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

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

[*]FAME, University of Geneva

[**]University of Maryland

HEC GENÈVE

FAME
INTERNATIONAL CENTER FOR FINANCIAL ASSET MANAGEMENT AND ENGINEERING

Outline

- Motivations
- Contribution & Results
- False Discovery Rate
- Performance Measurement & Data
- Empirical Results
- Conclusion

Motivations

- On average, the mutual fund industry underperforms
  - ➢ Lehman and Modest (1987), Elton et al. (1993), Pastor and Stambaugh (2002)…
- 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 γ
  3. The number of significant funds provides an estimator of the number of funds with positive or negative performance

  ➢ Proposed by Jensen (1968), Ferson and Schadt (1996), Ferson and Qian (2004)

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 γ 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 γ=0.05, 50 funds have positive significant alphas

➢ Do all of them truly yield positive alphas?

Question 2: Variation of the significance level γ?

- At γ=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 γ (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 γ?

- As γ 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,..,*M*):

$$H_0 \quad : \quad \alpha_i = 0,$$
$$H_A \quad : \quad \alpha_i > 0 \text{ or } \alpha_i < 0$$

   - α_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 γ

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(γ) of significant funds

- This approach cannot distinguish between luck and differential performance:
  - R(γ)= F(γ) + T(γ)

Funds with differential performance (i.e. $\alpha_i>0$ or $\alpha_i<0$), [points to T(γ)]

Lucky funds (or False Discoveries) [points to F(γ)]

INTERNATIONAL CENTER FOR FINANCIAL ASSET MANAGEMENT AND ENGINEERING

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:

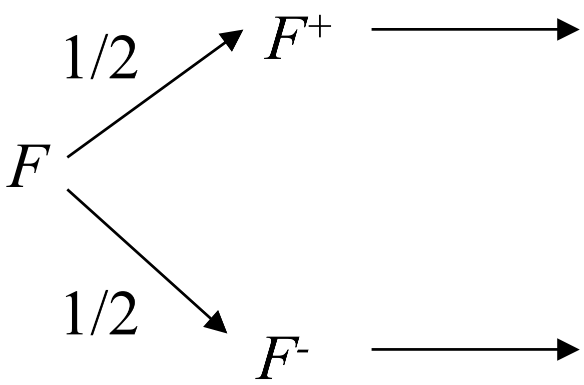
$$FDR(\gamma) = E\left(\frac{F(\gamma)}{R(\gamma)} \middle| R(\gamma) > 0 \right)$$

- It is a simple extension of the standard approach
  - ➢ We can then measure the impact of luck through $\widehat{F}(\gamma) = \widehat{FDR}(\gamma) \cdot \widehat{R}(\gamma)$
  - ➢ We can estimate the number of funds with non-zero alphas: $\widehat{T}(\gamma) = \widehat{R}(\gamma) - \widehat{F}(\gamma)$

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:



$$FDR^+(\gamma) = E\left(\frac{F^+(\gamma)}{R^+(\gamma)} \middle| R^+(\gamma) > 0\right)$$

$$FDR^-(\gamma) = E\left(\frac{F^-(\gamma)}{R^-(\gamma)} \middle| R^-(\gamma) > 0\right)$$

