Barra-Type Factor Structure

Lulu Wang

September 18, 2019
Overview

1. The Model Set Up
   - Analysis Portfolio with Factor Model
   - Style Exposure

2. Barra USE4 Model
   - Factor Returns
   - Factor Covariance Matrix
The intuition behind the model is that there exists a set of common factors that drive stocks return.

\[ r_n = \sum_k X_{nk} f_k + u_n \]  

where \( r_n \) is the return of stock \( n \).
\( f_k \) is the return of the factor \( k \).
\( X_{nk} \) is the exposure of stock \( n \) to factor \( k \).
\( u_n \) is the stock specific return, which cannot explain by the factors.
Consider a portfolio made by $N$ stocks, a weight of stock $n$ is $w_n$ then return of this portfolio is the weighted average of individual stock returns:

$$R_p = \sum_n w_n r_n$$ \hspace{1cm} (2)

The portfolio’s exposure to factor $K$ is given by the weighted average of the stock exposure, i.e.,

$$X^p_k = \sum_n w_n X_{nk}$$ \hspace{1cm} (3)

Therefore, the return of portfolio can be further expressed as the weighted form of single factor’s return, plus the weighted average of specific return.

$$R_p = \sum_k X^p_k f_k + \sum_n w_n u_n$$ \hspace{1cm} (4)

Here, return is computed daily
Since (1) factor returns are uncorrelated with specific returns. (2) specific returns are uncorrelated among themselves. This allow the variance of the portfolio to be expressed as:

$$\text{var}(R_p) = \sum_{kl} X_k^p F_{kl} X_k^p + \sum_n w_n^2 \text{var}(u_n)$$  \hspace{1cm} (5)$$

where $F_{kl}$ is the covariance between factors $k$ and $l$. 
The Model Set Up—Standardize the Exposure to the Factor

- Example: Exposure to earning factor.

<table>
<thead>
<tr>
<th>Style</th>
<th>Earnings Yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition:</td>
<td>0.68 \cdot \text{EPIBS} + 0.11 \cdot \text{ETOP} + 0.21 \cdot \text{CETOP}</td>
</tr>
</tbody>
</table>

**Components:**

- **EPIBS**
  - Analyst Predicted Earnings-to-Price
  - Earnings-to-price ratio forecasted by analysts.

- **ETOP**
  - Trailing earnings-to-price ratio
  - Computed by dividing the trailing 12-month earnings by the current market capitalization. Trailing earnings are defined as the last reported fiscal-year earnings plus the difference between current interim figure and the comparative interim figure from the previous year.

- **CETOP**
  - Cash earnings-to-price ratio
  - Computed by dividing the trailing 12-month cash earnings divided by current price.

EPIBS, ETOP and CETOP are descriptors of Earnings. We need to standardize descriptors to have a mean of 0 and standard deviation of 1.
The Model Set Up—Standardize the Exposure to the Factor

- We need to standardize descriptors to have mean of 0 and standard deviation of 1.

\[
d_{nl} = \frac{d_{nl}^{\text{Raw}} - \mu_l}{\sigma_l}
\]  

(6)

- Then, combine descriptors into factors.

\[
X_{nk} = \sum_{l \in k} w_l d_{nl}
\]

(7)

Additional to standardization, we also need to address multicollinearity and heteroscedasticity as well. Common methods are VIF (Variance Inflation Factor) and WLS.
In the Barra USE4 model, stock returns are explained by returns of country factor, industry factors and style factors.

\[ r_n = f_c + \sum_i X_{ni} f_i + \sum_s X_{ns} f_s + u_n \]  

(8)

where \( f_c \) is the return of the Country factor.

\( f_i \) is the return of industry factor \( i \).

\( f_s \) is the return of style factor \( s \).

\( X_{ni} \) are the exposures of stock \( n \) to industry \( i \), if the stock \( n \) in the industry \( i \), \( X_{ni} = 1 \), otherwise \( X_{ni} = 0 \).

\( X_{ns} \) are the exposures of stock \( n \) style \( s \).

\( U_n \) is the specific return.

Every stock has an exposure of 1 to the Country factor.
Notice for any stock, sum of industry factor exposures is 1:

$$\sum_i X_{ni} = 1.$$  \hspace{1cm} (9)

which is exactly the USE4 Country factor exposure. A constraint, therefore, must be applied to obtain a unique regression solution.

