The Cross-Section of Realized Stock Returns: The Pre-COMPUSTAT Evidence
Author(s): James L. Davis
Source: *The Journal of Finance*, Vol. 49, No. 5 (Dec., 1994), pp. 1579-1593
Published by: Blackwell Publishing for the American Finance Association
Stable URL: http://www.jstor.org/stable/2329263
Accessed: 25/11/2009 11:24

THE JOURNAL OF FINANCE • VOL. XLIX, NO. 5 • DECEMBER 1994

The Cross-Section of Realized Stock Returns: The Pre-COMPUSTAT Evidence

JAMES L. DAVIS*

ABSTRACT

Using a database that is free of survivorship bias, this article finds that book-to-market equity, earnings yield, and cash flow yield have significant explanatory power with respect to the cross-section of realized stock returns during the period from July 1940 through June 1963. There is a strong January seasonal in the explanatory power of these variables, even though small stocks are, by construction, excluded from the sample.

SEVERAL RECENT STUDIES HAVE documented the ability of certain variables to explain the cross-sectional variation in realized stock returns. Among these studies are Rosenberg, Reid, and Lanstein (1985), De Bondt and Thaler (1987), Chan, Hamao, and Lakonishok (1991), and Fama and French (1992), all of which find a relationship between book-to-market equity and stock returns. Other studies have shown variables such as earnings yield (see Basu (1977) and Jaffe, Keim, and Westerfield (1989)), cash flow yield (see Chan, Hamao, and Lakonishok (1991)), and historical sales growth (see Lakonishok, Shleifer, and Vishny (1993)) to be related to subsequent returns. Some studies find evidence of a January seasonal in the explanatory power of certain variables. For example, Jaffe, Keim, and Westerfield (1989, Table VI) find evidence of seasonality in the explanatory power of earnings yield, and Fama and French (1992, page 448) report a January seasonal in the magnitude of the regression coefficient on book-to-market equity.

A number of explanations for the observed predictive ability have been suggested. One explanation is that certain fundamental variables are measuring the riskiness of stocks, so that the correlation between the variables and subsequent returns reflects compensation for bearing risk (see Fama and French (1993)). Others have argued that the variables allow investors to identify stocks that are mispriced, thus creating opportunities for realized returns in excess of what is required to compensate investors for risk (see Lakonishok, Shleifer, and Vishny (1993)). A third hypothesis is that the observed predictive ability is an artifact of the research design and database

* Kansas State University. This article is based on part of my doctoral thesis at the University of Illinois. I am grateful to Louis Chan, Josef Lakonishok, Jay Ritter, René Stulz (the editor), two anonymous referees, and seminar participants at Kansas Sate University and the University of Illinois for their valuable comments. I am especially indebted to Jay Ritter for many helpful discussions.

used to conduct the study, and the predictive ability of certain variables would be reduced or vanish if different methodology and data were used (see Kothari, Shanken, and Sloan (1993)).

The present study addresses some of the methodological issues that have been raised in connection with this third hypothesis. Using a dataset that eliminates some of these concerns and seeks to minimize the effects of others, the analysis finds significant relationships between certain fundamental variables (such as book-to-market equity, cash flow yield, and earnings yield) and subsequent returns during the period from July 1940 to June 1963. The results suggest that the first two hypotheses (or a combination thereof) are likely candidates for explaining the observed predictive ability, rather than the third hypothesis.

Variables such as book-to-market equity, cash flow yield, and earnings yield are calculated using information that appears in a firm's financial statements. Accordingly, the availability of accounting information is crucial to studies that use these variables. Since Standard & Poor's COMPUSTAT service provides a readily available, machine readable source of accounting data for a large number of companies, most studies use the COMPUSTAT data to calculate the above-mentioned ratios.

The use of the COMPUSTAT data introduces at least four potential problems into any study of the behavior of stock prices. First, there is survivorship bias in the selection of firms that are included in the COMPUSTAT database. (See Davis (1994) for evidence that this problem is most serious in data for the period prior to the 1970s.) As shown in Banz and Breen (1986), this can have a significant effect on the results of studies that relate accounting data to stock returns. Second, Banz and Breen (1986) also discuss the problem of a look-ahead bias: the study may assume that information was available to investors at a point in time, when in fact the information was not yet publicly available. For example, it is not always appropriate to assume that a firm's December 31 financial statements are disseminated to the public by the following March 31 or April 30.

