Personal Indexes

One day, artificial intelligence will build them for each investor.

By Andrew W. Lo

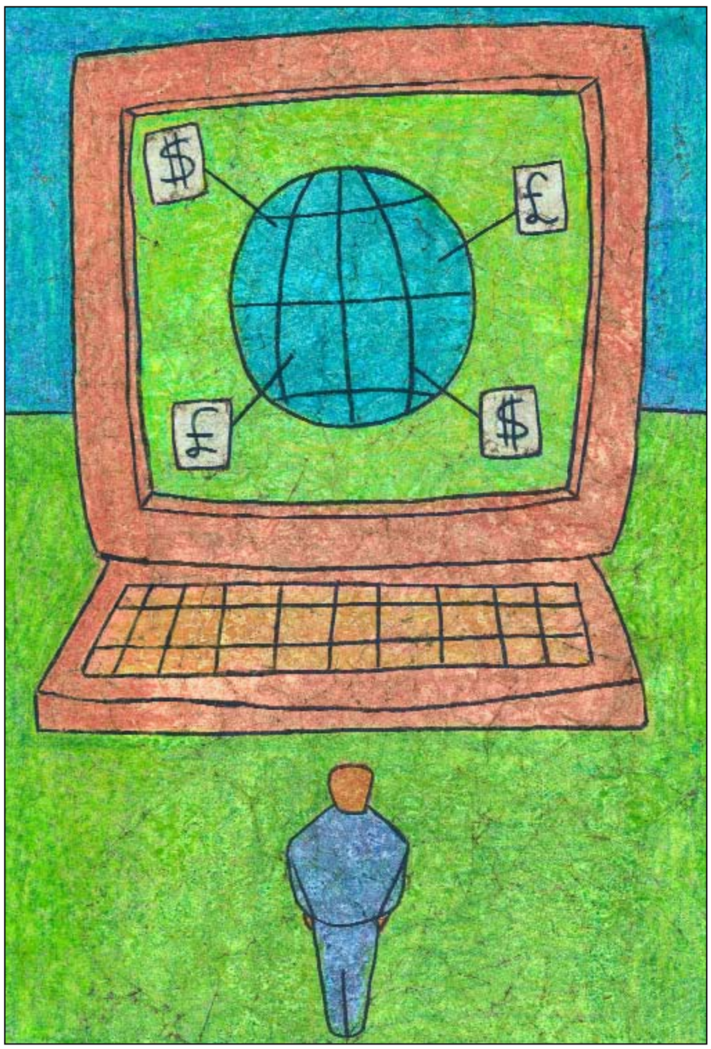


Illustration by Garrian Manning

One of the main purposes of an index of any kind is to facilitate the extraction and summary of information through an algorithmic process, e.g., averaging. Now typically done by computers, averaging is a simple example of artificial intelligence! Indeed, an important factor in the early popularity of the Dow Jones Industrial Average, first introduced in May 1896, was the "Dow Theory," a collection of heuristics proposed by Charles H. Dow and later expanded by William P. Hamilton in which specific economic meaning is attributed to certain time series patterns in the index.[1] The dream of developing automated processes for making better investment decisions is obviously not unique to our times.

But what *is* unique about our times is the confluence of breakthroughs in financial technology, computer technology, and institutional infrastructure that, for the first time in the history of modern civilization, makes automated personalized investment management a practical possibility. The combination of artificial intelligence and financial technology may one day render general market indexes obsolete: Each investor may have a "personal" index constructed specifically to meet his or her lifetime objectives and risk preferences, and a software agent to actively manage the portfolio accordingly. If this seems more like science fiction than reality, that is precisely the motivation for a review of some basic recent developments in artificial intelligence and their applications to financial technology.