In USE4, the model adopts a constraint that the cap-weighted industry factor returns sum to zero, which is:

$$\sum_i w_i f_i = 0.$$  \hspace{1cm} (10)

where $w_i$ is the capitalization weight in industry $i$. 
Consider a cap-weighted estimation universe (e.g.: all US stocks), if a weight of stock $n$ is $h^E_n$, then return of this stock market is

$$R_E = f_c + \sum_i w_i f_i + \sum_s X^E_s f_s + \sum_n h^E_n u_n$$  \hspace{1cm} (11)

Because the constrain we put before, the second sum is equal to 0, the standardization makes the third sum equal to 0, and the final term is the sum of specific returns of a broadly diversified portfolio, and is therefore approximately zero, we have

$$R_E \approx f_c$$  \hspace{1cm} (12)
We need solve a constrained Least Squares.

\[
\min \sum_n w_n \cdot (r_n - f_c - \sum_i X_{ni} f_i - \sum_s X_{ns} f_s)^2
\]  \hspace{1cm} (13)

such that \( \sum_i w_i f_i = 0 \). \hspace{1cm} (14)
EXAMPLE: CHINA STYLE FACTORS

- Fundamental Data Related Styles
  - Book-to-Price
  - Earnings Yield
  - Growth
  - Leverage
  - Non-linear Size
  - Size
Factor Returns USE4 — Example 2

EXAMPLE: CHINA STYLE FACTORS (CONT.)

- Market Data Related Styles
  - Beta
  - Residual Volatility
  - Liquidity
  - Momentum
STYLE ANALYSIS - A CHINA LARGE CAP EQUITY FUND

Data Source: Barra Aegis Portfolio Manager

- Large-Cap Equity
  - Positive Size exposure means large cap
  - High Earning Yield and Book-to-Price exposures both explain the value strategy.
Style Analysis - A China Growth Equity Fund

Data Source: Barra Aegis Portfolio Manager

- Growth strategy is shown by positive Growth factor exposure.
- The styles shifted a bit over time, but overall the fund style comply with what it claimed.
**STYLE ANALYSIS - A CHINA SMALL-CAP QUANT FUND**

- Significantly negative **Size** exposure reveals the small-cap strategy.
- Quant strategy normally intentionally has control over style exposures.

Data Source: Barra Aegis Portfolio Manager
Recall the variance of the portfolio is expressed as

\[
\text{var}(R_p) = \sum_{kl} X_k^p F_{kl} X_k^p + \sum_{n} w_n^2 \text{var}(u_n) \tag{15}
\]

Where \( F \) is the factor covariance matrix (FCM) of returns of factors, and \( u \) is the variance matrix of specific returns. The (FCM) predicts the volatilities and correlations of the factors, thus represents a second key pillar for constructing a high-quality risk model.

- How do we know if our risk model is accurate?
- A commonly used measure to assess a risk models accuracy is bias statistic. Conceptually, the bias statistic represents the ratio of realized risk to forecast risk.
Bias statistics is constructed as follows:

- First, assume perfect forecasts. $R_{t+h}$ be the return to a portfolio over period $h$, and let $\sigma_t$ be the beginning-of-period volatility forecast. Then the standardized return,

$$b_{t,h} = \frac{R_{t+h}}{\sigma_t}$$

(16)

has an expected standard deviation of 1.

- The bias statistic for portfolio $n$ is the realized standard deviation of standardized returns,

$$B = \sqrt{\frac{1}{T-1} \sum_{t=1}^{T} (b_t - \overline{b_t})^2}$$

(17)

where $T$ is the number of periods in the observation window.
Therefore

1. If the forecast standard deviation perfectly matches the realized one, we have $B_n = 1$.
2. If we underestimate risk, $B_n > 1$.
3. If we overestimate risk, $B_n < 1$.

Assuming normally distributed returns and perfect risk forecasts, for sufficiently large $T$, $B_n$ is approximately normally distributed with mean approximately to 1. Roughly 95 percent of the observations fall within the confidence interval,

$$B_n \in \left[ 1 - \sqrt{2/T}, 1 + \sqrt{2/T} \right].$$

(18)
The first step is to compute FCM from daily factor returns. The model employs exponentially weighted moving averages (EWMA).