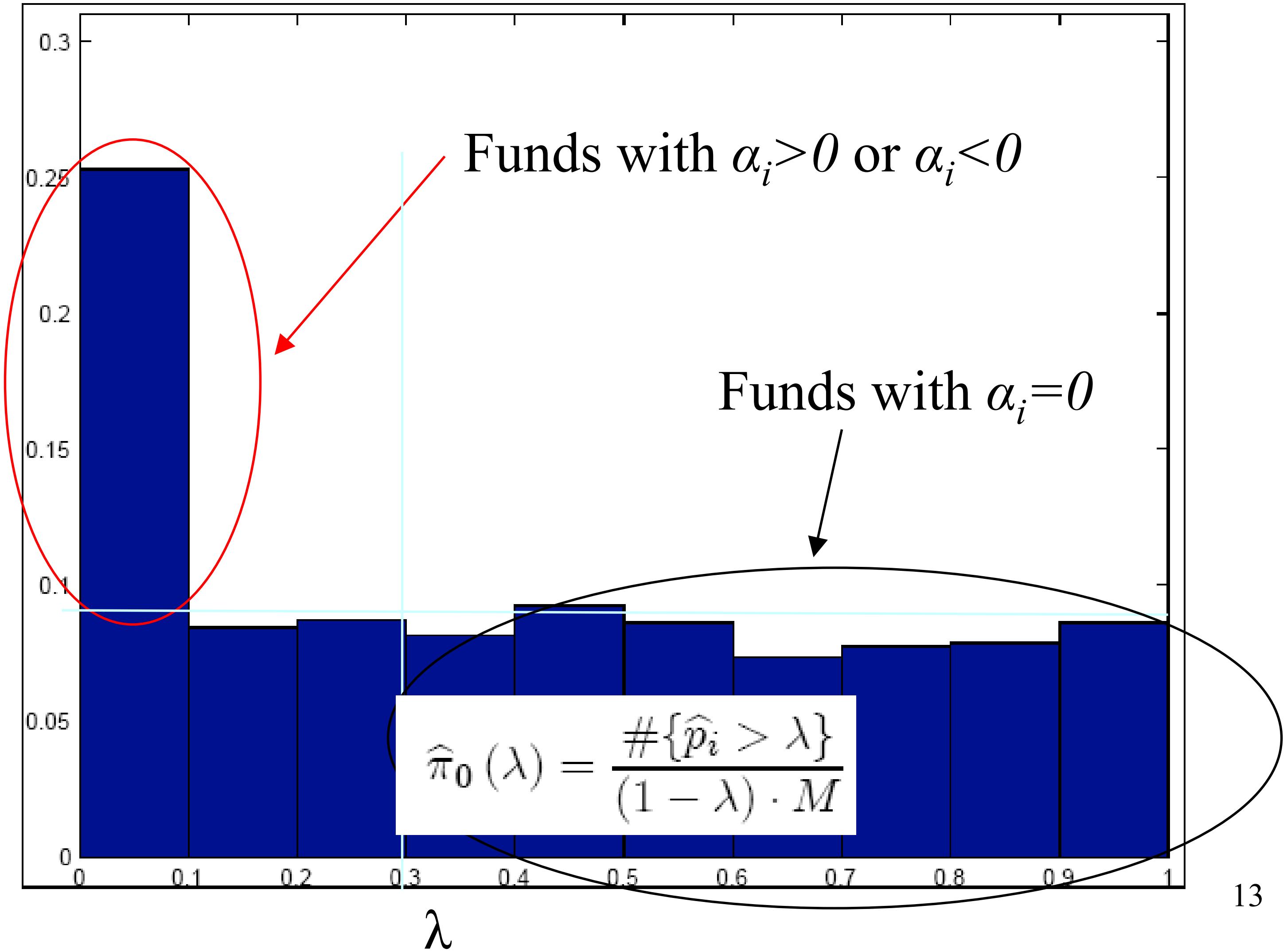
False Discovery Rate (FDR)

D. Estimation Procedure

- The estimator: $$\widehat{FDR}_{\lambda}(\gamma) = \frac{M \cdot \widehat{\pi}_0(\lambda) \cdot \gamma}{\#\{\widehat{p}_i < \gamma\}} = \frac{\widehat{F}(\gamma)}{\widehat{R}(\gamma)}$$
- M denotes the number of funds in the population
- π_0 is the proportion of funds with α_i=0
- γ is the significance level
- $\widehat{p}_i$ is the fund i estimation p-value
- The estimation procedure is trivial once we have π_0

False Discovery Rate (FDR)

Histograms of 1’500 fund *p*-values



Performance Measurement & Data

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

Performance Measurement & Data

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)

γ	0.05	0.10	0.15	0.20
R^+	52	83	112	140
$\widehat{R}^-$	104	165	234	282

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!

