Portfolio Construction in a Regression Framework

James Sefton

June 2007
Motivation

The approach can be used to answer the following questions:

- Are the weights in my portfolio the result of one or two abnormal events?
- How well are my portfolio weights estimated? Are some positions critical to the portfolio performance?
- Are my holdings sub-optimal? Is the evidence strong enough to warrant a trade?
- How different are two portfolios? Is one more efficient than the other? Or are they both efficient but just offer different risk-return trade-offs?
- Do I have significant tilts in my portfolio? Are these a result of stock picks or style tilts?
- What are the costs in terms of portfolio efficiency of my investment constraints?
- Is the new information sufficient to warrant a rebalancing?
What is so different about this approach?

♦ **Modified Mean-Variance Problem**: Given a time series of returns to a set of assets, estimate portfolio weights that maximise portfolio return for a given level of risk.

  – **Step 1.** Assume normality, and estimate from the time series data, the mean and covariance matrix of asset returns.
  – **Step 2.** Using an optimiser, estimate the optimal portfolio weights.

♦ Many authors (Jobson and Korkie, 1981; Jorian, 1992; Broadie, 1993; Michaud, 1989, Best and Grauer, 1991) have noted how sensitive the portfolio weights in Step 2 are to sampling errors in Step 1. Michaud coined the term that optimisers are ‘error maximisers’.

♦ The regression approach is 1 step procedure from data to portfolio. We can use all the developed regression diagnostics to analyse our portfolio.
Resampled Efficient Portfolios

- Michaud proposed an *ad hoc* procedure to limit the sampling error problem that has received considerable attention. It depends crucially on imposing a no-shorting constraint on all estimated portfolios.
  - Step 1. From the observed asset returns series, estimate a mean vector, $\mu_0$, and covariance matrix $V_0$ of the asset returns.
  - Step 2. Using the estimated distribution, generate another set of asset returns and re-estimate a new mean vector, $\mu_i$, and covariance matrix $V_i$ of the generated asset return series.
  - Step 3. Estimate a minimum variance and maximum return portfolio from the distribution $N(\mu_i, V_i)$. Also calculate the other 8 ranked decile portfolios with volatility equally spaced between the minimum and maximum.
  - Step 4. Go to step 2 and repeat $n$ times.
  - Step 5. The 10 resampled efficient ranked portfolios are then the simply the average of the $n$ estimated decile portfolios.
The resampled frontiers all lie below the estimated frontier estimated form $\mu_0$ and $V_0$.

These portfolios are averaged to get the 5th decile portfolio.

These portfolios are averaged to get the maximum return portfolio.
Resampling is a shrinkage estimator – Scherer (2004)

♦ The no-short constraint means that
  – average weights of low return/high volatility assets are biased up
  – average weights of high return/low volatility assets are biased down

---

<table>
<thead>
<tr>
<th>Portfolio weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean with no constraints</td>
</tr>
<tr>
<td>Mean with no-short constraint</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Assets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asset 1</td>
</tr>
<tr>
<td>Asset 2</td>
</tr>
<tr>
<td>Asset 3</td>
</tr>
</tbody>
</table>
The Regression Approach – Britten-Jones (1999)

- The issue is not the magnitude of the estimation error *per se* but rather the way it impacts upon the portfolio construction process – Broadie (1993).

- The advantages of a regression based approach is that
  1. It effectively maps returns directly into portfolio weights. A one step procedure!
  2. It is statistically rigorous and has a long pedigree!
  3. There is a substantial toolkit that can be brought to bear on regression problems.
  4. As it is simple, it can be easily extended to look at many practical problems e.g. incorporating priors, forecasts, sensitivity analysis.
The Regression Framework

