www.joim.com

Journal Of Investment Management, Vol. 13, No. 3, (2015), pp. 93–103

THE VALUE OF ACTIVE INVESTING*

Craig William French[a]

We examine whether the value of active investment management can exceed its cost, and find that it can, by a substantial margin. We consider the 0.67% average cost estimate in French (2008), comparing it with the expected value of a known active investment strategy. For a "passive" benchmark, we develop a 225 year index of monthly U.S. equity market returns, from July 1789 through June 2014. This timeframe encompasses the entire history of every U.S. exchange and includes all known periods of secondary market stock trading in the United States. We then estimate the long-run monthly returns of an active investment strategy based on an actual 11-year investment strategy. We present a new performance model extending the Treynor and Mazuy (1966) model and implement it to estimate monthly returns to the active strategy over the same 225 year period. We believe that this experiment offers a good example of how the value of well-constructed active investment strategies can be worth substantially more than their cost.

1 Introduction

"What is a cynic? A man who knows the price of everything and the value of nothing."

Oscar Wilde, *Lady Windemere's Fan* (1892)

The aggregate cost of active investing has been estimated at 0.67% annually by French (2008), which compares the costs "society pays to invest in the U.S. stock market with an estimate of what would be paid if everyone invested passively". Of course, this comparison contrasts a reality (actual active investing) with an impossibility (100% passive investing). When we define (at least some portion of) active investors as informed investors and passive investors as uninformed investors, it becomes clear that the latter—all passive investing—is a condition that cannot persist because it would lead to market failure, as shown in Grossman and Stiglitz (1980).

We interpret the French (2008) result in the context of the parsimonious "noisy rational information flow" model of Grossman and Stiglitz (1980)

*Views expressed in this article are those of the author, and may not represent the views of WBI Investments, Inc. or its affiliates.

[a]Director of Investments, WBI Investments, One River Centre, 331 Newman Springs Road, Suite 122, Red Bank, NJ 07701. Tel.: 732-842-4920, e-mail: crgfrench@gmail.com

to mean, 0.67% – or the 0.61%–0.74% general range found in French (2008)—is the cost of manager compensation plus the cost of information gathering and dissemination.[1] As Grossman and Stiglitz (1980) point out, "... prices perform a well-articulated role in conveying information from the informed to the uninformed ... the price system makes publicly available the information obtained by informed individuals to the uninformed [and] were it to do it perfectly, an equilibrium would not exist." While Grossman and Stiglitz (1980) focus on the impossibility of the efficient market hypothesis as developed and reviewed in Fama (1970), we take as a given that competitive equilibrium is established *because of* costly information and costly transactions, and address some normative questions that have been raised by others.

Both French (2008) and Grossman and Stiglitz (1980) explore positive economic aspects of the costs of active investing, leaving aside the social consequences and welfare effects of active versus passive investing. Others, including Malkiel (1973) and Bogle (2008), have made normative economic conclusions about the issue. For example, Bogle (2008) laments "the burdensome costs of financial intermediation" despite his own recognition that financial intermediation "...creates substantial value for our society. It facilitates the optimal allocation of capital... it enables buyers and sellers to meet efficiently; it provides remarkable liquidity; it enhances the ability of investors to capitalize on [market prices]; it creates [risk transfer benefits]..." Bogle (2008) asks "...whether, on the whole, the costs of obtaining those benefits have reached a level that overwhelms them." He points to "soaring costs," citing an increase in average weighted expense ratios of mutual funds from 0.60% in 1951 to 0.87% in 2007. In contrast, French (2008) finds that the total cost of active investing is relatively constant over the 27 years from 1980 to 2006, although "the components change a lot over time." The difference between French's result and Bogle's claim is due to the fact that Bogle cites only one component of the total fees and expenses of mutual fund investing (the expense ratio), whereas French considers both components of the total (the expense ratio and the annuitized load). French (2008) finds that, from 1980 to 2006 the value weighted average expense ratio increased from 70 basis points to 85 basis points, while simultaneously the weighted average load paid by investors declined from 149 basis points to only 15 basis points. French notes, "Driven by a steady decline in the loads open end fund investors pay, the [total] fees and expenses for mutual funds fall from 2.08%... in 1980 to 0.95% in 2006."

Jones and Wermers (2011) provide a comprehensive survey of the literature surrounding the active versus passive investing debate, and conclude that, while the *average* active manager does not outperform the market after all fees, *a significant minority* of active managers do add value. Jones and Wermers (2011) suggest four potential criteria for identification of superior asset managers, including evaluation of past performance, macroeconomic correlations, fund/manager characteristics, and fund holdings. Their conclusion is that identification of superior managers or strategies leads to expected positive alpha, or substantial return advantages relative to the market. In the present paper, we propose a new methodology to evaluate the dynamics of past performance, and apply it to a known investment strategy to estimate, for this investment strategy, the potential long-run benefits of active management. We find that the value of active management for this strategy is +1.68% per year, eclipsing the cost found by French (2008) by a substantial margin.

