Chapter 24 - Portfolio Performance Evaluation

Richard Herron

The Conventional Theory of Performance Evaluation

The difference between *arithmetic* and *geometric* averages

- Suppose we evaluate the performance of a portfolio over a 20-year holding period
- The *arithmetic average* is the sum of the 20 annual returns divided by 20:

$$\overline{r} = \frac{\sum_{t=1}^{20} r_t}{20}$$

- The *geometric average* is the constant annual return r_G that provides the same cumulative return:

$$r_G = [(1 + r_1)(1 + r_2) \cdots (1 + r_{20})]^{1/20} - 1$$

The difference between time-weighted and dollar-weighted averages

- The *time-weighted* average weights each period's return the same, which is the *geometric average*
- The *dollar-weighted* average weights each period's return by the amount invested that period, with is the internal rate of return (IRR)

Example: A stock sells for \$50. You purchase one share today and one more in one year. At the end of the second year, you sell both shares for \$54. Dividends of \$2 per share are paid annually at the end of each year (but before shares are sold).

Time	Outlays
0	\$50 to purchase the first share
1	\$53 to purchase a second share a year later
	Proceeds
1	\$2 dividend from initially purchased share
2	\$4 dividend from the 2 shares held in the second year, plus

Time	Outlays
	\$108 received from selling both shares at \$54 each

- The **time-weighted** average (geometric average) is 7.81%:

$$
\begin{aligned}
r_1 &= \frac{53+2-50}{50} = 0.1 \\
r_2 &= \frac{54+2-53}{=} 0.0566 \\
r_G &= (1.10 \times 1.0566)^{1/2} - 1 \implies r_G = 0.0781
\end{aligned}
$$

- The **dollar-weighted** average (IRR) is 7.12%:

$$0 = -50 + \frac{-53+2}{1+r} + \frac{112}{(1+r)^2} \implies r = 0.0712$$

We must risk-adjust returns to compare them

- Simplest and most popular risk adjustment is to compare rates of return with those of other investment funds with similar risk characteristics
- **Comparison universe:**
  - The set of money managers with similar investment styles used to assess the relative performance of a manager
  - For example, a 90^{th} percentile manager provides higher returns than 90% of managers in her comparison universe

- Comparison universes are simple and intuitive
- But have the following shortcomings:
  - Managers concentrate in subgroups within their comparison universe
  - Comparison universes are not investable (e.g., we cannot invest in the median manager)

There are several popular risk-adjusted performance measures

- **Sharpe ratio** $\left(\frac{\overline{r}_P - \overline{r}_f}{\sigma_P}\right)$ measures the reward to total risk trade-off
- **Treynor ratio** $\left(\frac{\overline{r}_P - \overline{r}_f}{\beta_P}\right)$ measures the reward to systematic risk trade-off
- **Jensen's alpha** $\left(\alpha_P = \overline{r}_P - \left[\overline{r}_f + \beta_P\left(\overline{r}_M - \overline{r}_f\right)\right]\right)$ is the average return on the portfolio over the CAPM-predicted return, given the portfolio's beta and the average market return
- **Information ratio** $\left(\frac{\alpha_P}{\sigma(e_P)}\right)$ divides portfolio alpha by its nonsystematic risk, so it measures abnormal return per unit of diversifiable risk

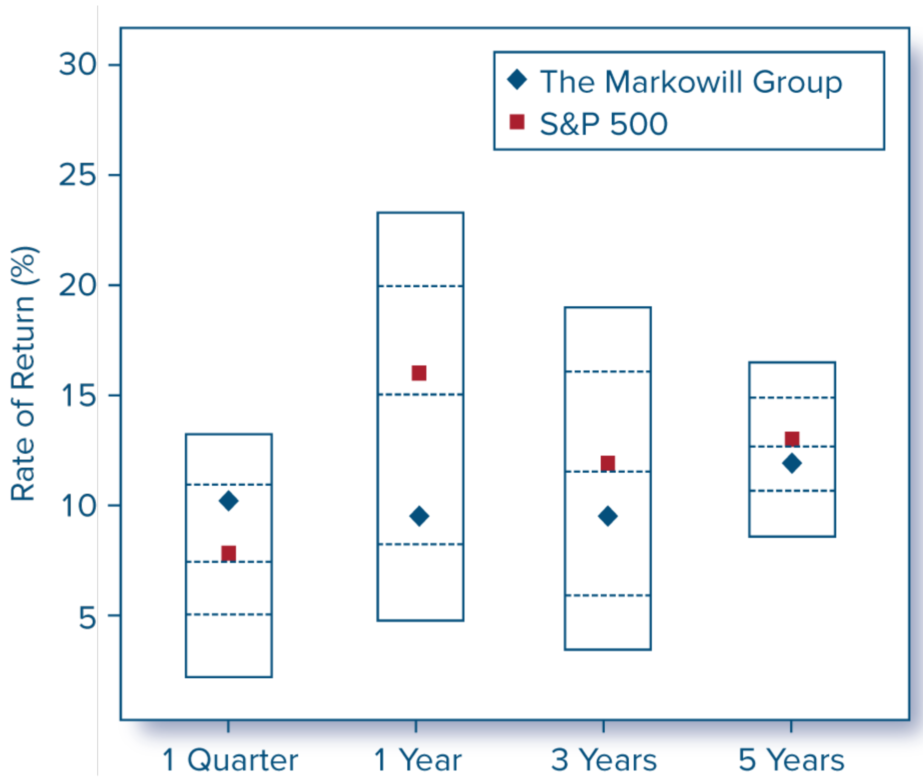


Figure 1: Universe comparison, periods ending December 31, 2028 (BKM 2023, Figure 24.1)

