Data-Snooping Biases in Financial Analysis

Andrew W. Lo
Harris & Harris Group Professor of Finance
Massachusetts Institute of Technology

Data-snooping biases can never be completely eliminated; they are an unavoidable aspect of nonexperimental inference and are a particular problem in nonlinear models because of the large degrees of freedom involved in these models. Awareness of the influence of data snooping is the most important step in dealing with the problem. Solutions to the problem will almost never be statistical in nature; rather, analysts should look to some kind of framework to limit the number of possibilities in their search. The framework might come from economic theory, psychological theory, or the analyst's intuition, judgment, and experience.

Data snooping—finding seemingly significant but, in fact, spurious patterns in the data—is a serious problem in financial analysis. Although it afflicts all nonexperimental sciences, data snooping is particularly problematic for financial analysis because of the large number of empirical studies performed on the same data sets. Given enough time, enough attempts, and enough imagination, almost any pattern can be teased out of any data set. In some cases, these spurious patterns are statistically small, almost unnoticeable in isolation. But because small effects in financial calculations can often lead to very large differences in investment performance, data-snooping biases can be surprisingly substantial.

An Example

Consider the following quantitative stock selection model that I call Θ–94.[1] It is based upon a mathematical result derived by the great French mathematician Pierre de Fermat in the arcane area of number theory. Fermat discovered that any prime number p—a number that is evenly divisible only by itself and 1 (e.g., 3, 5, 7, 11)—has the following property: When 2^{p-1} is divided by p, there is *always* a remainder of 1. For example, when $2^{13-1} = 2^{12} = 4{,}096$ is divided by 13, the result is 315 plus a remainder of 1. Remarkably, this property holds for *all* prime numbers (try to prove it!).

However, the converse is *not* true. That is, if some number q satisfies the property that 2^{q-1} divided by q has a remainder of 1, this does not necessarily mean that q is prime. But it is almost true; there are very few numbers q that satisfy this property and are *not* prime. In fact, they are so rare that they have their own name: "Carmichael" numbers. Of the numbers from 1 to 10,000, there are only seven Carmichael numbers: 561, 1,105, 1,729, 2,465, 2,821, 6,601, and 8,911. The stock selection strategy Θ–94 consists of selecting those stocks with one of these seven Carmichael numbers embedded in their CUSIP (Committee on Uniform Securities Identification Procedures) identifiers.

Now, before you laugh too hard, consider how Θ–94 performed on a backtest of NYSE and AMEX stocks over the period from 1926 to 1991. For simplicity, I consider only those stocks that were continuously listed during this period, so as to avoid stocks coming in and out of my sample at random dates. From this subsample, it turns out that there is exactly one stock with one of the seven Carmichael numbers embedded in its CUSIP: Ametek, Inc., with CUSIP 03110510. I had never heard of Ametek before my backtest, and I have no idea what its business is, but here are its performance statistics:

Mean monthly return	=	0.017
Standard deviation of monthly return	=	0.142
Skewness	=	1.4067
Kurtosis	=	11.3346
Total return	=	$320.75
Jensen alpha	=	5.15
Sharpe ratio	=	0.86
Treynor ratio	=	11.39
Appraisal ratio	=	0.13

[1] I am grateful to Professor Robert Merton for suggesting this example.

By any standards, these figures demonstrate extraordinary performance. The stock selection strategy Θ–94 is an incredible success.

Now, the problem with Θ–94 is that there is no reason why it should work. Despite Fermat's brilliance, prime numbers have nothing to do with stock selection. Or do they? Some people might argue that *why* it works does not matter, as long as it *does* work. This kind of logical positivism is a very misleading and potentially dangerous view, especially when applied to nonexperimental sciences (disciplines where you cannot perform controlled experiments). Why things work matters a great deal when you cannot test hypotheses by running repeated controlled experiments.

The fact is that Θ–94 has an important but subtle bias built into its backtest: The stock has survived for a 65-year period. If, in 1926, you obtained a list of stocks that were sure to survive over the next 65 years, you could throw a dart at that list and garner handsome returns.

