rapid and broader adoption of Bayesian methodology. The importance of Bayesian thought and methodology is manifest in the vignettes of Beck, Eisenberg, Lo, Raftery, Rossi and Allenby, Sobel, and Tsay.

One of the real pleasures of pulling this collection of vignettes together was to see a large number of themes recur through many of the vignettes. In addition to those just mentioned, latent variables, ecological regression, nonlinearity, long-range dependence, graphical models, and networks (social and neural) are among the topics that reemerge. Latent variables arise in a variety of contexts, including mixture models and latent trait models, and they figure into most of the vignettes. Beck highlights the role of political scientists in developing approaches to addressing the ecological inference problem, and Eisenberg notes the importance of this methodology in voting rights cases as well as the controversy that surrounds its use. Nonlinearity is a recurrent theme in various modeling contexts throughout the collection. For example, it arises in Browne's vignette in the context of generalizations of structural equation models and in Tsay's vignette in the context of nonlinear processes. Both authors point to these as important areas for future work. Tsay discusses long-range dependence in the context of data from communications networks and from financial markets, and Lo goes into some depth on this issue in his discussion of the stochastic nature of financial asset prices. Fienberg discusses graphical models in the context of log-linear model theory, and Sobel discusses them in the context of their usage for drawing causal inferences. Both Fienberg and Raftery note the importance of recent work on social networks, as well as the importance of work that remains to be done. The vignettes of Lo and of Rossi and Allenby make references to ways that neural networks are being used in business applications.

In summary, this collection points to the excitement of past and future developments arising from the interdigitation of statistics with business and social science. Though the types of questions arising in the various fields and the motivation behind them vary to some extent, it is clear that statistical thought and methodology is central to advancement of our understanding of human behavior and interactions. The opportunities presented by new and evolving technologies for collecting more and better data are abundant, and these will no doubt continue to motivate new statistical research and applications for many years to come. It is hoped that these vignettes will stimulate statistical scientists to become more deeply engaged in the challenges and problems of business and social science. The authors have pointed the way to a wide array of interesting challenges arising at the interstices of statistics with the economic, behavioral, and social sciences, and there are suggestions that we stand to profit by also bringing the biological and physical sciences to bear on some of these challenges. An attraction of the field of statistics has always been its broad applicability to interesting and important problems, and this collection demonstrates the numerous and intellectually challenging opportunities for making valuable contributions in various areas.

Finance: A Selective Survey

Andrew W. Lo

1. INTRODUCTION

Ever since the publication in 1565 of Girolamo Cardano's treatise on gambling, *Liber de Ludo Aleae* (*The Book of Games of Chance*), statistics and financial markets have been inextricably linked. Over the past few decades, many of these links have become part of the canon of modern finance, and it is now impossible to fully appreciate the workings of financial markets without them. In this brief survey, I hope to illustrate the enormous research opportunities at the intersection of finance and statistics by reviewing three of the most important ideas of modern finance: efficient markets, the random walk hypothesis, and derivative pricing models. Although it is impossible to provide a thorough exposition of any of these ideas in this brief essay, my less ambitious goal is to communicate the excitement of financial research to statisticians and to stimulate further collaboration between these two highly complementary disciplines. It is also impossible to provide an exhaustive bibliography for each of these topics—that would exceed the page limit of this entire article—and hence my citations are selective, focusing on more recent and most relevant developments for the readers of this journal. (For a highly readable and entertaining account of the recent history of modern finance, see Bernstein 1992.)

To develop some context for the three topics that I have chosen, consider one of the most fundamental ideas of economics, the principle of supply and demand. This principle states that the price of any commodity and the quantity traded are determined by the intersection of supply and de-

Andrew W. Lo is Harris & Harris Group Professor and director of the MIT Laboratory for Financial Engineering, MIT Sloan School of Management, Cambridge, MA 02142 (E-mail: *alo@mit.edu*). The author is grateful to Samantha Arrington and Mark Becker for helpful comments. This research was partially supported by MIT Laboratory for Financial Engineering and the National Science Foundation grant SBR-9709976.

Journal of the American Statistical Association
June 2000, Vol. 95, No. 450, Vignettes

mand curves, where the demand curve represents the schedule of quantities desired by consumers at various prices and the supply curve represents the schedule of quantities that producers are willing to supply at various prices. The intersection of these two curves determines an "equilibrium," a price-quantity pair that satisfies both consumers and producers simultaneously. Any other price–quantity pair may serve one group's interests, but not the other's.

