On the Robustness of Size and Book-to-Market in Cross-Sectional Regressions
Author(s): Peter J. Knez and Mark J. Ready
Source: *The Journal of Finance*, Vol. 52, No. 4 (Sep., 1997), pp. 1355-1382
Published by: Blackwell Publishing for the American Finance Association
Stable URL: http://www.jstor.org/stable/2329439
Accessed: 25/11/2009 10:11

THE JOURNAL OF FINANCE • VOL. LII, NO. 4 • SEPTEMBER 1997

On The Robustness of Size and Book-to-Market in Cross-Sectional Regressions

PETER J. KNEZ and MARK J. READY*

ABSTRACT

We use a robust regression estimator to analyze the risk premia on size and book-to-market. We find that the risk premium on size that was estimated by Fama and French (1992) completely disappears when the 1 percent most extreme observations are trimmed each month. We also show that the negative average of the monthly size coefficients reported by Fama and French can be entirely explained by the 16 months with the most extreme coefficients. We argue that further investigation of these results could lead to an understanding of the economic forces underlying the size effect, and may also yield important insights into how firms grow.

IN RELATED ARTICLES, FAMA and French (1992 and 1993) argue that size and book-to-market play a dominant role in explaining cross-sectional differences in expected returns.[1] In the first article, Fama and French show that there appear to be risk premia associated with these factors, and in the second they show that long-short portfolios constructed to maximize the "loading" on these factors are useful in explaining the common variation in returns. In this article, we investigate the robustness of the estimated risk premia for size and book-to-market. If the estimates are driven by a small subset of firms or months, then our robust techniques will allow us to isolate these influential observations. We believe that identifying and investigating these observations could deepen our understanding of the economic reasons for the risk premia associated with size and book-to-market, which is particularly important since neither of these factors is identified with any well articulated equilibrium asset pricing model.[2]

* Knez is from Northwestern University and Lincoln Capital Management. Ready is from University of Wisconsin-Madison. Conversations with Eugene Fama, Jed Frees, Doug Hanna, Lars Hansen, John Heaton, Robert Korajczyk, Doug Martin, Jonathan Paul, Mitch Petersen, Ken West, and Ray Zemon are gratefully acknowledged. We thank the anonymous referee and workshop participants at Northwestern University, the University of Wisconsin-Madison, and the Research Triangle for helpful comments and suggestions.

[1] There is a large body of work examining the relation between size and/or book-to-market and returns. A partial list includes Banz (1981), Blume and Stambaugh (1983), Breen and Korajczyk (1995), Keim (1983), Kothari, Shanken, and Sloan (1995), Frankel and Lee (1995), Lakonishok and Shapiro (1986), Roll (1983), and Stoll and Whaley (1983).

[2] Several authors, beginning with Ball (1978) and most recently including Berk (1995), have pointed out that since return is proportional to the reciprocal of initial price, size should proxy for missing factors in any test of a misspecified asset pricing model.

In Fama and French (1992) the factor risk premia are estimated by taking time series averages of the slope coefficients from monthly cross-sectional regressions as in Fama and MacBeth (1973). As is well known, the least squares (hereafter LS) objective used by Fama and MacBeth (hereafter FM) is sensitive to outliers in both the y-direction (outliers in the errors) and the x-direction (leverage points). Although Fama and French check their results using average returns on portfolios formed from bivariate sorts on size and book-to-market, these tests still focus only on the averages across firms in each portfolio and across months, potentially masking the importance of influential observations that may be driving the results. In this article, we incorporate a robust regression technique into the FM procedure that is not unduly sensitive to outliers or leverage points, and compare the results to those obtained using LS.

This robust regression technique, called least trimmed squares (LTS) trims a proportion of the influential observations and then fits the remaining observations using LS. We want to emphasize, however, that the outliers identified by the LTS procedure are not viewed here as "contaminates" that are to be discarded or deleted so that a proper inference may be conducted. Nor are we suggesting that these observations are not relevant for computing the conditional mean return associated with some portfolio formation strategy. On the contrary, we view them as "precious" in the sense that they may convey a lot of information about the return generating process and a proper model specification.[3] Accordingly, we use the LTS regressions to provide a diagnostic technique for evaluating the sensitivity of the inference conducted using LS and for revealing the possible economic role played by these regressors.