A third problem associated with the existing literature is related to the concept of data snooping (see Lo and MacKinlay (1990)). Any study that uses COMPUSTAT to analyze earnings yield or book-to-market effects in stock returns will be using the same data to study the phenomenon in which it has already been documented. It would be preferable to analyze the explanatory power of these variables using fresh data. Finally, COMPUSTAT is limited to a relatively short sample period; comprehensive accounting data are available only back to 1963. Little is known about the relationship between accounting variables and stock returns prior to 1963.[1]

This article studies the period from July 1940 to June 1963 using accounting data collected from the annual *Moody's Industrial Manuals* for this period. There is no survivorship bias in this dataset, since Moody's could not

[1] Earnings data are available on COMPUSTAT back to the early 1950s. However, as discussed in Fama and French (1992), there is a strong survivorship bias in the pre-1963 data.

exclude a company from its manuals because it ceases to exist after the manuals have been released. Also, since Moody's extracts the accounting information for each company from annual reports and Securities and Exchange Commission (SEC) reports, the information in the Moody's manuals had to be publicly available several weeks prior to the release of the manuals. Since the sample in this study is reformed each year at about the same time the industrial manuals are released, the look-ahead bias is effectively eliminated.[2] By focusing on the 1940 to 1963 time period, this article studies the cross-section of stock returns using fresh data that (to my knowledge) have not previously been used for such a purpose. Furthermore, by combining this study with the existing literature, the number of years in which the explanatory power of fundamental accounting variables can be analyzed is nearly doubled.

The variables that are the primary focus of this article are book-to-market equity, earnings yield (earnings/price), cash flow yield (cash flow/price), and historical sales growth. These are chosen from among all the variables that have been shown to have explanatory power with respect to returns because their calculation requires accounting information. Therefore, the results of studies that use these variables may be affected by the COMPUSTAT-related problems discussed above. Other variables—namely beta, firm size (market value of equity), and share price—are also included in the analysis because they appear in existing studies along with the accounting-based variables just mentioned.

The results indicate that the natural log of book-to-market equity has significant explanatory power with respect to subsequent stock returns in cross-sectional regressions. The quintile of high book/market stocks earns 6.8 percent per year more, on average, than the corresponding low book/market portfolio. The difference in betas for these portfolios is less than 0.3.

Earnings yield also displays explanatory power in the regression analysis, and the high earnings/price quintile earns 9.5 percent per year more, on average, than the low earnings/price quintile, with a difference in betas between these two portfolios of about 0.4. Cash flow yield exhibits explanatory power when book-to-market equity and historical sales growth are held constant, and the return and beta differences for the extreme cash flow yield quintiles are about the same as for the extreme earnings yield quintiles. A two-way sort using cash flow yield and historical sales growth produces an average return difference between extreme portfolios of more than 9.5 percent per year.

There is a January seasonal in the explanatory power of several of the independent variables; much of the book/market, earnings yield, and cash flow yield effects are in January. Since the sample consists of relatively large

[2] Although the survivorship and look-ahead biases are not present in the Moody's data, there may be other biases that are present. For example, the accounting information reported by Moody's is typically more complete for large, well-known firms. In some cases, this would prevent the inclusion of smaller firms in the sample.

firms, these results indicate that January seasonals are not restricted to the stocks of small firms in the 1940 to 1963 period.

The remainder of the article consists of three sections. Section I describes the data and methodology. Section II presents the results, and Section III concludes.

I. Data and Methodology

The data for the study are taken from two primary sources. Book value, earnings, cash flow, and sales figures are taken from the *Moody's Industrial Manuals* that were published from 1940 to 1962. Stock returns, stock prices, and market values of equity are taken from the University of Chicago's Center for Research in Security Prices (CRSP) monthly file of New York Stock Exchange (NYSE) (and American Stock Exchange (AMEX) after June 1962) firms.

As of June 30 of each year from 1940 to 1962, the stocks on the CRSP monthly file are ranked by market value of equity. Then a random sample of 100 firms is selected from the firms that are: (a) in the top half of the June 30 size ranking; and (b) listed in the *Moody's Industrial Manual* for the most recently ended year.[3] The study focuses on firms in the top half of the size spectrum so that problems with infrequent trading and bid-ask spreads are less likely to affect the results. The stocks in the sample are likely to be liquid issues with small relative spreads. The prices of the stocks are generally large enough to avoid problems associated with studying low-priced stocks. (In fact, there is only one stock in the entire sample with a price less than three dollars as of the annual June 30 sample selection date.) Because the contemporaneous manual is used for each year, all accounting information is publicly available as of the annual sample selection date, and no survivorship bias is induced.

Although the sample is restricted to the top half of the CRSP monthly file size ranking, the firms in the sample still exhibit nontrivial dispersion in size. For example, as of June 30, 1962 (the last sample selection date) the quintile of largest firms has an average market value of equity of approximately $1.4 billion, while the corresponding average for the smallest quintile is about $70 million.