Even in this simple description of a market, all the elements of modern finance are present. The demand curve is the aggregation of many individual consumers' desires, each derived from optimizing an individual's preferences subject to a budget constraint that depends on prices and other factors (e.g., income, savings requirements, borrowing costs). Similarly, the supply curve is the aggregation of many individual producers' outputs, each derived from optimizing an entrepreneur's preferences subject to a resource constraint that also depends on prices and other factors (e.g., costs of materials, wages, trade credit). Probabilities affect both consumers and producers as they formulate their consumption and production plans through time and in the face of uncertainty—uncertain income, uncertain costs, and uncertain business conditions.

It is the interaction between prices, preferences, and probabilities—sometimes called the "three p's of total risk management" (see Lo 1999)—that gives finance its richness and depth. Formal models of financial asset prices such as those of Breeden (1979), Lucas (1978), and Merton (1973a) show precisely how the three p's simultaneously determine a "general equilibrium" in which demand equals supply across *all* markets in an uncertain world where individuals and corporations act rationally to optimize their own welfare. Typically, these models imply that a security's price is equal to the present value of all future cashflows to which the security's owner is entitled. Several aspects make this calculation unusually challenging: individual preferences must be modeled quantitatively, future cashflows are uncertain, and so are discount rates. Pricing equations that account for such aspects are often of the form

$$P_t = E_t\left[\sum_{k=1}^{\infty} \gamma_{t,t+k} D_{t+k}\right], \tag{1}$$

and their intuition is straightforward; today's price must equal the expected sum of all future payments D_{t+k} multiplied by discount factors $\gamma_{t,t+k}$ that act as "exchange rates" between dollars today and dollars at future dates. If prices do not satisfy this condition, this implies a misallocation of resources between today and some future date, not unlike a situation in which two commodities sell for different prices in two countries even after exchange rates and shipping costs have been taken into account (a happy situation for some enterprising arbitrageurs, but not likely to last very long).

What determines the discount factors $\gamma_{t,t+k}$? They are determined through the equalization of supply and demand, which in turn is driven by the preferences, resources, and expectations of all market participants; that is, they are determined in general equilibrium. It is this notion of equilibrium, and all of the corresponding ingredients on which it is based, that lies at the heart of financial modeling.

2. EFFICIENT MARKETS

There is an old joke, widely told among economists, about an economist strolling down the street with a companion when they come upon a \$100 bill lying on the ground. As the companion reaches down to pick it up, the economist says "Don't bother—if it were a real \$100 bill, someone else would have already picked it up."

This humorous example of economic logic gone awry strikes dangerously close to home for proponents of the efficient markets hypotheses, one of the most controversial and well-studied propositions in all the social sciences. It is disarmingly simple to state, has far-reaching consequences for academic pursuits and business practice, and yet is surprisingly resilient to empirical proof or refutation. Even after three decades of research and literally hundreds of journal articles, economists have not yet reached a consensus about whether markets—particularly financial markets—are efficient or not.

As with so many of the ideas of modern economics, the origins of the efficient markets hypothesis 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." In an informationally efficient market, price changes must be unforecastable if they are properly anticipated; that is, if they fully incorporate the expectations and information of all market participants. In the context of the basic pricing equation (1), the conditional expectation operator $E_t[\cdot] \equiv E[\cdot|\Omega_t]$ is defined with respect to a certain set of information Ω_t; hence elements of this set cannot be used to forecast future price changes, because they have already been impounded into current prices. Fama (1970) operationalized this hypothesis—summarized in his well-known expression "Prices fully reflect all available information"—by specifying the elements of the information set Ω_t available to market participants; for example, past prices, or all publicly available information, or all public and private information.

This concept of informational efficiency has a wonderfully counterintuitive and "Zen-like" quality to it: The more efficient the market, the more random the sequence of price changes generated by such a market, and the most efficient market of all is one in which price changes are completely random and unpredictable. In contrast to the passive motivation that inspires randomness in physical and biological systems, randomness in financial systems is not an implication of the principle of insufficient reason, but instead is the outcome of many active participants attempting to profit from their information. Motivated by unbridled greed, speculators aggressively pounce 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 speculation. If this occurs instantaneously, which it must in an idealized world of "frictionless" markets and costless trading, then prices must always fully reflect all available information, and no profits can be garnered from information-based trading (because such profits have already been captured).

Such compelling motivation for randomness is unique among the social sciences and is reminiscent of the role that uncertainty plays in quantum mechanics. Just as Heisenberg's uncertainty principle places a limit on what we can know about an electron's position and momentum if quantum mechanics holds, this version of the efficient markets hypothesis places a limit on what we can know about future price changes if the forces of financial self-interest are at work.