One of the earliest and most enduring models of the behavior of security prices—and stock indexes in particular—is the Random Walk Hypothesis, an idea conceived in the sixteenth century as a model of games of chance.[2] As with so many of the ideas of modern economics, the first serious application of the Random Walk Hypothesis to financial markets can be traced back to Paul Samuelson (1965), whose contribution is neatly summarized by the title of his article, "Proof that Properly Anticipated Prices Fluctuate Randomly". Samuelson argued that in financial markets randomness is achieved through the active participation of many investors seeking greater wealth. An army of greedy investors trade aggressively on even the smallest informational advantages at their disposal, and in doing so, they incorporate their information into market prices and quickly eliminate the profit opportunities that gave rise to their trading.

While this argument for randomness is surprisingly compelling, a number of theoretical and empirical studies over the past 20 years have cast serious doubt on both its premises and its implications. For example, from a theoretical perspective, LeRoy (1973), Lucas (1978), and many others have shown in many ways and in many contexts that the Random Walk Hypothesis is neither a necessary nor a sufficient condition for rationally determined security prices. And empirically, numerous researchers have documented departures from the Random Walk Hypothesis in financial data.[3] Financial markets are predictable to some degree.

The rejection of the Random Walk Hypothesis opens the door to the possibility of superior long-term investment returns through disciplined active investment management. In much the same way that innovations in biotechnology can garner superior returns for venture capitalists, innovations in financial technology can, in principle, garner superior returns for investors. This is compelling motivation for the application of artificial intelligence in financial contexts.

Artificial Neural Networks

Recent advances in the theory and implementation of (artificial) neural networks have captured the imagination and fancy of the financial community.[4] Although they are only one of the many types of statistical tools for modeling nonlinear relationships, neural networks seem to be surrounded by a great deal of mystique and, sometimes, misunderstanding. Because they have their roots in neurophysiology and the cognitive sciences, neural networks are often assumed to have brain-like qualities: learning capacity, problem-solving abilities, and ultimately cognition and self-awareness. Alternatively, neural networks are often viewed as "black boxes" that can yield accurate predictions with little modeling effort.

In fact, neural networks are neither. They are an interesting and potentially powerful modeling technique in some, but surely not all, applications. To develop some basic intuition for neural networks, consider a typical nerve cell or “neuron.” A neuron has dendrites (receptors) at different sites that react to stimulus. This stimulus is transmitted along the axon by an electrical pulse. If the electrical pulse exceeds some threshold level when it hits the nucleus, this triggers the nucleus to react, *e.g.*, to make a particular muscle contract. This basic biological unit is what mathematicians attempt to capture in a neural network model.[5]

Even though neurobiologists have come to realize that actual nerve cells exhibit considerably more complex and subtle behavior, nevertheless AI researchers have found great use for the simple on/off or “binary threshold” model in approximating nonlinear relationships efficiently. In particular, neural network models have a very useful feature known as the “universal approximation property”. This property means that with enough neurons linked together *in an appropriate way*, a neural network can approximate any nonlinear relationship, no matter how strange.

Viewed as a statistical estimation technique, neural networks are a flexible model of nonlinearities. In this sense, they are just one of many techniques for modeling complex relationships. Examples of other nonlinear estimation techniques include: splines, wavelets, kernel regression, projection pursuit, radial basis functions, nearest-neighbor estimators and, perhaps the most powerful of all, human intuition.

Even simple neural networks can capture a variety of nonlinearities. Consider, for example, the sine function plus a random error term:

$$Y_t = \mathrm{Sin}(X_t) + 0.5\varepsilon_t \tag{1}$$

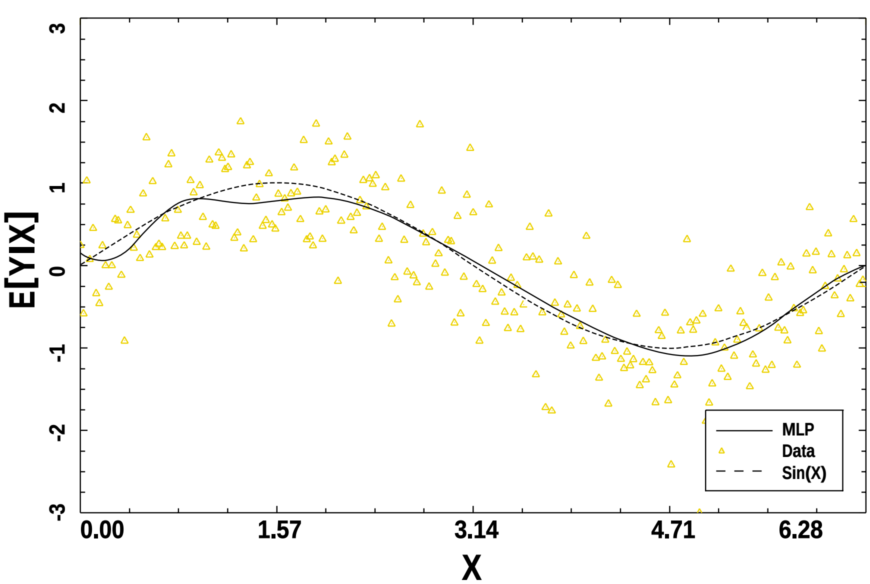
where ε_t is a standard normal random variable. Can a neural network extract the sine function from observations (X_t, Y_t)?

To answer this question, 500 (X_t, Y_t) pairs were randomly generated subject to the nonlinear relationship (3), and a neural network model was estimated using this artificial data (or in the jargon of this literature, a neural network was trained on this data set).[6] The following equation is the result of training a neural network on the data using nonlinear least squares:

$$\begin{aligned} Y_t = {} & 5.282 - 14.576\,\Theta(-1.472 + 1.869X_t) - \\ & 5.411\Theta(-2.628 + 0.642X_t) - \\ & 3.071\Theta(13.288 - 2.347X_t) + \\ & 6.320\Theta(-2.009 + 4.009X_t) + \\ & 7.892\Theta(-3.316 + 2.474X_t) \end{aligned} \tag{2}$$

where $\Theta(x)$ is the logistic function $1/(1 + \exp[x])$. This network has five identical activation functions $\Theta(x)$ (corresponding to the

Figure 1 MLP Estimator of Y = Sin(X) + .5Z



five nodes in the hidden layer) and a constant term. The network has only two inputs, X_t and 1.

Now (2) looks nothing like the sine function, so in what sense has the neural network "approximated" the nonlinear relation (1)? In Figure 1, the data points (X_t,Y_t) are plotted as triangles, the dashed line is the theoretical relation to be estimated (the sine function), and the solid line is the relation as estimated by the neural network (2). The solid line is impressively close to the dashed line, despite the noise that the data clearly contain. Therefore, although the functional form of (2) does not resemble any trigonometric function, its numerical values do. Add more data points, and it would likely get even closer.

Hutchinson, Lo, and Poggio (1994) proposed neural network models for estimating derivative pricing formulas. In particular, they take as inputs the primary economic variables that influence the derivative's price—current underlying asset price, strike price, time-to-maturity, etc.—and define the derivative price to be the output into which the neural network maps the inputs. When properly trained, the network "becomes" the derivative pricing formula, which may be used in the same way that formulas obtained from parametric pricing methods such as the well-known Black-Scholes formula are used: for pricing, delta-hedging, simulation exercises, etc.

These neural network models have several important advantages over the more traditional parametric models. First, since they do not rely on restrictive parametric assumptions such as lognormality or sample-path continuity, they are robust to the specification errors that plague parametric models. (In other words, you don't have to worry about having the right assumptions if you don't have to make any assumptions). Second, they are adaptive, and respond to structural changes in the data-generating processes in ways that parametric models cannot. Third, they are flexible enough to encompass a wide range of derivative securities and fundamental asset price dynamics, yet relatively simple to implement. And finally, they are easily parallelizable and may be computationally more efficient.

