THE JOURNAL OF FINANCE • VOL. XLIII, NO. 2 • JULY 1988

R^2

RICHARD ROLL*

ABSTRACT

Even with hindsight, the ability to explain stock price changes is modest. R^2s were calculated for the returns of large stocks as explained by systematic economic influences, by the returns on other stocks in the same industry, and by public firm-specific news events. The average adjusted R^2 is only about .35 with monthly data and .20 with daily data. There is little relation between explanatory power and either the firm's size or its industry. There is little improvement in R^2 from eliminating *all* dates surrounding news reports in the financial press. However, the sample kurtosis is quite different when such news events are eliminated, thereby revealing a mixture of return distributions. Nonnews dates also indicate the presence of a distributional mixture, perhaps due to traders acting on private information.

THE MATURITY OF A science is often gauged by its success in predicting important phenomena. Astronomy, the oldest science, is able to predict the positions of planets and the reappearance of comets with a high degree of accuracy. Astronomical phenomena are extraordinarily regular, and they permit the construction of forecasting models with only trivial prediction errors.

Financial science too can boast about very high explanatory power if the phenomenon is artfully selected. For instance, the daily change in an option's price can be accurately "predicted" by the *concurrent* change in the associated stock price. Similarly, a good predictor of tomorrow's asset price is today's price. But both of these examples are contrived; they are analogous to a meteorologist who might claim a high degree of predictive power because the weather can be forecast fairly accurately over the next hour!

The immaturity of our science is illustrated by the conspicuous lack of predictive content about some of its most intensely interesting phenomena, particularly *changes* in asset prices. General stock price movements are notoriously unpredictable and financial economists have even developed a coherent theory (the theory of efficient markets) to explain why they *should* be unpredictable.

However, many financial economists seem to believe that, with hindsight, they could explain most asset-price movements with authenticated information. The prevailing paradigm about stock price changes ascribes them to (1) unpredictable movements in pervasive economic factors,[1] (2) unpredictable changes in the firm's market environment, i.e., industry information, and (3) unpredictable

* Anderson Graduate School of Management, University of California, Los Angeles.

[1] There is a controversy within the paradigm about the identity and number of pervasive, or "systematic," economic factors, but almost everyone would agree that there is at least one, and some researchers have uncovered evidence that there are several.

Richard Roll
President of the American Finance Association
1987

events specific to the firm itself. By observing and measuring these influences *ex post*, one could presumably "explain" stock price movements with an astronomical degree of accuracy. The R^2 should be close to 1.0.

This paper offers a simple empirical investigation of the paradigm. Is it really true that we can explain the actual price movements of *individual* common stocks by broad economic influences, industry influences, and specific news events about the firm? The paper is organized by these categories of information: the second section examines pervasive factors, the third section industry factors, and the fourth section firm-specific factors.

The results are not very gratifying. With all explanatory factors included, less than forty percent of the monthly return volatility in the typical stock can be explained, and this is for a sample of the largest firms in the U.S. market. The explanatory success for daily return data is even less. The paucity of explanatory power represents a significant challenge to our science. We ought to discover either (a) measureable influences that will explain the remaining sixty percent, or (b) a coherent reason why it *should* forever remain unexplained.

I. The Data

All data in the paper refer to equities of corporations traded on the New York and American Stock Exchanges. Monthly returns are used in the first two sections below and daily returns are used in the fourth section. The monthly returns cover a five-year period, September 1982 through August 1987,[2] and they derive from two sources: through December 1986, the source was the CRSP monthly stock returns file, and from January through August 1987 the source was IDC.[3] The daily data covered the 1982–86 calendar years inclusive and were obtained from the CRSP daily returns file.

Data pertaining to news events were obtained from the Dow-Jones News Retrieval System. This data source contains *every* mention of most publicly held companies in a number of different publications and news services. The system was used here to retrieve every news item about a company that appeared either in the *Wall Street Journal* or on the Dow-Jones news wire (the Broad Tape) during the sample period. These news items included stories exclusively about a company and also stories of a more general nature in which the company was mentioned.

II. The R^2 of Systematic Factors

Systematic, or non-diversifiable, factors play a major role in the most widely studied theories of asset pricing: the Capital Asset Pricing Model (CAPM)[4] and the Arbitrage Pricing Theory (APT).[5] For our purposes here, the principal

[2] This was the most recent five-year period available at the time the calculations were made.

[3] CRSP is the acronym for the University of Chicago's Center for Research in Securities Prices, and IDC is the acronym for Interactive Data Corporation.

[4] See Sharpe [11], Lintner [5], and Mossin [7].

[5] See Ross [9].

distinction between the two theories is the number of non-diversifiable factors. The "market model" version of the CAPM implies a single pervasive market factor, while the APT allows more than one (though the APT is also consistent with just one).

Several factors may turn out to explain a larger proportion of intertemporal return volatility than a single factor, but this finding alone would not constitute evidence that a multiple-factor theory is better. *That* conclusion would also require an empirical finding that additional factors are indeed pervasive, non-diversifiable, and most important, that they are associated with additional risk premia. Such interesting and difficult questions are ignored in this paper.

Instead, the paper merely reports on the cross-sectional distribution of R,[2] *adjusted for degrees of freedom*,[6] from using a single factor *and* from using multiple factors in regressions of the type

$$r_{j,t} = a_j + b_{1,j} f_{1,t} + \cdots + b_{k,j} f_{k,t} + e_{j,t},$$

where $r_{j,t}$ is the total return on stock j in period t, $f_{i,t}$ is systematic factor i in period t, the a's and b's are estimated regression coefficients, and $e_{j,t}$ is the "unexplained" return. The adjusted R^2 from such a regression is defined as

$$R^2 = 1 - [(T - 1)/(T - k - 1)][s^2(e)/s^2(r)],$$

where T is the time-series sample size and $s(x)$ is the sample standard deviation of x.

For the CAPM single-factor market model ($k = 1$) and the APT five-factor market model ($k = 5$), Figure 1 presents the cross-sectional frequency distribution of R^2 for the 2030 individual stocks listed on the NYSE and AMEX as of September 1982, and for which there were at least thirty monthly observations through August 1987. For the CAPM, $f_{1,t}$ was defined as the equal-weighted index of all stocks in the sample available in month t. For the APT, f_1 through f_5 were factor scores obtained from a large-scale factor analysis using all stocks that were continuously listed during the five-year period. The first APT factor, f_1, is very highly correlated with the CAPM's single factor, and indeed the APT f_1 is tantamount to an equal-weighted market index. By construction, the second through the fifth APT factors are uncorrelated with f_1 and also with each other.

As Figure 1 reveals quite clearly, the entire distribution of R^2 is displaced to a somewhat higher level for the multiple-factor model (APT), relative to the single-factor model (CAPM). The mean R^2s were, respectively, .179 for the CAPM and .244 for the APT. Of the 2030 stocks, 1571 (77.4 percent) had a higher R^2 with the multiple-factor model. The second panel of Figure 1 shows the cross-sectional frequency distribution of the difference between the APT and CAPM R^2s for individual firms.

The *average* explanatory power of well-accepted market models is quite modest, but there are some firms with impressive R^2s (in either a single- or a multiple-factor regression), and perhaps a study of these firms would help us understand why the explanatory power is rather limited for the average firm.

[6] Every R^2 reported in the paper is adjusted for degrees of freedom, even if this is not explicitly stated.

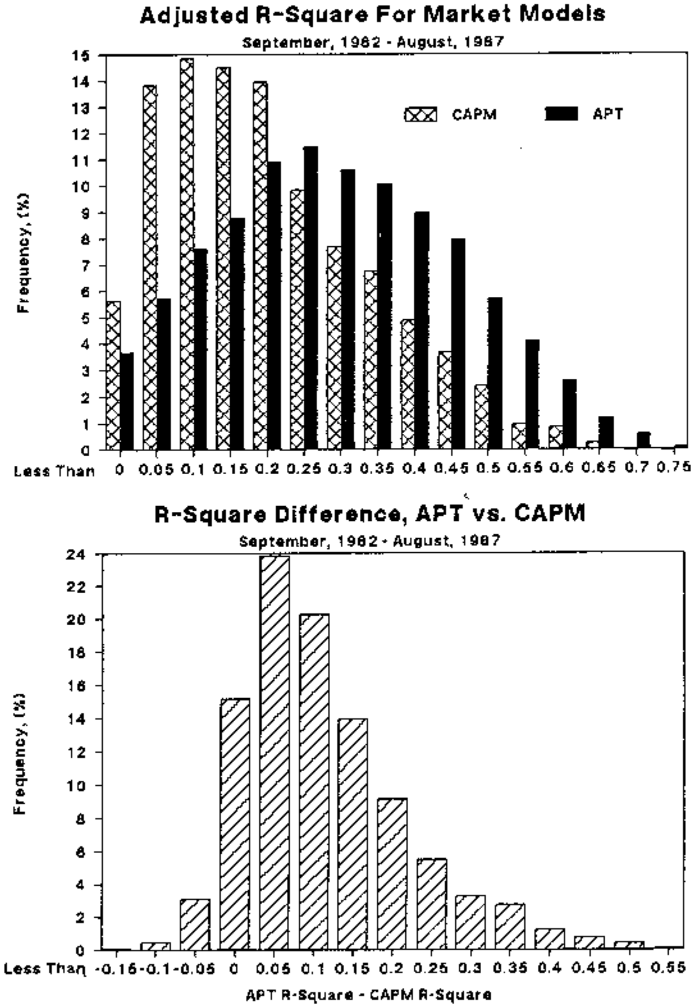

Figure 1.