However, one of the central tenets of modern finance is the necessity of some trade-off between risk and expected returns, and whether or not predictability in security prices is inefficient can be answered only by weighing it against the risks inherent in exploiting such predictabilities. In particular, if a security's price changes are predictable to some degree, then this may be just the reward needed to attract investors to hold the asset and bear the associated risks (see, e.g., Lucas 1978). Indeed, if an investor is sufficiently risk averse, then he might gladly *pay* to avoid holding a security that has unforecastable returns.

Despite the eminent plausibility of such a trade-off—after all, investors must be rewarded to induce them to bear more risk—operationalizing it has proven a formidable challenge to both finance academics and investment professionals. Defining the appropriate measures of risk and reward, determining how they might be linked through fundamental principles of economics and psychology, and then estimating such links empirically using historical data and performing proper statistical inference are issues that have occupied much of the finance literature for the past half-century, beginning with Markowitz's (1952) development of portfolio theory and including Sharpe's (1964) capital asset pricing model (CAPM), Merton's (1973a) intertemporal CAPM, Ross's (1976) arbitrage pricing theory, and the many empirical tests of these models. Moreover, recent advances in methods of statistical inference, coupled with corresponding advances in computational power and availability of large amounts of data, have created an exciting renaissance in the empirical analysis of efficient markets, both inside and outside the halls of academia; in earlier work (Lo 1997) I provided an overview and a more complete bibliography of this literature.

3. THE RANDOM WALK

Quite apart from whether or not financial markets are efficient, one of the most enduring questions of modern finance is whether financial asset price changes are forecastable. Perhaps because of the obvious analogy between financial investments and games of chance, mathematical models of financial markets have an unusually rich history that predates virtually every other aspect of economic analysis. The vast number of prominent mathematicians, statisticians, and other scientists who have applied their considerable skills to forecasting financial security prices is a testament to the fascination and the challenges that this problem poses.

Much of the early finance literature revolved around the random walk hypothesis and the martingale model, two statistical descriptions of unforecastable price changes that were (incorrectly) taken to be implications of efficient markets. One of the first tests of the random walk was devised by Cowles and Jones (1937), who compared the frequency of *sequences* and *reversals* in historical stock returns, where the former are pairs of consecutive returns with the same sign and the latter are pairs of consecutive returns with opposite sign. Many others performed similar tests of the random walk (see Lo 1997 and Lo and MacKinlay 1999 for a survey of this literature), and with the exception of Cowles and Jones (who subsequently acknowledged an error in their analysis), all reported general support for the random walk using historical stock price data.

However, some recent research has sharply contradicted these findings. Using a statistical comparison of variances across different investment horizons applied to the weekly returns of a portfolio of stocks from 1962 to 1985, Lo and MacKinlay (1988) found that the random walk hypothesis can be rejected with great statistical confidence (well in excess of .999). In fact, the weekly returns of a portfolio containing an equal dollar amount invested in each security traded on the New York and American Stock Exchanges (called an *equal-weighted* portfolio) exhibit a striking relation from one week to the next: a first-order autocorrelation coefficient of .30.

An autocorrelation of .30 implies that approximately 9% of the variability of next week's return is explained by this week's return. An equally weighted portfolio containing only the stocks of "smaller" companies, companies with market capitalization in the lowest quintile, has a autocorrelation coefficient of .42 during the 1962–1985 sample period, implying that about 18% of the variability in next week's return can be explained by this week's return. Although numbers such as 9% and 18% may seem small, it should be kept in mind that 100% predictability yields astronomically large investment returns; a very tiny fraction of such returns can still be economically meaningful.

These findings surprise many economists, because a violation of the random walk necessarily implies that price changes are forecastable to some degree. But because forecasts of price changes are also subject to random fluctuations, riskless profit opportunities are not an immediate consequence of forecastability. Nevertheless, economists still cannot completely explain why weekly returns are not a "fair game." Two other empirical facts add to this puzzle:

1. Weekly portfolio returns are strongly positively autocorrelated, but the returns to individual securities generally are not; in fact, the average autocorrelation—averaged across individual securities—is negative (and statistically insignificant).
2. The predictability of returns is quite sensitive to the holding period; serial dependence is strong and positive for daily and weekly returns but is virtually zero for returns over a month, a quarter, or a year.

For holding periods much longer than 1 week (e.g., 3–5 years), Fama and French (1988) and Poterba and Summers (1988) found negative serial correlation in U.S. stock returns indexes using data from 1926 to 1986. Although their estimates of serial correlation coefficients seem large in magnitude, there are insufficient data to reject the random walk hypothesis at the usual levels of significance. Moreover, a number of statistical biases documented by Kim, Nelson, and Startz (1991) and Richardson (1993) cast serious doubt on the reliability of these longer-horizon inferences.