But observe that Θ–94 didn't involve any obvious sorting or selection bias beyond the simple continuous-listing requirement. I just took those continuously listed stocks and picked randomly using the (highly nonlinear) Carmichael algorithm. Neural nets and other nonlinear techniques can all too easily give similar kinds of results. Unless you have some form of discipline to guide your analysis, you too can come up with some brilliant results that blow up out of sample (how do you think Θ–94 would do next year?).

A lot of what data snooping is about is asking the wrong question. In the next few sections, I hope to show you all sorts of ways of asking the wrong questions and getting some spectacular answers. Some counterexamples will demonstrate how to ask the right questions.

Psychological Foundations of Data Snooping

Data snooping can be loosely defined as finding patterns in the data that do not exist. Unfortunately, this kind of spurious pattern recognition is an unavoidable aspect of empirical work; the process by which we make genuine progress in understanding data is also the exact same process by which data snooping creeps in.

The data-snooping problem is unique to nonexperimental sciences, which must rely solely on statistical inference for their insights. The reason that physicists, biologists, and engineers perform *controlled* experiments is precisely to get rid of these data-snooping biases. They can set all the dials exactly where they want them, do the experiment, fiddle with the dials, and redo the experiment to see exactly what works and what does not. Unfortunately, we cannot do that, so data snooping is a problem that is going to be with us as long as we are using data.

Data-snooping biases are most likely to arise under several conditions. First, they are more likely when large amounts of data exist. Clearly this applies to financial markets. The second circumstance is when many analysts use exactly the same data set. That condition also applies to financial markets. The third condition is an absence of theory or controls. By theory, I mean not only economic theory but also intuition and judgment, and restrictions that come from experience. The fourth and perhaps most insidious source of data-snooping biases is the attitude that, if something works, it does not matter why. To avoid these kinds of data-snooping biases, you must raise many questions in nonexperimental situations and understand *why* a model works (think back to Θ–94).

Data snooping has very deep psychological roots. In fact, human perception and cognition are biased toward the unusual. If nothing happens, we do not notice. The very definition of "taking notice" means something unusual has happened. Professor Stephen Ross (1987) addressed this phenomenon in his study "Regression to the Max," in which he raises the Zen-like paradox: Do you notice an event because it is interesting, or is it interesting because you notice it? Alternatively, do we mistake causation for correlation?

The well-known psychologist Carl Jung called this human predilection toward coincidences "synchronicity." His idea of synchronicity is that, if two events are separated by time and space but the observer somehow feels that they are connected, perhaps supernaturally, then that perception of coincidence creates meaning, even though the two events are not causally linked. Jung used this theory to try to explain the supernatural—the fact that you could be sitting in your chair thinking of something, and suddenly, that thing happens. He believed that what people ascribe to supernatural causes is this human predilection toward synchronicity: We observe a coincidence and assign it meaning, and therefore, it takes on meaning in our minds.

Data-snooping biases can arise from synchronicity. For this reason, it is crucial in nonexperimental sciences to understand why something works. Whether or not it truly does work will always be in doubt because we cannot perform controlled experimentation. And if you do not understand why it works the way it does, you may ascribe meaning to it simply because of synchronicity.

There are several potential solutions to the problem of data snooping. One is not to deal with data, but for most of us, that solution is not practical. The second, and probably the most important, is to admit that the problem exists. As with most forms of substance abuse, the first sign of recovery is the recognition that you have a problem. The next step is to attempt to model the learning process explicitly. Typically, the way we have learned about data is by trial and error. We try something out, if it does not work, we throw it away, try something else, and go through the same process all over again. The trial-and-error process is very difficult to quantify in a model, but such a model could control for a great many forms of data snooping.

Another solution is to use proper statistical inference, the kind that takes simple selection biases into account (see the discussion in the next section on order statistics). However, we almost never learn about the data using these very simple kinds of data-snooping techniques. We snoop the data in all sorts of subtle ways (recall the backtest of Θ–94), and these subtleties are very difficult to capture statistically. Using proper statistical inference is certainly one way to handle it, but it is not a complete solution. In fact, I will argue in the next section that no complete solution to this problem exists.