The M^2 performance measure presents the Sharpe ratio as an excess return

- Lever or unlever portfolio P to match the volatility of the passive market index
- This portfolio P^* has the same volatility as the passive market index
- M_P^2 is the excess return on portfolio P^*:

$$M_P^2 = r_{P^*} - r_M$$

M_P^2 can be negative, even when $r_P > r_M$

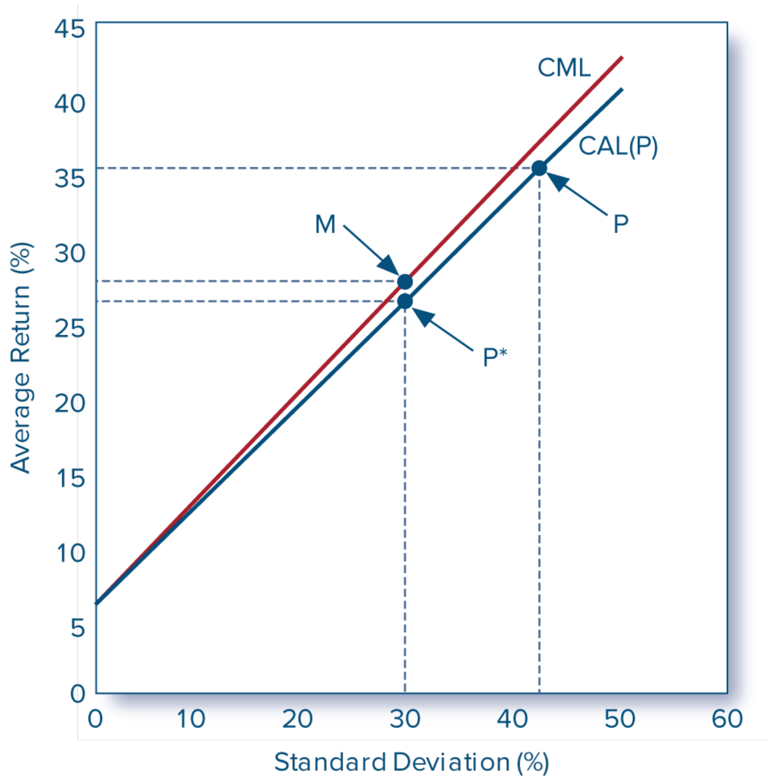


Figure 2: The M_P^2 of portfolio P is negative even though its average return was greater than that of the market index, M (BKM 2023, Figure 24.2)

The Treynor ratio considers *systematic* risk, so it is useful for assembling a diversified fund-of-funds

The information ratio measures the trade-off between alpha and diversifiable risk

- So, the information ratio is useful for adding an active portfolio to a passive portfolio

	Risk-free Asset	Portfolio Q	Portfolio U	Market Index, M
Beta	0	1.3	0.8	1.0
Average return	6	22.0	17.0	16.0
Excess return (%)	0	16.0	11.0	10.0
Alpha (%)	0	3.5	3.0	0.0

Note: Excess return = Average return − Risk-free rate

Alpha = Average return − Beta × (Market return − Risk-free rate)

Figure 3: Portfolio performance (BKM 2023, Table 24.1)

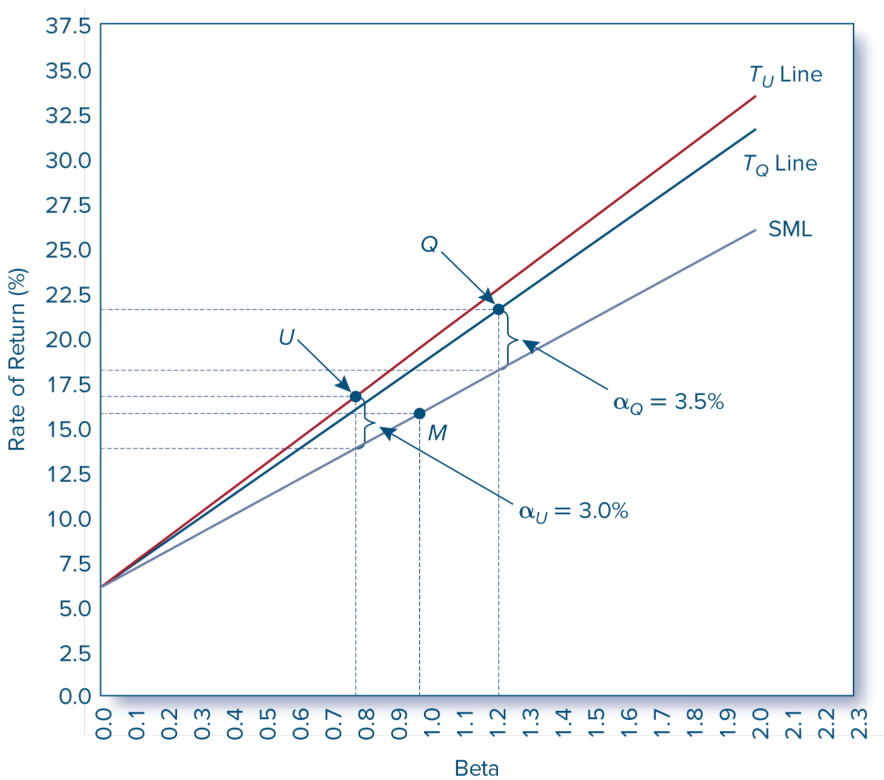


Figure 4: Treynor measures of two portfolios and the market index (BKM 2023, Figure 24.3)

- For example, *optimally() combining active fund H with market index M improves the Sharpe ratio as follows:

$$S_P^2 = S_M^2 + \left[\frac{\alpha_H}{\sigma(e_H)}\right]^2$$

When should we apply each risk-adjusted performance measure?