Despite these concerns, models of long-term memory have been a part of the finance literature ever since Mandelbrot (1971) applied Hurst's (1951) rescaled range statistic to financial data. Time series with long-term memory exhibit an unusually high degree of persistence, so that observations in the remote past are nontrivially correlated with observations in the distant future, even as the time span between the two observations increases. Nature's predilection toward long-term memory has been well documented in the natural sciences such as hydrology, meteorology, and geophysics, and some have argued that economic time series thus must also have this property.

But, using recently developed asymptotic approximations based on functional central limit theory, I (Lo 1991) constructed a test for long-term memory that is robust to short-term correlations of the sort uncovered by Lo and MacKinlay (1988, 1999), and concluded that despite earlier evidence to the contrary, there is little support for long-term memory in stock market prices. Departures from the random walk hypothesis can be fully explained by conventional models of short-term dependence for most financial time series. However, new data are being generated each day, and the characteristics of financial time series are unlikely to be stationary over time as financial institutions evolve. Perhaps some of the newly developed techniques for detecting long-term memory—borrowed from the statistical physics literature—will shed more light on this issue (see, e.g., Mandelbrot 1997; Pilgram and Kaplan 1998).

More recent investigations have focused on a number of other aspects of predictability in financial markets: stochastic volatility models (Gallant, Hsieh, and Tauchen 1997), estimation of tail probabilities and "rare" events (Jansen and de Vries 1991), applications of "chaos theory" and nonlinear dynamical systems (Hsieh 1991), Markov-switching models (Gray 1996), and mixed jump-diffusion models (Bates 1996). This research area is one of the most active in the finance literature, with as many researchers in industry as in academia developing tools to detect and exploit all forms of predictabilities in financial markets.

Finally, in contrast to the random walk literature, which focuses on the conditional distribution of security returns, another strand of the early finance literature has focused on the *marginal* distribution of returns, and specifically on the notion of "stability," the preservation of the parametric form of the marginal distribution under addition. This is an especially important property for security returns, which are summed over various holding periods to yield cumulative investment returns. For example, if P_t denotes the end-of-month-t price of a security, then its monthly continuously compounded return x_t is defined as $\log(P_t/P_{t-1})$, and hence its annual return is $\log(P_t/P_{t-12}) = x_t + x_{t-1} + \cdots + x_{t-11}$. The normal distribution is a member of the class of stable distributions, but the nonnormal stable distributions have a distinguishing feature not shared by the normal: they exhibit leptokurtosis or "fat tails," which seems to accord well with higher-frequency financial data, such as daily and weekly stock returns. Indeed, the fact that the historical returns of most securities have many more outliers than predicted by the normal distribution has rekindled interest in this literature, which has recently become part of a much larger endeavor known as "risk management."

Of course, stable distributions have played a prominent role in the early development of modern probability theory (see, e.g., Lévy 1937), but their application to economic and financial modeling is relatively recent. Mandelbrot (1960, 1963) pioneered such applications, using stable distributions to describe the cross-sectional distributions of personal income and of commodity prices. Fama (1965) and Samuelson (1967) developed the theory of portfolio selection for securities with stably distributed returns, and Fama and Roll (1971) estimated the parameters of the stable distribution using historical stock returns. Since then, many others have considered stable distributions in a variety of financial applications; McCulloch (1996) has provided an excellent and comprehensive survey.

More recent contributions include the application of invariance principles of statistical physics to deduce scaling properties in tail probabilities (Mandelbrot 1997; Mantegna and Stanley 1999), the use of large-deviation theory and extreme-value theory to estimate loss probabilities (Embrechts, Kluppelberg, and Mikosch 1997), and the derivation of option-pricing formulas for stocks with stable distributions (McCulloch 1996).

4. DERIVATIVE PRICING MODELS

One of the most important breakthroughs in modern finance is the pricing and hedging of "derivative" securities, securities with payoffs that depend on the prices of other securities. The most common example of a derivative security is a call option on common stock, a security that gives its owner the right (but not the obligation, hence the term "option") to purchase a share of the stock at a prespecified price K (the "strike price") on or before a certain date T (the "expiration date"). For example, a 3-month call option on General Motors (GM) stock with a \$90 strike price gives its owner the right to purchase a share of GM stock for \$90 any time during the next 3 months. If GM is currently trading at \$85, is the option worthless? Not if there is some probability that GM's share price will exceed the \$90 strike price some time during the next 3 months. It seems, therefore, that the price of the option should be determined in equilibrium by a combination of the statistical properties of GM's price dynamics and the preferences of investors buying and selling this type of security, as in the pricing equation (1).