Order Statistics

To see how one might deal with the data-snooping problem by proper statistical inference, we shall consider the use of "order statistics." Order statistics are exactly what they sound like: statistics that are ordered. An example will make this clearer.

Suppose we have a collection of n securities with (random) annual returns $R_1, R_2, \ldots, R_n$, respectively. For simplicity, assume that the returns are mutually independent and that they have the same probability distribution function $F_R(r)$.[2] That is, they have nothing to do with each other (if you observe one return, this gives you no information about the likely outcomes of the others), and they are governed by the exact same probability law.

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

$$\begin{aligned}\text{Prob}(R_i > 0.50) &= 1 - \text{Prob}\ (R_i \le 0.50) \\ &= 1 - \text{Prob}\left(\frac{R_i - 0.10}{0.20} \le \frac{0.50 - 0.10}{0.20}\right) \\ &= 1 - \text{Prob}(Z \le 2) = 1 - 0.9772 \\ &= 0.0228,\end{aligned}$$

where Z is a standard normal random variable (mean 0, standard deviation 1), and the second-to-last equality follows from any table of values of the cumulative normal distribution function. This tells us that it is very unlikely indeed that the return of security i will exceed 50 percent in any given year.

But suppose we focus not on any arbitrary security i but, rather, on the security that has the *largest* return among all n securities. Although we do not know in advance which security this will be, nevertheless we can characterize this best-performing security 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, whereas in our n-security example, even though there will always be a "best-performing" security or "winner," the *subsequent* performance of this winner will be identically distributed to all the other securities by assumption.

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 securities, the answer is no. Data-snooping biases arise when the distinction between these two examples is ignored.

To quantify the magnitude of the data-snooping bias in the n-security example, we begin by characterizing the statistical properties of the winner's return $R^*(n)$, where

$$R^*(n) \equiv \text{Max}(R_1, R_2, \ldots, R_n)\,. \tag{2}$$

$R^*(n)$ is often called the largest order statistic of the sample $\{R_i\}$ since its definition is based purely on the ordering of the sample. Its statistical distribution function $F_{R^*(n)}(r)$ can be derived explicitly as

$$F_{R^*(n)}(r) \equiv \text{Prob}\,[R^*(n) < r] = [F_R(r)]^n\,, \tag{3}$$

where $F_R(r)$ is the distribution function of each of the

[2]Recall that the distribution function $F_R(r)$ of a random variable R is defined as $F_R(r) \equiv \text{Prob}(R \le r)$.

individual securities' returns. Using Equation 3, we can now answer the following question: What is the probability that the annual return of the *best-performing* of the n securities exceeds 50 percent? Using Equation 3 in Equation 4, we have

$$
\begin{aligned}
\text{Prob}\,[R^*(n) > 0.50] &= 1 - [F_{R^*(n)}(0.50)] \\
&= 1 - [F_R(0.50)]^n \\
&= 1 - [\text{Prob}\,(R_i \le 0.50)]^n \\
&= 1 - (0.9772)^n . \qquad (4)
\end{aligned}
$$

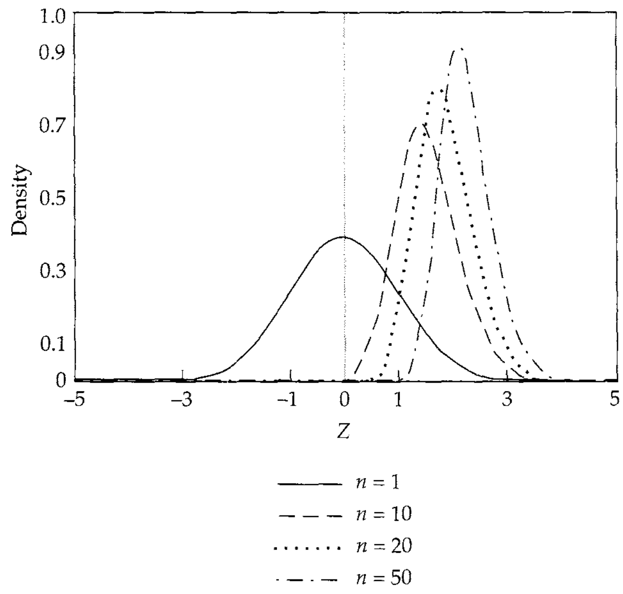
When $n = 1$, observe that Equation 4 yields the same probability that $R^*(n)$ exceeds 50 percent that we obtained for R_i exceeding 50 percent—2.3 percent; with only one security in the sample, $R^*(n)$ and R_i are one and the same.