♦ Define the matrices of asset returns and indicators as

\[
R = \begin{bmatrix}
r_{11} & r_{12} & \cdots & r_{1N} \\
r_{21} & r_{22} & \cdots & r_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
r_{T1} & r_{T2} & \cdots & r_{TN}
\end{bmatrix} \in \mathbb{R}^{T \times N} \quad e_T = \begin{bmatrix} 1 \\
1 \\
\vdots \\
1
\end{bmatrix} \in \mathbb{R}^T \quad e_N = \begin{bmatrix} 1 \\
1 \\
\vdots \\
1
\end{bmatrix} \in \mathbb{R}^N
\]

♦ And the portfolio weights were the least squares estimate in the regressions

\[
\gamma e_T = Rw + \varepsilon
\]

For some constant \( \gamma \) subject to the restriction that \( e_N'w = 1 \)
The Regression Framework

♦ As $\gamma$ varies the regression weights trace out the portfolios on the efficient frontier.

$$\varepsilon'\varepsilon = \sum_{t} (\gamma - \mu_E)^2 + w' \left( R - \mu \right)' \left( R - \mu \right) w$$

$$r_z = f(\gamma)$$
An Asset Allocation Example

♦ To illustrate the technique we build a portfolio of six equity funds of six markets. The returns to the funds are the returns to the relevant Dow Jones Global Index between Oct-96 and Sep-06.

♦ A summary of the fund returns is

<table>
<thead>
<tr>
<th></th>
<th>Annualised Mean Return</th>
<th>Annualised Volatilities</th>
<th>US</th>
<th>Canada</th>
<th>UK</th>
<th>Euro Area</th>
<th>Japan</th>
<th>Asia Pacific</th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
<td>9.6</td>
<td>17.4</td>
<td>1.00</td>
<td>0.74</td>
<td>0.65</td>
<td>0.69</td>
<td>0.34</td>
<td>0.53</td>
</tr>
<tr>
<td>Canada</td>
<td>14.3</td>
<td>18.5</td>
<td>1.00</td>
<td>0.58</td>
<td>0.66</td>
<td>0.42</td>
<td>0.42</td>
<td>0.61</td>
</tr>
<tr>
<td>UK</td>
<td>10.6</td>
<td>16.4</td>
<td>1.00</td>
<td>0.83</td>
<td>0.34</td>
<td>0.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Euro Area</td>
<td>12.2</td>
<td>20.4</td>
<td></td>
<td>1.00</td>
<td>0.38</td>
<td>0.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>3.9</td>
<td>22.6</td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asia Pacific</td>
<td>6.0</td>
<td>20.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
</tr>
</tbody>
</table>
Estimating the efficient frontier

- As we vary $\gamma$ so we estimate the portfolio along the frontier.

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>Minimum Variance $\chi/\beta = 0.16$</th>
<th>Tangency Portfolio $(1+\alpha)/\chi = 23.32$</th>
<th>46.65</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Portfolio Weights</strong></td>
<td>Std. Err</td>
<td>Std. Err</td>
<td>Std. Err</td>
</tr>
<tr>
<td>US</td>
<td>0.269 (0.06)</td>
<td>-0.263 (0.68)</td>
<td>-0.800 (1.37)</td>
</tr>
<tr>
<td>Canada</td>
<td>0.127 (0.05)</td>
<td>1.137 (0.65)</td>
<td>2.155 (1.31)</td>
</tr>
<tr>
<td>UK</td>
<td>0.645 (0.06)</td>
<td>0.656 (0.75)</td>
<td>0.667 (1.50)</td>
</tr>
<tr>
<td>Euro Area</td>
<td>-0.341 (0.06)</td>
<td>-0.010 (0.70)</td>
<td>0.324 (1.40)</td>
</tr>
<tr>
<td>Japan</td>
<td>0.180 (0.03)</td>
<td>-0.150 (0.36)</td>
<td>-0.482 (0.72)</td>
</tr>
<tr>
<td>Asia Pacific</td>
<td>0.121 (0.04)</td>
<td>-0.370 (0.50)</td>
<td>-0.864 (0.99)</td>
</tr>
<tr>
<td><strong>Average Return, $\mu_W$</strong></td>
<td>8.49</td>
<td>17.73</td>
<td>27.03</td>
</tr>
<tr>
<td><strong>Volatility, $\sigma_W$</strong></td>
<td>13.97</td>
<td>20.18</td>
<td>32.41</td>
</tr>
<tr>
<td><strong>Zero-Beta Return, $r_Z$</strong></td>
<td>$-\infty$</td>
<td>0.00</td>
<td>4.26</td>
</tr>
<tr>
<td><strong>Portfolio Sharpe Ratio</strong></td>
<td>N/A</td>
<td>0.88</td>
<td>0.70</td>
</tr>
</tbody>
</table>
Quantifies the observations of Broadie (1993)