A. R^2 *and Firm Size: The Portfolio-Diversification Effect*

Firm size is a possible explanation of why R^2s differ; larger firms generally have many divisions and often operate in more than a single industry and market. Thus, they superficially resemble diversified portfolios of smaller firms, and it is well known that diversified portfolios have high R^2s, at least with respect to the single CAPM factor and with respect to the broad market APT factor, f_1.

This possibility is supported to a modest extent by the results depicted in Figure 2, a cross-sectional scatter diagram of CAPM R^2 against the natural logarithm of firm size.[7] There is a discernible positive cross-sectional correlation.

[7] Firm size is defined as market price times the number of outstanding shares on the *beginning* date of the sample period, i.e., August 31, 1982.

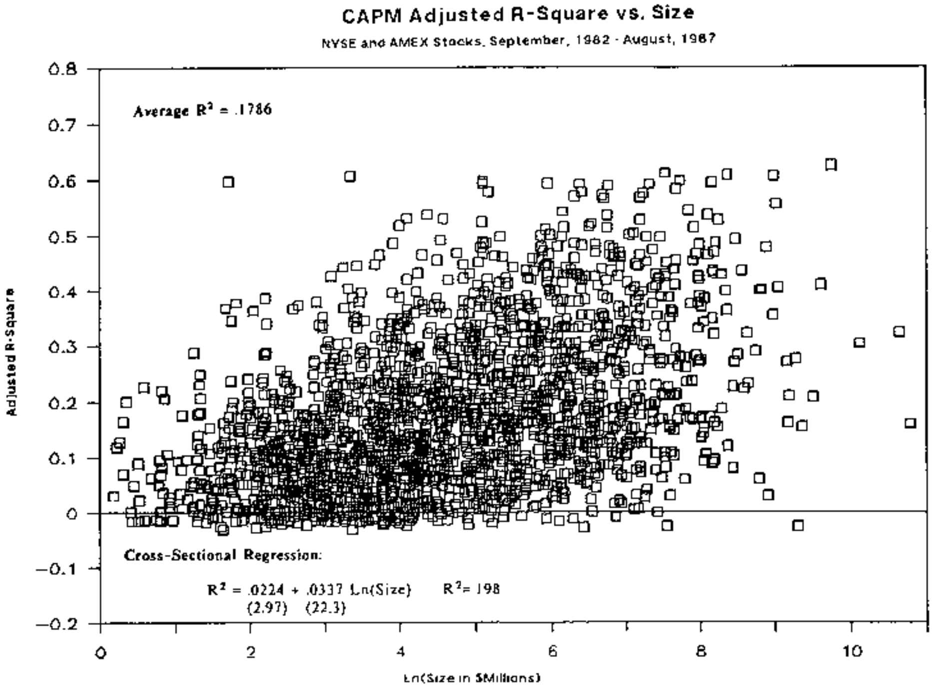

Figure 2.

A simple linear regression between R^2 and ln(Size) indicates a statistically significant connection (but this statistical significance is suspect because the observations are probably not cross-sectionally independent). When firms are relatively small, say below \$10 million, few R^2s exceed 0.3, while when firms are relatively large, say above \$150 million, there are many firms with R^2s above 0.4 and a few even above 0.5. Based on this evidence taken in isolation, diversification may indeed be an explanation for why larger firms display a greater degree of explanatory power.

However, some doubt about this conclusion can be derived from considering Figure 3, which presents the cross-sectional relation between the five-factor (APT) R^2 and ln(Size). There is a much less perceptible positive relation here, though the keen-eyed observer may still detect a small one. Aside from a higher R^2 level in general, the biggest difference between the multiple-factor and the single-factor scatter is a significant improvement in the R^2s for small firms. This may suggest that higher explanatory power of market models for large firms is due less to general diversification than to large firms being less susceptible, for some as yet unknown reason, to systematic risks that are *not* general market risks.

To investigate the diversification element in more detail, all of the firms in the sample were sorted by size on the beginning date of the period and divided into two groups, one group comprised of the decile of largest firms and the second

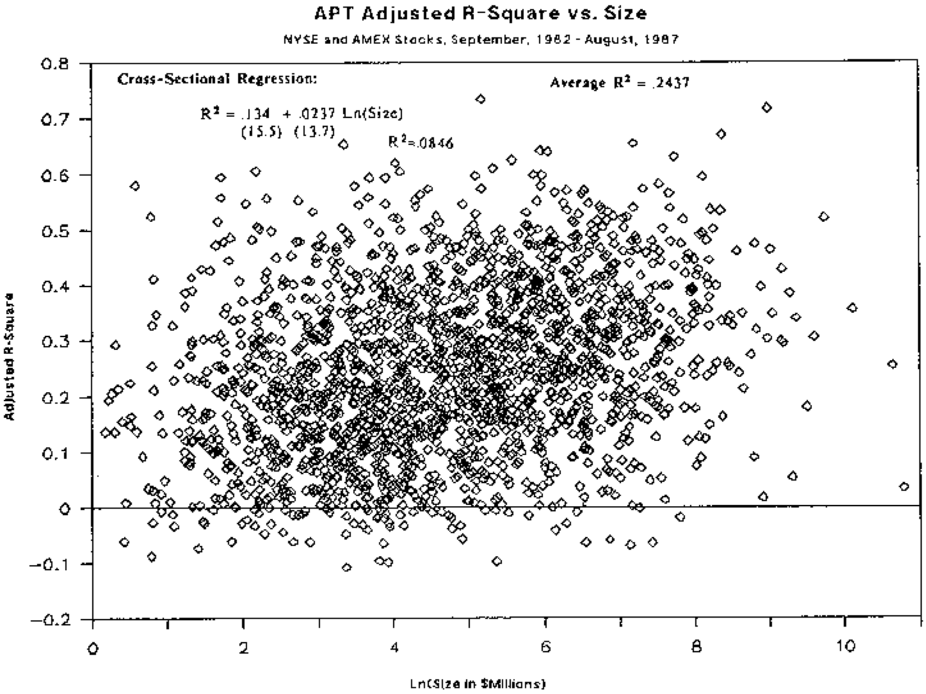

Figure 3.

group comprised of the nine small-firm deciles. For each large firm, an equal-size portfolio of smaller firms was constructed from the second group in the following fashion:

1. $Size_L$ was determined for large firm L.
2. Firms from the bottom nine deciles were selected at random, (without replacement), until, for the Jth firm so selected, their cumulative sizes satisfied

$$Size_{S,1} + \cdots + Size_{S,J} > Size_L.$$

3. A value-weighted portfolio was formed with the J small firms, where the weights were

$$w_j = Size_{S,j}/Size_L \qquad \text{for } j < J,$$

and

$$w_J = 1 - [Size_{S,1} + \cdots + Size_{S,J-1}]/Size_L.$$

4. The portfolio was rebalanced in each sample month to these original weights.

Figure 4 presents the relation between firm size and R^2 for both the decile of largest individual firms and the size-matched sample of portfolios. The first panel of Figure 4 is simply the largest size decile from the same data observations presented in Figures 2 and 3. There is little relation between size and R^2 for

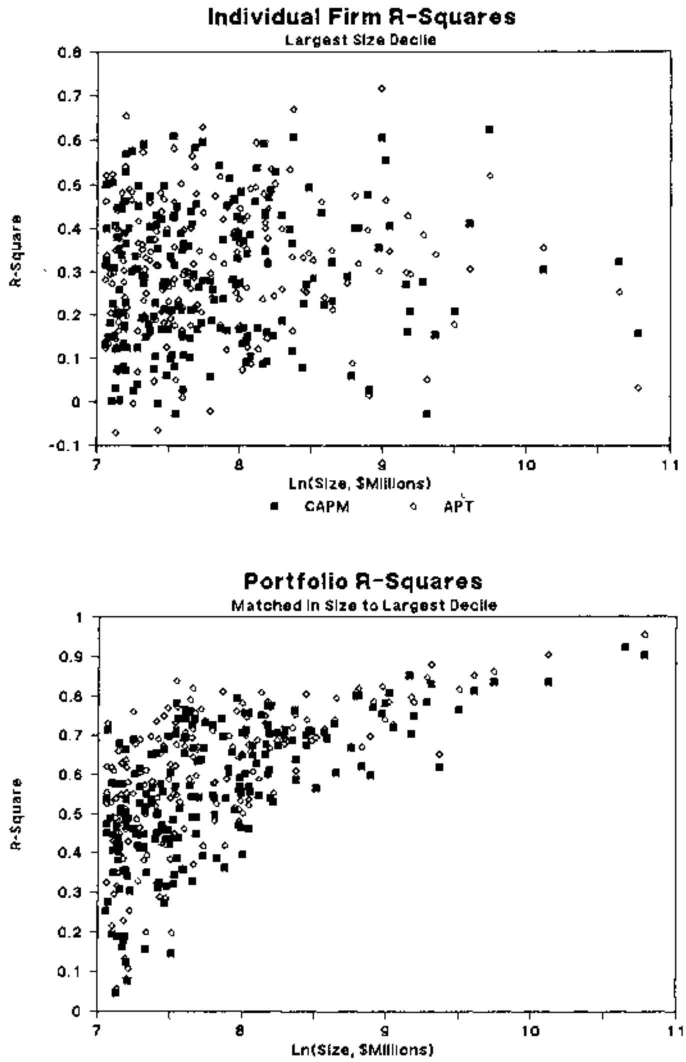

Figure 4.

