ON THE BIRTH OF THE ETF INDUSTRY

Investing and trading are two different concepts. Investors buy and hold for the long term, where the long term can be as short as a few months or as long as 30 years or more. Investors include individuals planning or saving for retirement or to buy a new home. Investors include university endowments whose (a) corpus is preserved in perpetuity and (b) annual payoff from investment is used for general and directed purposes. Investors include state governments whose tax collection coffers are used for various infrastructure projects. All the investors have in common is the need to allocate their wealth across asset classes by maximizing the expected return on their investment funds, subject to the constraints imposed by their planning horizons and risk tolerances. Traditionally, the allocation problem has involved three asset classes: stocks, bonds, and cash equivalents. Long-term investors buy and hold. No asset classes are shorted. Trading is passive and occurs only when there are incoming flows or outgoing disbursements. After making the initial portfolio allocation decision, the allocations are reconsidered periodically to determine whether adjustments are necessary. For example, suppose an investor allocated 60% to stocks and 40% to bonds two years ago. Today, he looks to find that a significant upturn in the stock market has moved his allocation to 80% stocks and 20% bonds. The change in allocation pushed him beyond his risk tolerance, so he may choose to re-optimize his portfolio allocations to reduce his stock market risk exposure. Investors do not constantly monitor the markets and trade infrequently.

Traders, on the other hand, are speculators. They place short-term directional bets based on the belief that they have some informational or modeling advantage (or just like to gamble). George Soros falls in the information category. His profits from macro speculation are legendary. While speculation is his investment strategy, he is not an investor in the traditional sense. He is a trader and, like other traders, has holding periods as short as a few hours to, perhaps, a few days.

Academics have focused on investment decision-making for many decades. It started with the simple behavioral assumption that (a) individuals grow happier with each dollar of wealth they accumulate, but (b) the last dollar received was more satisfying than the current one. If the individual is only concerned about wealth and consumption, his objective will be to maximize his expected satisfaction level.

The return/risk framework

Tobin (1958) cleverly transformed (mathematically) this objective function to show that expected satisfaction is maximized when an individual maximizes expected investment portfolio return for a given risk level. Risk is measured by the standard deviation of portfolio return or, in industry parlance, "volatility." Markowitz (1952) had already programmed computer routines to solve the portfolio allocation problem. At the time, computer technology was in its infancy and had to be performed on a mainframe computer. Today, the portfolio allocation problem is easily solved using Excel, as I will demonstrate shortly.

The aggregation of individuals' investment demands

Sharpe (1964) and Lintner (1965), who worked independently, were the next significant leap forward in understanding portfolio decision-making. They concluded that each investor's portfolio should include a passive, market value-weighted portfolio (i.e., the "market portfolio") of all risky assets together with cash equivalents. Low-risk tolerance individuals will invest some money in the market portfolio and some in cash equivalents. Since they are long cash equivalents, they are lending. High-risk tolerance individuals will invest more than all their investment wealth in the market portfolio by short-selling cash equivalents or borrowing.

The market portfolio in this framework must contain all risky assets. Practically speaking, this is not possible. At the same time, stocks and bonds have deep and liquid markets; markets for physical assets such as commodities do not. Confining the asset classes to stocks and bonds, consider each allocation in turn. What stocks should I hold? The answer is all of them. The stock market will produce a gross return over the next year. Some individuals might attempt to pick "winners" by selecting a subset of the stocks from the market. If they do win, others, also trying to "beat the market," must lose an equal amount. In essence, Sharpe said, "Why bother with the game?" Just hold a passive, market value-weighted stock market index portfolio. For bonds, do the same. Then, decide upon your allocation between the stock market, the bond market, and cash equivalents based on maximizing return, given your risk tolerance.

An application of the framework

The issue is, of course, how can I trade the entire stock market portfolio and the entire bond market portfolio? I will address that shortly. Using the abovementioned theory, I will first illustrate how an investor (e.g., a self-directed individual (retail) investor or a university endowment manager) should allocate investment funds. I assume the investor has $500,000 to invest and can allocate his funds to cash equivalents, bonds, and stocks. The expected return and volatility of each asset class and the expected correlation

between the returns of each asset class are hypothetical. The expected returns and volatilities increase monotonically from cash to bonds to stocks. The correlation between the return of the different asset classes is assumed to be 0.