But consider the probability that $R^*(n)$ exceeds 50 percent in a sample of $n = 100$ securities. In this case, according to Equation 4, the probability of this event is $1 - (0.9772)^{100}$ or 90.0 percent! Not surprisingly, the probability that the largest of 100 independent returns exceeds 50 percent is considerably greater than the probability that any individual return exceeds 50 percent, even though all securities have exactly the same expected return, the same variance, and indeed, the same distribution function.

The very different behavior of R_i and $R^*(n)$ is captured in **Figure 1**, in which the probability density function[3] of a standard normal random variable (the center curve) and the probability density functions of the largest order statistic of a sample of n standard normal random variables are plotted for various values of n (the curves to the right of the center curve). For example, the dashed curve is the probability density function that applies when you are selecting the largest of 10 standard normal random variables. The dotted-and-dashed density function at the extreme right is the density function that applies when selecting the largest of 50 standard normal random variables. Figure 1 confirms our intuition that the largest of 10 independent random security returns is smaller, on average, than the largest of 50 random security returns. As the number of securities increases, the probability density function of the largest return shifts out to the right.

How are the properties of order statistics 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 result of the selection procedure—picking the strategy or manager with the most successful track record—then past performance will *not necessarily* be 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."

Figure 1. Densities of N (0,1) Order Statistics



Source: Andrew W. Lo.

A more subtle example of data-snooping biases involves recent statistical tests of the capital asset pricing model (CAPM) based on data-driven anomalies such as the empirical relation between equity returns and factors such as market capitalization and the price-to-book ratio.[4] In these studies, the CAPM is tested by calculating the alphas of market-capitalization- or price/book-sorted portfolios, but how were these two characteristics selected? Sorting portfolios by a specific characteristic such as market capitalization can bias you toward rejecting the CAPM, even if the CAPM is true, if the choice of the sorting characteristic is purely data driven and not based on any a priori theoretical considerations.

Order statistics have a variety of applications in other sciences as well. For example, there is the "file drawer" problem and "meta-analysis" in biostatistics (see Iyengar and Greenhouse [1988] for an excellent review). At the conclusion of a typical clinical trial in which the effectiveness of a particular drug is

[3]Recall that a probability density function is a continuous-state analog to a histogram—the probability of the random variable taking on a value between points a and b is represented by the area of the region under the probability density function between a and b.

[4]See Black (1993) and Lo and MacKinlay (1990a) for more detailed discussions of data-snooping biases in testing the CAPM.

studied, if a significant effect is found, the results are published in a journal; if not, the results are placed in a file drawer, never to be seen by the general public. Therefore, what we see in the medical journals are only the "significant" clinical trials, and this is another selection procedure not unlike the one which generated $R^*(n)$ in our earlier example.

One solution that biostatisticians have proposed is to report the hypothetical number of *unpublished* studies containing insignificant results that would have to exist to reverse the significance of the published studies, and to allow readers to judge for themselves whether this number is plausible or not. For example, if the existence of only three unpublished studies containing insignificant findings were sufficient to reverse the significance of a published one, this raises some doubt about the genuine significance of the published study. If, however, a published study contained such strong results that it would require over 100 unpublished insignificant studies to reverse its findings, the file-drawer problem would not be as much of a concern. It remains to be seen whether this approach can be applied successfully to financial analysis.

