Sharpening the Arithmetic of Active Management

Lasse Heje Pedersen

To cite this article: Lasse Heje Pedersen (2018) Sharpening the Arithmetic of Active Management, Financial Analysts Journal, 74:1, 21-36, DOI: 10.2469/faj.v74.n1.4

To link to this article: https://doi.org/10.2469/faj.v74.n1.4
Sharpening the Arithmetic of Active Management

Lasse Heje Pedersen

Lasse Heje Pedersen is a principal at AQR Capital Management, Greenwich, Connecticut, and professor at Copenhagen Business School and NYU.

I challenge William F. Sharpe's famous equality that "before costs, the return on the average actively managed dollar will equal the return on the average passively managed dollar." This equality is based on the implicit assumption that the market portfolio never changes, which does not hold in the real world because new shares are issued, others are repurchased, and indexes are reconstituted—so even "passive" investors must regularly trade. Therefore, active managers can be worth positive fees in aggregate, allowing them to play an important economic role: helping allocate resources efficiently. Passive investing also plays a useful economic role: creating low-cost access to markets.

Sharpe's (1991) famous "arithmetic of active management" states that it must be the case that

(1) before costs, the return on the average actively managed dollar will equal the return on the average passively managed dollar, and

(2) after costs, the return on the average actively managed dollar will be less. . . .

These assertions will hold for any time period. Moreover, they depend only on the laws of addition, subtraction, multiplication and division. Nothing else is required. (p. 7; italics in the original)

Sharpe's arithmetic has been invoked by Warren Buffett, is often stated as incontrovertible fact by speakers at conferences (followed by a triumphant "QED"!), and is cited as proof that active management is "doomed" in aggregate (French 2008). If active management is doomed in aggregate, then so is our market-based financial system because we need someone to make prices informative. However, we may avoid doom on the basis of my arithmetic.

Sharpe's powerful insight is that one active investor's gain is another active investor's loss, which aggregates to zero for all active investors. This useful insight is correct when considering a fixed set of securities over a single time period, but in the real world, the set of securities in the market changes over time.
Sharpe’s argument thus abstracts from a key aspect of addition and subtraction—namely, the addition of new companies and shares and the subtraction of disappearing ones. Although seemingly minor, the market portfolio does change over time, such that even “passive” investors must regularly trade—for instance, in connection with issuances, share repurchases, index inclusions and deletions, and so on. Whenever passive investors trade in order to maintain their market-weighted portfolios, they may trade at less favorable prices than active managers do, which breaks Sharpe’s equality.

This turnover of the market portfolio is important for two reasons. First, the changes of the market portfolio may be large enough that active managers in aggregate can add modest, yet noticeable, returns relative to passive investors. Second, the issuance of securities is at the heart of a market-based economy: Capital markets are about raising capital. When we put these reasons together, we see that active management can be worth positive fees in aggregate, but whether it actually does and how much are empirical questions. On the basis of realistic arithmetic, we need to empirically evaluate the costs and benefits of active management. Investors should understand how fees diminish performance, but we should not expect to be able to allocate global capital in a market-based system without any active management.

Sharpe’s Hidden Assumption: A Market without Addition or Subtraction

Sharpe’s (1991, p. 7) active management arithmetic, in its beautiful simplicity, holds that each passive manager will obtain precisely the market return, before costs. From this, it follows (as the night from the day) that the return on the average actively managed dollar must equal the market return. Why? Because the market return must equal a weighted average of the returns on the passive and active segments of the market. If the first two returns are the same, the third must be also.

Sharpe’s argument relies on the notion of a “passive investor,” but what does this really mean? Two definitions seem plausible:

1. an investor who holds all securities according to their market-capitalization weights or
2. an investor who never trades.

Sharpe defines a passive investor using Definition 1, but people tend to make the implicit assumption that Definition 1 and Definition 2 are equivalent. Indeed, in a world with a fixed set of securities (i.e., no issuance or repurchases), Definition 1 does (sort of) imply Definition 2; that is, if you start with market-cap weights, your portfolio remains market-cap weighted no matter how prices change (which is a helpful property of market-cap weights).

But—as anyone who has tried this in practice knows—you first need to buy your portfolio, and you eventually need to sell it. Furthermore, holding a market-cap-weighted portfolio does require trading because securities come in and out of the market as shares are issued, shares are repurchased, and indexes are reconstituted.

If you read Sharpe’s argument carefully, he states that his conclusion holds for “any time period.” Strictly speaking, if passive investors somehow magically arrive at the beginning of any time period with all securities at their market-cap weights bought at the current midquote, then his conclusion follows and is correct. But real life is more than one time period, and it is important to note that “any time period” is not the same as “over all time periods.” Passive investors must trade to achieve their market-cap weights. Sharpe’s equality breaks down when we take this trading activity into account. Indeed, if passive investors, on average, buy at a premium and sell at a discount relative to active investors, then active managers can outperform passive investors before fees. Appendix A contains a simple equilibrium model in which active managers outperform passive managers before fees, while the after-fee performance
difference naturally depends on the level of fees relative to the added value.

So, the implicit assumption lies in Sharpe’s definition of passive and the abstraction from trading. Sharpe discusses trading in a footnote to his article, and although the footnote is precise and helps clarify the issue, others have seemingly missed this important point and incorrectly interpreted his result as a truth that relies on no assumptions other than the laws of arithmetic. Some researchers have even claimed that his argument “always” holds, even for subsets of the market and trading strategies with larger turnover. For instance, Fama and French (2009) wrote, “The same arguments apply whatever one takes to be the market, for example, value stocks, growth stocks. ... Active investors can only win at the expense of other active investors. In short, active investing in any sector is always a zero sum game.” Here, the potential error is clearly larger than in the case of a market-weighted equity index because of the larger turnover required to track a value or growth strategy. Indeed, “passive” value investors may lose from their trading as stocks regularly switch between being classified as value and growth.