Performance Measure	Definition	Application
Sharpe	$\frac{\bar{r}_P - \bar{r}_f}{\sigma_P}$	When choosing among portfolios competing for the overall risky portfolio
Treynor	$\frac{\bar{r}_P - \bar{r}_f}{\beta_P}$	When ranking many portfolios that will be mixed to form the overall risky portfolio
Information ratio	$\frac{\alpha_P}{\sigma(e_P)}$	When evaluating a portfolio to be mixed with a diversified benchmark portfolio

How does alpha relate to each risk-adjusted performance measure?

	Treynor	Sharpe	Information Ratio
Relation to alpha	$\frac{E(r_P)-r_f}{\beta_P} = \frac{\alpha_p}{\beta_P} + T_M$	$\frac{E(r_P)-r_f}{\sigma_P} = \frac{\alpha_p}{\sigma_P} + \rho S_M$	$\frac{\alpha_P}{\sigma(e_p)}$
Improvement compared to the market	$T_P - T_M = \frac{\alpha_p}{\beta_P}$	$S_P - S_M = \frac{\alpha_p}{\sigma_P} - (1-\rho) S_M$	$\frac{\alpha_P}{\sigma(e_p)}$

An application of risk-adjusted performance measures

- If P or Q represents the entire investment, Q is better because of its higher Sharpe ratio
- If P and Q are potential sub-portfolios, Q remains better because of its higher Treynor ratio
- However, if we seek an active portfolio to mix with an index portfolio, P is better because of its higher information ratio

Realized returns versus expected returns

- We must determine the statistical “significance level” of a performance measure to determine whether it reliably indicates any ability

Month	Portfolio P	Alternative Q	Index M
1	3.58%	2.81%	2.20%
2	−4.91	−1.15	−8.41
3	6.51	2.53	3.27
4	11.13	37.09	14.41
5	8.78	12.88	7.71
6	9.38	39.08	14.36
7	−3.66	−8.84	−6.15
8	5.56	0.83	2.74
9	−7.72	0.85	−15.27
10	7.76	12.09	6.49
11	−4.01	−5.68	−3.13
12	0.78	−1.77	1.41
Average	2.77	7.56	1.64
Standard deviation	6.45	15.55	8.84

Figure 5: Excess returns for portfolios P and Q and the market index M over 12 months (BKM 2023, Table 24.2)

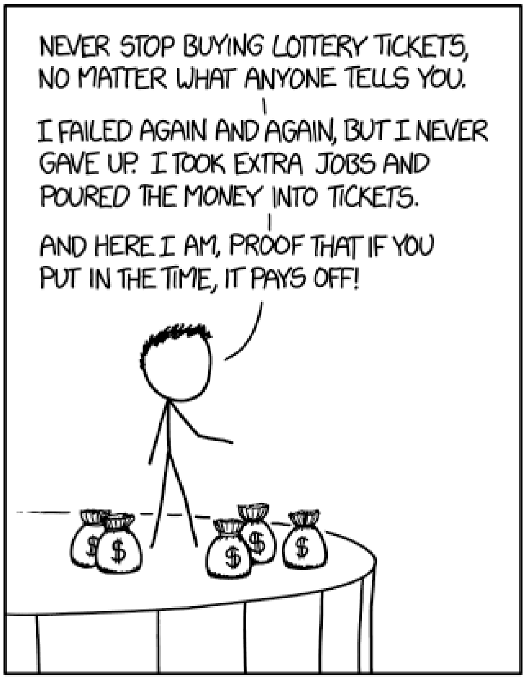
	Portfolio P	Portfolio Q	Portfolio M
Sharpe ratio	0.43	0.49	0.19
M^2	2.16	2.66	0.00
SCL regression statistics			
Alpha	1.63	5.26	0.00
Beta	0.70	1.40	1.00
Treynor	3.97	5.38	1.64
T^2	2.34	3.74	0.00
$\sigma(e)$	2.02	9.81	0.00
Information ratio	0.81	0.54	0.00
R-square	0.91	0.64	1.00

Figure 6: Performance statistics (BKM 2023, Table 24.3)

- Even moderate levels of statistical noise make it difficult to evaluate performance

Regardless of the performance criterion, some funds will outperform their benchmarks in any year, and some will underperform

- Performance in one period does not predict future performance
- Limiting a sample of funds to those with available returns over an entire sample period introduces **survivorship bias**



EVERY INSPIRATIONAL SPEECH BY SOMEONE SUCCESSFUL SHOULD HAVE TO START WITH A DISCLAIMER ABOUT SURVIVORSHIP BIAS.

