University of Colorado at Boulder
Colorado LEEDS School of Business
BURRIDGE CENTER FOR SECURITIES ANALYSIS AND VALUATION

The Mutual Fund Ratings Game

What is the "Risk" in Risk-Adjusted Performance?

Michael Stutzer
Professor of Finance and Director,
Burridge Center for Securities Analysis and Valuation
University of Colorado
Boulder, CO 80309

Published Fund Ratings

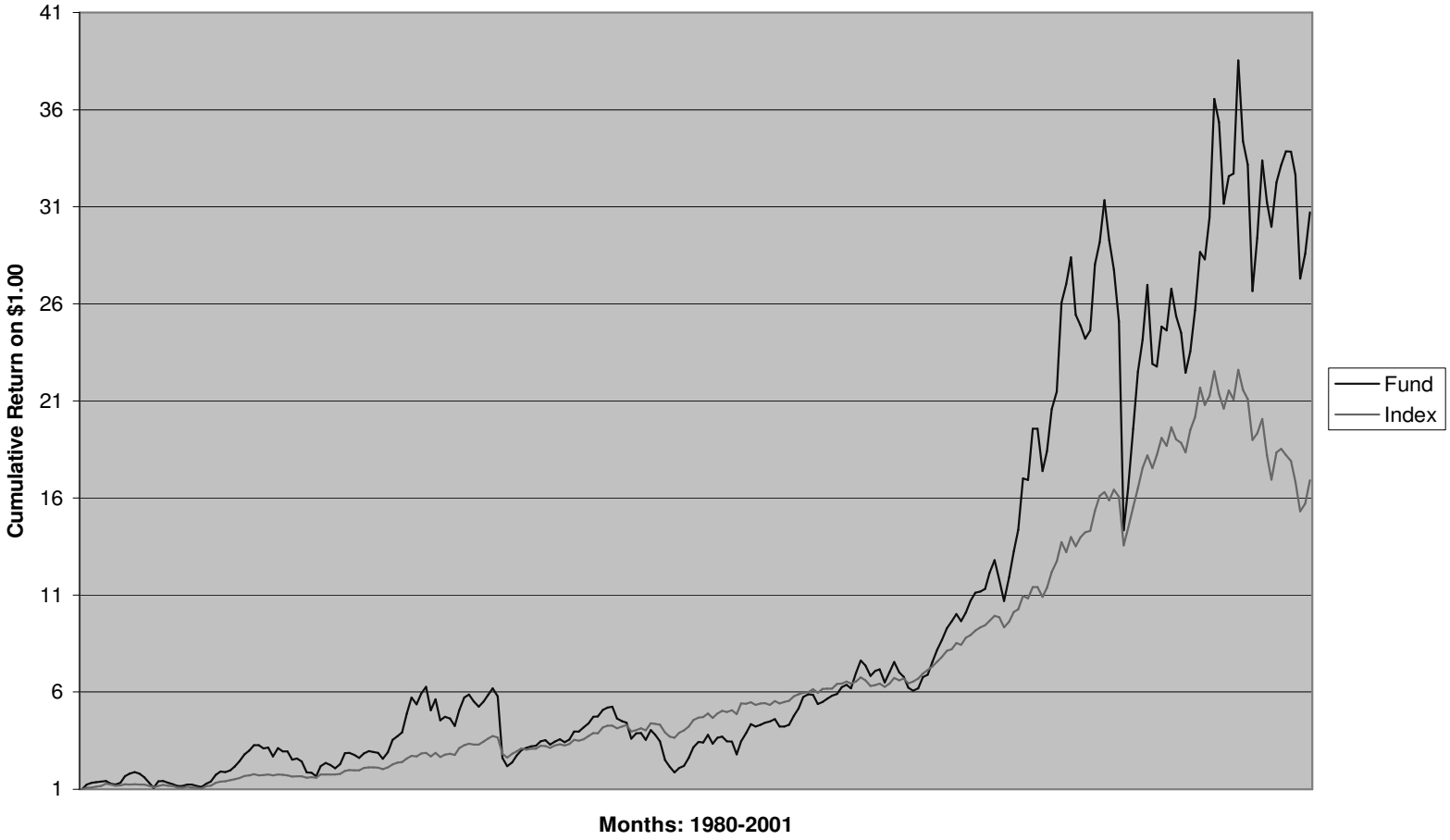
- Morningstar, Lipper, Standard and Poors and Schwab Use Different *Risk-Adjusted* Performance Measures To Rank Funds
  - These Ratings Can Influence Fund Flows; see Del Guercio and Tkac (2001)
  - Large Institutional Investors Also Use Risk-Adjusted Performance Measures, e.g. Sharpe Ratio.
  - The Meaning of These Measures is Obscure
    - What IS the “Risk” That Justifies Reducing the Importance of a Fund’s Historical Cumulative Return?
    - IS this relevant for all, some, or no investors?

