THE THEORY OF SECURITY PRICING AND MARKET STRUCTURE

by

Marshall E. Blume
Jeremy J. Siegel

20-91
(revision of 39-89)

RODNEY L. WHITE CENTER FOR FINANCIAL RESEARCH
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104-6367

First Draft
August 15, 1989
Current Revison
August 22, 1991

THE THEORY OF SECURITY PRICING AND MARKET STRUCTURE

by

Marshall E. Blume
Jeremy J. Siegel

The Rodney L. White Center for Financial Research
The Wharton School

We thank Robert Ariel, Myron Gordon, Allan Kleidon, Harry Markowitz, David Mayer, Haim Mendelson, Junius Peake, Marc Reinganum, Jay Ritter, Michael Rozeff, Daniel Seiver, H. Nejat Seyhun, and William Sharpe for comments on an earlier version of this manuscript. We especially thank Fischer Black, Paul Samuelson, and Robert Shiller for their extensive comments which have pointed the way towards substantial improvements of this manuscript; Suzanne Barrett and Todd Rosentover for their research assistance; Elizabeth Schmidt who coordinated the production of this manuscript; and the New York Stock Exchange for a grant in support of this study.

ABSTRACT

This manuscript reviews the evolving literature on the pricing of assets and the structure of financial markets. It begins with the formulation of early stock valuation models and proceeds to a description of the efficient-market hypothesis. The survey then examines recent empirical data which led to a reevaluation of the assumptions underlying an efficient market and ends with a description of the recent literature on market-making and how these models relate to the studies of efficient markets.

I. INTRODUCTION

A market in financial assets serves two major functions. The first is price determination--signaling investors and firms where the most profitable investments can be made and at what price. An increase in the price of a security is a signal to investors that the market values a particular productive activity more favorably, making it cheaper for the firm's owners to finance investment. Thus, prices of financial assets serve a similar function as prices in the standard economic models of supply and demand: allocating scarce resources to their best uses.[1]

The second function of the market in financial assets is to bring together buyers and sellers of existing securities and to enable firms to raise new capital by issuing new security to existing investors. The structure of this market, defined by such factors as the number of separate market makers, trading rules, dealer involvement, and the auction mechanism, determines the effectiveness of this second function of market-making.

Until recently, theoretical models assumed that the determination of price was independent of the specific market structure in which securities were traded. Prices were assumed to be functions of such variables as dividends, earnings, interest rates, and risk factors. The specific institutional structure of the market was assumed to have no effect on the price of a security. Recently a growing body of literature suggests that the institutional structure of the market can play a significant role in determining the price of a security.

---

[1]One should be mindful that the establishment of monopolies, foreign trade restrictions, and other restrictive practices can enhance the price of a security without increasing welfare. Similarly, an oil crisis may depress the price of existing shares while creating tremendous opportunities for new energy saving investment.

This research suggests that price determination must be analyzed in conjunction with the market structure, with different structures leading to different prices.

It is the purpose of this survey to review the evolving theory of the pricing of assets and the structure of the financial markets. Section II reviews the early theory of the pricing of risky assets known as fundamental analysis and then reviews the technical analysis or charting techniques which tried to find predictable dependencies in stock price changes.

The early tests on technical analysis found that price movements of stocks and bonds are essentially random. This implied that charting techniques, which relied on non-random prices, were useless. The apparent random character of changes in stock prices led to the development of the efficient-market hypothesis, reviewed in Section III. This hypothesis claimed that if all known pricing factors, such as earnings or past price patterns, are included in the price of a security, then the market price must generally follow a process that was to become known as a "random walk."

Further tests bore out the random character of changes in stock prices, which validated the efficient-market hypothesis. Few investors, even those who managed large investment pools such as mutual funds, outperformed the returns available from an investment in a random collection of securities. Coupled with the theory of efficient markets, these findings supported the case for investing in large diversified baskets of stocks which matched some broad based index, since it was difficult, if not impossible, to try to "beat the market." At the same time that researchers were carrying out these empirical tests of market efficiency, there were major theoretical advances in the theory of asset pricing leading to the development of the well-known capital asset pricing model (CAPM) used by both academics and practitioners.

Almost as soon as the evidence for an efficient market appeared to be overwhelming, there emerged a series of empirical results that were inconsistent with the predictions of the efficient-market hypothesis as it was understood at that time. These studies, reviewed in Section IV, showed that over long periods of time the returns of small firms significantly exceeded those of larger firms by more than could be explained by pricing theories developed under the efficient-market hypothesis. Many of these excess returns were perplexingly related to the calendar, mostly occurring early in the month of January. The further discovery of relations between price movements and the day of the week and even the time of the day provided further challenges to the efficient-market hypothesis.

There also emerged further evidence that was inconsistent with the efficient-market hypothesis. Some advisers appeared consistently to outperform the market and "insiders" earned significantly higher returns than predicted by the efficient-market hypothesis. Finally, studies of stock market returns over long time periods suggested that the stock market was excessively volatile relative to the fluctuations that the original versions of the efficient-market theory would predict. As a result, researchers tried to modify the existing asset pricing theory to add additional realism, but these modified models failed to provide adequate explanation for the observed pricing anomalies.

The empirical evidence against the efficient market caused a reexamination of the basic theoretical model underlying the hypothesis (Section V). This reexamination found a logical inconsistency in the efficient-market hypothesis itself. If the prices in the markets already reflected all the information about the security, as the efficient-market hypothesis claimed, there would never be any trading among individuals as a result of new information, since the prices of securities would adjust instantaneously to any such new information. Yet, if no trading took place, there would be no incentive to collect information on any security, since there would be no opportunity to profit in the absence of trading. If no information is produced, it is impossible for prices to reflect all information, so the very concept of an efficient financial market collapses.

In short, the "paradox" of an efficient market is that a market can only be efficient if some people think it isn't and invest resources in obtaining information. Although this observation may be self-evident to those who operate in financial markets, it was a challenge to theoreticians who modelled the prices of financial securities.

The new models of market pricing which arose from these insights included not only informed traders, who set prices in the early models of an efficient market, but a second group of investors who traded for reasons unrelated to new information. These latter traders became known as "liquidity" traders and helped provide a solution to the logical inconsistencies inherent in the efficient-markets hypothesis. The "liquidity" traders provided added "noise" to the trading of informed investors so that the contraparty in a trade never knew with certainty whether the initiator of a trade had new information or not.

The introduction of noise traders also influences the volatility of asset prices. Some models showed that noise traders could cause the prices of securities to deviate for long periods of time from "fundamental" or full information values. Furthermore, security prices could be a function of the information traders believe (but do not know for certain) other traders possess, a concept which harkened back to John Maynard Keynes' insights into speculative markets.

The introduction of liquidity traders also led to a reexamination of the nature of market making itself, reviewed in Sections VI and VII. The institutional structure of a financial market interacts with the demands of both informed and liquidity traders to determine the bids, offers, and prices of securities. The efficiency of the market in discovering prices depends not only upon the costs and competition inherent in market making, but also upon the types of information possessed by different traders.

New models showed that security prices are a function of the type of market and rules under which they are traded. The prices of securities traded in call markets, where orders are aggregated and transacted at a single point in time, might be different from the prices of the same securities traded in the more common continuous markets, characterizing most of the world's stock exchanges. The implication of this work is that the pricing function of the capital markets, which is of critical importance to the allocation of investment, could not be separated from the institutional structure of the market itself.

II. THE EARLY LITERATURE ON ASSET PRICING

A. Fundamental Analysis

Early writers on the subject of security analysis assumed that the essence of investing was to determine the "true," "intrinsic," or "fundamental" value of a security and that this value may differ from the current market price.[2] In 1934, Benjamin Graham and David L. Dodd first published *Security Analysis*,[3] which became a classic in investment analysis. In that work they asserted:

> [a security analyst] is concerned with the intrinsic value of the security and more particularly with the discovery of discrepancies between the intrinsic value and the market price. We must recognize, however, that intrinsic value is an elusive concept. In general terms it is understood to be that value which is justified by the facts, e.g., the assets earnings, dividends, definite prospects, as distinct, let us say, from market quotations established by artificial manipulation or distorted by psychological excesses.[4]

According to their view, the intrinsic value of a security is a function of the future earnings, or "earning power" of a company, rather than "book value," a concept used by earlier analysts. Graham and Dodd maintained that, while the market price of a security can deviate on occasion from its intrinsic value, over time it will move back to this value.[5]

Implicit in their approach to the evaluation of securities are the assumptions that some investors have better information than others and those investors with better information can accumulate underpriced securities without a significant, and self-defeating, impact on the market price of the stock. Only as other investors learn the true value of the stock will the price adjust to its intrinsic value. Graham and Dodd did not provide any explicit model of

---

[2]At this point, there was no formal theory as to why the market price and the intrinsic price of a security might differ.

[3]Benjamin Graham and David L. Dodd, *Security Analysis: Principles and Technique* (New York: McGraw-Hill Book Company, Inc., 1934).

[4]Graham and Dodd, *op. cit.*, p. 17.

[5]J. B. Williams, *The Theory of Investment Value*, published in 1939, also emphasized the "fundamental" and "intrinsic" valuation of securities.

the mechanism by which a stock price will adjust to its intrinsic value; they merely assumed that such an adjustment will ultimately take place.

In the jargon of the current academic literature, Graham and Dodd implicitly assumed that investors have "heterogeneous expectations." Stated simply, all investors do not have the same opinions about the future prospects of the company and some may be better informed than others. This assumption turns out to be critical in understanding both the development of the efficient- market hypothesis and the recent literature on market making mechanisms.

B. The Dividend Discount Model

In the 1960s, there was substantial debate in the academic world of what determined intrinsic value. Of the numerous models of asset pricing developed during these years, that of Myron Gordon has emerged over time as the prototypical model.[6] For the purposes of valuing a company, Gordon assumed that a company will pay a per share dividend of size D one year from now and that this per share dividend will grow in the future at a constant compound annual rate of growth, g. The true price of the stock is the value of this stream of future dividends discounted at an appropriate discount rate, r, which is related to the market rate of interest and the stock's risk. The price of the stock, P, is thus the present value of the future dividend stream and is given by the simple formula

---

[6]Myron J. Gordon, *The Investment, Financing and the Valuation of the Corporation* (Homewood, IL: R. D. Irwin, 1962).

$$P = \frac{D}{r - g}.$$

This formula implies that the value of the stock increases if either the current level of dividends or the growth rate of dividends increases, and it declines if the discount rate increases. The above model accords with the common intuition about how dividends, growth, and discount rates should influence security prices.[7] Wall Street often calls the Gordon model and its many variations the "dividend discount model."

A major limitation of this model is that it offers no insight into the source of the variation in the discount rate among different securities. Conceptually, the appropriate discount rate is the sum of an interest rate on some safe investment plus a risk premium that is positively related to the risk of the stock. However, in the 1950s and early 1960s, the concept of risk and the associated risk premium was not well developed. Further insights into the measurement of risk and the associated risk premium came with the theoretical development of the Capital Asset Pricing Model (CAPM) in the 1960s, a model that relies heavily on the efficient-market hypothesis and is described in Section III.D. below.

The type of analysis espoused by Graham and Dodd as well as Gordon is generally termed "fundamental analysis". These models made explicit those variables, such as

---

[7]One of the most popular uses of this model is to equate the expected return of a security with the discount rate. To do this, one rearranges the model by solving for r. In this rearranged form, r is the sum of the dividend yield and the expected growth rate in dividends, both of which can be estimated.

earnings and dividends, that need to be forecast. Therefore, determining the intrinsic value of a stock was an analytical exercise involving forecasts of specific company variables.

C. Technical Analysis

A second approach to choosing securities is termed technical analysis or "charting".[8] Chartists plot the history of past prices (and often volume) and try to discern some predictive pattern for future price movements.

Chartists give various reasons why there might be predictive patterns in past prices. Patterns might arise as investors with superior information about the future level of earnings of a company begin to accumulate its shares. This buying generates increased volume and an increase in price. Some chartists may be able to detect this pattern before others and start accumulating the security. This will ultimately drive the price to higher levels, resulting in profits both for the chartists and to the original investor with superior information.

Chartists themselves never worry explicitly about whether the price of a security is correct in terms of the fundamentals of the company. However, if chartists react to the trading of informed investors, chartists may facilitate the adjustment of a stock price to its new equilibrium. As a result many chartists, who sometimes refer to themselves as "micro-market economists," believe that their techniques do rely implicitly on sound fundamental grounds.

---

[8]The original classic in the field was published by John McGee and Robert Edwards, *Technical Analysis of Stock Trends* (Stock Trend Service, 1954). A recent summary of technical analysis is contained in Sumner N. Levine, ed., *Financial Analyst's Handbook* (Homewood, IL: Dow Jones Irwin, Inc., 1988).

III. THE EFFICIENT-MARKET HYPOTHESIS

A. Early Studies

Paralleling the development of the theory of valuation were empirical studies that examined the statistical properties of commodities and financial assets. In 1934 Holbrook Working published an examination of the statistical properties of wheat and found the changes to be essentially random.[9] But the significant breakthrough came in London in 1953, when the Royal Statistical Society met to examine a paper by the statistician Maurice Kendall about the behavior of speculative prices.

Kendall exhaustively examined the weekly price behavior of British stock prices and American commodity prices.[10] He concluded that *changes* in stock and commodity prices conform to a chance process, as if determined by the turn of a roulette wheel. The *level* of stock and commodity prices are simply the summation of these changes in price.

Harry Roberts[11] explored this difference between price levels and price changes through an experiment. He simulated random price changes for 52 weeks and then summed

---

[9]Holbrook Working, "A Random-Difference Series for Use in the Analysis of Time Series," *Journal of the American Statistical Association*, 29 (March 1934), 11-24.

[10]Maurice G. Kendall, "The Analysis of Economic Time Series. I", *Journal of the British Statistical Society* (Ser. A), 116 (1953), 11-25. What was particularly impressive about this work was the number of calculations that were undertaken without the benefit of modern computers.

