False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas

Laurent Barras*, Olivier Scaillet* & Russ Wermers**

*FAME, University of Geneva

**University of Maryland
Outline

• Motivations

• Contribution & Results

• False Discovery Rate

• Performance Measurement & Data

• Empirical Results

• Conclusion
Motivations

- On average, the mutual fund industry underperforms

- But do some funds generate differential performance, namely positive or negative alphas?

- Standard approach developed in the literature:
  1. Each fund estimated alpha is tested by computing its $p$-value
  2. A fund is significant if its $p$-value is smaller than a chosen significance level $\gamma$
  3. The number of significant funds provides an estimator of the number of funds with positive or negative performance
Motivations

• Every test on fund alpha is subject to luck
  - A lucky fund is a fund with a significant estimated alpha while its true alpha is equal to zero

• The standard approach implies multiple testing across all funds
  - If $\gamma$ is set to 0.05, the probability of finding at least one lucky fund is much higher than 5%!
  - Grinblatt and Titman (1995): «While some funds achieved positive abnormal returns, it is difficult to ascertain the implications of this for the efficient market hypothesis because of multiple comparison being made. That is, even if no superior management ability existed, we would expect some funds to achieve superior risk-adjusted returns by chance. »

• Therefore, the standard approach cannot account for luck!
Motivations

Test of differential performance among 1’500 funds

*Question 1: Impact of Luck on Performance?*
- At $\gamma=0.05$, 50 funds have positive significant alphas
  - Do all of them truly yield positive alphas?

*Question 2: Variation of the significance level $\gamma$?*
- At $\gamma=0.10$, 80 funds have positive significant alphas
  - Do the 30 new significant funds produce positive alphas?

*Question 3: Comparisons across investment categories?*
- Two categories: Growth and Growth & Income
  - The number of funds with positive significant alphas is identical
    - Is the real performance across the two categories the same?
Contributions & Results

- Measuring the impact of luck on mutual fund performance
- False Discovery Rate (FDR)
  - The proportion of lucky funds among any group of significant funds
  - Easy to compute from the individual fund $p$-values provided by the standard approach
- New methodology to measure the FDR among the best and worst funds
  - The best funds are funds with positive significant alphas (right tail)
  - The worst funds are funds with negative significant alphas (left tail)
- Answering the previous questions by computing the FDR:
  - Across different significance levels $\gamma$ (0.05, 0.10…)
  - Across different investment categories (All, G, AG, GI)
Contributions & Results

Answer to Question 1: Impact of Luck on Performance?

• Luck has a stronger impact on the performance of the best funds

Answer to Question 2: Variation of the significance level $\gamma$?

• As $\gamma$ rises, the FDR among the best funds increases quickly, while the FDR among the worst funds increases slowly

Answer to Question 3: Comparisons across investment categories?

• The AG funds obtain the best performance, while the GI funds generate the worst one
  
  ➢ The standard approach concludes that 7.7% of the GI funds have positive performance. Accounting for luck, none of them can achieve a positive performance. Clearly a False Discovery!!
False Discovery Rate (FDR)

A. The Standard Approach: A Three-Step Procedure

1. Test of differential performance for each fund $i$ ($i=1,\ldots,M$):

   $H_0 : \alpha_i = 0,$
   $H_A : \alpha_i > 0$ or $\alpha_i < 0$

   - $\alpha_i$ is computed with a given asset pricing model
   - The individual $p$-values can be computed with asymptotic theory or bootstrap techniques (Koswoski et al. (2005))

2. Fund $i$ is called significant if its $p$-value is smaller than $\gamma$
False Discovery Rate (FDR)

A. The Standard Approach: A Three-Step Procedure

3. The number of funds with non-zero alphas is estimated by the number $R(\gamma)$ of significant funds

• This approach cannot distinguish between luck and differential performance:

$$R(\gamma) = F(\gamma) + T(\gamma)$$

- Funds with differential performance (i.e. $\alpha_i > 0$ or $\alpha_i < 0$),
- Lucky funds (or False Discoveries)
False Discovery Rate (FDR)

