Journal of Mathematical Finance, 2025, 15(3), 666-686
https://www.scirp.org/journal/jmf
ISSN Online: 2162-2442
ISSN Print: 2162-2434

Applied Investment Research and the CRSP Stock Market Database: Celebrating 60 Years of Financial Research

John B. Guerard Jr.[1], Dimitrios Thomakos[2], Foteini Kyriazi[3], Bijan Beheshti[4]

[1]Pacific Spirit Investments, Bluffton, SC, USA
[2]Department of Business Administration, National and Kapodistrian University of Athens, Athens, Greece
[3]Department of Agribusiness and Supply Chain Management, Agricultural University of Athens, Athens, Greece
[4]FactSet Research Systems, San Francisco, CA, USA
Email: John.Guerard@pacificspiritinvestments.com, dthomakos@ba.uoa.gr, fkyriazi@aua.gr, bbeheshti@factset.com

How to cite this paper: Guerard Jr., J.B., Thomakos, D., Kyriazi, F. and Beheshti, B. (2025) Applied Investment Research and the CRSP Stock Market Database: Celebrating 60 Years of Financial Research. *Journal of Mathematical Finance*, **15**, 666-686.
https://doi.org/10.4236/jmf.2025.153026

Received: April 2, 2025
Accepted: August 22, 2025
Published: August 25, 2025

Abstract

The Center for Research in Security Prices (CRSP) database was funded and created to address the premier financial database for the academic and financial communities. At the time that the CRSP was funded in 1960, there were no large sets of databases for financial research. Additional questions were analyzed regarding the usefulness of daily stock returns in event studies. Academicians, such as James Lorie and Lawrence Fisher, the CRSP co-Directors, were concerned with relative stock and bond monthly returns, 1926-1964. William Sharpe, were presented his analysis of mutual fund returns at one of the first CRSP research meetings. Barr Rosenberg was an early prolific researcher of equity models and made numerous CRSP presentations in the 1970s and early-1980s. Harry Markowitz and his research group at Daiwa Securities experimented with CRSP data for monthly stock portfolio selection of equity portfolios, presenting at several CRSP research seminars in the early 1990s. In this article, we update the Markowitz Daiwa Securities articles presented at CRSP. Expected stock returns are a key input to portfolio selection. Stock selection models often use momentum, analysts' expectations, and fundamental data. We briefly trace the early economic theory of profits and stock prices, hypothesized by Wesley Clair Mitchell, the first Director of Research at the National Bureau of Economic Research, the NBER, 1910-1946. The Mitchell emphasis on corporate profits on business cycles was followed by the fundamental investing framework of the Graham and Dodd *Security Analysis* approach to investing. The fundamental approach to investing has been enhanced over the past 60 years by modeling stock data on momentum, analysts' expectations, and fundamental using CRSP data. This analysis specifically updates research of the past 30 years.

Keywords

Stock Prices and Returns, Relative Stocks and Bond Returns, Portfolio Selection, Robust Regression, Efficient Markets

1. Introduction

The Center for Research in Security Prices (CRSP) was funded by Merril Lynch, Pierce, Fenner, & Smith, Inc. in 1960, to conduct research and to disseminate the results to the academic and financial communities.[1] In this analysis, we review the development of the CRSP database by Lawrence Fisher and James Lorie, and its initial applications. The original stock database contained monthly closing prices of common stocks during the January 1926 to December 1960 period. The closing prices were adjusted for capital changes, cash and stock dividends, stock splits, rights, and exchange of shares. Professors Fisher and Lorie [1] stated in the first CRSP article that CRSP research was directed to controversies and unresolved issues regarding stock prices, including whether successive changes in the prices of common stocks are statistically independent or are serially correlated and whether dividends exerted pressure of stock prices. Research topics included the relationships between earnings and prices and the relevance of balance-sheet data in understanding and predicting the financial fate of corporations, and the relationship of stock prices to movements in the economy. Earnings were input on IBM cards into CRSP. Rates of returns including dividend reinvestment were created on the CRSP database such that research could be reported in 1964. Fisher [2] reported a new stock index computed from link relatives that are weighted averages of the arithmetic and geometric link relatives. During the entire thirty-five-year period, 1926-60, the rates of return, compounded annually, on common stocks listed on the New York Stock Exchange, with reinvestment of dividends, were 9.0 per cent for tax exempt institutions and 6.8 per cent for persons in the $50,000 income class.[2] The stock returns were substantially higher than for alternative investment media for which data are available. Professor Fisher [2] subtly stated that "probably worth noting here that a dollar earning 9.0 per cent per annum, compounded annually, would be worth over $20 in 35 years". The concept of investing in stocks for the long run was reborn in 1996.[3] Lorie, Dodd, and Hamilton Kimpton [3] reported in their well-known text that from 1926-1981, common stocks on CRSP had returned an arithmetic mean risk-premium (common stock returns, Treasury bill returns) of 8.3 percent with a 22.0 percent standard deviation; whereas the default risk premium on long-term corporate bonds was 0.5 percent with a corresponding 3.2 percent standard deviation. The

[1]Lawrence Fisher and James Lorie. 1964. "Rates of Return in Common Stocks", The *Journal of Business* 37, 1-21. A reprint of the Fisher and Lorie (1964) article was given to attendees of the CRSP Forum 2010, celebrating the 50th anniversary of the CRSP database.

[2]L. Fisher, 1966. "Some New Stock-Market Indexes", *The Journal of Business* 39, Supplement on Security Prices, 191-225.

[3]The authors use the term "reborn", because there is a long and substantial linkage between earnings, stock prices, and US business cycles, as originally found in the Wesley Clair Mitchell (1913).

maturity risk premium on long-term government bonds was 0.2 percent with a corresponding 6.5 percent standard deviation.[4] Some of the original research published using the CRSP database, primarily in *The Journal of Business*, involved factor analysis of stock returns, and tests of the Efficient Markets Hypothesis (EMH), for serial correlation in stock returns[5], stock splits[6], trading strategies of stock price trends[7], and merger profitability[8]. The overall impact of the CRSP database and its researchers supported the EMH, which stated that current stock prices reflected information. The EMH had three forms: stocks prices reflected historical stock prices and trading volume patterns, in its weak form of the EMH; public announcements of earnings, stock splits, and mergers, the semi-strong form of the EMH; and was somewhat mixed in its strong form, which held that all information is incorporated into the stock prices.[9] Financial researchers have used the Center for Research in Security Prices (CRSP) database and Compustat tapes since the mid-1960s as a source for financial data. The introduction of the I/B/E/S data in 1976 was another step forward in accessing financial research. I/B/E/S data is now included in the Wharton Research Data Services (WRDS) databases with CRSP. The overall summary is that historical earnings, profits, and earnings forecasts, earnings forecast revisions, and earnings forecast revisions have driven stock prices and returns during the 1960-2024 period.

The rise of the CRSP, Compustat, and I/B/E/S databases resulted from the financial community wanting to quantify by how stocks outperformed bonds and were there small stock premiums, initially reported in Mitchell [4]-[6] and Burns and Mitchell [7] continued. The post-World War II period (particularly 1960-2024) has been characterized by long-term prosperity, with booming stock prices reflecting the outstanding growth of corporate profits. The US economic growth of the past almost 80 years was not the for the United States during its first 175 years, particularly the 40 years prior to the establishment of the US Federal reserve system in 1913. We trace the early economic theory of profits and stock prices, hypothesized by Wesley Clair Mitchell, the first Director of Research at the National Bureau of Economic Research, the NBER, 1910-1946. Moreover, Graham and Dodd [8]-[10] and Williams [11], developed stock price valuation theories based on earnings and dividends. Markowitz [12] was particularly influenced by Williams.

In this analysis, we briefly address the historical role of profits in business cycles, discuss how the low earnings-to-price theories of valuation influenced Mar-

[4]See J. Lorie, P. Dodd, and M. Hamilton Kimpton, *The Stock Market: Theories and Evidence*, Homewood, Ill: Richard D. Irwin, Inc., Second Edition, 1985, Table 2-5.

[5]E. F. Fama. 1965. "The Behavior of Stock Prices*", The Journal of Business* 38. 34-105.

[6]E. F. Fama, L. Fisher, M. Jensen, and R. Roll. 1969. "The Adjustment of Stock Prices to New Information*", International Economic Review* 10, 1-21.

[7]E, F. Fama and M. E. Bulme, 1966. "Filter Rules and Stock Market Trading ", Security Prices" A Supplement, *The Journal of Business* 39, 226-241.

[8]See J. Lorie, P. Dodd, and M. Hamilton Kimpton, *The Stock Market: Theories and Evidence*, Chapter Figure 4-7, p. 71.