Of course, all these advantages do not come without some cost—the nonparametric pricing method is highly data-intensive, requiring large quantities of historical prices to obtain a sufficiently well-trained network. Therefore, such an approach would be inappropriate for thinly traded derivatives, or newly created derivatives that have no similar counterparts among existing securities.[7] Also, if the fundamental asset's price dynamics are well-understood and an analytical expression for the derivative's price is available under these dynamics, then the parametric formula will almost always outperform the network formula in pricing and hedging accuracy. Nevertheless, these conditions occur rarely enough that there may still be great practical value in constructing derivative pricing formulas by learning networks. To illustrate the practical relevance of their approach, Hutchinson, Lo, and Poggio (1994) apply it to the pricing and delta-hedging of S&P 500 futures options from 1987 to 1992. They show that neural network models perform well, yielding delta-hedging errors and option prices that are comparable to and, in some cases, better than traditional methods like Black-Scholes.

Data Mining and Data Snooping

A substantial portion of the recent literature in artificial intelligence is devoted to a discipline now known as "data mining" or "knowledge discovery in databases" (KDD). The combination of very large scale databases in many research and business contexts and the tremendous growth in computing power over the past several decades has naturally led to the development of computationally intensive methods for systematically sifting through large quantities of data.

In fact, Internet-based search engines are perhaps the most common examples of data mining applications, and there are many other prominent examples in marketing, financial services, telecommunications, and molecular biology. Perhaps the most challenging issue facing data miners today is a statistical one: how to determine whether the results from a data-mining search are genuine or spurious? For example, suppose we search a database of mutual funds to find the one with the most successful track record over the past five years, and the process yields XYZ Growth Fund. Does this imply that XYZ is a good fund, or is it possible that XYZ's performance is a fluke?

In searching for the presence of any effect, whether it is superior investment performance or a causal relationship between two characteristics, the dilemma will always be present: If the effect exists, data mining algorithms will generally detect it; if the effect does not exist, data mining algorithms will usually still find an "effect" anyway. The latter case is known as a "data-snooping bias." The problem is being aware of the kind of result you have.

To see how serious a problem this dilemma can be, suppose we have a collection of n mutual funds with (random) annual returns $R_1, R_2, \ldots, R_n$ respectively that are mutually independent and have the same probability distribution function $F_R(r)$.[8] That is, they have nothing to do with each other.

Now, for concreteness, suppose that these returns are normally distributed with an expected value of 10% per year and a standard deviation of 20% per year, roughly comparable to the historical behavior of the S&P 500. Under these assumptions, what is the probability that the return on fund i exceeds 50%? Because the distribution is normal, we know the probability in any given year is about 2.3%:

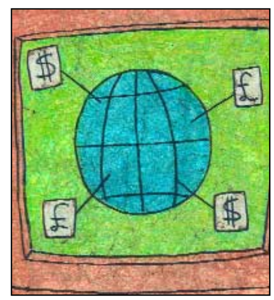
$$\text{Prob}(R > 0.50) = 1 - \text{Prob}(R \leq 0.50) = 0.0228\,. \tag{3}$$

But suppose we focus not on any arbitrary fund i, but rather on the fund that has the largest return among all n funds. Although we do not know in advance which fund this will be, nevertheless we can characterize this best-performing fund in the abstract, in much the same way that college admissions offices can construct the profiles of the applicants with the highest standardized test scores. However, this analogy is not completely accurate because on average, test scores do seem to bear some relation to subsequent academic performance. In our n-fund example, even though there will always be a "best-performing" fund or "winner," the *subsequent* performance of this winner will be statistically identical to all the other funds by assumption, i.e., the same expected return, the same volatility, and the same probability law.

This distinction is the essence of the data-snooping problem. There will always be a winner. The question is: Does winning tell us anything about the true nature of the winner? In the case of standardized test scores, the generally accepted answer is yes. In the case of the n independently and identically distributed mutual funds, the answer is no. The larger the sample, the larger (and therefore perhaps more tempting and hard to ignore) the largest score is likely to be.