[11]Harry V. Roberts, "Stock-Market 'Patterns' and Financial Analysis: Methodological Suggestions," *Journal of Finance*, 14:1 (March 1959), 1-10. At the same time, M. Osborne confirmed these results in "Brownian Motions in the Stock Market," *Operations Research*, 7:2 (March/April 1959), 145-73.

these price changes to obtain the price levels.[12] Although the price changes themselves in Roberts' simulation are perfectly random, the sum of these price changes, or the price levels themselves, appear to form patterns which chartists follow, such as "head-and-shoulder" formations, "flags," "pennants," "breakthroughs," and so on. However, because of their random construction, these patterns could not be used to predict future prices. After these studies, the academic literature began to reject the usefulness of technical analysis which relied on these patterns to predict future prices.

The term "random walk" refers to the process by which the sum of random changes generates the price level of a particular security or asset. Even before the empirical work cited above, theoretical literature had been published concerning the nature of random walk processes.[13] In a most prescient work, a French student named Louis Bachelier presented a Ph.D. dissertation in 1900 before the Faculty of Sciences of the Academy of Paris entitled *The Theory of Speculation.*[14] Bachelier described the random motion of speculative prices and derived the foundations of option theory. In his dissertation he wrote, "From the consideration of true prices, one could say: *At a given instant, the market believes in neither*

---

[12]The simplest type of random process can be simulated by moving prices up or down equal amounts depending on the flip of a fair coin. More complicated random behaviors allow for unequal probabilities of moving up or down and the existence of a trend. These more complex, but still random processes are called "martingales."

[13]It is believed that the term "random walk" originated in a note by Karl Pearson in a 1905 issue of *Nature* magazine, and was used to described the path of a drunk left to wander on an open field.

[14]The work was originally published in *Ann. Sci. Ecole Norm. Sup (3)*, No. 1018 (Paris: Gautheir-Villars, 1900) and dedicated to Monsier H. Poincare. It is found in its translated edition in Paul H. Cootner, ed., *The Random Character of Stock Market Prices* (Cambridge, MA: M.I.T. Press, 1964).

a rise nor a fall of true prices," and later, "Clearly the price considered most likely by the market is the current true price: if the market judges otherwise, it would quote not this price but another price higher or lower."[15] These sentences comprise the earliest known references to the workings of an efficient market. Bachelier was so ahead of his time that he recognized that the random walk process which security prices followed also possessed important properties that were to be rediscovered five years later by Albert Einstein in one of his early works on the Theory of Relativity.

Bachelier's thesis stood alone in the mathematical derivation of security prices for over sixty years until Paul Samuelson[16] and Eugene Fama[17] provided further theoretical frameworks that predicted the random walk behavior of stock prices.[18] The random properties of security returns were fundamental to the development of "the efficient-market hypothesis."[19]

---

[15]In Bachelier, *op. cit.*, reprinted in Paul H. Cootner, *The Random Character of Stock Market Prices* (Cambridge, MA: M.I.T. Press, 1964), pp. 26 and 28.

[16]Paul Samuelson, "Proof that Properly Anticipated Prices Fluctuate Randomly," *Industrial Management Review*, 6:2 (Spring 1965), 41-49.

[17]Eugene F. Fama, "The Behavior of Stock-Market Prices," *Journal of Business*, 38:1 (January 1965), 34-105.

[18]Paul Samuelson, "Proof That Properly Discounted Present Values of Assets Vibrate Randomly," *Bell Journal of Economics and Management Science*, 4:2 (Autumn 1973), 369-374. Samuelson also emphasized that the random walk properties applied only to the prices on securities adjusted for dividends, and not to prices (or price indices) which exclude dividends or interest. In fact, over time the price of securities excluding the income return will follow a negative autocorrelated process which he called "red noise." These observations become important for investors examining long-term stock and bond indices which typically exclude reinvested dividends.

[19]Burton Malkiel popularized this characteristic of stock prices with his book *A Random Walk Down Wall Street* (New York: W. W. Norton & Company, 1973).

B. Definitions of Efficient Capital Market

An "Efficient Capital Market" is a market in which the prices of all securities reflect all available information about the assets. This definition requires some elaboration. First, in order for the price to reflect all the information about the asset, there can be nothing which impedes the purchase and sale of securities, such as brokerage fees, taxes, etc. To the extent that impediments exist to the trading of assets, the prices will only imperfectly reflect information of relevance to the valuation of the securities.

Second, the phrase "all available information" needs to be explained. Does it mean *all* information about every security, whether possessed by insiders or not, or just publicly held information? These issues will be discussed in the next section.

Third, to say that "prices reflect" information implicitly assumes that one has in mind a specific pricing model of how information about risk and return should be incorporated into the prices of securities. One candidate for a pricing model is the dividend-discount model already discussed; another candidate, discussed in Section D below, is the Capital Asset Pricing Model (CAPM) which takes explicit account of the riskiness of individual assets. Ultimately, most tests of the efficient-market hypothesis are really joint tests of the validity of a specific pricing model and the correspondence of market prices to those implied by that pricing model.

In appreciating the more recent literature on market-making, it is important to note that this early literature on efficient capital markets assumed that the market prices incorporated new information instantaneously into stock prices. Casual observation suggests that it may take time for market prices to adjust to new information, but the assumption in this earlier literature is that this time is so short that there is no need to model the precise way in which this adjustment takes place. The more recent literature on market-making is directed towards the process by which new information is incorporated into stock prices. Specifically, the structure of a market itself--the technology and rules relating the activities of dealers and investors--may lead to different paths of prices. This new literature implies that the concept of an efficient market involves not only a pricing model but also the market structure itself, a concept which is elaborated in Sections VI and VII.

C. Levels of Market Efficiency

Since prices reflect all available information in an efficient market, the only reason for prices to change in such a market (excepting "fair" return consistent with holding the asset) is the arrival of *un*anticipated new information. Since such information is as likely to be favorable as unfavorable, the probability of a movement of price upward is as likely as downward.[20] Changes in the price of the security will therefore be unpredictable, and prices themselves will follow a "random walk." This is why the efficient-market hypothesis is so closely tied to the random walk theory of security prices.

---

[20]In a more general sense, the probability of a favorable or unfavorable news event weighted by their relative importance to the pricing of the security must be equal. Therefore, it is possible that there could be a large number of small favorable events counterbalanced with a smaller number of more unfavorable events. Prices of securities in these circumstances will follow a generalized version of a random walk called a "martingale.

In an unpublished paper presented to the Seminar on the Analysis of Security Prices at the University of Chicago in 1967, Harry V. Roberts defined three levels of market efficiency.[21]

C.1. Weak Form Efficiency

A market is defined as "efficient in the weak sense" if the current and past prices (and perhaps volume of trading) do not provide meaningful forecasts of future changes in prices. This definition is usually shortened to: A market is *weakly efficient* if the current price reflects all past price and volume information.

If capital markets are efficient in the weak sense, there are no dependencies in past price changes that a chartist could use to predict future changes. An indirect proof follows: Assume there does exist a positive run in prices, which implies that the next price change is more likely to be positive than negative. If such runs exist, investors will try to buy as soon as they see any evidence of a positive run. As investors attempt to buy the stock, the buy orders of these investors will cause a jump in price, the run will stop, and these investors will gain no further advantage. According to the efficient-market hypothesis, this process takes place so quickly that it would be impossible to exploit the prior run in prices. The price of the security adjusts to a new level instantaneously and thereby eliminates all

---

[21]H. V. Roberts, "Statistical Versus Clinical Prediction of the Stock Market," University of Chicago, May, 1967. Eugene Fama made Roberts's arguments more rigorous in "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance*, 25:2 (May 1970), 383-417.

statistical price dependencies. As a result, forecasting future price changes from past price changes becomes impossible, and the level of the stock price follows a random walk.

Early statistical tests, such as those of Alfred Cowles in the 1930s, provided strong and consistent support for this hypothesis.[22] Later M. Osborne, Eugene Fama, and others[23] confirmed the random, or Brownian motion of stock prices.[24] In the mid 1960s, Fama and Blume[25] examined the profits from a trading strategy that relied on the theory of relative strength and trading trends. Their tests, and studies that followed, found that there were no abnormal returns from this type of trading.

C.2. Semi-Strong Efficiency

Weak form efficiency was widely accepted by the financial community. Since most analysts relied on fundamental research, they had little sympathy with chartists and others who disdained the use of earnings and dividends and used price patterns to forecast stock prices.

---

[22]A. Cowles and H. Jones, "Some posteriori Probabilities in Stock Market Action," *Econometrica*, 5:3 (July 1937), 280-294.

[23]M. Osborne, "Brownian Motions in the Stock Market," *Operations Research*, 7:2 (March/April 1959), 145-173, and E. Fama, "The Behavior of Stock Prices, *Journal of Business*, 38:1 (January 1965), 34-105.

[24]There were two major types of tests used in estimating the degree of dependence among successive price changes and hence the random nature of security prices. One was correlation tests, and the other was runs tests. The run tests counted the number of runs of positive changes, of zero price changes, and of negative price changes and compared these numbers to those expected under the random walk hypothesis.

[25]Eugene F. Fama and Marshall E. Blume, "Filter Rules and Stock-Market Trading," *Journal of Business*, 39:1 part II (January 1966), 226-241.

After examining whether past prices could be used to predict the future, it was a natural step to study whether other types of information are discounted in security prices. A market is defined as "efficient in the *semi-strong* sense" if the current price of a stock reflects all *publicly* available information. If this is so, such information has no value in forecasting future price changes, since it is already discounted in the price. It is not necessary that everybody has the same information, but only that the price reflects all publicly available information, even if the trading is confined only to a few astute traders.

Semi-strong efficiency can be tested in two ways. First, researchers evaluated whether market professionals, such as investment managers who run private accounts or mutual funds tend to outperform the market. Again, Alfred Cowles was a pioneer in this field, concluding in a 1933 article entitled, "Can Stock Market Forecasters Forecast?" that

> A review of the various statistical tests, applied to the records for this period, of these 24 forecasters, indicates that the most successful records are little, if any, better than what might be expected to result from pure chance."[26]

Later, a 1962 Wharton study of mutual funds concluded that the performance of equity mutual funds was on average no better than randomly selected groups of stocks.[27] Following this study was a number of other studies of mutual funds, all reaching similar

---

[26]A. Cowles, "Can Stock Market Forecasters Forecast?," *Econometrica*, 1:3 (July 1933), 309-324.

[27]F. E. Brown, Irwin Friend, Edward S. Herman, and Douglas Vickers, "A Study of Mutual Funds," *Report of the Committee on Interstate and Foreign Commerce* (Washington: U. S. Government Printing Office, 1962).

conclusions.[28] These studies persuaded many academics that the market for equities was efficient in the *semi-strong* sense. The managers of mutual funds with all their resources for analyzing individual companies were unable, on average, to outperform randomly selected portfolios of stocks.

A second group of empirical tests of semi-strong efficient markets revolved around the speed of adjustment of market prices to new information. These tests examined whether the return to investors of holding stocks after the release of public information was significantly greater than the market taken as a whole. The events studied included stock splits, merger announcements, secondary offerings, and many other actions that might revise expectations.[29] These findings confirmed that stock prices rapidly incorporated new information and that no extraordinary return could be earned by investors acting on publicly available information.[30]

---

[28]Michael Jensen, "The Performance of Mutual Funds in the Period 1945-64," *Journal of Finance*, 23:2 (May 1968), 389-416; William F. Sharpe, "Mutual Fund Performance," *Journal of Business*, 39:1 part II (January 1966), 119-138; and Irwin Friend, Marshall Blume and Jean Crockett, *Mutual Funds and Other Institutional Investors* (New York: McGraw-Hill Book Company, 1970).

[29]E. Fama, L. Fisher, M. Jensen and R. Roll, "The Adjustment of Stock Prices to New Information, *International Economic Review*, 10:1 (February 1969), 1-21; P. Dodd, "The Effect on Market Value of Transactions in the Market for Corporate Control," *Proceedings of Seminar on the Analysis of Security Prices, CRSP* (University of Chicago, May 1981); and Myron S. Scholes, "The Market for Securities: Substitution versus Price Pressure and the Effects of Information on Share Prices," *Journal of Business*, 45:2 (April 1972), 179-211.

[30]The majority of Justices on the Supreme Court has recently endorsed the use of semi-strong definition of efficiency in legal cases by accepting the argument that the price of a stock reflects all relevant publicly available information, including any misleading information released by a company or an insider. Consequently, if there were such misleading information, an investor who purchased or sold such a stock could be damaged and hold the provider of the misleading information liable, even though the investor did not

Although convincing to many scholars, these academic studies of price dependencies, trading rules, and mutual funds had only marginal impact on the practice of investing money in the 1960s. However, combined with the many performance studies of institutional money managers,[31] these studies finally convinced many market practitioners that it is difficult to outperform the market. As a consequence, it is not surprising that institutional investors currently have over 10 percent of all domestic equities invested in index funds.[32]

The implications of a semi-strong efficient market upon investor behavior are profound. Since prices reflect all relevant publicly available information, investors should realize that all of the different views of the prospects of a firm are already incorporated into the price, thereby eliminating any potential abnormal profits from investing in any particular security. An efficient market should discourage investors from security analysis, which merely provides information that other investors already have. Since market prices already incorporate this information, gaining access to it will be of no value. Thus, even if some investors learned information that was new to them, these investors should not change their view as to the correct price of any security or make any change to their portfolios.

---

rely upon or even know about the misleading information. As long as some investors were aware of the information, the stock price would reflect this information, and any investor basing an investment decision on the market price would implicitly be relying on the misleading information. This theory of damages is termed "fraud on the market place." [Basic Incorporated vs Levinson, 99 L. Ed. 2d 194 (1988).]

[31]A. G. Becker and Merrill Lynch were early providers of such studies. There are now many firms that undertake these studies, but SEI is now the dominant firm through its purchase of the performance evaluation business of both A. G. Becker and Merrill Lynch.

[32]*Pension and Investment Age* (July 24, 1989), p. 2. There may be additional equities that are managed in styles closely resembling index funds not included in the figure of $138 billion.

In a semi-strong efficient market, an investor requires access to non-public or inside information to determine that a market price is wrong. Security analysis, traditionally defined, does not utilize inside information and hence would not benefit an investor.

C.3. Strong Form Efficiency

A market is efficient in the *strong* sense if the current price of a stock reflects *all* information, whether the information is publicly available or not. The additional feature of strong efficiency is the inclusion of "inside information" into the price of a stock, through either insider trading or the revelation of such information to other traders.