Performance Evaluation Applications

Perhaps the most direct application of order statistics is in performance evaluation. For example, suppose we have m mutual fund managers, each managing a portfolio over a 13-year period. On a risk-controlled basis, what is the probability that a given manager will outperform the S&P 500 in 11 of those 13 years, and would this be sufficient evidence of superior performance management?[5]

One approach to answering this question is to define a performance statistic, k_{it}, that is equal to 1 if manager i outperforms the market in year t and is equal to 0 if manager i does not outperform the market in year t (after adjusting for risk). Assume that there is *no* superior performance ability, so that

$$\text{Prob}(k_{it} = 1) = \frac{1}{2} \tag{5}$$

for all i and t. To apply the analysis of order statistics developed earlier, we also require that k_{it} is independently and identically distributed across both i and t.[6] If the managers have no ability whatsoever, then their performance is like a coin toss every year—a 50/50 probability of a positive alpha.

What about the aggregate performance of manager i over the 13-year period? This is measured by k_i:

$$k_i = \sum_{t=1}^{13} k_{it}. \tag{6}$$

The probability that any given manager i outperforms the market in every single year is simply the probability that $k_i = 13$. Under our assumption (Equation 5) of no performance ability, this probability is $(1/2)^{13}$, which is approximately 1 in 10,000—quite unlikely.

What about outperforming the market in 11 out of 13 years under the assumption of no performance ability? The probability that $k_i = 11$ turns out to be about 9 in 1,000; out of a sample of 1,000 mutual fund managers, none of whom have any skill, 9 of them will outperform in 11 out of 13 years purely by chance.

Although this is already a startling and somewhat counterintuitive statistic, there is an even more surprising statistic that comes from asking the more relevant question: What is the probability that the *best-performing* mutual fund manager out of a sample of m managers will outperform the market in 11 out of 13 years?

The distribution of $k^*(m)$—the maximum performance across all m mutual fund managers under the assumption that the performance statistics $\{k_i\}$ are independently and identically distributed and that no manager has superior skills—can be calculated explicitly using the theory of order statistics. **Table 1** reports the probability that the best-performing manager will outperform the market in more than k years out of 13 years, where k varies from 7 to 11 and m is the number of managers over which we are doing the selection. When m equals 200, for example, the probability that the best-performing fund outperforms the market in more than 10 out of 13 years is 89.6 percent! If you focus on the most successful manager, you are going to see some very anomalous results, but these results do not necessarily indicate great skill. Data-snooping biases can be very powerful indeed.

Table 1. Prob ($k^* > k$); $T = 13$

m	$k = 7$	$k = 8$	$k = 9$	$k = 10$	$k = 11$
50	1.000	0.999	0.906	0.431	0.082
100	1.000	1.000	0.991	0.677	0.157
200	1.000	1.000	1.000	0.896	0.290
300	1.000	1.000	1.000	0.966	0.401
400	1.000	1.000	1.000	0.989	0.495
500	1.000	1.000	1.000	0.996	0.575

Source: Andrew W. Lo.

[5]This example is motivated by the extraordinary record of Peter Lynch's Magellan Fund, which did outperform the S&P 500 in 11 out of 13 years.

[6]These requirements can be relaxed to some extent, but the analysis becomes considerably more complex. See David (1981) for further details.

The conclusion we should draw from this example is that, even if you do not have any skill, it may seem that you do and even if you do have skill, it may seem that you do not. This is why economic theory, in contrast to pure statistics, is needed to understand the outcome. Statistics alone can never answer the question that we as financial analysts are interested in. The essence of data snooping is that focusing on interesting events is quite different from trying to figure out which events are interesting. The former can be accomplished purely by statistics, while the latter requires additional theoretical discipline.

Size versus Power

The statistical inference in performance evaluation can yield two kinds of mistakes. One is concluding that there is superior performance when there is not; the other, concluding no superior performance when there is. The problem with statistical corrections for data snooping is that correcting for one type of error often increases the likelihood of the other type of error. There is generally a trade-off between the two.