individual firms within the decile. In contrast, the portfolios matched in size display a strong positive relation between size and R^2 (second panel of Figure 4). The striking dissimilarity between the two scatter diagrams implies that diversification *per se* cannot be the explanation of the larger R^2s of individual large firms. In addition, the average R^2 of the size-matched portfolios is somewhat higher than the average individual-firm R^2 in the largest size decile.[8]

Further doubt about the magnitude of the pure diversification effect can be obtained by studying Figure 5, which shows the relation between R^2 for each

[8] For the CAPM, the average individual-firm R^2 in the largest size decile was .2873 and the average size-matched portfolio R^2 was .5583. For the APT, the corresponding figures were .3159 and .5957.

individual firm in the largest size decile and its corresponding size-matched portfolio. The CAPM single-factor results are in the top panel and the multiple-factor APT results are in the bottom panel. They show an identical pattern, viz., there is no relation whatever between an individual firm's R^2 on systematic factors and the R^2 of a portfolio whose aggregate market capitalization is the same.

III. The Additional R^2 of Industry Events

Perhaps it is only mildly surprising that a large firm's R^2 should have virtually no relation to a portfolio's R^2 whose constituent stocks aggregate to the same market capitalization. After all, by choosing stocks randomly for inclusion in a

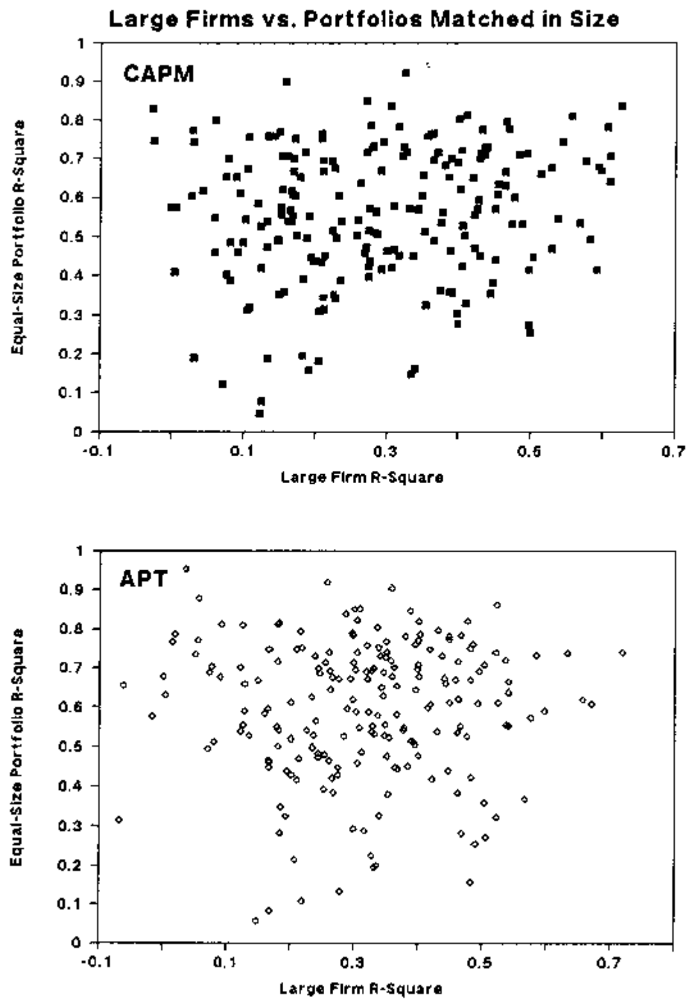

Figure 5.

size-matched portfolio, there is likely to be significant diversification across industries. To the extent that the larger firms are not similarly diversified by having divisions in several industries, their R^2s might naturally be lower.

To check this out, the four-digit SIC code enregistered in the CRSP file was used to categorize stocks into industries and then size-matched portfolios were constructed for as many large firms as possible *within* each industry. To be specific, the following procedure was employed:

1. All firms were sorted by four-digit SIC code.
2. For a *given* four-digit code, the largst firm was selected.
3. From among the smaller firms in the same four-digit industry, a portfolio was constructed, if possible, by randomly choosing firms, without replacement, until an *aggregate* market capitalization equal to that of the large firm had been obtained.[9]
4. For the next largest firm in the industry, another size-matched portfolio was constructed, and so on until as many portfolios as possible were constructed from the smaller firms in a given four-digit industry, each portfolio matched in size to one of the larger firms in the same industry.

Table I presents the resulting industries included, and the number of portfolios in each industry. There were ninety-six large firms, and thus ninety-six size- and industry-matched portfolios, in sixty-six different industries.

The sample of large firms, their industry codes, and the size- and industry-matched portfolios are presented in Table II, which also gives single-factor and multiple-factor R^2s and a few other pertinent statistics. For the ninety-six large firms, the average CAPM R^2 was .2394 and the APT R^2 was .2872; the corresponding portfolios had higher R^2s on average, .3614 and .4426, respectively. On average, 6.2 small firms in the same industry were required to construct a portfolio with that same aggregate size as the large firm.

Figure 6 presents evidence on the critical question as to whether explanatory power is improved significantly by controlling for industry. For both the CAPM in the upper panel, and the APT in the lower panel, there appears to be only a very slight cross-sectional relation between the R^2 of a large individual firm and an equal-size portfolio of smaller firms *within the same industry.* The scatter is almost as disperse as that observed in Figure 5 when there was no control on industry when forming the equal-size portfolio.

One might have anticipated significant differences across industries in the ability of pervasive economic factors to explain returns. Yet, the lack of inter-industry correlation between the R^2s of large firms and the R^2s of size- and industry-matched portfolios of smaller firms suggests that the explanatory power of pervasive factors is not very different from one industry to another.

To measure the extra explanatory power of industry events *in addition to* the explanatory power of general pervasive economic factors, further regressions were performed with each large firm in the sample using *that firm's* equal-sized industry-matched portfolio return as a supplemental regressor; i.e., the following

[9] The same method was employed in Section II to form size-matched portfolios without regard to industry.

Table I

Four-Digit Industries with Smaller Firm Portfolios Matched in Aggregate Size to the Largest Firms Within the Same Industry

Portfolios per Industry	Number of Industries	Frequency (%)
1	52	78.79
2	8	12.12
3	2	3.03
4	1	1.52
5	1	1.52
6	1	1.52
7	1	1.52
No. of Industries:		66
No. of Portfolios:		96

Industry	Number of Portfolios	Industry	Number of Portfolios	Industry	Number of Portfolios
1041	1	3541	1	4931	2
1211	1	3561	1	4941	1
1311	5	3569	1	5065	1
1382	1	3574	1	5311	2
2043	1	3612	1	5411	2
2211	1	3629	1	5912	1
2451	1	3662	3	5944	1
2621	1	3679	1	6025	2
2711	1	3694	1	6331	1
2819	1	3721	1	6411	1
2834	2	3732	1	6552	2
2869	1	3792	1	6711	7
2899	1	3822	1	6723	4
2911	2	3825	1	6792	1
3011	1	3841	1	6799	3
3069	1	4011	1	7011	1
3079	1	4511	2	7311	1
3312	1	4832	1	7379	1
3429	1	4899	1	7391	1
3494	1	4911	6	7392	1
3519	1	4923	1	8062	1
3533	1	4924	1	8911	1

regression was computed for each of the ninety-six large firms,

$$r_{j,t} = a_j + b_{1,j} f_{1,t} + \cdots + b_{k,j} f_{k,t} + c_j I_{j,t} + e_{j,t},$$

where $j = 1, \ldots, 96$, the f's are the systematic factors used previously ($k = 1$ for the CAPM and $k = 5$ for the APT), and $I_{j,t}$ is the return on the industry-matched portfolio of smaller firms. Table III gives the average increase in R^2 from adding the industry portfolio return, I_j. On average, the improvement is slightly greater for the CAPM than for the APT, which may indicate that the non-general-market, higher order APT factors, $f_2, \ldots, f_5$, have a differential impact across industries.