Style Analysis

Style analysis **measures the exposures of managed portfolios to asset classes**

- Introduced by William Sharpe
- Regress fund returns on indexes representing a range of asset classes
  - Coefficient on each index measures the fund's implicit allocation to each "style"
  - Intercept coefficient measures the average return from security selection
  - R^2 measures the percentage of return variability attributable to style choice instead of security selection

Style Portfolio	Regression Coefficient
T-bill	0
Small cap	0
Medium cap	35
Large cap	61
High P/E (growth)	5
Medium P/E	0
Low P/E (value)	0
Total	100
R-square	97.5

Figure 7: Style analysis for Fidelity's Magellan Fund (BKM 2023, Table 24.4)

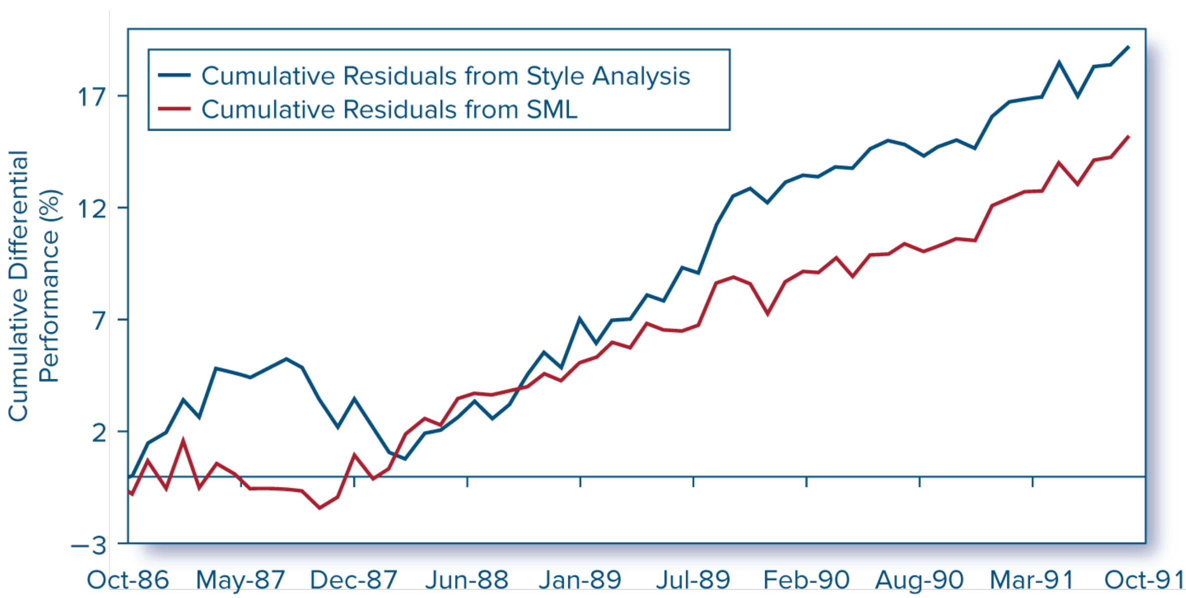


Figure 8: Fidelity Magellan Fund cumulative return difference: Fund versus style benchmark and fund verus SML benchmark (BKM 2023, Figure 24.4)

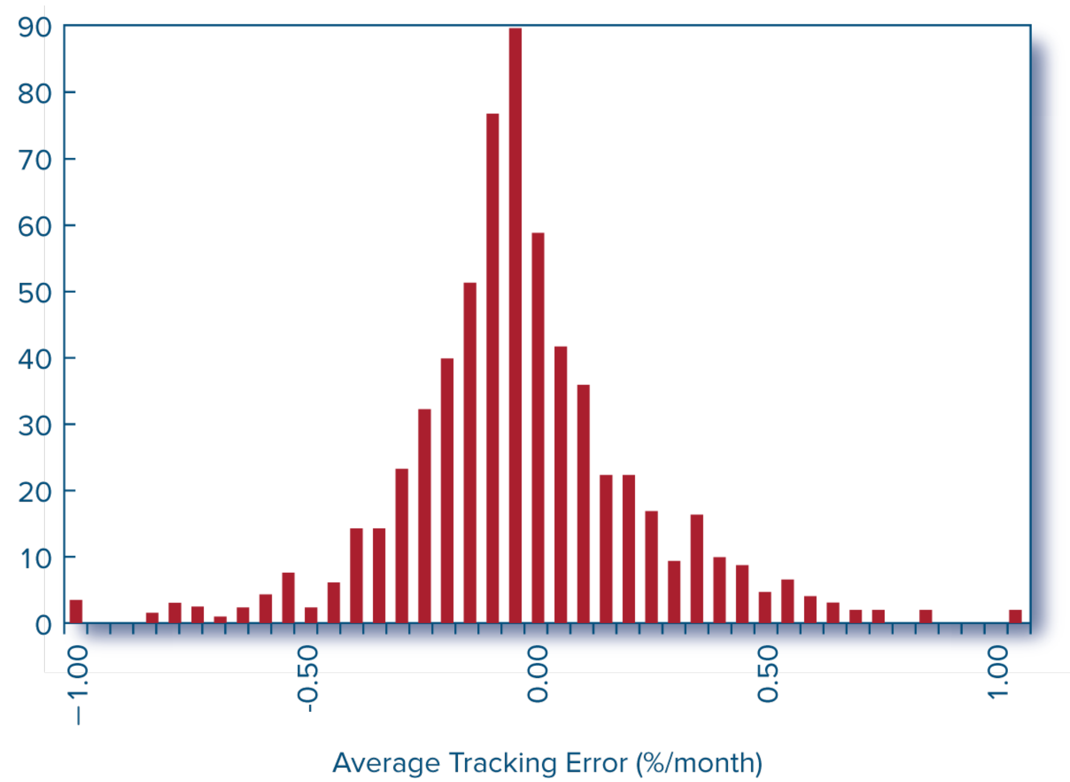


Figure 9: Average tracking error for 636 mutual funds, 1985–1989 (BKM 2023, Figure 24.5)

