MACHINE LEARNING AND ASSET MANAGEMENT

BENJAMIN BRUDER

ASSET MANAGEMENT BY
LYXOR
AGENDA

PORTFOLIO ALLOCATION PRINCIPLES

TREND ESTIMATION PROBLEM

PITFALLS

RISK FACTORS

DECEMBRE 2016
PORTFOLIO ALLOCATION PRINCIPLES
MODERN PORTFOLIO THEORY (1952)

Expectations over investment universe (i.e. probability distribution of asset prices)

Portfolio optimization problem

Expected returns

Expected risks (volatilities, correlations)

Maximize returns

Minimize risks

Portfolio constraints

Risk aversion

Optimal portfolio weights (optimal Sharpe ratio)

\[ w = \frac{1}{\lambda} \Sigma^{-1} \mu \]
MODERN PORTFOLIO THEORY (1952): PROBLEM LINKED WITH WITH MODERN PRACTICES

Expectations over investment universe (i.e. probability distribution of asset prices)

Expected returns

Expected risks (volatilities, correlations)

\[ \mu \] Very low available information, Strong overfitting bias

\[ \sum \] High dimension, Missing data, Different timezones

Optimization problem

Maximize returns

Minimize risks

Risk aversion

Portfolio constraints

Optimal portfolio weights (optimal Sharpe ratio)

\[ w = \frac{1}{\lambda} \left( \sum^{-1} \mu \right) \]

Potential instability

Statistics

Convex analysis

THE POWER TO PERFORM IN ANY MARKET
TREND ESTIMATION PROBLEM
WHAT IF PARAMETERS ARE CONSTANT?
SIMPLEST POSSIBLE MODELLING

\[
\frac{dS_t}{S_t} = \mu dt + \sigma dW_t
\]

**Average return**

Best estimation: Average long term return

Estimation quality depends on price history total length

\[
\mu = \frac{1}{T} \ln \left( \frac{S_T}{S_0} \right) + \frac{1}{2} \sigma^2 - \frac{\sigma}{T} W_T
\]

<table>
<thead>
<tr>
<th>Periode</th>
<th>Moyenne</th>
<th>Ecart Type</th>
<th>Proba&gt;0</th>
<th>Bruit/Signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1M</td>
<td>10%</td>
<td>54%</td>
<td>57%</td>
<td>541%</td>
</tr>
<tr>
<td>3M</td>
<td>10%</td>
<td>30%</td>
<td>63%</td>
<td>300%</td>
</tr>
<tr>
<td>6M</td>
<td>10%</td>
<td>21%</td>
<td>68%</td>
<td>212%</td>
</tr>
<tr>
<td><strong>1Y</strong></td>
<td><strong>10%</strong></td>
<td><strong>15%</strong></td>
<td><strong>75%</strong></td>
<td><strong>150%</strong></td>
</tr>
<tr>
<td>3Y</td>
<td>10%</td>
<td>9%</td>
<td>88%</td>
<td>87%</td>
</tr>
<tr>
<td>5Y</td>
<td>10%</td>
<td>7%</td>
<td>93%</td>
<td>67%</td>
</tr>
<tr>
<td>10Y</td>
<td>10%</td>
<td>5%</td>
<td>98%</td>
<td>47%</td>
</tr>
</tbody>
</table>

**Risk**

Best estimation: Quadratic variations

Estimation quality depends on overall numbers of returns

\[
\hat{\sigma}^2 = \frac{1}{t_n - t_0} \sum_{i=1}^{n} \left( \frac{S_{t_i} - S_{t_{i-1}}}{S_{t_{i-1}}} \right)^2
\]

