Finding outperforming managers

Randolph B. Cohen
Harvard Business School
Conventional wisdom holds that:

- Managers can’t pick stocks and therefore don’t beat the market
- It’s impossible to pick winning managers because there is no persistence in performance

Truth:

- Managers can pick stocks but fail because of institutional factors
- Winning managers can be identified in advance but doing so requires much more than simply looking at past average returns
Equity managers underperform S&P 500

Net return S&P 500 index

14.63%  15.40%
Can it be that no one can pick stocks?

- Start with top talent
- Then give them superb training
- Then put in place maximum incentives for hard work and performance

- Then it is claimed that these managers don’t pick stocks any better than someone throwing darts at a Wall Street Journal page!

- This makes no sense
- Reason: it’s not true
Managers can pick stocks

Gross and net performance of equity funds

Typical equity fund

- Fees: 0.79%
- Trading costs: 0.68%
- Cash drag: 0.20%
- Net return: 14.63%

S&P 500 index fund

- Fees: 0.20%
- Trading costs: 0.10%
- Cash drag: 0.20%
- Net return: 14.90%
Why doesn’t edge cover fees?

- Benchmark hugging
  - Research consistently shows institutions make the right calls
  - But timidity reduces their returns

- Long-only fees charge a lot for active management
Comparing fees

- $100 in a typical large-cap U.S. mutual fund costs $1/year (or more)
- Correlations with benchmark often very high
- Equivalent to
  - $90 indexed plus
  - $10 of long-short “bets”
- Fees on decomposed investment:
  - $.09 for the index piece @ 10 b.p./year
- So $.91 buys only $10 of active management
- Equivalent to 9.1% management fee
Can we find a subset who’ll outperform?

- Just because the industry as a whole doesn’t beat the index doesn’t mean there aren’t great managers to be found.

- But “past performance is no guarantee of future performance.”
No persistence in fund performance

Cumulative returns

<table>
<thead>
<tr>
<th>Past return</th>
<th>Future return</th>
</tr>
</thead>
<tbody>
<tr>
<td>-23%</td>
<td>12%</td>
</tr>
<tr>
<td>-8%</td>
<td>11%</td>
</tr>
<tr>
<td>17%</td>
<td>13%</td>
</tr>
<tr>
<td>32%</td>
<td>12%</td>
</tr>
</tbody>
</table>

Past return and future return.
Why is there so little persistence?

- A good manager should still be good a few years later
- But, track records can mislead
- This disguises true persistence
Performance attribution difficulties

- Example: leveraged buyout funds
- Buy small companies with average beta 1.25
- Lever the portfolio 4-to-1
- CAPM fair return would be
  \[ \text{T-Bill} + 5 \times (\text{Market} - \text{T-Bill}) \]
- If T-Bill = 5% and market premium = 7% ...
- 40% / annum is a fair return!
- If the asset is illiquid might demand more
- Many PE firms underperform this benchmark
Steady as she goes

- Many funds invest in illiquid securities
- Establishing valuations is a challenge
- If securities are marked low in up months, high in bad, results will be smoothed
- Volatility and beta can appear far lower than they are likely to be in the future
- Asness et. al. (2001) present evidence that such behavior may be widespread
- These problems can lead to explosive scenarios that are potentially devastating to investors
The pyramid

- Consistently overmark illiquid securities
- Three major benefits:
  - Creates good track record in the short run
  - Increases fees collected
  - “Sells” existing investors’ (including General Partners’) fund holdings to others at high prices
- Key is that the fund keeps growing
- Otherwise disaster is likely
Window dressing

- Standard window dressing story: buy winners at the end of the year to “dress up” the portfolio
- This makes little sense
- If performance is bad, why would the fact that you had lousy performance while owning good stocks?
- “Smart” window dressing means actually making performance look better
- Buy safe stuff - then can imply “numbers were X, and look - we did it it without buying risky assets”
- Musto (1997) shows this is common among money market managers
Where did the returns come from?