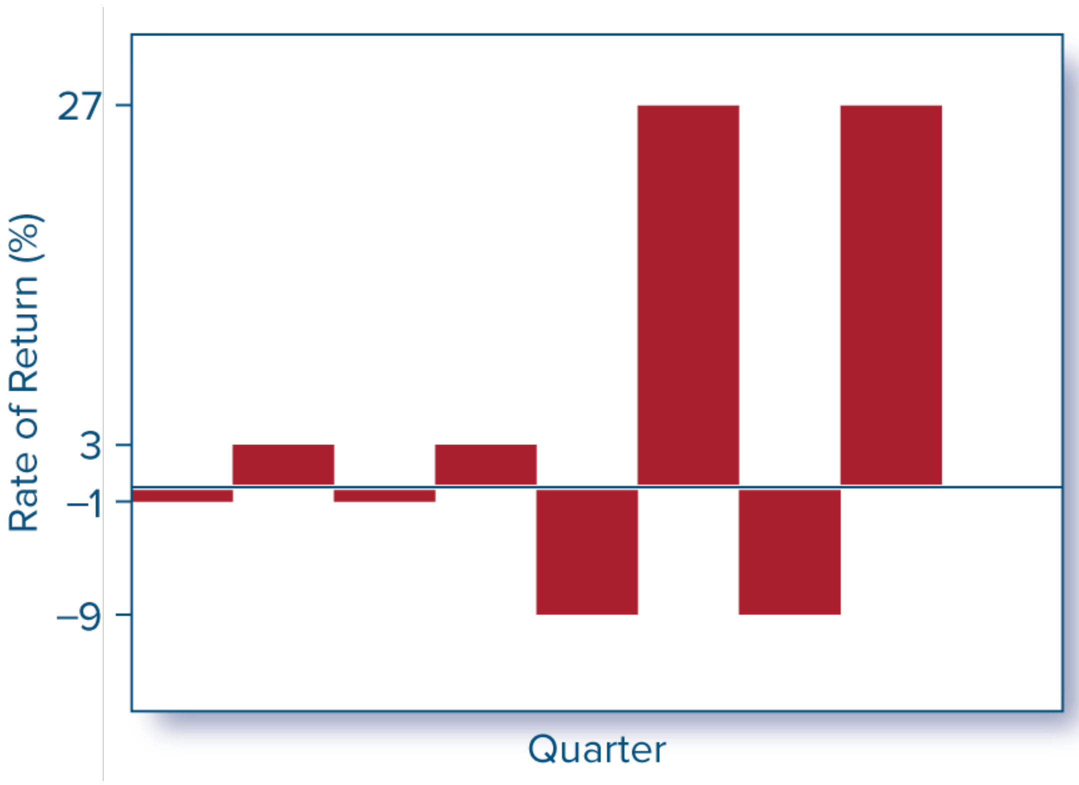


Figure 10: Portfolio returns: Returns in last four quarters are more variable than in the first four (BKM 2023, Figure 24.6)

Performance Measurement with Changing Portfolio Composition

Risk-adjusted performance measures assume constant portfolio risk, which is not necessarily true

Managers may game the risk-adjusted performance measures above

- The Sharpe ratio is invariant to the fraction y of the portfolio invested in the risky portfolio instead of the risk-free asset
- However, increasing leverage after the first period of an evaluation period increases the influence of second-period performance
  - If early returns are bad, increase leverage in the second period
  - If early returns are good, decrease leverage in the second period
  - This strategy creates a negative correlation between first-period and second-period returns, which reduces the Sharpe ratio denominator σ_P

The Morningstar risk-adjusted rating (MRAR) is impossible to manipulate!

- MRAR is the risk-free equivalent excess return of the portfolio for an investor with risk aversion γ:

$$\mathrm{MRAR}(\gamma) = \left[\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1+r_t}{1+r_{ft}}\right)^{-\gamma}\right]^{\frac{12}{\gamma}} - 1$$

- MRAR and Sharpe ratio rankings are similar

Market Timing

Market timing is shifting funds between a risky portfolio and a safe asset, depending on expected returns

- Tryenor and Mazuy estimate this regression:

$$r_p - r_f = a + b(r_M - r_f) + c(r_M - r_f)^2 + e_p$$

- Henriksson and Merton estimate this regression:

$$r_p - r_f = a + b(r_M - r_f) + c(r_M - r_f)D + e_p$$

  where $D = 1$ if $r_M > r_f$ else $D = 0$, so portfolio beta is b in bear markets and $b + c$ in bull markets
- A $c > 0$ indicates market-timing ability for both models, but neither model has found evidence of market-timing ability among managers