However, Black and Scholes (1973) and Merton (1973b) provided a compelling alternative to (1), a pricing model based only on arbitrage arguments and not on general equilibrium. [In fact, the Black and Scholes (1973) framework does rely on equilibrium arguments—it was Merton's (1973b) application of continuous-time stochastic processes that eliminated the need for equilibrium altogether (see Merton 1992 for further discussion).] This alternative is best illustrated through the simple *binomial option-pricing model* of Cox, Ross, and Rubinstein (1979), a model in which there are two dates, 0 and 1, and the goal is to derive the date-0 price of a call option with strike price K that expires at date 1. In this simple economy, two other financial securities are assumed to exist: a riskless bond that pays a gross rate of return of r (e.g., if the bond yields a 5% return, then $r = 1.05$) and a risky security with date-0 price P_0 and date-1 price P_1 that is assumed to be a Bernoulli random variable:

$$P_1 = \begin{cases} uP_0 & \text{with probability } \pi \\ dP_0 & \text{with probability } 1 - \pi, \end{cases} \tag{2}$$

where $0 < d < u$. Because the stock price takes on only two values at date 1, the option price takes on only two values at date 1 as well:

$$C_1 = \begin{cases} C_u \equiv \text{Max}[uP_0 - K, 0] & \text{with probability } \pi \\ C_d \equiv \text{Max}[dP_0 - K, 0] & \text{with probability } 1 - \pi. \end{cases} \tag{3}$$

Given the simple structure that has been assumed so far, can one uniquely determine the date-0 option price C_0? It seems unlikely, as we have said nothing about investors' preferences nor the supply of the security. Yet C_0 is indeed completely and uniquely determined and is a function of K, r, P_0, d, and u. Surprisingly, C_0 is not a function of π!

To see how and why, consider constructing a portfolio of Δ shares of stock and $\$B$ of bonds at date 0, at a total cost of $X_0 = P_0\Delta + B$. The payoff X_1 of this portfolio at date 1 is simply:

$$X_1 = \begin{cases} uP_0\Delta + rB & \text{with probability } \pi \\ dP_0\Delta + rB & \text{with probability } 1 - \pi. \end{cases} \tag{4}$$

Now choose Δ and B so that the following two linear equations are satisfied simultaneously:

$$uP_0\Delta + rB = C_u, \qquad dP_0\Delta + rB = C_d \tag{5}$$

which is always feasible as long as the two equations are linearly independent. This is assured if $u \neq d$, in which case we have

$$\Delta^* = \frac{C_u - C_d}{(u - d)P_0}, \qquad B^* = \frac{uC_d - dC_u}{(u - d)r}. \tag{6}$$

Because the portfolio payoff X_1 under (6) is identical to the payoff of the call option C_1 in both states, the total cost X_0 of the portfolio must equal the option price C_0; otherwise, it is possible to construct an *arbitrage*, a trading strategy that yields riskless profits. For example, suppose that $X_0 > C_0$. By purchasing the option and selling the portfolio at date 0, a cash inflow of $X_0 - C_0$ is generated, and at date 1 the obligation X_1 created by the sale of the portfolio is exactly offset by the payoff of the option C_1. A similar argument rules out the case where $X_0 < C_0$. Thus the following pricing equation holds:

$$C_0 = P_0\Delta^* + B^* = \frac{1}{r}\left[\left(\frac{r-d}{u-d}\right)C_u + \left(\frac{u-r}{u-d}\right)C_d\right] \tag{7}$$

$$= \frac{1}{r}\left[\pi^* C_u + (1 - \pi^*)C_d\right], \qquad \pi^* \equiv \frac{r-d}{u-d}. \tag{8}$$

This pricing equation is remarkable in several respects. First, it does not seem to depend on investors' attitudes toward risk, but merely requires that investors prefer more money to less (in which case arbitrage opportunities are ruled out). Second, nowhere in (8) does the probability π appear, which implies that two investors with very different opinions about π will nevertheless agree on the price C_0 of the option. Finally, (8) shows that C_0 can be viewed as an expected present value of the option's payoff, but where the expectation is computed not with respect to the original probability π, but with respect to a "pseudoprobability" π^*, often called a *risk-neutral* probability or *equivalent martingale measure*. [Contrast (8) with the pricing equation (1) in which the discount factors $\gamma_{t,t+k}$ are also present.]