B. The False Discovery Rate: Only One More Step

4. From the fund $p$-values, we simply compute the FDR

- The FDR is defined as the proportion of lucky funds among the significant funds:

$$FDR(\gamma) = E \left( \frac{F(\gamma)}{R(\gamma)} \mid R(\gamma) > 0 \right)$$

- It is a simple extension of the standard approach

  - We can then measure the impact of luck through $\hat{F}(\gamma) = \hat{FDR}(\gamma) \cdot \hat{R}(\gamma)$

  - We can estimate the number of funds with non-zero alphas: $\hat{T}(\gamma) = \hat{R}(\gamma) - \hat{F}(\gamma)$
False Discovery Rate (FDR)

C. FDR among the Best and Worst Funds

- We suggest to use a new methodology designed to measure the proportion of lucky funds among the best and worst funds:

  - $R^+$ is the number of funds with positive estimated alphas, namely the best funds
  - $R^-$ is the number of funds with negative estimated alphas, namely the worst funds
  - Because we use a equal-tailed, two-sided test, we expect that under $H_0$:

\[
FDR^+ (\gamma) = E \left( \frac{F^+ (\gamma)}{R^+ (\gamma)} \bigg| R^+ (\gamma) > 0 \right)
\]

\[
FDR^- (\gamma) = E \left( \frac{F^- (\gamma)}{R^- (\gamma)} \bigg| R^- (\gamma) > 0 \right)
\]
D. Estimation Procedure

- The estimator: \[ \widehat{FDR}_\lambda (\gamma) = \frac{M \cdot \hat{\pi}_0 (\lambda) \cdot \gamma}{\# \{ \hat{p}_i < \gamma \}} = \frac{\hat{F} (\gamma)}{\hat{R} (\gamma)} \]

- \( M \) denotes the number of funds in the population
- \( \pi_0 \) is the proportion of funds with \( \alpha_i = 0 \)
- \( \gamma \) is the significance level
- \( \hat{p}_i \) is the fund \( i \) estimation \( p \)-value
- The estimation procedure is trivial once we have \( \pi_0 \)
False Discovery Rate (FDR)

Histograms of 1’500 fund $p$-values

Funds with $\alpha_i > 0$ or $\alpha_i < 0$

Funds with $\alpha_i = 0$

\[
\hat{\pi}_0(\lambda) = \frac{\#\{\hat{p}_i > \lambda\}}{(1 - \lambda) \cdot M}
\]
A. Performance Measurement

• Baseline asset pricing model (Carhart model):

\[ r_{i,t} = \alpha_i + b_i \cdot r_{m,t} + s_i \cdot r_{smb,t} + h_i \cdot r_{hml,t} + m_i \cdot r_{mom,t} + \varepsilon_{i,t} \]

• Fund \( p \)-values are computed by bootstrap technique
  
  ➢ Kosowski et al. (2005)

• Use of the \( t \)-stat instead of the alpha
  
  ➢ Better statistical properties for the bootstrap (higher order improvements)
  ➢ Reduce the impact of extreme alphas
B. Data

- Monthly returns of U.S. open-end equity funds from CRSP between 1975 and 2002
  - Wermers (2000), Kosowski et al. (2005)

- Investment objectives by Thomson Financial
  - Wermers (2000)

- 1,472 All // 1,025 G // 234 AG // 310 GI funds
Empirical Results

A. The Standard Approach (All funds)

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>0.05</th>
<th>0.10</th>
<th>0.15</th>
<th>0.20</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{R}^+$</td>
<td>52</td>
<td>83</td>
<td>112</td>
<td>139</td>
</tr>
<tr>
<td>$\hat{R}^-$</td>
<td>104</td>
<td>165</td>
<td>234</td>
<td>283</td>
</tr>
</tbody>
</table>

1) Do these significant funds truly yield non-zero alphas?  
2) Are these new significant funds all performing?