To quantify this effect, we can derive the probability law of the return R^* of the "best-performing" fund:

$$R^* = \text{Max}\,[\,R_1, R_2, \ldots, R_n\,] \tag{4}$$



which is given by the following distribution function:

$$F_{R^*}(r) = [F_R(r)]^n. \tag{5}$$

After data snooping, the probability of observing performance greater than 50% is given by:

$$\text{Prob}(R^* > 0.50) = 1 - \text{Prob}(R^* \leq 0.50) = 1 - [F_R(r)]^n = 1 - (0.9772)^n. \tag{6}$$

When $n = 1$, the probability that R^* exceeds 50% is the same as the probability that R exceeds 50%: 2.3%. But among a sample of $n = 100$ securities, the probability that R^* exceeds 50% is $1 - (0.9772)^{100}$ or 90.0%! Not surprisingly, the probability that the *largest* of 100 independent returns exceeds 50% is considerably greater than the probability that any individual fund's return exceeds 50%. However, this has no bearing on future returns since we have assumed that all n funds have the same mean and variance, and they are statistically independent of each other.

How are the properties of R^* related to data-snooping biases in financial analysis? Investors often focus on past performance as a guide to future performance, associating past successes with significant investment skills. But if superior performance is the unavoidable result of the selection procedure—picking the strategy or manager with the most successful track record, for example—then past performance is not necessarily an accurate indicator of future performance.

In other words, the selection procedure may bias our perception of performance, causing us to attribute superior performance to an investment strategy or manager that was merely "lucky."

There are statistical procedures that can partially offset the most obvious types of data-snooping biases,[9] but the final arbiter must inevitably be the end-user of the data-mining algorithm. By trading off the cost of one type of error (detecting effects that do not exist) with the other (not detecting effects that do exist), a sensible balance between data mining and data snooping needs to be struck.[10] The necessity of at least some human intervention at the decision-making point may yet prove to an insurmountable limitation of artificial intelligence.

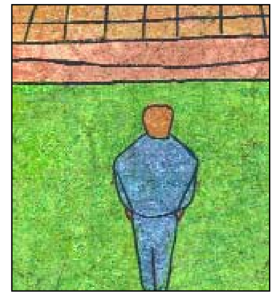
Pattern Recognition

One of the biggest rifts that divide academic finance and industry practice is the separation between technical analysts and their academic critics. In contrast to fundamental analysis, which was quick to be adopted by the scholars of modern quantitative finance, technical analysis has been an orphan from the very start. In some circles, technical analysis is known as "voodoo finance." In his influential book *A Random Walk Down Wall Street*, Burton Malkiel (1996) concludes that "[u]nder scientific scrutiny, chart-reading must share a pedestal with alchemy."

One explanation for this state of controversy and confusion is the unique and sometimes impenetrable jargon used by technical analysts. Campbell, Lo, and MacKinlay (1997, pp. 43–44) provide a striking example of the linguistic barriers between technical analysts and academic finance by contrasting these two statements:

> *The presence of clearly identified support and resistance levels, coupled with a one-third retracement parameter when prices lie between them, suggests the presence of strong buying and selling opportunities in the near-term.*

> *The magnitudes and decay pattern of the first 12 autocorrelations and the statistical significance of the Box-Pierce Q-statistic suggest the presence of a high-frequency predictable component in stock returns.*



Despite the fact that both statements have the same basic meaning—that past prices contain information for predicting future returns—most academics find the first statement puzzling and the second plausible.

These linguistic barriers underscore an important difference: Technical analysis is primarily visual, while quantitative finance is primarily algebraic and numerical. Technical analysis employs the tools of geometry and pattern recognition, while quantitative finance employs the tools of mathematical analysis and probability and statistics. In the wake of recent breakthroughs in financial engineering, computer technology, and numerical algorithms, it is no wonder that quantitative finance has overtaken technical analysis in popularity. The principles of portfolio optimization are far easier to program into a computer than the basic tenets of technical analysis.