That π^* is a probability is not immediately apparent and requires further argument. A necessary and sufficient condition for $\pi^* \in [0, 1]$ is the inequality $d \leq r \leq u$. But this inequality follows from the assumption of the coexistence of stocks and riskless bonds in our economy. Suppose, for example, that $r < d \leq u$; in this case, no investor will hold bonds, because even in the worst case, stocks will yield a higher return than r. Hence bonds cannot exist; that is, they will have zero price. Alternatively, if $d \leq u < r$, then no investor will hold stocks, and hence stocks cannot exist. Therefore, $d \leq r \leq u$ must hold, in which case π^* can be interpreted as a probability. The fact that the option price is determined not by the original probability π, but rather by the equivalent martingale measure π^*, is a deep and subtle insight that has led to an enormous body of research in which the theory of martingales plays an unexpectedly profound role in the pricing of complex financial securities.

In particular, Merton's (1973b) derivation of the celebrated Black–Scholes formula for the price of a call option makes use of the Itô calculus, a sophisticated theory of continuous-time stochastic processes based on Brownian motion. Perhaps the most important insight of Merton's (1973b) seminal paper—for which he shared the Nobel prize in economics with Myron Scholes—is the fact that under certain conditions, the frequent trading of a small number of long-lived securities (stocks and riskless bonds) can create new investment opportunities (options and other derivative securities) that otherwise would be unavailable to investors. These conditions—now known collectively as *dynamic spanning* or *dynamically complete markets*—and the corresponding financial models on which they are based have generated a rich literature and a multitrillion-dollar derivatives industry in which exotic financial securities such as caps, collars, swaptions, and knock-out and rainbow options are synthetically replicated by sophisticated trading strategies involving considerably simpler securities.

This framework has also led to a number of statistical applications. Perhaps the most obvious is the estimation of the parameters of Itô processes that are the inputs to derivative pricing formulas. This task is complicated by the fact that Itô processes are continuous-time processes, whereas the data are discretely sampled. The most obvious method, maximum likelihood estimation, is practical for only a handful of Itô processes—those for which the conditional density functions are available in closed form; for example, processes with linear drift and diffusion coefficients. In most other cases, the conditional density cannot be obtained analytically but can only be characterized implicitly as the solution to a particular partial differential equation, the Fokker–Planck or "forward" equation (see Lo 1988 for further discussion). Therefore, other alternatives have been developed, including generalized method-of-moments estimators (Hansen and Scheinkman 1995), simulation estimators (Duffie and Singleton 1993), and nonparametric estimators (Aït-Sahalia 1996).

Because the prices of options and most other derivative securities can be expressed as expected values with respect to the risk-neutral measure [as in (8)], efficient Monte Carlo methods have also been developed for computing the prices of these securities (see Boyle, Broadie, and Glasserman 1997 for an excellent review). Moreover, option prices contain an enormous amount of information about the statistical properties of stock prices and the preferences of investors, and several methods have been developed recently to extract such information parametrically and nonparametrically (e.g., Aït-Sahalia and Lo 1998, 2000; Jackwerth and Rubinstein 1996; Longstaff 1995; Rubinstein 1994; Shimko 1993).

Finally, the use of continuous-time stochastic processes in modeling financial markets has led, directly and indirectly, to a number of statistical applications in which functional central limit theory and the notion of *weak convergence* (see, e.g., Billingsley 1968) are used to deduce the asymptotic properties of various estimators, such as long-horizon return regressions (Richardson and Stock 1989), long-range dependence in stock returns (Lo 1991), and the approximation errors of continuous-time dynamic hedging strategies (Bertsimas, Kogan, and Lo 2000).

5. CONCLUSIONS

The three ideas described here should convince even the most hardened skeptic that finance and statistics have much in common. There are, however, many other examples in which statistics has become indispensable to financial analysis (see Campbell, Lo, and MacKinlay 1997 and Lo and MacKinlay 1999 for specific references and a more complete survey). Multivariate analysis, especially factor analysis and principal components analysis, are important aspects of mean-variance models of portfolio selection and performance attribution. Entropy and other information-theoretic concepts have been used to construct portfolios with certain asymptotic optimality properties. Nonparametric methods such as kernel regression, local smoothing, and bootstrap resampling algorithms are now commonplace in estimating and evaluating many financial models, most of which are highly nonlinear and based on large datasets. Neural networks, wavelets, support vector machines, and other nonlinear time series models have also been applied to financial forecasting and risk management. There is renewed interest in the foundations of probability theory and notions of subjective probability because of mounting psychological evidence regarding behavioral biases in individual decisions involving financial risks and rewards. And Bayesian analysis has made inroads into virtually all aspects of financial modeling, especially with the advent of computational techniques such as Markov chain Monte Carlo methods and the Gibbs sampler.