Summary of Presentation

- Motivating Example: Equity Fund vs. Index
- Risk of Underperforming a Benchmark
  - Importance of Benchmark Choice
  - An Underperformance Risk-Adjusted Performance Measure: LMIR
- Description of Published Performance Measures
  - Enlightening Approximations
  - Comparison to LMIR
- Practical Problems Plaguing Published Performance Rankings
  - Nonexistent or Irrelevant Benchmarks
  - Historical Average Returns Inaccurately Estimate Long-Run Averages
- Proposed Palliatives
  - Adopt Benchmarks Closely Correlated With Ranked Funds
  - Use Filtering Techniques to Improve Estimation of Long-Run Averages
  - Adopt LMIR, so that the “Risk” Adjustment is for the Risk that a Fund Will Underperform a Relevant Benchmark

Will This Fund Beat The Index In the Future?

Figure 1: Volatile Fund Beats Market Index

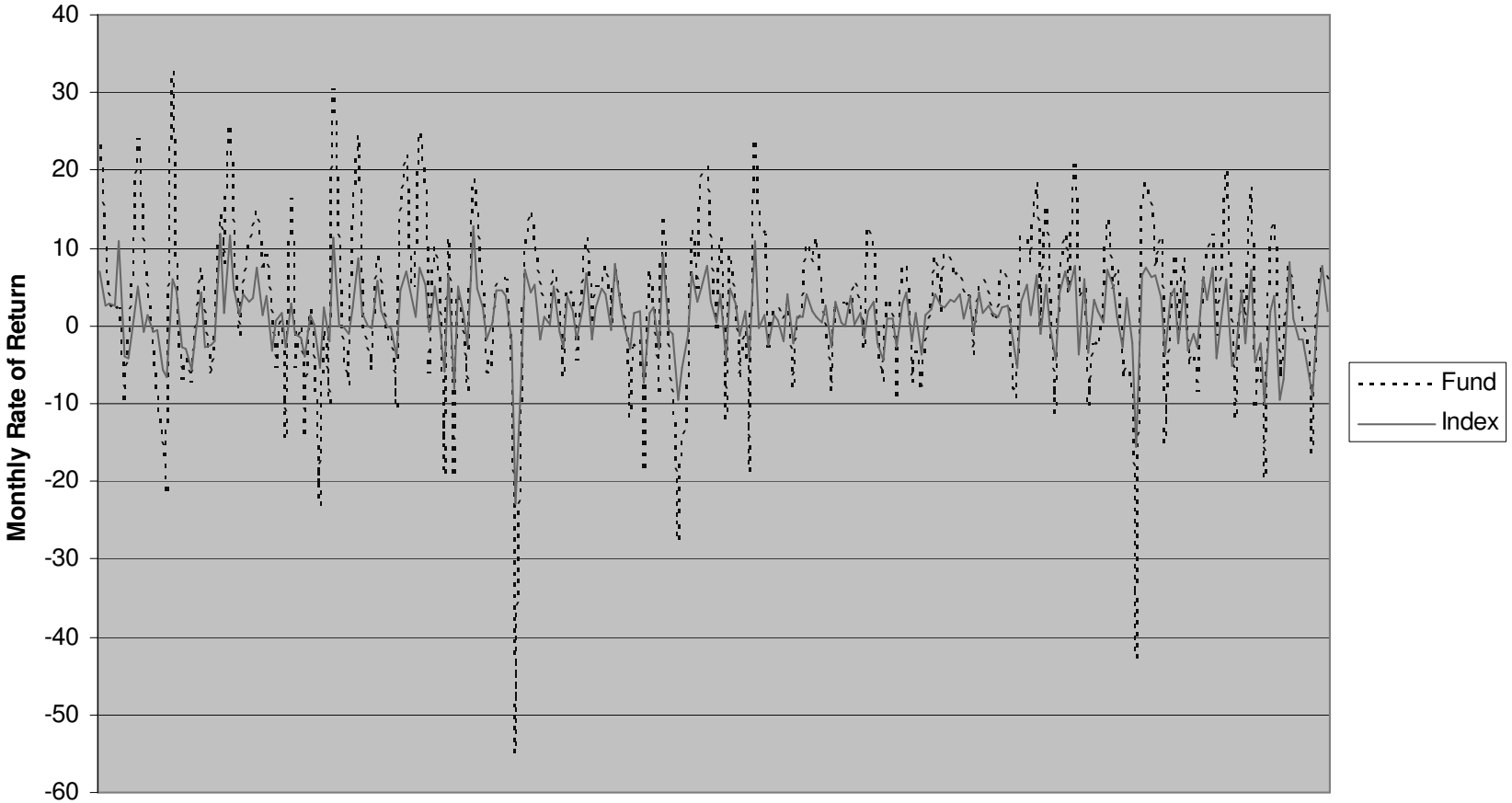


-- Fund Beat Index Historically, But is More Volatile

-- Could Higher Volatility Lead to Underperformance?