[9]H. V. Roberts. 1959. "Stock Market 'Patterns' and Financial analysis: Methodological Suggestions", The Journal of Finance 14, 1-10 for the original EMH statement. See R, Brealey. 1969*. An Introduction to Risk and Return to Common Stocks*, Boston: MIT Press, for an excellent, very readable monograph on common stocks and market efficiency.

kowitz and his investment research during the 1991-2021 period, and report research on how the earnings and balance sheet data of the CRSP database was enhanced by the inclusion of the Institutional Brokerage Estimation System (I/B/E/S) database of analysts' forecasts of earnings per share, created in 1976. In 2024, researchers can download 60 years of common stock earnings, book value, and other income statement, balance sheet, cash flow statements, and earnings forecast historical items to forecast stock prices and returns. Just as CRSP data was used to compare common stock returns, corporate bond returns, government bond returns, and Treasury Bill yields, since 1926, researchers can develop, estimate, and stock selection models with which asset managers can possibly "beat the market" with active management. The CRSP database is used in Section 2 to (1) incorporate the forecasted profits of businessmen as a key driver of investment and growth; (2) building models of expected returns for stocks, with an emphasis on historical and expected corporate earnings per share; and (3) discussing Markowitz Mean-Variance Efficient Portfolios and Multi-Factor Risk Models in Section 3; and (4) reporting Markowitz Mean-Variance Efficient Portfolios using the Axioma Multi-Factor Risk Model for the 1994-2020 period using CRSP data in Section 4. Summary and conclusions are presented in Section 5.

2. Mitchell's Early NBER Business Cycles Analysis: What Investors Need to Know about the Cumulation of Prosperity

The study of management expectations of enterprise profits, business cycles, and stock prices can be traced back to Wesley Clair Mitchell [4]. Stock prices and business cycles have long been associated. Mr. Wesley Clair Mitchell, the first Director of Research at the National Bureau of Economic Research, NBER. Mr. Mitchell was elected the NBER Director of Research on February 2, 1920, shortly after its chartering.[10] He held that position for over 25 years. In that capacity, Mr. Mitchell produced the first two NBER volumes of research, on Business Cycles.[11]

Let us begin with Part III of Mr. Mitchell's *Business Cycles* (1913) and review prosperity, crises, and depressions, his stages of economic activity and the implications for stock prices. Mr. Mitchell wrote his book with an intended audience of businessmen. In this monograph, we stress money profits by business enter-

[10]The NBER was chartered on December 29, 1919. The National Bureau was formed "to encourage, in the broadest and most liberal manner, investigation, research and discovery, and the application of knowledge to the well-being of mankind; and in particular to conduct, or assist in the making of, exact and impartial investigations in the field of economic. Social, and industrial science". National Bureau Charter and By-Laws. The quote is taken from *Wesley Clair Mitchell: The Economic Scientist*, Arthur F. Burns, Editor (New York: National Bureau of Economic Research, 1952), pages 30-31. Mr. Burns edited a volume shortly after Mr. Mitchell passed on October 29, 1948. Contributors to the volume included Joseph Dorfman, John Maurice Clark, Milton Friedman, Joseph Schumpeter, and Alvin Hansen. A personal sketch was included by his wife, Lucy Spague Mitchell.

[11]In fact, prior to his NBER appointment, Mr. Mitchell had conducted extensive research on the history of the greenbacks, currency issued during the U.S. Civil War, and business cycles, producing several large monographs on these topics. Wesley Clair Mitchell, *A History of the Greenbacks, with Special References to the Economic Consequences of Their Issue,* 1862-1865 (The Decennial Publications of the University of Chicago, 2d. Series, V.9, the University of Chicago Press, 1903). was born in 1874, in Rushville, Illinois, and entered the University of Chicago, under President Harper at its creation in 1982. Upon completion of *Business Cycles*, Mr. Mitchell joined the faculty at Columbia University, in 1913.

prises, as did Mr. Mitchell when he stated, “Since the quest of money profits is the controlling factor among the economic activities of men who live in a money economy, the whole discussion must center about the prospects of profits”.[12]

Mr. Mitchell began his discussion of business cycles with the revival of business activities between 1890 and 1910.[13] U.S. revivals began with highly profitable grain harvests in 1891 and 1897 and a successful defense of the gold reserve standard in 1895. Depressions eventually create conditions for recovery, such as falling prime and supplemental costs of manufacturing commodities (inputs) and in inventory of wholesale and retail merchants, low rates of interest, and a liquidation of business debts. Falling costs and lower interest rates tend to widen profit margins and facilitate bank borrowing. Once started, a revival of economic activity spreads across most, if not all, of the business world. The industries producing the raw materials and supplemental supplies are the first industries stimulated. Transportation, railroads and banking industries start to boom. Employees earn higher salaries and proprietors earn higher profits, enabling workers and owners to pay off debts incurred in depressions and expand their purchases. Better quality food is substituted for lesser quality eaten during depressions and clothing demand increases. Furniture, entertainment, and luxury items are purchased. Industries are stimulated as the revival expands across the economy. Optimism endues and loans are provided for business enterprise expansion. Optimism is reenforced by an increase in the volume of goods ordered.

Prices begin to rise, but tend to lag in the revival, and rising prices create a stronger incentive for obtaining larger supplies to sell at wider profit margins. Eager bidding allows suppliers to exact higher prices. Mr. Mitchell observed that (1) retail prices rose less than wholesale prices of the same commodities; (2) wholesale prices of finished goods lag behind the prices of partially manufactured goods in the same commodities; (3) wholesale consumer goods prices rose less than wholesale producers’ goods prices; and (4) wholesale prices of raw materials responded to changes in business conditions with greater accuracy and certainty than whole prices of raw farm or forest products.[14] Men control more completely the production of coal, iron, copper, and zinc more than the production of beef, pork, mutton, and wool.to meet the increasing demand.

Workers traditionally think of making a living, rather than making money. The prices of labor rise less in a revival than wholesale commodity prices. Mitchell attributes some of the lag in wages to nonexistent or weak trade-unions.[15] Despite a lagging wage, working class members are better off as business conditions im-

[12]Wesley Clair Mitchell, *Business Cycles and Their Causes* (Berkeley and Los Angeles: University of California Press, 1963, Fifth printing), p. xi. Mitchell, 1963, is the fifth printing of Wesley Clair Mitchell *Business Cycles and Their Causes*, 1941, reprinting of Part III of the original *Business Cycles: University of California Press, 1913*. The reader can read the original 1913 on the National Bureau of Economic Research website. See Mr. Clark’s discussion of Mr. Mitchell and his “money economy” in John Maurice Clark, “Theory of Business Cycles”, in *Wesley Clair Mitchell: The Economic Scientist*, Arthur F. Burns, Editor (New York: National Bureau of Economic Research, 1952), pages 198-199.

[13]Mr. Mitchell frequently uses business fluctuations in lieu of business cycles. Joseph A. Schumpeter, “The General Economist”, in *Wesley Clair Mitchell: The Economic Scientist*, Arthur F. Burns, Editor (New York: National Bureau of Economic Research, 1952), pages 302-333.

[14]Mitchell, *Business Cycles*, 1963, p.13.

[15]Mitchell, *Business Cycles*, 1963, p. 17.

prove. Interest rates lag in a revival. In fact, discount rates are usually lower in the first year of a revival than the rates were in the last year of a business depression. Bank loans increase in a revival and bankers often have liberal reserves. During a recovery from a depression, the ratio of capital liabilities to total liabilities falls due to depositors' inflows.

The net effect of rising prices and lagging wages and interest costs are increased profits.[16] Supplemental costs rise slowly with the physical volume of business. Wages, freight costs, prices of prime and supplemental rise slower than prices such that net profits rise with business volume. Profitability increases vary greatly across industries, due to relative differences in industry prime, supplemental, labor, and transportation cost ratios. Mitchell turns his attention to the stock market, and states plainly that the market price of a business enterprise rests primarily on the capitalized value of its current and prospective profits. That is, stock prices "vary roughly" with the rate of profits.[17] In 1913, as Mitchell wrote his *Business Cycles* monograph, railroads were the one group of business enterprises with data existed on stock prices and profits.[18] "Dividend smoothing" was noted in 1913 as Mr. Mitchell noted that dividends had been kept more stable than net income. Interest rates at which stock prices are capitalized are subject to variations. Further complications, known in 1913, included stock manipulation, speculation, and contests for control. Mr. Mitchell's further observations included that low-priced stocks rose more than high-priced stocks; common stocks rose more than preferred stocks; and irregular dividend-paying stocks rose more rapidly than stable dividend-paying stocks.[19] Business revival conditions lead to increases to business expansion in size and the creation of new enterprises. New investment never ceases entirely, but falls to very low level in a depression, and becomes large again after the recovery is well established. The creation of new enterprises increases demand for buildings, machinery and furnishings, which further increases the demand for materials, labor, equipment, and loans. Increases in prices, volume of trade, and profits make optimists of entrepreneurs and every convert to optimism makes new converts, further favoring business expansion and further price increases. Workers now demand, and employers conceded to higher wages, resulting in higher family income that widens the market for consumer goods. Higher wages increase labor costs of retail and wholesale commodities. Interest rates rise for similar reasons and increase the cost of production. The increases in wages and interest rates are why prosperity does not continue indefinitely. Mr. Mitchell refers to prosperity as "The Business Equilibrium". It is the increasing costs of business that disrupt business equilibrium. The economic condition of the US economy has been one of prosperity, on average since the creation of CRSP. The US economy has never boomed continuously. Mr. Mitchell's re-

[16]Mitchell, *Business Cycles*, 1963, p. 20.