A comment regarding the starting date of the study (June 30, 1940) is in order. It would have been possible to start the sample period as early as 1931 and still have sufficient data (five years) for the computation of historical betas. However, 1940 is a preferable starting date for two reasons. First, the period from 1940 to the early 1960s is a period of fairly constant volatility in the stock market, at least compared to the more turbulent 1930s. Second, the

[3] If a firm's most recent financial information as reported in Moody's is for a fiscal year-end that predates the sample selection date by eighteen or more months, the firm is replaced with another randomly selected firm that has more up-to-date information.

accounting information available in the Moody's manuals is more complete after the Securities Act of 1933 and the Securities Exchange Act of 1934 created the SEC. Up until the mid-to-late 1930s, historical sales data are not reported in Moody's for the majority of firms, and depreciation expense is not reported during this period for some firms. Few firms reported extraordinary items separately on the income statement during the 1930s. This lack of data is problematic for analyses that use sales, earnings, and cash flow to compute explanatory variables. Consequently, the sample period starts on June 30, 1940.

As of each June 30, the following variables are computed for each firm in the sample:

BETA = the beta coefficient assigned to each stock, calculated as follows. As of each June 30 from 1940 to 1962, the stocks on the CRSP monthly file are assigned to beta deciles according to their historical betas, which are calculated using returns to the stock and the CRSP value-weighted index for the previous 24 (up to 60 when available) months. Then full-period betas are calculated for each of the beta deciles by regressing equally weighted annual buy-and-hold returns to the decile on the returns to the CRSP value-weighted index from July 1940 to June 1963. Each stock in the sample during a given year is assigned the full-period beta of the beta decile to which it belongs that year. Portfolio betas are used in an effort to reduce the effects of measurement errors in the calculation of betas. This procedure for assigning betas is a modified version of the method used by Fama and French (1992). (In the Fama-MacBeth (1973) analysis described below, regressions are also run using the full-period betas from twenty beta-ranked portfolios, and a third set of regressions is run using the individual-firm historical betas. The results from all three methods are qualitatively similar, so only the results using betas from ten beta-ranked portfolios are reported below.)

LBM = the natural log of the ratio of book value of common equity to market value of equity as of the sample selection date, based on the most recent fiscal year-end book value of common equity reported in the contemporaneous *Moody's Industrial Manual*. It should be noted that no firm in the sample has a negative book value of equity. Accordingly, the treatment of negative book values is not an issue in this study.

LMV = the natural log of the market value of equity for each stock as of the sample selection date (June 30 of each year).

E/P = the ratio of earnings per share to stock price (earnings yield) as of the sample selection date, based on the earnings for the most recent fiscal year reported in the contemporaneous

Moody's Industrial Manual. In the regressions, $E/P+$ and $EPNEG$ are used, where:

$$E/P+ = \begin{cases} E/P, & \text{if } E/P \geq 0 \\ 0, & \text{if } E/P < 0 \end{cases}$$

$$EPNEG = \begin{cases} 1, & \text{if } E/P < 0 \\ 0, & \text{if } E/P \geq 0 \end{cases}$$

Earnings are defined as net income before extraordinary items, less taxes and preferred dividends.

CF/P = the ratio of cash flow per share to stock price (cash flow yield) as of the sample selection date, based on the cash flow for the most recent fiscal year reported in the contemporaneous *Moody's Industrial Manual*. In the regressions, $CF/P+$ and $CFPNEG$ are used, where:

$$CF/P+ = \begin{cases} CF/P, & \text{if } CF/P \geq 0 \\ 0, & \text{if } CF/P < 0 \end{cases}$$

$$CFPNEG = \begin{cases} 1, & \text{if } CF/P < 0 \\ 0, & \text{if } CF/P \geq 0 \end{cases}$$

Cash flow is equal to earnings as defined above plus depreciation, depletion, and amortization.

$GROWTH$ = the five-year compound annual average sales growth rate (in decimal form), based on sales for the most recent fiscal year reported in the contemporaneous *Moody's Industrial Manual* and the five previous fiscal years.

$LPRICE$ = the natural log of the stock price as of the June 30 sample selection date.

The empirical analysis takes two forms. First, the stocks in the sample are ranked by the variables, and quintiles are formed based on these rankings. Then the returns, systematic risk, and other characteristics of these quintiles are analyzed. In addition to the univariate rankings, portfolios are also formed from bivariate sorts as suggested by the existing literature. Specifically, a portfolio is formed that consists of stocks that are in the top third of the cash flow yield ranking and in the bottom third of the sales growth ranking (value stocks). Another portfolio is formed that consists of stocks that are in the bottom third of the cash flow yield ranking and in the top third of the sales growth ranking (glamour stocks). These correspond to the portfolios that had the best and worst overall performance, respectively, in Lakonishok, Shleifer, and Vishny (1993). According to the theory presented in that article, the former portfolio realizes high returns because it consists of stocks that have performed poorly in the past, and investors extrapolate that poor performance into the future, underestimating the probability that the firms' fortunes will turn around. By similar reasoning, the latter portfolio experiences low returns because it is made up of overpriced stocks that have had recent good performance but subsequently disappoint investors.