Assume a normal distribution for 5 assets. Generate 24 months of return data.
- True Efficient Frontier – the assumed frontier
- Estimated Frontier – the estimated frontier
- Actual Frontier – The achieved performance from the estimated portfolios

Source: reproduced with permission from Mark Broadie (1993), Annals of Operations Research
Leverage Statistics enables the detection of outliers

- Leverage and Influence statistic suggest September 1998 has undue impact. Dropping this one observation increases the underweight position on Asia by 20% and reduce the UK by 10%.
Testing Portfolio Efficiency

For a given $\gamma$, we can test the efficiency of a portfolio $P$ by an F-test of whether the estimated portfolio weights $w=w_E$ are equal to the weights of portfolio $P$.

$$Z(\gamma) = \left( \frac{T - N - 1}{N - 1} \right) \left( \frac{SSR(w = w_p) - SSR(w_E)}{SSR(w_E)} \right) \sim F_{N-1,T-N-1}$$

$$Z(\gamma) \approx -\frac{\left( \frac{(\mu_E - r_Z)}{\sigma_E} \right)^2}{\left( \frac{(\mu_p - r_Z)}{\sigma_p} \right)^2} 1 + \frac{\left( \frac{(\mu_p - r_Z)}{\sigma_p} \right)^2}{\left( \frac{(\mu_p - r_Z)}{\sigma_p} \right)^2}$$

However $\gamma$ is unknown. We therefore propose as a statistic the minimum over all $\gamma$

$$Z = \min_{\gamma} Z(\gamma)$$
If a portfolio $P$ is on the efficient frontier then there exists a zero-beta return, $r_Z$ such that for any asset $i$

$$E(r_i) - r_Z = \beta_i \left( E(r_P) - r_Z \right)$$
Shanken’s Test of Portfolio Efficiency

♦ Hence if a portfolio $P$ is efficient frontier, there exists a zero-beta return, $r_Z$ such that for all assets the constant $\alpha_i$ in the regression

$$r_{it} - r_Z = \alpha_i + \beta_i (r_{Pt} - r_Z) + \varepsilon_{it}$$

is equal to zero.

♦ Hence Shanken proposes that the efficiency of portfolio $P$ can be tested by an F-test of the null hypothesis that the constants $\alpha_i$ are all zero. This test statistic is

$$W(r_z) = \left( \frac{T - N}{N - 1} \right) \left( \frac{\alpha' \Sigma^{-1} \alpha}{1 + S_P^2} \right) \sim F_{N-1,T-N}$$

where $S_P = \left( \frac{\mu_P - r_Z}{\sigma_P} \right)$ is the Sharpe ratio of $P$

where $\Sigma$ is the covariance matrix of the residuals $\varepsilon$, $T$ is the number of periods and $N$ the number of assets.
Shanken’s Test of Portfolio Efficiency

♦ This supposes that the zero beta return, $r_z$, is known.

♦ Shanken therefore proposes taking the minimum of this statistic overall returns $r_z$.

$$W = \min_{r_z} W(r_z)$$

He shows that this minimum exists and gives a strict test of efficiency – in the sense that if portfolio $P$ fails this test, then we can reject the hypothesis that $P$ is efficient.

