Accounting for non-normality and luck in **FUND PEER PERFORMANCE** evaluation

Vignette

David Ardia[1], **Samuel Borms**[2], **Kris Boudt**[3]

December 16, 2017

CFE 2017, London

[1]Université de Neuchâtel, Université Laval (david.ardia@unine.ch)
[2]Université de Neuchâtel, Vrije Universiteit Brussel (samuel.borms@unine.ch)
[3]Vrije Universiteit Brussel, Vrije Universiteit Amsterdam (kris.boudt@vub.be)

Important question in practice

Morningstar Rating for Funds

This is a proprietary Morningstar data point.

Morningstar rates mutual funds and ETFs from 1 to 5 stars based on how well they've performed (after adjusting for risk and accounting for sales charges) in comparison to similar funds and ETFs.

Within each Morningstar Category, the top 10% of funds and ETFs receive 5 stars and the bottom 10% receive 1 star. Funds and ETFs are rated for up to three time periods-three-, five-, and 10-years and these ratings are combined to produce an overall rating. Funds and ETFs with less than three years of history are not rated.

Ratings are objective, based entirely on a mathematical evaluation of past performance. They're a useful tool for identifying funds and ETFs worthy of further research, but shouldn't be considered buy or sell signals.

Problem

Peer performance evaluation is not as easy as it seems:

- If all funds have equal performance, the ranking is a random number depending on how lucky the fund is.
- Testing the performance differential can give statistically significant results, even if the true differential is zero. Luck is involved.

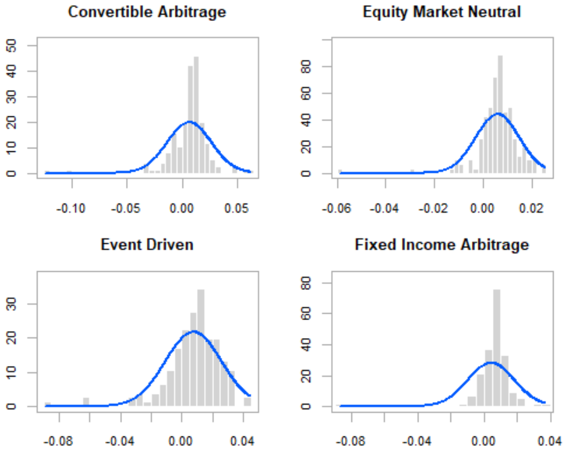
The problem is more generally described as that of **false discoveries**.

- False positive: Detected outperformance, if in fact equal-performance or underperformance
- False negative: Detected underperformance, if in fact equal-performance or outperformance

Performance estimators robust to false discoveries are desired.

Non–normality of returns

Fund returns are non–normal.



Source: Dataset Edhec in PerformanceAnalytics

Alternatives to Sharpe ratio need to be considered.

The peer performance framework as a solution

Ardia & Boudt (2018) propose three new estimators, for each focal fund i:

1. Outperformance ratio (π_i^+)
   the *proportion* of funds fund i outperforms
2. Equal–performance ratio (π_i^0)
   the *proportion* of funds with equal performance as fund i
3. Underperformance ratio (π_i^-)
   the *proportion* of funds that outperform fund i

The emphasis is on the performance of a given fund relative to all other funds (peers) in a universe (e.g. investment style).

Two–step estimation procedure (1/2)

Estimation of the peer performance ratios occurs in two steps.

Step 1: Estimate and test the differential in performance for each pair

- Flexibility to choose the measure the performance of funds: Sharpe ratio, modified Sharpe ratio, information ratio, Jensen's alpha, to name a few.
- Compute $\hat{\Delta}_{i-j}$ as the performance differential between fund i and fund j for one of these measures.
- Perform two–sided test based on studentized t–statistic to get pairwise p–values $\hat{p}_{i-j}$ under H_0: $\Delta_{i-j} = 0$.
- These p–values are uniformly distributed if null hypothesis of equal performance is true, and else below a (low) treshold value λ_i. We thus expect $n_i^0(1-\lambda)$ p–values to exceed λ_i.