The remainder of this paper proceeds as follows: Section 2 develops a two and a quarter century index of monthly U.S. equity market returns.

This index serves as our "passive" benchmark. Section 3 then estimates the monthly returns of an active investment strategy based on an actual 11-year investment strategy. We present a performance model extending the Treynor and Mazuy (1966) model and implement it to estimate monthly returns to the active strategy over the same 225 year period covered by the passive index. Section 4 concludes with some thoughts about the results, implications for investors and inferences about the costs of active investing versus the potential value of active investing.

2 Passive returns

"The sea, the sea… Man alone, Passive, unaware in his elemental sadness."

Scott Hastie

We develop a 225 year index of monthly U.S. equity market total returns, from July 1789 through June 2014. This timeframe encompasses the entire history of every U.S. exchange and includes all known periods of secondary market stock trading in the United States. The 225 year index serves as our passive benchmark for the following section, in which we evaluate an active investment strategy. For a long-run view, our index of U.S. stock market returns begins prior to the inception of the PHLX in 1790 and the NYSE in 1792. Two long-run studies, Schwert (1990) and Siegel (1994), employ data going back to 1802. These both use Smith and Cole (1935), Macaulay (1938) and Cowles and Associates (1939) data for early periods, and then CRSP and other data for later periods. We utilize the same early sources and prepend more than an additional decade of monthly returns derived from Foundation for the Study of Cycles (1975) as described in Shirk (1978). After 1927, we employ the S&P 500 Index total return monthly series from Bloomberg.

We develop the passive index as a complex chained series as described in Schwert (1990). We choose price series from the sources listed in Table 1. Without splicing, we adjust these price series by adding in estimated dividend yields to those series which lack them. Schwert (1990) notes that during this early period, "…there is no evidence of a secular change in the level of dividend yields… thus, there is reason to believe that dividend yields before 1871 were similar to those measured after 1871." We therefore utilize the mean monthly dividend estimates from the Cowles column in Table 3 of Schwert (1990), adding them to each monthly return prior to February 1871. We do not employ the filter of Schwert (1990) to correct for the effects of time averaging noted in Working (1960) because (1) the magnitude of the problem is unknown and unknowable; (2) the series are not random

Table 1 Return sources of the passive index.

Source	*Begin*	*End*	*Dividends*
Foundation for the Study of Cycles (1975)	July 1789	December 1801	No (added Schwert Table 3)
Smith and Cole (1935, Table 61)	January 1802	December 1820	No (added Schwert Table 3)
Smith and Cole (1935, Table 62)	January 1821	December 1833	No (added Schwert Table 3)
Smith and Cole (1935, Table 69)	January 1834	December 1842	No (added Schwert Table 3)
Smith and Cole (1935, Table 70)	January 1843	January 1857	No (added Schwert Table 3)
Macaulay (1938, Table 10)	February 1857	January 1871	No (added Schwert Table 3)
Cowles and Associates (1939, Table C-1)	February 1871	January 1928	Yes
Bloomberg, S&P 500 TR	February 1928	June 2014	Yes

chains; and (3) as noted in Schwert (1990), the univariate filters do not resolve the potential problem. Although the inception of our index precedes that of Schwert (1990) by more than a decade, over the 124 year common period (1,488 months) January 1802 through December 1925, the differences in our estimates of monthly total return estimates have a median of zero and a mean of −0.025% per month (our series has a slightly lower mean return and variance than Schwert's, mainly because we do not employ the Working effect filters[2]). Our index grows at a very similar rate as Schwert's, with a mean of 0.575% monthly return versus Schwert's 0.600% and a sample standard deviation of 3.27% per month versus Schwert's 3.82% over the same period. So, as expected, our passive index is quite similar to that of Schwert (1990).

The approach to estimating early returns taken by Schwert (1990) and Siegel (1994) has been criticized by Zweig (2009). We, adopting similar methods as Schwert (1990) and Siegel (1994), obviously disagree with Zweig (2009). Zweig's criticisms include an assault on the validity of the Smith and Cole (1935) indices as "cherry-picked" and subject to survivorship bias, as well as being non-comprehensive and/or unrepresentative of stocks commonly held by investors. For example, Smith and Cole (1935) used 7 of 38 stocks for the index in their Table 60. Zweig finds great fault with this selectivity; we merely note that $7/38 = 18.42\%$ coverage, which is far more comprehensive than one of the most popular current stock market indices, the Dow Jones Industrial Index. The Dow incorporates 30 out of 9,095 U.S. listed equities, representing $30/9{,}095 = 0.03\%$ coverage. Even allowing for the exclusion of so-called micro-capitalization stocks, those with market capitalizations less than, say $300 million, the Dow Industrial Index represents $30/2,906 = 1.03\%$ coverage of the U.S. equity market. Perhaps the indices of Smith and Cole (1935) are not so bad after all.