How Active Managers Can Outperform: Market Is Not Buy and Hold

First, I will provide some conceptual examples of why active management can outperform owing to the necessary trading of passive investors. Then, in the next sections, I will discuss the magnitude of this passive trading in the real world and the resulting performance impact. Appendix A contains a simple equilibrium model in which active managers beat the market in aggregate.

Example 0. Who Are the “Active Managers”? Most active managers typically suggest that they add value by selecting good securities that outperform the market, not necessarily exploiting price moves related to changes in the market portfolio. Sharpe’s powerful argument shows that the activity of selecting good securities within a fixed set of securities is a zero-sum game, so not all managers can win this game. Although some managers do, in fact, emphasize their added value in connection with changes in the market (e.g., the so-called event-driven hedge funds), let us nevertheless first dig further into Sharpe’s zero-sum case.

The most obvious reason that “informed active managers” can outperform in aggregate is that they trade against “non-informational investors” who are motivated by liquidity needs, institutional constraints, or hedging or are influenced by behavioral biases. This argument is ruled out, however, by Sharpe’s definition of “active managers” as everyone who is simply not passive. In other words, Sharpe’s approach groups together informed investors along with all these non-informational traders. For example, if naive investors buy glamour stocks and informed active managers benefit from value stocks, then these gains do not “count” because Sharpe treats all these investors as one group. Indeed, since the loss by naive investors equals the gain by informed managers, the net profit of the entire group is zero. Similarly, if leverage-constrained investors tilt toward risky stocks and less constrained managers profit from betting against beta, this is counted as a net of zero. If a central bank intervenes in the foreign exchange or bond markets for purposes of managing the macroeconomy and informed asset managers profit as a result, this is also counted as a zero. If pension funds hedge their asset/liability mismatch and fixed-income traders profit from providing liquidity, this is a net zero according to Sharpe’s definition.

The main point here is that active managers may systematically profit from other nonpassive investors with special motives to trade. This part of the debate is well-known and recognized by Sharpe (1991). It depends on semantics and the measurement of the relative importance of different types of nonpassive investors. Therefore, I will not go into this debate, and I leave this example at number “0” because Sharpe would rightly say that it does not count by his method.

Instead, let us focus on the indisputable cases where even passive investors must trade and all nonpassive investors can win in aggregate. The idea is simple: Because passive investors must trade to hold the market, nothing ensures that they trade at the same prices as active investors do and, therefore, nothing ensures that the two groups get the same return.

Example 1. IPOs, SEOs, and Share Repurchases. Suppose that we define the market portfolio as all securities traded on any exchange in the world, and recall that each year, many new securities are listed and some other securities are delisted. For example, companies that go public sell their shares in an initial public offering (IPO) before they are floated on the exchange.
First, suppose that passive investors do not participate in the primary market (i.e., the market for new shares, such as the IPO). Research has shown that IPO securities are, on average, sold at a discount in the IPO relative to the price in the secondary market, when the shares start trading on the exchange. This finding means that informed investors can buy the new shares cheaply in the IPO and then sell some of the shares in the secondary market to other (passive) investors at a premium. In this case, the group of informed, active investors clearly can outperform the group of noninformed, passive investors before fees.

Some may ask, What happens if passive investors participate in the IPO? In this case, they will ask for the same fraction of the shares in any IPO. For example, if half the investors are passive, they would ask for half the shares being issued; let’s keep this assumption for clarity. Active investors will seek out shares in the IPOs that they deem to be priced cheaply on the basis of their security analysis and will avoid those in overpriced IPOs. Hence, an under-priced IPO will be oversubscribed, and therefore, all investors (including the passive ones) will get fewer shares than they asked for. When the price jumps in the secondary market, the passive investors, in seeking to get to their market-cap weighting, must buy the additional shares at a price above the IPO value, again losing to the informed, active investors. For an overpriced IPO, the passive investors end up with all the shares they asked for, and the company (or its bank) retains some of the shares it planned to issue, which will later be sold to the active investors at a discount in the secondary market following a decline in price to fundamental value once listed on the exchange. Again, the active investors obtain the shares at a more favorable price than the passive investors do.

In addition to IPOs, a passive investor must also trade in connection with seasoned equity offerings (SEO) and share repurchases. Here, similar arguments apply, and passive investors face a cost because of adverse selection whereas active management can outperform.

Some skeptics may question whether active investors can systematically profit from IPOs and other changes in the market portfolio. The general point is that passive investors are not guaranteed the same IPO performance as the group of active investors because they trade at different prices and quantities, thereby breaking Sharpe’s equality. Once the equality is broken, is it so hard to believe that those who spend resources collecting information are compensated for their costs in the form of better prices, as theory and evidence suggest?

**Example 1b. What Happens If Everyone Is Passive?** To level the playing field for passive investors, can we simply ban active management from IPOs? Well, even if we could, it would not work. To see why, suppose that only passive investors could participate in IPOs. Then, they would get 100% of the shares, and as discussed previously, they would lose to active investors if the passive investors needed to sell part of the shares in the secondary market.

What if we ensured that everyone was passive in both the primary and the secondary markets? In this case, all shares would be bought in every IPO at whatever the offer price is because no investor would perform security analysis and every investor would simply request his fraction of shares (the same fraction he owns of any security in the market). This indiscriminate buying might initially lead to a fantastic IPO boom at high prices because most anybody could take a company public at any price. Ultimately, if many of these new, “opportunistic” companies went bankrupt, confidence in the financial system would quickly vanish as investors exited the market, leading to a collapse in security prices and a complete halt in new issues, even for good companies. The economy would then come to a grinding halt.

In summary, in order for investors to be willing to buy new securities, these securities need to be sold at fair prices. To set fair prices, some investors must actively collect information about the securities. Hence, when we take into account that capital markets are also about raising capital, we see that informational efficiency, which requires active investors, can have a significant positive impact on the real economy.

**Example 2. Indexes.** Suppose, instead, we define the market portfolio as some index—for example, the S&P 500 Index, the MSCI World Index, or a combination of indexes across global equities and global bonds. In this case, the market portfolio also clearly has turnover because securities are added to and deleted from the index in so-called index reconstitutions (in addition to share issuance and repurchases, as discussed earlier).