B. Using the FDR (All funds)

Best funds

γ	0.05	0.10	0.15	0.20
$\widehat{FDR}^+$	55.5%	69.5%	76.8%	82.1%
$\widehat{R}^+$	52 ?	83	112	139
$\widehat{F}^+$	29	58	87	116
$\widehat{T}^+$	23	25	25	25

+25%

Constant!

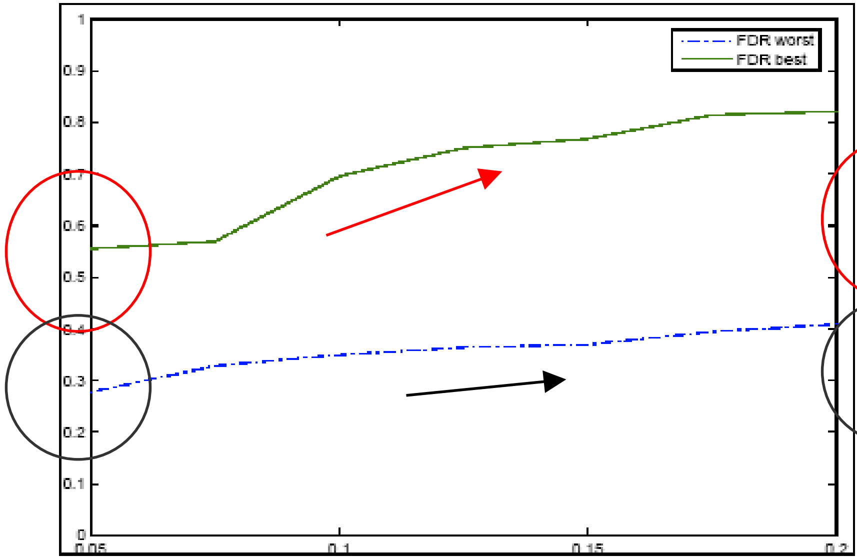
Worst funds

γ	0.05	0.10	0.15	0.20
$\widehat{FDR}^-$	27.7%	34.9%	37.0%	40.9%
$\widehat{R}^-$	104 ?	165	234	283
$\widehat{F}^-$	29	58	87	116
$\widehat{T}^-$	75	107	147	167

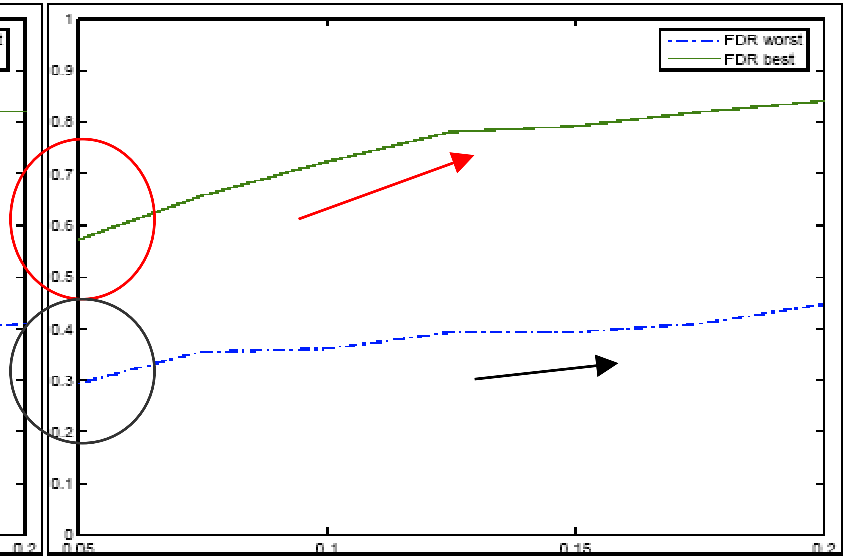
+12%

+7%!