♦ Note: Finding the minimum amounts to solving a quadratic equation. We shall return to this later.
Theorem

After a great deal of algebra, it is possible to show that

\[ W(r_z) = Z(\gamma) \]

when

\[ r_z = \frac{(1 + \alpha - \chi \gamma)}{(\chi - \beta \gamma)} \]

where

\[ \alpha = \mu'V^{-1}\mu \quad \beta = e_N'V^{-1}e_N \quad \chi = \mu'V^{-1}e_N \]

Hence \( W = Z \). Further we can give this minimum a geometric interpretation in the next slide.
Geometric Interpretation

- The Shanken’s quadratic condition for the minimum can be rephrased as requiring that the line $r_z - P$ be perpendicular to the line $\gamma - P$.

- As we know that angles subtended on the circumference of a circle are $90^0$, the circle with diameter $\gamma - r_z$ connects all ‘close’ portfolios.
Testing the efficiency of investing only in the US

- We can test whether a portfolio invested entirely in the US is efficient, i.e. a null of \( w = [1 \ 0 \ 0 \ 0 \ 0 \ 0] \), at different point on the frontier.

<table>
<thead>
<tr>
<th>( \gamma )</th>
<th>US Portfolio</th>
<th>Japanese Portfolio</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma = \frac{\chi}{\beta} )</td>
<td>0.16</td>
<td>92.83</td>
</tr>
<tr>
<td>( \frac{1+\alpha}{\chi} )</td>
<td>23.32</td>
<td>0.22</td>
</tr>
</tbody>
</table>

| Statistic \( Z(\gamma) \) | 0.55 | 0.0096 | 0.0075 | 1.63 |
| No. of Restrictions, \( P \) | 5 | 5 | 5 | 5 |
| \( \frac{(T-N-1)}{P} Z(\gamma) \) | 57 | 0.99 | 0.77 | 167.9 |
| \( p \)-value | 0.00 | 0.42 | 0.57 | 0.00 |

| Zero-Beta Return, \( r_Z \) | \(-\infty\) | 0 | 6.37 | -3589 |
| Sharpe Ratio of \( w_P \) | N/A | 0.55 | 0.19 | N/A |
| Sharpe Ratio of \( w_E \) | N/A | 0.88 | 0.65 | N/A |
The $Z$-distance induces a geometry

- We map the risk-return space into a return-efficiency space.
- Portfolios are different because they are more or less efficient or offer a different return.
Bayesian Extensions

♦ The approach can be extended very easily by extending the regression

\[ \gamma e_{ext} = R_{ext} \omega + \epsilon \]

where

\[
e_{ext} = \begin{bmatrix} e_T \\ 1 \\ 0 \end{bmatrix} \quad R_{ext} = \begin{bmatrix} R \\ r_{\text{forecast}} \\ V^{\frac{1}{2}}_{\text{Prior}} \end{bmatrix} \quad \text{Var}(\epsilon) = \sigma^2 \begin{bmatrix} I & 0 & 0 \\ 0 & t_0^{-1} & 0 \\ 0 & 0 & t_0^{-1} \end{bmatrix}
\]

♦ The regressors are extended by the forecast returns and the prior covariance matrix. The parameter \( t_0 \) is a measure of confidence in the priors in data units.
Bayesian Extensions 2

♦ The risk matrix is now a weighted average of the prior and the sample risk matrix

\[ V_{ext} \approx \frac{T}{T + t_0} V + \frac{t_0}{T + t_0} t_0 V_{Prior} \]

and the posterior estimates of the returns is

\[ \mu_{ext} = \frac{T}{T + t_0} \mu + \frac{t_0}{T + t_0} r_{\text{forecast}} \]

♦ Further we could find the best portfolio \( w \) subject to a set of constraints on the holdings, e.g. no-short or industry neutral.
Inclusion of priors shrink the portfolios back

We assume a single factor model and the implied equilibrium returns as our prior. As we increase $t_0$ so the shrinkage increases and std errors improve.