Before solving the portfolio allocation problem, I must provide intuition regarding risk tolerance. Risk tolerance is just that—the maximum volatility that an investor is willing to tolerate. It depends on many factors. The investor's investment horizon is an important one. Suppose an individual is about to retire and live on his retirement funds. In that case, he is likely to have a relatively low-risk tolerance, say 2% to 5%. Suppose an individual is beginning her career and is just beginning to set aside investment funds for retirement 30 years down the road. In that case, she can tolerate a higher level of risk, say 10% to 15%. Risk tolerance levels generally range from 2% (extremely conservative) to 20% (extremely aggressive).

	Expected return/risk parameters		
Asset class	Cash	Bonds	Stocks
Expected return	2.00%	5.00%	10.00%
Volatility	0.50%	7.00%	22.00%
	Expected correlations		
Asset class	Cash	Bonds	Stocks
Cash	1	0.000	0.000
Bonds	0.000	1	0.000
Stocks	0.000	0.000	1

Countless software packages and services can solve the problem of maximizing expected returns across asset classes for a given level of risk tolerance. I solved this example using Excel multiple times to illustrate the effects of risk tolerance on decision-making. The results are in the table below.

Each row in the table contains the optimal allocations across asset classes for a given risk tolerance, and the risk tolerance levels range from 1% (extremely conservative) to 22% (extremely aggressive). With a highly conservative 1% risk tolerance, the allocations are 87.5% cash, 9.9% bonds, and 2.6% in stocks. As I move from extreme conservatism to moderate conservatism at 5%, the allocations change to 30.7% cash, 54.6% bonds, and 14.6% stocks. At 15%, the allocations are 32.6% in bonds and 67.4% in stocks. The highest level of risk tolerance possible without borrowing is 22%, where all the investment funds are in stocks.

Expected portfolio return for given risk tolerances				
Expected return	Risk tolerance	Asset class allocations		
		Cash	Bonds	Stocks
2.51%	1.00%	0.875	0.099	0.026
3.39%	2.50%	0.657	0.271	0.073
4.81%	5.00%	0.307	0.546	0.146
7.07%	10.00%	0.000	0.585	0.415
8.37%	15.00%	0.000	0.326	0.674
9.54%	20.00%	0.000	0.091	0.909
10.00%	22.00%	0.000	0.000	1.000

The epiphany

With the mechanics of long-term decision-making, the focus turned to discussing what asset classes are available for investment and how they trade. In the 1960s and early 1970s, individuals and institutions mainly delegated long-term investment management. They bought actively managed mutual funds. Portfolio managers tried identifying and buying under-priced securities to earn above-market returns. For this service, investors paid various fund fees (e.g., load and management fees) and the trading costs from active stock picking (e.g., commissions, bid/ask spreads, and price impact). At the time, it was common for investors to pay the fund manager 300 basis points or more per year. For every dollar $1,000 an individual might have in his actively managed retirement fund, he would give up $30 annually.

Relying upon the wisdom of Sharpe (1964) and Lintner (1965), Samuelson (1974) recognized the folly of active portfolio management. The stock market produces only one gross return – a market value-weighted average return that includes all stocks in the marketplace. Suppose investors delegate their stock investment decisions to active portfolio managers. In that case, common sense dictates that the gross return across managers must be equal to the gross stock market return (i.e., the return of winning managers must equal the return of losing managers). The system has leakage, however. Investors earn the gross market return only after trading costs and management fees. Samuelson's advice? Get rid of the active managers (which he colloquially calls "gunslingers"). Create passive index mutual funds by buying all the stocks in the market. Again, the gross market return is the same. The difference is that the investor avoids most of the costs and fees. Amusingly, Samuelson (1974, p. 18) goes on to suggest "… that most portfolio decision-makers (i.e., active portfolio managers) should go out of business – taking up plumbing, teach Greek, or help produce the annual GNP by serving as corporate executives."