We find that the estimated relation between size and returns is significantly affected by using a robust FM procedure. For example, whereas the mean slope coefficient (monthly risk premium) on size is -12 basis points using LS regressions, it is $+33$ basis points using LTS regressions that trim 5 percent of the observations each month. Even when only 1 percent of the observations are trimmed each month, the mean slope coefficient is $+14$ basis points, which implies that the so called "size effect" (see Banz (1981)) is entirely driven by extreme observations that represent less than 1 percent of each month's data. Intuitively, the LTS regressions are designed to fit *most* of the data very well. Accordingly, the LTS regression results imply that most small firms actually do worse than larger firms.

We also investigate bivariate regressions of returns on both size and book-to-market and univariate regressions of returns on book-to-market. In bivariate regressions, the mean slope on size is -8 basis points using LS and $+37$ basis points using LTS regressions that trim 5 percent of the observations. This difference between the two techniques is identical to that from the univariate regressions discussed above. For book-to-market, on the other hand, the impact of the trimming inherent in the LTS techniques depends on whether the regressions also include size. The LTS technique produces a

[3] See Huber (1991) for a discussion of this perspective.

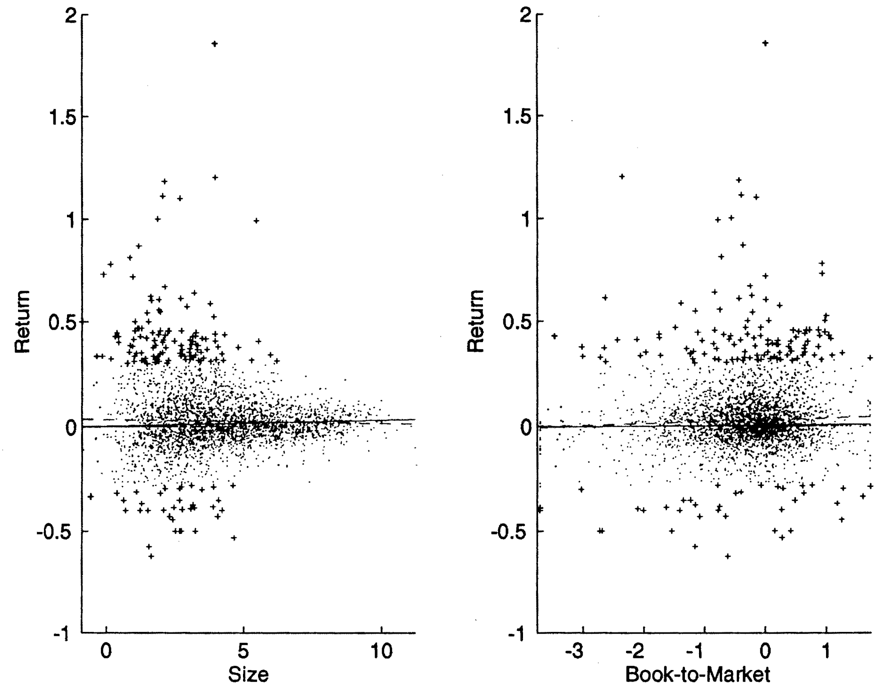

Figure 1. Scatterplots and fitted regression lines for least squares and least trimmed squares for March, 1989. Size is the natural logarithm of market value of equity in millions and Book-to-Market is the natural logarithm of the ratio of the book value of equity to market value of equity. The dashed lines are fit using least squares (LS) and the solid lines are fit using least trimmed squares (LTS) with 5 percent of the data trimmed. The observations trimmed by LTS are shown with a "+" symbol, and the remainder of the observations are shown as dots.

somewhat lower mean slope coefficient than LS when book-to-market is used as the only explanatory variable, but this difference between the robust and LS techniques reverses in the bivariate regressions. Thus, rather than indicating that the book-to-market results are driven by extreme observations, the fact that the univariate book-to-market robust and LS regressions differ seems to be explained by the fact that size is an omitted variable. On balance, the results indicate that the size effect is driven by a few extreme observations, but that the risk premium on book-to-market is not affected by extreme observations once you control for size.