Two–step estimation procedure (1/2)

Step 2: Estimate percentage of peer funds with equal performance to get $\hat{\pi}_i^0$

- This gives the estimator:

$$\hat{\pi}_i^0 \equiv \frac{1}{n}\hat{n}_i^0 \equiv \frac{1}{n}c_i^0 \min\left\{\frac{\sum_{j\neq i} I\{\hat{p}_{i-j} \geq \lambda_i\}}{1-\lambda_i}, n\right\}$$

The correction factor c_i^0 accounts for the truncation when ensuring unbiasedness of the estimator.

Based on insights by Storey (2002) and Barras, Scaillet & Wermers (2010).

Two–step estimation procedure (2/2)

Step 2: Attribute $1 - \hat{\pi}_i^0$ to $\hat{\pi}_i^+$ and $\hat{\pi}_i^-$

- Based on the number of performance differences using statistics $\hat{\tau}_{i-j}$ and γ–quantile of the estimated distribution $\hat{q}_{i-j}^{\gamma}$ of $\hat{\tau}_{i-j}$ under the null.
- The percentage of funds being outperformed, π_i^+, is estimated after correction for false positives as follows:

$$\hat{\pi}_i^+ \equiv \frac{1}{n} \max \left\{ \sum_{j \neq i} I\{\hat{\tau}_{i-j} \geq \hat{q}_{i-j}^{\gamma^+}\} - \hat{n}_i^0 (1 - \gamma^+), 0 \right\}$$

- The estimator for π_i^- follows from the sum of the ratios being one:

$$\hat{\pi}_i^- \equiv 1 - \hat{\pi}_i^0 - \hat{\pi}_i^+$$

In sum

- Three new, related, performance metrics that account for false discoveries.
- Closed–form non–parametric estimation.
- Good finite sample properties of estimators.
- Entire procedure may be computationally demanding if many peers (see `R` package).
- Applications beyond fund evalution not hard to imagine (e.g. trading rules, herding, hedging, peer group construction).

PeerPerformance R Package

Package's overview

Package implements the peer universe fund screening as described, while also offering simple performance measurement and comparison functionalities.

Available on `CRAN`.

Straightforward in use; only seven functions:

- Sharpe analysis: `sharpe`, `sharpeTesting` and `sharpeScreening`.
- modified Sharpe analysis: `msharpe`, `msharpeTesting` and `msharpeScreening`.
- alpha screening: `alphaScreening`.

Main interest lies in screening functions.

```
# load package
R> library("PeerPerformance")
```

Function's control arguments

Main control arguments:

- `type`: asymptotic or studentized circular bootstrap (Ledoit & Wolf, 2008).
- `ttype`: testing based on ratio (1) or product (2).
- `hac`: heteroscedastic–autocorrelation consistent s.e. or not.
- `nBoot`: number of bootstrap replications.
- `bBoot`: block length in bootstrap, if `0` optimized.
- `pBoot`: symmetric (1) or asymmetric (2) p–value.

Additional arguments (screening functions):

- `nCore`: number of cores for parallelization.
- `minObs`: minimum number of concordant observations to compute ratios.
- `minObsPi`: minimum number of observations for computing the p-values.
- `lambda`: threshold value to compute $\hat{\pi}^0$, if `NULL` data–driven.

Not all arguments available to all functions yet.

Sharpe screening

Let's start with a hypothethical example.

Simulate 200 low-return funds, 700 medium-return funds and 100 high-return funds, and do Sharpe screening.