When a stock in one of the quintile portfolios is delisted during the year, the return to that stock for the remainder of the year is equal to the average of the returns to the two stocks nearest to the delisted stock in the appropriate ranking. This is done in an effort to replace delisted stocks with stocks that are as close as possible to the delisted stock in terms of the variable of interest. For the High $CF/P \times$ Low $GROWTH$ and Low $CF/P \times$ High $GROWTH$ portfolios, the two replacement stocks are selected at random from among the remaining stocks in the portfolio.

The second part of the analysis involves a Fama-MacBeth (1973) cross-sectional regression analysis to determine which factors have explanatory power with respect to the cross-section of realized returns. For each month in the sample period (July 1940 to June 1963), a set of cross-sectional regressions is run using various combinations of the above-defined variables as regressors and the monthly returns to each stock (in percent format) as the dependent variable. The time-series average and standard deviation of each series of parameter estimates are then used to assess the statistical and economic significance of the independent variables.

The results of the monthly regressions are presented for all months, for January only, and for February–December only. The sample period for this study is known to have a January seasonal in returns, and existing studies, such as Jaffe, Keim, and Westerfield (1989) and Fama and French (1992), find evidence of seasonality in the relationship between certain fundamental variables and returns during more recent periods. Thus, it would not be surprising to find seasonality in the explanatory power of these variables during the 1940 to 1963 period.

II. Empirical Results

A. Portfolio Results

Table I presents summary statistics for the stocks in the sample, sorted into quintiles by each of the variables. Summary statistics are also presented for the portfolios formed by the two-way sorts described above. The returns shown in Table I are averages of equally weighted annual buy-and-hold returns, measured from June 30 to June 30, and the betas are calculated by regressing the annual returns to each portfolio on the returns to the CRSP value-weighted index. These full-period portfolio betas are reported instead of average historical betas to address the concerns of Chan (1988) and Ball and Kothari (1989) that historical betas are unreliable estimates of systematic risk for stocks that have been extreme performers in the recent past.

Table I indicates that ranking on certain variables produces dispersion in returns. Earnings/price and cash flow/price each produce annual return differences (quintile 5 versus quintile 1) greater than 9.5 percent, as does the

Table I

Summary Statistics for Portfolios of Stocks Formed by Sorting on Various Factors, July 1940 to June 1963

As of June 30 of each year from 1940 to 1962, a random sample of 100 firms is selected from the firms that are: (1) in the top half of the June 30 CRSP monthly file firm size ranking; and (2) listed in the *Moody's Industrial Manual* for the most recently ended year. The firms are sorted by various factors, and summary statistics are calculated for the quintiles formed from these sorts. Also shown are summary statistics for two portfolios that are formed based on independent two-way sorts. The *HI CF/P × LO GROWTH* portfolio consists of firms that are in the top third of the cash flow yield ranking and the bottom third of the sales growth ranking. The *LO CF/P × HI GROWTH* portfolio consists of firms that are in the bottom third of the cash flow yield ranking and the top third of the sales growth ranking. Reported statistics include: (1) *PRICE*, the average June 30 share price; (2) *B/M*, the average book-to-market equity ratio (book value as of fiscal year-end, market value as of June 30); (3) *E/P*, the ratio of earnings to stock price; (4) *CF/P*, the ratio of cash flow to stock price; (5) *GROWTH*, the five-year annual average growth rate in sales; (6) *RETURN*, the equally weighted annual buy-and-hold return to the portfolio for the sample period; and (7) *BETA*, calculated by regressing annual returns to the portfolio on the corresponding returns to the CRSP value-weighted index. *PRICE, B/M, E/P, CF/P*, and *GROWTH* are equally weighted averages across the 23 sample selection dates from June 30, 1940 to June 30, 1962.

Portfolios Sorted By:	*PRICE* ($)	*B/M*	*E/P*	*CF/P*	*GROWTH*	*RETURN* (%)	*BETA*
Earnings/Price							
0 (Negative)	29.76	1.27	−0.09	−0.00	0.02	11.91	0.85
1 (Low)	58.73	0.54	0.05	0.08	0.10	13.48	1.02
2	44.84	0.71	0.08	0.11	0.10	15.41	0.87
3	39.99	0.84	0.09	0.14	0.10	15.01	1.07
4	36.64	1.01	0.12	0.17	0.11	19.25	1.19
5 (High)	33.94	1.32	0.18	0.27	0.15	23.04	1.45
Cash Flow/Price							
0 (Negative)	24.83	1.33	−0.19	−0.14	−0.12	13.46	1.51
1 (Low)	60.22	0.44	0.05	0.07	0.10	12.54	0.98
2	46.45	0.68	0.08	0.11	0.09	15.23	0.83
3	38.92	0.83	0.10	0.14	0.11	17.62	1.10
4	35.39	1.02	0.12	0.18	0.12	18.72	1.29
5 (High)	32.99	1.46	0.17	0.28	0.14	22.07	1.40
5-Year Sales Growth							
1 (Low)	37.34	1.04	0.09	0.14	−0.00	15.67	1.04
2	45.25	0.89	0.09	0.14	0.06	17.02	1.06
3	45.33	0.90	0.10	0.15	0.10	18.85	1.09
4	47.31	0.86	0.11	0.16	0.14	18.35	1.17
5 (High)	40.01	0.76	0.11	0.16	0.28	15.58	1.22
Firm Size							
1 (Small)	31.80	1.12	0.12	0.18	0.11	18.13	1.16
2	33.79	1.03	0.11	0.17	0.12	17.49	1.16
3	37.90	0.82	0.10	0.14	0.12	17.03	1.04
4	45.98	0.79	0.10	0.15	0.12	18.43	1.25
5 (Large)	64.27	0.67	0.08	0.13	0.11	15.03	1.01