- Some strategies pay off a small amount often but have a large loss rarely
- Famously selling “put” options has this property
- Following such strategies can create a spectacular track record right up until the surprise bad event occurs
- Such a track record is hard to distinguish from that of a manager who is generating consistent alpha
- Especially confusing because put-selling strategies are in fact often good strategies
Hatching, matching and dispatching

- Investors are only shown returns of living funds
- Thus all fund companies (hedge funds, mutual funds, funds of funds) may find it optimal to start many funds, then show investors the results of funds that were lucky
- Past returns investors observe are likely to be higher than what they should anticipate in the future
- The extreme of this is “incubation”
- But continually adding new products and marketing the winners works too
“The money management industry in a nutshell”

Cumulative returns

<table>
<thead>
<tr>
<th>Year</th>
<th>S&amp;P 500</th>
<th>Fund</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 year</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td>5 year</td>
<td>140</td>
<td>180</td>
</tr>
<tr>
<td>10 year</td>
<td>380</td>
<td>460</td>
</tr>
</tbody>
</table>
The roach motel

- Bidding up the fund’s own positions

- At the end of any given month, managers have incentive to buy more of what they already own, and not necessarily at the lowest possible price

- Musto, Carhart, Kaniel and Reed (2004) has evidence of this behavior
What maximizes outperformance?

- Better fee structures
- Concentration
- Focus
- Illiquid/overlooked/inefficient markets
- Staying within capacity
Managers can outperform net of fees

- Hedge funds do appear to outperform
- Data is messy, but:
- Even HF skeptics, using
  - Data cleaned of survival and selection biases
  - Recent data to exclude “good old days”
  - Data that excludes many top performers
- Still find 6% gross and 3% net alpha
- Compare fraction of alpha taken by HF and MF managers
Best ideas

- New research shows that the best ideas of managers outperform - by a lot

- This is true of “typical” managers, not just superstars
Best idea portfolio returns

- Top 20: 0.92
- Top 50: 0.72
- Top 100: 0.49
- All with tilt > 5%: 0.43
- All with tilt > 3%: 0.37

Monthly alpha
Expected return gap across holdings

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>4%</td>
<td>1</td>
</tr>
<tr>
<td>12%</td>
<td>1</td>
</tr>
<tr>
<td>20%</td>
<td>2</td>
</tr>
<tr>
<td>30%</td>
<td>1</td>
</tr>
<tr>
<td>50%</td>
<td>2</td>
</tr>
</tbody>
</table>
Return cost of doubling AUM (%)
Best ideas - and worst

- Why do managers add mediocre stocks to “round out” the portfolio?

- Asset gathering - more assets demand more stocks if price impact is to be minimized

- Volatility reduction to improve Sharpe ratio and other measures that have little relevance to diversified investors
Focus

- Focus is about building specialized expertise

- Country, sector, deal type etc.

- Alternative case for opportunistic generalists

- But empirical evidence supports a preference for specialization
Inefficiency

- Research findings are very consistent
- Almost all tested strategies work better in markets/situations that:
  - Are less institutional
  - Are less developed (e.g. emerging markets)
  - Are less followed (e.g. by analysts)
  - Require short selling
  - Require complex operations/data gathering/legal support/etc.
- Need to find the dark little corners of markets
Capacity

- Strong evidence in long-only that smaller funds do better
The Effect of Fund Size in Performance

Log total net assets

Alpha

0 2 4 6 8 10 12 14 16

0 0.1 0.2 0.3

-0.1 -0.2 -0.3
Data is less clear in hedge funds

- Small funds perform about the same as big

- But is this a fair comparison
The company they keep

- Past performance is too blunt a tool to use to pick managers who will win in the future
- How can we sharpen it?
- “Judging fund managers by the company they keep”
- We show a way to identify a group that outperforms their less-skilled peers by 5-10% / year
- This approach works for short-track-record managers as well
Our approach to performance evaluation

- A manager’s stock-picking ability is judged by the extent to which his investment decisions resemble the decisions of managers with distinguished performance records.
- Similar decisions are assumed to be made by managers with similar stock holdings.
  - A manager is skilled if his holdings are similar to those of managers who have done well, and different from those of managers who have done poorly.
  - Example:
    - Two managers with equally impressive past performance
    - Manager 1 holds a lot of Intel, which is held mostly by managers with good track records
    - Manager 2 holds a lot of Microsoft, which is held mostly by managers who have done poorly
    => Manager 1 is likely to be skilled; Manager 2 is likely to have been lucky.
- Can also use changes in stock holdings rather than levels
Measure Based on Levels of Holdings

- $M$ managers, $m = 1, \ldots, M$  
  $\delta_m$ ... reference measure of skill of manager $m$

- $N$ stocks, $n = 1, \ldots, N$  
  $w_{m,n}$ ... weight on stock $n$ in manager $m$'s portfolio