Efficiency considerations alone might suggest the desirability of having the current market price of a stock reflect inside information. If the information were favorable, but not reflected in the price, a seller of the stock would be harmed. Similarly, if the information were unfavorable, but not reflected in the price, the buyer would be harmed. However, legal statutes, reflecting social policy in the United States as well as many other countries, dictate that it is unfair for insiders to profit from their inside information,[33] and laws and regulations have been enacted to restrain insider trading.

Early studies of insider trading showed that inside information was valuable in pricing securities. Jaffe found that insiders earn about six percent more per year, before transaction costs, than investors without special information.[34] Since insiders often trade for reasons

---

[33]The case for insider trading is made by Hayne Leland, "Insider Trading: Should It Be Prohibited?," mimeo, October 1990.

[34]Jeffrey Jaffe, "Special Information and Insider Trading," *Journal of Business*, 47:3 (July 1974), 410-428, is an early example of the study of insider trading. A more recent

not related to inside information, and since the data on insider trading does not distinguish between informationally motivated trading and other types of trading, the additional returns that insiders earn on informationally motivated trades are probably in excess of those reported in these studies.

This research opened the question of whether outsiders could use the insider trading reports that the SEC collects to make superior returns. Insiders are required to report their trading to the SEC within ten days of the end of the month in which they trade, producing significant delays between the actual insider trading and the reporting of these trades. With such delays, it would be surprising if these reports had any predictive value under a semi-strong version of the efficient market. Yet some subsequent studies of insider trading found that, even with a lag, outsiders could mimic insider trading to make additional returns.[35] While not as great as those earned by insiders, the additional returns may still be significant.

Studies of insider trading disproved market efficiency in the strong sense. However, these studies did not disturb some proponents of the efficient-market hypothesis. After all, the strong form of the efficient market is an extreme concept, much like a perfect vacuum.

---

study is H. Nejat Seyhun, "The Information Content of Aggregate Insider Trading," *Journal of Business*, 61:1 (January 1988), 1-24.

[35]Michael S. Rozeff and Mir A. Zaman, "Market Efficiency and Insider Trading: New Evidence," *Journal of Business*, 61:1 (January 1988), 25-44, argue that this finding that outsiders can make additional returns by mimicking insider trading is due to an improper definition of the normal rate of return from which additional return is measured. By using another definition of normal return, they conclude that outsiders cannot make additional returns.

Since the set of relevant information is theoretically infinite, it is unreasonable to expect that the market would literally incorporate all information into stock prices at every point in time.

D. Asset Pricing in an Efficient Market

One of the implications of an efficient market is that, under weak assumptions, the risky portion of the portfolio of every investor (without access to insider information) should be as diversified as possible.[36] This prescription for investing stems from the statistical property that total risk can be reduced, for a given expected return, by holding a large quantity of smaller risks. That diversification pays has often been summarized by the popular expression "Don't put all your eggs in one basket," which has strong theoretical and empirical support.[37]

In 1952 Harry Markowitz provided a formal model of diversification that identified efficient portfolios--portfolios which minimize risk at each level of expected return.[38] Markowitz argued that risk averse investors would only hold these efficient portfolios.

---

[36]One set of assumptions needed to make this statement correct is that the distribution of returns is not too skewed and that the returns of any non-marketable asset, such as human capital, are uncorrelated with the returns on marketable assets. If returns of non-marketable assets are correlated with the returns of marketable assets, investors would hold not only the market portfolio but an additional portfolio to hedge the risk of the non-marketable assets. See David Mayers, "Nonmarketable Assets and Capital Market Equilibrium Under Uncertainty," in Michael C. Jensen, ed., *Studies in the Theory of Capital Markets* (New York: Praeger Publishers, 1972).

[37]Harry M. Markowitz, *Portfolio Selection* (New York: John Wiley & Sons, Inc., 1959); Paul A. Samuelson, "General Proof that Diversification Pays," *Journal of Financial and Quantitative Analysis*, 2:1 (March 1967), 1-13; and J. Tobin, "Liquidity Preference as Behavior Towards Risk," *Review of Economic Studies*, 25:67 (February 1958), 65-86.

[38] H. Markowitz, "Portfolio Selection," *Journal of Finance*, 7:1 (March 1952), 77-91.

Markowitz's model showed that the demand for a risky asset depended not only upon the return characteristics of each security individually but also upon the interrelation among the return characteristics of all securities. This analysis gained Markowitz the 1990 Nobel Prize in Economics.

Markowitz demonstrated that securities with lower than average expected returns, and possibly even negative expected returns, would be included in a portfolio as long they provided sufficient diversification of risk. He also showed that risk which was uncorrelated with the market influenced the demand for the asset very little, especially if there were a large number of assets. His model was not totally at odds with traditional fundamental analysis, since asset demands were a function of the analyst's estimates of expected returns, but the treatment of risk was far more sophisticated (and less intuitive) than had been previously believed.

It was not long before the fundamentals of portfolio selection developed by Markowitz were applied to concepts of an efficient market. William Sharpe, who was awarded the Nobel Prize along with Markowitz, described the nature of portfolio selection under the assumption that all available information was already incorporated in prices.[39] Since in an efficient market security prices reflect such information, there is no reason to substitute one's own estimate of expected return and variance for that already implicit in the market. Sharpe demonstrated that if one accepted the market's judgment about the returns on each asset, then

---

[39]William Sharpe, "Capital Asset Prices: A Theory of Market Equilibrium under Condition of Risk," *Journal of Finance*, 19:3 (September 1964), 425-442. Sharpe began developing his analysis in a Ph.D. dissertation accepted at the University of California at Los Angeles in 1961. Jan Mossin further described the nature of the equilibrium in "Equilibrium in a Capital Asset Market," *Econometrica*, 34:4 (October 1966), 768-783.

the best portfolio of risky assets to hold is one that is totally diversified and where each security is held in proportion to its value in the market. What was surprising is that the above conclusions holds *no matter what the risk preferences of the investor*, as long as the investor was risk averse. If an investor tilted his portfolio towards any particular asset, it would increase the investor's risk without any compensating increase in expected return.

The theory of portfolio selection combined with that of an efficient capital market was a powerful incentive for the development of fully diversified portfolios, or "index funds." These funds hold assets in proportion to their market value and are managed at low cost and without any regard to the fundamental factors influencing security prices, assuming that such information is already reflected in prices.

Total risk could be controlled by moving between such an index fund and "risk-free" assets, or short-term money market instruments. Investors with a high risk tolerance would hold a greater fraction of their wealth in the index fund (perhaps even borrowing funds) while more risk averse individuals would hold a smaller fraction of wealth in such a fund. Under no circumstances would it be optimal for such individuals to overweight his or her portfolio with high or low risk *individual* stocks. By changing the proportion of risky and risk-free assets, the investor can adjust the overall risk level of his portfolio.[40]

Sharpe exploited the concept of diversification to develop a formal model to explain the differences in risk premiums of individual stocks, which came to be known as the Capital Asset Pricing Model or CAPM. The risk premium is defined as the difference between the

---

[40] This property became to be known as the "mutual fund separation theorem." The separation theorem was first analyzed by Tobin, *op. cit.*, in 1968.

expected return on a risky asset and the return on a safe asset, such as a treasury bill. Sharpe demonstrated that in an efficient market the risk premium on each stock would be proportional to the risk premium on the entire stock market (the expected return on the market minus the risk-free rate), where the constant of proportionality (called the *beta* coefficient of the stock) is related to the covariance of the individual stock's return with that of the market. Note that the beta coefficient, and hence the risk premium is only associated with the *co*variance of the asset's return with the market, not the variability of its individual return which can be diversified away in a large portfolio.

For the first time, the CAPM gave investors an operational way to determine the risk premium on individual stocks. By regressing the historical return of a stock on the return of the overall market, one can estimate the beta coefficient.[41] The premium on the entire market could also be obtained from historical studies of aggregate stock and bond returns. The popularity of beta as a means of measuring the risk of a portfolio rapidly spread to portfolio managers who themselves were increasingly in touch with academic research.

It was not long before researchers began to find some troubling discrepancies between the predictions of an efficient market and the empirical data. Low risk assets seemed to overperform relative to the predictions of the CAPM and high risk assets tended to underperform.[42] And, as Richard Roll strongly maintained, there were empricial

---

[41]Marshall E. Blume, "On the Assessment of Risk," *Journal of Finance*, 26:1 (March 1971), 1-10, showed that beta coefficients of portfolios of stocks were sufficiently stationary over time to make historically estimated beta coefficients a reliable predictor of future risk, suggesting that the CAPM may have some empirical content.

[42]Irwin Friend and Marshall E. Blume, "Measurement of Portfolio Performance Under Uncertainty," *American Economic Review*, 60:4 (September 1970), 561-575; Marshall E.

difficulties in any empirical test of the CAPM.[43] In theory, the market portfolio in the CAPM consists of all risky assets including real estate, bonds, and other non-equity assets. In practice, the empirical tests of the CAPM measured the return on the market portfolio of only a subset of risky assets--often proxying this market portfolio by a stock index such as the S&P 500. Roll showed that the use of an incomplete measure of the market portfolio resulted in mismeasuring the beta coefficients of individual assets and made tests of the CAPM meaningless. Despite these problems, by the end of the 1970s, the CAPM was the most sophisticated pricing mechanism available to portfolio managers and was widely used in investment analysis.

Blume and Irwin Friend, "A New Look at the Capital Asset Pricing Model," *Journal of Finance*, 28:1 (March 1973), 19-33; E. F. Fama and J. D. MacBeth, "Risk Return and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81:3 (May/June 1973), 607-636; Fischer Black, Michael Jensen, and Myron Scholes, "The Capital Asset Pricing Model: Some Empirical Tests," in Michael Jensen, ed., *Studies in the Theory of Capital Markets* (New York: Praeger Publishers, 1972).

[43]Richard Roll, "A Critique of the Asset Pricing Theory's Tests: Part 1: On Past and Potential Testability of the Theory," *Journal of Financial Economics*, 4:2 (March 1977), 129-176.

IV. EMPIRICAL CONTRADICTIONS TO THE EFFICIENT MARKET

Despite the large body of evidence in support of the early version of the efficient-market hypothesis, a growing number of studies raised questions about its validity.

A. Value Line

In 1973, Fischer Black published an article entitled "Yes, Virginia, There is Hope: Tests of the Value Line Ranking System".[44] This paper presented empirical evidence that the security recommendations of Value Line had some value in forecasting future prices. Since Value Line's recommendations were widely circulated, this finding violated the semi-strong version of the efficient market. Subsequent studies of the Value Line recommendations have reached similar conclusions.[45]

The prediction record of Value Line is impressive. From April 16, 1965 through December 30, 1990, the stocks in their top-rated "Group 1" realized compound annual capital gains of 14.1 percent. In contrast, the value of the stocks in the lowest rated group

---

[44]Fischer Black, "Yes, Virginia, There is Hope: Tests of the Value Line Ranking System," *Financial Analysts Journal*, 29:5 (September/October 1973), 10-14.

[45]Scott E. Stickel, "The Effect of Value Line Investment Survey Rank Changes on Common Stock Prices," *Journal of Financial Economics*, 14:1 (March 1985), 121-143, is one of the most recent studies and contains a bibliography of previous articles on this subject.

(Group 5) realized compound annual capital losses of 0.6 percent.[46] Groups 2 through 4 experienced appropriately ranked intermediate returns.[47]

Proponents of the efficient-market hypothesis found these results disturbing. Most of the evidence, at least through the 1970s, had lent support to the efficient-market hypothesis. After all, the typical mutual fund with all its resources did not outperform the market. Some suggested that an investor could not use the Value Line recommendations to invest significant amounts of money, or that the tests of these recommendations were flawed in some unknown way. Nonetheless, the apparent success of the Value Line recommendations marked the first significant deviation from the predictions of the efficient-market hypothesis.

B. Closed End Mutual Funds

A persistent puzzle for efficient-market advocates was the presence of funds whose only assets were traded securities and whose value differed markedly from the value of these underlying securities. These funds, referred to as closed end mutual funds, trade like stocks on organized exchanges and almost always sell at a discount to net asset value. Over the last thirty years the discount on the seven largest funds has averaged about fifteen percent and, in the mid 1970s, has been as high as nearly 30%.

---

[46]These returns only measure capital gains and do not include the additional return due to cash dividends.

[47]These returns assume that the investor constructs five portfolios at the beginning of each year based upon the last set of recommendations in the prior year and holds these portfolios for one year. In actual fact, Value Line publishes recommendations weekly. If one adjusts the five portfolios within a year for these intermediate changes in recommendations, the results are even more impressive.

A fundamental question facing efficient-market advocates was why would anyone buy the underlying shares of such funds when such shares could be purchased at a discount by buying the fund itself? Alternatively, why would anyone buy such funds when they are first issued, since, although originally sold at full net asset value (plus commissions), they almost always sank quickly to a discount?[48] The lack of answers to these elementary questions posed a dilemma to efficient-market advocates.

Many possible explanations have been offered ranging from potential tax liabilities, lack of liquidity of the underlying stocks, management fees, etc. In 1977 Malkiel[49] found that none of these explanations was sufficient to resolve the puzzle.

C. The Calendar Anomaly Literature

Perhaps some of the most persuasive evidence against the efficient-market hypothesis comes from the "anomaly" literature, which has discovered unusual patterns in the price behavior of securities. Some of the most puzzling price anomalies are related to seasonal and calendar patterns in the movements of stock prices. Other anomalies relate to returns

---

[48]Indeed, these discounts point to a potentially profitable investment through the purchase of a closed end fund at a discount and the simultaneous short sale of the underlying securities. Abstracting from transactions costs, the dividends received from the fund would exceed the dividends paid on the short position, yielding a riskless profit. However, shorting the underlying stocks requires margin and exposes the investor to a paper loss if the discount should widen. Another potential way to profit is to gain control of the fund and liquidate the underlying securities, eliminating the discount.