Table II

Large Firms and Portfolios Matched in Size and in Industry

	Largest Four-Digit Industry Code Firms						Size-Matched Portfolios			
Name	Size ($ Mill.)	SIC Code	No. of Months	CAPM Beta	CAPM R^2	APT R^2	No. of Matched Firms	CAPM R^2	APT R^2	Variance Ratio Vj/Vp
Mean Values	1845		57	1.025	0.2394	0.2872	6.20	0.3614	0.4426	1.8523
DOME MINES LTD	642	1041	56	1.084	0.0538	0.3133	3	0.0577	0.3031	2.4110
PITTSTON CO	531	1211	60	0.884	0.1642	0.3335	4	0.2281	0.1482	2.2960
ALLIED CORP	1077	1311	36	1.028	0.2128	0.2285	10	0.3628	0.6064	0.7915
CANADIAN PAC	2143	1311	39	1.035	0.2950	0.3186	18	0.2217	0.4717	1.0200
MESA PETE CO	952	1311	52	0.286	−0.0104	−0.0580	2	0.0766	0.3368	2.0640
OCCIDENTAL PETE	2008	1311	60	0.699	0.1079	0.3035	10	0.2233	0.4755	1.1720
TEXAS OIL & GAS CORP	2489	1311	41	1.233	0.2355	0.4756	15	0.2736	0.6085	1.1370
PENNZOIL CO	1409	1382	60	0.420	0.0269	−0.0017	12	0.2653	0.4519	0.9012
GENERAL MILLS INC	2231	2043	60	0.874	0.2639	0.3487	2	0.3036	0.3588	1.1170
BURLINGTON INDS INC	659	2211	60	0.898	0.1821	0.2675	6	0.3963	0.4834	1.3140
REDMAN INDS INC	159	2451	60	1.540	0.2923	0.3740	2	0.2431	0.3761	0.6720
INTERNATIONAL PAPER	2066	2621	60	1.218	0.3606	0.2957	3	0.4998	0.3784	1.2570
GANNETT INC	2115	2711	60	1.215	0.4104	0.5661	4	0.6534	0.6604	1.0960
UNION CARBIDE CORP	3385	2819	60	0.889	0.1714	0.1237	7	0.3710	0.3724	1.4010
AM. HOME PRODS.	6317	2834	60	0.808	0.2907	0.2749	7	0.4711	0.4167	0.8678
SMITHKLINE BECKMAN	5692	2834	60	0.791	0.2323	0.2137	3	0.5053	0.4811	0.9491
INTL. FLAVORS	896	2869	60	1.077	0.2747	0.2349	4	0.3897	0.3811	1.3050
ETHYL CORP	514	2899	60	1.098	0.2238	0.3281	3	0.4724	0.3281	1.6000
EXXON CORP	24770	2911	60	0.731	0.3047	0.3568	11	0.2290	0.4343	0.6568
SHELL OIL CO	11040	2911	33	0.163	−0.0265	0.0538	5	0.1826	0.4600	1.2950
GOODYEAR TIRE & RUBR	1983	3011	60	1.501	0.3504	0.2933	9	0.5635	0.5873	2.5940
TRINOVA CORP	261	3069	60	1.106	0.3847	0.3984	8	0.4200	0.5114	1.3080
RUBBERMAID INC	312	3079	60	1.245	0.2929	0.3465	9	0.1795	0.2511	0.4749
U S X CORP	1782	3312	60	0.636	0.0611	0.1268	7	0.2866	0.4571	0.8974
PARKER HANNIFIN	461	3429	60	1.825	0.4976	0.4497	5	0.5488	0.5950	2.7560
KEYSTONE INTL INC	268	3494	60	0.968	0.2046	0.3846	4	0.4169	0.5668	1.5130
COLT INDS INC DEL	692	3519	60	0.478	0.0002	−0.0629	3	0.5795	0.5717	5.8910
BAKER INTL CORP	1498	3533	54	1.115	0.1944	0.3630	8	0.2513	0.4919	1.0260

Table II—*Continued*

Name	Size ($ Mill.)	SIC Code	No. of Months	CAPM Beta	CAPM R^2	APT R^2	No. of Matched Firms	CAPM R^2	APT R^2	Variance Ratio Vj/Vp
Largest Four-Digit Industry Code Firms							Size-Matched Portfolios			
SUNDSTRAND CORP	634	3541	60	1.379	0.4270	0.4695	6	0.3429	0.3773	1.0150
TRICO INDS INC	69	3561	54	0.774	0.0792	0.1531	2	0.4071	0.4667	2.5620
PALL CORP	432	3569	60	1.180	0.3057	0.3063	8	0.2190	0.3091	1.5170
WANG LABS INC	2002	3574	60	1.674	0.3655	0.3987	3	0.4948	0.4860	1.5840
ANALOG DEVICES INC	198	3612	60	1.432	0.2350	0.4328	4	0.3997	0.6238	2.3300
TECH SYM CORP	62	3629	60	1.762	0.3321	0.3950	2	0.1996	0.3752	2.4340
I T T CORP	3589	3662	60	1.130	0.3208	0.3789	12	0.6077	0.5958	0.9131
R C A CORP	1519	3662	45	1.071	0.2868	0.3068	11	0.4345	0.6176	0.3854
TELEDYNE INC	1993	3662	60	0.929	0.1511	0.2167	10	0.4081	0.5717	0.9747
TDK CORP	1634	3679	60	0.822	0.0753	0.0486	9	0.4584	0.6098	1.9570
CHAMPION SPARK PLUG	312	3694	60	0.919	0.2102	0.2885	2	0.4280	0.4370	1.0270
BOEING CO	2292	3721	60	1.172	0.2812	0.4366	5	0.4333	0.4409	1.8280
CHRIS CRAFT INDS INC	86	3732	60	0.572	0.0555	0.1044	2	0.3643	0.5042	0.9588
FLEETWOOD ENT	247	3792	60	1.549	0.3062	0.4083	3	0.4420	0.5691	0.6618
JOHNSON CTLS INC	369	3822	60	0.840	0.1860	0.3571	3	0.3843	0.5429	1.2920
PERKIN ELMER CORP	985	3825	60	1.450	0.3056	0.3220	4	0.4710	0.5930	1.4930
AMERICAN HOSP SUP	1754	3841	38	1.259	0.2132	0.1829	6	0.4269	0.4280	3.0860
CANADIAN PAC LTD	1774	4011	60	1.283	0.3889	0.3992	3	0.4108	0.4740	1.2880
DELTA AIR LINES INC	1248	4511	60	0.896	0.1229	0.1456	7	0.3099	0.3841	1.4630
NWA INC	704	4511	60	0.900	0.1239	0.1515	2	0.2203	0.3415	0.8521
ABC	1291	4832	40	0.514	0.0058	0.0022	2	0.2659	0.3866	2.3220
COM SAT	501	4899	60	0.975	0.1924	0.1921	4	0.1915	0.3386	1.6100
AMERICAN ELEC PWR	2950	4911	60	0.572	0.1683	0.2377	8	0.2143	0.4199	1.6570
COMMONWLTH EDISON	3201	4911	60	0.446	0.1051	0.0888	7	0.2088	0.3035	1.2760
PACIFIC GAS & ELEC CO	3489	4911	60	0.428	0.0894	0.2368	5	0.1834	0.2801	1.3000
SO CAL EDISON	3114	4911	60	0.428	0.0920	0.3535	7	0.1521	0.3593	1.2780
SOUTHERN CO	2820	4911	60	0.636	0.2833	0.3869	5	0.0969	0.1720	0.5750
TEXAS UTILS CO	2447	4911	60	0.625	0.1876	0.2418	5	0.1721	0.1715	0.5900
HOUSTON NAT GAS	1252	4923	34	0.702	0.0314	−0.0689	4	0.2590	0.3479	2.0390

HIRAM WALKER RES	1108	4924	49	0.770	0.1545	0.2116	4	0.3578	0.2738	3.5840
PHILADELPHIA ELEC CO	1835	4931	60	0.536	0.0992	0.1930	8	0.1162	0.4113	2.5840
PUBLIC SVC ENT	1883	4931	60	0.443	0.0815	0.1781	3	0.2117	0.4026	1.9290
ENTERRA CORP	207	4941	60	0.946	0.0652	0.1691	5	0.1491	0.2004	6.4350
ARROW ELECTRS INC	54	5065	60	1.190	0.0913	0.1072	4	0.3554	0.4929	1.8470
J C PENNEY INC	3110	5311	60	1.003	0.3429	0.4287	5	0.3208	0.3645	0.8780
SEARS ROEBUCK & CO	7979	5311	60	1.671	0.6060	0.7173	21	0.4807	0.5488	1.9340
KROGER CO	1099	5411	60	0.434	0.0611	0.0528	4	0.1300	0.1834	0.9061
WINN DIXIE STORES	961	5411	60	0.865	0.2714	0.2489	3	0.3641	0.3181	0.5295
ECKERD JACK CORP	817	5912	44	1.217	0.3539	0.3813	5	0.5523	0.6553	0.8864
ZALE CORP	132	5944	53	0.748	0.1651	0.1876	2	0.2422	0.3096	0.8217
CHASE MANHATTAN	1216	6025	60	1.633	0.5041	0.5241	3	0.4368	0.4466	1.5980
MFCTR HANOVER	916	6025	60	1.267	0.2932	0.3739	3	0.4989	0.5435	1.3200
GENERAL RE CORP	1925	6331	60	0.950	0.2247	0.2352	3	0.4093	0.4460	1.6520
MARSH & MCLENNAN	1258	6411	60	0.818	0.2010	0.2124	3	0.4312	0.4000	1.0860
LENNAR CORP	124	6552	60	1.901	0.3216	0.3240	6	0.3851	0.5629	1.7420
PULTE HOME CORP	139	6552	60	2.770	0.4296	0.5472	4	0.3207	0.4364	4.2960
BANKAMERICA CORP	2454	6711	60	1.569	0.2606	0.3355	6	0.4618	0.4300	3.6460
C I G N A CORP	2894	6711	60	1.140	0.3781	0.3911	15	0.6346	0.5895	1.7910
C S X CORP	1863	6711	60	1.571	0.6111	0.5829	8	0.4697	0.5510	2.2270
G T E CORP	5251	6711	60	1.018	0.4364	0.4616	15	0.7076	0.7299	1.7770
J P MORGAN & CO INC	2175	6711	60	1.205	0.4554	0.4806	9	0.3922	0.3949	0.7903
NORFOLK SOUTHERN	3546	6711	60	0.955	0.4366	0.3566	13	0.7394	0.7884	0.8488
UNION PAC CORP	3689	6711	60	1.197	0.4881	0.5368	9	0.6611	0.7078	1.4130
A S A LTD	373	6723	60	0.509	0.0161	0.1510	7	0.7078	0.7011	11.7400
LEHMAN CORP	546	6723	60	0.658	0.4857	0.4257	7	0.4170	0.4027	1.9410
TEXAS EASTN CORP	1205	6723	60	1.026	0.1835	0.2056	14	0.5416	0.5254	10.4900
TRI CONTL CORP	552	6723	60	0.827	0.5699	0.5366	7	0.2423	0.3141	2.6080
SOUTHLAND RTY CO	759	6792	39	1.116	0.1393	0.2915	5	0.2386	0.4049	2.1480