<table>
<thead>
<tr>
<th>periode</th>
<th>nb prix</th>
<th>Moyenne</th>
<th>Ecart type</th>
<th>Bruit/Signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>10D</td>
<td>10</td>
<td>14.6%</td>
<td>3.3%</td>
<td>23%</td>
</tr>
<tr>
<td>1M</td>
<td>20</td>
<td>14.8%</td>
<td>2.4%</td>
<td>16%</td>
</tr>
<tr>
<td>3M</td>
<td>60</td>
<td>14.9%</td>
<td>1.4%</td>
<td>9%</td>
</tr>
<tr>
<td>6M</td>
<td>130</td>
<td>15.0%</td>
<td>0.9%</td>
<td>6%</td>
</tr>
<tr>
<td><strong>1Y</strong></td>
<td><strong>260</strong></td>
<td><strong>15.0%</strong></td>
<td><strong>0.7%</strong></td>
<td><strong>4%</strong></td>
</tr>
</tbody>
</table>

In practice, parameters vary with time (sure for volatility, we suppose so for trends…)

In practice, parameters vary with time (sure for volatility, we suppose so for trends…)

THE POWER TO PERFORM IN ANY MARKET
WHICH ARE THE UNDERLYING TRENDS OF THESE PRICES?
HISTORICAL VERIFICATION OF TREND PERSISTENCE

- Distribution of 1M GSCI returns conditionally the past 3M return

<table>
<thead>
<tr>
<th>Trend</th>
<th>Positive</th>
<th>Negative</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eurostoxx 50</td>
<td>1.1%</td>
<td>0.2%</td>
<td>0.9%</td>
</tr>
<tr>
<td>S&amp;P 500</td>
<td>0.9%</td>
<td>0.5%</td>
<td>0.4%</td>
</tr>
<tr>
<td>MSCI WORLD</td>
<td>0.6%</td>
<td>-0.3%</td>
<td>1.0%</td>
</tr>
<tr>
<td>MSCI EM</td>
<td>1.9%</td>
<td>-0.3%</td>
<td>2.2%</td>
</tr>
<tr>
<td>TOPIX</td>
<td>0.4%</td>
<td>-0.4%</td>
<td>0.9%</td>
</tr>
<tr>
<td>EUR/USD</td>
<td>0.2%</td>
<td>-0.2%</td>
<td>0.4%</td>
</tr>
<tr>
<td>USD/JPY</td>
<td>0.2%</td>
<td>-0.2%</td>
<td>0.4%</td>
</tr>
<tr>
<td>GSCI</td>
<td>1.3%</td>
<td>-0.4%</td>
<td>1.6%</td>
</tr>
</tbody>
</table>

No threshold | Threshold = 5% | Threshold = 10% | Threshold = 15%

- Negative trend
- Positive trend
PITFALL: OVERFITTING WITH FEW INFORMATION
ARE STATISTICAL LEARNING CONCEPTS INSIGHTFUL OR MISLEADING IN ASSET MANAGEMENT?

Underfitting vs overfitting (aka bias vs variance)

Typical pattern on a machine learning problem:

Dimension: if we consider thousands of parameters, there has to be some strategies that performed in the past
PORTFOLIO OPTIMIZATION WITH STANDARD ESTIMATORS

OPTIMIZATION WITHOUT CONSTRAINTS

Initial asset universe n asset with (estimated) covariance matrix $\Sigma$

Perform a PCA, obtain independent portfolios, with unit variance. The new covariance matrix is the identity matrix.

Equivalent problem: Allocation on those portfolios which are (drifted) Brownian motions $(W^1, \ldots, W^n)$

Standard estimation of the drift on the interval $[0, T]$:
$$\hat{\mu} = \frac{1}{T} (W_T^1, \ldots, W_T^n)$$

Perform Markowitz optimization problem, i.e. maximize:
$$\alpha' \hat{\mu} - \frac{1}{2} \lambda \alpha' I \alpha$$

Optimal portfolio composition is given by:
$$\alpha^* = \frac{1}{\lambda} \hat{\mu} = \frac{1}{\lambda T} W_T$$

Sharpe ratio of the optimal portfolio:
$$\frac{\alpha^* . \hat{\mu}}{\sqrt{(\alpha^*)' I \alpha^*}} = \frac{\sqrt{\sum_{i=1}^n (W_T^i)^2}}{T \sum_{i=1}^n}$$
WHAT CAN WE “LEARN” FROM A WHITE NOISE?
SUPPOSE THAT ALL OUR ASSETS ARE ZERO-MEAN BROWNIAN MOTIONS

In sample bias:
Maximize the ex post Sharpe ratio of a combination of n Brownian Motions.