Table I—*Continued*

Portfolios Sorted By:	*PRICE* ($)	*B/M*	*E/P*	*CF/P*	*GROWTH*	*RETURN* (%)	*BETA*
Stock Price							
1 (Low)	18.31	1.16	0.11	0.18	0.12	20.17	1.42
2	28.54	1.02	0.11	0.17	0.11	18.43	1.16
3	37.19	0.90	0.11	0.16	0.11	16.85	1.05
4	47.53	0.77	0.10	0.14	0.12	15.26	0.95
5 (High)	82.18	0.59	0.08	0.12	0.11	15.42	1.01
Book/Market							
1 (Low)	62.31	0.32	0.06	0.08	0.14	14.24	1.00
2	45.38	0.56	0.09	0.12	0.13	15.46	1.02
3	39.46	0.78	0.10	0.15	0.12	17.55	1.11
4	36.55	1.03	0.11	0.17	0.10	17.70	1.19
5 (High)	30.04	1.73	0.14	0.24	0.09	21.06	1.29
Historical Beta							
1 (Low)	49.98	0.71	0.08	0.12	0.09	13.47	0.71
2	42.19	0.89	0.10	0.15	0.10	18.93	1.02
3	42.21	0.91	0.10	0.15	0.11	16.66	1.06
4	39.57	0.91	0.11	0.17	0.13	18.74	1.29
5 (High)	39.79	1.00	0.11	0.18	0.14	18.19	1.52
2-Way Sorts							
LO CF/P × HI GROWTH	58.49	0.44	0.06	0.08	0.20	12.74	0.99
HI CF/P × LO GROWTH	32.27	1.52	0.15	0.24	0.03	22.30	1.65

two-way sort. The book/market sort produces a return dispersion of 6.8 percent.[4] (Also, the book/market and cash flow/price sorts are monotonic in returns across quintiles 1 to 5.) None of the other sorts produce return dispersions greater than 5 percent. (Note, however, that the 3.1 percent return difference for the firm size sort represents only the top half of the size ranking from the CRSP monthly file.)

In addition to documenting a relationship between certain variables and subsequent returns, inspection of Table I also indicates that there are correlations among the variables. For example, earnings/price and cash flow/price are highly correlated at the individual firm level, as are cash flow/price and book-to-market equity. It is not surprising to find high correlation among these variables, since each is a scaled version of price. These relationships are confirmed by examining the average cross-sectional correlations between the variables. Specifically, $\bar{\rho}(E/P, CF/P) = 0.83$, $\bar{\rho}(CF/P, LBM) = 0.66$, $\bar{\rho}(E/P, LBM) = 0.50$, $\bar{\rho}(LP, LMV) = 0.45$, and $\bar{\rho}(LP, LBM) = -0.42$, where $\bar{\rho}(\cdot, \cdot)$ is the average of the cross-sectional correlations of the indicated variables for the 23 years in the sample period. None of the other pairs of variables have $|\bar{\rho}| > 0.32$. In particular, $\bar{\rho}(BETA, LMV) = -0.12$, so the

[4] The return differences between quintiles 1 and 5 are significant at the 5 percent level for the *E/P*, *CF/P*, and *B/M* sorts.

high correlation between beta and size that is found in other studies is not present here. (As pointed out by Jegadeesh (1992), the correlation between beta and size is close to 1 in absolute value at the portfolio level in previous studies. The correlation is lower here due to individual firms being used instead of portfolios, and because only the top half of the size ranking is included.)

A comment about the negative earnings/price and cash flow/price portfolios is in order. For 14 of the 23 years in the sample, there are no firms with negative earnings, and there are no firms with negative cash flow in 17 of the years. In the remaining years, these portfolios typically contain few stocks. The statistics reported for these portfolios should be interpreted with this in mind.