[49]Burton Malkiel, "The Valuation of Closed-End Investment Company Shares," *Journal of Finance*, 32:3 (June 1977), 847-859. See Charles M. C. Lee, Andrei Shleifer and Richard Thaler, "Investor Sentiment and the Closed-End Fund Puzzle," *Journal of Finance*, 46:1 (March 1991), 75-109 for an up-to-date survey. The latter authors indicate that investor sentiment, which is also linked to the small stock returns, strongly influences the discount on these funds.

dependent upon the size of a firm and the dividend yield. Most surprisingly, many of these anomalies seem to occur in January, a phenomenon which became known as the "January Effect."

In the mid 1970s, Blume and Friend[50] showed that there were substantial differences in the returns between large and small firms that could not be explained by the CAPM model. From 1928 through 1968, the returns on stocks of small firms far exceeded the returns on those of large firms, although the reverse occurred during some subperiods. Recent articles have reached similar conclusions about the "size" effect,[51] and there is even evidence that this effect is present in foreign markets.[52]

The existence of a size effect in explaining stock market returns may not be that surprising. After all, it is possible that the accepted models of equilibrium had omitted some components of risk that were correlated with size. What is surprising, however, is that virtually all of the differences in the returns between large and small companies occurred in the month of January. Donald Keim has classified New York and American Stock Exchange stocks by deciles of market value into ten portfolios and then calculated "abnormal"

---

[50]Marshall E. Blume and Irwin Friend, "Risk, Investment Strategies and the Long-Run Rates of Return," *The Review of Economics and Statistics*, 56:3 (August 1974), 259-269.

[51]R. W. Banz, "The Relationship between Return and Market Value of Common Stock," *Journal of Financial Economics*, 9:1 (March 1981), 3-18; S. Basu, "Investment Performance of Common Stock in Relation to their Price/Earnings Ratios: A Test of the Efficient Market Hypothesis," *Journal of Finance*, 32:3 (June 1977), 663-682; and S. Basu, "The Relationship Between Earnings, Yields, Market Value and the Returns for NYSE Stocks: Further Evidence," *Journal of Financial Economics*, 12:1 (June 1983), 129-156.

[52]Donald B. Keim, "The CAPM and Equity Return Regularities," *Financial Analysts Journal*, 42:3 (May/June, 1986), 19-34.

returns.[53] Abnormal return was defined as the difference between the actual realized return on a stock over a period of time and a benchmark return over the same period. In Keim's work, the benchmark return is the return that an investor would have expected over that period on a stock of similar risk but without knowledge of the company's size.

As Keim discovered, abnormal returns and the market value of the firm's equity are strongly correlated with the January effect. Moreover, further work has revealed that most of the difference in the returns between small and large companies occurs in the first few days of January.[54] As yet, no one has given a satisfactory rationale for these results.[55]

Another peculiar pattern is the "day-of-the-week effect." Monday returns, measured from Friday close to Monday close, are on average negative, and virtually all of this negative return occurs from Friday close to Monday open.[56] The greatest daily returns occur on Fridays with the smallest companies realizing the greatest return. Other return irregularities

---

[53]Don Keim, *op. cit.*

[54]Donald B. Keim, "Size-Related Anomalies and Stock Return Seasonality: Further Empirical Evidence," *Journal of Financial Economics*, 12:1 (June 1983), 13-32.

[55]Keim's data show that there is also some size effect in the remaining 11 months of the year. However, this apparent evidence of a size effect in these months is due to a statistical problem in measuring returns. When this problem is corrected, the size effect is negligible in these 11 months. See Marshall E. Blume and Robert F. Stambaugh, "Biases in Computed Returns: An Application to the Size Effect," *Journal of Financial Economics*, 12:3 (November 1983), 387-404.

[56]F. Cross, "The Behavior of Stock Prices on Fridays and Mondays," *Financial Analysts Journal*, 29:6 (November/December 1973), 67-69; K. French, "Stock Returns and the Weekend Effect," *Journal of Financial Economics*, 8:1 (March 1980), 55-69; M. Gibbons and P. Hess, "Day of the Week Effects and Asset Returns," *Journal of Business*, 54:4 (October 1981), 579-596; and D. Keim and R. Stambaugh, "A Further Investigation of the Weekend Effect," *op. cit.*

have been reported during a trading day[57] and during a calendar month, with returns in the first half of each month exceeding those in the second half.[58]

Still another anomaly is the dividend yield effect. Marshall Blume reported a U-shaped relation between dividend yield and excess returns, with the greatest excess returns accruing to those stocks with either a zero yield or a high yield.[59] Keim showed that virtually all of this relation is due to the returns in January.[60]

There has not been a satisfactory explanation of these irregularities. In very recent years some of these calendar anomalies have been reduced or disappeared. Whether this is the result of investors acting on the results of these studies, or is just a statistical happenstance, is yet to be determined.[61]

---

[57]Lawrence Harris, "A Transactions Data Study of Weekly and Intradaily Patterns in Stock Returns," *Journal of Financial Economics*, 16:1 (May 1986), 99-118 and Michael Smirlock and Laura Starks, "Day-of-the-Week and Intraday Effects in Stock Returns," *Journal of Financial Economics*, 17:1 (September 1986), 197-210.

[58]Robert A. Ariel, "A Monthly Effect in Stock Returns," *Journal of Financial Economics*, 18:1 (March 1987), 161-174.

[59]Marshall E. Blume, "Stock Returns and Dividend Yields: Some More Evidence," *The Review of Economics and Statistics*, 62:4 (November 1980), 567-577.

[60]Keim, "The CAPM and Equity Return Regularities," *op. cit.* A recent paper by William G. Christie, "Dividend Yield and Expected Returns: The Zero-Dividend Puzzle," *Journal of Financial Economics*, 28:1&2 (November/December 1990), 95-126, suggests that the apparently greater excess return on stocks with zero yields is sample sensitive, and with a different sample, he finds that the excess returns on stock with zero yields are negative, not positive.

[61]The January Effect, which has been most widely publicized, appears to have diminished in recent years, particularly in futures markets where the transactions costs are small.

D. Volatility Tests

A further setback to the efficient-market hypothesis consists of the growing body of research on the volatility of financial markets. While casual observation may suggest that markets are often too volatile, proponents of the efficient-market hypothesis claim that rapid price movements are just a consequence of new information rapidly incorporated into the valuation of securities.

The value of any asset is the present value of its future cash flows. However, in an important article, Robert Shiller found statistical evidence that financial markets, and particularly the stock market, are too volatile to be explained by the behavior of dividends.[62] Shiller studied the aggregate dividends and earnings of the S&P 500 index from 1871 through 1979. He used this information to calculate what the "intrinsic" value of the S&P 500 stocks should be, for a wide range of discount rates, if investors knew with certainty the future path of dividends and earnings. Shiller called these intrinsic values the "perfect foresight," or "*ex post* rational" value of the S&P 500 index.[63] He found that the actual value of the S&P index fluctuated far more than could be explained by subsequent cash flows to investors and attributed this "excess volatility" of the stock market, as the

---

[62]Robert J. Shiller, "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?," *American Economic Review*, 71:3 (June 1981), 421-436, reprinted in Robert Shiller, *Market Volatility* (Cambridge, MA: M.I.T. Press, 1989). Related work was published by Stephen LeRoy and Richard Porter, "Stock Price Volatility: Test Based on Implied Variance Bounds," *Econometrica*, 49:3 (May 1981), 555-574.

[63]The concept of the perfect foresight price (or yield) was introduced earlier in a similar study of the British "consol" (or long-term bond) market over a two hundred year span. This study also showed excess reaction to short-term trends. See Robert Shiller and Jeremy Siegel, "The Gibson Paradox and Historical Movements in Real Interest Rates," *Journal of Political Economy*, 85:5 (October 1977), 891-907.

phenomenon came to be known, to such "irrational" behavior as investor over-reaction to short-run fluctuations in earnings or other variables.

There have been a number of criticisms of Shiller's methodology.[64] Some have questioned the use of standard statistical tests of long-term series which tend to behave like random walks. Others have indicated that the rather arbitrary decision of firms to pay dividends or retain earnings can account for the apparent excessive smoothness of dividend series. Advocates of Shiller's results have produced a great number of articles demonstrating how speculators, feeding on "noise" or liquidity traders, can make the market excessively volatile, a phenomenon we describe in Section V. The outcome of this debate is not yet settled.

An issue related to the excess volatility of the market was evidence that the returns to equity prices display "mean reversion," i.e., periods of high returns followed by periods of low returns and vice versa. If stock prices actually followed a random walk, there would be no tendency for stock returns to revert to some statistical mean. However, a number of studies demonstrated that there was a tendency for stock returns to revert to some average value over long periods of time.[65] Eugene Fama and Ken French found that between 25%

---

[64]See A. Kleidon, "Variance Bounds Tests and Stock Price Valuation Models," *Journal of Political Economy*, 94:5 (October 1986), 953-1001 and Terry A. Marsh and Robert C. Merton, "Dividend Variability and Variance Bounds Tests for the Rationality of Stock Market Prices, Aggregate Dividend Behavior and its Implications for Tests of Stock Market Rationality" *American Economic Review*, 76:3 (June 1986), 483-498.

[65]See James M. Poterba and Lawrence H. Summers, "Mean Reversion in Stock Prices: Evidence and Implications," *Journal of Financial Economics*, 22:1 (October 1988), 27-60 and Fischer Black, "Mean Reversion and Consumption Smoothing, *Review of Financial Studies*, 3:1 (1990), 107-114.

and 40% of the returns to equities over long periods could be attributable to mean reversion.[66]

Another challenge to the efficient-market hypothesis came from Ken French and Richard Roll, who confirmed that the variability of market returns when the market is closed is far lower than when the market is open.[67] This result suggests that traders extract information from the movement of prices that in turn motivates further trading and would be inconsistent with the efficient-market hypothesis. This may also indicate a relation between the structure of trading and the pricing of securities, the subject of much of the current literature on market-making.

The stock crash of October 1987 also troubled the proponents of the efficient market. The record one-day drop of 22% in the Standard and Poor's 500 Stock Index on October 19, 1987 (and even greater drop in the index-related futures price) has still not been associated with any identifiable event.[68] There is some recent research that suggests that many large

---

[66]Eugene Fama and Ken French, "Permanent and Temporary Components of Stock Prices," *Journal of Political Economy*, 96:2 (April 1988), 247-273.

[67]K. French and R. Roll, "Stock Return Variances: the Arrival of Information and the Reaction of Traders," *Journal of Financial Economics*, 17:1 (September 1986), 5-26.

[68]Mark L. Mitchell and Jeffry M. Netter, "Triggering the 1987 Stock Market Crash: Antitakeover Provisions in the Proposed House Ways and Means Tax Bill," *Journal of Financial Economics*, 24:1 (September 1989), 37-68, argue that restrictions on takeovers proposed by the House Ways and Means Committee caused the Crash. However, these actions occurred on October 14, 1987, not on October 19.

movements in the overall market occur even when there are no events to cause major changes in expected future dividends or discount rates.[69]

Theoretical and empirical problems with the efficient-market hypothesis suggested to academicians that the pricing of securities was a far more complicated process than the random walk described by the efficient-market hypothesis.

E. Further Developments in Asset Pricing

In an attempt to explain the empirical anomalies found in security prices, theoretical work continued to extend the capital asset pricing model. Fischer Black extended the CAPM to an economy which restricts short selling or does not possess a risk-free asset.[70] The latter was important since uncertain inflation precludes the existence of an asset which is riskless in real terms. Michael Brennan studied the effects of taxation,[71] and Haim Levy examined the impact of transactions costs on asset pricing.[72] Although these refinements led to somewhat different allocations of wealth between assets, they did not change any substantive results of the capital asset pricing model.

---

[69]See David Cutler, James Poterba, and Lawrence Summers, "What Moves Stock Prices," *Journal of Portfolio Management*, 15 (Spring 1989), 4-12 and Richard Roll, "R-Squared," *Journal of Finance*, 43:3 (July 1988), 541-566.

[70]F. Black, "Capital Market Equilibrium with Restricted Borrowing," *Journal of Business*, 45:3 (July 1972), 444-454.

[71]M. J. Brennan, "Taxes, Market Valuation, and Corporate Financial Policy," *National Tax Journal*, 23:4 (December 1970), 417-427.

[72]H. Levy, "Equilibrium in an Imperfect Market: A constraint on the number of Securities in a Portfolio," *American Economic Review*, 68:4 (September 1978), 643-658.

In response to Richard Roll's critique of the lack of a well defined market portfolio, Douglas Breeden developed a consumption based capital asset pricing model, where the betas were computed as the covariance of security returns with an individual investor's consumption, without needing to specify the "market" portfolio.[73] Although such a formulation has strong theoretical support, empirical tests of the model have proved to be quite disappointing.[74]

The original development of the CAPM assumed that the distribution of future returns was known and did not change over time. An important extension to the CAPM was to allow the distribution of returns to change unpredictably over time. Robert Merton was the first to introduce such changing opportunity sets by allowing the future risk free rate to be random. His model showed that the expected return of an asset depends not only upon the covariance of the asset's return with the market but also with the covariance of the asset's return with *future* expected returns.[75] This model was the first to suggest that the correlation with a single market index might be inadequate to describe asset returns.

The empirical inadequacies of the traditional CAPM and later the difficulty in identifying *the* market portfolio led Steve Ross to offer an alternative theory of asset

---

[73]D. Breeden, "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," *Journal of Financial Economics*, 7:3 (September 1979), 265-296.

[74]See Douglas T. Breeden, Michael R. Gibbons, and Robert H. Litzenberger, "Empirical Tests of the Consumption-Oriented CAPM," *Journal of Finance*, 44:2 (June 1989) 231-262.

[75]Robert Merton, "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41:5 (September 1973), 867-887. More recently, see John Campbell, "Intertemporal Asset Pricing without Consumption," mimeo, Princeton University, 1990.

pricing.[76] Ross proposed that certain economic and financial factors, such as production, interest rates, inflation, etc., drive asset returns instead of some aggregate market measure, such as the S&P 500 Index. Ross suggested that these factors could be used to fashion low variance portfolios in a way that was more effective than the tradition CAPM.

