Linear models for style analysis

Style analysis: modelli lineari e stili di gestione per fondi comuni di investimento

Francesco Corielli                  Attilio Meucci
Istituto di Metodi Quantitativi     Bain & Company
Università Bocconi, Milano           attilio.meucci@bca.it
francesco.corielli@uni-bocconi.it


Keywords: Linear models, Financial statistics.

1. Sharpe model for style analysis

The analysis of Sharpe (Sharpe 1992) tries to determine a mutual fund composition by observing only the level of the fund performance and the performance of some market indexes. Sharpe’s assumptions are:

1. There exists a set of \( l \) exhaustive, mutually exclusive (in a sense to be defined) easy to track indexes.
2. The prices \( I_1(t), \ldots, I_l(t) \) at time \( t \) of these indexes are publicly available.
3. Each portfolio \( \Pi(t) \) can be seen as the sum of a passive and an active portfolio:
   \[
   \Pi(t) = P(t) + A(t)
   \]  

4. The passive portfolio is a combination of the indexes,
   \[
   P(t) = \sum_{i=1}^{l} n_i(t) I_i(t),
   \]

where \( n_i \) is the unknown number of shares of the \( i \)-th index.

5. The active portfolio at time 0 is a zero-value sum of bets
   \[
   A(t) = \sum_{j \in \text{assets}} m_j(0) S_j(0) \equiv 0,
   \]

where \( m_j \) is the (possibly negative) unknown number of shares of the \( j \)-th security, whose price we denote by \( S \).

Sharpe’s purpose is to fairly evaluate the performance of the manager. In order to do this we have to determine the composition of the passive portfolio at each time \( t \), i.e., the unknowns \( n_i(t) \). Equivalently, defining the relative weights
as follows: \( \omega_i (t) = \frac{n_i(t)I_i(t)}{P(t)} \), we have to determine the unknowns \( \omega_i (t) \), which must satisfy the constraints

\[
\begin{align*}
1 & = \sum_{i=1}^{l} \omega_i (t) \\
0 & \leq \omega_i (t) \leq 1
\end{align*}
\]

By the definition of return of a price \( p \) between time \( t_{k-1} \) and time \( t_k \): \( R_p(t_k) = \frac{p(t_k)}{p(t_{k-1})} - 1 \), it follows that \( R_{P} (t_k) = R_{P} (t_k) = \sum_{i=1}^{l} \omega_i (t_k) R_{I_i} (t_k) \). Therefore Sharpe collects the equally spaced (monthly) time series for \( I_1, \ldots, I_l \) and \( \Pi \), runs the regression

\[
R_{\Pi} (t_k) = \sum_{i=1}^{l} \omega_i R_{I_i} (t_k) + \epsilon (t_k),
\]

and estimates \( \hat{\omega}_i \) using constrained OLS. From the assumptions we have:

6. The \( \omega_i \)'s are constant in time.

7. \( \epsilon (t_k) \) are not correlated with \( R_{I_i} (t_k) \), their expected value is 0 and the \( u_i (t) \) are approximately equal to 1.

2. A criticism of Sharpe model

1. Sharpe’s model is incorrectly specified even in the case where the portfolio manager follows a buy and hold strategy and invests only in indexes, that is \( n_i (t) \) is constant and \( A(t) = 0 \) for all \( t \). In this case, in fact, an exact relation holds:

\[
P (t) = \sum_{i=1}^{l} n_i I_i (t)
\]

and the \( n_i \) can be computed with no error by regressing the index prices with the price of a quota of the fund. Given the numbers of shares it is easy to compute the value weight of the component. On the other hand in this case \( \omega_i (t_k) \) changes with time depending on each index story.

2. In the case of an active strategy, that is a strategy where the fund manager changes the fund composition in time, both the weights and the numbers of shares change in time.

3. In the general case where \( A(t) \) is not identically zero hypotheses 6. and 7. seem quite unlike to hold. Notice, e.g., that a ”successful” fund manager shall show,
on average, positive values of $A(t)$ and that these values shall likely be correlated with changes in $n$ and $I$.