- The idea of EWMA is that recent events have bigger impact on risk.
- This approach gives more weight to recent observations and is an effective method for dealing with non-stationary data.

\[ F_{kl}^{EWMA} = \text{cov}(f_k, f_l)_t = \sum_{s=t-h}^{t} \lambda^{t-s} (f_k^s - \bar{f}_k)(f_l^s - \bar{f}_l)/ \sum_{s=t-h}^{t} \lambda^{t-s} \]  

(19)

where, \( \lambda = 0.5^{1/\tau} \), and \( \tau \) is factor correlation half-life.
Factor Covariance Matrix — Newey-West (deal with auto-correlation)

We calculate the factor covariance matrix from daily factor returns, so we need to consider auto-correlated between factor returns.

- Use the Newey-West approach, we have:

\[
F_{NW}^{NW} = 22[F_{EWNA}^{NW} + \sum_{\Delta=1}^{D} (1 - \frac{\Delta}{D+1})(C_{+\Delta}^{(d)} + C_{-\Delta}^{(d)})]
\]

\[
C_{+\Delta}^{(d)} = \text{cov}(f_{k}^{t-\delta}, f_{l}^{t}) = \sum_{s=t-h+\Delta}^{t} \lambda^{t-s}(f_{k}^{s-\Delta} - \bar{f}_{k})(f_{l}^{s} - \bar{f}_{l}) / \sum_{s=t-h+\Delta}^{t} \lambda^{t-s}
\]

\[
C_{-\Delta}^{(d)} = \text{cov}(f_{k}^{t}, f_{l}^{t-\delta}) = \sum_{s=t-h+\Delta}^{t} \lambda^{t-s}(f_{k}^{s}-\bar{f}_{k})(f_{l}^{s-\Delta} - \bar{f}_{l}) / \sum_{s=t-h+\Delta}^{t} \lambda^{t-s}
\]

where \(D\) is Newey-West correlation lags.
Factor Covariance Matrix — Newey-West (deal with auto-correlation)

Factor covariance matrix parameters for the USE4 model.

<table>
<thead>
<tr>
<th>Model</th>
<th>Factor Volatility Half-Life</th>
<th>Newey-West Volatility Lags</th>
<th>Factor Correlation Half-Life</th>
<th>Newey-West Correlation Lags</th>
<th>Factor VRA Half-Life</th>
</tr>
</thead>
<tbody>
<tr>
<td>USE4S</td>
<td>84</td>
<td>5</td>
<td>504</td>
<td>2</td>
<td>42</td>
</tr>
<tr>
<td>USE4L</td>
<td>252</td>
<td>5</td>
<td>504</td>
<td>2</td>
<td>168</td>
</tr>
</tbody>
</table>
In 1952, Markowitz (1952) established the mean-variance framework for constructing efficient portfolios. This paradigm provided the foundation upon which the modern theory of finance was built.

But Muller (1993) found that risk models have a systematic tendency to under predict the risk of optimized portfolios.

Recently, Shepard (2009) derived an analytic result for the magnitude of the bias.
Under assumptions of normality, stationarity, and many assets (i.e., the large N limit), he found:

$$\sigma_{true} \approx \frac{\sigma_{pred}}{1 - (K/T)}$$

where, $\sigma_{true}$ is true volatility of the optimized portfolio. $\sigma_{pred}$ is predicted volatility from the risk model. $K$ is the number of factors, and $T$ is the effective number of observations used to compute the covariance matrix.

For example: if we estimate the covariance matrix by using 60 trading days and 30 stocks, then the predicted volatility from the risk model is half of the true risk.

An important innovation in the Barra USE4 model is to identify these biases, and to correct the biases by adjusting the factor covariance matrix.
The underestimation of risk for optimized portfolios is closely linked with the concept of eigenfactors.

- For FCM, eigenfactors are the eigenvectors of the matrix;
- They represent uncorrelated portfolios of pure factors.
Let $F_0 = F^{NW}$ denote the $K \times K$ sample factor covariance matrix (FCM), which can be expressed in diagonal form as:

$$D_0 = U_0^T F_0 U_0$$  \hspace{1cm} (24)

where $U_0$ is the $K \times K$ rotation matrix whose columns are given by the eigenvectors of $F_0$.

$$U_0 = [v_1, v_2, \ldots, v_k].$$  \hspace{1cm} (25)

It’s worthy to point out that an element in this matrix, $u_{ij}$, represent a weight of factor $i$ in the portfolio $j$. 