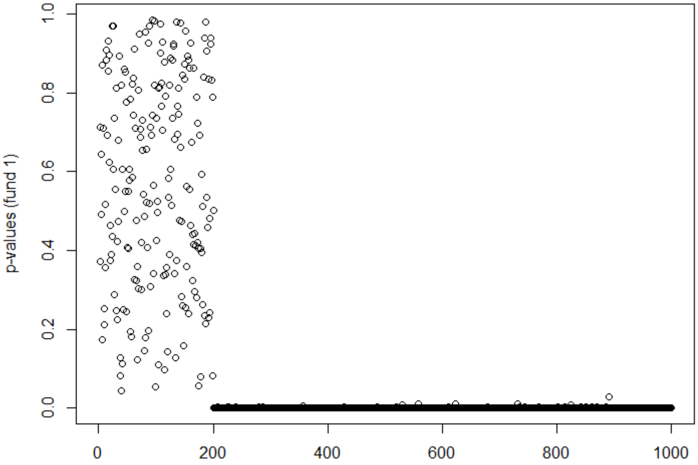
```
# simulate fund data
R> set.seed(123)
R> t <- 60
R> gr1 <- matrix(rnorm(t*200, mean=0, sd=0.01), nrow=t)
R> gr2 <- matrix(rnorm(t*700, 0.01, 0.01), nrow=t)
R> gr3 <- matrix(rnorm(t*100, 0.05, 0.01), nrow=t)
R> data <- cbind(gr1, gr2, gr3)

# do a Sharpe screening
R> ctr <- list(nCore=4, hac=TRUE) # to speed up computations
R> screen <- sharpeScreening(data, control=ctr)
```

Below set of graphs illustrate the difference in behavior of p–values, which is exploited by the peer performance measures.

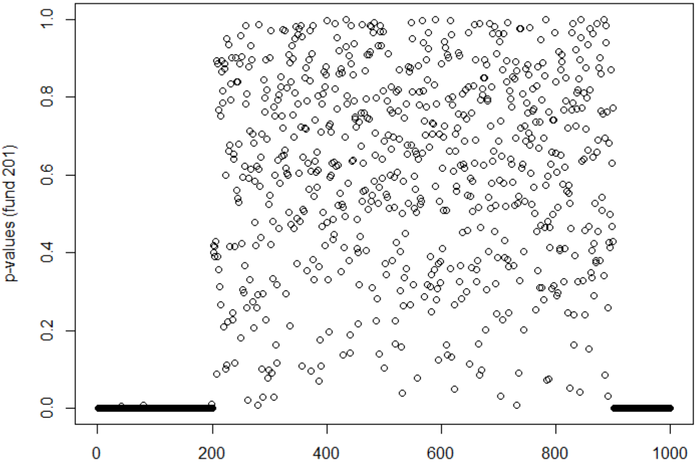
Distribution of p-values (fund 1)

```
R> plot(screen$pval[1, ], ylab="p-values (fund 1)")
```



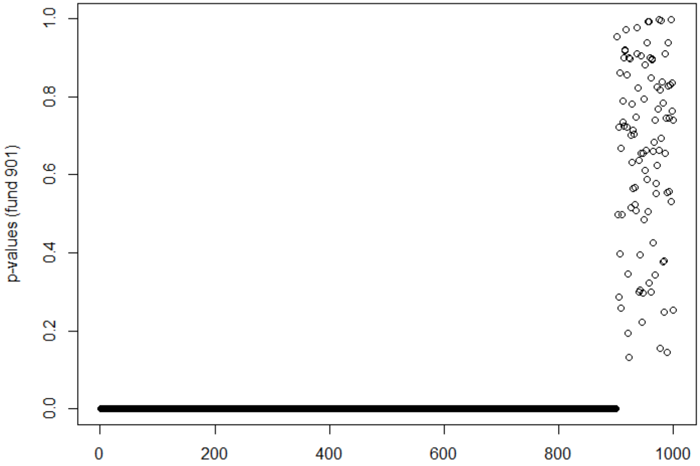
Distribution of p-values (fund 201)

```
R> plot(screen$pval[201, ], ylab="p-values (fund 201)")
```



Distribution of p-values (fund 901)

```
R> plot(screen$pval[901, ], ylab="p-values (fund 901)")
```



