FINANCIAL SERVICES REVIEW, 1(2):101-108
ISSN: 1057-0810

Should Individual Investors Avoid the Stock Market Outside of January?

Steven V. Mann
Donald P. Solberg

Recent studies suggest that there is no reward for bearing risk outside of January, implying that individuals should invest in common stocks only in January. The purpose of this study is to demonstrate that this conclusion is far too strong given existing empirical evidence. Our results suggest that inferences drawn from the evidence can be altered greatly through small changes in the way the empirical question is addressed. There is sufficient evidence to doubt the conclusion that individuals are not compensated for the risk of participating in the stock market outside of January.

Recent studies present a disturbing picture of the relationship between risk and return.[1] Simply put, researchers suggest that there is no reward for bearing risk outside of January. The return foregone by holding Treasury bills rather than risky stocks in months other than January is virtually zero.

If true, the message to individual investors is clear, albeit surprising. Based on these studies, the rational investor would invest in common stocks during the month of January, then shift funds into riskless assets for the remainder of the year. The validity of this strategy is especially important to the individual due to transaction costs incurred during portfolio rebalancing.

Clearly, if there is no reward to risk outside of January, it is puzzling that rational investors would choose to hold stocks in the remaining eleven months of the year. In this paper, we demonstrate that this conclusion is far too strong given existing empirical evidence.

The most troubling studies are by Chang and Pinegar (1988a and 1988b), which examine excess monthly returns of long-term corporate over government bonds, stocks over government bonds, and individual stocks over bonds of the same firm. They conclude that investors receive no risk premia February through December. Chang and Pinegar's results appear especially powerful because their tests rely

Steven V. Mann • College of Business Administration, University of South Carolina, Columbia, SC 29208. **Donald P. Solberg** • Federal Home Loan Mortgage Corporation, McLean, VA 22102.

on the least restrictive model imaginable—that investors demand incremental returns for taking greater risk.

Our analysis suggests that this approach is deceptively simple. The test of whether a risk premium differs significantly from zero is really a test of whether the average risk premium is "large" relative to its standard error. This test is sensitive to both the market price of risk (which varies over time) and the size of the return sample. Furthermore, relatively minor changes in the research question can produce the opposite conclusion.

Given these conflicting results, one must attempt to make a sensible interpretation of the available evidence. One could conclude that investors are irrational and require compensation for risk bearing only in January, even if it runs counter to our intuition. Alternatively, the empirical tests and interpretation of these tests may be at fault. Our own results place in doubt the conclusion that equity risk premia are zero in February through December. We leave it to the reader to decide if investors are irrational for holding equity in non-January months.

Data

We examine excess returns for bearing risk with two equity portfolios. The first, Common, is the Standard and Poor's Composite Index. The risk premium is the difference between monthly returns on this index and returns on one-month Treasury bills. The second portfolio, Small, is a value-weighted portfolio of the smallest quintile stocks in terms of market value on the New York Stock Exchange.[2] Similarly, the risk premium is monthly returns on this portfolio less returns on one-month Treasury bills. Return series for these portfolios are taken from the *Stocks, Bonds, Bills and Inflation 1989 Yearbook.*[3]

Empirical Results

Table 1 presents average risk premia and their standard deviations for each month over the period 1926-1988. Table 1 also presents t-statistics for the null hypothesis that mean monthly risk premia are zero using a one-tail test. Statistical significance at the five percent level requires t-statistics of at least 1.645. While January returns safely meet this benchmark, the other months meet the standard infrequently. Generally, the results of Table 1 are similar to those reported in other studies.

Although the methodology in these tests appears straightforward, similar statistical tests can produce widely differing conclusions. We use two different approaches. The first examines the question: Are monthly risk premia positive if we pool the eleven monthly risk premia for all non-January months into a single sample? This method increases the number of observations in the test and increases the precision of the sample estimates accordingly. Furthermore, this test answers

TABLE 1.
Summary Measures for the Individual Monthly Risk Premia

(Sample period: 1926–1988)