Fund is Volatile but Highly Correlated

Figure 2: Volatile Mutual Fund vs. Market Index



Monthly Returns: 1980 - 2001

-- Typifies Funds that Beat Index Historically

-- Fund Will Get Hammered in Bear Markets

Simulating Possible Futures

- Randomly Resample Past Months
- String Resampled Months' Returns Together to Produce Possible Future Cumulative Returns for Fund and Index
  - Repeat Many Times

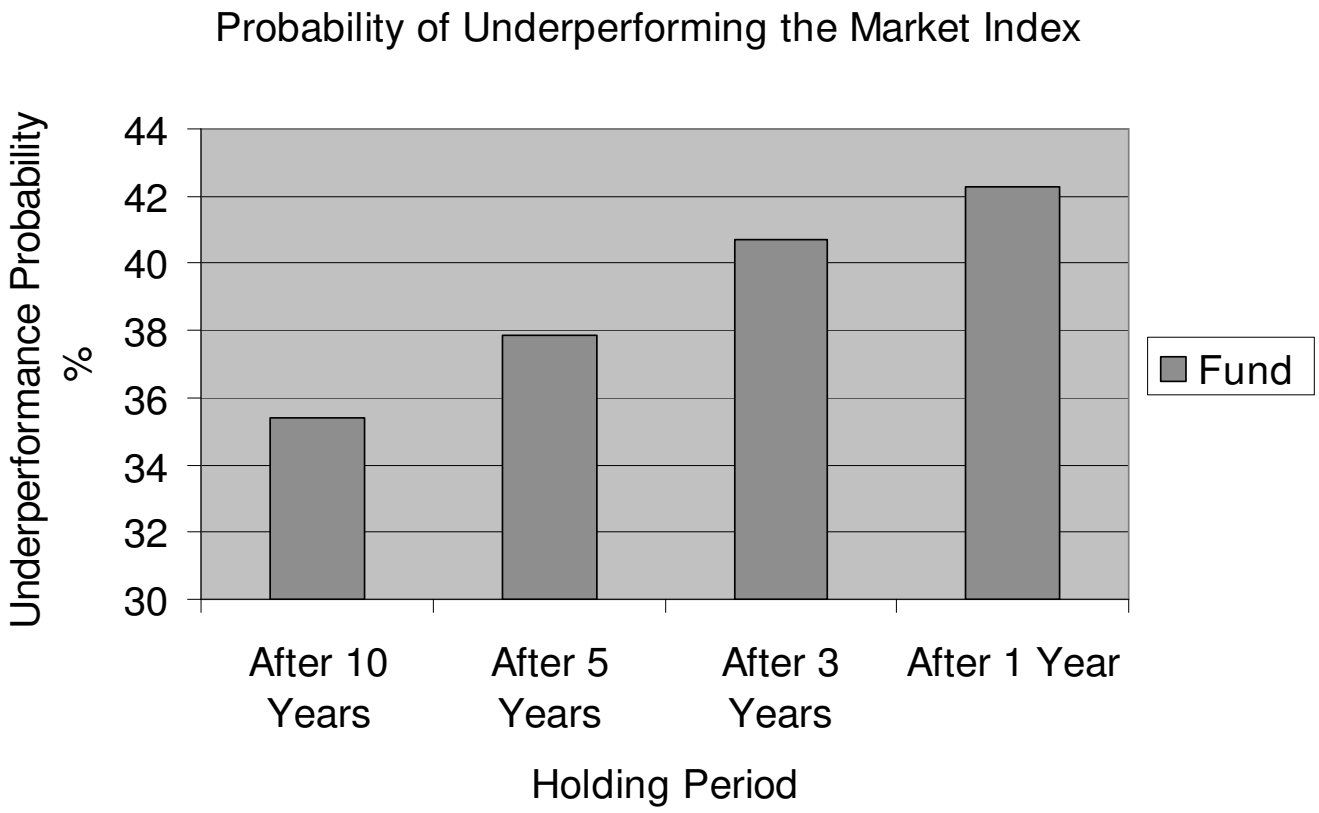
- Compute Fraction of Simulations Where Fund Loses to Index. This Estimates the Underperformance Risk.

<u>Alternatively</u>:

-- One Could Do Fancier "Block" Version of This

-- One Could Build Parametric Model for the Differential Return of the Fund from the Index