3 Active returns

"Price is what you pay. Value is what you get."

Warren Buffett

Models of active investing abound. Some examples include Williams (1938), Livermore (1940), Graham (1949), Markowitz (1959), Thorpe (1967), Treynor and Black (1973), Klarman (1991), and Grinold and Kahn (1995). Lo and MacKinlay (1988) offer an econometric counterpoint to Malkiel (1973), providing theoretical justification for active management. In this section, we develop an estimate of the value of active investing for an active investment strategy. The active strategy is philosophically similar to that of Graham (1949) and Klarman (1991), and is described in Schreiber and Stroik (2011), with further detail provided in Schreiber and French (2014). Our goal is to determine whether the value of active investing exceeds the cost, and, if so, by potentially how much.

Data for the active strategy is derived from the historical monthly gross performance composite of the WBI Dividend Income strategy from its inception, July 1, 2003 through December 2013. To model the market capture dynamics of the active strategy, we extend the quadratic model of Treynor and Mazuy (1966) into a cubic form. The following discussion outlines the development of our approach.

The CAPM as originally developed in Treynor (1961) and Sharpe (1964) follows the linear form:

$$Y = bX + a \tag{1}$$

The performance model as developed in Treynor and Mazuy (1966) follows the quadratic form:

$$Y = cX^2 + bX + a \tag{2}$$

The performance model we utilize in the present study follows the cubic form:

$$Y = dX^3 + cX^2 + bX + a \qquad (3)$$

The intuition for Equation (1) is well-known, this is the linear single-factor model estimated via ordinary least-squares regression. This approach predominates in the field of finance; other early influential examples of its use includes Black *et al.* (1972) and Fama and French (1993). We can best interpret Equations (2) and (3) as polynomial regressions of order 2 (quadratic) and order 3 (cubic), respectively. While all of the regression functions are linear in terms of the unknown parameters $a, b, c,$ and d, for the present purpose both Equations (2) and (3) are superior to Equation (1) because of their ability to model nonlinearities in the relationship between X and Y. Polynomial regressions are technically nothing more than special cases of multiple linear regressions. Mathematically, these models use basis functions to model the functional relationship between two quantities—in this case, the monthly returns of a passive investment strategy and an active investment strategy. Appendix A provides a brief treatment of Equations (1)–(3).

The active strategy is designed to attempt to realize nonlinear "upcapture" and "downcapture"—it seeks to maximize the former while minimizing the latter, relative to the passive strategy. A model such as Equation (1) cannot accommodate such dynamics. Treynor and Mazuy (1966) showed that Equation (2) can be a useful model to detect nonlinearities. We find Equation (3) to be an even more utile model because it can generate more inflection points. The characteristic lines of best fit from Equation (3) are intuitive to readers familiar with derivatives—they tend to appear very similar to options profit & loss charts, allowing the analyst to easily visualize gamma characteristics in the relationship if we consider Y to be a derivative of X. Equation (2) also allows such visualizations, however the benefit of Equation (3) over Equation (2) for the analyst is that while Equation (2) allows for the visualization of single options (long or short a put or call), Equation (3) additionally allows for the interpretation of combinations of options (collars, spreads, etc).

Least-squares minimization with active monthly total returns of the WBI Dividend Income strategy as the dependent (Y) variable and S&P 500 Index monthly total return data from Bloomberg as the independent (X) variable over the 10.5 year period from inception 7/1/2003 through 12/31/2013 yields the following model parameters for Equation (3):

$$a = +0.0007; \quad b = +0.6448;$$
$$c = +1.2912; \quad \text{and} \quad d = -3.3789$$

This regression is illustrated in Figure 1. Notice the clear convexity in the characteristic line. This curvature is identified primarily in the c parameter, which exhibits a t-statistic of 2.79, significant at the 99.9% level with a p-value of 0.006. The b parameter is even stronger, with a t-statistic of 15.23, statistically significant at the 100.0% level with a vanishingly small p-value of 2e$-$16. In this particular case, the negative value of the d parameter did not add significantly to the power of the model, with an insignificant t-value of -0.783 (had the cubic term been excluded from our model, the compound annualized rate of total return shown in Table 2 for the active strategy would have increased by 10 bp annually, to 9.48%; in this case, the effect of including the cubic term was to slightly reduce the return, relative to a quadratic model). Overall, the model itself is extremely robust, with an F-statistic of 146.5 and a p-value of 2.2e$-$16. We are therefore essentially certain that the active strategy is convex, meaning it behaves like a call option, limiting downside risk.