Table II—*Continued*

Name	Size ($ Mill.)	SIC Code	No. of Months	CAPM Beta	CAPM R^2	APT R^2	No. of Matched Firms	CAPM R^2	APT R^2	Variance Ratio Vj/Vp
	Largest Four-Digit Industry Code Firms						Size-Matched Portfolios			
GENERAL GROWTH	117	6799	36	0.418	0.0254	−0.0185	2	0.0240	0.0403	1.4540
PERMIAN BASIN RTY	460	6799	60	0.644	0.1097	0.2327	13	0.4060	0.4845	4.7720
SAN JUAN BASIN RTY	443	6799	60	0.684	0.1552	0.2692	14	0.3728	0.5049	1.7210
HOLIDAY CORP	1159	7011	60	0.777	0.1336	0.1244	4	0.3431	0.3965	1.6410
INTERPUBLIC GROUP CO	167	7311	60	1.265	0.3980	0.2912	2	0.1909	0.1874	1.0570
COMDISCO INC	213	7379	60	2.342	0.3244	0.3310	4	0.3639	0.4408	3.0870
FLOW GEN INC	85	7391	60	1.805	0.2150	0.3962	2	0.3237	0.3853	1.2050
EQUIFAX INC	244	7392	60	1.159	0.3535	0.3479	4	0.4314	0.3965	0.6341
HOSPITAL CORP AMER	2262	8062	60	0.943	0.1838	0.2081	2	0.4192	0.4075	0.9686
E G & G INC	584	8911	60	1.290	0.3809	0.3171	12	0.5645	0.6136	1.7770

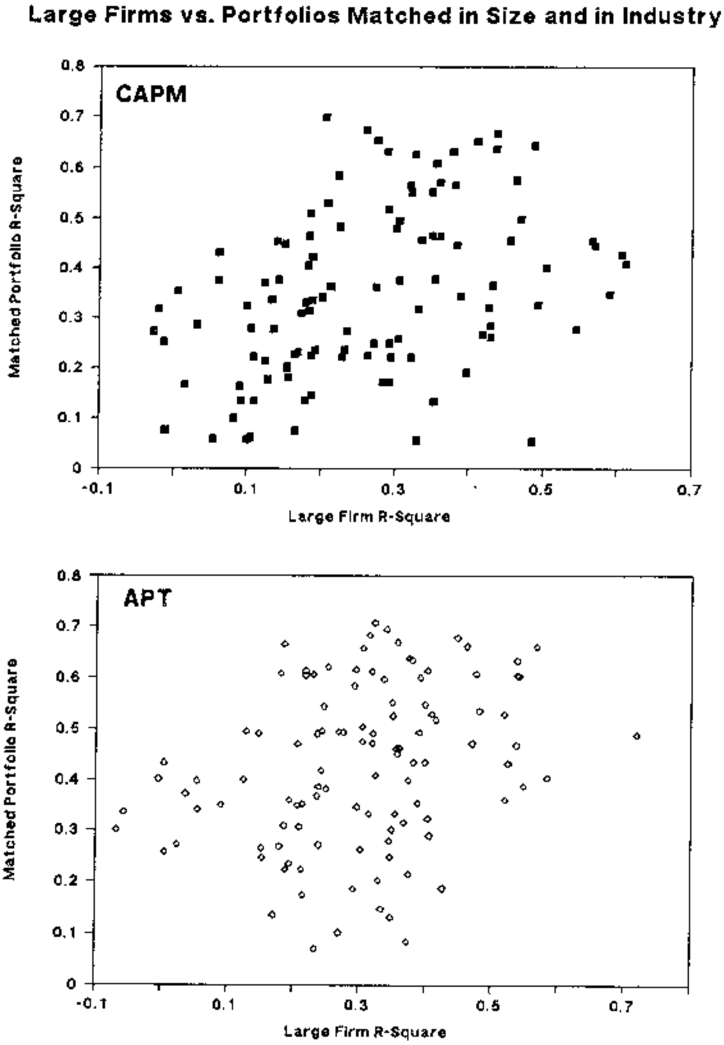

Figure 6.

Table III

Average R^2s for a Sample of Ninety-Six Large Firms with a Size- and Industry-Matched Smaller Firm Portfolio Used as an Additional Regressor

Model	Without Industry Factor	With Industry Factor	Difference
CAPM	.2394	.3438	.1043
APT	.2872	.3607	.0735

The upper and lower panels of Figure 7 show the CAPM and APT R^2s without and with the industry portfolio included in the regression. The addition of an industry "factor" improves the relation between the explanatory powers of single and multiple systematic factors. Figure 8 shows that increases in R^2 obtained by adding an industry regressor are related between the two models.

In Figure 9, the difference in explanatory power between the multiple-factor APT and the single-factor CAPM is compared with and without the industry regressor. The scatter has a curious pattern: there are a number of seeming outliers that have a much bigger APT R^2 − CAPM R^2 without the industry regressor than with the industry regressor. These points are located around the abscissa toward the right-hand side of the figure.

Most of these "outliers" are in just two industries, petroleum and regulated utilities. The names of some of the companies are indicated on the figure for easy

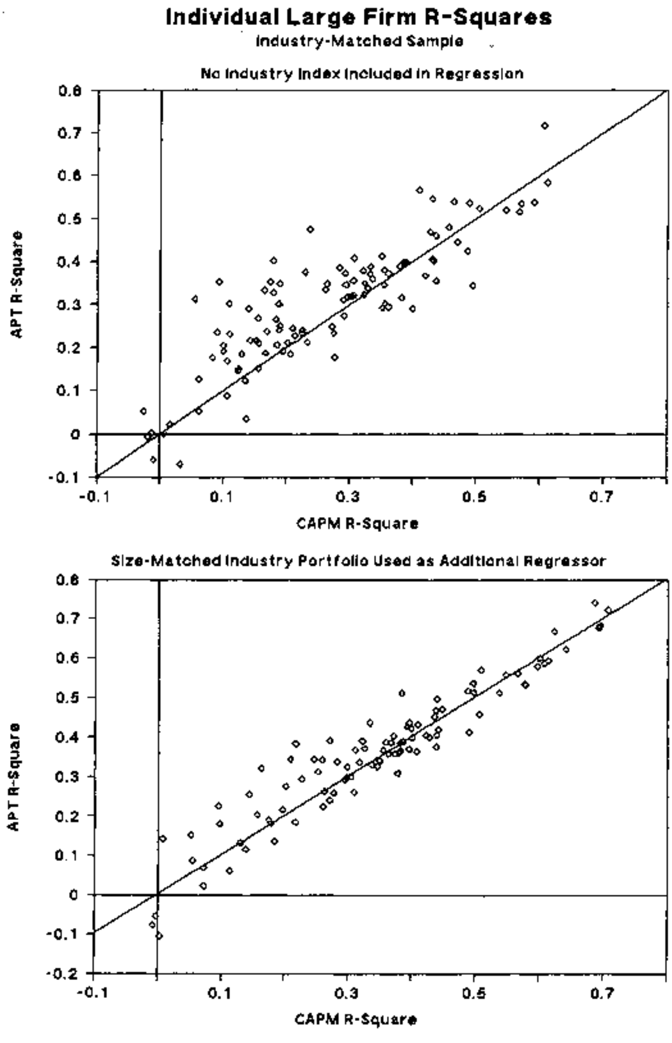

Figure 7.