Normalized weights:  
$$ u_{m,n} = \frac{w_{m,n}}{\sum_{m=1}^{M} w_{m,n}} $$

- For each stock $n$, define its quality measure $\bar{\delta}_n$ as  
  $$ \bar{\delta}_n = \sum_{m=1}^{M} u_{m,n} \delta_m $$  
  (average skill of managers holding stock $n$)

- Our performance measure:  
  $$ \delta_m^* = \sum_{n=1}^{N} w_{m,n} \bar{\delta}_n $$  
  (average quality of stocks held by manager $m$)

$\Rightarrow$ Manager quality is a weighted average of the quality of the stocks they hold.
Simulations

- We create an artificial world in which managers have different ability to pick stocks
- Then we run 10,000 simulations of this world
- Managers with better “true” ability have, on average, greater alpha in simulations
- But sometimes managers with skill perform poorly
- Similarly managers with high ability tend to score high on our delta measure
- Key finding: unless simulation runs for decades, delta correlates more highly with true ability than alpha
<table>
<thead>
<tr>
<th>M</th>
<th>(\hat{\alpha})</th>
<th>(\hat{\delta^*})</th>
<th>(\hat{\delta^{**}})</th>
<th>(\hat{\alpha})</th>
<th>(\delta^*)</th>
<th>(\delta^{**})</th>
<th>(\hat{\alpha})</th>
<th>(\delta^*)</th>
<th>(\delta^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.26</td>
<td>0.34</td>
<td>0.35</td>
<td>0.80</td>
<td>0.80</td>
<td>0.82</td>
<td>0.46</td>
<td>0.64</td>
<td>0.65</td>
</tr>
<tr>
<td>100</td>
<td>0.27</td>
<td>0.40</td>
<td>0.42</td>
<td>0.81</td>
<td>0.82</td>
<td>0.85</td>
<td>0.47</td>
<td>0.76</td>
<td>0.77</td>
</tr>
<tr>
<td>300</td>
<td>0.27</td>
<td>0.44</td>
<td>0.45</td>
<td>0.82</td>
<td>0.83</td>
<td>0.85</td>
<td>0.47</td>
<td>0.80</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T=1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.53</td>
<td>0.64</td>
<td>0.63</td>
<td>0.94</td>
<td>0.86</td>
<td>0.87</td>
<td>0.77</td>
<td>0.89</td>
<td>0.90</td>
</tr>
<tr>
<td>100</td>
<td>0.54</td>
<td>0.72</td>
<td>0.74</td>
<td>0.95</td>
<td>0.85</td>
<td>0.88</td>
<td>0.78</td>
<td>0.93</td>
<td>0.95</td>
</tr>
<tr>
<td>300</td>
<td>0.54</td>
<td>0.76</td>
<td>0.78</td>
<td>0.95</td>
<td>0.84</td>
<td>0.88</td>
<td>0.79</td>
<td>0.94</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T=5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.66</td>
<td>0.75</td>
<td>0.75</td>
<td>0.96</td>
<td>0.86</td>
<td>0.88</td>
<td>0.86</td>
<td>0.93</td>
<td>0.93</td>
</tr>
<tr>
<td>100</td>
<td>0.68</td>
<td>0.81</td>
<td>0.83</td>
<td>0.97</td>
<td>0.85</td>
<td>0.89</td>
<td>0.88</td>
<td>0.94</td>
<td>0.96</td>
</tr>
<tr>
<td>300</td>
<td>0.68</td>
<td>0.82</td>
<td>0.86</td>
<td>0.97</td>
<td>0.85</td>
<td>0.88</td>
<td>0.88</td>
<td>0.94</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T=10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.79</td>
<td>0.82</td>
<td>0.83</td>
<td>0.98</td>
<td>0.87</td>
<td>0.89</td>
<td>0.92</td>
<td>0.94</td>
<td>0.95</td>
</tr>
<tr>
<td>100</td>
<td>0.80</td>
<td>0.84</td>
<td>0.87</td>
<td>0.98</td>
<td>0.85</td>
<td>0.89</td>
<td>0.93</td>
<td>0.95</td>
<td>0.96</td>
</tr>
<tr>
<td>300</td>
<td>0.80</td>
<td>0.84</td>
<td>0.88</td>
<td>0.99</td>
<td>0.85</td>
<td>0.89</td>
<td>0.93</td>
<td>0.95</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T=20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.84</td>
<td>0.84</td>
<td>0.85</td>
<td>0.98</td>
<td>0.87</td>
<td>0.89</td>
<td>0.94</td>
<td>0.94</td>
<td>0.95</td>
</tr>
<tr>
<td>100</td>
<td>0.86</td>
<td>0.84</td>
<td>0.88</td>
<td>0.99</td>
<td>0.85</td>
<td>0.89</td>
<td>0.95</td>
<td>0.95</td>
<td>0.96</td>
</tr>
<tr>
<td>300</td>
<td>0.86</td>
<td>0.84</td>
<td>0.88</td>
<td>0.99</td>
<td>0.85</td>
<td>0.89</td>
<td>0.96</td>
<td>0.95</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T=30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.84</td>
<td>0.84</td>
<td>0.85</td>
<td>0.98</td>
<td>0.87</td>
<td>0.89</td>
<td>0.94</td>
<td>0.94</td>
<td>0.95</td>
</tr>
<tr>
<td>100</td>
<td>0.86</td>
<td>0.84</td>
<td>0.88</td>
<td>0.99</td>
<td>0.85</td>
<td>0.89</td>
<td>0.95</td>
<td>0.95</td>
<td>0.96</td>
</tr>
<tr>
<td>300</td>
<td>0.86</td>
<td>0.84</td>
<td>0.88</td>
<td>0.99</td>
<td>0.85</td>
<td>0.89</td>
<td>0.96</td>
<td>0.95</td>
<td>0.96</td>
</tr>
</tbody>
</table>
Empirical tests