[17]Mitchell *Business Cycles*, 1963. p. 22.

[18]Mitchell *Business Cycles*, 1963. p. 22.

[19]Mitchell, *Business Cycles*, 1963, p. 23. See also W. C. Mitchell, "The Prices of American Stocks: 1890-1909", *Journal of Political Econo*my 18 (1910), 345-380, and his "The Prices of Preferred and Common Stocks, 1890-1909", *Journal of Political Econo*my 18 (1910), 513-524.

search centers on the earnings expectations of businessmen and their investment decisions. Expectations of enterprise, or corporate profits, were associated with higher stocks prices and an expanding economy, culminating in prosperity. The expanding US economy was driven by corporate profits and expectations of those profits.

3. Expected Returns Modeling and Stock Selection Models: The Role of Corporate Profits

Do earnings matter in determining stock prices? The consensus among most economists is yes. Benjamin Graham and David Dodd, in their classical *Security Analysis* [8], are presently considered by many, including Warren Buffet, the preeminent financial contributors to the theory of stock valuation. In Chapter 27, "The Theory of Common-Stock Investment," Graham and Dodd discussed their explanation for the departure of the public from rational common valuation during the 1927-1929 period. Graham and Dodd attributed much of the valuation departures to the instability of intangibles and the dominant importance of intangibles (p. 301). Investment in the pre-war (World War I) was confined to common stocks with stable dividends and fairly stable earnings, which would lead to stable market levels (p. 303). Graham and Dodd hold that the function of (security) analysis is to search for weakness in pre-war conditions such as improperly stated earnings, a poor current condition of its balance sheet, or debt growing too rapidly (p. 303). Moreover, new competition, deteriorating management and/or market share must condemn the common stock from the possible valuation of a "cautious investor." Graham and Dodd attributed speculation on future prospects or expectations. They taught their students at Columbia University that buying common stock could be viewed as taking a share in the business (p. 305). In the "new-era" theory of common stock valuation, (1) dividend rate had little bearing upon value; (2) there was no relationship between earning power and asset value; and (3) past earnings were significant only to the extent that they indicated what changes in earnings might occur in the future (p. 307). Graham and Dodd chastised 1927-1929 investors for valuation analysis that emphasized: (1) "the value of a common stock depends on what it can earn in the future; (2) good common stocks will prove sound and profitable investments; and (3) good common stocks are those which have shown a rising trend of earnings" (p. 309). The new-era investment theory was held to be the equivalent of pre-war speculation. Graham and Dodd attributed much of the speculation to the work of Edgar Smith and his *Common Stocks as Long-Term Investments* (1924) text. Graham and Dodd held that Smith postulated that stocks increased in value more than was justified as a result of reinvestment earnings capitalization. That is, if a company earned nine percent, paid dividends of six percent, and added three percent to surplus (retained earnings), then good management should lead the stock value to increase with its book value, with the three percent compounded. Graham and Dodd assaulted the new-era investment theory for paying 20 - 40 times earnings. Smith, with his reinvestment of surplus earnings theory built up asset values and thus created the growth of common-stock values (Graham and Dodd, p. 313).[20] John Burr Williams [12], in his dissertation at Harvard, proposed that the value of a stock should equal the present value of its expected future dividends, which are paid earnings. Williams included the Graham and Dodd low price-to-earnings strategy and the Graham and Dodd net current asset value (buying stocks for their "liquidation" or break-up value) strategy.

Graham and Dodd [9] discussed the concept of intrinsic value in the 1951 edition of their *Security Analysis*, and approximated intrinsic value by including the past level of earnings, earnings trends, and the pattern of dividends in their equation. A capitalization rate based on standard levels modified for various qualitative factors affecting the particular firm is applied to a weighted combination of earnings and dividends. The result is an intrinsic or normal value.

$$\text{Intrinsic Value} = M\left(\frac{E}{3} + D\right) \pm \text{asset factor}, \tag{1}$$

where M represents a "normal capitalization multiplier" which ran from 8 to 15 for ordinary industrials with lower multipliers given for highly leveraged and volatile companies and the higher multipliers reserved for stable and blue-chip companies.[21] E is the projected earnings per share, generally based on the average earnings of the last five years corrected for any pronounced trend. The figure for earnings is divided by 3, and then D, the projected dividend, is added to the multicand. The amount for D is generally based on the dividends of the last five years. When a company's payout rate approximated 66% per cent, M approaches the pure earnings multiplier. The asset correction is used to reduce the intrinsic value if it exceeds book value by too large an amount; a plus correction is made if intrinsic value based on earnings falls below the rule-of-thumb "liquidating value" per share. Graham and Dodd admitted the difficulty of applying the formula to very new companies, to high-risk situations, to highly leveraged companies, and to strong growth companies.[22] Corporate profits dominate the Graham and Dodd intrinsic value calculation, the primary calculation in fundamental investing.[23] The Graham and Dodd investing framework dealt with selecting stocks on the basis of historical earnings growth, with little to no consideration of risk. Harry Markowitz read Graham and Dodd and Williams and realized that their system was not complete. Portfolio selection, including risk, is forthcoming in our next

[20]Benjamin Graham and Jerome Newman created the Graham and Newman Corporation in 1936 to buy stocks in companies whose price was less than their intrinsic value. Graham and Newman engaged in hedging these opportunities. In their 1946 letter to stockholders, Benjamin Graham wrote, on January 31, 1946, that their 4 million portfolio, composed of $2.18 million of bonds, $0.86 million of preferred stock, and $1.13 million of common stock had produced a return over its 10 years of existence of 17.56 percent while the S&P had produced a corresponding return of 10.1 percent and the Dow has returned 10.0 percent. The Graham-Newman Corporation was liquidated upon the retire of Benjamin Graham as its President in 1955. Clearly Benjamin Graham, as a faculty member at Columbia University, and as President at the Graham-Newman Corporation believed that his valuation techniques, based upon purchasing low PE stocks worked. Earnings mattered.

[21]For the basic statement of this concept, see Graham Dodd, *Security Analysis,* McGraw-Hill, 1951, pp: 410-11. See Chapter 8 of Guerard and Schwartz (2007) for a more complete introduction to Graham and Dodd valuation and the pricing of common stock.

[22]In the fall of 1974, when stock prices had fallen substantially, Benjamin Graham wrote in the *Financial Analysts Journal*, *FAJ,* that he believed that the low PE and low PB, book value-to-price stocks, looked very attractive and was convinced that future portfolio returns would verify his belief.

[23]See Latane, Tuttle, and Jones (1975) and Malkiel (1973 and 2023).

two sections.

Here the authors discuss issues of databases and the inclusion of variables, particularly net income (profits) and expectations of future net income, by security analysts, in composite models to identify undervalued securities in a United States stock universe. The database for this analysis is created by the use of all securities listed on the Center for Research in Security Prices, CRSP, Compustat, and I/B/E/S databases during the January 1995-December 2020 period. There are a seemingly infinite number of financial variables that may be tested for statistical association with monthly security returns. Bloch, Guerard, Markowitz, Todd, and Xu [13] tested a set of fundamental variables for the U.S. during the 1975-1990 period. The individual variables tested in this study are the original Bloch *et al.* [13] variables, plus I/B/E/S earnings per share forecast variables:

EP = earnings per share/price per share;

BP = book per share/price per share;

CP = cash flow per share/price per share;

SP = sales per share/price per share;

PM = price momentum, or $\text{Price}_{t-1}/\text{Price}_{t-12}$;

FEP1 = one-year-ahead forecast earnings per share/price per share;

FEP2 = two-year-ahead forecast earnings per share/price per share;

RV1 = one-year-ahead forecast earnings per share monthly revision/price per share;

RV2 = two-year-ahead forecast earnings per share monthly revision/price per share;

BR1 = one-year-ahead forecast earnings per share monthly breadth;

BR2 = two-year-ahead forecast earnings per share monthly breadth;

and CTEF=equally-weighted FEP1, FEP2, BR1, BR2, RV1, and RV2.