When a security is added to an index, index investors simultaneously buy large numbers of shares, pushing up the price. Conversely, when securities are deleted
from an index, index investors sell those securities, resulting in price drops. These price moves translate into costs for index investors and profit opportunities for active managers. For example, when a security is added to an index, active investors can buy the security ahead of the rebalancing date (or they already own it) and then sell it to the passive investors at a higher price when the security is added. Likewise, when a security is deleted from the index, active investors can buy the security cheaply after the deletion and perhaps even try to short-sell the security before the deletion. These effects mean that active investors have a chance to outperform indexes.

Even a passive investor who perfectly tracks the performance of an index incurs an implicit cost—because the cost is built into the index itself. Indeed, the index is defined as buying added securities at high prices and selling deleted ones at low prices.

**Example 3. What “Market”? Passive Investors Own Only a Subset of Assets.**

Another issue is that no one, not even Sharpe, actually knows what the market portfolio is in practice. Indeed, in Sharpe’s Nobel lecture (1990, p. 329), he stated that “no financial futures contract corresponds to the overall market portfolio.” Hence, even those who seek to follow Sharpe’s advice and buy the market portfolio would probably have differing interpretations of what constitutes the market portfolio, and this interpretation would probably change over time. Some people might buy only domestic stocks; others might buy various indexes of global stocks; others could focus on stocks and bonds only in developed markets; and others might include emerging markets, corporate bonds, and so on. These differing interpretations of what defines the market portfolio mean that the resulting portfolios need not add up to the true market portfolio, so the residual creates an opportunity for active managers.

As a simple example, if passive investors buy only stocks included in the S&P 500, then these stocks may become expensive and active investors who also hold the nonincluded stocks may earn higher average returns as a result.

Further, since the market portfolio is the portfolio of all investable assets, it should also include private equity, venture capital, and real estate, among other private assets. However, you cannot be a passive investor in the private markets because you cannot demand to co-invest in every private deal at the same terms as other investors. Therefore, passive and active investors clearly obtain different outcomes when we include all the private assets.

Note that Examples 1–3 are related because the typical life cycle of a successful company is to start as a private company, later be listed on an exchange in an IPO, and later yet be added to an index (and possibly issue other corporate securities or participate in mergers, spinoffs, etc.). Whereas passive investors hold the stock only when it is added to the index (to keep their strategy simple and avoid complications with respect to private companies and IPOs), active investors may benefit from participating in the full life cycle.

**Example 4. Rebalancing and Market Timing.** Passive investors must decide during each time period how much to invest in risk-free securities versus the market portfolio of risky assets. How should they make this choice? An investor’s allocation to the market should be based on her risk tolerance and the perceived risk versus reward (Sharpe ratio) of the market; so, the portfolio allocation is an active choice even for the passive!

Hence, passive investors must rebalance their portfolio allocation over time as their risk preferences change and as the market’s risk–return profile changes—and for other reasons as well. Indeed, they must initially buy their portfolio and will eventually need to sell it, as discussed in Berk and van Binsbergen (2015). Along the way, they need to decide whether to reinvest dividends or sell some fraction of their portfolio to pay for a new house, a car, or other expenses.6

Such rebalancing by passive investors involves buying or selling all securities in proportion to their respective benchmark weights. If there is an equal number of passive buyers and sellers at a given time, these portfolio trades could occur at little or no cost to both parties. However, there will probably be times with more sellers than buyers and vice versa, and therefore, passive investors will buy or sell in aggregate. At such times, someone must take the other side of the aggregate trade by the passive, and this someone has to be from the group of active investors. Prices must necessarily respond, which translates into a transaction cost for passive investors and a corresponding trading gain for active investors.

The rebalancing costs of passive investors can arise at different time scales. At a high frequency, there may be costs associated with trades that create intraday imbalances. At the other end of the spectrum, there
may be years when passive investors herd into risky assets, potentially making (their preferred version of) the market portfolio overvalued. In other years, passive investors may panic and move toward risk-free securities, potentially making the market cheap. Active investors doing the opposite (because the market must clear) may then benefit from buying the market when it is cheap and selling when it is high (a form of market timing). Hence, even if active and passive investors have similar returns on their risky investments in each year, active investors could over time realize higher dollar-weighted returns if their market-timing decisions are better in aggregate.\(^7\)

Active managers also incur transaction costs. Given that active managers trade more than passive ones, active managers may in fact incur greater transaction costs. However, transaction costs consist of both pure commissions/fees, which are a drag on the whole universe of investors, and market impact costs, which are zero sum. In other words, one active investor’s market impact cost may be another active investor’s trading profit. Hence, the aggregate effect of rebalancing on active investors is the sum of (1) minus the loss from commissions, (2) plus zero coming from the zero-sum market impact game among active investors, and (3) plus the potential profits from providing liquidity when passive investors trade.

The next question is whether the arguments just discussed are mostly theoretical or have significant practical relevance. In other words, is Sharpe’s assumption so close to being true that we can simply ignore the “error term”?

To address this issue, let us first consider the turnover of passive investing before we turn to the return implications. As a simple first look, suppose that you owned all stocks at the major US exchanges in 1926—that is, all common stocks in the Center for Research in Security Prices (CRSP) database. Further, suppose that you stayed passive in the inactive sense of the word (Definition 2), not Sharpe’s sense (Definition 1), meaning that you never participated in any IPOs, SEOs, or share repurchases. Figure 1 shows how this portfolio evolved as a fraction of the market portfolio of all listed stocks.\(^8\) As the figure shows, the investor who never trades gradually owns a smaller and smaller fraction of the market because she never buys the new shares, and over time, this effect is quite large. In just 10 years, the inactive investor’s portfolio dropped to about 60% of the market.

