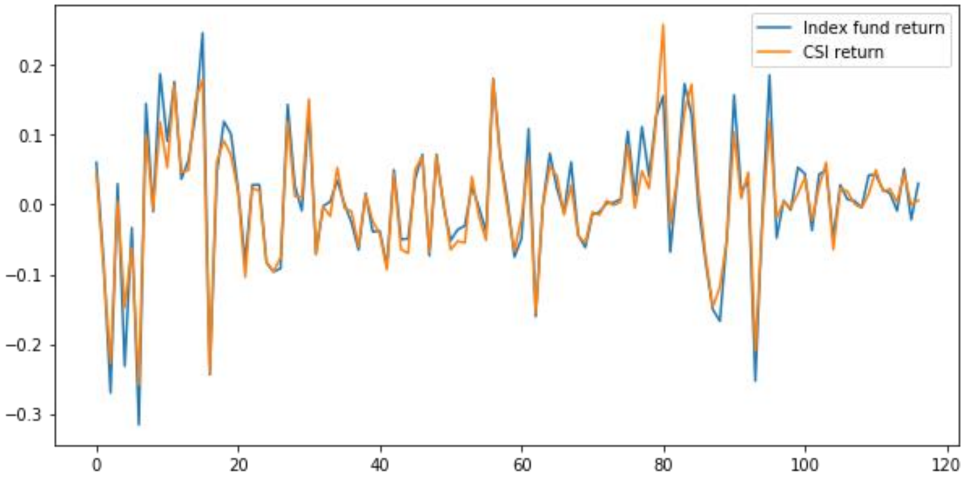

Figure 1 – The tracking performance of constructed index fund in terms of monthly return (using factor model and EWMA)

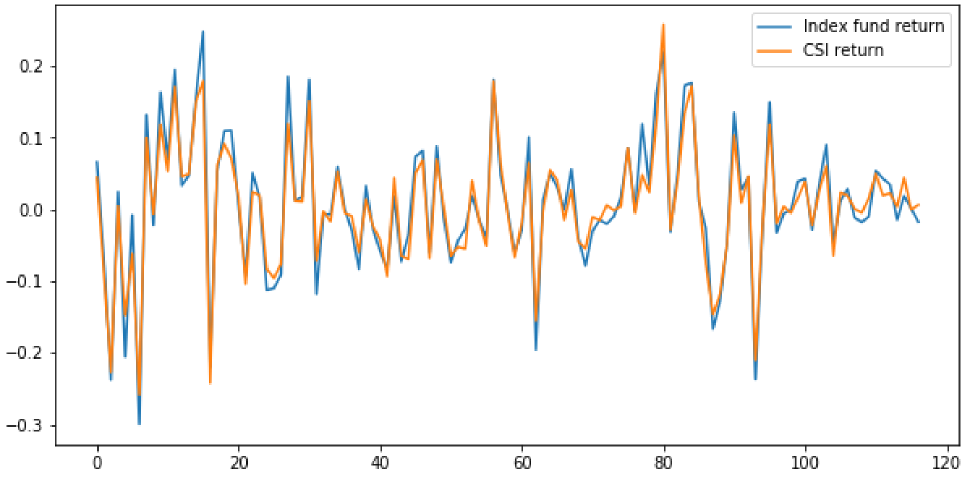

Figure 2 – The tracking performance of constructed index fund in terms of monthly return (using simple historical sample mean and covariance)

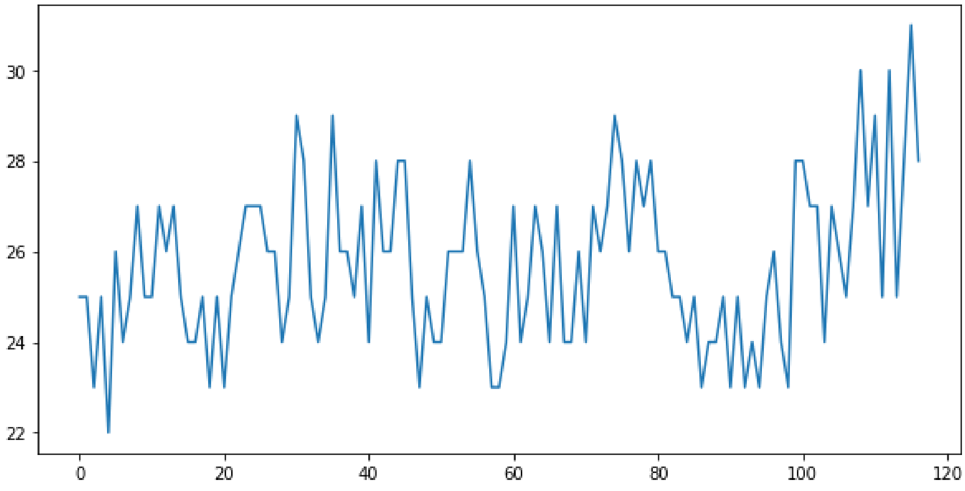

Figure 3 – The number of stocks in the fund with weights more than 0.005 (using factor model and EWMA)