Simulated Underperformance Probabilities

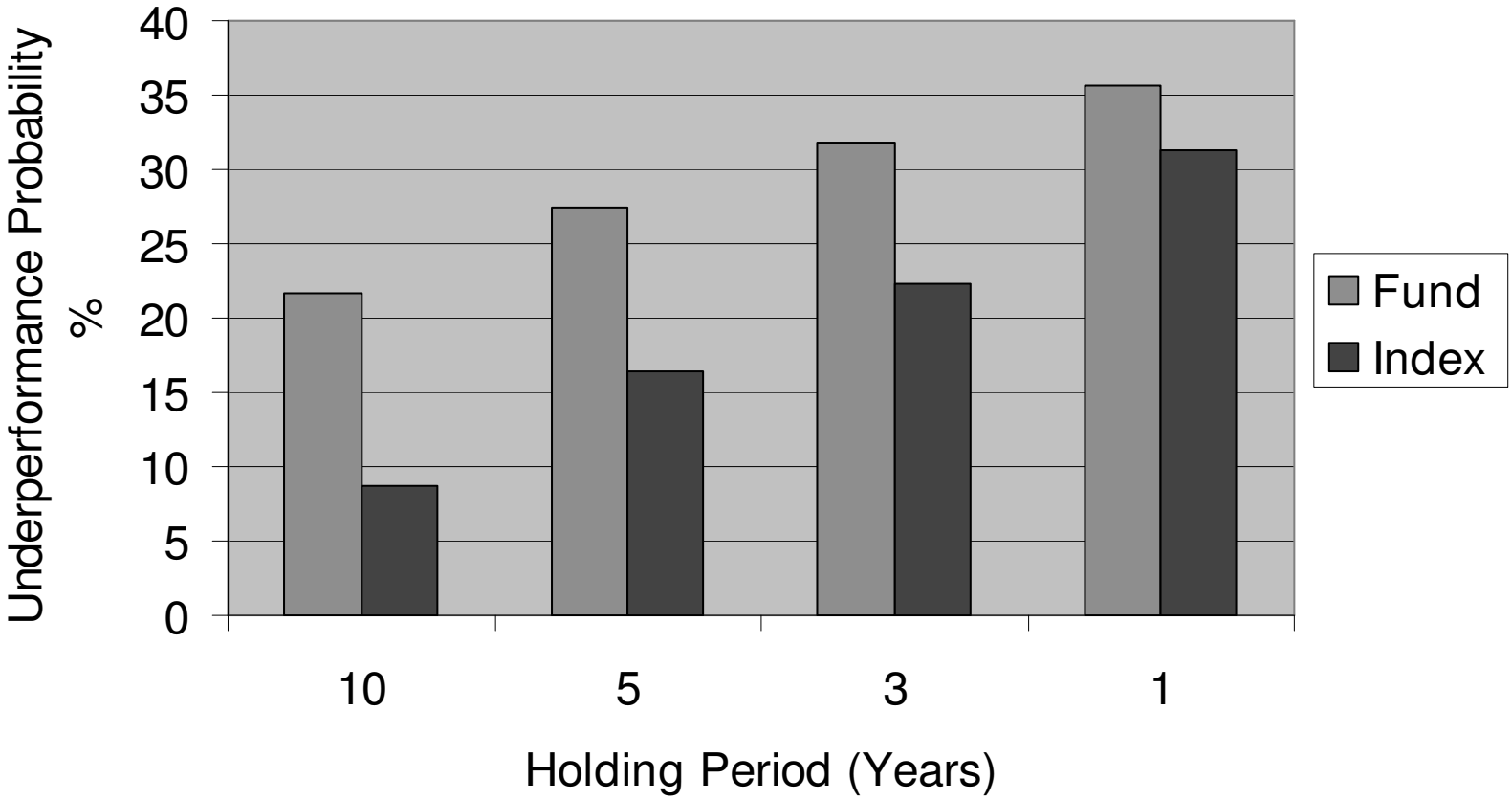


-- So Probability That Fund Beats Index > ½

-- <u>But</u> Underperformance Prob Persists For a Long Time!!

Compare Fund and Index to T-Bills

Figure 4 : Probability of Underperforming T- Bills



- If Investor Benchmark is T-Bills, Then the Index is Better than the Fund
  -- Index is Better Because Its Underperformance Probs. Decay Faster

Outperformance Ranking: Lesson 1

- **The Choice of Benchmark is Critical**
  - Fund Has a Better Than Even Chance of Beating an Index Benchmark
  - But Index Has a Higher Chance of Beating a T-Bill Benchmark Than the Fund Does
- **Significant Underperformance Probabilities Persist Over Surprisingly Long Holding Periods**
  - But the Probs. Decay to Zero as the Holding Period Lengthens to Infinity.
  - So it is Important for the Underperformance Probabilities Decay to Zero Quickly (i.e. a High Rate of Decay) as the Holding Period Lengthens.
- **An Underperformance Risk-Adjusted Performance Measure Will Be Directly Related to the Underperformance Probability Decay Rate.**

Underperformance Risk-Adjusted Performance Ranking Measure:

Let $\text{Log}[1+R^b]$ = Log Gross Monthly Return From Investor's Benchmark

Rank Funds In Accord With the Size of:

$$\frac{\textbf{Average Log}[1+R^{fund}] - \textbf{Average Log}[1+R^b]}{\textbf{Std.Deviation}(\text{Log}[1+R^{fund}] - \text{Log}[1+R^b])}$$

= ***Log-Modified*** **Information Ratio**

Log-Modified Information Ratios

$$\frac{\text{Average Log}[1+R^{fund}] - \text{Average Log}[1+R^{index}]}{\text{Std.Deviation}(\text{Log}[1+R^{fund}] - \text{Log}[1+R^{index}])}$$

$= (1.35\% - 1.10\%) \div 7.83\% = 3.2\% > 0$

So The Fund Will Eventually Outperform the Index, i.e. Its Underperformance Probs Decay to Zero as Holding Period Lengthens. If the Number Were Larger than 3.2%, the Probs Would Decay Faster.