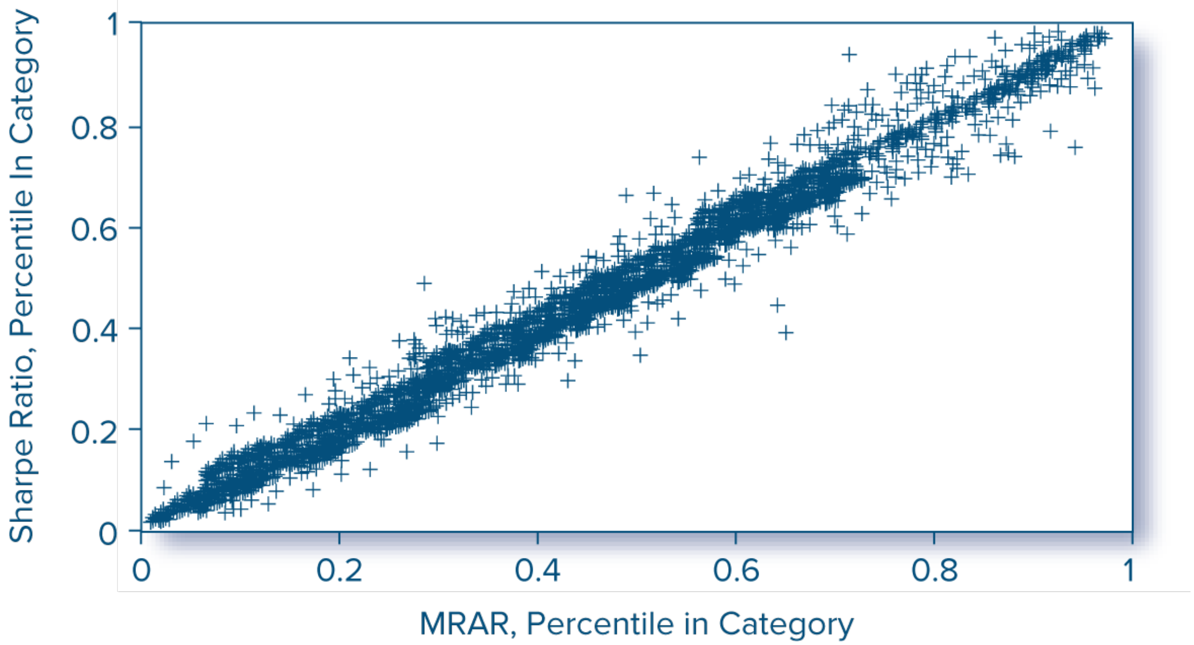


Figure 11: Rankings based on Morningstar's RAR versus Sharpe ratio (BKM 2023, Figure 24.7)

	Portfolio	
Estimate	***P***	***Q***
Alpha (*a*)	1.77 (1.63)	–2.29 (5.26)
Beta (*b*)	0.70 (0.70)	1.10 (1.40)
Timing (*c*)	0.00	0.10
R-square	0.91 (0.91)	0.98 (0.64)

What is the potential of perfect market timing?[1]

[1] Performance of bills, equities, and perfect (annual) market timers. Initial investment = $1 (December 31, 1926 to December 31, 2021)

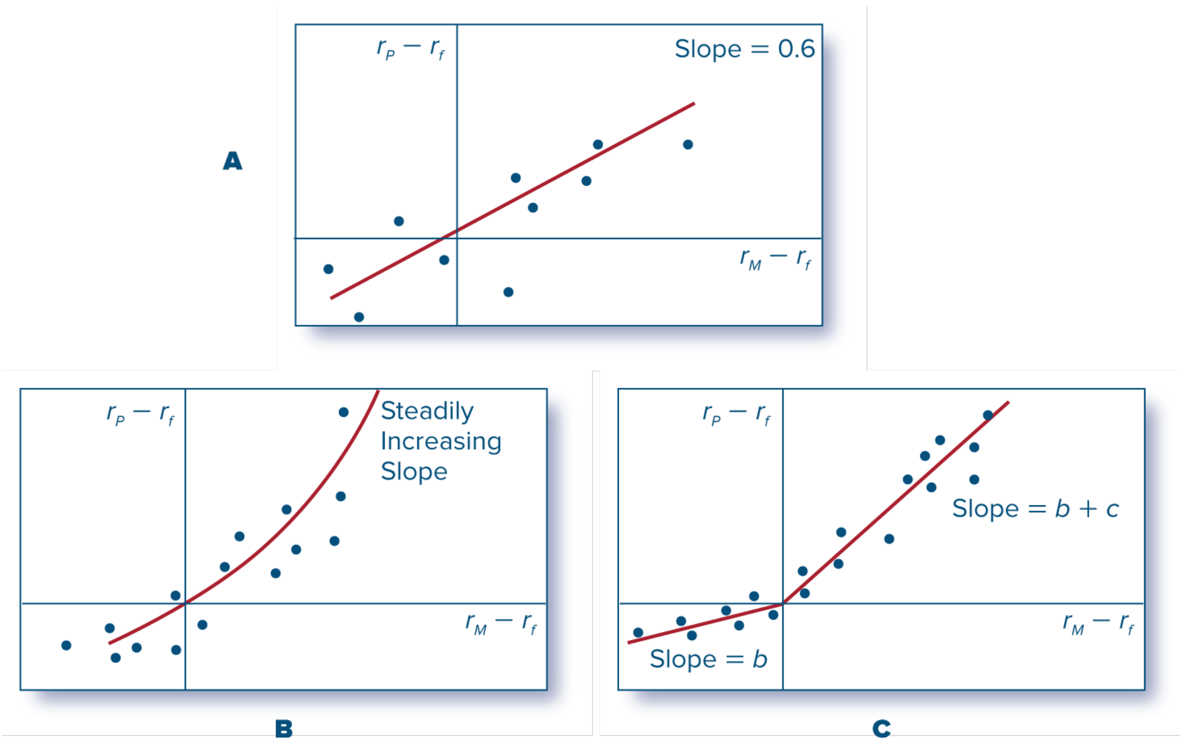


Figure 12: Characteristic lines. *Panel A:* No market timing, beta is constant. *Panel B:* Market timing, beta increases with expected market excess return. *Panel C:* Market timing with only two values of beta. (BKM 2023, Figure 24.8)