B. Regression Results

Table II presents the average coefficients from the monthly regressions for all months from July 1940 to June 1963. Univariate regressions are presented for each of the variables (except bivariate regressions are run for *CF/P* and *E/P* to reflect the firms with negative cash flow and earnings), and multivariate regressions, as suggested by existing studies, are also shown. *LMV* and *LPRICE* are not included as independent variables in Tables II and III because the results of univariate and multivariate regressions indicate little explanatory power for these variables. These results are clearly not comparable to other studies, since the current study excludes small firms with low stock prices from the sample. Existing research indicates that these variables do have explanatory power during the 1940 to 1963 period when the entire CRSP monthly file is used. However, when the sample is restricted to relatively large firms, these variables do not exhibit a significant relationship with stock returns.

It should be noted that the average coefficients on *CFPNEG* and *EPNEG* are computed using fewer parameter estimates than the other variables, due to the previously mentioned absence of firms with negative earnings and cash flow in several years. Of the 276 months in the sample period, there are 168 for which *EPNEG* coefficients cannot be estimated and 204 for which *CFPNEG* coefficient estimates are absent. The average coefficients on these variables should be interpreted accordingly.

The univariate *LBM* regression indicates a statistically significant relationship between book-to-market equity and subsequent returns. The point estimate on *LBM* (0.26) is smaller than those found in studies of the post-1963 era (for example, Fama and French (1992, Table III) report a coefficient of 0.50), but the statistical significance is present here as in those studies. When *LBM*, *CF/P+*, *CFPNEG*, and *GROWTH* are included as independent variables, *LBM* loses its significance. Since book-to-market equity and cash flow/price are highly correlated, it is difficult to make definitive statements about the marginal explanatory power of either variable. However, the results in Table II suggest that, controlling for differences

Table II

Average Slopes (t-Statistics) from Monthly Regressions of Stock Returns on Beta, Book-to-Market Equity, *CF/P*, *E/P*, and Sales Growth: July 1940 to June 1963

As of June 30 of each year from 1940 to 1962, a random sample of 100 firms is selected from the firms that are: (1) in the top half of the June 30 CRSP monthly file firm size ranking; and (2) listed in the *Moody's Industrial Manual* for the most recently ended year. A cross-sectional regression is run for each of the following twelve months, with the return to each stock (in percent format) as the dependent variable. The regressors include: (1) *BETA*, the postranking beta coefficient for the beta-ranked decile of each stock, calculated using annual buy-and-hold returns to the decile and the CRSP value-weighted index; (2) *LBM*, the natural log of the book-market ratio (book value as of fiscal year-end, market value as of June 30); (3) $CF/P+$, the ratio of cash flow per share to stock price if cash flow is positive, and zero if cash flow is negative; (4) *CFPNEG*, equal to zero if cash flow is positive and one if cash flow is negative; (5) $E/P+$, the ratio of earnings per share to stock price if earnings are positive, and zero if earnings are negative; (6) *EPNEG*, equal to zero if earnings are positive and one if earnings are negative; and (7) *GROWTH*, the five-year annual average growth rate in sales. Parameter estimates and t-statistics (in parentheses) are based on the averages and standard deviations of the time series of 276 regression coefficients.

BETA	*LBM*	$CF/P+$	*CFPNEG*	$E/P+$	*EPNEG*	*GROWTH*	Avg. Adj. R^2
0.15							0.029
(0.91)							
	0.26						0.020
	(2.38)						
		1.64	0.29				0.021
		(1.55)	(0.27)				
				4.35	0.52		0.023
				(2.31)	(0.72)		
						−0.46	0.010
						(−0.96)	
		1.56	0.43			−0.45	0.031
		(1.43)	(0.40)			(−0.93)	
	−0.05	2.55	0.20			−0.56	0.040
	(−0.35)	(2.09)	(0.19)			(−1.16)	

in book-to-market equity, cash flow yield has predictive ability with respect to subsequent realized returns.

Another variable that displays explanatory power is earnings yield. This explanatory power is found both in the bivariate ($E/P+$ and *EPNEG*) regression and in multivariate regressions (not shown) that include various combinations of beta, size, and price. Furthermore, the magnitude of the $E/P+$ coefficient in Table II (4.35) is fairly close to the corresponding point estimate of 4.72 for the 1963 to 1990 period in Fama and French (1992, Table III), even though the current study covers a different time period and focuses on relatively large firms.

Beta does not fare as well in the univariate regression as one might expect. Although it may not be surprising to see insignificant explanatory power for beta in certain multivariate regressions, the existing literature (e.g., Fama and MacBeth (1973) and Chan and Lakonishok (1993)) finds evidence of a univariate relationship between beta and returns during the period that is studied here. The evidence in favor of beta would undoubtedly be stronger here if smaller firms were included in the sample (see Ritter and Chopra (1989)). However, for the firms in this study, the relationship between beta and returns is essentially flat. Multivariate regressions that include beta as an independent variable are not shown, since beta adds no explanatory power to such regressions.