This model, called Arbitrage Pricing Theory or APT, has been subject to wide debate and can be shown to be equivalent to the CAPM theory under certain assumptions. The underlying theory is that the factors against which investors wish to hedge are far too complicated to be summarized by some ill-defined market portfolio. Macroeconomic factors cannot be summarized solely by the "market" portfolio but will appear in individual asset pricing equations. Although empirical tests have been promising, the theory in fact represents a groping for the fundamental determinants of asset prices and does not rest on the strong theoretical foundations that the CAPM does.

All these models of asset pricing still rely on the assumptions of rational participants in an efficient market. Overall, they failed to explain the growing empirical chinks in the efficient-market model: the excess volatility of the market, the pricing of closed end mutual funds, seasonal and other anomalies in stock prices and, to some extent, the predictive power of certain advisory service.

---

[76]Steven Ross, "Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13:3 (December 1976), 341-360.

V. HETEROGENEOUS EXPECTATIONS AND FINANCIAL MARKETS

A. Early Work

The early development of the efficient-market hypothesis recognized that investors had different views about the future values of dividends, earnings and other factors, but these differences in the end, were not important to security valuation. As investors attempted to profit from these different expectations, prices in the market would adjust immediately to incorporate all relevant information. Out of this process would emerge a consensus view of the returns and risks of each security.

In an important study of security pricing under heterogeneous expectations, John Lintner showed that the equilibrium value of securities is a weighted average of each investor's expectations, with greater weight given to those investors with greater wealth and to those investors with a greater tolerance to risk.[77] Risk tolerance matters because investors who are more tolerant of risk are willing to place more of their wealth in assets that they perceive to be mispriced.

Lintner showed that if the view of each investor is replaced by this weighted average, or "consensus," the equilibrium level of stock prices is identical to the level reached when each investor has distinct expectations. Thus was born the artifice of a "representative investor" with consensus expectations. Instead of incorporating the different views of each

---

[77]John Lintner, "The Aggregation of Investor's Diverse Judgments and Preferences in Purely Competitive Security Markets," *Journal of Financial and Quantitative Analysis*, 4:4 (December 1969), 347-400.

investor into the determination of security prices, one could instead use the representative or consensus view and obtain the same set of equilibrium prices.[78]

B. Rational Expectations

The incorporation of differential information into the pricing structure of risky securities thus seemed to present no real conceptual difficulties. However, Sanford Grossman in 1976 showed that earlier studies on heterogeneous expectations contained a critical flaw.[79] In the process of analyzing market equilibrium, Grossman noted that stock prices could not reflect all information as required under the strong or semi-strong form of the efficient-market hypothesis.

The flaw in the earlier research was the failure to recognize that market prices themselves contain information. Since the aggregate supply and demand of stocks determine prices, and each investor's supply and demand reflect the specific information available to that investor, prices must contain information about the views of others. In an efficient market, individuals would deduce as much information as possible from the bids and offers of other investors, incorporate that information into the information that they may have obtained independently, and then recalculate their security demands accordingly. The resultant set of prices is termed a "rational expectation equilibrium." Lintner's model of

---

[78]Mark Rubinstein showed in a very general context that there will always exist a consensus belief which, if held by all investors, would lead to the same set of equilibrium prices. See Mark Rubinstein, "Securities Market Efficiency in an Arrow-Debreu Economy," *American Economic Review*, 65:5 (December 1975), 812-824.

[79]Sanford Grossman, "On the Efficiency of Competitive Stock Markets Where Trade[r]s Have Diverse Information," *Journal of Finance*, 31:2 (May 1976), 573-585.

equilibrium assumed that investors did not revise their expectations in light of the bids and offers of other investors, and thus was not a rational expectation equilibrium.

If all participants in the market tried to infer the information of others from their bids and offers, the market price of each stock would ultimately embed all the relevant information about the security. In these circumstances, the market prices are said to be "fully revealing" in that the prices contain the special information and opinions of all investors. The market price of every security is said "to aggregate" all public and private information.

The concept of a fully revealing equilibrium is similar to the strong form of the efficient-market hypothesis. As noted earlier, a strongly efficient market reflects all information whether it is public or private. If the market is strongly efficient, the price of a security already incorporates all individuals' information relevant to determining the price of a security.

C. Auction Processes

C.1. Definitions

The price setting structure that underlies Grossman's rational expectations model is a recontracting auction. In this type of auction, all investors negotiate back and forth, taking into account the reaction of all other participants, until every investor is satisfied with the final price. No offers are binding until everyone agrees with the final configuration of quantities and prices. Such an outcome is called a "Nash Equilibrium," insofar as each participant has reacted optimally given the reaction of every other participant.[80]

This type of equilibrium contrasts with that found in standard economic models, which is termed a "Walrasian equilibrium" after the nineteenth century French economist Leon Walras. In this equilibrium, participants reveal a schedule of desired purchases or sales at a wide array of prices and then an auctioneer finds the price which clears the market. A financial market which comes close to a Walrasian market is the bidding for government securities in the periodic treasury auctions. Participants submit a schedule of quantities and prices at which they wish to acquire certain securities and recontracting is not permitted.[81]

C.2. Example

The following example illustrates a "rational expectation" equilibrium under the auction process postulated by Grossman in which no offer is binding until there is full agreement by all parties.

> Suppose that the current market price of ABC Corp. is $20 a share and that you obtain some private or special information that indicates to you that the price should really be $22 a share.
>
> Being "rational", you look at the market price of $20 and conclude that others may have other information that tells them that the stock is not worth as much as your information would suggest.

---

[80]This most famous strategic equilibrium was first described by John F. Nash, "The Bargaining Problem," *Econometrica*, 18:2 (April 1950), 155-162.

[81]Often a single price does not clear this market (the "tail" of the distribution) as the government distributes securities along the demand curve. The public is able to participate by submitting non-competitive bids, which are allocated at the average price.

Nonetheless, you determine that, although ABC Corp. may not be worth $22 a share, it is worth more than $20, perhaps as high as $20 1/2. In trying to extract the maximum profit, you place an order to buy ABC Corp. at a price no higher than $20 1/8.

But in a rational expectation equilibrium, a potential seller will go through the same thought process. He will ask, "Why is someone suddenly bidding up to $20 1/8 for ABC Corp.? Does he know something that I do not?" Thus, the potential seller is concerned that the potential buyer may have some private information that indicates a value for ABC Corp. of more than $20.

In the first round of bidding, the seller will utilize the current price of the stock, his own information, and the information that another is willing to buy at $20 1/8 to raise his offer price to somewhere above $20. By doing so he will avoid selling to, or avoid being "picked off" by the knowledgeable investor.

Let's say the potential seller counters with a tentative offer to sell at $20 1/4. The potential buyer will use his own information and the information that another is willing to sell at $20 1/4 to determine whether he should accept this offer.

If the buyer still decides to accept the offer, the seller knows that the offering price is still too low and will revise it upwards again.

Ultimately the price will adjust to a new equilibrium level that reveals all of the special information that individual investors possess. In a rational expectations equilibrium, no investor will have the incentive to trade at the final price, since there will be no profit from a trade once the price reflects all the information.

D. Inconsistency of Rational Expectation Equilibria

Grossman noted that the theory has a serious flaw in this version of a rational expectation equilibrium. If there is any cost associated with obtaining private information, investors in this model have no incentive to collect and process this type of information. As a consequence, all investors will abandon the search for costly information and hence no new information will be produced. Under these circumstances, capital markets will no longer be efficient.[82]

Thus, the presence of costly information and a strongly efficient capital market are logically inconsistent. Since an investor derives no benefit from securing private information, there will be no resources devoted to security analysis and prices would not, and could not reflect all information.

A key result of the rational expectations equilibrium with informed traders is that the very process of making an offer to buy or sell reveals "too much" information to the other side of the potential trade. If a trade is to occur, an informed trader must be able to make an offer without fully revealing his private infórmation, otherwise the price will rise to the point where the trade becomes unprofitable.

E. "Noise" Traders

Model builders had to find some device to add more "uncertainty" to the model to prevent the price from revealing all the information that informed traders possessed about a security. Initially, it was assumed that there was exogenous uncertainty in the supply of the

[82]Fama's original development of the efficient market hypothesis assumed that information was available to all and at no cost. If there is no cost to gathering information, there is no logical inconsistency in the efficient market hypothesis.

security which prevented investors from knowing whether trades were informationally motivated or not.[83]

To enable informed traders to hide partially their information from other informed traders, later work introduced "noise" or "liquidity" traders who buy and sell for reasons other than the possession of special information.[84] If an informed trader is unable to determine with certainty whether a bid or offer is from another informed trader or a noise trader, that trader cannot know in fact whether special information motivates the bid or offer. Since there is a possibility that a bid or offer comes from a noise trader, an informed traders has the possibility of profiting from making a trade. Thus, the existence of non-informationally motivated traders provides the motivation for spending resources on the acquisition of information about the value of securities and the opportunity to profit from this activity.

E.1. Early References

The first reference to liquidity trading was made in a 1971 article written under the pseudonym of Walter Bagehot.[85] Bagehot posited three types of investors: "one,

---

[83]See Sanford Grossman and Joseph Stiglitz, "On the Impossibility of Informationally Efficient Markets," *American Economic Review*, 70:3 (June 1980), 393-408.

[84]Among the reasons for "liquidity" trading are tax considerations, changes in wealth levels, changes in risk preferences, the accumulation or decumulation of assets for consumption, or for any reason not related to information about the value of the security.

[85]Walter Bagehot, "The Only Game in Town," *Financial Analysts Journal*, 27:2 (March/April 1971), 12-14, 22. The real author of this paper is Jack Treynor. Walter Bagehot himself lived from 1826 to 1877 and was a noted English banker and journalist. This article is extremely prophetic but was ignored until recently. Indeed, some recent work can be viewed as formalizing the essential insights of this article.

transactors possessing special information; two, 'liquidity-motivated' transactors who have no special information but merely want to convert securities into cash or cash into securities; three, transactors acting on information that they believe has not yet been fully discounted in the market price, but which has."[86]

The third type of investor is not consistent with a rational expectation equilibrium. These investors by failing to make profits and suffering transactions costs should eventually learn that their information is already discounted into the price and stop trading.[87] Possibly, as a consequence, recent research omits this third type and use only the "liquidity" investors to induce noise into the trading process.

Models which do not include this third group may be ignoring a large number of investors. The trading volume of many institutional traders, often exceeding one hundred percent per year, is too large to be attributable to liquidity trading alone. This observation, coupled with their inability to outperform the market, is suggestive that there are large number of investors who believe, incorrectly, that they have some special information about the true value of a security.

The introduction of noise traders permits the existence of a rational expectations equilibrium with costly information, since investors can be sure that not all of the securities they buy and sell come from individuals who have special information about the price.

---

[86]Bagehot, *ibid.*, p.13. One might add to the third type those who believe the market has *in*correctly discounted known information.

[87]The assumption that traders eventually learn their mistakes may not be true in a finite horizon, overlapping generation model, where, as P.T. Barnum had said, "There's a sucker born every minute."

Therefore it may pay some investors to acquire special information about the value of a security because such an investor will be able to acquire some shares from liquidity traders without increasing the price so much that the trade generates no profit.

E.2. Market Volatility

The presence of noise traders has also led to a growing literature concerning how they may influence the volatility of financial markets. Speculators that trade on signals which contain no information about the value of a security can move prices away from the fundamental, or full information valuation. Milton Friedman in 1953 had persuasively argued that destabilizing speculation could not be profitable and such speculators, by losing wealth, would eventually be driven from the market.[88] However, in a series of articles by De Long, Shleifer, Summers, and Waldman, it is argued that noise traders can destabilize financial markets, causing excess volatility and other pricing anomalies.[89] This is the case since rational investors often maximize over a finite horizon, rather than over the infinite future. This behavior restricts arbitrage over time since assets may become even more mispriced when the rational investor must liquidate his holdings.

---

[88]Milton Friedman, "The Case for Flexible Exchange Rates," *Essays in Positive Economics* (Chicago: University of Chicago Press, 1953).

[89]See, for example, J. Bradford De Long, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann, "Positive Feedback Investment Strategies and Destabilizing Rational Speculation," *Journal of Finance*, 45:2 (June 1990), 379-395. Andrei Shleifer and Lawrence Summers review this approach in "The Noise Trader Approach to Finance," *Journal of Economic Perspectives*, 4:2 (Spring 1990), 19-33.

Alan Kraus and Maxwell Smith have obtained a similar result that investors who trade for reasons which are not relevant to determining the asset price can influence the trading behavior of informed investors.[90] This will especially be true if different groups of traders are uncertain as to whether the other group possesses special information or not. Each group may interpret the willingness of the other group to hold positions as indicating some probability that the other group may have received favorable private information, even though no one has actually received a favorable signal. As the authors note, it is sometimes impossible for a group of speculators to know whether the demands of other traders are due to the possession of private information or to beliefs about the information of the first group.

This line of reasoning is very reminiscent of Keynes' discussion of the behavior of capital markets in the *General Theory* more than fifty years earlier. Keynes wrote:

> This battle of wits to anticipate the basis of conventional valuation a few months hence, rather than the prospective yield of an investment over a long term of years, does not even require gulls amongst the public to feed the maws of the professional; -- it can be played by professionals amongst themselves. . . . We have reached [the point] where we devote our intelligences to anticipating what average opinion expects the average opinion to be.[91]

VI. MARKET MAKING STRUCTURE

The introduction of noise or liquidity traders also began the process of interrelating the asset pricing literature with the market making literature. The interaction of informed

---

[90]Alan Kraus and Maxwell Smith, "Beliefs About Beliefs: The Effect on Equilibrium of Non-Payoff-Relevant Uncertainty," mimeo, September 1990.

[91]John Maynard Keynes, *The General Theory of Employment, Interest, and Money*, 1936, (1965 Harbinger Edition), pp. 155-56.

and non-informed traders not only changed the equilibrium in the markets but was critical in the development of models describing the way in which prices are posted by dealers and other market participants. The separation of asset pricing and market making, which had long been a hallmark of efficient-market literature, was breaking down.

If all traders are informationally motivated, there would be no trading and an efficient capital market would not exist.[92] However, when investors trade for liquidity or other reasons, it pays some investors to acquire costly information and trading takes place. Recent literature raises the possibility that the rules and organizational structure of a market could make a substantial difference in how security prices are determined. This section describes different types of market structures for the purpose of understanding some of the more widely known models of price determination at the micro-market level.