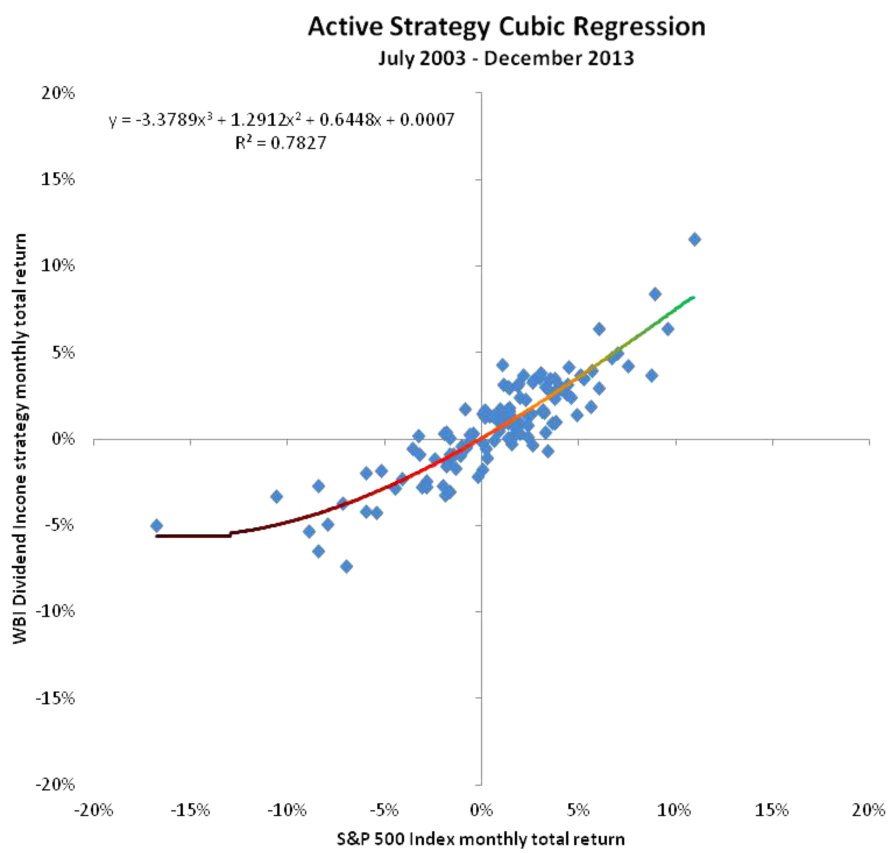

Figure 1 Cubic regression of an active investment strategy.

We then estimate the hypothetical historical active return series by applying this model, which has a coefficient of determination of 0.7827 (adjusted = 0.7773), to the monthly passive return series developed previously in Section 2 over the entire 225 year period from July 1789 through June 2014. Table 2 details the risk and return characteristics of both the passive and the active strategies, while Figure 2 displays the potential long-run value of the active investment strategy. The difference is large: the passive index generates a compound annualized 7.69% rate of total return with 15.59% annualized standard deviation, while the active strategy generates a compound annualized 9.38% rate of total return with 9.55% annualized standard deviation. The difference in rates of return is +1.68% annually, with risk reduction of 39%. Even adjusting

Table 2 Active and passive strategy risk and return characteristics, July 1789–June 2014.

	Compound Annualized Rate of Total Return	Annualized Standard Deviation	Linear OLS Beta	Annualized Linear OLS Alpha
Passive index	7.69%	15.59%	1.00	0.00%
Active strategy	9.38%	9.55%	0.59	4.44%
Active minus passive	1.68%	(6.04%)	(0.41)	4.44%