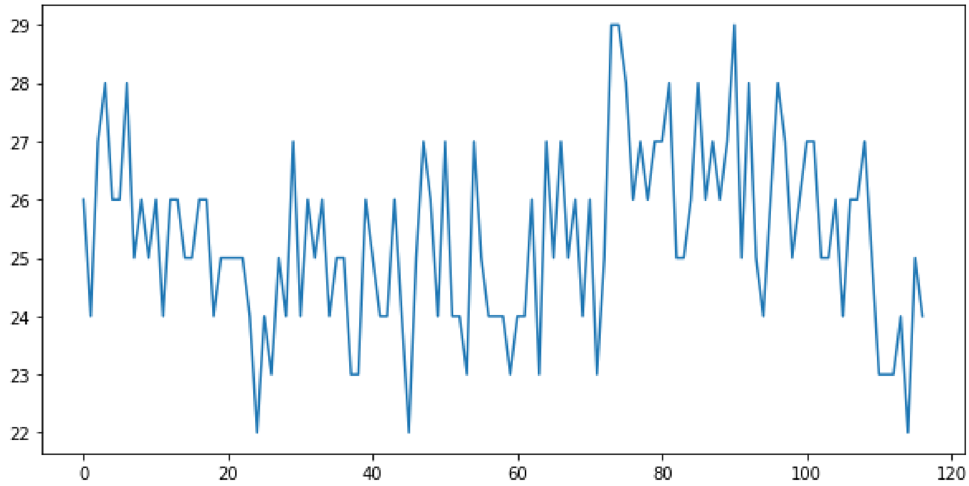

Figure 4 – The number of stocks in the fund with weights more than 0.005 (using simple historical sample mean and covariance)

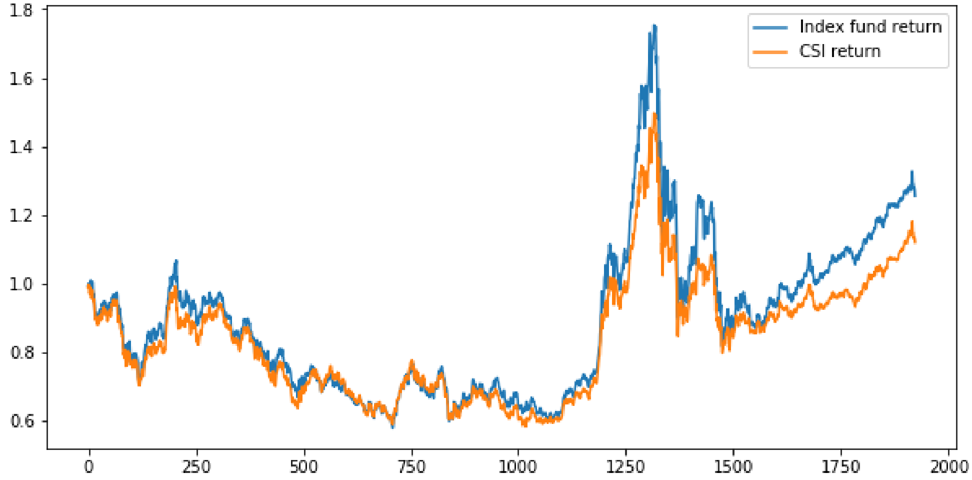

Figure 5 – The tracking performance of constructed index fund in terms of cumulative daily return (using factor model and EWMA)

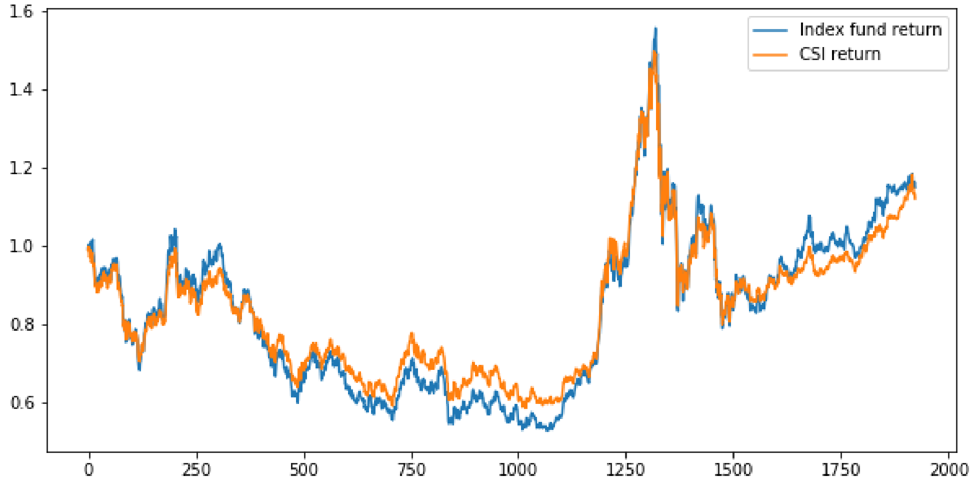

Figure 6 – The tracking performance of constructed index fund in terms of cumulative daily return (using simple historical sample mean and covariance)

Reference

[1] Zakamulin, V. (2015). A Test of Covariance-Matrix Forecasting Methods. The Journal of Portfolio Management, 41(3), 97–108.