	Bills	Equities	Perfect Timer
Terminal value	\$21	\$10,546	\$15,12,355
Arithmetic average	3.31%	11.96%	16.72%
Standard deviation	3.12%	19.89%	13.34%
Geometric average	3.26%	10.24%	16.16%
Maximum	14.71%	57.35%	57.35%
Minimum	−0.02%	−44.04%	0.00%
Skew	1.08	−0.45	0.69
Kurtosis	1.09	0.11	−0.14
LPSD	0	12.80%	0

Perfect market-timing ability is a call option on market returns

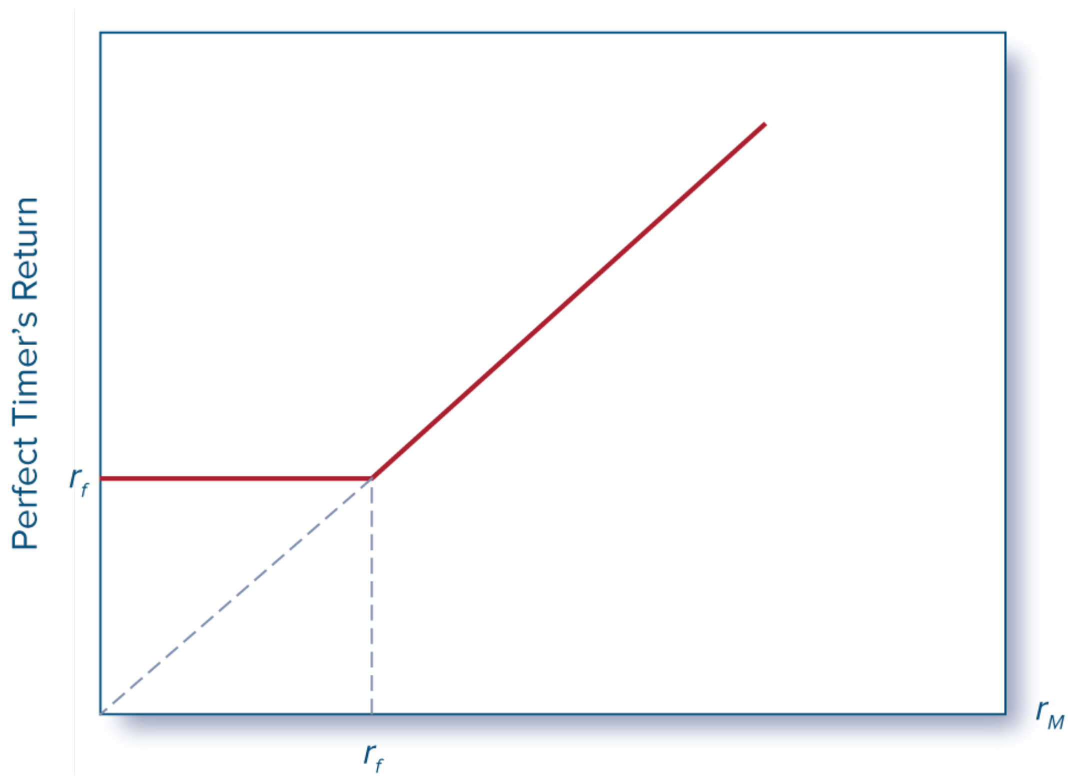


Figure 13: Rate of return of a perfect market timer as function of the rate of return on the market index (BKM 2023, Figure 24.9)

	$S_T < X$	$S_T \geq X$
Bills	$S_0(1+r_f)$	$S_0(1+r_f)$
Call	0	$S_T - X$
Total	$S_0(1+r_f)$	S_T

- The table on the previous slide gives the payoff of this call option
- X is the strike price, and $X = S_0(1+r_f)$
- The value of this perfect market-timing call option is $C = 7.92\%$ of the equity portfolio value
- The value of imperfect market timing is $(P_1 + P_2 - 1) \times C$ where:
  - P_1 is the proportion of correct forecasts of bull markets
  - P_2 is the proportion of correct forecasts of bear markets

Performance Attribution Procedures

Performance attribution decomposes overall performance into discrete components

- One common attribution approach decomposes performance into three components:
  1. Broad asset allocation choices across equity, fixed-income, and money markets
  2. Industry (sector) choice within each market
  3. Security choice within each sector
- This attribution approach explains the difference between a managed portfolio P and a benchmark portfolio B
  - Portfolio B, or **bogey**, measures the returns on a completely passive strategy, removing *asset allocation* and *security selection* decisions
  - Therefore, asset allocation and security selection explain all differences between portfolios P and B

	Bogey Performance and Excess Return	
Component	**Benchmark Weight**	**Return of Index during Month (%)**
Equity (S&P 500)	0.60	5.81
Bonds (Barclays Aggregate Index)	0.30	1.45
Cash (money market)	0.10	0.48
Bogey = (0.60 × 5.81) + (0.30 × 1.45) + (0.10 × 0.48) = 3.97%		
Return of managed portfolio	5.34%	
− Return of bogey portfolio	3.97	
Excess return of managed portfolio	1.37%	

Figure 14: Performance of the managed portfolio (BKM 2023, Table 24.6)