How does one develop and estimate a stock selection model? One can survey the academic and practitioner literature, as we have done, and calculate information coefficients for the equity universe that one seeks to manage assets within.

There is strong support for fundamental variables (particularly earnings and cash flow), earnings expectations variables, and the momentum variable. An objective examination of the reported ICs leads one to identify CTEF, PM, EP, and CP as leading variables for inclusion in stock selection models.

Bloch, Guerard, Markowitz, Todd, and Xu [13] used an eight-factor model that was previously reported in Equation (2) denoted as REG8.

$$\text{TR}_{t+1} = a_0 + a_1\text{EP}_t + a_2\text{BP}_t + a_3\text{CP}_t + a_4\text{SP}_t + a_5\text{REP}_t + a_6\text{RBP}_t + a_7\text{RCP}_t + a_8\text{RSP}_t + e_{t+1} \quad (2)$$

If one adds price momentum, PM, and the consensus analysts' earnings forecasts and revisions variable, CTEF, to the stock selection model, we can estimate an expanded stock selection model to use as an input to an optimization analysis. See Guerard, Markowitz, and Xu [14] [15]. The stock selection model estimated in this study, often denoted as United States Expected Returns, USER, or REG10, is:

$$\text{TR}_{t+1} = a_0 + a_1\text{EP}_t + a_2\text{EP}_t + a_3\text{EP}_t + a_4\text{EP}_t + a_5\text{REP}_t + a_6\text{RBP}_t + a_7\text{RCP}_t + a_8\text{RSP}_t + a_9\text{CTEF}_t + a_{10}\text{PM}_t + e_{t+1} \quad (3)$$

where: EP = [earnings per share]/[price per share] = earnings-price ratio;
BP = [book value per share]/[price per share] = book-price ratio;
CP =[cash flow per share]/[price per share] = cash flow-price ratio;
SP = [net sales per share]/[price per share] = sales-price ratio;
REP = [current EP ratio]/[average EP ratio over the past five years];
RBP = [current BP ratio]/[average BP ratio over the past five years];
RCP = [current CP ratio]/[average CP ratio over the past five years];
RSP = [current SP ratio]/[average SP ratio over the past five years];
CTEF = consensus earnings-per-share I/B/E/S forecast, revisions and breadth, PM = Price Momentum; and e = randomly distributed error term.

The USER model is estimated using weighted latent root regression analysis on Equation (6) to identify variables statistically significant at the 10% level; uses the normalized coefficients as weights; and averages the variable weights over the past twelve months. The 12-month smoothing is consistent with the four-quarter smoothing in Bloch *et al.* [13]. The WLRR technique produces the largest and most statistically significant IC; a result consistent with the previously noted studies and the GPRD example. The t-statistics on the composite model exceed the t-statistics of its components. The purpose of a composite security valuation model is to identify the determinants of security returns, and produce a statistically significant out-of-sample ranking metric of total returns.[24] A detailed statistical appendix discussing latent root regression was created for Guerard, Takano, and Yamane [16] to address the concern of our referee.

Guerard, Gultekin, and Stone, GGS, [17] studied the intersection of Compustat, CRSP and I/B/E/S databases, building on the fundamental forecasting work in Bloch, Guerard, Markowitz, Todd, and Xu [13] in two ways: (1) adding to the Bloch *et al* eight-variable regression equation a growth measure; and (2) adding three measures of analysts' forecasts and forecast revisions from the I/B/E/S database, namely consensus analysts' forecasts, forecast revisions, and the direction (net up or down) of the forecast revisions. The CTEF variable is more complex and more statistically significant analyst forecast variable than the one-year-ahead and two-year-ahead forecasts and forecast revisions of the I/B/E/S database, as reported in Elton. Gruber, and Gultekin [18] because it includes breadth of the direction of analysts' revisions. The breadth variable does not have significant exposures to the value variable in multi-index models, see Guerard and Mark [19] [20].

Using monthly CRSP data, we estimate REG8 and REG10 models for all CRSP stocks. What do we learn from our CRSP investigation? First, REG8 and REG10 are highly statistically significant model, 1977-2020. We use the Beaton-Tukey Bisquare [21] procedure to identify outliers, as the author did with Harry Marko-

[24]See Latane, Tuttle, and Jones, Security Analysis and Portfolio Management (New York: The Ronald Press, 1975) and Malkiel (1973 and 2023), A Random Walk Down Wall Street (New York: W.W. Norton), First and 50th Anniversary Editions.

witz in 1993 in Bloch *et al.* [13]. The REG8 US model, estimated with data 30 years past its initial estimation was a t-value of 4.10, which is highly statistically significant. The REG10 model has an IC of .052, with an estimated t-statistic of 7.984, again highly statistically significance. See **Table 1**.

Table 1. Information coefficients IC and their t-statistics, averages from 1977:05 to 2020:12.

		Pred	EP/FEP1	BP	CP	SP	REP	RBP	RCP	RSP	CTEF	PM
REG8	IC	0.021	0.029	0.004	0.024	0.000	0.021	−0.018	0.015	−0.023		
	t-IC	4.101	7.893	0.669	6.190	−0.065	6.592	−3.174	4.205	−4.233		
REG10	IC	0.052	0.029	0.004	0.024	0.000	0.021	−0.018	0.015	−0.023	0.042	0.021
	t-IC	7.984	7.893	0.669	6.190	−0.065	6.592	−3.174	4.205	−4.233	13.758	2.563

The reader is referred to Martin, Guerard, and Xia [22] for a presentation of mOpt and more recent robust regression analysis. A definitive reference is Maronna, Martin, Yohai and Salibian-Barrera [23]. Does the predictive power of the REG10 model surpass REG8 model in creating excess returns from optimized portfolios? The authors address this question in sections 3 and 4.

4. Markowitz Mean-Variance Analysis and Alternative Models of Portfolio Selection

Portfolio selection has been recognized by the Nobel Prize committee in Economic Sciences to have begun with Tobin [24], Markowitz [25]-[27] and Sharpe [28] [29]. The Markowitz [25] [26] portfolio construction approach seeks to identify the efficient frontier, the point at which returns are maximized for a given level of risk, or risk is minimized for a given level of return. The portfolio expected return, $E(Rp)$, is calculated by taking the sum of the security weights multiplied by their respective expected returns. The portfolio standard deviation is the sum of the weighted covariances.

$$E\left(R_p\right)=\sum_{i=1}^{N}x_iE\left(R_i\right)=\sum_{i=1}^{N}x_i\mu_i \tag{4}$$

$$\sigma_p^2=\sum_{i=1}^{N}\sum_{j=1}^{N}x_ix_jc_{ij} \tag{5}$$

where μ is the expected return vector, C is the variance-covariance matrix, x is the portfolio weights.

The efficient frontier can be traced out by

$$minimize_{\{x_i\geq 0,x_i\leq \bar{u}\}}\ x^TCx-\lambda\mu^Tx \tag{6}$$

where λ is the risk-return tradeoff parameter and $\bar{u}$ is the fixed upper bound. Bloch, Guerard, Markowitz, Todd, and Xu created efficient frontiers using a purely full historical covariance matrix.

Sharpe [28] developed the Diagonal Model which reduced the computing time of the RAND QP (quadratic programming) code on the IBM 7090 computer from 33 minutes to 30 seconds for a 100-stock example. In the Diagonal Model, Sharpe modeled stock returns are related through common relationships of an underlying factor or index. Sharpe specifically suggested variables such as the level of the stock market, Gross National Product (GNP), or a price index.[25]

The Tobin [24] and Markowitz [25] [26] total risk analysis was divided into systematic and unsystematic risk in the Sharpe [29], Lintner [30], and Mossin [31] Capital Asset Pricing Model, CAPM. The reader is referred to Mossin [32] for the theory of financial markets. In the mid-1970s, Barr Rosenberg pioneered a commercially available multi-factor risk model (MFM) known as the Barra risk model, to create optimized portfolios for institutional investors. Rosenberg documented the Barra model extensively in Rosenberg and Marathe [33]. Rudd and Rosenberg [34] published a paper, Realistic Portfolio Optimization", that enhanced portfolio optimization by introducing transactions costs into the portfolio optimization analysis. Moreover, and more importantly, Rudd and Rosenberg introduced the use of the BARRA multi-factor risk model parameters in lieu of using only the historic covariances among stock returns. The BARRA optimization model was an industry standard and Rudd and Clasing [35] is a standard reference to the modeling system and its application.