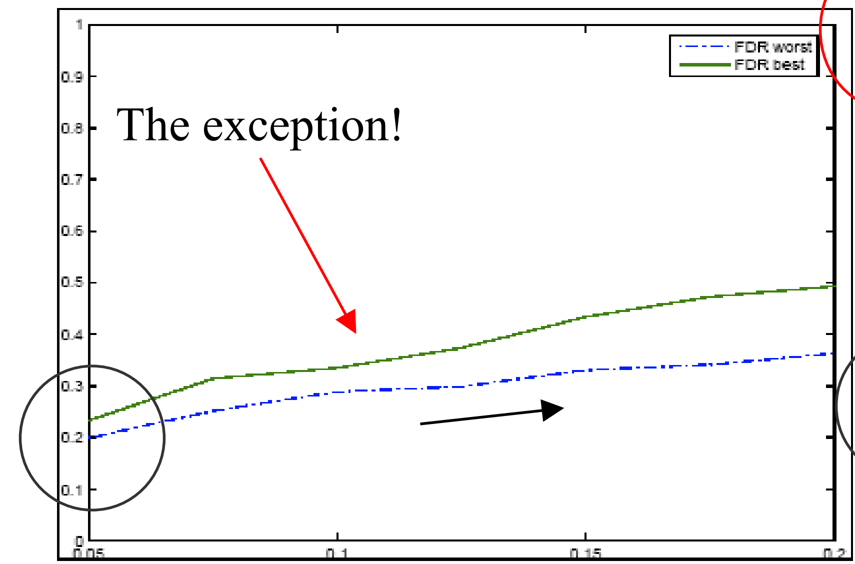
False Discovery Rates among the Best and the Worst Funds



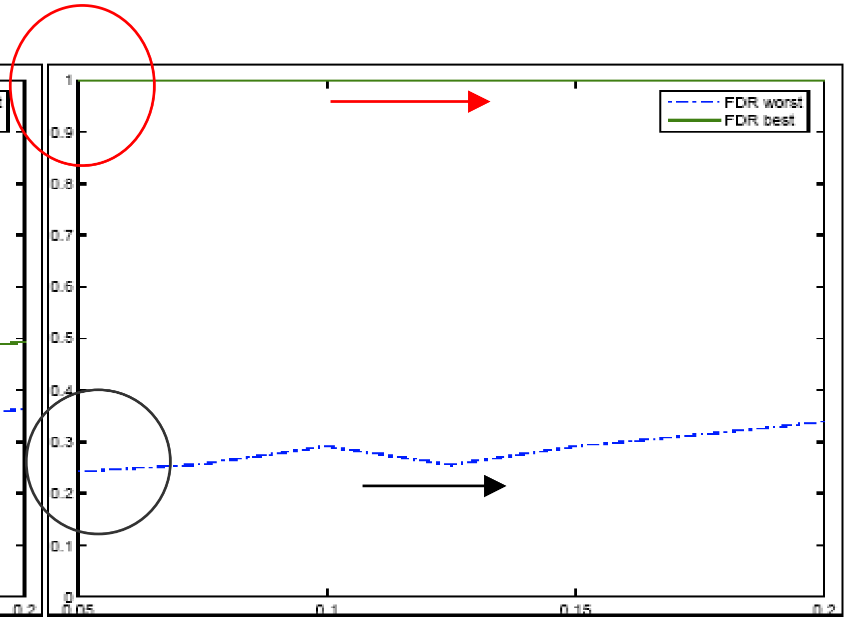
(a) *All* funds



(b) *G* funds



(c) *AG* funds



(d) *GI* funds

Empirical Results

Implications for Mutual Fund Performance Analysis

	Positive performance $\widehat{\pi}_A^+$	Negative performance $\widehat{\pi}_A^-$
All funds	1.9%	19.6%
G funds	1.5%	18.0%
AG funds	8.1%	20.3%
GI funds	0.0%	24.3%

- 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 γ

  - 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))