Contributions of asset allocation and selection

A. Contribution of asset allocation to performance					
Market	(1) Actual Weight in Market	(2) Benchmark Weight in Market	(3) Active or Excess Weight	(4) Index Return (%)	(5) = (3) × (4) Contribution to Performance (%)
Equity	0.70	0.60	0.10	5.81	0.5810
Fixed-income	0.07	0.30	−0.23	1.45	−0.3335
Cash	0.23	0.10	0.13	0.48	0.0624
Contribution of asset allocation					0.3099
B. Contribution of selection to total performance					
Market	(1) Portfolio Performance (%)	(2) Index Performance (%)	(3) Excess Performance (%)	(4) Portfolio Weight	(5) = (3) × (4) Contribution (%)
Equity	7.28	5.81	1.47	0.70	1.03
Fixed-income	1.89	1.45	0.44	0.07	0.03
Contribution of selection within markets					1.06

Figure 15: Performance attribution (BKM 2023, Table 24.7)

Contributions of *sector* and *security* selections

	(1) Beginning-of-Month Weights (%)	(2)	(3)	(4)	(5) = (3) × (4)
Sector	Portfolio	S&P 500	Active Weight (%)	Sector Return (%)	Sector Allocation Contribution
Basic materials	1.96	8.3	−6.34	6.9	−0.4375
Business services	7.84	4.1	3.74	7.0	0.2618
Capital goods	1.87	7.8	−5.93	4.1	−0.2431
Consumer cyclical	8.47	12.5	−4.03	8.8	0.3546
Consumer noncyclical	40.37	20.4	19.97	10.0	1.9970
Credit sensitive	24.01	21.8	2.21	5.0	0.1105
Energy	13.53	14.2	−0.67	2.6	−0.0174
Technology	1.95	10.9	−8.95	0.3	−0.0269
Total					1.2898

Figure 16: Sector selection with the equity market (BKM 2023, Table 24.8)

Sector	(1) Beginning-of-Month Weights (%) Portfolio	(2) S&P 500	(3) Active Weight (%)	(4) Sector Return (%)	(5) = (3) × (4) Sector Allocation Contribution
Basic materials	1.96	8.3	−6.34	6.9	−0.4375
Business services	7.84	4.1	3.74	7.0	0.2618
Capital goods	1.87	7.8	−5.93	4.1	−0.2431
Consumer cyclical	8.47	12.5	−4.03	8.8	0.3546
Consumer noncyclical	40.37	20.4	19.97	10.0	1.9970
Credit sensitive	24.01	21.8	2.21	5.0	0.1105
Energy	13.53	14.2	−0.67	2.6	−0.0174
Technology	1.95	10.9	−8.95	0.3	−0.0269
Total					1.2898

Figure 17: Portfolio attribution: summary (BKM 2023, Table 24.9)

Putting it all together

Summary from BKM (2023)

Page 855

1. The simplest performance measure compares average return to that on a benchmark such as an appropriate market index or even the median return of funds in a comparison universe. Alternative measures of the average return include the arithmetic and geometric average and time-weighted versus dollar-weighted returns.
2. The appropriate performance measure depends on the role of the portfolio to be evaluated. Appropriate performance measures are as follows:

   a. Sharpe: When the portfolio represents the entire investment fund.

   b. Information ratio: When the portfolio is an active portfolio to be optimally mixed with the passive portfolio.

   c. Treynor: When the portfolio is one subportfolio of many.

   d. Jensen (alpha): All of these measures require a positive alpha for the portfolio to be considered attractive.
3. Many observations and long sample periods are required to eliminate the effect of the "luck of the draw" from the evaluation process because portfolio returns commonly are very noisy.
4. Style analysis uses a multiple regression model where the factors are category (style) portfolios such as bills, bonds, and stocks. The coefficients on the style portfolios indicate a passive strategy that would match the risk exposures of the managed portfolio. The return difference between the managed portfolio and the matching portfolio measures performance relative to similar-style funds.
5. Shifting mean and risk of actively managed portfolios makes it difficult to assess performance. An important example of this problem arises when portfolio managers attempt to time the market, resulting in ever-changing portfolio betas.
6. One way to measure timing and selection success simultaneously is to estimate an expanded security characteristic line, for which the slope (beta) coefficient is allowed to increase as the market return increases. Another way to evaluate timers is based on the implicit call option embedded in their performance.
7. Common attribution procedures decompose portfolio performance to asset allocation, sector selection, and security selection decisions. Performance is assessed by calculating departures of portfolio composition from a benchmark or neutral portfolio.

Key equations from BKM (2023)

Geometric time-weighted return: $1 + r_G = [(1 + r_1)(1 + r_2) \cdots (1 + r_n)]^{1/n}$

Sharpe ratio: $S_P = \dfrac{r_P - r_f}{\sigma_P}$

M^2 of portfolio P given its Sharpe ratio: $M^2 = \sigma_M(S_P - S_M)$

Treynor measure: $T_P = \dfrac{r_P - r_f}{\beta_P}$

Jensen's alpha: $\alpha_P = \bar{r}_P - \left[\bar{r}_f + \beta_P\left(\bar{r}_M - \bar{r}_f\right)\right]$

Information ratio: $\dfrac{\alpha_P}{\sigma(e_P)}$

Morningstar risk-adjusted return: $\text{MRAR}(\gamma) = \left[\dfrac{1}{T}\sum_{t=1}^{T}\left(\dfrac{1 + r_t}{1 + r_{ft}}\right)^{-\gamma}\right]^{\frac{12}{\gamma}} - 1$

References

Bodie, Zvi, Alex Kane, and Allan J. Marcus (2023). *Investments*. 13th ed. New York: McGraw Hill.