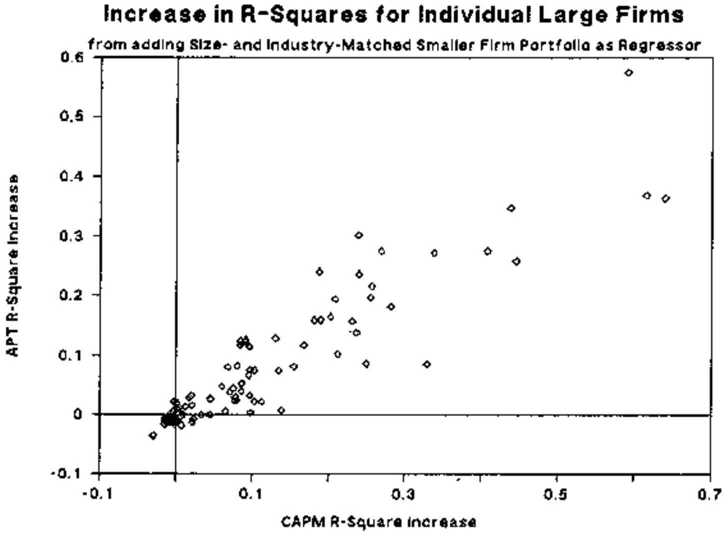

Figure 8.

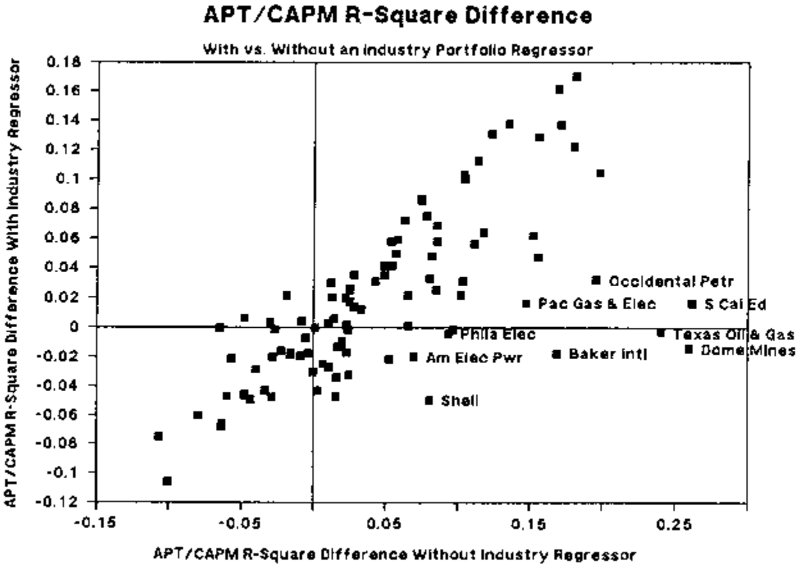

Figure 9.

reference. Evidently, some of the higher order APT factors are especially important for oils and utilities, and these industries must have characteristics that are unlike those of a typical stock in a broad market index.

IV. R^2 and Firm-Specific News

After removing the explanatory influences of broad, pervasive economic influences and of industry influences, the current paradigm of financial theory attributes the remaining volatility to *unique* information about the firm. With hindsight, such information should be discoverable by tracing its appearance in the financial press, an implicit presumption being that anything insignificant enough to be ignored by the media is probably also immaterial in its impact on market prices.

It should, therefore, be possible to increase greatly the R^2 of pervasive factors by considering only periods when there is *no* reported news about the firm or about its industry. When there is no firm-specific news, *all* of the observed changes in prices would presumably be explained by pervasive factors. We can investigate this part of the paradigm by expurgating those data observations that coincide with news stories in the financial press and then running the same regressions as before on the remaining information-cleansed observations.

In the U.S. market, there are two prominent sources of financial news, the Dow-Jones news service, or "Broad Tape," which contains a real-time record of major news developments of all kinds, and *The Wall Street Journal*, which records not only major developments but significant analyses of these events by staff writers and others. Given the available resources, it was not possible to collect news dates from these sources for every listed firm. Thus, for ease of comparison with the previous sections of the paper, these news events were collected for the ninety-six large firms whose industry influences were analyzed in Section III.

Daily data were used in the analysis of this section because of the high frequency with which large firms are mentioned in the financial press. For many large firms, few months pass *without* a mention; there are not enough news-free *months* to constitute an adequate sample of observations. In contrast, consistent daily mention is unusual, so a large number of no-news daily observations for most firms can be collected even if a few days are excluded around the publication date of each news item. On average over all ninety-six stocks, 23.7 percent of the daily observations were excluded by being either the day of a news event or the preceding day; this means that about 965 of the 1264 trading days in the 1982–86 calendar years are available for the non-news analysis.

It is probably important to expurgate the date of the news publication *and* at least one day before publication. For items that appear only in the *Wall Street Journal* and not on the Broad Tape, the news might actually have been publicly available on the previous day. Even for items on the Broad Tape, insiders sometimes receive forewarning, and the price may move in response to their trades. We report below on the consequences of widening the window of excluded dates around each news appearance.

Figure 10 presents the resulting R^2s contrasting two effects, the CAPM single-factor model versus the APT multiple-factor model and the presence or absence of firm-specific news. The top panel shows R^2s with all available daily observations; the bottom panel shows R^2s for regressions excluding the day of a news appearance and the previous trading day. The 45-degree line lies below *all* observations, which means that *every* stock had a higher R^2 in the multiple-factor daily regressions.[10]

[10] This contrasts with only about three quarters of the stocks with higher multiple-factor R^2s using monthly data.

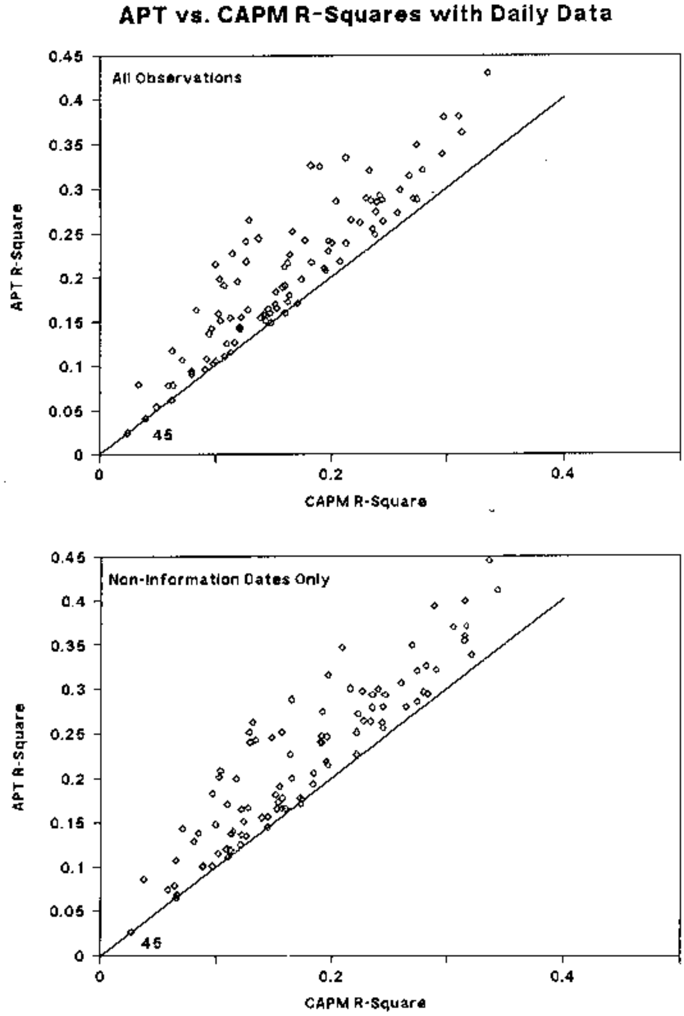

Figure 10.

Although it is not ocularly obvious in Figure 10, the R^2s are slightly higher when news dates are excluded. The average CAPM R^2 is .163 with all data and .177 with non-news data. The average APT R^2 is .205 with all data and .221 with non-news data. This is extremely disappointing in that the average R^2, when there is no public news of any kind, not even a mention of the firm in the text of a story about *any* subject, is hardly different from the average R^2 using all observations.