The equity factor returns f_k in the Barra U.S. Equity Risk Model, denoted USE3, are estimated by regressing the excess returns r_n against the factor exposures, X_{nk},

$$r_n = \sum_{k=1}^{K_E} X_{nk} f_k + u_n \tag{6}$$

USE3 uses weighted least squares, assuming that the variance of specific returns is inversely proportional to the square root of total market capitalization. The USER model is our approximation of the expected return, or the forecast active return, α, of the portfolio.[26] Researchers in industry most often apply the Markowitz mean/variance framework to active management, as described in Grinold and Kahn [36].

$$U = \alpha \cdot h - \lambda \cdot \omega^2 \cdot h^2 \tag{7}$$

[25]There is extensive literature on the impact of individual value ratios on the cross section of stock returns. We go beyond using just one or two of the standard value ratios (EP and BP) to include the cash-price ratio (CP) and/or the sales-price ratio (SP). The major practitioner papers on combination of value ratios to predict stock returns (that include at least CP and/or SP) include Jacobs and Levy (1988 and 2017), and Guerard, Gultekin, and Stone (1997). We reviewed these papers in some detail, see Guerard, The Leading Economic Indicators and Business Cycles in the United States (Switzerland: Palgrave macmillan, 2022).

[26]The USER score was converted into an expected returned by a five-step process at each point in time: (1) each asset's residual volatility (residual to the benchmark) is calculated and scaled to a monthly figure; (2) the asset's raw score is divided by its monthly residual volatility, creating a score over volatility; (3) the mean and standard deviation of the score over volatility are calculated; (4) the score over volatility is normalized, in which the raw score over volatility, less its mean, is divided by its standard deviation; and (5) the expected excess return, or alpha, is calculated by multiplying the annualized multiplier (12) times its monthly residual volatility times the information coefficient times the standard deviation of the score over volatility. The five-step process creates an expected return consistent with Grinold and Kahn "Fundamental Law of Active Management," Chapter 6.

Here α is the forecast active return (relative to a benchmark which can be cash), ω is the active risk, and h is the active holding (the holding relative to the benchmark holding). By varying the tolerance or risk-version, one can create the efficient frontier using the Barra Optimizer, as was done in Bloch [13], by changing the variable m. The risk aversion parameter, λ, captures individual investor preference.[27] Grinold and Kahn [36] introduce the Information Ratio (IR) as a portfolio construction objective to be maximized, which measures the ratio of residual return to residual risk:

$$IR = \frac{\alpha}{\omega} \tag{8}$$

There are several commercially available multi-factor risk models. The Advanced Portfolio Technologies (APT) system of Blin and Bender, documented **in** Blin, Bender, and Guerard [37] and Blin, Guerard, and Mark [38] was an outstanding system for portfolio implementation. The Axioma Robust Risk Model[28] is a multi-factor risk model, in the tradition of the Barra model. Axioma offers both U.S. and world fundamental and statistical risk models. The Axioma Risk Models use several statistical techniques to efficiently estimate factors. The ordinary least squares residuals (OLS) of beta estimations are not of constant variance; that is, when one minimizes the sum of the squared residuals to estimate factors using OLS, one finds that large assets exhibit lower volatility than smaller assets. Axioma uses a weighted least squares (WLS) regression, which scales the asset residual by the square root of the asset market capitalization (to serve as a proxy for the inverse of the residual variance). Robust regression, using the Huber M Estimator, addresses the issue and problem of outliers. (Asymptotic) Principal components analysis (PCA) is used to estimate the statistical risk factors. A subset of assets is used to estimate the factors and the exposures and factor returns are applied to other assets. Guerard, Markowitz, and Xu [15] tested CTEF and a ten-factor regression-based model of global expected returns, GLER, during the 1997-2011 time period. The authors reported that the geometric means and Sharpe ratios increase with the targeted tracking errors; however, the information ratios are higher in the lower tracking error range of 3% - 6%, with at least 200 stocks, on average, in the optimal portfolios. They reported that statistically-based risk models using principal components, such as Sungard APT and Axioma, produce more efficient trade-off curves than fundamentally-based risk model using our varia-

[27]The Markowitz (1976) Geometric Mean maximization is consistent with the work of Fisher (1966a). Miller, Xu, and Guerard (2013) constructed an Efficient Frontier varying the risk-aversion levels. Their portfolio construction process uses 8 percent monthly turnover, after the initial portfolio is created, and 125 basis points of transactions costs each way. The USER-optimized portfolios outperform the market, here defined as the Russell 300 Growth Index. The portfolio that maximizes the Geometric Mean (Markowitz, 1976) occurs at a risk-aversion level of 0.02. The Sharpe Ratio is also maximized at a risk-aversion level of 0.02 with 89 stocks in the efficient portfolio. The Information Ratio, defined as the ratio of portfolio excess return relative to estimated tracking error, is maximized at a risk-aversion level of 0.01 (0.80). Asset selection of 749 basis points was produced with an estimated t-statistical of 3.28, highly statistically significant. Asset selection was accompanied by positive and statistically significant small stock exposures and a higher earnings yield exposure. The size and earnings yield exposures were the only statistically significant BARRA multi-factor risk exposures, Overall, the portfolio total managed returns came from asset selection.

[28]*Axioma Robust Risk Model Handbook*, January 2010.

bles.[29] See Blin, Guerard, and Mark [38] for a more complete discussion of commercially available risk models. Furthermore, Guerard, Xu, and Markowitz [39] reported that the WLRR applications passed the Markowitz-Xu [40] data mining corrections test for MSCI All Country World (ACW) constituent universes. The statistically significant results were not the product of data mining. Guerard and Mark validated the continued statistical significance of CTEF and REG9 using Axioma Medium-Term US Model, version 3.[30]

Markowitz, Guerard, and Xu (2019) reported at the Q-Group meeting in La Jolla, and Markowitz, Guerard, Xu, and Beheshti [40] reported that CTEF, the most statistically significant model used for stock selection, published in 1997, has been one of the 2 - 3 most dominant variables in the MCM in all MSCI universes up till September 2019.[31] The authors used data only as it was known (or more exactly, our portfolios were tested out-of-sample). Their simulation conditions assumed 8 percent monthly turnover, 35 basis point threshold positions, an upper bound in Mean-Variance optimization of 4 percent on security weights, and ITG

[29]John Blin, Steve Bender, and Blin, Guerard, and Mark (2022) demonstrated the effectiveness of the APT, Sungard APT, and FIS APT systems in portfolio construction and management. The estimation of security weights, w, in a portfolio is the primary calculation of Markowitz's portfolio management approach. The issue of security weights will be now considered from a different perspective. The security weight is the proportion of the portfolio's market value invested in the individual security. The active weight of the security is calculated by subtracting the security weight in the (index) benchmark, b, from the security weight in the portfolio, p. The marginal security systematic volatility is the partial derivative of the systematic volatility of the portfolio relative to the security weight. In the King's English, the marginal tracking error measures the sensitivity of the tracking error relative to the marginal change in the security active weight. If a position taken in a security leads to an increase in the portfolio's volatility, then the security is said to create a positive contribution to risk. A negative contribution to risk occurs when a security reduces the portfolio volatility such as a long position on a security with a negative beta or a short position on a security with a positive beta. Obviously, the contribution to risk depends upon the security weight and the security's beta to the overall portfolio. The security contribution to tracking error, reflects the security's contribution to the tracking error of a portfolio considering the security return that is undiversified at the active portfolio level. The portfolio Value-at-Risk (VaR) is the expected maximum loss that a portfolio could produce over one year. The APT measure of portfolio risk estimating the magnitude that the portfolio return may deviate from the benchmark return over one year is referred to as TaR, or "Tracking-at-Risk". TaR is composed of systematic and specific components. What is the economic importance of tracking error at risk? First, TaR helps the assert manager assess downside risk. Second, by optimizing portfolios where systematic risk is more important than specific risk, one produces high Information Ratios, IRs, than equally-weighting systematic and specific risk or using only total risk (Markowitz, 1959). TaR specifically addresses fat tails in stock return distributions. Third, as portfolios becomes diversified, the R-squared statistics of portfolio returns rise, and the optimal TaR ratio to relative tracking errors rise, to 1.645 (unsystematic risk is weighted 0.345). Guerard, Rachev and Shao (2013) and Guerard, Markowitz, and Xu (2015) reported the highly statistically significant excess returns (and specific returns) effectiveness of an APT MVTaR optimization analysis of CTEF in global markets during the 1997-2011 time period and Guerard, Markowitz, and Xu (2014) reported CTEF effectiveness in U.S. markets over the corresponding time period. See Blin, Guerard, and Mark (2022) for a summary of commercially available risk models. Guerard, Rachev and Shao (2013) and Guerard, Markowitz, and Xu (2015) reported the highly statistically significant excess returns (and specific returns) effectiveness of an APT MVTaR optimization analysis of CTEF in global markets during the 1997-2011 time period and Guerard, Markowitz, and Xu (2014) reported CTEF effectiveness in U.S. markets over the corresponding time period.

