

THE JOURNAL OF INVESTING

FALL 2009

www.ijoi.com

VOLUME 18 NUMBER 3

When Indexing Works and When It Doesn't in U.S. Equities: *The Purity Hypothesis*

WILLIAM R. THATCHER



Hammond Associates
INSTITUTIONAL FUND CONSULTANTS, INC.

IIJ Institutional
Investor
Journals

The Voices of Influence | ijournals.com

When Indexing Works and When It Doesn't in U.S. Equities: *The Purity Hypothesis*

WILLIAM R. THATCHER

WILLIAM R. THATCHER
is a senior consultant
at Hammond Associates
in St. Louis, MO.
bthatcher@haifc.com

Many investors think that domestic equity indexing tends to outperform active management in efficient markets like large stocks and that it tends to underperform active managers in less-efficient areas like small stocks. We call this idea the Efficiency Hypothesis. However, this concept contradicts Sharpe's observation that active management within an asset class is a zero-sum game (Sharpe [1991]). Further, performance data from 1998–2007 does not support the Efficiency Hypothesis.

From 1998–2007, domestic equity indexing was more likely to beat active managers in asset classes with high relative returns, and it was more apt to lose to active managers in categories with low relative performance. For example, if mid-cap value stocks were performing best and large-cap growth stocks were doing worst, then indexing would tend to beat active management in mid-cap value, and active management would tend to outdo indexes in large cap growth. William Bernstein started writing about this concept in 1999.¹ We at Hammond Associates call this idea the Purity Hypothesis.

The Purity Hypothesis gets its name from the fact that indexes are more style-pure than comparable active funds, i.e., active managers invest differently than the benchmark, which gives them less style purity. These differences create performance disparities between active and passive funds. Below are three examples of how domestic equity active

managers are less “pure” than their benchmark indexes:

1. On average, active domestic equity large-cap funds hold smaller stocks than similar large-cap indexes.
2. Active domestic equity small-cap managers generally invest in larger companies than those found in the Russell 2000 index.²
3. Active domestic equity funds have cash balances and corresponding indexes don't.³

The Purity Hypothesis hinges on the following idea: the more style-pure a fund is, the better it performs when its style does well, and the worse it performs when its style does poorly. For instance, if the mid-cap value asset class were beating all other domestic equity categories, then a mid-cap value index would tend to outperform mid-cap value active managers—many of which have a diluted style versus the benchmark (a style beta less than one). However, mid-cap value active managers that take more concentrated style bets than the index (a style beta greater than one) would be apt to beat the benchmark. Similarly, if the large-cap growth asset class were losing to all other domestic equity groups, then the large-cap growth index would tend to underperform large-cap growth active funds—many of which have watered-down styles versus the benchmark

(a style beta less than one). At the same time, the large-cap growth index would likely beat large-cap growth active managers that have more concentrated style positions vis-à-vis the benchmark (a style beta greater than one).

EXAMPLE OF THE PURITY HYPOTHESIS

Exhibit 1 illustrates the Purity Hypothesis. It ranks from best to worst the performance of S&P's nine domestic equity indexes versus comparable active managers for the 10-year period ended 2007. Data comes from Morningstar, and results are net of fees.

The column titled "Index Outperformance Versus Active Managers" references the percentage of active managers that a particular index outperformed. For instance, the S&P Mid-Cap 400 index beat 78% of mid-cap core

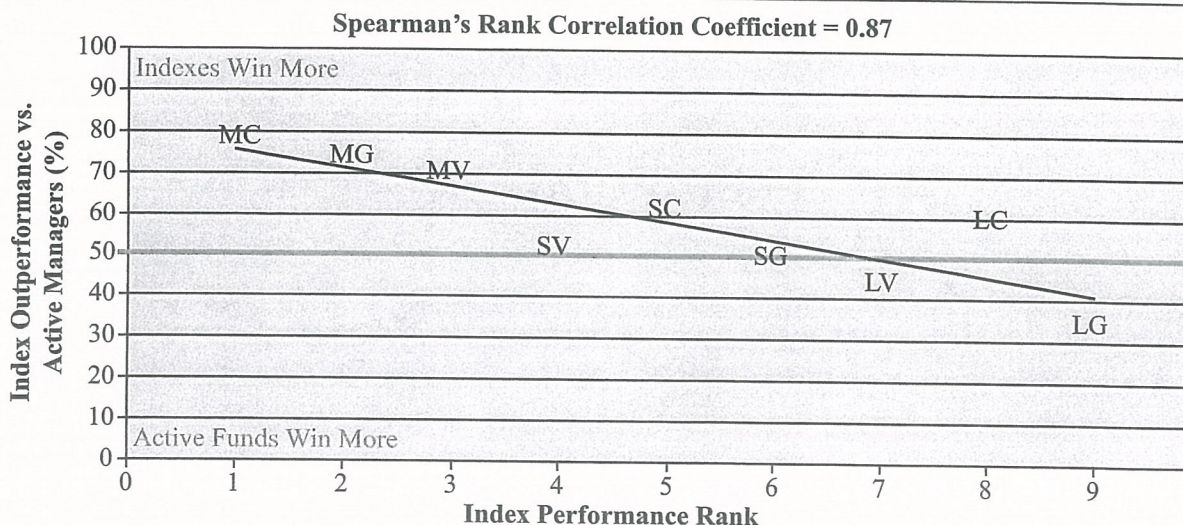
active managers. Note the strong correlation between the numbers in the "Index Outperformance Versus Active Managers" column and those in the "Index Performance Rank" column. Generally, the higher the "Index Outperformance Versus Active Managers" percentage, the higher the "Index Performance Rank" is. The degree of correlation between these two measures is gauged by Spearman's rank correlation coefficient. As shown in the graph's title, Spearman's rank correlation coefficient is 0.87. "Index Performance Rank" and the "Index Outperformance Versus Active Managers" percentage are correlated to a statistically significant degree (0.05 level) if Spearman's rank correlation coefficient is 0.60 or greater. So, if Spearman's rank correlation coefficient is 0.60 or more, then relative index returns help explain much of an index provider's performance versus active managers.