On the flip side of this issue, we can also focus on Sharpe’s definition of a passive investor and ask

---

**Figure 1.** An Inactive Investor Is Different from Sharpe’s “Passive” Investor

Notes: This figure considers what happens to an investor who starts off with the market portfolio but never trades after that. The solid green line represents an investor who bought the entire US stock market in 1926 and did not participate in any IPOs, SEOs, or share repurchases and did not reinvest any dividends, showing the resulting fraction of the market that is owned over time. The other lines show the fraction of the market owned by passive investors who start in 1946, 1966, 1986, and 2006, respectively.
how much trading is required to continue to own all the securities in the same market-cap weights. For example, how much trading is required to own 1% of all securities day in and day out, taking into account securities that are added and deleted? Stated differently, how much trading is required to be “passive” in the sense of Sharpe? This question is answered by Figure 2, which shows the turnover for various definitions of the market portfolio.  

Panel A of Figure 2 shows that the average annual turnover for all equities in the CRSP database from 1926 to 2015 is 7.6%. Turnover is calculated as the sum of absolute changes in shares outstanding as a percentage of the total market value in the previous year. Figure 2 shows the average yearly turnover for a “passive” investor who keeps market-cap weights in a given equity or bond investment universe. Panel A shows the turnover for all US-listed stocks in the CRSP database and for US municipal bonds, Treasury bonds, mortgage-related bonds, corporate debt, federal agency securities, and asset-backed securities. Panel B shows the turnover for equity indexes (S&P 500 and Russell 2000) and corporate bond indexes (Bank of America Merrill Lynch investment-grade and high-yield indexes).

Notes: This figure shows the average yearly turnover for a “passive” investor who keeps market-cap weights in a given equity or bond investment universe. Panel A shows the turnover for all US-listed stocks in the CRSP database and for US municipal bonds, Treasury bonds, mortgage-related bonds, corporate debt, federal agency securities, and asset-backed securities. Panel B shows the turnover for equity indexes (S&P 500 and Russell 2000) and corporate bond indexes (Bank of America Merrill Lynch investment-grade and high-yield indexes).
month. Of this turnover, 5.3% consists of new listings (i.e., IPOs), SEOs, delistings, and share repurchases. The “Other” category, at 2.4%, includes mergers that may not require trading.

The figure also shows the 4.2% dividend rate, which leads to trading for investors who reinvest all proceeds. Further, the turnover number used is the more conservative so-called one-sided turnover, meaning that it assumes that the investor buys new shares with cash and invests the proceeds from repurchases in cash. If instead we assumed that the investor would sell shares to finance the purchase of new shares and buy shares with proceeds from sales, then we would get a “two-sided turnover,” which could be as high as double the 5.3% one-sided turnover number reported. But because cash inflows and outflows can sometimes be netted, the true two-sided turnover would be somewhere in between.

Whereas equities have the property of being perpetual, bonds have finite maturity. Hence, new bonds are continually issued, which makes it all the more obvious that trading is required, which can also be seen in Panel A of Figure 2. For example, even if the set of companies does not change, companies regularly issue new corporate bonds and other types of securities. Likewise, government bonds are continually issued, as are mortgage bonds, municipal bonds, and so on. Holding the market portfolio of all these fixed-income securities requires a turnover of 20% to buy new issues (which basically means that bonds have an average maturity of around five years, such that one-fifth of bonds mature each year).

Panel B of Figure 2 shows the turnover of various major equity and bond indexes. The turnover of the indexes is large owing to various types of additions to and deletions from the indexes and because tracking an index requires a two-sided turnover, as discussed previously. The Russell 2000 Index has a much larger turnover than the S&P 500 because stocks are both added when they are large enough and deleted when they become too large (and are moved to the Russell large-cap index). The bond indexes have even greater turnover because of the large issuance shown in Panel A and because of rating changes, maturity dropping below one year, defaults, and other events.

Some index funds try to limit their turnover costs in various ways; for example, bond funds often hold only a subset of the index. Nevertheless, when looking at the reported turnover rates of actual index funds, we do see meaningful trading. For example, S&P 500 index funds often report turnover rates above 3%. The reported turnover for a US mutual fund is the lesser of its purchases or its sales, excluding the trading of derivatives and in-kind transactions. Hence, the actual trading is at least twice that number, and it could be significantly more—sometimes even larger than 10% for some of the best-run funds. Bond index funds routinely report very high turnover, sometimes with percentages in the hundreds.

Lastly, let us not forget that the true turnover of passive investors is likely to be higher than reported in Figure 2, for several reasons. First, investors must buy and sell when they decide to save or consume or change their risk aversion. Second, investors tend to change their definition of the market. Emerging markets are added and deleted, frontier markets open up and are closed down (e.g., owing to the relaxing or tightening of capital controls), and countries and asset classes are added and deleted as they are deemed appropriate or inappropriate investments for various reasons (e.g., the risk of a war or systemic risk in the financial system). And active investors may benefit from all these changes. Third, passive investors often trade to hedge foreign exchange exposures, and they trade to roll over futures contracts and other derivatives.

How Large Is the Hidden Cost of Passive Investing?

We have seen that passive investing entails a non-trivial amount of trading. We next consider the cost associated with the required trading by passive investors or, equivalently, the expected outperformance of active management resulting from that trading.

Starting with equity issuance and repurchases as discussed in Example 1, note that IPOs have been underpriced by 10%–20% on average in the United States over long time periods as well as in a number of other countries (see, e.g., Ljungqvist 2008). Given the new listings of 1.2% per year reported in Figure 2, these numbers translate into a performance difference for a passive versus an informed participant in the new listings of about 1.2% × 15%—that is, about 18 bps (although not all these new listings in the CRSP database are IPOs). Similarly, if SEOs are underpriced by about 2% on average, then they give rise to a performance effect of about 6 bps. Share purchases are harder to evaluate, but they are a smaller group and passive investors may more easily
avoid adverse selection. As shown in Figure 2, the trading of bonds required by passive investors is substantial, which adds to their costs. IPO underpricing also occurs with corporate bonds, with an average magnitude of 0.47% for high yield and 0.02% for investment grade (Cai, Helwege, and Warga 2007).