Nevertheless, technical analysis has survived, perhaps because its visual mode of analysis is more conducive to human cognition, and because pattern recognition is one of the few repetitive activities for which computers do not have an absolute advantage (yet). However, this is changing. Artificial intelligence has made admirable progress in the automation of pattern detection, hence the possibility of automating technical analysis is becoming a reality. In particular, Lo, Mamaysky, and Wang (2000) have proposed an algorithm for detecting technical indicators such as "head-and-shoulders" and "double bottoms," and have applied it to the daily prices of several hundred U.S. stocks over a 30-year period to evaluate the information content of such patterns.

This process is motivated by the general goal of technical analysis, which is to identify regularities in the time series of prices by extracting nonlinear patterns from noisy data. Implicit in this goal is the recognition that some price movements are significant—they contribute to the formation of a specific pattern—and others are merely random fluctuations to be ignored. In many cases, the human eye can perform this "signal extraction" quickly and accurately, and until recently computer algorithms could not.

However, "smoothing" estimators such as kernel regression are ideally suited to this task because they extract nonlinear relations by "averaging out" the noise. Lo, Mamaysky, and Wang (2000) use these estimators to mimic, and in some cases sharpen, the skills of a trained technical analyst in identifying certain patterns in historical price series.

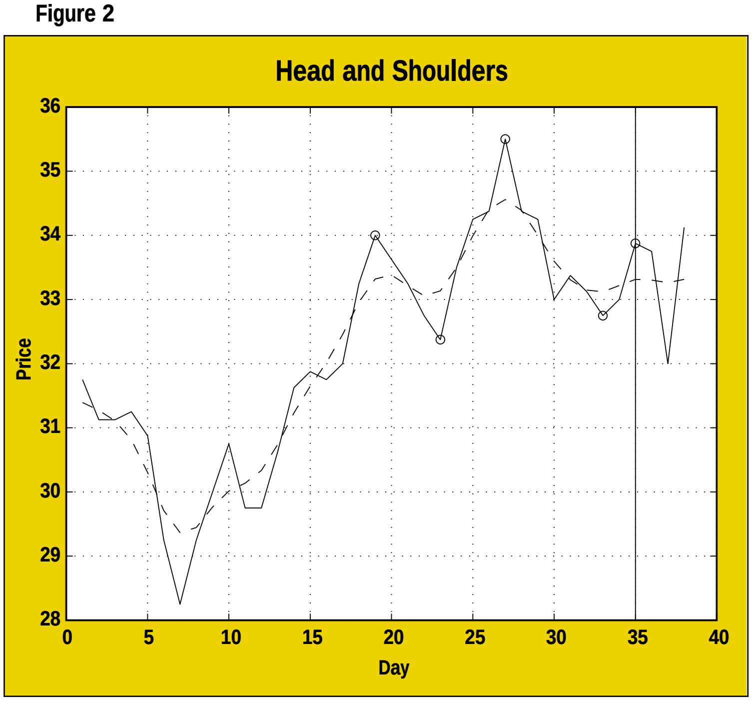
Armed with a mathematical representation of the time series of historical prices from which geometric properties can be characterized in an objective manner, they construct an algorithm for automating the detection of technical patterns consisting of three steps:

1. Define each technical pattern in terms of its geometric properties, e.g., local extrema (maxima and minima) so that algorithms for identifying its occurrence can be developed.
2. Construct a kernel-regression estimator of a given time series of prices so that its extrema can be determined numerically.
3. Analyze the fitted curve of this estimator for occurrences of each technical pattern.