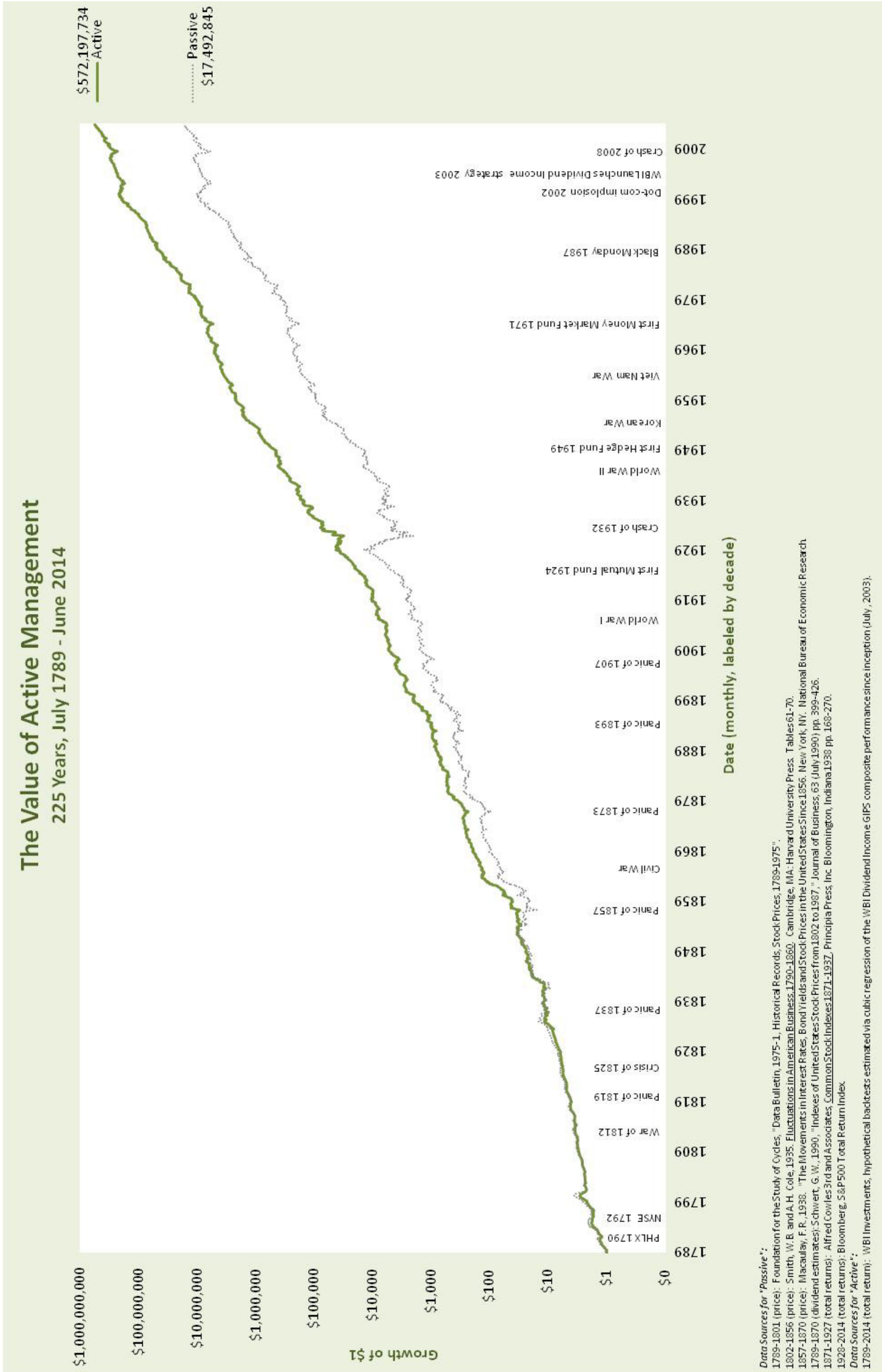

Figure 2 The long-run value of an active investment strategy.

for the 0.67% cost of active investing found in French (2008), this active strategy outperforms the passive approach by 1.01% annually. Another way to interpret this result is that rational investors would be indifferent between paying 1.68% annually for the active strategy and the passive index—at any fee less than 1.68%, the active strategy dominates. In risk-adjusted terms, the difference is even more striking, and risk-averse rational investors would potentially find even higher utility for the active strategy than 1.68% annually.

4 Conclusions

"The cost of a thing is the amount of what I will call life which is required to be exchanged for it, immediately or in the long run."

Henry David Thoreau, *Walden* (1854)

Is the *value* of active investing worth its *cost*? On average across all active strategies, French (2008) finds the *cost* to be (0.67%) annually. We find the *value* to be +1.68% annually over the long run for the active strategy examined. Rational economic agents would pay sixty-seven cents to receive \$1.68. Clearly there is social value in active investing, even for the average active investment strategy. The benefits of active investing include optimal allocation of capital; provision of an efficient marketplace; remarkable liquidity facility; enhancement of investors' ability to accumulate and grow wealth; and creation of important risk transfer benefits. The French (2008) finding implies that U.S. equity investors are willing to pay about two-thirds of a percentage point each year for such benefits. A different interpretation of the French (2008) cost estimate could be as follows: "On average, active investment strategies underperform a passive investment approach by 0.67% each year." Obviously there is dispersion around an average; Jones and Wermers (2011) suggest that it can be profitable to find managers offering active investment strategies with positive expected return. We agree.