$$\frac{\text{Average Log}[1+R^{index}] - \text{Average Log}[1+R^{bill}]}{\text{Std.Deviation}(\text{Log}[1+R^{index}] - \text{Log}[1+R^{bill}])}$$

$= (1.10\% - 0.54\%) \div 4.56\% = 12.4\%$

versus the fund's:

$(1.35\% - 0.54\%) \div 11.55\% = 7.0\%$

So With a T-Bill Benchmark, The Index is Ranked Higher Than the Fund

Comparison to Sharpe Ratio

- The Most Common Risk-Adjusted Measure:

$$SR = \frac{\textbf{Average } R^{fund} - \textbf{Average } R^{bill}}{\textbf{Std.Deviation}(R^{fund} - R^{bill})}$$

  - **Always** Uses T-Bill ("Riskless") Benchmark

  - **Substitutes** Net Return **R** for **Log[1+R]**
    - Critical Difference When Fund is Volatile

- Traditional SR Motivation is 1960's One-Period Model

Example Comparison

$$\textbf{SR} = \frac{\textbf{Average } R^{\textbf{fund}} - \textbf{Average } R^{\textbf{bill}}}{\textbf{Std.Deviation}(R^{\textbf{fund}} - R^{\textbf{bill}})}$$

SR^{fund} = (1.99% - 0.54%) ÷ 10.83% = 13.4%

SR^{index} = (1.21% - 0.54%) ÷ 4.52% = 14.9%

So SR Ranks Index Higher Than Fund, Despite Fund's Higher Probability of Beating the Index, Seen From

$$\textbf{Log Modified Information Ratio} = \frac{\textbf{Average Log}[1+R^{\textbf{fund}}] - \textbf{Average Log}[1+R^{\textbf{index}}]}{\textbf{Std.Deviation}(\textbf{Log}[1+R^{\textbf{fund}}] - \textbf{Log}[1+R^{\textbf{index}}])}$$

= (1.35% - 1.10%) ÷ 7.83% = 3.2% > 0

Note: *Log Modified Information Ratio Ranking ≠ Value-At-Risk Ranking*

Outperformance Ranking: Lesson 2

- In Practice, a Positive Log-Modified Information Ratio (LMIR) Indicates That the Fund Will Eventually Outperform Its Benchmark.
- The Size of the LMIR Is Directly Related to the Rate At Which The Underperformance Probability Decays to Zero as the Holding Period Lengthens.
- The Sharpe Ratio (SR) Appears Similar, But Isn't the Same.
  - The Sharpe Ratio Has Similar Properties When:
    - The Relevant Benchmark is T-Bills
    - The Fund isn't Too Volatile, So Log[1+R] ≈ R
- Both LIMR and SR Depend on the Ratio of an Average to a Standard Deviation.
- As Will Be Seen, Many Published Rating Systems Depend (De-Facto) On the Difference of an Average and a Standard Deviation.

Published Fund Rating Systems

- Standard and Poors
  - Partition Funds into Investment Style Bins
    - Use Returns-Based Statistical Analysis to Do This
  - Compute Fund Annual Return and Sharpe Ratio (Using 3 yr. Std. Dev.) For Each of 3 Previous Years
    - Sort the Fund's Bin by Annual Return into Percentiles in Each of 3 Previous Years; then Average the 3 Percentile Rankings to Place Each Fund in a Decile
    - Do the Same Sorting and Averaging by Sharpe Ratio
  - Average the Above Two Decile Rankings
  - Separately From This, They Interview Managers' "Quality"

Published Fund Rating Systems

- Morningstar
  - Partition Funds into 48 Bins
    - Representing Combos of Size, Style, Asset Class, Sector, etc.
  - Compute Separate 120, 60, and 36 month **Ranking Statistic** Equivalent To :

Average of $-[(1+R^{fund})/(1+R^{bill})]^{-\gamma}$ Using $\gamma = 2$

- Always Uses T-Bill Benchmark, Like Sharpe Ratio Does
- Sort the Fund's Bin Separately for 120, 60, and 36 month stats
- Top 10% Get ***** Rating, Bottom 10% Get * Rating
- Overall Rating is a Declining Weighted Average of Separates

Published Fund Rating Systems

- Lipper : Uses Several Different Systems
  - Lipper “Consistent Return”
    - Partitions Funds Into Lipper’s Bins
    - Like S & P and Schwab, Combines **Two Stats**
      - **Hurst Exponent**: Weird Measure of Return Smoothness
      - **Effective Return** = Average Exponential Utility, Using Two Coefficients of Risk Aversion: Higher Coefficient is Used to Evaluate Losses (see Dacorogna, Gencay, et.al., Olson and Associates, 1999).
      - Uses Un-Weighted Average of 120,60,and 36 month Percentile Rankings.
      - Top 20% Are “Lipper Leaders For Consistent Return”