The first step is the most challenging since this is the heart of the pattern-recognition algorithm and where much of the creativity of human technical analysts comes into play. For example, note that only five consecutive extrema are required to identify a head-and-shoulders pattern (although its completion requires two more, where it initially and finally crosses the "neckline"). This follows from the formalization of the geometry of a head-and-shoulders pattern: three peaks, with the middle peak higher than the other two. Because consecutive extrema must alternate between maxima and minima for smooth functions,[11] the three-peaks pattern corresponds to a sequence of five local extrema: maximum, minimum, highest maximum, minimum, and maximum.

Lo, Mamaysky and Wang (2000) define nine other technical patterns in similar fashion, and once they have been given mathematical precision in this way, the detection of these patterns can be readily automated. An illustra-

Figure 2

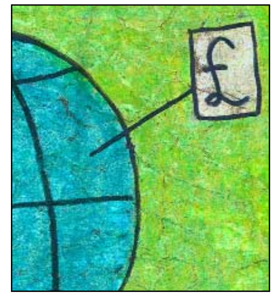


tion of their algorithm at work is given in Figure 2. When they apply their algorithm to daily prices of over 300 US stocks from 1962 to 1996, they find that certain technical indicators do provide incremental information, and that technical indicators tend to be more informative for NASDAQ stocks than for NYSE or AMEX stocks.

While pattern-recognition techniques have been successful in automating a number of tasks that were previously considered to be uniquely human endeavors—fingerprint identification, handwriting analysis, and face recognition, for example—nevertheless, it is possible that no machine algorithm is a perfect substitute for the skills of an experienced technical analyst. However, if an algorithm can provide a reasonable approximation to *some* of the cognitive abilities of a human analyst, such an algorithm can be used to leverage the skills of any technician. Moreover, if technical analysis is an art form that can be taught, then surely its basic precepts can be quantified and automated to some degree. And as increasingly sophisticated pattern-recognition techniques are developed, a larger fraction of the art will become a science.

Conclusions

In this article, I have only scratched the surface of the many applications of artificial intelligence that will be transforming financial technology over the next few years. Other emerging technologies include artificial markets and agent-based models of financial transactions, electronic market-making, modeling emotional responses as computational algorithms ("affective computing"), the psychophysiology of risk preferences, and financial visualization. Artificial intelligence will undoubtedly play a more central role in active investment management, but this does not imply that indexation will become less relevant for investors.



Artificial intelligence and active management are not at odds with indexation, but instead imply the evolution of a more sophisticated set of indexes and portfolio management policies for the typical investor, something investors can look forward to, perhaps within the next decade. Imagine a software program that constructs a custom-designed index for each investor according to his or her risk preferences, financial objectives, insurance needs, retirement plans, tax bracket, etc. This "SmartIndex" will serve to guide each investor towards a path of long-term financial security—a path that is unique to each investor—so that if an investor's portfolio return differs by more than a certain margin from the return of his or her SmartIndex, this will serve as an "early-warning signal" to change the investment policy to get back on track. Whether or not an investor's portfolio outperforms the S&P 500 in any given year will not be as relevant as whether it outperforms the investor's SmartIndex, hence such an innovation will change the very nature of indexation and the role of index funds and benchmark returns. This concept may well be science fiction today, but the technology for SmartIndexes already exists. As with the transformation of all great ideas from theory into practice, it is only a matter of time.

References

Ait-Sahalia, Y. and A. Lo, 1998, " Nonparametric Estimation of State-Price Densities Implicit In Financial Asset Prices" , *Journal of Finance* 52, 499–548.

Campbell, J., Lo, A., and C. MacKinlay, 1997, *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press.

Hald, A., 1990, *A History of Probability and Statistics and Their Applications Before 1750*. New York: John Wiley and Sons.

Hertz, J., A. Krogh, and R. Palmer, 1991, *Introduction to the Theory of Neural Computation*. Reading, MA: Addison-Wesley Publishing Company.

Hamilton, W., 1922, *The Stock Market Barometer*. New York: Harper and Row (republished in 1998 by John Wiley & Sons).