The impact of the extreme observations on the slope coefficients can be seen by examining scatter plots for individual months. Figure 1 shows the scatter plots of returns against size and returns against book-to-market for March, 1989, as well as the estimated univariate regression lines using both the LS and LTS techniques. This month was chosen to illustrate the results because all of the regression slopes are reasonably close to the overall sample averages.

From the left hand scatter plot of return against size in Figure 1, it is readily apparent that most of the trimmed observations (shown with "+" signs) are extreme positive returns on small firms. If the returns were normally distributed, the 5 percent most extreme should be evenly distributed above and below the regression line. In contrast, about 77 percent of the trimmed observations in this month are above the line, so the distribution of the small firm returns is positively skewed. Also, as can be seen from the figure, the 5 percent of the observations that are trimmed have returns that are from about 27 percentage points to more than 150 percentage points from the regression lines. About 50 percent of the data lie within 6 percentage points of the regressions lines, so the trimmed observations are extreme both in economic terms and relative to the rest of the data (there is excess kurtosis). If the data were normally distributed and had this same interquartile range, we would expect only two or three observations to have residuals as large as 27 percentage points, and the chance of observing any much larger than that would be virtually nil.

These extreme positive observations pull the left hand (small firm) side of the LS line upward, but do not affect the LTS line. For this month, there is a -0.17 correlation between size and book-to-market (the average correlation for all months is -0.27), so these same extreme returns tend to cluster toward the right-hand side of the book-to-market plot, pulling the univariate LS line up for high book-to-market firms. Because it ignores the trimmed observations, the univariate LTS slope on book-to-market is lower than for LS, but this is partly due to the fact that the many small firms that are *not* trimmed also cluster on the right hand side of the plot and these firms tend to have *lower* returns. The bivariate LTS regression controls for the effect of firm size and the book-to-market slope is slightly higher than from the LS regression.

We investigate whether our results are related to the well known January seasonal in small firm returns (see Keim (1983)). We find that the difference between the LS and LTS slopes on size is more pronounced in January than in other months, but we also find that there is still a strong January effect in the LTS slopes. Thus, an analysis of extreme returns may ultimately help in understanding the January effect, but it is unlikely that these firms alone can provide the entire explanation. We also confirm that our results are not driven by low price firms, because we obtain similar results when the regressions are run on a sample restricted to firms with share prices above 5 dollars.

As can be seen from Figure 1, the observations trimmed from the LTS regressions are quite extreme. On average, we find that the observations that are trimmed above the regression line are returns of about 37 percent (per month). Moreover, although the LTS methodology trims based only on the absolute difference from the regression line, about 74 percent of the trimmed observations lie above the line each month. These extreme observations are primarily small firms, and we speculate that, in part, these extreme returns represent the infrequent transition from a small regional firm to a larger national or multinational firm. Small firms that are poised for this transition are no doubt quite different from other small firms with stable businesses and prospects. For one thing, we would expect these potential transition firms would be younger. In fact, we show that the firms whose extreme returns are trimmed in the LTS regressions are younger than those in the rest of the sample, even after controlling for size.

The article is organized as follows. In Section I we briefly discuss the FM procedure as used by Fama and French (1992) and present our extension to include robust regression techniques. This essentially amounts to a *robust* Fama-MacBeth procedure. In Section II the regression results are presented and the existence of influential observations and influential months is explored. Section III contains a discussion of possible economic interpretations of the influential observations, and Section IV concludes.

I. Econometric Method

In this section we briefly discuss the FM procedure as applied in Fama and French (1992), and briefly describe the data and variable definitions. We then describe the LTS regression procedure, its traditional motivation as a way to deal with contaminated data, and our use of the procedure as a diagnostic tool to isolate the effect of influential observations within each month. Finally, we discuss the aggregation of the monthly slopes from both the LS and LTS regressions, and describe our technique for identifying months that have a large effect on the averages.