Most passive investors actually follow specific indexes, as discussed in Example 2. Stock price moves around index reconstitutions imply costs for passive investors, and Petajisto (2011, p. 271) reported that for “additions to the S&P 500 and Russell 2000, we find that the price impact from announcement to effective day has averaged +8.8% and +4.7%, respectively, and −15.1% and −4.6% for deletions.” On the basis of these price moves and the turnover reported in Figure 2, Petajisto (2011, p. 271) estimated the lower bound of “the index turnover cost” to be “21–28 bp annually for the S&P 500 and 38–77 bp annually for the Russell 2000.” Note that this cost is embedded in the index itself (so it is separate from whether a given passive index fund delivers the same return as the index). Stated differently, if you pay a manager 10 bps to track the S&P 500, you could be paying an explicit 10 bps on top of an implicit cost of more than 25 bps. Further, index funds often try to recover their costs via securities lending, but this revenue may also be earned by active managers.

The cost of passive investing in fixed-income securities is also nontrivial. Index inclusion or exclusion moves prices from 0.20% to 5%, depending on the type of event (Dick-Nielsen 2012). And as shown in Panel B of Figure 2, the turnover is large, so this can lead to a meaningful effect on returns.

In addition to these costs, passive investors incur rebalancing costs when they seek to simultaneously buy or sell, as discussed in Example 4, as do active investors. An exhaustive analysis of the aggregate costs of passive investing is beyond the scope of this article, and such an analysis should also consider the magnitude of the gains to active investing relative to the size of the portfolio (see the model in Appendix A). The preliminary findings suggest that active management can, in aggregate, add a modest but nontrivial amount of value relative to passive investing before fees. Further, even if the cost of passive investing is modest in today’s market, this cost would grow if most investors became passive. Nevertheless, the current added value for active managers may be below their typical asset management fees, which would lead the average active manager to underperform after fees and expenses, as I will discuss next.

Evidence on the Performance of Active Managers

Now I turn to the empirical evidence on performance. Depending on the sample and benchmark, active mutual funds have on average performed similarly to the market before fees and underperformed after fees (Fama and French 2010) or have outperformed before fees and performed similarly to the market after fees (Berk and van Binsbergen 2015). Hence, active mutual fund fees have been so large on average that they have consumed all or more than the managers’ alpha, so the average retail investor would have been well served by following Sharpe’s advice to use low-cost, passive investing. At the same time, according to Gerakos, Linnainmaa, and Morse (2016, p. 29), institutional asset managers “outperform strategy benchmarks by 86 basis points gross, or 42 basis points net of fees,” so institutions appear to have benefited from active investing in aggregate.

Hence, the average manager across both segments may outperform slightly before fees, but finding a manager who can be expected to outperform after fees is difficult. This evidence suggests that Sharpe’s equality could be violated but, nevertheless, also supports his idea that low-cost index funds are extremely useful, especially for individual investors.

More broadly, the evidence is consistent with the model of Gârleanu and Pedersen (forthcoming 2017)—namely, that investors looking for good asset managers face a problem similar to asset managers looking for good securities. In both cases, you can hope to outperform only if you spend resources locating a good manager or good security. Naturally, large institutional investors can better afford to spend resources on a manager selection team. So, it is not a surprise that institutional investors have been more successful in their active management than smaller investors.

It is time to sharpen the debate on active versus passive management. Some commentators have interpreted Sharpe’s equality as a law that holds without assumptions and have called for everyone to be passive, whereas active investors have an incentive to “talk up” their profits; some have even claimed that “passive investing is worse than Marxism” (Sanford C. Bernstein & Co., 2016). Instead, we should find a common ground by first recognizing that both active and passive management can have value and then attempting to empirically evaluate their costs and benefits.
Do We Need All the Trading?

We have seen that active management plays a crucial role in companies’ ability to efficiently finance new projects by issuing new securities. However, some people might still argue that most of the active trading is useless. The argument goes something like this: “Sure, trading can be good when it helps companies raise capital to make real investments—for example, to build a factory that produces medicine—but this is only about x% (see Figure 2). So, the majority of active trading is useless.” This argument might be correct if we assume that, despite little trading in the secondary markets, (1) companies could still issue new securities at the same prices and (2) the liquidity needs of passive investors were still met. But these assumptions run counter to reason and a lot of research.

Active management makes secondary market prices more informative, and investors value the ability to trade in liquid markets (Acharya and Pedersen 2005). As a result, a company’s cost of issuing new securities is lower when secondary markets are more liquid (Butler, Grullon, and Weston 2005; Ellul and Pagano 2006).

In other words, if we eliminate all the “extraneous” trading of active management, then companies’ cost of issuing securities increases. For instance, the pharmaceutical company may no longer build the new factory and produce the medicine if the cost of raising capital increases. To see why at a more practical level, suppose that you are buying a house. Would you be more comfortable bidding a large amount of money if you knew that you could easily sell the house later? Surely, the answer is yes for most people, so liquidity matters. Would you be more comfortable bidding a large amount of money if you knew the prices of a lot of comparable houses and knew that these prices had been set efficiently? Again, surely, the answer is yes, so efficient and transparent secondary markets matter.

If most investors were passive, the liquidity in individual securities would vanish because investors would trade only overall indexes. The lack of liquidity and active management would make prices much less informative. When the secondary market is illiquid and uninformative, buying in the primary market becomes much riskier, which, in turn, increases companies’ cost of finance.

So, yes, we need more trading than the trading in the primary markets, even if we may not need as much turnover and as much active management as we have today, as more and more investors realize the benefits of low-cost, passive investing.13

Conclusion: The Future of Asset Management

Sharpe’s arithmetic has profound implications for the future of active management: Because active managers lose after fees, investors should move money out of active and into passive. Where does this process end? One would think that less active management means less efficient markets, which makes profiting from active management easier, restoring equilibrium (as in “normal” supply/demand economics). Not so, according to Sharpe’s arithmetic. Sharpe’s arithmetic means that active managers lose regardless of how inefficient the market is. So, eventually, all capital should leave the doomed active management industry, spelling doom for capital markets, according to Sharpe’s arithmetic.