A. Types of Market Structures

There are two basic types of markets for common stocks: continuous and batch markets. In a continuous market, buyers and sellers continually interact with each other or through an agent, called a dealer, and trades can take place at any point in time. In a batch market, orders are accumulated and executed together at specific points in time. The execution of a batch market can be through a dealer, auctioneer, or automated computer system. Real financial markets often involve combinations of these two types of markets.

---

[92]Cf. Section V. D.

B. Continuous Markets: Auctions and Dealers'

There are two variations of a continuous market. One is an auction market and the other is a dealers' market. An auction market brings together all potential buyers and sellers at one physical location (or through a computer network). If the auction market takes place in a physical location, there is a limitation on the number of individuals that can gather together in what is termed the "crowd" or the collectivity of buyers and sellers. Where there are a large number of buyers and sellers, the market structure has to impose some restrictions on access to the crowd, and in this case investors must use agents to act on their behalf. An example of an "auction" market in a physical location is the market for most commodity and financial futures where the auction takes place in what is called the trading pit.

In contrast to a market at a physical location where the number of traders is limited, advances in computer networking now make it possible to include a larger number of potential buyers and sellers and theoretically all potential buyers and sellers.[93] An example, though not a pure one, of an "auction" market using a computer network is Autranet, a system used by institutional investors to trade amongst themselves. This system is a billboard system in which institutions can advertise their desires to buy or sell individual securities. It is not a pure "auction" because the institutions typically only indicate the approximate size of their intentions, and the system itself only introduces a prospective buyer

---

[93]A practical problem in permitting wide access to a computer network is providing assurance to the participants that contraparties will honor their buy and sell commitments. Steven Wunsch has recently implemented a batch system in which institutions can enter their orders directly, and the orders are guaranteed by Banker's Trust.

and seller. Once paired, the two institutions leave the system and negotiate the price and volume over the telephone.

A dealers' market, in contrast, consists of professionals who set prices at which they will buy or sell securities on their own account. Dealers post their bid and ask prices along with the number of shares available at each price for one or more stocks. A public investor (an investor who is not a dealer) can only execute a trade by notifying the dealer, who utilizes his own inventory to satisfy the order. In contrast to an "auction" market, public investors, or their agents, never trade between themselves. The National Association of Securities Dealers Automated Quote system (NASDAQ), part of the over-the-counter market, is for the most part a dealers' market.[94] Dealers utilize a computer system to display their quotes.[95] A public investor places a buy or sell order with an over-the-counter broker,[96] who either acts as a dealer himself or, as an agent, executes the order with another dealer.

C. Batch Markets: Sealed and Open

As with continuous markets, there are various types of batch markets. In one type of batch market, investors' bids and offers are kept secret, or "sealed" from the eyes of other investors. These bids can be either open market orders for a given number of shares or an

---

[94]At least one large institutional investor circumvents the dealer by having the dealer place a bid or ask into NASDAQ as if it were the dealer's quote, while in fact it is the quote of the investor. For this service, the dealer receives a commission.

[95]The quotes include the bid and ask prices and the number of shares available at these prices. In practice, the posted numbers of shares are nominal amounts and may understate the actual numbers of shares available at the quoted prices.

[96]An over-the-counter broker must be registered with the National Association of Securities Dealers or NASD.

entire schedule of quantities to be traded depending upon the actual price at which the market clears. As mentioned above, an example of this type of market is the new issue market for U.S. government securities.

If investors are able to view the orders of other investors before the auction takes place and to revise their orders before the actual auction, the market is termed open or "unsealed." The important difference between these two types of batch markets is that investors only know about their own demands in the sealed process, whereas in the unsealed process, investors learn about the demands of others. With different amounts of information available to investors, it is highly possible that the clearing price in each type of market could differ, and recent work has confirmed this intuition.

A distinguishing characteristic of a continuous market, in contrast to a batch market with sealed orders, is that buyers and sellers agree on a price and quantity before a trade takes place. A distinguishing characteristic of a batch market with sealed orders is that a trader does not know *both* the price and quantity before the auction. With unsealed orders, a trader may know both price and quantity just before the auction.

D. Actual Markets

Actual markets do not fall neatly into these highly stylized compartments and indeed are much more complex. As an example, the usual pattern of trading on the New York Stock Exchange (NYSE), the largest equity market in the U.S., is to begin or "open" trading with a batch market and then to change to a continuous market with characteristics of both a dealer market and a market where investors can trade directly with each other through floor traders. The Appendix provides more detail about the actual trading practices of the three major North American equity markets: the New York Stock Exchange, the Over-the-Counter Market, and the Toronto Stock Exchange.

VII. THE THEORY OF MARKET MAKING

A. Inventory Models

In the spirit of the original versions of the efficient market, the early literature on market structure assumed a continuous dealer market, in which the dealer's sole role was to facilitate non-informationally motivated trades. The dealer played no role in the price-discovery process. The only purpose of the dealer was to provide transaction services to the public, with the "bid-ask spread" as compensation for these services. In short, the dealer maintained an inventory to facilitate the trading of public investors.

In a pioneering article, Harold Demsetz viewed trading costs "as the costs of exchanging titles."[97] Demsetz then went on to explain that the bid-ask spread is "the markup that is paid for predictable immediacy of exchange in organized markets."[98] Immediacy is described as the service of providing an investor with an immediate execution of a buy or sell order. Thus, the bid-ask spread is very much like the inventory markup of a normal merchant, a markup charged to cover operating costs and the required return on working capital.

---

[97]Harold Demsetz, "The Cost of Transacting," *Quarterly Journal of Economics*, 82:1 (February 1968), 35.

[98]*Ibid.*, 35-36.

Since, in Demsetz's view, trading through the dealer is just like any other merchant trading activity, the principles of classical economics apply. The amount of competition that the dealer faces has a direct influence on the size of the bid-ask spread, just as it would in any market. Other markets for the stock or other market makers provide competition to the specialist. In the spirit of a competitive model, Demsetz also argued that as the volume of "trading" increased, the bid-ask spread would decline. This conclusion is drawn from the analogy with physical goods, where rapid turnover of inventory is associated with smaller margins. Demsetz presented some empirical analysis in support of his model.

But the parallel of the trading of stocks to the usual trading of goods by a merchant is not perfect. In the case of trading goods, the merchant must finance inventory by borrowing money or selling assets and must receive compensation for the interest costs incurred or revenues foregone. In the case of securities, however, there is an offsetting revenue stream in that stocks themselves are productive assets and provide a return to the holder.

In the spirit of a merchant trading model, Hans Stoll,[99] and later Thomas Ho and Stoll,[100] proposed a rationale for a bid-ask spread that explicitly recognizes that an inventory of securities provides a revenue stream. Their model assumes a single monopolistic dealer who is risk averse and has invested his wealth in three types of assets: inventory of stock for market making, cash or short-term safe assets, and other risky assets. By standing ready to buy or sell individual stocks, the dealer's inventory will change over

---

[99]Hans Stoll, "The Supply of Dealer Services in Securities Markets," *Journal of Finance* 33:4 (September 1978), 1133-1151.

[100]Thomas Ho and Hans Stoll, "Optimal Dealer Pricing under Transactions and Return Uncertainty," *Journal of Financial Economics*, 9:1 (March 1981), 47-73.

time in a random way. Sometimes, the dealer will own too much or too little of a given stock relative to the desired allocation of his wealth, or "optimal portfolio." As a result, the risk and return characteristics of his overall portfolio will often be driven to suboptimal levels because of randomly changing inventory positions. The bid-ask spread emerges in this model as the compensation necessary to induce the dealer to hold a non-optimal portfolio.

An implication of this model is that the bid-ask spread is a function of the dealer's risk aversion and level of wealth. The bid-ask spreads of two equivalent stocks could differ if they are assigned to dealers with different risk tolerances and wealth levels.

A major contribution of these types of inventory models is that they provide insight into why there may exist a bid-ask spread even though a dealer in equities does not bear the usual carrying costs of inventory. The bid-ask spread arises as compensation to induce a dealer to hold a non-optimal portfolio of assets. However, it should be noted that these inventory models provide no insight into how new information or differences in investors' information affects the price of a stock.

B. Theory of Continuous Markets

B.1. Setting the Bid-Ask Spread

One of the first models incorporating informed traders was that of Glosten and Milgrom.[101] They assumed a continuous dealer market with risk-averse informed and

---

[101]Lawrence R. Glosten and Paul R. Milgrom, "Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders," *Journal of Financial Economics*, 14:1 (March 1985), 71-99. Glosten and Milgrom build on an earlier model by Thomas Copeland and Dan Galai, "Information Effects on the Bid-Ask Spread," *Journal of Finance*, 38:5 (December 1983), 1457-1469.

uninformed traders. The dealer is risk neutral and behaves competitively. As a consequence, the objective of the dealer is to obtain an expected profit of zero. The rules of the market structure of Gloston and Milgrom require that the dealer posts a bid and offer price that is binding for one trading unit. The traders arrive randomly, each with reservation prices for buying or selling. The informed trader bases his reservation prices on both public information and his own private information, whereas the uninformed trader bases his reservation prices only upon public information.

Importantly, the dealer does not know whether the next trader is informed or not. Since the dealer has no costs, competition forces the dealer to set the bid and offer prices in such a way that the expected gain each trade is zero. The following illustrates how the dealer sets the bid ask spread in such a model.

> Let us assume that your best estimate of the true value of the stock is 20 1/8. You, the dealer need to set your bid and offer prices around the true price in order to make zero profit from a trade. Assume that if the next trader buys the stock and is an informed trader then you believe the value of the stock is 20 3/8.
>
> The problem is that you do not know whether the next buyer will be informed or uninformed--only the probability of being informed or uninformed. Let us assume that you believe that there is a fifty-fifty chance that the buyer will be informed. Since your goal is only to break even on average, you would set your ask price to the unconditional expected price given that the next trade is a buy. In this case, you would set the ask price at 20 1/4, gaining 1/8 if the buyer is uninformed and losing 1/8 if the buyer is informed.
>
> Similarly you would go through the same exercise on the assumption that the next trader decides to sell stock to you. Taking into account your expectation of the price given the trader was informed, your expectation of the price given that the trader was uninformed, and the probability of a seller being informed or uninformed, you set an offer price so as obtain zero expected profit.

Regardless of whether the next trader is informed or not, you will have already incorporated into your bid and ask price the loss if the trader happened to be informed.

In this example, the dealer does not know whether the offer is from a liquidity or informed trader. If the dealer knows that the buyer is a liquidity trader, he would sell at a lower price, so liquidity traders have an incentive to identify themselves as such. Moreover, informed traders have an interest in being mistaken for liquidity traders. For example, if the market believes that liquidity traders transact in smaller lots than informed traders, large and informed traders will split their orders into smaller orders in order to mimic liquidity traders. Informed traders may also try to hide their identity by using a number of brokers to execute a trade, as well as other techniques.[102] Therefore, both liquidity and informed traders face a number strategic decisions in executing trades. It should be noted that Glosten and Milgrom's model is not rich enough to take into account this strategic behavior.

Glosten and Milgrom go on to show that the market can fail in that no trades take place. Specifically, as the proportion of informed traders increases, the bid-ask spread will widen. At some point, the bid-ask spread may become so large that trading ceases.[103] In this case, the market fails since it does not reveal in any meaningful sense the true price of the security.

---

[102]See Fischer Black, "Trading in Equilibrium with Bluffing, Credits, and Debits," Goldman, Sachs & Co. Working Paper, March 1991.

[103]If in the limit, all traders become informed, there will be no trading under the rational expectation equilibrium, as discussed previously.

B.2. Monopoly Dealers and Price Discovery

One of the reasons that the market may fail in the model of Glosten and Milgrom is their assumption that the dealer behaves competitively by always choosing bid and ask prices that produce zero expected gain per transaction. They conjecture that a monopolistic dealer would in some circumstances keep the market open, while competitive dealers would let it close.[104] The argument is the following: A monopolistic dealer may decide to post a quotation, knowing he will lose to an informed trader, but since the quotation is only good for one trading unit, his losses will be limited. As a result of this experiment, the dealer will learn something about the true price of the security and be able to make profits from future liquidity traders.

In a competitive model, a dealer is unable to recoup earlier losses since other dealers will have also learned about the first dealer's experiment. Hence, in this setting a dealer will not experiment by posting a quote to discover the correct price. Therefore, under some circumstances, the granting of a monopoly to a dealer may encourage the faster discovery of correct prices, which benefits society.

---

[104]Christopher Leach and Ananth Madhavan, "Price Experimentation and Market Structure," mimeo, University of Pennsylvania, 1989, provide a formal proof of this conjecture.

C. Sequential Batch Markets

For the purposes of analyzing strategic behavior, Albert Kyle[105] has analyzed a sequential batch market in which there are also three types of participants: liquidity traders, informed traders, which he calls "monopolistic news traders," and dealers, who take advantage of the information in the order flow and in the sequence of prices. Kyle extends the Glosten and Milgrom model by deriving the exact sequence of prices assuming both dealers and informed traders behave strategically. He shows that the price path is still a random walk although the informed trader may be accumulating shares over time.

Kyle rigorously derives the dynamic optimizing trading strategy of informed traders as these traders try to profit from their special or inside information. The problem facing the informed trader is how to minimize the effect of his trading on the price of the security and obtain the greatest gain as the market price adjusts to the new information. Kyle shows that the informed trader will prefer to trade small quantities in each batch using the transactions of liquidity traders to camouflage his trading. Not surprisingly, as the number of liquidity traders increase, the profits of the informed trader will also increase. Kyle also examines the profits of informed traders as the number of batch markets approaches infinity and becomes a continuous market and finds that the informed traders prefer fewer batch markets.

Interestingly, in Kyle's model, market prices are almost never equal to the true price except at the end of the series of auctions when all information is revealed. Even under these circumstances, he shows that the past history of market prices has no value for

---

[105]Albert Kyle, "Continuous Auctions and Insider Trading," *Econometrica*, 53:6 (November 1985), 1315-1335.

predicting future changes, just as in the earlier theory of efficient markets. Thus, the earlier evidence offered in support of the efficient-market hypothesis, namely that historical changes in market price have no predictive value and the inability of most investors to outperform the market, is also consistent with the hypothesis that investors have heterogeneous expectations and that market prices can deviate from their true values.