	Common			*Small*		
	Arithmetic mean	*Standard deviation*	*t-statistic*	*Arithmetic mean*	*Standard deviation*	*t-statistic*
Jan	0.014	0.049	2.19	0.068	0.091	5.93
Feb	0.002	0.043	0.43	0.014	0.064	1.74
Mar	0.001	0.054	0.09	-0.001	0.083	-0.14
Apr	0.011	0.075	1.12	0.010	0.103	0.77
May	-0.001	0.062	-0.14	-0.001	0.112	-0.04
Jun	0.012	0.059	1.68	0.009	0.085	0.82
Jul	0.017	0.068	2.00	0.022	0.086	2.01
Aug	0.018	0.066	2.15	0.017	0.108	1.28
Sep	-0.012	0.063	-1.55	-0.010	0.102	-0.80
Oct	-0.002	0.065	-0.23	-0.013	0.088	-1.17
Nov	0.010	0.054	1.47	0.009	0.074	0.95
Dec	0.013	0.038	2.64	0.006	0.064	0.71

TABLE 2.
Summary Measures for the Monthly Non-January Risk Premia

(Sample period: 1926–1988)

Portfolio	*Arithmetic mean*	*Standard deviation*	*t-statistic*
Common—Bills	0.0062	0.0599	2.72
Small—Bills	0.0056	0.0895	1.64

TABLE 3.
Summary Measures for the Eleven Month February-December Risk Premia

(Sample period: 1926–1988)

Portfolio	*Arithmetic mean*	*Standard deviation*	*t-statistic*
Common—Bills	0.070	0.203	2.74
Small—Bills	0.068	0.327	1.66

more directly the question of whether risk premia are positive on average over all non-January months.

Table 2 presents means and standard deviations for monthly non-January risk premia of our two equity portfolios. Utilizing a one-tail test, Table 2 also presents t-statistics for the null hypothesis that mean risk premia are zero. The mean risk premia for the Common and Small equity portfolios are .62% and .56%, respectively; they are positive and statistically significant at conventional levels. Contrary to the findings of Chang and Pinegar and others, these results suggest that equity investors *are* rewarded on average for bearing risk in non-January months.

A second set of tests asks if investors are rewarded for risk bearing over eleven-month holding periods, February through December. A series of eleven-month returns are computed for the equity portfolios for each year 1926-1988.[4] Returns for investing in the corresponding series of eleven one-month Treasury bills are also computed. Finally, we compute mean differences between equity returns and Treasury bill returns over these eleven-month periods.

Table 3 presents means and standard deviations for the eleven-month return spreads for the two equity portfolios. The t-statistics, under the null hypothesis that the mean differences are zero using a one-tail test, indicate that mean differences are positive and statistically significant. On average, equity investors earn an excess return over Treasury bills of approximately seven percent.[5] Like the pooled one-month returns, these eleven-month returns also suggest that equity investors are compensated for bearing risk outside the month of January. Previous researchers' conclusions that there is no reward to bearing risk in non-January months are sensitive to the specific hypothesis being tested. However, our results do not settle the issue about whether investors earn returns appropriate for the level of risk they bear, period. In the next section, we argue that researchers are asking for more than empirical tests (such as those used here and in previous studies) can deliver.

In Search of the Risk-Return Relationship

Results in Table 1 lead to opposite conclusions from results in Tables 2 and 3 for equity risk premia, even though the samples are drawn from the same raw return series. Surely, the tests used to detect whether investors are compensated for bearing risk deserve more careful scrutiny.

The question motivating these tests is simple enough. Is the difference between the return on risky assets and the return on a "riskless" asset positive on average in each calendar month of the year? Unfortunately, there are some pitfalls in this seemingly direct test.

The t-statistic in our hypothesis tests is given by

$$E(RP) / [\sigma(RP) / \sqrt{n}] \tag{1}$$

where $E(RP)$ is the sample mean excess return or risk premium of a particular portfolio, $\sigma(RP)$ is the estimated standard deviation of the risk premium, and n is the number of sample observations. The denominator in this expression is the standard error, which is simply the standard deviation divided by $\sqrt{n}$.

Rewrite the expression above as

$$[E(RP)] / [\sigma(RP)] \cdot \sqrt{n}) \tag{2}$$

The ratio $E(RP)/\sigma(RP)$ can be viewed as the reward to risk demanded by investors. The other component of the t-statistic, $\sqrt{n}$, is important from a purely statistical perspective. As sample size increases, one can obtain more precise estimates of the sample mean risk premium. Therefore, any ratio of $E(RP)$ to $\sigma(RP)$ is more likely to be statistically different from zero if the two parameters are estimated with a large number of observations. In principle, even a relatively small reward to risk would be judged significant if it persisted over a long period of time. In contrast, even a "substantial" reward to risk might not be significantly different from zero if there were very few observations.