With these developments in mind, can there be any doubt that the intersection between finance and statistics will become even greater and more active over the next few decades, with both fields benefiting enormously from the association?

REFERENCES

Aït-Sahalia, Y. (1996), "Nonparametric Pricing of Interest Rate Derivative Securities," *Econometrica*, 64, 527–560.

Aït-Sahalia, Y., and Lo, A. (1998), "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices," *Journal of Finance*, 53, 499–548.

——— (2000), "Nonparametric Risk Management and Implied Risk Aversion," *Journal of Econometrics*, 94, 9–51.

Bates, D. (1996), "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options," *Review of Financial Studies*, 9, 69–107.

Bernstein, P. (1992), *Capital Ideas*, New York: Free Press.

Bertsimas, D., Kogan, L., and Lo, A. (2000), "When Is Time Continuous?," *Journal of Financial Economics*, 55, 173–204.

Billingsley, P. (1968), *Convergence of Probability Measures*, New York: Wiley.

Black, F., and Scholes, M. (1973), "Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637–654.

Boyle, P., Broadie, M., and Glasserman, P. (1997), "Monte Carlo Methods for Security Pricing," *Journal of Economic Dynamics and Control*, 21, 1267–1321.

Breeden, D. (1979), "An Intertemporal Capital Pricing Model With Stochastic Investment Opportunities," *Journal of Financial Economics*, 7, 265–296.

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

Cowles, A., and Jones, H. (1937), "Some A Posteriori Probabilities in Stock Market Action," *Econometrica*, 5, 280–294.

Cox, J., Ross, S., and Rubinstein, M. (1979), "Option Pricing: A Simplified Approach," *Journal of Financial Economics*, 7, 229–264.

Duffie, D., and Singleton, K. (1993), "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61, 929–952.

Embrechts, P., Kluppelberg, C., and Mikosch, T. (1997), *Modelling Extremal Events for Insurance and Finance*, New York: Springer-Verlag.

Fama, E. (1965), "Portfolio Analysis in a Stable Paretian Market," *Management Science*, 11, 404–419.

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

Fama, E., and French, K. (1988), "Permanent And Temporary Components of Stock Prices," *Journal of Political Economy*, 96, 246–273.

Fama, E., and Roll, R. (1971), "Parameter Estimates for Symmetric Stable Distributions," *Journal of the American Statistical Association*, 66, 331–338.

Gallant, R., Hsieh, D., and Tauchen, G. (1997), "Estimation of Stochastic Volatility Models With Diagnostics," *Journal of Econometrics*, 81, 159–92.