Recent research by Admati and Pfleiderer[106] has added a fourth class of investor to Kyle's model. These participants are liquidity traders who can anticipate their liquidity needs and have some flexibility in the time they participate in batch auctions. The term used for these traders is "discretionary liquidity traders".

The discretionary liquidity traders will find it in their interest to trade together by concentrating their orders in the same auction. With an increased presence of liquidity traders, a price change will be less sensitive to the size of the net order imbalance. Liquidity traders have an incentive, as in the case of continuous markets, to reveal their presence to a dealer. Because of the increased presence of liquidity traders, there will also be more informationally motivated volume. Order flow attracts order flow.

The empirical implication of this theory is provocative. On exchanges, such as the NYSE, which open with a batch auction, price changes should be less sensitive to net order imbalances than during the day, when a continuous auction prevails. In the terminology of the Street, the market is "deeper" at the morning opening, when liquidity traders concentrate

---

[106]Anat R. Admati and Paul Pfleiderer, "A Theory of Intraday Patterns: Volume and Price Variability," *Review of Financial Studies*, 1:1 (Spring 1988), 3-40.

their orders.[107] Furthermore, those investors who have some discretion as to the timing of their purchases would find they receive better prices during this auction.[108]

A recent article by Ananth Madhavan[109] extends the analysis of batch markets by analyzing the differences between the prices set in sealed and unsealed batch auctions. In a sealed auction, buyers and sellers place orders that are not subject to revision. In an unsealed auction, traders are able to revise their orders before the final price is determined. As noted in Section V. C., an unsealed auction is closer to the recontracting equilibrium that characterizes the literature on an efficient market.[110]

In an unsealed batch auction, an informed trader can observe the orders of others and may learn something about the number of liquidity traders as well as special information of the informed traders. As a result, the informed trader may also revise his orders and hence influence the price set in the batch auction. In contrast, in a sealed batch auction, the

---

[107]Of course trading on the opening is influenced by investors with different interpretations of the news that has taken place since the previous close. Sometimes the morning price may be particularly volatile because of the greater uncertainty of the intrinsic value of the stock. As the demands are revealed throughout the day, volatility may decrease.

[108]Another reason for liquidity traders to concentrate their market orders at the opening of the NYSE is that the trading costs may be less. First, since all stocks trade at the same price at the open, a buyer of stock will sometimes be buying at essentially a bid price, an ask price, or within these two prices. Thus, on average, the trade price for a buy or sell will be between the bid and the ask price. Second, there is no floor brokerage on orders executed at the open.

[109]Ananth Madhavan, "Price Formation in Speculative Markets under Rational Expectations and Imperfect Competition," mimeo, University of Pennsylvania, 1989.

[110]To the extent that specialists reveal the order imbalance and likely opening price before actually opening a stock, the opening procedure on the New York Stock Exchange where investors are able to change their bids and offers before the official opening resembles closely an unsealed batch auction.

submitted orders reflect only the original information available to each investor. Since the order flow in these two types of batch auctions is based upon different sets of information, the prices set in sealed and unsealed auctions can differ. Hence, the rules and procedures of a stock exchange may have significant impacts on the types of information that stock prices reveal.

Grossman and Miller[111] have contrasted batch markets with continuous markets and have reached the conclusion that in some circumstances one type of market will dominate the other. Specifically, with normal news flow, liquidity traders will prefer a continuous market, since for small orders, a trader knows in advance the price and quantity that can be traded and receives immediate execution. However, in the presence of significant new information, they suggest that a batch market, with its greater concentration of orders, may be the more efficient market. It is interesting to note that the trading process on the NYSE parallels these observations. A continuous market accommodates small trades, but in the presence of significant news, or a large order imbalance, trading may be halted and the market reopens with a batch auction.

D. Market Fragmentation

Markets are becoming increasingly fragmented as the growth of alternative ways of exchanging equities in any part of the world allows traders to bypass national exchanges, avoiding either taxes, uncompetitive commissions, or outright detection. Furthermore, some

---

[111]Sanford J. Grossman and Merton H. Miller, "Liquidity and Market Structure," *Journal of Finance*, 43:3 (July 1988), 617-637.

dealers have successfully attracted liquidity traders by "feeding off" the quotations published in central exchanges and attempting to avoid trading with informed traders. The literature on market fragmentation is limited, but growing.

Marco Pagano[112] has shown that, under certain conditions, there could emerge two or more separate markets: one for large orders and one for small orders. For instance, institutions may find it profitable to engage in a costly search for a trading partner rather than placing a large order on an exchange where it might have an adverse price impact. Small investors, however, avoid this high fixed cost of search by using an exchange. Thus, separate markets with different allocations between fixed and variable costs could arise as a natural outcome of the competitive process.

Pagano's thesis is consistent with the development in the United States of the so-called "upstairs market" where, in response to large institutional orders, brokers search out the contraparties. Of interest, Grossman, using somewhat different logic, provides a rationale for an "upstairs" market.[113] Sometimes, a broker will know of a customer who is willing to sell or buy a particular security at the right price, but who has not submitted a formal order. For example, a broker might know of a large institution who wants to accumulate some industry group without being specific as to the actual names. If an institution wants to sell stock in that same industry, the upstairs market may provide an ideal arena in which to trade.

---

[112]Marco Pagano, "Trading Volume and Asset Liquidity," *The Quarterly Journal of Economics*, 104:2 (May 1989), 255-274.

[113]Sanford J. Grossman, "The Informational Role of Upstairs and Downstairs Trading," Rodney L. White Center for Financial Research Working Paper No. 22-90.

VIII. CONCLUSION

Academic research on the pricing of securities has come nearly full circle from the days of Graham and Dodd. In those days, most researchers took it for granted that investors had different expectations about the future and that observed security prices could and often did differ from their true or intrinsic values. Although there may be times when there is a gap between the market price of a security and its intrinsic value, Graham and Dodd assumed, without explicitly modelling the adjustment process, that over time this gap will diminish.

Graham and Dodd viewed the intrinsic value of a security as a function of the future earning prospects of the firm, and later Gordon formalized this view in his well known dividend-discount model. What was missing in Gordon's model was a formal theory to determine the appropriate risk-adjusted discount rate. Indeed, the very concept of risk was not well defined at the time.

The theoretical development of portfolio theory was advanced in the 1960s by Harry Markowitz and William Sharpe. Markowitz's model allowed heterogeneous judgments about the distribution of future returns in that each investor could utilize this model to form efficient portfolios. By imposing the additional, theoretically crucial assumption that all investors had the same view of the future, William Sharpe developed the Capital Asset Pricing Model, or CAPM. This model provided, for the first time, a formal theory for measuring the risk of an individual security. This risk was summarized by a "beta" coefficient which could be used to determine the discount rate appropriate to any security.

Empirical research found that publicly available information had little predictive value in forecasting security prices. This evidence, combined with research that showed that the typical professional investor turned in a lackluster performance, suggested that any differences in expectations would not play an important role in determining market prices. Thus was born the "Efficient-Market Hypothesis" which asserted that market prices properly reflected all the relevant information. For the purposes of obtaining tractable theories of security pricing, one might just as well assume a "representative investor" and attribute this investor's expectations to everyone. Hence the assumptions needed to derive the CAPM did not seem to be refuted by market data.

Despite the early empirical work supporting the efficient-market hypothesis, there emerged a large body of empirical literature that was inconsistent with this hypothesis as epitomized in the CAPM and its variations. Numerous studies found various anomalies, particularly those connected to calendar time, that were difficult to reconcile with any reasonable variant of most researchers' views of an efficient market. Additionally, some researchers concluded that the stock market was too volatile relative to that implied by a dividend discount model in an efficient market with homogeneous expectations.

At the same time that empirical chinks appeared in the efficient market and representative investor models of security pricing, there reappeared theoretical literature, reminiscent of Graham and Dodd, that explicitly modelled investors with heterogenous expectations. Within this framework, recent literature explores the interrelation between market structure and the incorporation of new information into the observed price. It analyzes questions such as: What are optimal trading strategies for different investors?

What are the characteristics of market prices as they adjust to the new information? How does the totality of knowledge about the true price become incorporated into the market price?

Unlike Graham and Dodd, who believed that there was a true value for a security and that this value was driven by the future earning prospects of a company, the current literature on market-micro structure is silent on what determines the true or intrinsic value of a security. It is assumed that the payoff value comes from some arbitrary distribution function.

The above simplifications may be adequate to analyze the reaction of the market to a single piece of new information. But it is clearly inadequate to understand market valuation in the face of real information flows. There is rarely a time when the "true" price is revealed in the market or when full information prevails. In the earlier literature on market efficiency, the assumption of homogeneous expectations gave meaning to the concept of a "true" price. When investors are heterogeneously informed and full information is never revealed, the very concept of a true price may not be defined.

To come a full circle from the days of Graham and Dodd, researchers will need to model the way in which investors process new information about earnings, dividends, and underlying asset values of a firm. Perhaps new pricing models will be developed that will bring us back to some of the same fundamental variables that Graham and Dodd advocated while explicitly incorporating the market structure in which the firm's securities are traded.

Some of the current skepticism expressed by academics and practitioners about the explanatory power of existing models is a necessary precursor to the development of new paradigms in the theory of asset pricing.

APPENDIX[114]

THE STRUCTURE OF WORLD EQUITY MARKETS

Actual security markets are much more complex than those assumed in the theoretical literature on market microstructure. To provide an overview of this complexity, this appendix contains a description of the three major North American security markets: the New York Stock Exchange, the Over-the-Counter Market organized under the NASD, and the Toronto Stock Exchange.

New York Stock Exchange (NYSE)

The trading mechanisms and procedures of the New York Stock Exchange (NYSE) have evolved over a time period of almost 200 years and do not fall into any simple scheme of classification. At some points in time, trading takes place in a batch market; at other times, trading takes place in a dealer's market. Trading in each stock is controlled by an assigned dealer, called a specialist. In principle, the dealer in a particular stock is a monopolist, but he in fact does face competition from traders on the floor of the NYSE, dealers and traders in other markets, and even the general public.

At the opening, the NYSE operates as a batch market where buying and selling interests are matched at a common price. After the opening, the NYSE normally operates as

---

[114]We express our gratitude to Deborah B. Sosebee of the New York Stock Exchange, Gene Finn of the NASD, John Carson for the Toronto Stock Exchange, and Eric Kirzner of the University of Toronto for their careful reading of this appendix to verify the accuracy of the institutional descriptions that it contains. Still any errors that remain are our responsibility.

a continuous market with features of both an auction market and a dealer market. On occasion, when there is impending news or a large order imbalance, the continuous market can be suspended or "halted" and then reopened in the same way as at the beginning of the day. Recently, the NYSE introduced "circuit breakers" to stop trading of all NYSE-listed stocks in response to major declines in the Dow-Jones Industrial Average.[115]

In return for being granted the position of a monopolistic dealer, the specialist has both "affirmative" and "negative" obligations that limit his trading activities. According to the NYSE, the principal affirmative obligations of the specialist are to smooth prices and facilitate the execution of public order in the case of a temporary order imbalances, to act as a catalyst to bring buyers and sellers together, to serve as an auctioneer, and to manage the "book". Negative obligations require that the specialist refrain from trading with the trend or stepping in front of public orders.

Formerly, the book was a spiral ring notebook in which the specialist recorded both limit and stop orders[116] as well as more specialized orders.[117] Currently, a computer screen has replaced the physical book for limit and stop orders, although specialized orders

---

[115]The Circuit Breakers halt all trading on the NYSE for one hour if the DJIA falls 250 point during the day and an additional two hours if the DJIA falls another 150 points.

[116]There are two types of stop orders: One is an order that becomes a market sell order if the price of the security falls to a particular level. The other is an order that becomes a market buy order if the price of the security increases to a particular level. A market order is an order to buy or sell immediately at the prevailing market price. A limit buy order is an order to buy at or below a prespecified price. A limit sell order is an order to sell at or above a prespecified price.

[117]An example is a percentage order. Robert A. Schwartz, *Reshaping the Equity Markets* (New York: Harper Business, 1991) contains a description of this and other special types of orders.

are still recorded manually. On occasion, the NYSE specialist will reveal the contents of the book to floor traders. Indeed, some specialists position the screen of their electronic book in such a way that any person on the floor of the exchange can see it. An investor off the floor may be able to learn of the contents of the book by asking a floor trader for that information, but this is cumbersome.

The opening process proceeds as follows: The specialist has the responsibility for setting an opening price at which all trades--be they buy or sell--will be executed. In setting this price, the specialist will take into account public market orders to buy or sell at the open, previously submitted limit, stop, and specialized orders, interest in the crowd, and finally his own inventory position. Until the stock is actually opened, investors can cancel or change their orders.[118] The specialist often provides (but is under no obligation) information about the volume of orders as well as a likely opening price to NYSE members on the floor of the Exchange. The members in turn have the option of relaying this information to their customers, and the specialist has the option of publishing a pre-opening indication of the likely opening price.

If the specialist can find a price not too far from the previous close,[119] he may open the stock at that price and cover the order imbalance from his own inventory. Otherwise, he

---

[118]Previously, the specialist matched all the orders by hand to determine the order imbalance at each possible opening price. Nowadays, Opening Automated Report Service (OARS), an electronic system integrated with the electronic book, facilitates this matching process by calculating the imbalance of market at open orders.

[119]If the specialist wants to open a stock with too large a price change from the prior close, the specialist must seek permission from a floor official. Specifically, approval is required if the price change is more than one dollar from a last sale of a stock under $20 a share or two dollars for a stock over $20.

may delay the opening to give time for other traders and investors to place orders. If the specialist plans to set the opening price substantially different from the previous close, the rules of the Exchange encourage, but do not require, the specialist to give an indication of the possible opening range on the tape and wait an appropriate interval of time before opening the stock. As a consequence, the opening on the NYSE has characteristics of both a sealed and an unsealed batch process.[120]

After the opening, the market for stocks on the NYSE is usually a continuous market. The specialist for a stock quotes a bid and an ask price as well as the depth of the market--that is, the number of shares available at each price. The offers to buy and sell are a combination of the limit orders on the specialist's book and the offers from the specialist himself.