- We need to assess the impact of luck, i.e. the proportion of lucky funds among the different groups of significant funds!\(^{16}\)
B. Using the FDR (All funds)

### Best funds

<table>
<thead>
<tr>
<th>( \gamma )</th>
<th>0.05</th>
<th>0.10</th>
<th>0.15</th>
<th>0.20</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \overleftarrow{FDR}^+ )</td>
<td>55.5%</td>
<td>69.5%</td>
<td>76.8%</td>
<td>82.1%</td>
</tr>
<tr>
<td>( \hat{R}^+ )</td>
<td>52</td>
<td>?</td>
<td>83</td>
<td>112</td>
</tr>
<tr>
<td>( \hat{F}^+ )</td>
<td>29</td>
<td>58</td>
<td>87</td>
<td>116</td>
</tr>
<tr>
<td>( \hat{T}^+ )</td>
<td>23</td>
<td>25</td>
<td>25</td>
<td>25</td>
</tr>
</tbody>
</table>

### Worst funds

<table>
<thead>
<tr>
<th>( \gamma )</th>
<th>0.05</th>
<th>0.10</th>
<th>0.15</th>
<th>0.20</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \overrightarrow{FDR}^- )</td>
<td>27.7%</td>
<td>34.9%</td>
<td>37.0%</td>
<td>40.9%</td>
</tr>
<tr>
<td>( \hat{R}^- )</td>
<td>104</td>
<td>?</td>
<td>165</td>
<td>234</td>
</tr>
<tr>
<td>( \hat{F}^- )</td>
<td>29</td>
<td>58</td>
<td>87</td>
<td>116</td>
</tr>
<tr>
<td>( \hat{T}^- )</td>
<td>75</td>
<td>107</td>
<td>147</td>
<td>167</td>
</tr>
</tbody>
</table>

(+92) Constant!
False Discovery Rates among the Best and the Worst Funds

(a) All funds

(b) $G$ funds

The exception!

(c) $AG$ funds

(d) $GI$ funds
### Empirical Results

#### Implications for Mutual Fund Performance Analysis

<table>
<thead>
<tr>
<th></th>
<th>Positive performance $\hat{\pi}^+_A$</th>
<th>Negative performance $\hat{\pi}^-_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All funds</strong></td>
<td>1.9%</td>
<td>19.6%</td>
</tr>
<tr>
<td><strong>G funds</strong></td>
<td>1.5%</td>
<td>18.0%</td>
</tr>
<tr>
<td><strong>AG funds</strong></td>
<td>8.1%</td>
<td>20.3%</td>
</tr>
<tr>
<td><strong>GI funds</strong></td>
<td>0.0%</td>
<td>24.3%</td>
</tr>
</tbody>
</table>

- The negative average mutual fund performance is caused by the poor performance of 20% of the funds!
- AG funds perform well, while GI funds represent a striking evidence of false discovery!!
Conclusion

• Can some funds achieve differential performance?

• To answer this question, we need to measure the impact of luck on performance due to multiple testing

• This is done by using the False Discovery Rate (FDR)
  
  - Proportion of lucky funds among a given group of significant funds
  
  - Straightforward extension of the standard approach developed in the literature: very easy to implement while giving much insight into the individual performance of mutual funds
Conclusion

• Luck has a stronger impact on the performance of the best funds
  - The FDR among the best funds is high and rises quickly
  - The FDR among the worst funds is low and rises slowly

• However, a tiny fraction of funds yield positive performance
  - 1.5% of the All and G funds (more pronounced for AG funds)
  - These funds are located at the extreme right tail of the cross-sectional distribution of alpha
Additional Results

• Is this tiny evidence sufficient to form portfolios of the best funds which generate positive alphas?

• Yes, because the best funds are located at the extreme right tail of the cross-sectional alpha distribution
  ➢ They can be separated from the non-performing funds by setting a low $\gamma$
  ➢ Implications for the selection procedure in the fund of fund industry

• Moreover, the FDR has wide application: it can be used every time a test is run a large number of times
  ➢ Test of predictability
  ➢ Performance of technical trading rules (Sullivan et al. (1999))