The influence of sample size explains why the inferences drawn from examining the pooled set of February through December risk premia in Table 2 would differ from examining each of the eleven calendar months individually, as in Table 1. There are only sixty-three observations for each calendar month reported in Table 1, but 693 observations for each risk premium series reported in Table 2. In contrast, Chang and Pinegar's (1988b) sample has twenty observations per return series; their study would require a very high risk premium to standard deviation ratio to produce statistical significance.

Sample size, however, does not explain the contradictory results in Tables 1 and 3. The return series in Table 1 represent eleven series of sixty-three one-month risk premia while the return series in Table 3 represents sixty-three eleven-month risk premia. This apparent contradiction is resolved if we consider the relationship between return interval and the standard deviation of returns. For return series that are well-behaved statistically (zero serial correlation and constant variance), the standard deviation of the eleven-month returns will be approximately $\sqrt{11}$ or 3.16 times as large as the standard deviation of monthly returns.[6]

The standard deviations increase at the rate of $\sqrt{t}$ as the interval over which returns are computed increases, while returns increase at the rate of t. As a result, average eleven-month returns will be approximately eleven times as large as average one-month returns. Since the numerator grows at a rate of t, while the denominator grows at $\sqrt{t}$, the ratio $E(RP)/\sigma(RP)$ should increase at the rate of $\sqrt{t}$ as the return interval increases and t-statistics should reflect this behavior. For example, the t-statistic for eleven-month returns would be $\sqrt{11}$ times the t-statistic for one-month returns if the number of observations is held constant. This is precisely what we see in Table 3. Both the one-month average risk premia in Table 1 and the eleven-month average risk premia in Table 3 have sixty-three observations. However, the eleven-month average risk premia in Table 3 are positive and statistically significant while most of the one-month average risk premia are not.

Standard deviation has a dual role in these tests. First, standard deviation is simply a measure of dispersion. Obviously, the lower the standard deviation, the more likely a researcher will reject the hypothesis that a given average risk premium is zero. This raises the question, "how high should we expect the average risk premium to be relative to its standard deviation?"

There is no simple answer. The ratio $[E(RP) / \sigma(RP)]$ measures the excess return per unit of standard deviation. From portfolio theory, we know that efficient portfolios would maximize this ratio for each standard deviation. For an efficient portfolio, this reward to total risk measure is the slope of the Capital Market Line.[7] The slope of the Capital Market Line depends on investors' aggregate risk aversion, the reward demanded per unit of risk. Theory provides no direct guidance on the size of this slope—only that investors demand a positive risk premium.

Let us suppose we have identified a mean-variance efficient portfolio and wish to test whether the average risk premium is positive in a particular month, say August. In this context, we can state a corollary research question, "is the slope of the Capital Market Line (the market price of risk) sufficient to yield a t-statistic of 1.645 in each calendar month?"[8] The expectation that the slope will be sufficient to yield a t-statistic of 1.645 over all arbitrary calendar months and all arbitrary time periods seems to be an unreasonable restriction on the data.

Moreover, since the portfolios in our study and others are not necessarily efficient, the risk premium to standard deviation ratios will almost certainly be smaller than those of efficient portfolios. In well functioning capital markets, investors will price such portfolios to yield a risk premium sufficient to compensate for risk. The failure to reject the hypothesis that average risk premia equal zero says more about the power of the statistical tests to distinguish between zero and some positive number in small samples than it does about the true risk premium.

Another important component of empirical tests is the level of statistical significance demanded by researchers. A five or ten percent level of significance represents an aversion on the part of researchers to making Type I errors (rejecting the null hypothesis when it is true). Apparently, researchers would prefer to avoid inferring that there is a positive risk premium when, in fact, there is none. However, investors may be more concerned about making Type II errors (accepting the null hypothesis of a zero average risk premia when, in fact, there is a positive average risk premium). One could plausibly conclude that investors might take a chance of falsely believing there is a common stock risk premium if they can expect to earn an additional return of seven percent in the February through December period. This is certainly more plausible than investors choosing to forego additional return because it is not statistically different from zero at the five percent level.[9]

CONCLUSION

Our research indicates that reports of the death of the risk/return relationship are premature. Inferences drawn from the evidence can be altered greatly through small changes in the way the empirical question is addressed. The precision of the mean risk premia can be increased by expanding the number of observations. Since the test statistic is sensitive to the number of observations, a high excess return to risk ratio is required to find statistical significance when there are few observations.