The coefficients on historical sales growth (*GROWTH*) are negative, as in Lakonishok, Shleifer, and Vishny (1993), but, unlike that study, the estimates are not statistically significant. Since Lakonishok, Shleifer, and Vishny (1993) use the firms' average sales growth ranking instead of the average sales growth rate in their regressions, their results may be less affected by outliers than those reported here. However, the regression results and the evidence in Table I suggest that the relationship between sales growth and returns may have been weaker (except in January, as seen below) during the years prior to 1963.[5]

C. Seasonal Patterns

Table III presents average regression coefficients for January (Panel A) and February–December (Panel B). It can be seen from this table that the significance of *LBM*, *E/P+*, and *CF/P+* in Table II is due mostly to January. Focusing on the univariate point estimates in Panels A and B, one can see that more than half of the book/market effect on subsequent returns occurs in January. (This January seasonal is much stronger than that reported by Fama and French (1992, p. 448) for the 1963 to 1990 period.) Nearly half of the *E/P+* effect is in January, as are more than four-fifths of the *CF/P+* and *GROWTH* effects. The results in Table III indicate that January seasonals are not restricted to small firms during the 1940 to 1963 period.[6] Further, since there are few low-priced stocks in the sample, the results are not likely to be unduly influenced by penny stocks selling at the bid price at the end of the year.

III. Conclusions

This article analyzes the ability of certain fundamental variables to explain the cross-section of realized returns in the period from July 1940 to June 1963. In doing so, it presents evidence that is independent of studies from the

[5] The point made by Lakonishok, Shleifer, and Vishny (1993) is not that sales growth should have explanatory power by itself, but that sales growth and cash flow yield considered together should be able to predict returns. The two-way sorts in Table I indicate that this is the case, although the dispersion in returns produced by the cash flow yield ranking is nearly as large as that in the two-way sort.

[6] For the univariate/bivariate regressions, the null hypothesis of no difference between the January and February–December average coefficients is rejected at the 1 percent level for all variables except *BETA*, *CFPNEG*, and *EPNEG*.

Table III

Average Slopes (t-Statistics) from Monthly Regressions of Stock Returns on Beta, Book-to-Market Equity, CF/P, E/P, and Sales Growth: July 1940 to June 1963; January Compared to February–December

As of June 30 of each year from 1940 to 1962, a random sample of 100 firms is selected from the firms that are: (1) in the top half of the June 30 CRSP monthly file firm size ranking; and (2) listed in the *Moody's Industrial Manual* for the most recently ended year. A cross-sectional regression is run for each of the following twelve months, with the return to each stock (in percent format) as the dependent variable. The regressors include: (1) *BETA*, the post-ranking beta coefficient for the beta-ranked decile of each stock, calculated using annual buy-and-hold returns to the decile and the CRSP value-weighted index; (2) *LBM*, the natural log of the book-market ratio (book value as of fiscal year-end, market value as of June 30); (3) $CF/P+$, the ratio of cash flow per share to stock price if cash flow is positive, and zero if cash flow is negative; (4) *CFPNEG*, equal to zero if cash flow is positive and one if cash flow is negative; (5) $E/P+$, the ratio of earnings per share to stock price if earnings are positive, and zero if earnings are negative; (6) *EPNEG*, equal to zero if earnings are positive and one if earnings are negative; and (7) *GROWTH*, the five-year annual average growth rate in sales. Panel A shows average slopes and t-statistics for January only (23 observations), and Panel B shows the corresponding statistics for February to December (253 observations). Parameter estimates and t-statistics (in parentheses) are based on the averages and standard deviations of the respective time series of regression coefficients.

BETA	*LBM*	$CF/P+$	*CFPNEG*	$E/P+$	*EPNEG*	*GROWTH*	Avg. Adj. R^2
			Panel A: January Only (23 observations)				
0.20							0.028
(0.35)							
	1.95						0.043
	(5.64)						
		15.83	3.77				0.030
		(3.65)	(1.10)				
				25.54	4.73		0.033
				(3.22)	(1.71)		
						−4.59	0.014
						(−2.58)	
		16.31	4.45			−4.60	0.043
		(3.99)	(1.50)			(−2.78)	
	1.15	9.12	5.00			−3.41	0.061
	(2.50)	(1.97)	(2.15)			(−2.03)	
			Panel B: February–December (253 observations)				
0.15							0.029
(0.85)							
	0.11						0.018
	(0.98)						
		0.35	−0.02				0.020
		(0.34)	(−0.02)				
				2.42	0.14		0.022
				(1.29)	(0.19)		
						−0.08	0.010
						(−0.17)	
		0.22	0.06			−0.07	0.030
		(0.20)	(0.05)			(−0.14)	
	−0.15	1.95	−0.24			−0.30	0.038
	(−1.17)	(1.55)	(−0.22)			(−0.60)	

COMPUSTAT era of the mid-sixties to the present. The results are not tainted by the survivorship or look-ahead biases discussed in Banz and Breen (1986), and fresh data are introduced into the study of the cross-section of stock returns. Extremely low-priced stocks are virtually excluded from the study, so microstructure-induced biases are not likely to be a major problem.