In this paper we have provided an example of one active investment strategy that we believe has a positive expected return, and built a model to estimate the long-run value that could have been created by such an approach. To do so, we developed an index of early U.S. stock total returns that precedes the CRSP data by more than a century, and precedes the early data used by Siegel (1994) and Schwert (1990) by more than a decade. We have extended the model of Treynor and Mazuy (1966) from a quadratic model to a cubic model and applied it to estimate the dynamics of the active investment strategy over the entire history of secondary market securities trading in the United States, over the 225 year period July 1789 through June 2014.

Future directions for similar work might include a multifactor analysis, seeking to distill components of so-called "priced factors" using multifactor models such as Fama and French (1993) or Carhart (1997). For our present purpose, in which we seek to compare the value of an active investment strategy with the cost, the single-factor approach is appropriate and sufficient.

This example is only one of many possible active investment strategies that may exist on the right side of the distribution of active investment strategy returns. Our results indicate that the value of well-constructed active investment strategies can be worth substantially more than their cost. We conclude that the value of active investing can, for certain strategies at least, substantially exceed the cost.

Acknowledgments

I would like to thank Don Schreiber, Jr., Gary Stroik, Matt Schreiber, Bob Confessore, Yuxin

Zhang, Wenhao Zhou, Do Yon Kwon, Jack Hain, Tracey Crespo, and an anonymous referee. I am also grateful to Shirley French, Kiely French, Connor French, Jenna French and Hannah French for giving me inspiration and joy. I alone am responsible for any remaining errors.

Notes

[1] Readers, especially those familiar with mutual fund fees, may anecdotally find that the French (2008) estimates seem low. Cost data may be particularly time-dependent, as noted in French (2008). Higher cost estimates would argue against active management. Recall that French (2008) estimates the cost of investing across all market participants (including direct holdings, open-end funds, closed-end funds, DB plans, DC plans, ESOPs, public funds, nonprofits, institutional investors, hedge funds and foreign investors), not just mutual funds. In the aggregate, all of these investors pay lower fees than would seem to be indicated by considering only mutual funds.

[2] Our approach is very similar to, but not identical to, that of Schwert (1990). Differences include:

(1) we do not employ univariate filters to correct for the Working effect of time-series averaging;

(2) we do not splice any of the original series, rather we select independent, abutting time series and chain them;

(3) our selection of the "best" index(es) over each early period differs slightly from Schwert's, although we believe the differences are essentially immaterial to the results; interested readers may compare our Table 1 with Schwert's Table 5 on pp. 414 of Schwert (1990);

(4) Schwert's first month, February 1802, is derived from Smith and Cole (1935, Table 61); Smith and Cole's index value for January 1802 is 97. Schwert calculates a February return from the 98 February index value and the 97 January index value; we assume that the Smith and Cole index begins at 100 in the beginning of January 1802 and the 97 January value is as of the end of that month. For the January 1802 return, our estimate of negative (2.49%), is comprised of −3% price return implied by Smith and Cole (1935, Table 61) (under our assumption) plus the 51 bp January mean dividend estimate from Schwert (1990, Table 3).

References

Anscombe, F. J. (1973). "Graphs in Statistical Analysis," *The American Statistician* **27**(1), 17–21.

Black, F. S., Jensen, M. C., and Scholes, M. S. (1972). "The Capital Asset Pricing Model: Some Empirical Tests," in M. C. Jensen, *Studies in the Theory of Capital Markets*, New York, NY: Praeger Publishers.

Bogle, J. C. (2008). "A Question So Important That It Should Be Hard To Think of Anything Else," *Journal of Portfolio Management* **34**(2), 95–102.

Carhart, M. M. (1997). "On Persistence in Mutual Fund Performance," *Journal of Finance* **52**(1), 57–82.

Cowles, A. III, and Associates (1939). *Common Stock Indexes*, 2nd edn., Cowles Commission Monograph #3. Bloomington, IN: Principia Press.

Fama, E. F. (1970). "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance* **25**(2), 383–417.

Fama, E. F. and French, K. R. (1993). "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics* **33**(3), 3–56.

French, K. R. (2008). "The Cost of Active Investing," *Journal of Finance* **63**(4), 1537–1573.

Foundation for the Study of Cycles (1975). *Data Bulletin, 1975-1, Historical Records, 1789–1975.*

Graham, B. (1949). *The Intelligent Investor.* New York: Harper & Brothers.

Grinold, R. C. and Kahn, R. N. (1995). *Active Portfolio Management*. New York, NY: McGraw-Hill.

Grossman, S. J. and Stiglitz, J. E. (1980). "On the Impossibility of Informationally Efficient Markets," *The American Economic Review* **72**(4), 393–408.

Jones, R. C. and Wermers, R. (2011). "Active Management in Mostly Efficient Markets," *Financial Analysts Journal* **67**(6), 29–45.

Klarman, S. A. (1991). *Margin of Safety*. New York, NY: HarperCollins Publishers.