Published Fund Rating Systems

- Lipper “Preservation”
  - Partitions Funds into One of Three Bins: Equity, Mixed Equity, or Fixed Income
  - Compute Separate 120, 60, and 36 Month **Ranking Statistic** Equivalent To:

    Average of Min[0, R^{fund}]

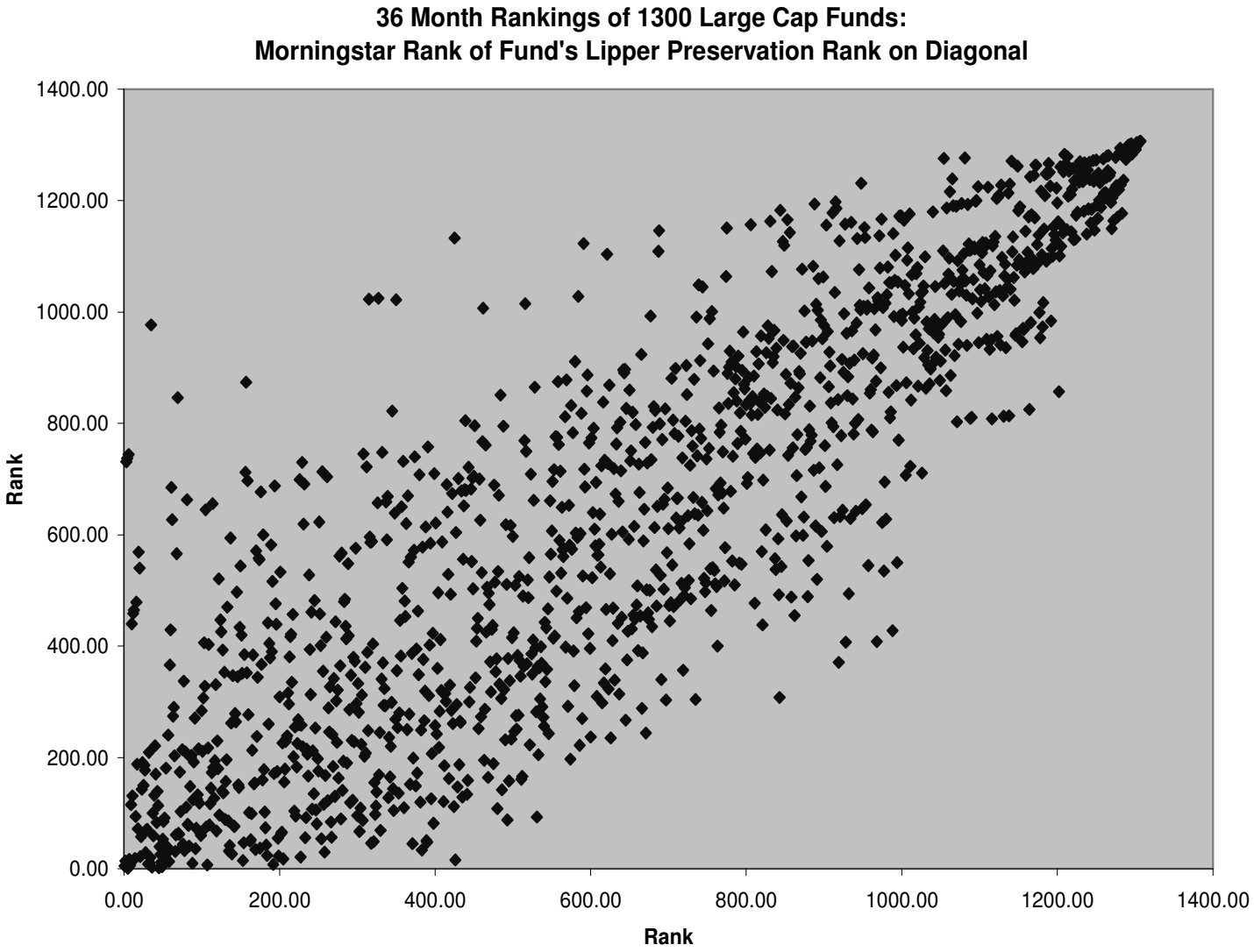
    • So Funds With Fewer and/or Smaller Negative Returns Are Ranked Higher
  - Uses Un-Weighted Average of 120,60,and 36 month Percentile Rankings. Top 20% Are “Lipper Leaders For Preservation”

Comparison of Rating Systems

- Facilitated By the Following Approximations:
  - Lipper Preservation: Average R^{fund} - (Std. Dev. R^{fund})
  - Morningstar: Average Log [1+R^{fund}] -(Std. Dev. Log [1+ R^{fund}])^2
    - Both Companies Confirm the Accuracy of Approximations
  - Standard and Poors: Average R^{fund} + Sharpe Ratio
  - Schwab: Similar to Standard and Poors

- All Systems Reward High Average Return and Penalize High Volatility
  - As a Result, Different Firms' Rankings Would Be Fairly Similar Within the Same Bin, Using the Same Number of Historical Monthly Returns