Sharpe computation

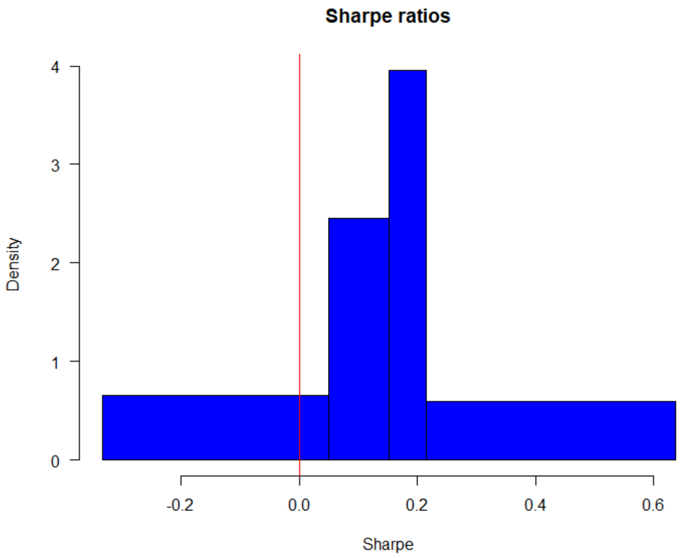
Now on to the built–in dataset of the package.

```
# load data
R> data("hfdata")
R> dim(hfdata)
[1] 60 100
```

Easy calculation of (modified) Sharpe ratios.

```
R> sharpes <- sharpe(hfdata)
```

Distribution of Sharpe ratios



Sharpe comparison and screening

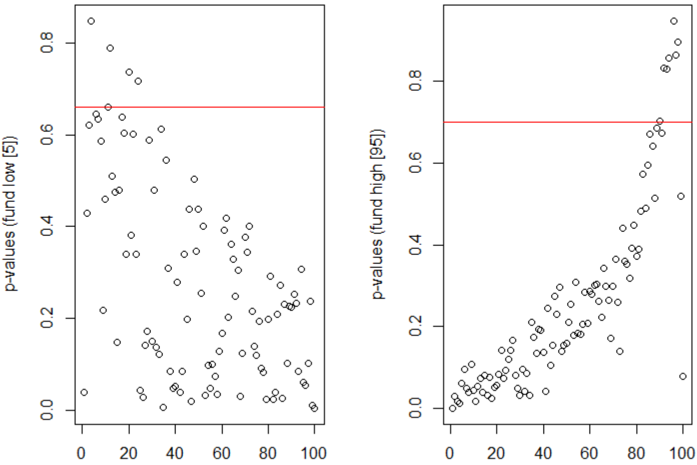
Testing or screening of performance measures requires only one function call.

```
# significance testing of Sharpe ratios
R> hfdataOrd <- hfdata[, order(sharpes)]
R> test <- sharpeTesting(hfdataOrd[, 5], hfdataOrd[, 95])
R> test$pval
[1] 0.01349261

# screening
R> screenS <- sharpeScreening(hfdataOrd, control=ctr)
```

We show the p–values of a low–Sharpe and a high–Sharpe fund.

P-values for low and high Sharpe funds



Alpha screening vs. modified Sharpe screening

Different risk–adjusted performance measures give difference peer performance ratios.

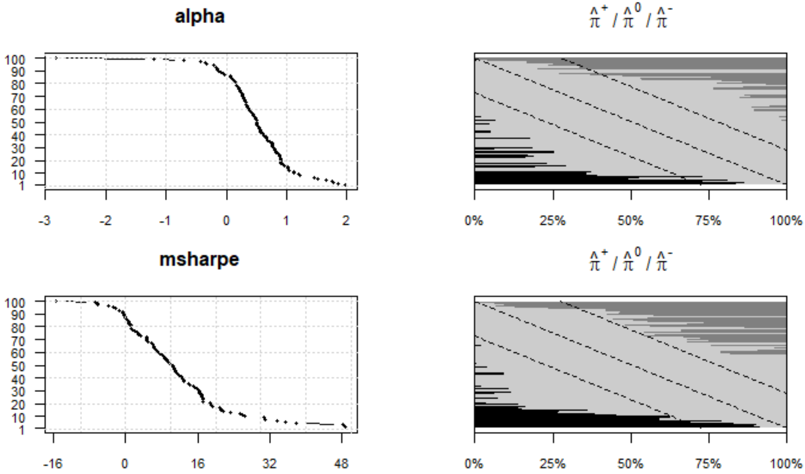
Alpha screening with Fung–Hsieh (2004) factors.

```
R> fh <- tail(factors, dim(hfdata)[1]) # Fung-Hsieh seven factors
R> screenA <- alphaScreening(hfdata, factors=fh, control=ctr)
```