Livermore, J. L. (1940). *How to Trade in Stocks.* New York, NY: Duell, Sloan and Pierce.

Lo, A. W. and MacKinlay, A. C. (1988). "Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test," *Review of Financial Studies* **1**(1), 41–66.

Macaulay, F. R. (1938). *The Movements of Interest Rates, Bond Yields and Stock Prices in the United States Since*

1856. New York, NY: National Bureau of Economic Research.

Malkiel, B. G. (1973). *A Random Walk Down Wall Street*. New York, NY: W.W. Norton & Co.

Markowitz, H. M. (1959*). Portfolio Selection: Efficient Diversification of Investments.* New York: John Wiley & Sons.

Schreiber, D. Jr. and French, C. W. (2014). *Benefits of Optimizing Portfolio Capture Ratios*. Red Bank, NJ: WBI Investments, Inc.

Schreiber, D. Jr. and Stroik, G. E. (2011). *All About Dividend Investing*, 2nd ed. New York, NY: McGraw-Hill.

Schwert, G. W. (1990). "Indexes of U.S. Stock Prices from 1802 to 1987," *Journal of Business* **63**(3), 399–442.

Sharpe, W. F. (1964). "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *Journal of Finance* **19**(3), 425–442.

Shirk, G. (1978). "The Trend and Long Cycle in Stock Prices," *Cycles* **29**(3), 53–59.

Siegel, J. J. (1994). *Stocks for the Long Run*. New York, NY: McGraw-Hill.

Smith, W. B., and Cole, A. H. (1935). *Fluctuations in American Business, 1790-1860*. Cambridge, MA: Harvard University Press.

Thorpe, E. O. (1967). *Beat the Market, A Scientific Stock Market System*. New York, NY: Random House.

Treynor, J. L. (1961). *Market Value, Time and Risk*. Unpublished draft, August 8.

Treynor, J. L. and Black, F. S. (1973). "How to Use Security Analysis to Improve Portfolio Selection," *Journal of Business* **46**(1), 66–88.

Treynor, J. L. and Mazuy, K. (1966). "Can Mutual Funds Outguess the Market?" *Harvard Business Review* **44**(4), 131–136.

Williams, J. B. (1938). *The Theory of Investment Value*. Cambridge, MA: Harvard University Press.

Working, H. (1960). "A Note on the Correlation of First Differences of Averages in Random Chains," *Econometrica* **28**(4), 916–918.

Zweig, J. (2009). "Does Stock Market Data Really Go Back 200 Years?" *The Wall Street Journal*, July 11.

Keywords: Active investing; passive investing; efficient markets; informed investors; uninformed investors; positive economics; normative economics; polynomial regression; chained price series; pre-CRSP stock returns; dividends; market timing; investment value

Appendix A

Here we illustrate the utility of cubic regression versus quadratic and linear regression. Figure A.1 displays 15 scenarios (a pentadectet) with a variety of data values, all of which have (approximately) identical linear OLS regression coefficients satisfying $y = 0.5x + 3$. The top equation in each chart is the linear fit as in Equation (1), the middle equation is the quadratic fit as in Equation (2) and the bottom equation is the cubic fit as in Equation (3).

In Figure A.1, the top panel of four sets replicates Anscombe's quartet as described in Anscombe (1973). The remaining eleven sets present all possible combinations of Anscombe's four original sets. In all cases, by construction, the cubic regression fits the data as well as or better than the quadratic regression, which in turn fits the data as well or better than the linear regression.

(1) In Anscombe's set I, cubic regression with a coefficient of determination $[R^2] = 68.87\%$ fits the data better than quadratic (68.73%), which fits better than linear (66.65%).
(2) In set II (a quadratic function), both cubic and quadratic regressions yield $R^2 = 100\%$ versus the linear fit of 66.62%.
(3) In set III (the function is $y = 0.346x + 4$ plus a single outlier), the cubic R^2 of 68.84% exceeds the quadratic R^2 of 68.47%, which exceeds the linear R^2 of 66.63%.
(4) In set IV (an orthogonal group with a single outlier), the fit is linear and all three regression methods yield $R^2 = 66.67\%$.