Best ex-post portfolio allocation: \( (W_T^1, \ldots, W_T^n) \)
Best Sharpe ratio:
\[
\sqrt{\frac{\sum_{i=1}^{n} (W_T^i)^2}{T}} \sim \frac{\sqrt{X^2(n)}}{\sqrt{T}}
\]

Out of sample bias:
Try n (Brownian) strategies.
Keep the best out of sample performer on a given test set of 4 years.
Best Sharpe ratio:
\[
\max_{i} (X_i) \quad \text{where} \quad X_i \sim \mathcal{N} \left(0, \frac{1}{\sqrt{T}}\right)
\]
A large number of « reasonable » investment strategies that performed in the past

EXAMPLE: INVESTMENT BANK INDICES
POSITIVE RESULT: RISK FACTORS
Can we find variables (price/earning, past returns, volatility…) explaining the covariance structure?
• Built a portfolio weighted from those variables (rescaled…)

Indicator
(Price / Earnings, volatility, Past performance, Market capitalization…)

Portfolio weights

Conversion to weight (e.g. Z SCORE rescaling or ranking)

Multiply weights by asset returns

Factor portfolio returns
See if the factor portfolios explain the returns (regression of given portfolios vs factor portfolio returns)

\[ R_{t}^{\text{Given Portfolio}} = \beta_{\text{factor}1} R_{t}^{\text{factor1}} + \beta_{\text{factor}2} R_{t}^{\text{factor2}} + \ldots + \beta_{\text{factor}n} R_{t}^{\text{factor}n} + \epsilon_{t} \]

Factor Selection: LASSO regression

Dynamic factor Exposure: Ordinary least square on rolling window or Kalman filtering
Example: Hedge fund replication
Defining the right risk factors

» Decomposition of equity portfolio returns
  » Market risk factor
  » Size factor
  » Value factor
  » Momentum factor
  » Low Risk factor
  » Quality factor

» Large proportion of stock returns explained by those factors

» Example: Fund analysis
“Standard predictive regressions fail to reject the hypothesis that the party of the U.S. President, the weather in Manhattan, global warming, El Nino, sunspots, or the conjunctions of the planets, are significantly related to anomaly performance. These results are striking, and quite surprising. In fact, some readers may be inclined to reject some of this paper’s conclusions solely on the grounds of plausibility. I urge readers to consider this option carefully, however, as doing so entails rejecting the standard methodology on which the return predictability literature is built.”
CONCLUSION
CONCLUSION
DATA SCIENCE IS APPEALING, BUT INFORMATION MAY BE TOO SMALL COMPARED TO NOISE

• The general mindset of machine learning (training/validation/test) gives good insights.

• Techniques apply well when studying covariances (time series or cross sectional).

• Need to be parsimonious, especially when estimating expected returns.

<table>
<thead>
<tr>
<th></th>
<th>Bond Scoring</th>
<th>Stock Picking</th>
<th>Trend Filtering</th>
<th>Mean Reverting</th>
<th>Index Tracking</th>
<th>HF Tracking</th>
<th>Stock Class.</th>
<th>Technical Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lasso</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
</tr>
<tr>
<td>NMF</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
</tr>
<tr>
<td>Boosting</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
</tr>
<tr>
<td>Bagging</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
</tr>
<tr>
<td>Random forests</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
</tr>
<tr>
<td>Neural nets</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
</tr>
<tr>
<td>SVM</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
</tr>
<tr>
<td>Sparse Kalman</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
</tr>
<tr>
<td>K-NN</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
</tr>
<tr>
<td>K-means</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
</tr>
<tr>
<td>Testing protocols</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
<td>🙆</td>
</tr>
</tbody>
</table>

😊 = encouraging results
😊 = disappointed results