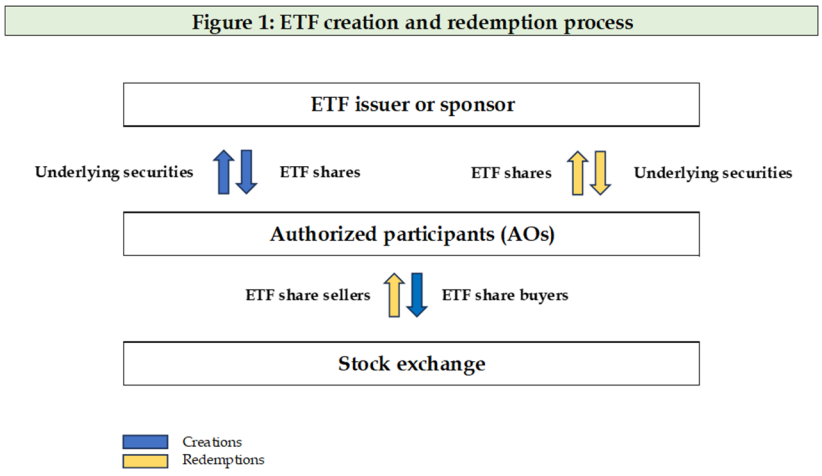
John Bogle picked up Samuelson's idea and formed The Vanguard Group in 1974, creating the first index mutual fund.[1] The fund was benchmarked to the S&P 500 index portfolio, passively managed, market value-weighted, and transparent. Trading costs were minimal, and management fees were only about 30 basis points, $3 annually for every dollar $1,000 invested, a whopping reduction of 90%. The Vanguard 500 Index Fund launched on August 31, 1976, with a mere $11.3 million from investors. Today, the fund's value is $544 billion.[2] But this is a single fund. To appreciate the true impact of indexing, consider that no mutual funds were indexed in 1976. Indexing accounts for over half the $30 trillion-plus mutual fund industry today.[3]

After decades of accelerating growth in the index mutual fund industry, the idea of trading indexes like stocks came. Index mutual funds receive orders to buy and sell shares of the fund throughout the day; however, the trades take place at end-of-day prices. The reason is simple. There must be a mechanism for ensuring the market price of the fund's shares is in line with the net asset value (NAV) of the stocks in the fund when the accumulated investment demand from the day meets supply. This arbitrage mechanism ensures that the market price equals net asset value.

Exchange-traded funds (ETFs) and the creation/redemption process

Demand to trade index funds continuously throughout the day arose in the early 1990s. The American Exchange (AMEX) responded by developing a *creation/redemption* (CR) mechanism involving the *ETF sponsor (*i.e., *issuer* or *provider*) and *authorized participants* (APs). During the day, APs create ETF shares in large increments (say, 50,000) called *creation units*. Each creation unit contains all the securities in the index portfolio in their appropriate weights.[4] The AP delivers those securities to the ETF sponsor. In return, the ETF sponsor "wraps" the bundles of securities into ETF shares and then gives the ETF shares to the AP. The new ETF shares, created in the primary market, are then traded in the secondary market (i.e., on a securities exchange). When demand increases, more ETF shares are created. The creation process is illustrated by the blue arrows in Figure 1.

[1] Bogle (2016) recounts the colorful history of the birth of the passive index mutual industry.
[2] See VOO Fact Sheet (September 30, 2023).
[3] See ICI (2022).
[4] In the U.S., the exact composition of the ETF is published by the issuer on its website each morning.



APs can also redeem ETF shares by reversing the process. Bundles of ETF shares—known as redemption units (again, typically 50,000 shares)—are collected in the secondary market and delivered to the ETF sponsor in exchange for the underlying securities. Redemption is the opposite of creation, denoted by the gold arrow in Figure 1. When demand decreases, the ETF shares are "unwrapped" and again become single securities.

