An overview of Mutual Fund Performance Evaluation Studies

by Nik Tuzov

www.stat.purdue.edu/~ntuzov/
Major sources (not all)

- Early studies: Jensen’68, Grossman & Stiglitz’80
- Ippolito 1989
- Grinblatt et al. 1989
- Brown et al’92
- Hendricks et al’93
- Elton et al’93
- Grinblatt & Titman’93
- Malkiel’95
- Carhart’97 (Goldman Sachs Asset Mgmt)
- Russ Wermers, University of Maryland: 1997 – present
  (check out his website: http://www.smith.umd.edu/faculty/rwermers)
Typical issues of interest

- Overall performance of MF industry (e.g., TNA-weighted portfolio of all funds)
- Relative performance of groups of funds sorted by: investment objective, size, expenses, etc.
- The difference between MF overperformance (if any) and fees/expenses
- Persistence in performance
- Survivorship bias (not so important after 1997)
- Cross-correlation in performance measures (hard to determine)
- How to single out good managers (relatively new)
How is the performance measured?

1968 – 1993 : Multifactor model’s alpha:

\[ R_{it} - R_{Ft} = \alpha_i + \beta_1^i F_{1,t} + \ldots + \beta_p^i F_{p,t} + \varepsilon_{it} \]

\( i = 1, N \) – number of funds or portfolios of funds
\( t = 1, T \) - number of time periods

- Examples: CAPM, Fama-French 3-factor, Carhart 4-factor
- Conditional factor model (Ferson & Schadt) : loadings at time t may change based on macroeconomics observed at time (t-1) - corresponds to performing on public information
Multifactor model pitfalls:

- Factor loadings assumed fixed -> possible bias in alpha if they are time-dependent
- Factors are not always well known
- Conditional factor model (Ferson & Schadt) : too many parameters (while the time frequency is monthly at most)
- Too much noise in alpha
- Unknown cross-correlation structure (especially “pure” cross-correlation)
Holdings-based performance measures

Grinblatt & Titman’93

• Idea: for a good manager, today’s portfolio weights have to have a high covariance with tomorrow’s stock returns (move in the same direction)

• The covariance an be estimated as:

\[
\text{Cov}(w_{t-1}, R_t) \approx \frac{1}{T} \sum_{t=1}^{T} GT_t, \\

GT_t = \sum_{j=1}^{N} (w_{j,t-1} - w_{j,t-k-1}) R_{j,t},
\]

\( N \) – number of stocks
\( T \) - number of time periods,
\( k \) - weight lag

• Self-benchmarking: (constant) factor exposure is modeled perfectly
• Statistical arbitrage interpretation: return on a zero initial investment, zero systematic risk portfolio
Holdings-based performance measures

- Clear and easy performance attribution – can be decomposed in many ways

Daniel et al’97 performance decomposition:

Gross return at time $t$: $GT_t = CS_t + CT_t + AS_t$

$CS_t = \sum_{j=1}^{N} w_{j,t-1} (R_{j,t} - R_{Bj,t-1}^{B}) - \text{stock-picking skills}$

$CT_t = \sum_{j=1}^{N} (w_{j,t-1} R_{t-1}^{B} - w_{j,t-k} R_{t}^{B} - w_{j,t-k} R_{t-1}^{B}) - \text{"benchmark-timing" skills}$

$AS_t = \sum_{j=1}^{N} w_{j,t-k} R_{t-1-k}^{B} - \text{average style (no credit to the manager)}$

- $w_{j,t}, R_{j,t} - \text{weight and return of stock j at time t}$
- $B_{j,t} - \text{benchmark portfolio for the stock j at time t (changes with time)}$
- $k$ is taken to be 4 (quarters)

- 125 characteristic-based portfolios (based on the size, B/M, momentum)
Holdings-based performance measures

- Higher precision than multifactor
- Do not have to worry about the non-stock holdings
- Measuring the performance of “new stocks” VS “old stocks”
- Measuring the future performance of today’s “buys” and “sells”
- Measuring “buying on momentum” and “selling on momentum”
- Conditioning on pre-observed macroeconomic variables (similar to conditional multifactor, see Ferson & Khang’02)
How do MF perform, after all?

- Lots of mixed evidence, especially in the early papers

- Sources of inconsistency:

  1. Different size and quality of datasets

  2. The RHS of multifactor model misspecified: Ippolito’89 VS Elton’93, Hendricks’93 VS Carhart’97

  3. Transaction costs (such as price impact) are not observed and have to be estimated

  4. Not performing multiple regression diagnostics (skewed residuals, influential observations etc – see Kosowski, 2006) leads to incorrect p-values for the performance measure

  5. Not considering False Discovery Rate – lots of “garbage” in the identified groups of both “significantly good” and “significantly bad” performers
How do MF perform, after all?

- The performance of an average fund, at best, only covers its fees and expenses
- Superior MF do exist, but should be selected with caution:

  Barras et al, 2006 (used FDR method of Benjamini and Hochberg’95):
  
  - 1456 open-end, domestic equity MF observed on 1975 – 2002
  - 3.8 % of all funds exhibit positive & significant (at 5% level) alphas
  - Half of those 3.8% exhibit “good” alphas due to luck
  - In the subgroup of 310 “Growth and Income” MF 1.6% of funds exhibit positive & significant alphas, but…
  - ALL of those 1.6% exhibit “good” alphas due to luck!