To counter problems 1. and 2. Sharpe suggests to apply a ”window” constrained OLS estimate on returns where the ”weights” at time $t$ are estimate using only data in a time neighborhood of $t$.\footnote{For a detailed analysis of Sharpe model see (DeRoon and TerHost 2000).} We suggest a different methodology.

We specify the model in price levels, as opposed to returns. We first estimate the number of shares $\hat{n}_i(t)$. Then we define the weight estimate

$$\hat{\omega}_i(t) = \frac{\hat{n}_i(t) I_i(t)}{P(t)}.$$ 

The use of prices, instead of returns, seems natural and can help avoiding pitfalls 1. and 2. an clarifying 3. Indeed, by (1), (3) and (2) we see that:

$$\Pi(t_k) = \sum_{i=1}^{l} n_i(t_k) I_i(t_k) + \sum_{j \in \text{assets}} m_j(t_k) S_j(t_k),$$

where $n_i(t_k)$ is unknown. If we suppose that $\sum_{j \in \text{assets}} m_j(t_k) S_j(t_k) = u(t_k)$ is uncorrelated with $\sum_{i=1}^{l} n_i(t_k) I_i(t_k)$ we can estimate $n_i(t_k)$ using Kalman filter with observation equation $\Pi(t_k) = \sum_{i=1}^{l} n_i(t_k) I_i(t_k) + u(t_k)$ and state equation $n_i(t_k) = n_i(t_{k-1}) + e(t_k)$. $u(t_k)$ and $e(t_k)$ are independent errors, $u(t_k)$ can be autocorrelated while $e(t_k)$ is uncorrelated in time. The unknown parameters in the covariance matrices are estimated using nonlinear maximum likelihood. Alternatively we present results from kernel regressions where the model specification is $\Pi(t_k) = \sum_{i=1}^{l} n_i(t_k) I_i(t_k) + u(t_k)$ as in the above observation equation. The $n_i(t_k)$’s at each $t_k$ are estimated using quadratic programming with a kernel weight for time $t_i$ given by $\exp(-(t_i - t_k)^2/(2h))$. The bandwidth $h$ is set to 10 which implies 99% of data weight is concentrated in the $\pm 10$ months interval around $t_k$.

3. A simulation exercise

In order to avoid the problems in point 3. above we decided to compare Sharpe’s methodology to our variant using simulated data. We start from Sharpe’s data on indexes and we suppose a fund manager only invests in these indices. Two situations are compared. First: the manager follows a buy and hold strategy. Second: the fund managers abruptly changes the fund composition in the middle of the sample period. The results are displayed in tables 1 and 2. The numbers correspond to the mean absolute deviation between the true weights of each index (row) in the portfolio and those estimated with each suggested method (column). The indexes are as in (Sharpe 1992). The methods we compare are: Sharpe OLS and Window OLS on returns, OLS on prices, Kalman Filter on prices and kernel regression on prices.

In Table 1 we see how the methods we suggest work better than Sharpe’s methods by several orders of magnitude when the fund manager follows a buy and hold strategy. In Table 2 the best performing method is the kernel regression but all price level based methods perform better than Sharpe’s return models.
Table 1: No jumps in portfolio composition.\(^2\)