Rank Correlation of Seemingly Dissimilar Systems



Rank Correlation is 87%

Summary of Comparisons

Summary of De-Facto Weightings of Averages and Standard Deviations

Performance Measure	Benchmark	Return Average	Standard Deviation
Sharpe Ratio	T-Bill	Net	Divided Into
Morningstar	T-Bill	Log Gross	Squared and Subtracted From
Lipper Preservation	Zero	Net	Subtracted From
Log Information Ratio	User-Selected	Log Gross	Divided Into

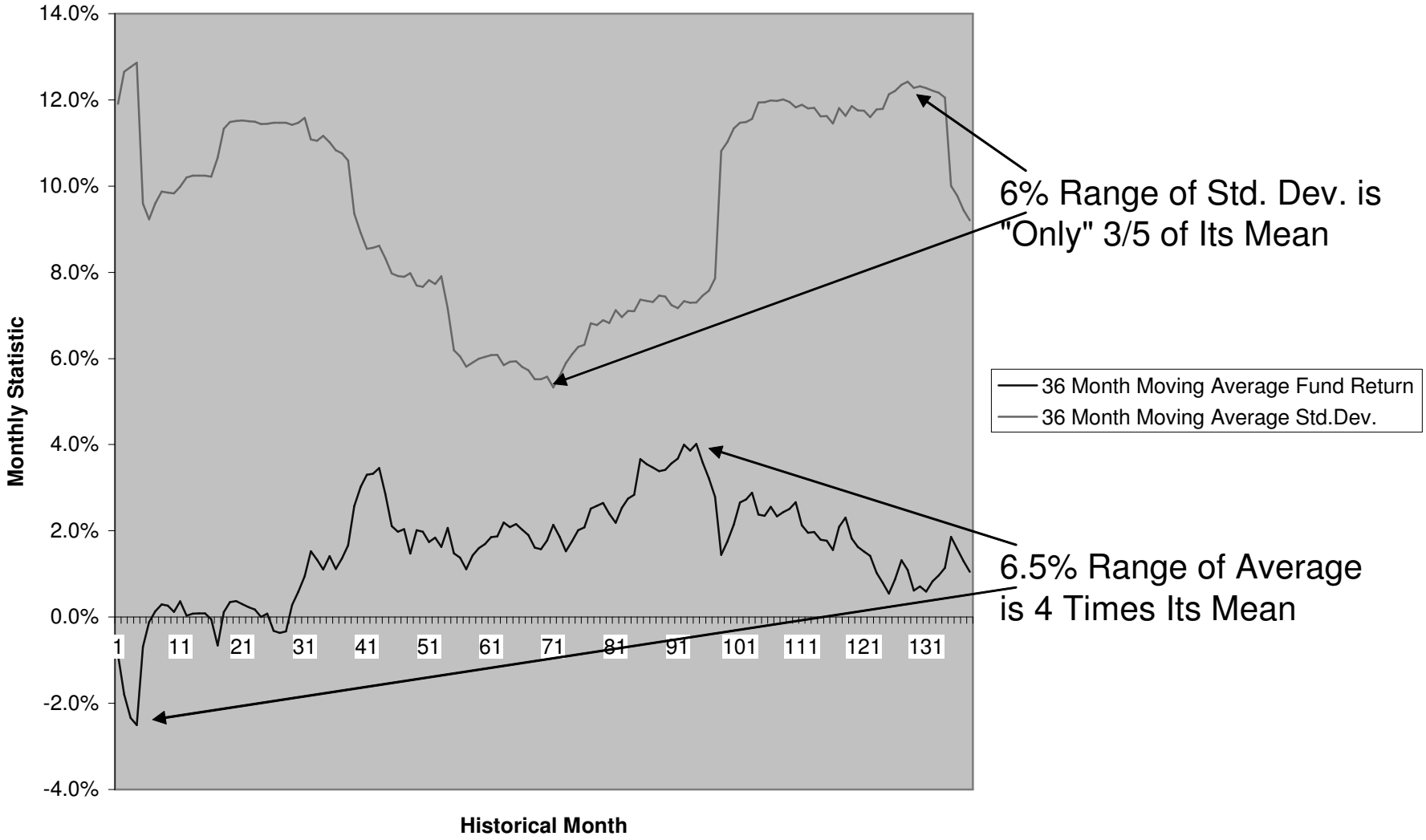
Morningstar's Ranking Can Also Be Transformed Into LMIR Ranking By Changing its Fund-Independent Value of $\gamma = 2$ to the Value of γ that Maximizes It (see the Paper's Appendix).

Implementation Problems

- A Fund's Bin Assignment and Benchmark Is Critical
  - Published Rating Systems Assign Funds to Bins Differently
  - Published Rating Systems Use Either a T-Bill or No Benchmark
    - Investors Want to Beat Relevant Benchmarks, Not T-Bills

- 36 Months of Historical Returns Is Too Few To Get An Accurate Estimate of Long-Run Average Returns
  - All Rating Systems (De-Facto) Depend Heavily on Avg. Return
  - This Inaccuracy Could Lead to Spurious Instability of Fund Rankings
  - Standard Deviations Not As Problematic: More Frequent Measurement (e.g. weekly, daily, etc.) Improves Accuracy

Instability of 36 Month Historical Averages vs. Std. Devs.