The predicted variances of the eigenfactors (portfolios) are given by the diagonal elements of $D_0$.

$$D_0 = \begin{bmatrix}
var(v_1) & \cdots & \\
\cdots & \ddots & \\
& & var(v_k)
\end{bmatrix}$$
Bias statistics of eigenfactors using the unadjusted covariance matrix (95% CI is indicated by the two dashed horizontal lines) Figure 28

The lower volatility eigenfactors have realized volatilities higher than their predicted volatilities and fall well outside the 95% CI. The larger eigenfactors, by contrast, fall mostly within the confidence interval.
Since we identified biases, next step is to compute how much we need to adjust according them. We estimate the magnitude of the empirical eigenfactor biases via Monte Carlo simulation.

- We consider the sample FCM $F_0$ as the true FCM.
- Step 1: generate a $K \times T$ matrix $b_m$ that simulate eigenfactors (portfolios) return.

\[
b_m = \begin{bmatrix}
  b_{11} & b_{12} & b_{13} & \cdots & b_{1T} \\
  b_{21} & b_{22} & b_{23} & \cdots & b_{2T} \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  b_{k1} & x_{K2} & x_{K3} & \cdots & x_{KT}
\end{bmatrix}
\]

(26)

i.e. $b_1 \sim N(0, \text{var}(v_1))$

That is, the elements of row $K$ of $b_m$ are drawn from a random normal distribution with mean zero and variance given by the diagonal element $D_0(k)$ of matrix $D_0$.  


Step 2: we can generate a set of factor returns for simulation $m$ as:

$$f_m = U_0 b_m$$

(27)

where $U_0$ is the $K \times K$ rotation matrix whose columns are given by the eigenvectors of $F_0$.

We can think of eigenvectors vs in $U_0$ as different portfolios. Recall Each element in a vector $\nu$ is the weight of a corresponding asset. For $b_m$, we can think of it as the return of the asset.
Step 3: Compute estimated FCM:

\[ F_m = \text{cov}(f_m, f_m) \]  \hspace{1cm} (28)

where \( F_m \) is unbiased.

Step 4: Diagonalize the simulated FCM

\[ D_m = U_m' F_m U_m \]  \hspace{1cm} (29)

where \( U_m \) denotes the simulated eigenfactors with estimated variances given by the diagonal elements of \( D_m \), i.e., \( D_m(k) \).
Step 5: Since we know the true distribution that governs the simulated factor returns, we can compute the true FCM of the simulated eigenfactors,

\[ \tilde{D}_m = U'_m F_0 U_m \]  

(30)

Note that since \( U_m \) is not composed of the true eigenfactors, the matrix \( \tilde{D}_m \) is not diagonal. Nevertheless, current focus is on the diagonal elements of the matrix.

Repeat step 1 to step 5 \( M \) times.
Compute the simulated volatility biases according to

\[ v(k) = \sqrt{\frac{1}{M} \sum_{m} \frac{\tilde{D}_m(k)}{D_m(k)}} \]  

Mean simulated volatility bias computed per Equation (33)  
Figure 33
Qualitatively, Figure 33 is in good agreement with Figure 28. However, our simulation assume both normality and stationarity. Real financial data, of course, violate both of these assumptions. In practice, therefore, additional scaling is required to fully remove the biases of the eigenfactors.

\[ v_s(k) = a[v(k) - 1] + 1 \]  

(32)

where \( v_s(k) \) is the scaled value. \( a = 1.4 \) is an empirically determined constant.
We now assume that the sample FCM, $F_0$, which uses the same covariance estimator as the simulated FCM, $F_m$, also suffers from the same biases. Let $\tilde{D}_0$ denote the diagonal FCM whose eigenvariances have been adjusted

$$\tilde{D}_0 = \nu_s^2 D_0$$ (33)

where $\nu_s^2$ is a diagonal matrix whose elements are given by $\nu_s^2(k)$. The FCM in Equation (33) is now rotated from the diagonal basis to the pure factor basis using the sample eigenfactors. That is,

$$\tilde{F}_0 = U_0 \tilde{D}_0 U_0'$$ (34)

where $\tilde{F}_0$ denotes the eigen-adjusted factor covariance matrix.
Bias statistics of eigenfactors using the eigen-adjusted covariance matrix of Equation (34)

Figure 36