EXHIBIT 1

S&P Domestic Equity Indexes Ranked by Performance 10 Years Ended 2007

Index Versus Active Performance Analysis

Index	Annualized Performance %	Index Performance Rank	Index Outperformance Versus Active Managers %	Asset Class Abbreviation
S&P Mid-Cap 400	11.2	1	78	MC (Mid Core)
S&P Mid-Cap 400 Growth	11.1	2	72	MG (Mid Growth)
S&P Mid-Cap 400 Value	11.1	3	70	MV (Mid Value)
S&P Small-Cap 600 Value	9.0	4	53	SV (Small Value)
S&P Small-Cap 600	9.0	5	61	SC (Small Core)
S&P Small-Cap 600 Growth	8.2	6	50	SG (Small Growth)
S&P 500/Citigroup Value	6.6	7	46	LV (Large Value)
S&P 500	5.9	8	60	LC (Large Core)
S&P 500/Citigroup Growth	4.8	9	35	LG (Large Growth)



The graph's vertical axis shows "Index Outperformance Versus Active Managers" in percentage terms—the higher the number, the better indexes did against active managers (80 means the index outperformed 80% of similar active managers). The horizontal axis presents "Index Performance Rank," i.e., the nine indexes' performance is ranked from highest (1) to lowest (9). In an effort to quantify and better present relationships among the nine asset classes, a trend line fitted to the data points is shown. Lastly, observe that data points above the horizontal 50% line represent indexes that beat over half of comparable active managers. Data points under the 50% line signify indexes that lost to more than half of similar active funds.

The Purity Hypothesis' validity is assessed by identifying its predictions and seeing how many are right. We test the Purity Hypothesis by assuming it makes six predictions:

1. Indexing will beat active management in the highest-returning asset class.
2. Indexing will beat active management in the second-highest-returning asset class.
3. Indexing will beat active management in the third-highest-returning asset class.
4. Active management will beat indexing in the lowest-returning asset category.
5. Active management will beat indexing in the second-lowest-returning asset category.
6. Active management will beat indexing in the third-lowest-returning asset category.

We define the Purity Hypothesis as working if four or more of these predictions are correct and if Spearman's rank correlation coefficient is equal to or greater than 0.6. In our view, the data in Exhibit 1 support the Purity Hypothesis because five of the six predictions are right: the data points for the three top-performing asset classes lie above the 50% horizontal line, and two of the three data points for the worst-performing asset classes are below that line. In addition, Spearman's rank correlation coefficient is 0.87.

On the other hand, the S&P data provide little support for the Efficiency Hypothesis. We test the Efficiency Hypothesis by assuming it makes six predictions:

1. In the small-cap core asset class, active management will beat indexing.
2. In the small-cap value asset class, active management will beat indexing.

3. In the small-cap growth asset class, active management will beat indexing.
4. In the large-cap core asset category, indexing will beat active management.
5. In the large-cap value asset category, indexing will beat active management.
6. In the large-cap growth asset category, indexing will beat active management.

We define the Efficiency Hypothesis as working if four or more of these predictions are right. However, the data in Exhibit 1 do not endorse the Efficiency Hypothesis since four of the six predictions are wrong: the data points for two of three small-cap asset classes lie above the 50% horizontal line, and two of three large-cap data points are below that line.

Also, note that survivor bias in the Morningstar data overstates active management performance versus indexing, i.e., when the active manager returns used in Exhibit 1 are calculated, funds that have gone out of business or merged into others are not counted. Since liquidated or merged funds typically have poor returns, the historical performance of remaining funds is inflated versus comparable indexes. Carhart et al. [2002] found that active equity managers' performance advantage due to survivor bias over 10 years totaled 66 basis points.