An AP may be a market maker, a specialist, or a major bank such as JPMorgan, Goldman Sachs, or Morgan Stanley. Whatever the entity, it must have significant financial resources to acquire the ETF's creation unit securities. The number of APs for a particular ETF varies depending on the size and activity of the fund. The fund pays nothing for the APs service. The AP earns profit through arbitrage. Because an ETF trades like any other security, its market price is affected by supply and demand. If many investors want to buy an ETF, the ETF's share price might rise above the fair value of its underlying securities or trade *at a premium*. When the ETF becomes "overpriced," the AP sells the ETF shares, buys the securities that underlie the ETF, and redeems the underlying securities for ETF shares with the issuer. Competition among APs drives the ETF price down (and the prices of the underlying securities up) until the premium disappears.

Conversely, suppose the ETF trades *at a discount* relative to its securities. In that case, the AP buys the ETF shares, redeems them for the underlying securities, and then sells them in the marketplace. Purchasing the "underpriced" ETF shares drives its price toward fair value while providing a costless arbitrage profit. Competition for the earnings from the costless arbitrage opportunities keeps the ETF's price in line with the NAV of its underlying securities.

An essential benefit of the creation/redemption mechanism is that the issuer has a cost-efficient way of acquiring new securities. When investors pour new money into mutual funds, the fund company must take that money and go into the marketplace to buy securities. In doing so, it pays bid/ask spreads and commissions, which drag on the fund's return. The same thing happens when investors remove money. With ETFs, APs do the buying and selling. When APs see new demand to buy ETF shares (i.e., the ETF share price moves to a premium), they enter the market and create new shares. When the APs see demand to sell (i.e., the ETF share prices fall to a discount), they process redemptions. Unlike mutual funds, the ETF issuer benefits from the AP paying all the trading costs and fees. In addition, the issuer receives a payment from the AP to process each creation/redemption.

The current landscape of the U.S. ETF industry

To give a sense of the landscape of the U.S. ETF industry, I downloaded some summary data from ETFdb.com for the year-end 20221229 and reported it in Table 1. Currently, there are 3,105 funds. Of these, 76.1% are equity ETFs, 19.6% are bond ETFs, and the remaining 4.3% are other asset classes. What is striking is the sheer amount invested in traditional asset classes. As discussed later, experiments into creating ETFs for non-traditional asset classes have had limited success.

Table 1: $AUM (in billions) by asset class on 20221229

Asset class	No. of ETPs	$AUM	% of total
Equity	2,122	4,907.9	76.1%
Bond	512	1,261.3	19.6%
Commodity	105	131.5	2.0%
Real Estate	45	68.5	1.1%
Multi-Asset	143	34.7	0.5%
Preferred Stock	16	30.1	0.5%
Alternatives	35	5.8	0.1%
Currency	20	4.0	0.1%
Volatility	17	3.0	0.0%
Total	3,015	6,446.8	

Table 2 shows the top 10 ETFs by billions of dollars in assets under management ($AUM). Several things are noteworthy. First, the top 10 funds account for over 27% of total $AUM. These funds are traditional (i.e., physical replication) ETFs with passive, transparent, market value-weighted index construction. Second, the top 3 are benchmarked to the S&P 500 portfolio, accounting for nearly 14% of all $AUM. AMEX launched SPY in January 1993. BlackRock's iShares IVV joined in May 2000. VOO, Vanguard's entrant, started trading in September 2010. While Vanguard entered the ETF much earlier (VTI was launched in 2001), Bogle refused to list VOO for fear of

cannibalizing the $AUM of his original Vanguard 500 Index Fund. Third, these funds' expense ratios (ERs) are a meager 0.03%, down 100 times from the 3% of actively managed mutual funds. Fourth, while theory says all U.S. stocks should be included in the benchmark, the S&P 500 has the most assets under management. The prevalent use of the S&P 500 index portfolio is attributable to several factors.[5] VTI, Vanguard's Total Stock Market ETF, is based on a benchmark index that is transparent, passive, and a market value-weighted combination of all U.S. stocks. Its $AUM has been growing year-by-year relative to SPY, IVV, and VOO. Fifth, only two ETFs are benchmarked to non-stock indexes, and both are bond funds. Vanguard's BND and BlackRock's AGG are written on the Bloomberg U.S. Aggregate Bond Index, a market value-weighted index of virtually all the 11,000+ investment-grade bonds traded in the U.S.