My arithmetic paints a less gloomy picture. As I have shown, active managers can profit in aggregate, and more so if little capital is allocated to active. Therefore, the future of active management is not doom. At the same time, the historical record of average active manager returns after fees is not impressive, so more and more investors will surely recognize the benefits of passive investing. Hence, I expect that the fraction of passive investors will continue to grow but will end up at a number less than 100%.

More broadly, capital markets are not a zero-sum game; they are a positive-sum game. Companies benefit from access to capital markets, passive investors benefit from low-cost access to investing in diversified markets, and active managers benefit from their information collection efforts through potentially even higher investment returns before costs. Active management is socially valuable when it helps finance promising new companies, which increases the collective wealth.

Said differently, much of financial economics considers what is called a “pure exchange economy” (Lucas 1978), which basically means that the set of companies and what they produce is taken as given, so capital markets are used only for risk sharing—that is, a big insurance scheme—which works best if no one is informed. In the real world, however, capital markets are also used for raising capital, and this
function, which requires active investors, is crucial for economic growth.

In summary, active and passive investors can coexist in an efficiently inefficient market equilibrium where passive investors benefit from low costs and active investors benefit from their efforts to make markets close to efficient.\textsuperscript{14}

Appendix A. Stylized Model: Passive Matches the "Market" and Active Beats It

In this appendix, I present the simplest possible model in which the market portfolio changes over time, as it does in the real world, showing how active managers can beat the market in aggregate. I first present the performance results (assuming price moves around inclusions/deletions) and then show how these results arise naturally in equilibrium.

Consider two kinds of risky securities: those included in passive investors’ definition of the "market portfolio" (denoted by \( i \), for “included”) and those that are not included (denoted by \( n \)). For example, the included securities can be (1) the members of an index, such as the S&P 500, or (2) stocks traded on a major exchange. Nonincluded stocks in Interpretation 1 are those outside the S&P 500, and in Interpretation 2, they are companies that are yet to be listed on the major exchange.

Securities switch between being included in the index and not being included. Specifically, a fraction \( s^u \) of nonincluded securities “switch up” to being included at the end of each time period. For example, we can imagine that these securities are added to a stock index (or in the alternative interpretation, they are listed on the exchange). Likewise, a fraction of the included securities are deleted from the market at the end of each time period, denoted by \( s^d \), for “switching down.”

Included securities have a price premium of \( x \) relative to securities that are not included (as shown in the equilibrium model here and empirically by, e.g., Petajisto 2011). That is, when a security is included, its price will be \((1 + x)\) times what it would have been outside the index. For simplicity, assume that all securities have similar risk characteristics such that the only difference between them is whether they are included or not. The return of included securities between time \( t \) and \( t + 1 \) is then

\[
r^i_{t+1} = \frac{D^i_{t+1} + (1 - s^d)p^i_{t+1} + s^dp^i_{t+1}}{p^i_t} / (1 + x) - 1.
\]

Here, the first term is the dividend, the second term is the value of the shares that remain included in the market, and the third term is the value of shares deleted from the market. We can write the expected return in terms of the dividend yield, \( \delta_t = \frac{E_t(D^i_{t+1})}{p^i_t} \), and the expected price appreciation, \( \mu_t = \frac{E_t(p^i_{t+1})}{p^i_t} \):

\[
E_t(r^i_{t+1}) = \delta_t + (\mu_t - 1) - s^d \mu_t \frac{x}{1 + x}.
\]

In other words, the expected return on included securities is the return in the absence of deletions (the dividend yield plus the price appreciation) minus the negative impact of deletions (which depends on the fraction of deletions, \( s^d \), and the price drop, \( x \)).

Similarly, the return on nonincluded securities is

\[
r^n_{t+1} = \frac{D^n_{t+1} + (1 - s^u)p^n_{t+1} + s^up^n_{t+1}(1 + x)}{p^n_t} - 1,
\]

so the expected return is

\[
E_t(r^n_{t+1}) = (1 + x)\delta_t + (\mu_t - 1) + s^u x \mu_t.
\]

Here, the first term means that nonincluded securities earn a higher dividend yield because they have a lower price for a given level of dividends. The second term states that nonincluded securities earn the same relative price appreciation. The third term says that nonincluded securities enjoy a benefit from the price increases associated with inclusions in the market. Hence, this simple model can capture the idea that IPOs increase in value upon listing and stock prices increase when stocks are added to a major index (even though we do not consider information asymmetries and details of the IPO process).

The return difference between nonincluded and included securities is, therefore,

\[
E_t(r^n_{t+1} - r^i_{t+1}) = x(\delta_t + \frac{s^d \mu_t}{1 + x} + s^u \mu_t).
\]
which can be written more simply by using $\mu_{t+1} \approx 1$ (which holds exactly in the steady-state equilibrium derived below):

$$E_t (r^n_{t+1} - r_i^f_{t+1}) \approx x(\delta_t + s^d + s^u).$$

Note that the nonincluded securities earn higher returns because of increased dividend yields, inclusion effects, and deletion effects. The inclusion/deletion (or issuance/repurchase) effects correspond to Examples 1 and 2 earlier in the article, whereas the dividend-yield effect corresponds to Example 3 (i.e., it arises because passive investors do not hold the full market).

For example, if the price premium of included securities is 10%, the dividend yield is 3%, 1% of included shares are deleted from the index, and 2% of nonincluded shares get included, then the approximate return difference is $10\% \times (3\% + 1\% + 2\%) = 0.60\%$.

More broadly, the return difference is greater when the market portfolio has more turnover (larger $s^d + s'^d$) and when the price premium, $x$, is greater.

We can also consider an investor $a$ who holds both included and nonincluded securities, with the fraction of nonincluded securities represented by $f$. The return difference between this “active” investor who holds both kinds of securities and a “passive” investor who holds only the included securities is

$$E_t (r^n_{t+1} - r_i^f_{t+1}) = fE_t (r^n_{t+1} - r_i^f_{t+1}) \approx fx(\delta_t + s^d + s^u).$$

Suppose as before that the price premium is $x = 10\%$, the dividend yield is $\delta_{t+1} = 3\%$, and $s^d = 1\%$ and, further, that the active investor invests only $f = 2\%$ of her capital in nonincluded securities but does so very selectively by trying to buy only those that are about to be included. If all the nonincluded securities that the active investor buys end up being included, then $s^u = 100\%$, and in that case, the return difference is $2\% \times 10\% \times (3\% + 1\% + 100\%) = 0.21\%$.