Exhibit 1's results comport with those found in Appendix 1 (Appendices are found at <http://www.haifc.com/PressRoom/articles/PurityHypothesisAppendices.pdf>). Appendix 1 has identical analyses using MSCI, Russell, and Morningstar domestic equity indexes for the 10 years ended 2007. These analyses show that the Purity Hypothesis held true for these indexes too. Appendix 2 uses data from these S&P, MSCI, Russell, and Morningstar analyses to show that the Purity Hypothesis worked better than the Efficiency Hypothesis for the 10-year period ended 2007.

THE PURITY HYPOTHESIS OVER SINGLE-YEAR PERIODS

Exhibit 1 and Appendices 1–2 endorse the Purity Hypothesis for the 10 years ended 2007. During this period, however, the Purity Hypothesis did not work in every individual year for S&P, MSCI, and Russell, but it did work every year for Morningstar. We define the Purity Hypothesis as working if four or more of its six predictions

EXHIBIT 2

Years that the Purity Hypothesis Did Not Work By Index Provider, 1998–2007

	Years That the Purity Hypothesis Did Not Work
S&P	1999–2004, 2007
MSCI	1998, 1999, 2001, 2006, 2007
Russell	1999, 2001, 2002, 2006, 2007
Morningstar	None

per index are right and if Spearman's rank correlation coefficient is 0.6 or more. Appendix 3 shows when the Purity Hypothesis worked and didn't work in each year for S&P, MSCI, Russell, and Morningstar over the 1998–2007 period. Exhibit 2 identifies the individual years that the Purity Hypothesis did not work by index provider.

Note that for S&P, MSCI and Russell, as the time period for analysis lengthened from 1 year to 10 years, the Purity Hypothesis became more evident. We think the Purity Hypothesis worked more consistently for Morningstar due to its greater style purity versus S&P, MSCI, and Russell.⁴

CONCLUSION

Conventional wisdom says that domestic equity indexing is more likely to outperform active managers in efficient markets like large stocks and is more apt to underperform active managers in less-efficient areas like small stocks. This idea is called the Efficiency Hypothesis, but data from 1998–2007 do not support it. However, 1998–2007 data do show that domestic equity indexing tends to outperform active management in the highest-returning asset classes and tends to underperform active managers in the lowest-returning asset categories. This concept is called the Purity Hypothesis. The Purity Hypothesis held true over the 10-year period ended 2007 for the S&P, MSCI, Russell, and Morningstar domestic equity indexes. But its effectiveness on a single-year basis during this time varied by index provider.

ENDNOTES

¹Bernstein [1999, 2000] wrote that his colleague, Steven Dunn, noticed that when an asset class did well, an index fund in that asset class did even better. Bernstein called this Dunn's Law.

²Waid [2008] presents evidence that small-cap managers buy larger stocks than those found in the Russell 2000.

³Active manager cash is sometimes offered as the sole cause of the Purity Hypothesis. The explanation goes like this: if cash underperforms the top equity asset classes' returns by a wide margin, then this cash drag accounts for active managers' losing to indexes in high-return equity categories. In the same vein, if cash greatly outperforms the worst equity groups' returns, then this cash return explains active manager outperformance versus indexes in the bottom-performing equity classes. Rekenhaller [1999] tested this explanation using some simplifying assumptions and found that manager cash had only a small effect. Our tests on this topic incorporated simplifying assumptions and yielded similar results. See Appendix 4 for details.

⁴Morningstar's index construction differs from that of MSCI, Russell, and S&P in two key ways, which helps make Morningstar's indexes more style-pure. First, stocks in Morningstar's core indexes are not found in its growth and value indexes. On the other hand, MSCI, Russell, and S&P core indexes hold the stocks found in their respective growth and value indexes. Second, Morningstar has no overlapping holdings between growth and value indexes. However, the same stock can be held in both growth and value indexes at MSCI, Russell, and S&P. These differences may help explain why the Purity Hypothesis worked in Morningstar indexes on a single-year basis but not in those from MSCI, Russell, or S&P.

REFERENCES

- Bernstein, William. "When Indexing Fails." April 1999. Available at <http://www.efficientfrontier.com/ef/499/indexing.htm>.
- . "The Dunn's Law Review." Spring 2000. Available at <http://www.efficientfrontier.com/ef/400/dlr.htm>.
- Carhart, Mark M., Jennifer M. Carpenter, Anthony W. Lynch, and David K. Musto. "Mutual Fund Survivorship." *Review of Financial Studies*, Vol. 15, No. 5 (Winter 2002), pp. 1439–1463.
- Rekenhaller, John. "When to Index? Dunn's Law Trumps the Efficiency Theory." *Journal of Financial Planning*, June 1999, pp. 34–36.
- Sharpe, William F. "The Arithmetic of Active Management." *Financial Analysts Journal*, January/February 1991, pp. 7–9.
- Waid, Robert. "Examining Small-Cap Manager Investable Universes: Why the Russell 2000 Missed the Mark." *Journal of Indexes*, May/June 2008, pp. 30–33.

To order reprints of this article, please contact Dewey Palmieri at dpalmieri@ijournals.com or 212-224-3675.