Partial Solutions

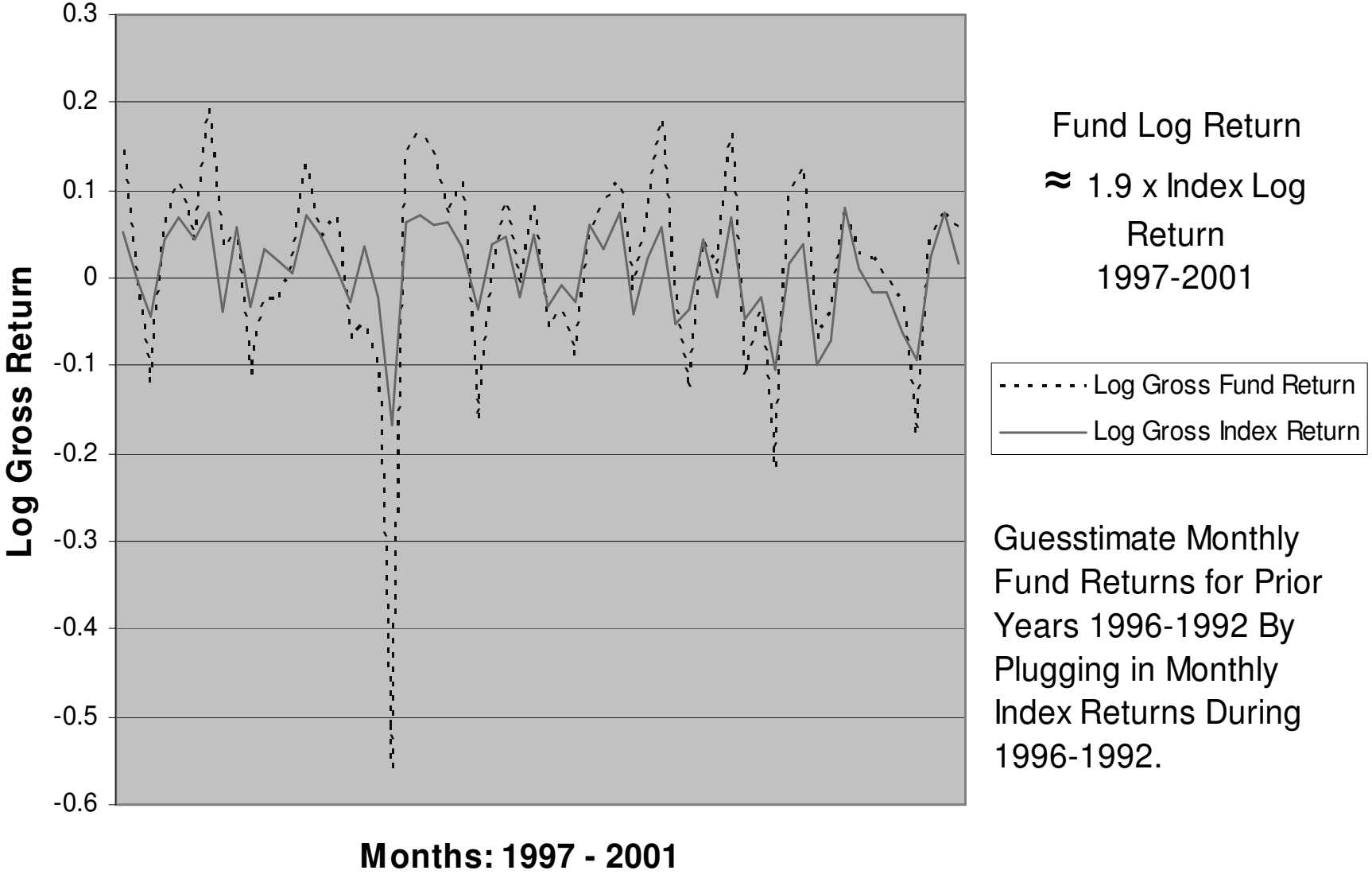
- Both Problems Can Be Partially Solved By Using a Bin-Specific Benchmark, Highly Correlated with Funds in Bin

  – Historical Average *Difference* of Fund from Benchmark's Return May Be a Better Estimate of its Long-Run Difference than the Fund's Average Return is of its Long-Run Return.
    - It is Important That the Fund is Highly Correlated with Its Benchmark

  – If a Highly Correlated Benchmark Has a Longer Return History Than the Fund, *Filtering* Methods Can Be Used to Improve Estimates of the Fund's Long Run Average.

    - A Simple Approach: Regress Fund Returns on Benchmark, and Use Regression Equation to "Backcast" Fund Returns for Prior Months When the Fund Didn't Exist!! Works Better Than it Sounds!

Figure 9: Fund and Index Log Gross Returns



Using the Longer History of a Highly Correlated Benchmark to Improve Fund Estimates