[30]Guerard and Mark (2020) reported monthly Axioma attribution statistics which, in the case of CTEF, indicates that the forecasted earnings acceleration variable loads on Medium-Term Momentum (0.257), Growth (0.151), and Value (0.469) and that Mean-variance CTEF and REG10 portfolios produced approximately 300-350 basis points of Specific Returns for the 20-year time periods. In the U.S. portfolios, equally-weighted 125 stock portfolios outperform Mean-variance (MV) four percent portfolios. In the Non-U.S. and EAFE universes, Guerard and Mark (2018) reported that the CTEF ICs were higher than the REG10 or GLER ICs in their 10, 5, 3, and one-year time sub-periods. The CTEF and REG10 produced approximately 400 - 500 basis points of Active Returns and about 250 basis points of Specific Returns. The Non-US portfolios offer more stock selection than U.S. portfolios with the addition of the REG8 plus CTEF (denoted REG9) and REG10 factors. The t-statistic on the risk stock selection effect in Non-US portfolios is maximized with ranked CTEF. The t-statistics on the risk stock selection effect is statistically significant for REG10, although the t-statistic on the risk stock selection effect in the Non-US portfolios is only statically significant at the 10 percent level. Guerard and Mark (2018) reported that only ranked CTEF is statistically significant in the U.S. whereas globally, ranked CTEF and REG10 are statistically significant in Total Active Returns and Risk Stock Selection Returns.

[31]The models discussed in Guerard and Markowitz (2019), written for 84,000 Certified Financial Planners, stressed that the models identified by the authors well before the 2003-11/2018 were highly statistically significant post-publication. The models that passed the Markowitz-Xu (1994) Data Mining Corrections test in 1994 and 1997, continued to be statistically significant for another 25 years.

transactions costs.[32] We use the Mean-Variance (MV) optimization techniques. Our portfolio approximates index benchmarks on 15 dimensions, domestically, and 20 dimensions, globally. The financial anomalies of EP, BP, CP, SP, CTEF, and PM were reflected in the REG10 (or USER) reported results. In the CTEF-optimized portfolios outperform in 70% and 77% of the years, respectively. Financial anomalies, as published in 1993, 2003, and 2021 continue to outperform, with slightly reduced winning percentages, from 83% in 1993 to 77% in 2018. Is that enough for investors? Should that be enough for investors? The excess returns reported in Markowitz, Guerard, Xu, and Beheshti [40], about 400 basis points, annually, particularly for targeted tracking errors of 6 and 8 percent and the Axioma risk model, should satisfy investors who seek to maximize the Geometric Mean and the utility of terminal wealth.[33]

5. Markowitz Mean-Variance Analysis and Alternative Models of Portfolio Selection in CRSP Data, 1994-2020

In optimized portfolios, using a 20% quarterly turnover, the REG8, REG10, and CTEF portfolios were constructed using a four percent Equal Active Weight, EAW4, and produced excess returns, total effects, subtracting 125 basis points of round-trip transactions costs, in the range of 300 - 700 basis points, annualized, for CTEF, REG8, and REG10 portfolios. The reader is referred to **Table 2**. The WLRR and forecasted earnings portfolios. portfolios of 1993 work to outperform the benchmark during the 1994-2020 period. Sharpe and Information Ratios often rise with higher tracking errors 6% - 8%, as opposed to 2% - 4%, more of an index-enhanced level. Both stock selection (the risk stock specific effect) is highly statistically significant, as far the risk factor returns. Higher earnings yield and size, smaller stocks), are associated with the higher factor risk returns, as noted earlier in Guerard, Thomakos, Foteini, and Beheshti [41].

The EAW4 IRs are very similar to the model IRs reported in Martin, Guerard, and Xia [42] [43]. The EAW4 portfolios substantially outperform the EAW2 portfolios of Guerard *et al.* [41] in Sharpe Ratios, and by 300 - 750 basis points, annualized, after transaction costs. The authors have always pushed to maximize the Sharpe Ratios and geometric means, as advocated in Markowitz [26], Latane [44] [45], and Bloch *et al.* [13]. Why do we report the EAW4 results? The stock selec-

[32]ITG estimated our transactions costs to be about 60 basis points, each-way, for 2011-2015.

[33]Fama (1991) hypothesized that anomalies could not be effectively tested because of changing asset pricing models and the number of factors in multi-factor models. Barillas and Shanken (2018) addressed the issue of changing risk models and factors. We have addressed these issues in two previous studies and are currently addressing a third issue. First in Guerard and Mark (2003), we based what BARRA USE models were known at that time; there was no look-ahead bias in risk models. Second, in Guerard and Mark (2018) we used the Boolean signal portfolio construction process in which we buy attractively ranked stocks and sell them when they fell through pre-determined level (buy stocks with CTEF or REG10 scores of at least 85 and higher, where 0 is least preferred and 99 is most preferred and sell at 70, holding in equally-weighted portfolios). The Boolean signal teats re-enforced the mean-variance portfolio results. The mean-variance portfolios, with a specified upper bound on stock weights and positive holdings for long-only portfolios produced very reasonable (possible) weights. We do not believe the Brennan-Lo (2011) footnote that repeats the impossible mean-variance optimized portfolios tales told by Wall Street portfolio managers. The Axioma attribution of the Boolean signal portfolios attributed all Active Returns to stock selection in the CTEF model and the majority of the REG10 model Active Returns were due to Specific Returns, or asset selection. Finally, we test are testing whether it makes a difference as to whether we use a (1) mean-variance tracking error at risk model, stressing systematic risk minimization; (2) a mean-variance model without factors (MVM59), using only total risk; or (3) a mean-variance model using only systematic risk.

tion effects are statistically significant in an EAW4 model whereas stock selection was not statistically significant in the EAW2 portfolio in Guerard *et al.* [41]. Furthermore, Harry Markowitz told me in 2013 that he was my Chief Investment Officer "You need to put on your big boy investment pants when you work with John's Weighted Latent Root Regression model. It is not perfect, and it works in 75% of the years (beating the benchmark), and, on average, and over the long run, you will win!" The stock weights are the primary decision variables in Markowitz analysis. If you have confidence in your market, then invest larger percentages of the portfolio in higher expected return assets.

Table 2. (a) CRSP Universe Optimized vs the US R3000 Dashboard, 1994-2020. (b) Portfolio Selection FactSet Risk Model Multi-Factor Risk Factor Returns.

(a)

Portfolio Weights: EAW4

Portfolios	Risk Stock Specific Effect	Risk Stock Specific Effect T-Stat	Risk Factors Effect	Risk Factors Effect T-Stat	Total Effect
US_R3000_CTEF_2TE	5.11	13.47	1.03	5.05	6.14
US_R3000_CTEF_4TE	5.53	7.11	1.09	3.95	6.62
US_R3000_CTEF_6TE	6.20	5.44	1.08	3.36	7.28
US_R3000_CTEF_8TE	6.23	4.24	0.54	2.75	6.77
US_R3000_REG10_2TE	4.82	13.81	0.60	2.92	5.41
US_R3000_REG10_4TE	5.53	7.54	0.54	2.72	6.07
US_R3000_REG10_6TE	6.63	5.88	0.11	2.48	6.75
US_R3000_REG10_8TE	7.43	5.42	−0.21	2.08	7.22
US_R3000_REG8_2TE	5.91	16.80	−0.34	1.23	5.57
US_R3000_REG8_4TE	7.63	9.95	−1.43	1.07	6.19
US_R3000_REG8_6TE	7.77	6.87	−2.26	0.56	5.51
US_R3000_REG8_8TE	5.78	4.42	−2.68	0.02	3.09

(b)

Portfolios	Earnings Yield	Medium-Term Momentum	Size	Value	Sharpe Ratio	Information Ratio	Historical Tracking Error
US_R3000_CTEF_2TE	0.80	0.53	0.78	0.10	0.730	0.263	4.21
US_R3000_CTEF_4TE	1.28	0.94	1.56	0.17	0.702	0.158	7.02
US_R3000_CTEF_6TE	1.64	1.35	2.22	0.20	0.688	0.122	9.23
US_R3000_REG10_2TE	0.97	0.01	1.11	0.50	0.675	0.260	4.20
US_R3000_REG10_4TE	1.41	0.07	2.01	0.82	0.672	0.156	7.09
US_R3000_REG8_2TE	0.79	−0.76	1.16	0.58	0.667	0.230	4.77
US_R3000_REG8_4TE	0.95	−1.26	2.04	0.86	0.658	0.143	7.73
US_R3000_REG10_6TE	1.61	0.12	2.55	1.04	0.655	0.116	9.57
US_R3000_REG10_8TE	1.69	0.16	2.47	1.21	0.655	0.103	10.93
US_R3000_CTEF_8TE	1.80	1.48	2.61	0.33	0.617	0.099	11.27
US_R3000_REG8_6TE	0.76	−1.47	2.54	0.98	0.571	0.105	10.44
US_R3000_REG8_8TE	0.39	−1.55	−0.49	2.98	0.403	0.081	12.790

BOLD denotes statistically significant stock selection (RiskStock Specific Effect) and Risk Factors Effect. Green denotes Enhancing Factors; Red denotes Lesser Factors; Yellow denotes Indifferent Factors.