Modified Sharpe screening based on a 95% Value–at–Risk level.

```
R> screenMS <- msharpeScreening(hfdata, level=0.95, control=ctr)
```

Comparison of screening plots



Overestimation

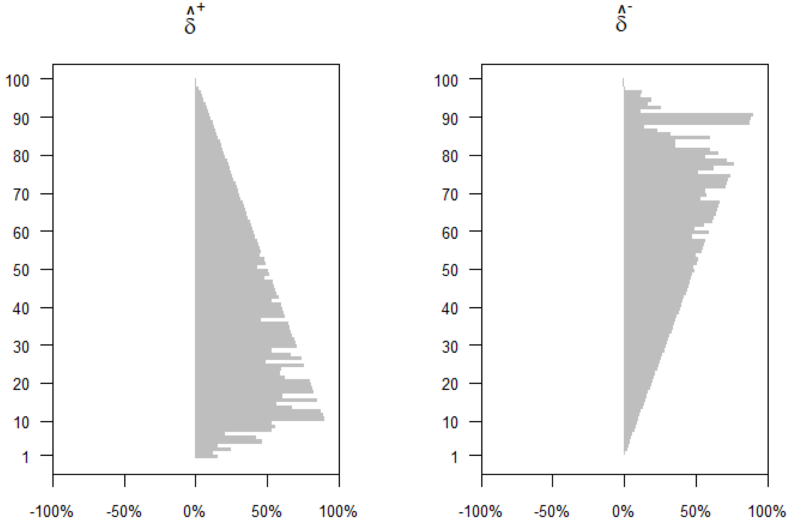
Overestimation of out– and underperformance in rank–based approaches:

$$\hat{\delta}_i^+ \equiv \frac{n-k+1}{n} - \hat{\pi}_i^+ \qquad \hat{\delta}_i^- \equiv \frac{k-1}{n} - \hat{\pi}_i^-$$

```
# funds ranked by percentile
percOrder <- order(screenA$alpha, decreasing=TRUE)
n <- length(percOrder)
percOut <- (n - 1:n + 1)/n
percUnd <- (1:n - 1)/n

# out- and underperformance correction
deltaP <- percOut - screenA$pipos[percOrder]
deltaM <- percUnd - screenA$pineg[percOrder]
```

Overestimation plot for alpha screening



Conclusion

New approach to fund performance evaluation:

- Out– and underperformance ratios robust to luck/false discovery.
- Improved ranking tool, explicitly based on performance relative to others.

Convenient `R` package to implement the peer performance framework. Benefits to reap:

- Implementation of matrix operations in `Rcpp`.
- Add Ledoit & Wolf (2008) bootstrapping for `alphaScreening` function.
- Plotting and enhanced output functionalities.
- Additional performance measures.

References

Ardia & Boudt (2018). "The Peer Performance Ratios of Hedge Funds". *Journal of Banking and Finance* 87, 351–368, doi:10.1016/j.jbankfin.2017.10.014.

Ardia & Boudt (2018). "Accounting for Non-Normality and Luck in Fund Peer Performance Evaluation: The `PeerPerformance` `R` Package". *In preparation*.

Barras, Scaillet & Wermers (2010). "False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas discoveries". Journal of Finance 65(1), 179–216, doi:10.1111/j.1540-6261.2009.01527.x.

Fung & Hsieh (2004). "Hedge Fund Benchmarks: A Risk Based Approach". *Financial Analysts Journal* 60(5), 65–80, doi:10.2469/faj.v60.n5.2657.

Ledoit & Wolf (2008). "Robust Performance Hypothesis Testing with the Sharpe Ratio". *Journal of Empirical Finance* 15(5), 850-859, doi:10.1016/j.jempfin.2008.03.002.

Storey (2002). "A Direct Approach to False Discovery Rates". *Journal of Royal Statistical Society Series B* 64(3), 479-498, doi:10.1111/1467- 9868.00346.