In principle, an active investor could also add value by avoiding holding the included securities that are about to get deleted, which would further raise his return by approximately $(1-f)xs^d$ (i.e., the numerical example, which is $98\% \times 10\% \times 1\% = 0.10\%$).

Note, however, that such trades that occur before the deletion are not consistent with the equilibrium derived below, in which passive investors trade only upon deletion (i.e., such trades would be zero sum among the active investors).

**Equilibrium**

To show how the previous insights can arise in equilibrium, consider an economy in which a fraction $I$ of all securities are included in the “market.” Passive investors buy a fraction $\theta$ of all shares included in the “market,” and the supply of shares of each security is normalized to 1. Active investors choose their portfolio $\pi$ (which is a vector of positions in included and nonincluded securities) in order to maximize their mean–variance preferences:

$$\max_{\pi} \pi \gamma = (1 + r^f)P_t - \gamma \pi^2,$$

where $r^f$ is the risk-free rate and $\gamma$ can be thought of as the variance of each security’s return multiplied by investors’ risk aversion parameter. The solution is

$$\pi = \frac{1}{\gamma} [E_t (D_{t+1}^i + P_{t+1}^i) - (1 + r^f)P_t].$$

In equilibrium, active investors must choose a position of $\pi^i = 1 - \theta$ for included securities and $\pi^n = 1$ for nonincluded securities. Hence, in a steady-state equilibrium, we have two equations for the two unknowns—namely, the prices of included securities ($P^i$) and nonincluded securities ($P^n$):

$$(1 + r^f)P^i = D + P^n - s^d (P^i - P^n) - \gamma (1 - \theta);$$

$$(1 + r^f)P^n = D + P^i + s^u (P^i - P^n) - \gamma,$$

where the expected dividend, $D = E(D_{t+1}^i)$, is the same for all securities. After combining these equations, the equilibrium price difference, $\Delta P = P^i - P^n$, is given by

$$\Delta P = \frac{\gamma \theta}{r^f + s^d + s'^u},$$

and the equilibrium prices are

$$P^i = \frac{D - \gamma (1 - \theta) - s^d \Delta P}{r^f}.$$
Sharpening the Arithmetic of Active Management

and

\[ p^n = \frac{D - \gamma + s^u \Delta P}{r^f}. \]

Intuitively, the price difference, \( \Delta P \), increases in the amount bought by passive investors, \( \theta \).

From these equilibrium prices, we can compute all the numbers discussed earlier. For example, the price premium is \( x = \Delta P / p^n \). Further, since active investors hold all the nonincluded stocks and \( 1 - \theta \) of the included stocks, the fraction of nonincluded stocks in their portfolio is \( \frac{1 - l}{1 - l + l(1 - \theta)} \), and in a value-weighted sense, it is

\[ f = \frac{(1 - l)p^n}{(1 - l)p^n + l(1 - \theta)p^f} = \frac{1 - l}{1 - l + l(1 - \theta)(1 + x)}. \]

Lastly, note that in a steady state, where the same number of companies move up and down in each period, we have \( ls^d = (1 - l)s^u \); that is, \( s^u = s^d l/(1 - l) \).

We are ready to consider a numerical example for illustration. We normalize the expected dividend to be \( D = 1 \), the risk-free rate is \( r^f = 2\% \), half the securities are included (\( l = 50\% \)), the fraction of deletions is \( s^d = 2\% \), the fraction of nonincluded stocks that are added is \( s^u = 2\% \), passive investors buy \( \theta = 40\% \) of the included shares,\(^{17}\) and the active investors have a risk aversion corresponding to \( \gamma = 0.5 \) (chosen to have a reasonable equilibrium dividend yield of around 3\%).

Using these inputs, the price of included and nonincluded securities can be computed as \( P^i = 31.7 \) and \( P^n = 28.3 \), respectively. Hence, the dividend yield of included securities is \( \delta = 3.2\% \), and the price premium is \( x = 12\% \). The expected return difference for nonincluded versus included stocks is

\[ E_t(v^n_{t+1} - r^i_{t+1}) = 0.82\%. \]

Given that the active investors hold \( f = 60\% \) of assets in nonincluded securities, the excess return of active relative to passive is \( E_t(v^a_{t+1} - r^i_{t+1}) = 0.49\% \) (before fees).

We can consider the effect of changing various parameters. First, if the fraction of included securities is higher—for example, \( l = 80\% \)—then the return difference between included and nonincluded securities remains the same.\(^{18}\) Nevertheless, the return difference between active and passive is smaller—\( E_t(v^a_{t+1} - r^i_{t+1}) = 0.22\% \)—because active investors now hold a smaller fraction, \( f \), of their portfolio in nonincluded securities.

Second, Panel A of Figure A1 shows how the return difference increases with the size of the passive investors. We can see that if more investors are passive, the price difference between included and nonincluded securities is larger, and active investors can add more value from changes in the market portfolio. Panel B shows how the return difference decreases with the size of active investors. Intuitively, having more active investors means that the price differences between included and nonincluded securities diminish, leading to lower return differences.

One could try to determine the equilibrium size of active versus passive management on the basis of their relative costs and benefits. That is, rather than separately varying the sizes of active and passive as in Figure A1, we could try to determine these endogenously on the basis of an entry decision. See Gârleanu and Pedersen (forthcoming 2017) for such an equilibrium level of active management in the context of a more sophisticated model with private information, search, and equilibrium asset management fees; see Petajisto (2009) for theory and evidence on how active management leads to downward-sloping demand curves for stocks.