Portfolio theory places no constraints on the size of the risk premium to standard deviation measure. One reasonable constraint is that it is positive, which is true even outside of January for the two equity portfolios examined in this study. Investors should also note that the evidence suggests that the reward for bearing risk does vary over the calendar year. However, for equity investors that have longer holding periods, this variability in risk premia is of little importance and a rational investor would participate in the market throughout the year, not only during the month of January. In summary, we believe there is sufficient evidence to be skeptical about the conclusion that the risk/return relationship is severed outside of January.

Acknowledgments: The helpful comments of Robin Grieves, Scott Harrington and Ted Moore are gratefully acknowledged. The usual disclaimer applies.

NOTES

1. See, for example, Tinic and West (1984), Gultekin and Gultekin (1987), and Chang and Pinegar (1988a, 1988b). Tinic and West find that January is the only month in which there is a reliable relationship between beta and expected returns in the context of the Capital Asset Pricing Model. Gultekin and Gultekin find an analogous result for risk measures associated with the Arbitrage Pricing Theory. Of course, the empirical difficulties associated with testing these models makes these conclusions more tenuous. See Roll (1977) and Shanken (1982) for discussions of these empirical difficulties.
2. Starting in 1982, this portfolio of small capitalization stocks contains some stocks that are listed on the American Stock Exchange and on NASDAQ.
3. *Stocks, Bonds, Bills and Inflation* is published annually by Ibbotson and Associates. For further details on the construction of these portfolios, the interested reader is directed to the 1989 *Yearbook*.
4. Each monthly return (in decimal form) is augmented by one. The eleven-month return is then simply the product of these eleven individual monthly returns.
5. Since the term structure of interest rates is generally upward sloping, the choice of one-month bills may bias our results in favor of finding a positive and statistically significant equity risk premium. An investor would be able to capture an additional term premium in a typical year by investing in longer term government securities. The choice of one-month bills is made purely as a matter of convenience due to data availability.
6. The properties of the relationship between standard deviation of returns and time are discussed by Young (1971) and McEnally (1985). First order serial correlations for the Common and Small equity portfolios are .10 and .16, respectively. Positive serial correlation in returns will lead t-period standard deviations to be slightly more than $\sqrt{t}$ times as large as single period standard deviations.
7. Of course, this is not literally true since the Capital Market Line is developed in the context of a single period model. The slope of the theoretical capital market line is given by $(E(R_M) - R_F) / \sigma_{RM}$. It measures the ex ante reward to risk of the market portfolio. Our measure is an ex

post measurement approximating this slope. If the risk-free rate were non-stochastic over all return intervals and R_M represented the true market portfolio, our measure would provide a sample estimate of the slope of the Capital Market Line. For our purposes, these differences are relatively minor.

8. Recall that 1.645 is the critical value for a one-tail test at the five percent level of significance. Thus, the t-statistic must be at least 1.645 or we would fail to reject the null hypothesis that the risk premium is zero.
9. For a general discussion of how researchers misuse tests of significance, see McCloskey (1985).

References

Chang, Eric, and J. Michael Pinegar. 1988. "A Fundamental Study of the Seasonal Risk-Return Relationship: A Note," *Journal of Finance*, 43: 1035-1039.

Chang, Eric, and J. Michael Pinegar. 1988. "Does the Market Reward Risk in Non-January Months?" *Journal of Portfolio Management*, 15: 55-57.

Gultekin, Bulent, and Mustafa N. Gultekin. 1987. "Stock Return Anomalies and Tests of the APT," *Journal of Finance*, 42: 1213-1224.

McCloskey, Donald. 1985. "The Loss Function Has Been Mislaid: The Rhetoric of Significance Tests," *American Economic Review*, 75: 201-205.

McEnally, Richard W. 1985. "Time Diversification: The Surest Route to Lower Risk?" *Journal of Portfolio Management*, 11: 24-26.

Roll, Richard W. 1977. "A Critique of the Asset Pricing Theory's Tests: Part I: On the Past and Potential Testability of the Theory," *Journal of Financial Economics*, 4: 129-176.

Shanken, Jay. 1982. "The Arbitrage Pricing Theory: Is It Testable?" *Journal of Finance*, 37: 1129-1140.

Tinic, Seha M., and Richard West. 1984. "Risk and Return: January vs. the Rest of the Year," *Journal of Financial Economics*, 13: 561-574.

Young, William. 1971. "Random Walk of Stock Prices: A Test of the Variance-Time Function," *Econometrica*, 39: 797-812.