Did we get "lucky"? No, the Shao, Guerard, and Xu [46] estimated the Markowitz and Xu [40] data mining corrections model and reported that both Mean-Variance (MV) and Mean-Expected Tail Loss (M-ETL) sets of 16 portfolios, constructed using our CRSP database, generated excess returns that were statistically different than the average estimated portfolio. The reader should read Harvey, Lin, and Zhu [47] and Harvey [48] for a recent outstanding analyses of expected returns modeling and data mining testing.

In this section, the authors have updated Bloch *et al.* [13] and Guerard, Markowitz, and Xu [14] [15]. Harry Markowitz, who passed in 2023, created much of the field of financial economics. The reader is referred to Guerard [49] for a brief tribute to Mr. Markowitz, and to Guerard, Thomakos, Kyriazi, and Beheshti [41] and Martin, Guerard, and Xia [43] for 30-year and 40-year verifications of Bloch *et al.* [13] using CRSP and FactSet data, respectively. The authors believe that Financial Engineering, a subset of Financial Economics, has advanced because of many studies of the stocks in the CRSP database. The database is of great benefit to researchers and special thanks are made to Lawrence Fisher and James Lorie, who created the database with their 1960 Merrill Lynch grant. Conner, Goldberg, and Koraczyk [50] specifically state that Financial Engineering does not use of portfolio risk analysis, and the authors believe that they are correct. Financial Engineering perhaps uses the more commercially advantageous aspects of the Markowitz analysis. WLRR is a form of Machine Learning. Shu and Mulvey [51] and de Prado, Simonian, Fabozzi, and Fabozzi [52] report recent advances in statistical jump modeling in Regime Forecasting and Machine Learning in Financial Engineering, respectively.

6. Conclusions

NBER business cycle research was associated with the expectations of profits at its creation by Professor Wesley Clair Mitchell in 1927. Investing with fundamental data, particularly earnings. expectations, and momentum variables is a good investment strategy over the long-run. Stock selection models often incorporate momentum, analysts' expectations, and fundamental data. Earnings, historical, and expected, must be considered. We find support for composite modeling using these sources of data. We celebrate a 40-year-backtest of fundamental and expectational data using the CRSP database, created by Professors Fisher and Lorie in 1960. We find additional evidence to support the use of MSCI Barra and APT multi-factor models for portfolio construction and risk control, as initially reported in Bloch *et al.* (1993), Blin, Bender, and Guerard (1997), Miller, Xu, and Guerard (2013), and Guerard, Markowitz, and Xu (2014). The inclusion of consensus analysts' forecasts, revisions, and breadth variable has become more important in picking global and U.S. stocks. Robust regression models of fundamental factors and earnings forecasting factors continue to enhance portfolio returns. We report that the Markowitz mean-variance optimization is particularly efficient for producing efficient frontiers using a forecasted earnings acceleration model, CTEF, and a composite, robust-regression based ten-factor model, REG10. Have markets and stock selection models changed since Bloch, Guerard, Markowitz, Todd and Xu (1993) Miller, Xu, and Guerard [53], Markowitz and Guerard [54], and Ziemba and Schwartz [55] published their studies? Yes and no, and REG8, and REG10 still dominate most other models. The 10-factor models continue to pass the Markowitz-Xu (1994) Data Mining Corrections Tests and more recent tests. The uses of multi-factor risk-controlled portfolio returns allow us to reject the data mining corrections test null hypothesis. The anomalies literature can be applied in real-world portfolio construction in academia, with the CRSP database, and in industry, with FactSet. It has paid to be bullish in equity investing. The stock weights are the primary decision variables in Markowitz analysis. If you have confidence in your model in its market of application, then invest larger percentages of the portfolio in higher expected return assets. For the past 40 - 50 years, active returns have been earned with more aggressive equity strategies.[34]

Acknowledgements

The corresponding author acknowledges access to the WRDS, Wharton Research and Data Services, data, used in this study, provided for the Computational Finance and Risk Management (CFRM) 502 and 509 classes, the University of Washington, 2020-2021.

We appreciate the comments of (the late) Harry Markowitz, Ganlin Xu, and DeWitt Miller on earlier versions of this paper. Ganlin Xu was kind enough to have read several versions. Campbell Harvey suggested a substantial reduction to the Mitchel NBER analysis and the necessity of including CRSP data event studies. Any errors remaining are the responsibility of the corresponding author.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Fisher, L. and Lorie, J.H. (1964) Rates of Return on Investments in Common Stocks. *The Journal of Business*, **37**, 1-21. https://doi.org/10.1086/294647

[2] Fisher, L. (1966) Some New Stock-Market Indexes. *The Journal of Business*, **39**, 191-225. https://doi.org/10.1086/294848

[34]Fama (1991) hypothesized that anomalies could not be effectively tested because of changing asset pricing models and the number of factors in multi-factor models. We have addressed this issue in two previous studies and are currently addressing a third issue. First in Guerard and Mark (2003), we based what BARRA USE models were known at that time; there was no look-ahead bias in risk models. Second, in Guerard and Mark (2022) we used the Boolean signal portfolio construction process in which we buy attractively ranked stocks and sell them when they fell through pre-determined level (buy stocks with CTEF or REG10 scores of at least 85 and higher, where 0 is least preferred and 99 is most preferred and sell at 70, holding in equally-weighted portfolios). The Boolean signal teats re-enforced the mean-variance portfolio results. The mean-variance portfolios, with a specified upper bound on stock weights and positive holdings for long-only portfolios produced very reasonable (possible) weights. The Axioma attribution of the Boolean signal portfolios attributed all Active Returns to stock selection in the CTEF model and the majority of the REG10 model Active Returns were due to Specific Returns, or asset selection. Finally, we test are testing whether it makes a difference as to whether we use a (1) mean-variance tracking error at risk model, stressing systematic risk minimization; (2) a mean-variance model without factors (MVM59), using only total risk; or (3) a mean-variance model using only systematic risk.

[3] Lorie, J., Dodd, P. and Hamilton Kimpton, M. (1985) The Stock Market: Theories and Evidence. 2nd Edition, Richard D. Irwin, Inc.

[4] Mitchell, W.C. (1913) Business Cycles. Burt Franklin Reprint.

[5] Mitchell, W.C. (1927) Business Cycles: The Problem and Its Setting. NBER.

[6] Mitchell, W.C. (1951) What Happens during Business Cycles: A Progress Report. NBER.

[7] Burns, A.F. and Mitchell, W.C. (1946) Measuring Business Cycles. NBER.

[8] Graham, B. and Dodd, D. (1934) Security Analysis: Principles & Technique. McGraw-Hill Book Company.

[9] Graham, B. and Dodd, D. (1951) Security Analysis: Principles & Technique. McGraw-Hill Book Company.

[10] Graham, B., Dodd, D., Cottle, S., Block, E. and Murray, R.F. (1988) Security Analysis: Principles & Technique. 5th Edition, McGraw-Hill Book.

[11] Williams, J.B. (1938) The Theory of Investment Value. Harvard University Press.