The regression results indicate that book-to-market equity and earnings yield each have explanatory power with respect to subsequent returns in univariate regressions. Portfolios formed by ranking on these variables confirm their ability to predict returns. Two-way sorts using cash flow yield and sales growth also produce dispersion in average returns, as does the univariate cash flow yield ranking. Regardless of whether the observed return differences can be explained by differences in risk, the magnitudes of the return differences are noteworthy, especially when it is kept in mind that the firms are all in the largest half of NYSE firms.

Assessing the marginal explanatory power of CF/P, E/P, and LBM is difficult, because they are each a scaled version of price and thus highly correlated with one another. There is no clear winner among the three with respect to explaining the cross-sectional variation in stock returns. Nevertheless, each is shown to reliably predict returns during the 1940 to 1963 period, especially in January.

The lack of explanatory power shown by firm size and share price in the overall results must be viewed in light of the sample selection procedure. These variables would have been significant if small firms with low stock prices had not (by construction) been excluded from the sample. However, book-to-market equity, earnings yield, and cash flow yield have been shown to have explanatory power with respect to subsequent returns during the "pre-COMPUSTAT era," even when the sample is restricted to relatively large firms (and such firms comprise the vast majority of publicly traded equity).

While there is no guarantee that these fundamental accounting variables will continue to be related to subsequent stock returns, the evidence presented here, together with the evidence in Fama and French (1992) and other studies of the period from 1963 forward, demonstrates that such relationships have been present for at least the past half-century. These relationships are not artifacts of data snooping or a survivorship bias in the accounting data. Furthermore, the results from 1940 to 1963 indicate that the results from the COMPUSTAT era of 1963 forward are not merely a manifestation of the takeover wave of the 1980s.

REFERENCES

Ball, Ray, and S. P. Kothari, 1989, Nonstationary expected returns: Implications for tests of market efficiency and serial correlation in returns, *Journal of Financial Economics* 25, 51–74.

Banz, Rolf W., and William J. Breen, 1986, Sample-dependent results using accounting and market data: Some evidence, *Journal of Finance* 41, 779–793.

Basu, S., 1977, Investment performance of common stocks in relation to their price earnings ratios: A test of the efficient market hypothesis, *Journal of Finance* 32, 663–682.

Blume, Marshall, and Frank Husic, 1973, Price, beta and exchange listing, *Journal of Finance* 28, 283–299.

Blume, Marshall, and Robert F. Stambaugh, 1983, Biases in computed returns: An application to the size effect, *Journal of Financial Economics* 12, 387–404.

Chan, K. C., 1988, On the contrarian investment strategy, *Journal of Business* 61, 147–163.

Chan, Louis K. C., Yasushi Hamao, and Josef Lakonishok, 1991, Fundamentals and stock returns in Japan, *Journal of Finance* 46, 1739–1764.

Chan, Louis K. C., and Josef Lakonishok, 1993, Are the reports of beta's death premature?, *Journal of Portfolio Management* 19, 51–62.

Davis, James L., 1994, The cross-section of stock returns and survivorship bias: Evidence from delisted stocks, Working paper, Kansas State University.

De Bondt, Werner F. M., and Richard M. Thaler, 1987, Further evidence on investor overreaction and stock market seasonality, *Journal of Finance* 42, 557–581.

Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

———, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F., and James D. MacBeth, 1973, Risk return and equilibrium: Empirical tests, *Journal of Political Economy* 81,607–636.

Jegadeesh, Narasimhan, 1992, Does market risk really explain the size effect?, *Journal of Financial and Quantitative Analysis* 27, 337–351.

Jaffe, Jeffrey, Donald B. Keim, and Randolph Westerfield, 1989, Earnings yields, market values, and stock returns, *Journal of Finance* 44, 135–148.

Keim, Donald B., 1983, Size-related anomalies and stock return seasonality, *Journal of Financial Economics* 12, 13–32.

Kothari, S. P., Jay Shanken, and Richard G. Sloan, 1993, Another look at the cross-section of expected returns, Working paper, University of Rochester.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1993, Contrarian investment, extrapolation, and risk, Working paper, University of Illinois.

Lo, Andrew, and A. Craig MacKinlay, 1990, Data-snooping biases in tests of financial asset pricing models, *Review of Financial Studies* 3, 431–467.

Ritter, Jay R., and Navin Chopra, 1989, Portfolio rebalancing and the turn-of-the-year effect, *Journal of Finance* 44, 149–166.

Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein, 1985, Persuasive evidence of market inefficiency, *Journal of Portfolio Management* 11, 9–17.