<table>
<thead>
<tr>
<th></th>
<th>OLS rets</th>
<th>WOLS rets</th>
<th>OLS price</th>
<th>K.Filt. price</th>
<th>Kernel price</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBills</td>
<td>0.0561(0)</td>
<td>0.0394(0)</td>
<td>0.2388(−3)</td>
<td>0.1623(−12)</td>
<td>0.1197(−3)</td>
</tr>
<tr>
<td>IntBds</td>
<td>0.0690</td>
<td>0.0512</td>
<td>0.1602</td>
<td>0.1685</td>
<td>0.1700</td>
</tr>
<tr>
<td>LngBds</td>
<td>0.0390</td>
<td>0.0324</td>
<td>0.2357</td>
<td>0.1015</td>
<td>0.0852</td>
</tr>
<tr>
<td>CrpBds</td>
<td>0.0621</td>
<td>0.0515</td>
<td>0.1746</td>
<td>0.2316</td>
<td>0.1113</td>
</tr>
<tr>
<td>ValStx</td>
<td>0.0343</td>
<td>0.0250</td>
<td>0.4447</td>
<td>0.1391</td>
<td>0.0512</td>
</tr>
<tr>
<td>MedVal</td>
<td>0.0293</td>
<td>0.0316</td>
<td>0.3408</td>
<td>0.1719</td>
<td>0.0646</td>
</tr>
<tr>
<td>MedGth</td>
<td>0.0468</td>
<td>0.0268</td>
<td>0.1103</td>
<td>0.1232</td>
<td>0.0462</td>
</tr>
<tr>
<td>SmlStx</td>
<td>0.0491</td>
<td>0.0288</td>
<td>0.2382</td>
<td>0.1430</td>
<td>0.0543</td>
</tr>
<tr>
<td>ForBds</td>
<td>0.0347</td>
<td>0.0236</td>
<td>0.1147</td>
<td>0.1096</td>
<td>0.0689</td>
</tr>
<tr>
<td>EurStx</td>
<td>0.0184</td>
<td>0.0153</td>
<td>0.3279</td>
<td>0.0876</td>
<td>0.0632</td>
</tr>
<tr>
<td>JpnStx</td>
<td>0.0434</td>
<td>0.0241</td>
<td>0.3049</td>
<td>0.0420</td>
<td>0.0368</td>
</tr>
</tbody>
</table>

Table 2: Sudden jump: sell 50% of each of (MedGth, SmlStx, ForBds, EurStx, JpnStx) to buy 50% more of each of (IntBds, LngBds, CrpBds, ValStx, MedVal). The change in TBills makes the trade self-financing.

<table>
<thead>
<tr>
<th></th>
<th>OLS rets</th>
<th>WOLS rets</th>
<th>OLS price</th>
<th>K.Filt. price</th>
<th>Kernel price</th>
</tr>
</thead>
<tbody>
<tr>
<td>TBills</td>
<td>0.0452</td>
<td>0.0578</td>
<td>0.0547</td>
<td>0.0493</td>
<td>0.0311</td>
</tr>
<tr>
<td>IntBds</td>
<td>0.0909</td>
<td>0.0596</td>
<td>0.0265</td>
<td>0.0349</td>
<td>0.0268</td>
</tr>
<tr>
<td>LngBds</td>
<td>0.0441</td>
<td>0.0428</td>
<td>0.0113</td>
<td>0.0328</td>
<td>0.0297</td>
</tr>
<tr>
<td>CrpBds</td>
<td>0.0602</td>
<td>0.0542</td>
<td>0.0381</td>
<td>0.0482</td>
<td>0.0380</td>
</tr>
<tr>
<td>ValStx</td>
<td>0.0482</td>
<td>0.0465</td>
<td>0.0511</td>
<td>0.0573</td>
<td>0.0456</td>
</tr>
<tr>
<td>MedVal</td>
<td>0.0760</td>
<td>0.0550</td>
<td>0.0509</td>
<td>0.0590</td>
<td>0.0456</td>
</tr>
<tr>
<td>MedGth</td>
<td>0.0565</td>
<td>0.0374</td>
<td>0.0121</td>
<td>0.0210</td>
<td>0.0117</td>
</tr>
<tr>
<td>SmlStx</td>
<td>0.0792</td>
<td>0.0457</td>
<td>0.0134</td>
<td>0.0322</td>
<td>0.0202</td>
</tr>
<tr>
<td>ForBds</td>
<td>0.0470</td>
<td>0.0186</td>
<td>0.0542</td>
<td>0.0422</td>
<td>0.0260</td>
</tr>
<tr>
<td>EurStx</td>
<td>0.0239</td>
<td>0.0146</td>
<td>0.0139</td>
<td>0.0167</td>
<td>0.0151</td>
</tr>
<tr>
<td>JpnStx</td>
<td>0.0250</td>
<td>0.0256</td>
<td>0.0295</td>
<td>0.0182</td>
<td>0.0207</td>
</tr>
</tbody>
</table>

References


\(^2\)The numbers in brackets are power of 10 to be multiplied by the entries in each column.