Although the average R^2 is only trivially increased by excluding news dates, there are a few firms with substantial increases. Figure 11 plots, for the CAPM in the upper panel and the APT in the lower panel, the cross-firm relation between R^2s with and without news dates (defined as the date of publication plus the preceding day). The biggest improvements in explanatory power, not surpris-

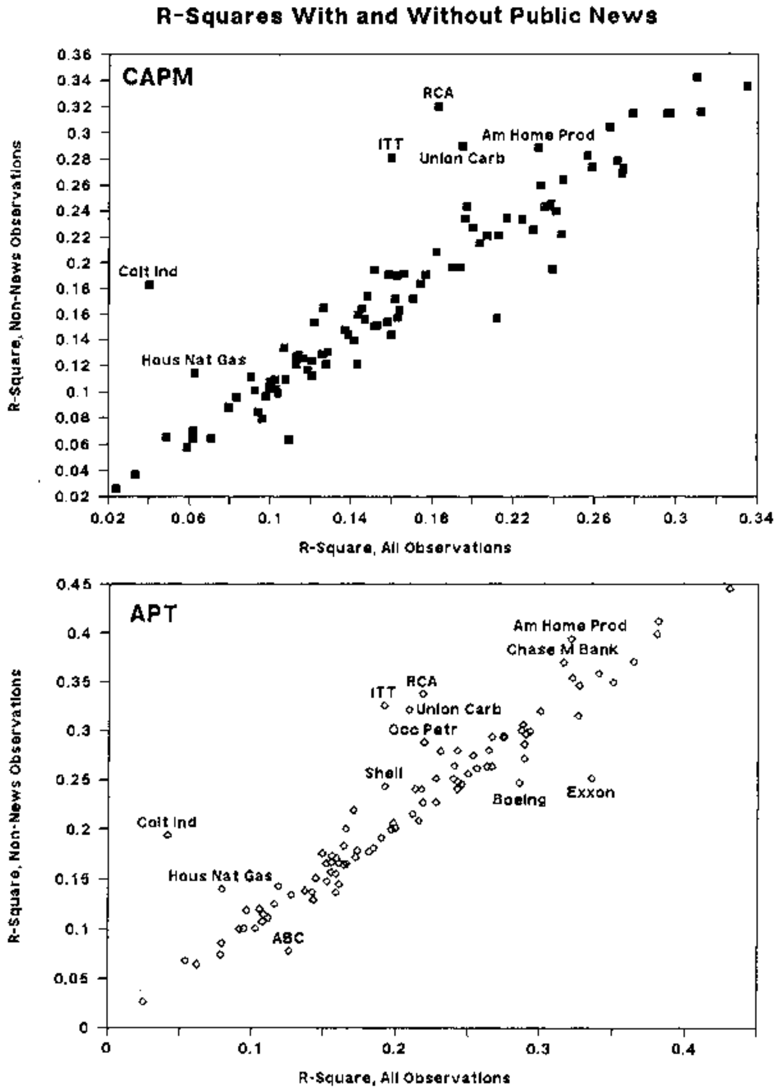

Figure 11.

ingly, were associated with firms involved in takeover situations (e.g., RCA, Colt, ITT), or firms such as Union Carbide that suffered major disasters. But perhaps the most striking result portrayed in Figure 11 is the rather strong connection between the R^2s with and without news dates. Except for the few outliers just mentioned, the degree of explanatory power seems to be similar, firm by firm, regardless of the particular observations used in the regression.

Consistent with this observation, Figure 12 shows that there is little relation between the increase in R^2 obtained by excluding news dates and the total number of dates within the sample on which news about the firm appeared in the financial press. There is a very slight positive relation in Figure 12, but if

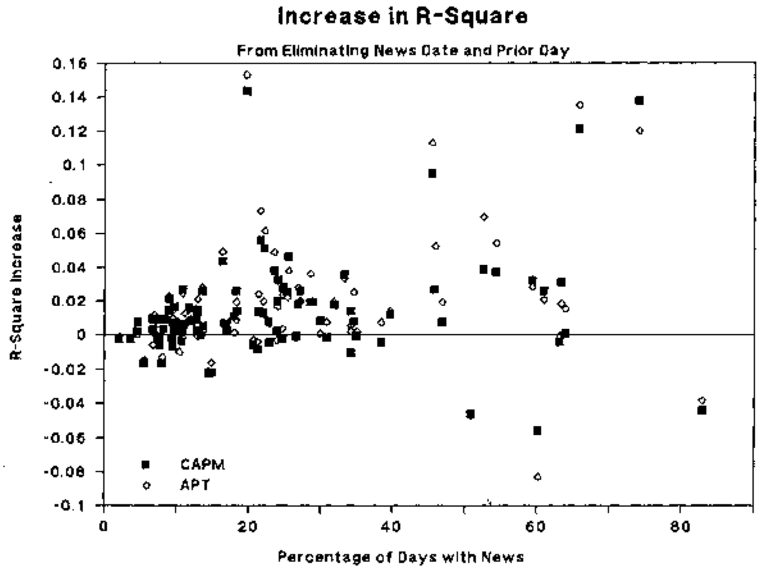

Figure 12.

non-news-date returns had been easier to explain with the systematic factors, the depicted relation would have been stronger.

A. Estimating the News Process with a Simple Mixture Model

Even though the results are disappointing, there is some very slight evidence, particularly from those firms that have experienced extraordinary events such as merger, that public news reduces the explanatory power of systematic factors. Thus, it seems worth investigating the phenomenon a bit further, if only to help in understanding why the power is so poor on average.

It is natural to model the overall stochastic process of price changes with a mixture of probability models, i.e., a different probability distribution during news and non-news periods. This idea has been investigated previously by a number of different authors, most of whom have presented relatively sophisticated statistical models.[11] The principal feature of mixed distributions emphasized by most authors involves the higher moments, particularly the fourth moment or the sample kurtosis. Kurtosis can reveal something about the probability of information and the difference between the information-related distribution and the non-information-related distribution of returns. Damodaran [1], for instance, finds that kurtosis is closely correlated to such measures of information as the number of analysts following a stock and the number of *Wall Street Journal* stories, *inter alia.*

[11] One of the first, and still one of the most interesting models was by Press [8]. Press' model postulated a mixture of normal distributions; the number of distinct normals in the mixture was itself a Poisson-distributed random variable. There are many papers in this tradition (e.g., Epps and Epps [2], Morgan [6], and Westerfield [12]). Harris [3], Damodaran [1], and Harris [4] are recent contributions of the same genre.

A very simple model of this type can be constructed as follows. Assume that a market-model residual, e_t, can be decomposed as

$$e_t = x_t + g_t y_t,$$

where

$$x = \text{background trading noise (no news)},$$

$$y = \text{news-related residual return},$$

$$g_t = 1 \text{ if there is news on date } t, \quad \text{otherwise } 0.$$

Assume $\mathrm{E}(e) = \mathrm{E}(x) = \mathrm{E}(y) = \mathrm{E}(xy) = \mathrm{E}(xg) = \mathrm{E}(gy) = 0.$

Define

$$V_j = \text{variance of } j, \quad (j = e, x, y)$$

$$p = \mathrm{Prob}\{g_t = 1\}.$$

If x and y are normally distributed,

$$\mathrm{E}[x + gy]^4 = 3[V_x^2 + pV_y^2 + 2pV_xV_y],$$

and the "kurtosis" is

$$K = \mathrm{E}(e^4)/(V_e^2) = 3\{1 + [p(1-p)]/[V_x/V_y + p]^2\} > 3.$$

Define

$$Q = K/3 - 1 = [p(1-p)]/[R + p]^2, \tag{1}$$

where $R = V_x/V_y$ is the ratio of noise to news variances. Q can be directly estimated, but there is no straightforward way to decompose the estimate of Q into its components R and p. However, provided that Q is positive, R *can* be bounded. As it turns out, for every one of the ninety-six firms in our sample, the sample value of Q is positive (which is a result typical of daily stock-return data).

Solving equation (1) for p gives

$$p = \{1 - 2RQ \pm [1 - 4RQ(1 + R)]^{1/2}\}/[2(Q + 1)],$$

and, to obtain a non-imaginary p,

$$R + R^2 < 1/(4Q),$$

which implies that the maximum value of R is given by

$$R_{\max} = \{[1 + 1/Q]^{1/2} - 1\}/2. \tag{2}$$

This is, therefore, the minimum ratio of news variance to noise variance. At this minimum ratio of variances, the probability of news is given by

$$p_{\max} = [1 - 2R_{\max}Q]/[2(Q + 1)]. \tag{3}$$

The cross-sectional frequency distribution of the reciprocal of $R_{\max}$ is given in the top panel of Figure 13. The bottom panel gives the corresponding frequency distribution for $p_{\max}$. The average values of these statistics, along with several

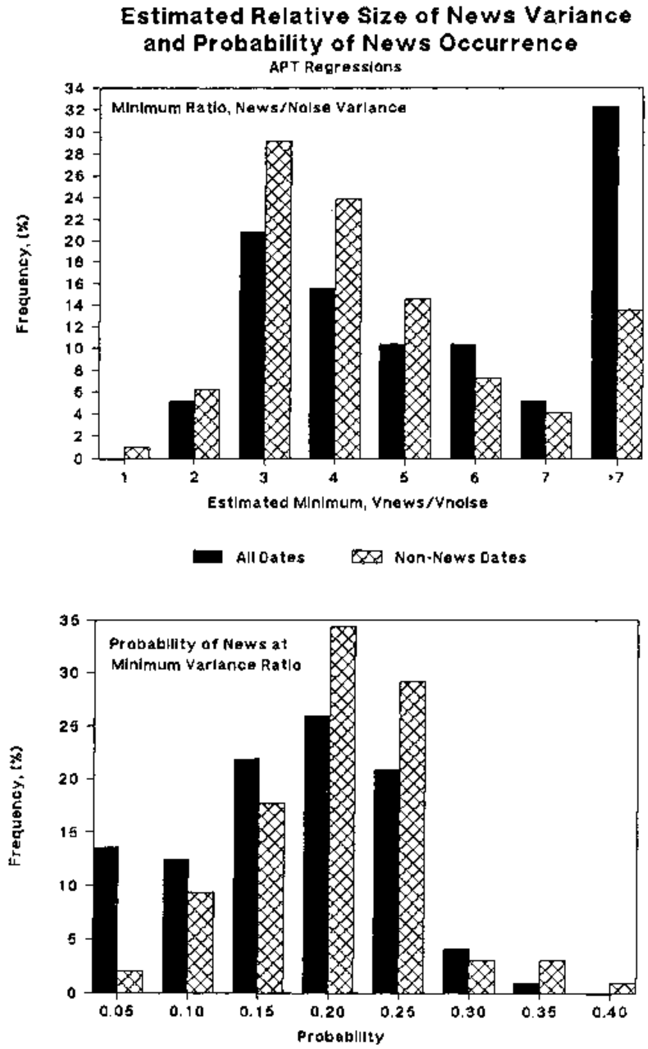

Figure 13.

other pertinent average values over the ninety-six sample firms, are listed in Table IV.