For example, if the cutoff for determining whether a manager has skill is a perfect record—outperformance in all 13 years—then in addition to excluding those managers with no skills, this unrealistic hurdle will also exclude those managers who do have skills. Unfortunately, such a high hurdle will always exclude managers with some skill—such as those who can outperform the market 51 percent of the time—because the standards are so high. However, lowering the hurdle to outperformance in 11 out of 13 years will include a certain number of managers who do not have any skill. If we try to reduce one kind of error, more of the other kind creeps in.

Consider another example, where we are trying to evaluate the performance of a portfolio manager who happened to call the biotechnology boom. Suppose that, 20 years ago, this portfolio manager said, "Biotechnology is likely to revolutionize the pharmaceuticals and agricultural industries. I am going to invest 100 percent of my assets in biotechnology stocks." So she bet on this industry, and 20 years later, her performance looks fantastic. Her risk-adjusted alpha is 200 basis points.

Now, academics have documented that, over the past 20 years, small-capitalization stocks seem to have done better than large-capitalization stocks. So, in evaluating the performance of this biotechnology fund, in addition to taking out the market factor, you also take out a small-cap index, and now the performance of this portfolio manager looks dismal. Does this imply the absence of superior performance?

Not necessarily. Since the biotechnology industry is relatively new, even the successful companies will have relatively small capitalization. Therefore, the biotechnology fund manager's performance will be highly correlated with a small-cap index, and once this index is statistically removed from the biotechnology fund's returns, there may be very little left; that is, the alpha may be negative. We have "overcorrected," controlling for characteristics of the data that are correlated with genuine performance, and despite the presence of superior performance, we may come to the opposite conclusion.

These kinds of data-snooping corrections are too conservative. While they guard against incorrectly attributing skill to managers without any, for those managers who do have skill, we will never detect it. Therefore, while data-snooping corrections can be useful, they are not without their own limitations. And while there are more sophisticated alternatives to correct for data-snooping biases—Monte Carlo simulations, Bayesian analysis, asymptotic theory, etc.—the basic problem is that, when we deal with nonexperimental data, we will always be subject to mistakes of one kind or another.

Conclusion

Data-snooping biases cannot be completely eliminated. They are an unavoidable aspect of nonexperimental inference. Awareness of the problem is the first and probably biggest step towards minimizing the influence of data snooping. Simple biases such as picking the best-performing mutual fund manager out of a large pool can be corrected, but some of the more subtle biases are still going to exist. The solutions to these problems are almost never going to be statistical in nature, because no statistical procedure is free of data-snooping biases. A partial solution is to provide theoretical restrictions, some kind of framework that will limit the total number of possibilities that we search over, whether the framework be from economic theory, psychological theory, or our own intuition, judgment, or experience.

Because of the larger degrees of freedom, nonlinear models are more likely to succumb to data-snooping biases than linear ones. If we think about a nonlinear model as picking over a large number of independent models or parameters, nonlinear models have many more of those than do linear models, so it is much easier to come up with significant results that are not really there.

Question and Answer Session

Andrew W. Lo

Question: Is it not more appropriate to ask if any investor has the ability to select a fund manager who can perform 11 out of 13 years rather than to ask whether such a manager exists?

Lo: That is a difficult issue. My own feeling is that picking such a manager purely from performance statistics alone will not be very successful. Despite the numerical example of spurious mutual fund performance, I do believe that Peter Lynch has superior skills, but I do not believe it because he outperformed the market in 11 out of 13 years. The reason I believe he has some skill is because of what I know about what he did: how he went about the investment process, the size and growth of the Magellan Fund, the turnover of the fund, the magnitude by which it outperformed the market, market conditions at the time, etc.

The short answer to your question is that individuals would have difficulty in picking out successful managers. We have all done the calculation: Given the signal-to-noise ratio of equities, you would need 2,000 years of data to tell whether or not a manager has a genuine alpha of 50 basis points. As a portfolio manager, telling your clients that they must leave their money with you for 2,000 years before they can fire you with just cause does not go over very well. The only alternative is to try to bring in other information, such as bits of intuition or judgment about what the manager is doing. A clean test would be to ask a manager to tell you in advance how he or she is going to make money, exactly what procedure will be used to earn excess returns and then see whether the manager actually did that after the fact. That approach also takes time—five or ten years, perhaps—but that is the essence of a fair test.