[12] Markowitz, H.M. (1991) Foundations of Portfolio Theory. *The Journal of Finance*, **46**, 469-477. https://doi.org/10.1111/j.1540-6261.1991.tb02669.x

[13] Bloch, M., Guerard, J., Markowitz, H., Todd, P. and Xu, G. (1993) A Comparison of Some Aspects of the U.S. and Japanese Equity Markets. *Japan and the World Economy*, **5**, 3-26. https://doi.org/10.1016/0922-1425(93)90025-y

[14] Guerard, J.B., Markowitz, H. and Xu, G. (2014) The Role of Effective Corporate Decisions in the Creation of Efficient Portfolios. *IBM Journal of Research and Development*, **58**, Paper 11. https://doi.org/10.1147/jrd.2014.2326591

[15] Guerard, J.B., Markowitz, H. and Xu, G. (2015) Earnings Forecasting in a Global Stock Selection Model and Efficient Portfolio Construction and Management. *International Journal of Forecasting*, **31**, 550-560. https://doi.org/10.1016/j.ijforecast.2014.10.003

[16] Guerard, J.B., Takano, M. and Yamane, Y. (1993) The Development of Efficient Portfolios in Japan with Particular Emphasis on Sales and Earnings Forecasting. *Annals of Operations Research*, **45**, 91-108. https://doi.org/10.1007/bf02282043

[17] Guerard, J.B., Gultekin, M. and Stone, B.K. (1997) The Role of Fundamental Data and Analysts' Earnings Breadth, Forecasts, and Revisions in the Creation of Efficient Portfolios. In: Chen, A., Ed., *Research in Finance*, Vol. 15, Emerald JAI, 69-91.

[18] Elton, E.J., Gruber, M.J. and Gultekin, M. (1981) Expectations and Share Prices. *Management Science*, **27**, 975-987. https://doi.org/10.1287/mnsc.27.9.975

[19] Guerard, J.B. and Mark, A. (2003) The Optimization of Efficient Portfolios: The Case for an R & D Quadratic Term. In: Chen, A., Ed., *Research in Finance*, Vol. 20, JAI Press, 213-247.

[20] Guerard, J.B. and Mark, A. (2022) Earnings Forecasts and Revisions, Price Momentum, and Fundamental Data: Further Explorations of Financial Anomalies. In: Lee, C.F., Eds., *Handbook of Financial Econometrics*, *Mathematics*, *Statistics*, *and Machine Learning*, World Scientific Publishing, 1151-1209.

[21] Beaton, A.E. and Tukey, J.W. (1974) The Fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectroscopic Data. *Technometrics*, **16**, 147-185. https://doi.org/10.1080/00401706.1974.10489171

[22] Martin, R.D., Guerard, J.B. and Xia, D.Z. (2024) Earnings to Price Analysis with mOpt versus Bisquare Robust Regression. *Journal of Mathematical Finance*, **14**, 243-249. https://doi.org/10.4236/jmf.2024.142014

[23] Maronna, R.A., Martin, R.D., Yohai, V.J. and Salibian-Barrera, M. (2019) Robust Statistics; Theory & Methods with R. Wiley.

[24] Tobin, J. (1958) Liquidity Preference as Behavior Towards Risk. *The Review of Economic Studies*, **25**, 65-86. https://doi.org/10.2307/2296205

[25] Markowitz, H. (1952) Portfolio Selection. *The Journal of Finance*, **7**, 77-91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x

[26] Markowitz, H.M. (1959) Portfolio Selection: Efficient Diversification of Investment. Cowles Foundation Monograph No. 16, John Wiley & Sons.

[27] Markowitz, H. (1987) Mean-Variance Analysis in Portfolio Choice & Capital Markets. Basil Blackwell.

[28] Sharpe, W.F. (1963) A Simplified Model for Portfolio Analysis. *Management Science*, **9**, 277-293. https://doi.org/10.1287/mnsc.9.2.277

[29] Sharpe, W.F. (1964) Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance*, **19**, 425-442. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x

[30] Lintner, J. (1965) The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics*, **47**, 13-37. https://doi.org/10.2307/1924119

[31] Mossin, J. (1966) Equilibrium in a Capital Asset Market. *Econometrica*, **34**, 768-783. https://doi.org/10.2307/1910098

[32] Mossin, J. (1977) The Economic Efficiency of Financial Markets. Lexington Books.

[33] Rosenberg, B. and Marathe, V. (1979) Tests of Capital Asset Pricing Hypotheses. In: Levy, H., Ed., *Research in Finance*, 115-224.

[34] Rudd, A. and Rosenberg, B. (1979) Realistic Portfolio Optimization. In: Elton, E. and Gruber, M.J., Eds., *Portfolio Theory*, 25 *Years after*, North-Holland, 21-46.

[35] Rudd, A. and Clasing, H.K. (1982) Modern Portfolio Theory: The Principals of Investment Management. Dow-Jones-Irwin.

[36] Grinold, R. and Kahn, R. (1999) Active Portfolio Management. McGraw-Hill/Irwin.

[37] Blin, J.M., Bender, S. and Guerard Jr. (1997) Earnings Forecasts, Revisions & Momentum in the Estimation of Efficient Market-Neutral Japanese & U.S. Portfolios. In: Chen, A., Ed., *Research in Finance*, Vol. 15, JAI Press, 93-114.

[38] Blin, J., Guerard, J. and Mark, A. (2022) A History of Commercially Available Risk Models. In: Lee, C.F., Ed., *Encyclopedia of Finance*, Springer International Publishing, 1-38. https://doi.org/10.1007/978-3-030-73443-5_99-2

[39] Guerard, J.B., Xu, G. and Markowitz, H. (2020) A Further Analysis of Robust Regression Modeling and Data Mining Corrections Testing in Global Stocks. *Annals of Operations Research*, **303**, 175-195. https://doi.org/10.1007/s10479-020-03521-y

[40] Markowitz, H.M. and Xu, G.L. (1994) Data Mining Corrections. *The Journal of Portfolio Management*, **21**, 60-69. https://doi.org/10.3905/jpm.1994.409494

[41] Guerard, J.B., Thomakos, D., Kyriazi, F., Xu, G. and Beheshti, B. (2025) Earnings Forecasting and Mean-variance Efficient Portfolios in the US and Japan: 30 Years After the Markowitz and Ziemba Applications. *Annals of Operations Research*, **346**, 393-414. https://doi.org/10.1007/s10479-024-06432-4

[42] Martin, R.D., Guerard, J.B. and Xia, D.Z. (2024) Earnings to Price Analysis with mOpt versus Bisquare Robust Regression. *Journal of Mathematical Finance*, **14**, 243-249. https://doi.org/10.4236/jmf.2024.142014

[43] Martin, R.D., Guerard, J.B. and Xia, D.Z. (2024) Resurrecting Earnings-to-Price with

Robust Control for Outliers. *The Journal of Portfolio Management*, **50**, 60-81. https://doi.org/10.3905/jpm.2024.1.626

[44] Latané, H.A. (1959) Criteria for Choice among Risky Ventures. *Journal of Political Economy*, **67**, 144-155. https://doi.org/10.1086/258157

[45] Latane, H.A., Tuttle, D.L. and Jones, C.P. (1975) Security Analysis and Portfolio Management. The Ronald Press Company.

[46] Shao, B.P., Guerard, J.B. and Xu, G. (2024) Mean-Variance and Mean-Etl Optimizations in Portfolio Selection: An Update. *Annals of Operations Research*, **346**, 657-671. https://doi.org/10.1007/s10479-024-06337-2

[47] Harvey, C.R., Liu, Y. and Zhu, H. (2015) … and the Cross-Section of Expected Returns. *Review of Financial Studies*, **29**, 5-68. https://doi.org/10.1093/rfs/hhv059

[48] Harvey, C.R. (2017) Presidential Address: The Scientific Outlook in Financial Economics. *The Journal of Finance*, **72**, 1399-1440. https://doi.org/10.1111/jofi.12530

[49] Guerard, J. (2023) Harry Markowitz: An Appreciation. *International Journal of Forecasting*, **39**, 1496-1501. https://doi.org/10.1016/j.ijforecast.2023.07.004

[50] Conner, G., Goldberg, L.R. and Korajczyk, R. (2010) Portfolio Risk Analysis. Princeton University Press.

[51] Shu, Y., Yu, C. and Mulvey, J.M. (2024) Dynamic Asset Allocation with Asset-Specific Regime Forecasts. *Annals of Operations Research*, **346**, 285-318. https://doi.org/10.1007/s10479-024-06266-0

[52] López de Prado, M., Simonian, J., Fabozzi, F.A. and Fabozzi, F.J. (2024) Enhancing Markowitz's Portfolio Selection Paradigm with Machine Learning. *Annals of Operations Research*, **346**, 319-340. https://doi.org/10.1007/s10479-024-06257-1

[53] Miller, D., Xu, G. and Guerard Jr., J.B. (2014) Portfolio Construction and Management in the Barra Aegis System: A Case Study using the USER Data. *Journal of Investing*, **23**, 111-120. https://doi.org/10.3905/joi.2014.23.4.111

[54] Markowitz, H.M. and Guerard Jr., J.B. (2018) The Existence and Persistence of Financial Anomalies: What Have You Done for Me Lately? *Financial Planning Review*.

[55] Ziemba, W.T. and Schwartz, S.L. (1992) Invest Japan. Probus Publishing Company.