Referring first to Figure 13, notice that the entire distribution of the news/noise minimum-variance ratio, $1/R_{max}$, is much farther to the right when all dates are included in the regressions than when the regressions are calculated excluding the dates of news events plus one preceding day. With all dates included, the *average* value of $1/R_{max}$ is slightly over 20 (for both the CAPM and APT). When the news dates in the financial press are excluded, the average ratio drops to about 7.75. (See Table IV.)

The top panel of Figure 13 reveals that almost every stock has an empirical

Table IV

Volatility, Kurtosis, Estimated Information Statistics, and R^2 for Windows Around Public News Dates (Average Values over Ninety-Six Industry-Matched Large Firms)

	Standard Deviations		Kurtosis	Minimum V_{News}/V_{Noise} $(1/R_{max})$	P\{News\} (p_{max})	R^2
	Returns	Residuals				
Including All Daily Observations						
CAPM	23.454	21.488	17.725	20.457	0.14393	0.16282
APT	23.454	20.927	17.935	20.738	0.14357	0.20532
Excluding One Day Before and the Day of the News						
CAPM	22.491	20.427	8.228	7.750	0.17594	0.17650
APT	22.491	19.843	8.256	7.785	0.17780	0.22102
Excluding Two Days Before and the Day of the News						
CAPM	22.419	20.344	7.829	7.209	0.18150	0.17821
APT	22.419	19.738	7.800	7.164	0.18522	0.22367
Excluding Two Days Before Through One Day After the News						
CAPM	22.263	20.192	7.371	6.582	0.18162	0.17880
APT	22.263	19.557	7.312	6.502	0.18520	0.22548

estimate of news volatility, V_y, several times the volatility of background trading noise, V_x. This is the case for both the complete sample including every daily observation and the sample censored of news dates in the financial press.

Table IV presents sample averages for several variables and for three different windows around news dates. All of the censored samples display much lower kurtosis and much lower estimated news volatility than the full (uncensored) sample, but widening the window to censor more days before and after the news-publication date seems to have only a minor additional effect.

The results in Table IV indicate that the sample volatility for the total return or for the residual return is almost as large in the censored samples as in the complete sample. This is in striking contrast with kurtosis, and it reveals immediately that the probability is rather small of observing the more volatile news-related member of the mixture of distributions. In keeping with this inference, the average value of p_{max} for the complete sample is about .14.

The average probability of "news" is even higher in the censored samples than in the complete sample. This seems to imply that the financial press misses a great deal of relevant information generated privately. However, the volatility induced by private information is lower than that induced by the big newsworthy events. Evidently, the very simple model presented here should be generalized at least enough to subsume a three-distribution mixture, say y_t for public news, x_t for noise, and an additional z_t for private information.

B. Caveats

There are a few things that were considered but omitted from the paper because of space or time limitations.

1) In daily regressions of stock returns on indicia (including APT factors),

there is a serious problem of coefficient bias because of non-synchronous trading.[12] The regressions above were estimated also with leading and lagged values of the explanatory variables, but this had very little impact on the R^2s. Large firms are less susceptible to this problem anyway, and apparently the problem is more material for slope coefficients than for R^2s.

2) The CAPM market index and the APT factors are composed of returns on traded securities. A certain proportion of these securities would experience firm- or industry-specific news on any given day; thus, the indexes and factors must contain extra volatility arising from such unique events (relative to the volatility induced by purely general economic influences unpolluted by specific firm news). In principle, one could construct better factors by eliminating every stock with news on each day. It seems doubtful that this would materially increase R^2, but one cannot know for sure until it is tried.

3) Several authors have suggested that volatility of asset prices can be better explained by psychological factors, fads, etc., than by information. The results above are actually consistent with such a view. After all, the unexplained volatility on non-news dates could conceivably indicate not that private information is being incorporated into prices but that mania is periodically gripping investors. Perhaps the components of the distributional mixture consist of low-variance background liquidity traders and high-variance traders striken by either panic or euphoria. It would be nice to have a method for detecting the difference.

V. Summary and Conclusions

Most scholars and practitioners have resigned themselves to poor *ex ante* forecasting power for stock price changes. However, the current paradigm of financial markets implies much better explanatory power *ex post*. With hindsight, stock price changes should be explainable by general systematic influences, industry influences, and events unique to the firm. This paper attempted an empirical investigation of the paradigm.

Regressions of individual monthly stock returns on either a single market index or on multiple factors produced explanatory power, as measured by the average adjusted R^2, in the neighborhood of 0.30. Adding an industry factor increased the average R^2 to around 0.35.

For the decile of the largest AMEX and NYSE firms, portfolios of smaller firms were constructed to match each large firm in aggregate size. These portfolios had much higher R^2s than their corresponding size-matched large individual firms, thereby indicating that diversification by the large firm is not much of an explanation for the slightly larger explanatory power. There was no perceptible cross-sectional relation between the R^2 of the large firm and the R^2 of its aggregate size-matched portfolio.

Also, portfolios were constructed to match a sample of large firms *both* in size *and* in industry. Again, there is little relation between size and explanatory power. Perhaps most surprising, there was again virtually no cross-sectional relation between the R^2 of a large firm and the R^2 of its size-matched and industry-

[12] See, e.g., Scholes and Williams [10].

matched companion portfolio. This indicates that explanatory power by systematic economic factors is not very different across industries.[13]

Daily data were employed here to investigate the incidence and impact of unique news about the firm. With daily data, the average R^2 of the same sample of large firms dropped to around 0.20. Every mention of each firm in either the *Wall Street Journal* or over the Dow-Jones Broad Tape was defined as an information event. Regressions on systematic factors were conducted *only* with *non*-information dates. Even with this information-censored data, the average explanatory power was only marginally better. For example, using the multiple-factor APT pervasive factors as regressors, the average R^2 increased from .205 using all days in calendar years 1982–86 to only .225 after excluding the information event, two days before, and one day afterward.

There was, however, a dramatic decline in sample kurtosis from excluding public news events. A simple mixture-of-distributions model (i.e., one distribution for a non-news date and a second, higher variance distribution for a news date) yielded two suggestive results: (1) the average minimum ratio of news variance to background noise variance was around 20 for all sample dates but only around 7 for non-news dates, and (2) the estimated probability of news was modest but material for *both* identifiable public news dates and for other dates. This seems to imply the existence of either private information or else occasional frenzy unrelated to concrete information.

[13] Although *explanatory power* is not very industry related, the coefficients in the regressions *are* very different in different industries.

REFERENCES

1. Aswath Damodaran. "Information Frequency: Measures and Implications." Working paper, Graduate School of Business Administration, New York University, September 1986.
2. T. W. Epps and M. L. Epps. "The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implications for the Mixture-of-Distributions Hypothesis." *Econometrica* 44 (March 1976), 305–21.
3. Lawrence Harris. "Cross-Security Tests of the Mixture of Distributions Hypothesis." *Journal of Financial and Quantitative Analysis* 21 (March 1986), 39–46.
4. ———. "Transaction Data Tests of the Mixture of Distributions Hypothesis." *Journal of Financial and Quantitative Analysis* 22 (June 1987), 127–41.
5. John Lintner. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." *Review of Economics and Statistics* 47 (February 1965), 13–37.
6. Ieuan Morgan. "Stock Prices and Heteroskedasticity." *Journal of Business* 49 (October 1976), 496–508.
7. Jan Mossin. "Equilibrium in a Capital Asset Market." *Econometrica* 34 (October 1966), 768–83.
8. James S. Press. "A Compound Events Model for Security Prices." *Journal of Business* 40 (July 1967), 317–35.
9. Stephen A. Ross. "The Arbitrage Theory of Capital Asset Pricing." *Journal of Economic Theory* 13 (December 1976), 341–60.
10. Myron Scholes and Joseph Williams. "Estimating Betas from Nonsynchronous Data." *Journal of Financial Economics* 5 (December 1977), 309–27.
11. William F. Sharpe. "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk." *Journal of Finance* 19 (September 1964), 425–42.
12. Randolph Westerfield. "The Distribution of Common Stock Price Changes: An Application of Transactions Time and Subordinated Stochastic Models." *Journal of Financial and Quantitative Analysis* 12 (December 1977), 743–65.