Question: Doesn't the performance problem have two sides? Just as you cannot prove performance, clients cannot prove you underperformed, unless you are very bad.

Lo: Yes, but clients aren't always as rational as finance professors when it comes to performance! I've heard it said that investors only worry about downside volatility, and this produces an asymmetry in how your clients deal with under- and overperformance. By the way, knowing the probability of underperformance is just as valuable as knowing the probability of overperformance. If you can identify somebody who will consistently give you bad advice, that is incredibly valuable. Just take whatever the person says, put a minus sign in front of it, and watch your investments grow!

Question: What about evaluating performance on a shorter horizon to have more data to analyze?

Lo: Certainly, the solution is to bring more data to bear, but if you have lots of data and other people have exactly the same data, you all are going to be subject to the same biases. Out-of-sample data would work. If today you decide what investment strategy you want to use, put it in place, and see how it performs, that is a clean test. The problem with that approach is you have to wait, but for most short-horizon trading strategies, it is actually feasible. For intraday trading, it is possible to do a hold-out sample of 20 days and see how it works. For asset allocation, that kind of a clean experiment is virtually impossible to do.

Question: I am also under the impression that shorter data will compress your alpha correspondingly. Is it true that your actual tests will not change if you shorten the period?

Lo: Whether the tests change depends upon which market you are in. For equities, the signal-to-noise ratio is going to be roughly proportional to the time horizon. For nonlinear instruments, such as options or highly leveraged securities, the tests could differ dramatically over different horizons.

Question: If I wanted to prove statistically that the CAPM does not work, what kinds of tests would I use and what results would I have to demonstrate?

Lo: I believe that disproving the CAPM statistically is impossible. MacKinlay (1987) wrote a paper showing the impossibility of distinguishing statistically among multifactor linear models of equity returns. His basic conclusion is that the signal-to-noise ratio in equity returns is so low that, using statistics alone, you cannot reliably distinguish between a one-factor CAPM and, say, a two-factor arbitrage pricing theory model. The two models will look virtually identical in a sample of 10 or 20 years of monthly equity data calibrated to the kinds of means, variances, and covariances that we typically see in historical returns.

Merton (1973) has developed a dynamic version of the CAPM in which a stochastic investment opportunity set—for example, time-varying interest rates—will create a second factor that investors use to hedge changes in the opportunity set. That is an example of a theory that will tell you what to look for. For example, a potential problem with the Fama and French (1992) analysis that Black (1993) pointed out is that there is no theory as to why the price-to-book ratio is an important factor in determining expected equity returns; Fama and French simply document its statistical significance. A recent study by Breen and Korajcyzk (1994) shows that Compustat's tendency to "backfill" data can create spurious significance for the price-to-book ratio when, in fact, none exists. In a data set that is free from such backfilling, Breen and Korajcyzk found considerably less explanatory power for the price-to-book ratio. So, once again, such an analysis has to return to first (theoretical) principles. Can you come up with a better theory for determining expected equity returns than the single-factor CAPM?

Question: Would you comment on the use of systems of differential equations that specify structure and values for a neural net versus back-propagation?

Lo: Your question is somewhat misleading, in that back-propagation is a specific technique for estimating the connection strengths of a neural network. So, regardless of whether you impose structure on the network or not, you can use back-propagation or you can use a "batch" or "off-line" method such as nonlinear least squares to estimate connection strengths.

Perhaps a more relevant question is: Is it valuable to have systems of structural equations that tell you what is going on and provide some guidance for the network topology? The answer to that question is yes, absolutely. The "topology" of a neural network must have some kind of structure because, otherwise, anything goes. And if anything goes, you will find something significant eventually if you look hard enough. Having some kind of restrictions is exceedingly important, and perhaps systems of differential equations that come from structural models are the way to go. It is not the only way to go, but it is certainly one way.