We extend Anscombe's quartet to a pentadectet by producing all remaining combinations of the four original sets, for a total of 15 sets. In every case, linear OLS yields the same fit, $y = 0.5x + 3$, with a coefficient of determination of 66.65% (shown as 0.67 on each chart). Equation (1) is incapable of distinguishing among these 15 sets

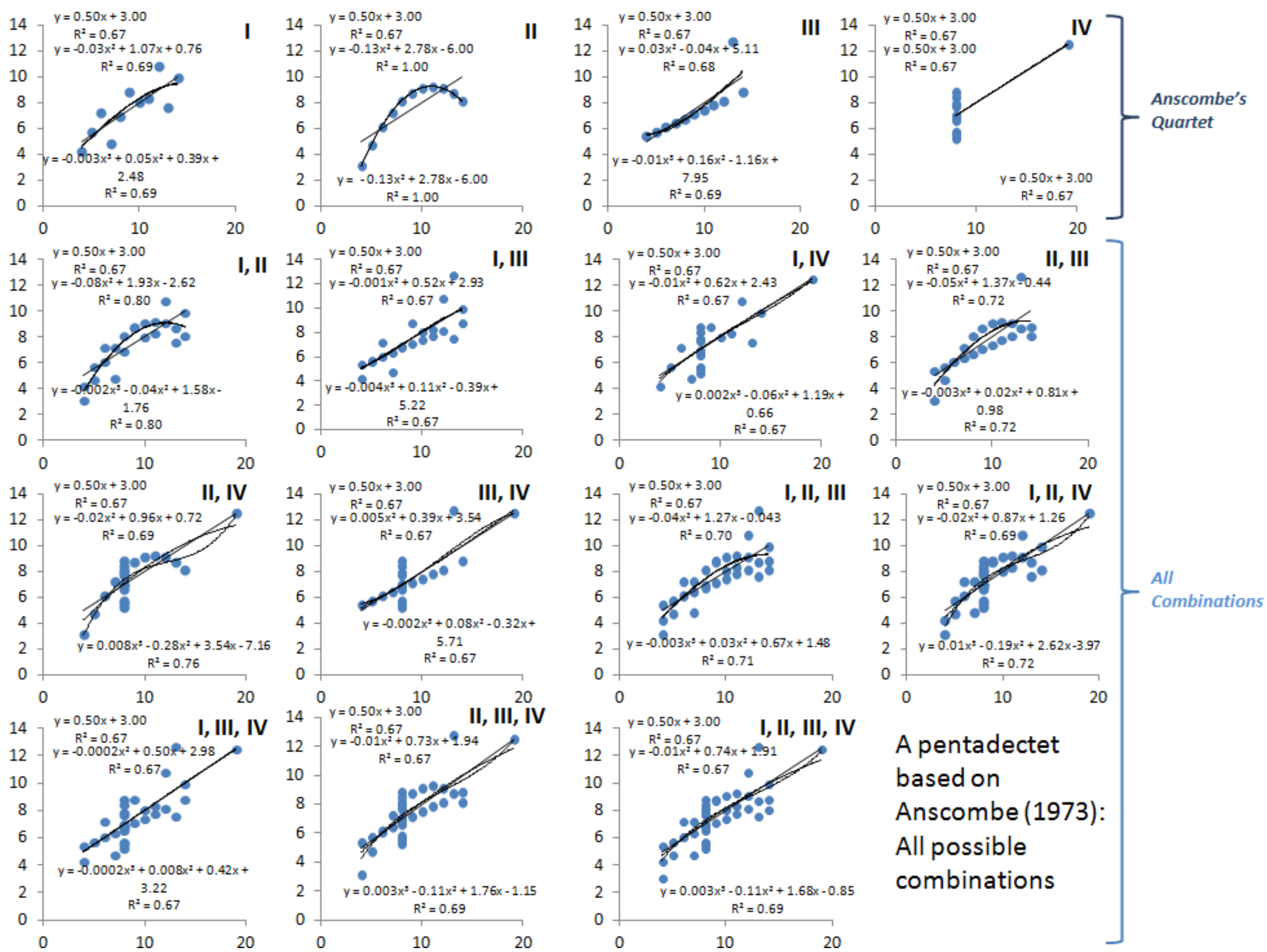

Figure A.1

of data. In contrast, Equations (2) and (3) can. Cubic regression significantly outperforms linear regression in 10 of the 15 sets [I, II, III, I&II, II&III, II&IV, I&II&III, I&II&IV, II&III&IV, and I&II&III&IV], and at least matches linear in the remaining 5 cases. Quadratic regression significantly outperforms linear regression in 8 of the 15 sets, while at least matching linear in the remaining 7 cases. Cubic regression significantly outperforms quadratic regression in 6 of the 15 sets [III, II&IV, I&II&III, I&II&IV, II&III&IV, and I&II&III&IV], and at least matches quadratic in the remaining 9 cases.

Of note, when considering these charts as options profit and loss graphs, in every case Equation (1) would seem to indicate a long underlying position. However, when fitted by Equation (2) or Equation (3), set II appears to be a written straddle; sets I&II, II&III, and I&II&III appear to be written puts. Further, when fitted by Equation (3), sets II&IV, I&II&IV, II&III&IV and I&II&III&IV appear to be split strike synthetic long positions, whereas Equation (2) would less accurately identify them as short put positions.