Gray, S. (1996), "Modeling the Conditional Distribution of Interest Rates

as a Regime-Switching Process," *Journal of Financial Economics*, 42, 27–62.
Hansen, L., and Scheinkman, J. (1995), "Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes," *Econometrica*, 63, 767–804.
Hsieh, D. (1991), "Chaos and Nonlinear Dynamics: Application to Financial Markets," *Journal of Finance*, 46, 1839–1877.
Hurst, H. (1951), "Long-Term Storage Capacity of Reservoirs," *Transactions of the American Society of Civil Engineers*, 116, 770–799.
Jackwerth, J., and Rubinstein, M. (1996), "Recovering Probability Distributions From Contemporary Security Prices," *Journal of Finance*, 51, 1611–1631.
Jansen, D., and de Vries, C. (1991), "On the Frequency of Large Stock Returns: Putting Booms and Busts Into Perspective," *Review of Economics and Statistics*, 73, 18–24.
Kim, M., Nelson, C., and Startz, R. (1991), "Mean Reversion In Stock Prices? A Reappraisal of the Empirical Evidence," *Review of Economic Studies*, 58, 515–528.
Lévy, P. (1937), *Théorie de l'Addition des Variables Aléatoires*, Paris: Gauthier-Villars.
Lo, A. (1988), "Maximum Likelihood Estimation of Generalized Itô Processes With Discretely Sampled Data," *Econometric Theory*, 4, 231–247.
——— (1991), "Long-Term Memory in Stock Market Prices," *Econometrica*, 59, 1279–1313.
——— (Ed.) (1997), *Market Efficiency: Stock Market Behaviour in Theory and Practice*, Vols. I and II, Cheltenham, U.K.: Edward Elgar.
——— (1999), "The Three P's of Total Risk Management," *Financial Analysts Journal*, 55, 13–26.
Lo, A., and MacKinlay, C. (1988), "Stock Market Prices do not Follow Random Walks: Evidence From a Simple Specification Test," *Review of Financial Studies*, 1, 41–66.
——— (1999), *A Non-Random Walk Down Wall Street*, Princeton, NJ: Princeton University Press.
Longstaff, F. (1995), "Option Pricing and the Martingale Restriction," *Review of Financial Studies*, 8, 1091–1124.
Lucas, R. (1978), "Asset Prices in an Exchange Economy," *Econometrica*, 46, 1429–1446.
Mandelbrot, B. (1960), "The Pareto-Lévy Law and the Distribution of Income," *International Economic Review*, 1, 79–106.
——— (1963), "The Variation of Certain Speculative Prices," *Journal of Business*, 36, 394–419.
——— (1971), "When Can Price be Arbitraged Efficiently? A Limit to the Validity of the Random Walk and Martingale Models," *Review of Economics and Statistics*, 53, 225–236.
——— (1997), *Fractals and Scaling in Finance*, New York: Springer-Verlag.
Mantegna, R., and Stanley, E. (1999), *Scaling Approach to Finance*, Cambridge, U.K.: Cambridge University Press.
Markowitz, H. (1952), "Portfolio Selection," *Journal of Finance*, 7, 77–91.
McCulloch, H. (1996), "Financial Applications of Stable Distributions," in *Handbook of Statistics, Volume 14: Statistical Methods in Finance*, eds. G. Maddala and C. Rao, Amsterdam: Elsevier Science.
Merton, R. (1973a), "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41, 867–887.
——— (1973b), "Rational Theory of Option Pricing," *Bell Journal of Economics and Management Science*, 4, 141–183.
——— (1992), *Continuous-Time Finance* (rev. ed.), Oxford, U.K.: Basil Blackwell.
Pilgram, B., and Kaplan, D. (1998), "A Comparison of Estimators for $1/f$ Noise," *Physica D*, 114, 108.
Poterba, J., and Summers, L. (1988), "Mean Reversion in Stock Returns: Evidence and Implications," *Journal of Financial Economics*, 22, 27–60.
Richardson, M. (1993), "Temporary Components of Stock Prices: A Skeptic's View," *Journal of Business and Economics Statistics*, 11, 199–207.
Richardson, M., and Stock, J. (1990), "Drawing Inferences From Statistics Based on Multiyear Asset Returns," *Journal of Financial Economics*, 25, 323–348.
Ross, S. (1976), "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13, 341–360.
Rubinstein, M. (1994), "Implied Binomial Trees," *Journal of Finance*, 49, 771–818.
Samuelson, P. (1965), "Proof that Properly Anticipated Prices Fluctuate Randomly," *Industrial Management Review*, 6, 41–49.
——— (1967), "Efficient Portfolio Selection for Pareto–Lévy Investments," *Journal of Financial and Quantitative Analysis*, 2, 107–122.
Sharpe, W. (1964), "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *Journal of Finance*, 19, 425–442.
Shimko, D. (1993), "Bounds of Probability," *RISK*, 6, 33–37.

Statistics and Marketing

Peter E. Rossi and Greg M. Allenby

Statistical research in marketing is heavily influenced by the availability of different types of data. The last 10 years have seen an explosion in the amount and variety of data available to market researchers. Demand data from scanning equipment have now become routinely available in the packaged goods industries. Data from e-commerce and direct marketing are growing at an exponential rate and provide coverage to a wide assortment of different products. Web-based technology has dramatically lowered the cost of survey research. Web-browsing data provide an important new source of information about consumer tastes and preferences, which is becoming available for a large fraction of the total consumer population. In this vignette we explore some of the implications of this data explosion for the development of statistical methodology in marketing, with primary emphasis on the explosion in demand data.

Scanning equipment has provided the market researcher with a national panel of stores in addition to panels of households, altering the focus of marketing research. These data have stimulated a large literature on applied demand and discrete choice modeling. Demand models at the store level typically take the form of multivariate regression models in which demand for a vector of products is related to marketing variables such as prices, displays, and various forms of advertising. At the household level, demand is

Peter E. Rossi is Joseph T. Lewis Professor of Marketing and Statistics, Graduate School of Business, University of Chicago, IL 60637 (E-mail: *peter.rossi@gsb.uchicago.edu*). Greg M. Allenby is Helen C. Kurtz Chair in Marketing and Professor of Marketing and Statistics, Fisher College of Business, Ohio State University, Columbus, OH 43210 (E-mail: *allenby.1@osu.edu*).

Journal of the American Statistical Association
June 2000, Vol. 95, No. 450, Vignettes