Hutchinson, J., Lo, A., and T. Poggio, 1994, " A Nonparametric Approach to Pricing and Hedging Derivative Securities via Learning Networks" , *Journal of Finance* 49, 851–889.

Iyengar, S. and J. Greenhouse, 1988, " Selection Models and the File Drawer Problem" , *Statistical Science* 3, 109–135.

Jin, L. and A. Lo, 2001, " How Anomalous Are Financial Anomalies?" , unpublished working paper, MIT Sloan School of Management.

Leamer, E., 1978, *Specification Searches*. New York: John Wiley & Sons.

LeRoy, S., 1973, " Risk Aversion and the Martingale Property of Stock Returns" , *International Economic Review* 14, 436–446.

Lo, A., 1994a, " Neural Networks and Other Nonparametric Techniques in Economics and Finance" , in H. Russell Fogler, ed.: *Blending Quantitative and Traditional Equity Analysis*. Charlottesville, VA: Association for Investment Management and Research.

Lo, A., 1994b, " Data-Snooping Biases in Financial Analysis" , in H. Russell Fogler, ed.: *Blending Quantitative and Traditional Equity Analysis*. Charlottesville, VA: Association for Investment Management and Research.

Lo, A. and C. MacKinlay, 1990, " Data Snooping Biases in Tests of Financial Asset Pricing Models" , *Review of Financial Studies* 3, 431–468.

Lo, A. and C. MacKinlay, 1999, *A Non-Random Walk Down Wall Street*. Princeton, NJ: Princeton University Press.

Lo, A., Mamaysky, H., and J. Wang, 2000, " Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation" , *Journal of Finance* 55, 1705–1765.

Lucas, R., 1978, " Asset Prices in an Exchange Economy" , *Econometrica* 46, 1429–1446.

Malkiel, B., 1996, *A Random Walk Down Wall Street*. New York: W.W. Norton.

Samuelson, P., 1965, " Proof that Properly Anticipated Prices Fluctuate Randomly" , *Industrial Management Review* 6, 41–49.

White, H., 1992, *Artificial Neural Networks: Approximation and Learning Theory*. Cambridge, MA: Blackwell Publishers.

Footnotes

1 See Hamilton (1922) for a thorough exposition of the Dow Theory.
2 See, for example, Hald (1990, Chapter 4).
3 See, for example, Lo and MacKinlay (1988, 1999).
4 The adjective " artificial" is often used to distinguish mathematical models of neural networks from their biological counterparts. For brevity, I shall omit this qualifier for the remainder of this article although it will be implicit in all of my references to neural networks.
5 Since this is meant to be an overview, I will not hesitate to sacrifice rigor for clarity. A more extensive overview is given in Lo (1994). More mathematically-inclined readers are encouraged to consult Hertz, Krogh, and Palmer (1991) for the general theory of neural computation, White (1992) for the statistics of neural networks, and Ait-Sahalia and Lo (1994) and Hutchinson, Lo, and Poggio (1994) for financial applications.
6 Specifically, the particular neural network is a " single hidden-layer feedforward perceptron" with five nodes. As of yet, there are no formal rules for selecting the optimal configuration or " topology" of a neural network, and this is one of the primary drawbacks of neural network models. Currently, experience and heuristics are the only guides we have for specifying the network topology.
7 However, since newly created derivative securities can often be replicated by a combination of existing derivatives, this is not as much of a limitation as it may seem at first.
8 Recall that the distribution function $F_R(r)$ of a random variable R is defined as $F_R(r) = \text{Prob}(R < r)$.
9 See, for example, Leamer (1978), Iyengar and Greenhouse (1988), Lo and MacKinlay (1990, and Lo (1994).
10 See Jin and Lo (2001) for an example of this kind of analysis in a financial context.
11 After all, for two consecutive maxima to be local maxima, there must be a local minumum in between, and vice versa for two consecutive minima.