- **Data**
  1. Quarterly fund holdings are from the Spectrum Mutual Fund Database, 1980Q1–2002Q2
  2. Monthly stock and fund returns are from CRSP
  3. Intersection of CRSP and Spectrum mutual fund databases

- For each fund and each quarter, we compute $\alpha$ and $\delta$
  - Three versions of $\alpha$ are computed: the CAPM alpha, the Fama-French alpha, and the four-factor alpha of Carhart (1997)
  - Three versions of delta
    - The alphas are estimated using look back period of 12, 24, and fund’s complete return history.
    - 27 different metrics for evaluating funds.
Empirical tests

- Each quarter, funds are sorted into decile portfolios by $\alpha$ and $\delta$.
- Decile portfolio returns (equal-weighted) are tracked over the following three-months.
  - The three-month return series are linked across quarters to form a series of returns on the decile portfolios covering April 1980 - September 2002.
- Single sorts to assess persistence.
- Double sorts to assess incremental information contained in our measures.
- Double sorts with delay to investigate investor response to information contained in our measures.
- Upshot: delta does a lot alone, alpha very little; but together they are most effective.
## Delta predicts outperformance

<table>
<thead>
<tr>
<th>delta</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1.55</td>
<td>-0.95</td>
<td>0.05</td>
<td>-0.76</td>
<td>0.40</td>
<td>-0.56</td>
</tr>
<tr>
<td>2</td>
<td>-1.28</td>
<td>-0.23</td>
<td>-1.08</td>
<td>-0.01</td>
<td>1.18</td>
<td>-0.28</td>
</tr>
<tr>
<td>3</td>
<td>-2.05</td>
<td>1.21</td>
<td>0.58</td>
<td>0.81</td>
<td>1.28</td>
<td>0.37</td>
</tr>
<tr>
<td>4</td>
<td>-0.85</td>
<td>1.20</td>
<td>0.95</td>
<td>1.81</td>
<td>2.29</td>
<td>1.08</td>
</tr>
<tr>
<td>5</td>
<td>1.22</td>
<td>2.27</td>
<td>1.91</td>
<td>2.78</td>
<td>5.41</td>
<td>2.72</td>
</tr>
</tbody>
</table>

| 5-1  | 2.77 | 3.22 | 1.86 | 3.54 | 5.01 | 3.28    |
| t-statistic | 1.57 | 1.66 | 1.09 | 2.00 | 2.66 | 2.06    |
Key insight

- Because our measure is based on holdings, it does not require a track record to be effective.
- Emerging managers can be judged by their similarity-of-approach to seasoned managers.
- Since there is much reason to think managers with less capital are most likely to succeed, and since these managers tend to have short histories, this method is potentially extremely useful.
Conclusions

- Typical money managers can pick stocks
- But Wall St. captures the value they add, leaving little for investors
- Outperformance thus requires identifying better-than-typical managers
- But past performance gives little guidance here

Conclusion: need a more sophisticated approach to selecting managers who will outperform in the future