<table>
<thead>
<tr>
<th>Portfolio Weights</th>
<th>$t_0 = 520$</th>
<th>$t_0 = 2080$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\gamma = \chi/\beta = 0.17$</td>
<td>$\gamma = (1+\alpha)/\chi = 27.29$</td>
</tr>
<tr>
<td></td>
<td>Std. Err</td>
<td>Std. Err</td>
</tr>
<tr>
<td>US</td>
<td>0.21 (0.03)</td>
<td>0.34 (0.43)</td>
</tr>
<tr>
<td>Canada</td>
<td>0.15 (0.03)</td>
<td>0.41 (0.42)</td>
</tr>
<tr>
<td>UK</td>
<td>0.26 (0.03)</td>
<td>0.24 (0.43)</td>
</tr>
<tr>
<td>Euro Area</td>
<td>0.05 (0.03)</td>
<td>0.20 (0.42)</td>
</tr>
<tr>
<td>Japan</td>
<td>0.18 (0.03)</td>
<td>-0.07 (0.33)</td>
</tr>
<tr>
<td>Asia Pacific</td>
<td>0.15 (0.03)</td>
<td>-0.12 (0.39)</td>
</tr>
</tbody>
</table>

Average Return, $\mu_W$
- $9.10$
- $11.21$
- $9.04$
- $10.03$

Volatility, $\sigma_W$
- $15.84$
- $17.58$
- $16.41$
- $17.29$

Zero-Beta Return, $r_Z$
- $-\infty$
- $0.00$
- $-\infty$
- $0.00$

Portfolio Sharpe Ratio
- $N/A$
- $0.64$
- $N/A$
- $0.58$
We can estimate the cost of an investment constraint

- The constrained frontier lies within the efficient frontier.
- The Z-distance is now between a point of this frontier and its nearest neighbour on the efficient frontier.
Testing for the cost of investment constraints

- We look at two constraints. Investing only in EAFE markets and the imposing a no-short constraint.

<table>
<thead>
<tr>
<th>Statistic Z($\gamma$)</th>
<th>EAFE Portfolio</th>
<th>No Short Constraint</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma = \frac{\chi}{\beta} = \frac{(1+\alpha)}{\chi} \times 100$</td>
<td>0.103 0.0066 0.0047</td>
<td>0.066 0.0024 0.0011</td>
</tr>
<tr>
<td>2 0.0066 0.0047</td>
<td>0.066 0.0024 0.0011</td>
<td></td>
</tr>
<tr>
<td>26.68 1.00 0.30</td>
<td>6.79 0.25 0.1143</td>
<td></td>
</tr>
<tr>
<td>$p$-value</td>
<td>0.00 1.70 1.20</td>
<td>6.79 0.25 0.1143</td>
</tr>
<tr>
<td>Zero-Beta Return, $r_Z$</td>
<td>$-\infty$ 0.00 8.47</td>
<td>$-\infty$ 0.00 -8.30</td>
</tr>
<tr>
<td>Sharpe of Constrained Portfolio</td>
<td>N/A 0.66 0.39</td>
<td>N/A 0.80 1.31</td>
</tr>
<tr>
<td>Sharpe Ratio of Efficient Portfolio</td>
<td>N/A 0.88 0.63</td>
<td>N/A 0.88 1.34</td>
</tr>
</tbody>
</table>
Conclusions

♦ We have rephrased our original questions into the regression framework.

♦ Is there sufficient new information to justify a rebalancing of my portfolio?
  – This can be rephrased as is my current portfolio, statistically different from an optimal portfolio designed using all my current information?

♦ Are the tilts or positions in my portfolio significantly different from the consensus or efficient set of portfolios?
  – This could be rephrased as simply whether my portfolio is significantly different from the efficient set.
  – Or more informatively, what is the confidence I need to have in my forecasts to justify my current positions?
References