Table 2: Top 10 ETFs by $AUM (in billions) on 20221229

Symbol	Name	Asset class	$AUM	Inception	ER	% of total
SPY	SPDR S&P 500 ETF Trust	Equity	353.5	19930122	0.0945%	5.5%
IVV	iShares Core S&P 500 ETF	Equity	285.5	20000515	0.03%	9.9%
VOO	Vanguard S&P 500 ETF	Equity	258.4	20100907	0.03%	13.9%
VTI	Vanguard Total Stock Market ETF	Equity	256.8	20010524	0.03%	17.9%
QQQ	Invesco QQQ Trust	Equity	143.5	19990310	0.20%	20.1%
VEA	Vanguard FTSE Developed Markets ETF	Equity	99.6	20070720	0.05%	21.7%
VTV	Vanguard Value ETF	Equity	98.4	20040126	0.04%	23.2%
IEFA	iShares Core MSCI EAFE ETF	Equity	88.3	20121018	0.07%	24.6%
BND	Vanguard Total Bond Market ETF	Bond	84.7	20070403	0.03%	25.9%
AGG	iShares Core U.S. Aggregate Bond ETF	Bond	82.5	20030922	0.03%	27.2%
Total across all ETFs			6,446.8			

Finally, I prepared a summary of ETFs by issuer or sponsor and listed the top 10 by $AUM. The top 3 are often called the "Big Three" and account for 77% of all $AUM. Interestingly, all their products are traditional funds. The $AUMs of the other sponsors are much smaller. Of the 213 different issuers, the top 10 accounts for 92.3% of the total $AUM.

[5] Several factors come to mind. First, Bogle's success with the S&P 500 index mutual fund was likely the target of attack. He had succeeded on a grand scale with the Vanguard 500 Index Fund through the 1980s and early 1990s. Second, the most well-known, market value-weighted U.S. stock index in the early 1990s was the S&P 500. Total U.S. stock market indexes drew little attention. Third, the S&P 500 has been an actively traded futures market since March 1982. Arbitrage between the S&P 500 futures market and the S&P 500 ETF market promotes depth and liquidity in both.

Table 3: Market share by ETF issuer on 20221229				
	No. of		Market share	
Issuer	ETFs	$AUM	%	% of total
BlackRock Financial Management	384	2,183.3	33.9%	33.9%
Vanguard	81	1,866.8	29.0%	62.8%
State Street	139	914.0	14.2%	77.0%
Invesco	237	324.7	5.0%	82.0%
Charles Schwab	28	256.8	4.0%	86.0%
First Trust	197	132.0	2.0%	88.1%
JPMorgan Chase	47	88.8	1.4%	89.4%
Dimensional	29	71.4	1.1%	90.6%
World Gold Council	2	58.4	0.9%	91.5%
ProShares	137	56.8	0.9%	92.3%
Total no. of issuers	213			

References

Bogle, John C., 2016, David and Goliath: Who wins the quantitative battle? *Journal of Portfolio Management* 43, 127-137.

ICI, 2022, *Investment Company Fact Book: A review of trends and activities in the investment company industry*. Investment Company Institute, www.icifactbook.org.

Lintner, John, 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics* 47, 13-37.

Markowitz, Harry, 1952, Portfolio selection, *Journal of Finance* 12, 77-91.

Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing," *Journal of Economic Theory* 13, 341–360.

Samuelson, Paul A., 1974, Challenge to judgment, *Journal of Portfolio Management* 1 (Fall), 17-19.

Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425-442.

Sharpe, William F., 1992, Asset allocation: Management style and performance measurement, *Journal of Portfolio Management* 18, 7-19.

Tobin, James, 1958, Liquidity preference as behavior towards risk, *Review of Economic Studies* 25 (February), 65-86.