Finally, note that the numbers I have used should be viewed only as illustrative examples. Future research should analyze realistic calibrations, considering different extensions that take into account the complexities of real markets. First, I have assumed here that the price impacts at inclusion and deletion events are permanent price changes. One could make a similar calculation if the price impact is (partly) temporary.\(^{19}\) Second, although I have assumed that all passive investors use the same index, the model can be extended to capture many different indexes. Third, I have ignored seasoned equity offerings, repurchases, rebalancing trades, new companies, bankruptcies, private information, noise trading, and many other effects where active investors could, in principle, add value (these complexities are part of the reason that passive investors choose to track a simple index and why such indexes must change over time). Fourth, I have assumed a simple structure for risk (constant prices, independently and identically distributed dividend risk, and diversifiable inclusion/deletion risk), but in the real world, discount rates are time varying, adding significant risk because of such things as changing price levels. Fifth, Petajisto (2009) argued that because some active investors specialize in certain trades, the risk aversion implicit in the aggregate equity premium (e.g., the dividend yield)
Figure A1. How Much Can Active Beat the Market?

A. Active minus Passive Return vs. the Size of Passive Investing

B. Active minus Passive Return vs. the Size of Active Investing

Notes: This figure plots the expected return to the active investors in excess of the expected return of the securities included in the "market" portfolio, \( E_t(r^A_{t+1} - r^P_{t+1}) \), based on the numerical example discussed in the text. Given the base-case parameters in the text, Panel A varies the fraction bought by passive investors, \( \theta \), and Panel B varies the size of the active investors by varying their risk tolerance, \( 1/\gamma \) (a higher number of active investors means a smaller aggregate absolute risk aversion).

is lower than the risk aversion implicit in the slope of demand curves (e.g., both active and passive investors hold equities, but only active investors provide liquidity when a security leaves an index).

Editor's Note
Submitted 10 January 2017
Accepted 4 July 2017 by Stephen J. Brown

Notes
2. Footnote 4 in Sharpe (1991, p. 8) states, “We assume here that passive managers purchase their securities before the beginning of the period in question and do not sell them until after the period ends. When passive managers do buy or sell, they may have to trade with active managers; at such times, the active managers may gain from the passive managers, because of the active managers’ willingness to provide desired liquidity (at a price).”
3. Event-driven hedge funds specialize in trading around such corporate events as mergers, new issues, seasoned equity offerings, spinoffs, and so on.
4. For an exploration of leverage constraints and betting-against-beta strategies, see Frazzini and Pedersen (2014).
5. Ljungqvist (2008) documented an average IPO underpricing of 10%–20%. The IPO underpricing is needed to compensate passive investors for adverse selection
and give active investors an incentive to take part in the underwriting process and spend resources in determining the value of the securities (Rock 1986).

6. Furthermore, the passive trades in connection with IPOs and index reconstitutions discussed previously are more complex than many investors realize. When a passive investor buys shares in an IPO, where does the money for these shares come from? If he does not want to use cash, he needs to sell parts of all his other security holdings.

7. If passive investors hold only publicly traded securities and these securities are mispriced, passive investors may also be exploited through companies’ decisions to list on exchanges or delist. That is, the effects discussed in Example 1 may be exacerbated by endogenous decisions to add and subtract securities from the public market at opportune times.

8. Figure 1 reports the total market value of the buy-and-hold strategy as a fraction of the total market value of all shares, adjusted for stock splits by assuming that the passive investor is treated like other investors in any stock split. Some readers may be surprised to learn that doing nothing is not enough to be passive, whereas others may be surprised that the investor from 1926 continues to hold almost 10% of the market today (this is due to old giants like Standard Oil, GE, Chevron, and Coca-Cola).


10. These turnover rates can be computed by looking at the separate numbers for purchases and sales in the footnotes to the annual reports. Turnover could be higher still because certain types of trades are not included in these statistics—for example, derivatives trading.

11. See also Chen, Noronha, and Singal (2006, p. 45), who found that the “loss to an investor in the Russell 2000 may be about 120 bps a year and can be as high as 184 bps a year, and S&P 500 investors may lose as much as 12 bps a year. Consistent with this finding, we found that the Russell 2000 underperformed other small-cap indices by more than 3 percentage points a year in the 1995–2002 period, even though comparable indices did not entail greater risk.”

12. This debate has been intense ever since passive investing was introduced (see, e.g., Langbein and Posner 1977).

13. Pástor and Stambaugh (2012) studied the size of the active management industry.

14. Small investors are likely to perform better with low-cost, passive investing, whereas large investors are more likely to benefit from being active. The marginal investor should be indifferent between passive investing and spending resources to find an active manager who is worth more than his fees. See the formal model of asset management by Gârleanu and Pedersen (forthcoming 2017) and practical examples in Pedersen (2015).

15. Example 4 from earlier in the article corresponds to a time-varying premium \( x \) (and a corresponding time-varying \( \theta \) in the equilibrium model), leading to additional dynamic effects not considered here.

16. I assume that securities have independently and identically distributed dividend risk, I consider a steady-state equilibrium with constant prices, and I assume that risk associated with inclusions and deletions is fully diversifiable. Hence, the variance of a diversified portfolio is \( \text{var}_t [\pi_t (1 + r_t)] = \sigma^2 \pi_t \); that is, the variance from dividend risk \( \sigma^2 \) times the squared portfolio weights. In other words, in each period, a given fraction \( s^2 \) of securities are deleted from the market so that there is no aggregate deletion risk—and similarly for additions.

17. For an estimate of the size of passive investors as a group, see, for example, Morningstar (2017, Exhibit 10).

18. The choice of \( l \) depends on the setting. For US exchange-traded equities, S&P 500 stocks constitute about 80% of the market value of listed US stocks. But historically, that number is lower, and it is even lower if you assume that active investors can also hold some foreign equities, convertibles, private equities, and other assets.

19. If the price impact of inclusion/deletion is temporary, included securities continue to suffer from a low return because of the price drop associated with deletions and nonincluded securities continue to benefit from a price jump upon inclusion. But several other effects arise. Newly included securities have a low return because their recent price increase reverses, recently deleted securities have a high return because their price drop reverses, and the dividend-yield effect discussed previously disappears if price levels are equal for the two groups.

References