---

[120]Cf. Section VI. C.

Example

The Specialist's Book
Limit Orders

Buy Price	Shares	Sell Price	Shares
		20 1/2	3000
		20 3/8	500
		20 1/4	10000
20	2500		
19 7/8	400		

The specialist might quote a bid price of 20 for 2500 shares and an ask price of 20 1/4 with 10000 shares. In this case, non-specialist limit orders represent both sides of the quote.

Alternatively, the specialist might quote a bid price of 20 for 2500 shares and an ask price of 20 1/8 for 2000 shares, with the ask price representing an offering by the specialist. The specialist could also participate at the bid by, for instance, quoting a bid price of 20 for 5000 shares, adding 2500 shares to the limit order on the book. The specialist is required to execute the limit order before he executes his own order.

With a minor exception,[121] the earliest entered orders at a given price are executed first, a procedure called "time-price priority." A specialist's offer to buy or sell is always executed after all other orders at the same price, even if the other orders are entered after the specialist's order. The time priority of limit orders allows an exchange member in the crowd or even a public investor to step ahead of a specialist's own orders.

---

[121]If the primary CME S&P 500 future contracts fall by 12 points, the NYSE's computers will delay for a minimum of five minutes the transmission to the specialist of electronically submitted market orders which are identified with program trading in the NYSE-listed components of the S&P 500. This is called the "Sidecar."

Example (Continued)

Assume the same book as above and that the quote is a bid of 20 for 2500 shares and an ask of 20 1/8 for 2000 shares. If a member in the crowd submits a limit order to sell 2000 shares at 20 1/8, that member has effectively supplanted the specialist. A public investor could also submit a limit order to sell 2000 shares at 20 1/8 and supplant the specialist.

The specialist is obligated to honor the quoted bid and ask prices and the number of shares available at each quote. Nonetheless, the quoted bid and ask prices and the available shares do not fully describe the market. For example, a specialist or anyone in the crowd may execute a market order within the bid and the ask price. Likewise, the specialist may decide to buy or sell more shares at the quoted bid or ask price than he has guaranteed in his quote. Furthermore, a specialist may hold or "stop" an incoming market order in order for the seller or buyer to obtain a price better than the posted bid or ask. In these circumstances, the specialist will guarantee that the buyer and seller will do no worse than the posted quote at the time the order is stopped.[122]

---

[122]If a limit order represents the side of a quote for which a market order is stopped and if that stopped order is subsequently executed at a better price, the original limit order will not have been not executed. Thus, the improvement in price for a stopped order may occur at the expense of a previously entered limit order.

Example (Continued)

> Assume the quote is a bid of 20 for 2500 shares and an offer of 20 1/4 for 10,000 shares and the book is given as above, so that both the bid and offer represent public limit orders. An investor is considering placing a market order to buy 100 shares. The published quote indicates that the price of this purchase will be 20 1/4, but there is the possibility that the price will be 20 1/8 if someone in the crowd steps in or the specialist chooses to sell the stock at this price.

When an investor places a market order to buy or sell on the NYSE, there is always the possibility of an execution at a price or depth better than the published quotation, either with someone in the crowd or with the specialist. On occasion, the SEC and Congress have highlighted this possibility of price improvement as a desirable feature of an exchange. This possibility of price improvement means that the published quotes are not necessarily the best prices at which trades can actually take place. In effect, the NYSE rules permit unpublished bids and offers on the part of floor traders and specialists that are better than the published bids and offers.

There are two ways to transmit a small order to the specialist post. The first, and traditional way is for the broker to call a clerk on the edge of the floor of the exchange who in turns gives the order to a floor trader. The floor trader walks the order to specialist's post for that stock and acts as an agent to obtain the best price. The second is for the broker to transmit electronically the order to the specialist's post. The specialist then acts as an agent to obtain the best possible price. The electronic system known as SuperDot expedites the transmission of the order to the specialist and the confirmation of the execution back to the broker.[123]

SuperDot has facilitated program trading[124] by allowing a trader to prepare in advance an entire program of market orders for execution at some later unspecified point and then to transmit this program to SuperDot in a matter of seconds. This ability to transmit almost instantaneously an entire program of orders to the floor of the Exchange is a major tool in reducing the trading risk of a program trade. There still remains some risk in that during the delay between the transmission of a market order to the Specialist and its actual execution, market prices might change.

SuperDot does not provide for the automatic execution of small trades, and indeed as long as there is a possibility of obtaining a better execution than at the published quotes, it would not be possible for SuperDot to provide such automatic execution.[125] As described below, other markets do provide for automatic execution of small trades.[126]

---

[123]Currently, a broker can use SuperDot to transmit market order of up to 30,099 shares and limit orders of up to 99,999 shares.

[124]The NYSE defines a program trade as a simultaneous trade of 15 or more stocks with a total market value of over one million dollars.

[125]A change to SuperDot that would allow automatic execution is the creation of an electronic file containing all unexposed bids and offers, and orders could only be executed against the exposed and unexposed bids and offers.

[126]The NYSE has on occasion experimented with automatic execution for small trades. The first experiment (1971) was the Automated Trading System (ATS) covering two or three stocks in which market orders of 100 shares were executed automatically against the published quote. The goal was to determine whether a specialist could monitor an automatic executing system and at the same time carry on his other responsibilities. The second experiment (1982) was the Registered Representative Rapid Response system (R4) in which a broker could confirm an order for 100 shares at the quoted price before sending it to the floor for execution. The SEC reluctantly gave permission for this experiment with the strong

The execution of large orders frequently occurs in what has become known as the "upstairs" market. Since a large order to buy or sell could have a substantial impact on the price of the security, institutional investors try to hide the true extent of their planned trades. Institutional investors often use the upstairs market to discreetly match a large order with orders from other investors, or against a broker's own account as a principal. Once matched, these orders are generally[127] transmitted to the floor of a registered exchange where they are executed according to the rules of that exchange. Under these circumstances, if the order were to hit any limit orders, these limit orders would first be satisfied from one side of the matched order. To date, no major market provides automatic computerized execution of large orders.

Rule 390 of the NYSE restricts the arenas in which members of the NYSE can execute trades on NYSE stocks. The interpretation of this rule (formerly Rule 394) has varied over time. Today, it requires that NYSE members execute all trades for their own account on a registered exchange when the NYSE is opened. The only exception are "Rule 19c-3 stocks," which are stocks first listed on the NYSE after April 26, 1979. The SEC instituted Rule 19c-3 as a modest experiment to enhance competition in the trading of listed equities.

The intention of Rule 390 is to keep member firms from making markets in their own back office in competition with the specialists on registered exchanges. It also prevents firms from internalizing customer orders--thereby bypassing competing dealers and markets. As some large brokerage firms have acquired specialists on both the regional exchanges and on

---

reservation that the system negated the possibility of price improvement.

[127]The exception is Rule 19c-3 stocks as discussed below.

the NYSE itself, Rule 390 is less effective at preventing the internalization of order flow and maintaining the separation of broker and dealer activities. In the case of Rule 19c-3 stocks, NYSE members are free to make markets in their own offices and, importantly, can bypass the specialist book.

NASDAQ[128]

The trading mechanisms for Over-the-Counter (OTC) stocks have evolved over time from a loosely organized network of dealers to highly sophisticated computerized network. Operated by the National Association of Security Dealers (NASD),[129] this network offers a large range of support facilities.

The OTC currently provides different trading mechanisms for stocks as a function of their size and volume. In the 1960s, the trading of stocks not listed on any registered exchange took place in a loosely organized telephone network of dealers, termed the "Over-the-Counter" market. Each morning a stack of "pink sheets" was published which listed many thousands of OTC stocks, the dealers who made a market in each stock, and representative bids and offers of these dealers as of 2:00 pm of the prior day. These "pink sheets" served two purposes: They allowed investors to price OTC stocks and they provided a listing of the brokers who made markets in each OTC stock. To trade a particular issue, a

---

[128]NASDAQ is the acronym that is always used for this market. It stands for National Association of Security Dealers Automated Quote System.

[129]The NASD is also the self-regulatory body for the OTC market.

broker was required to obtain and compare three different quotes before, if possible, executing the order. This process required a minimum of three phone calls.

A major change occurred in the OTC market in 1971. In that year, the NASD developed NASDAQ, an automated quotation system, for the larger, more actively traded OTC stocks. NASDAQ is a network of computer terminals in which dealers in OTC stocks enter current quotations for dissemination through the network. NASDAQ provides the same function of displaying prices and dealers as the "pink sheets," but on a current basis.

In addition to the bid and offer price, the quotation of each dealer contains a binding indication of the number of shares available at the bid and the offer price, which is termed the "depth" of the market. In practice, dealers have for the most part displayed only the minimum mandated number of shares (200 to 1000 shares) rather than the actual depth of the market. Thus, the true depth of the market is often greater than displayed. To obtain a quote for a order larger than the nominal size still requires a telephone inquiry and perhaps some negotiation.[130]

It is of interest that dealers usually only post the minimal mandated depth even though the design of NASDAQ allows a larger depth to be published. Apparently, dealers do not wish to reveal their true intentions. This observation demonstrates that, while highly advanced computerized trading system may be technically feasible, a system will only succeed if it meets the needs of its participants.

---

[130]In November 1990, the NASD introduced SelectNet, a computerized network that allows dealers and brokers to bypass the telephone and broadcast an order of any size to all the other dealers for instantaneous execution. According to the NASD, this system now executes several million shares a day.

The largest and most actively traded stocks on NASDAQ,[131] which capture 95 percent of the dollar volume, have been designated National Market Stocks (NMS)[132] and as such, are subject to "last trade reporting." The effect of this requirement is that both the price and volume of any trade in a NASDAQ/NMS stock must be reported within 90 seconds.[133]

In December 1984, NASD introduced the Small Order Execution System (SOES) that allowed the automatic execution against the best bid and offer of market orders of up to 1000 shares for NASDAQ/NMS stocks and up to 500 shares for the other stocks quoted in NASDAQ.[134] In January 1989, SOES was expanded to include limit orders. Public orders are executed rotationally among those dealers with the best bid and ask prices. Importantly, this rotational assignment of orders gives no priority to those market makers who were first with the best bid or offer nor to those who provided the greatest depth. A

---

[131]There are two methods for trading the NASD stocks that are not part of NASDAQ: First, pink sheets for the least active NASD stocks are still published daily. Second, NASD has developed a computer bulletin board where brokers and dealers can advertise an interest in trading a specific stock where this interest can be binding or non-binding.

[132]The NASD was the first to use the term NMS, but it now applies to any stock with last trade reporting requirements.

[133]Because of the number of dealers and 90 second delays, the ordering of some of the reported trades may not reflect the time sequence of actual executions although most small trades are now automatically executed and reported within seconds of the actual time of execution.

[134]Additionally, there are several competing automated proprietary systems operated by broker/dealers that often provide greater depth.

market maker can jump into the rotational trading at any time and have virtually the same access to order flow as any other market maker.[135]

To provide guaranteed liquidity to SOES, any registered market maker in a NASDAQ stock now must participate in this system. During the Crash of October 19, 1987, when participation in SOES was voluntary, many market makers withdrew from SOES, drawing much criticism. As a result, the NASD adopted a rule that requires market makers who withdraw from making a market in a specific stock to wait twenty days before they can resume making a market in that stock.[136]

The Toronto Stock Exchange

The Toronto Stock Exchange trades stocks in two different ways: on a trading floor and through the Computer Assisted Trading System (CATS). Trading in the more active stocks generally occurs through a melding of the traditional crowd on the floor of the exchange with an electronic system that allows brokers off the floor to enter orders. Each of these active stocks has a Designated Market Maker (DMM), who has responsibilities similar to those of the NYSE specialist. Importantly, members of the exchange have open access to the book of limit orders as well as the identity of the brokerage house submitting each limit

---

[135]This rotational allocation of order flow is likely to reduce the probability that one market maker will come to dominate the trading of smaller orders, particularly since the minimum spread for all but the lowest priced stocks is one eighth.

[136]Since NASD market makers are under no affirmative obligation to maintain a tight spread, one must question the efficacy of this new regulation. After all, market makers can effectively withdraw from SOES by increasing their published spreads by a sufficient amount to discourage any trading through SOES.

order. Quote vendors supply the contents of the limit order book to subscribers as well, but without the identity of the brokerage houses. Although an investor can sometimes obtain knowledge of the content of the book on the NYSE, it is a very cumbersome procedure. In contrast, the open system of the Toronto Exchange allows an investor to examine readily any book without first having to contact the floor.

In integrating the floor traders with an electronic screen, Toronto delays the transmission of electronically submitted orders to the floor by 30 seconds. One might conjecture that this delay gives some trading advantage to floor traders. Floor traders do not see these orders during the delay, but the designated marker maker is aware of these orders and has the right to cause an additional delay in order to enter a previously declared order or to improve an execution price.

Trading in the less active stocks generally takes place through the Computer Assisted Trading System (CATS),[137] which provides access to the Market Order System of Trading (MOST) and to the Limit Order Trading System (LOTS). A DMM is assigned to manage the trading of the more active stocks included in CATS with the obligation to preserve an orderly market. A market surveillance computer program monitors the trading activity of the very thinly traded stocks in CATS, and is programmed to stop trading in the event of unusual price movements. In 1990, stocks traded through CATS represented 23.1 percent of the total

[137]A few actively traded issues, such as Petro Canada and TIPS (Toronto Index Participation Shares) are included in CATS.

trades on the Toronto, 22.4 percent of total share volume, and 17.2 of market value volume.[138]

In entering a large order, a trader in CATS or in the electronic system that is integrated with floor trading has the option of not disclosing to other investors that part of the order in excess of 5000 shares. However, Toronto gives priority at a given price to disclosed orders, so that any undisclosed orders are executed only after the disclosed orders have been executed.

Like the U.S. market, there is an upstairs market that matches large trades. The rules of the Exchange require that these matched trades be executed on the exchange.[139]

---

[138]The office of John Carson at the Toronto Stock Exchange supplied this information.

[139]When the spread between the bid and offer is greater than the minimum tick size, the price of a matched order can be adjusted through offsetting changes in commissions to insure execution within the bid and the offer. To the extent that a cross occurs at a bid or an offer, the book must first be cleared as on the NYSE.