A. The Fama-MacBeth (FM) Procedure

This section provides only a cursory review of the FM procedure. For a more complete discussion, see Fama (1976) and Shanken (1992). The FM procedure involves a time series of cross-sectional LS regressions of realized returns on one or more explanatory variables that are assumed to be possible candidates for risk factors. For each period $t = 1, \ldots, T$ the following regression is run using LS,

$$\tilde{R}_{j,t} = \lambda_0 + \sum_{i=1}^{F} \lambda_{i,t} x_{i,j,t} + \epsilon_{j,t} \qquad \text{for} \quad j = 1, 2, \ldots, N_t \tag{1}$$

where N_t is the number of firms in period t, $\tilde{R}_{j,t}$ is the realized return on stock j in period t, F is the number of potential risk factors, and $x_{i,j,t}$ is the realization of risk factor i for firm j in period t. The null hypothesis is that the time-series average of the monthly regression slopes is zero, and statistical significance is established using a standard t-test.

In Fama and MacBeth (1973), the right-hand side variables in equation (1) are capital asset pricing model (CAPM) beta, beta squared, and residual risk. They find that the mean slope coefficient on beta is significantly positive, and that the mean slope coefficients on beta squared and residual risk are insignificant. These results are consistent with the CAPM predictions that expected return is a linear function of beta risk, and that no other risk factors are important. In contrast, Fama and French (1992) use size, book-to-market, earnings-to-price, and leverage in addition to beta as variables in equation (1). They conclude that size and book-to-market subsume the effects of the other variables, including beta, in explaining the cross-sectional variation in expected returns.

In this article, we use the same data as in Fama and French (1992), and use both LS and LTS procedures to estimate equation (1). Based on their success in the Fama and French regressions, we examine only size and book-to-market as predictors. Briefly, the data include all New York Stock Exchange (NYSE), American Stock Exchange (AMEX), and Nasdaq nonfinancial firms in the monthly Center for Research in Security Prices (CRSP) tapes, for the period from July 1963 through December 1990. Book-to-market is the natural logarithm of the ratio of COMPUSTAT book value to CRSP market values as of year end, whereas size is the natural logarithm of CRSP market value as of June 30. Following Fama and French, the smallest and largest 0.5 percent of the values for book-to-market are set equal to the 0.005 and 0.995 fractiles. These values for the regressors are matched with the firm's returns from July through June of the following year. Fama and French require firms to have 24 nonmissing monthly returns in the five years prior to inclusion in the sample in order to allow calculation of "pre-ranking" betas. Although we do not calculate any betas in this article, we retain the same data selection procedures to allow direct comparison of the results.[4]

B. A Robust Fama-MacBeth Procedure

While LS has many attractive features, it has the undesirable feature of being sensitive to outliers and leverage points. This is particularly problematic if there is some chance that the data contain errors. One way to formulate the sensitivity of an estimator to errors in the data is with the notion of the *breakdown point*.[5] Let S denote a sample of N observations and let the regression estimator using sample S be denoted by $\theta(S)$. The maximum bias that can be induced by considering all the possible corrupted samples S' obtained by replacing m of the N sample points by arbitrary values is given by

$$\mathrm{b}(m;\ \theta,\ S) = \sup_{S'} \|\theta(S') - \theta(S)\|.$$

Note that if the bias b is infinite then a set of m erroneous observations can have an arbitrarily large effect on the estimator θ and the estimator is said to break down. The breakdown point of an estimator is given by

$$\mathrm{bp}(\theta,\ S) = \min\left\{\frac{m}{N} \text{ such that } b(m;\ \theta,\ S) = \infty\right\}.$$

[4] Although our procedure is designed to match Fama and French (1992) exactly, there is a slight difference in our estimated slopes for size. One possible explanation for the discrepancy is that we use more recent versions of the CRSP tapes, and CRSP continually updates missing information on the returns for delisted firms.

[5] See Hampel, Ronchetti, Rousseeuw, and Stahel (1986), Rousseeuw and LeRoy (1987), and Rousseeuw (1984) for a complete discussion.

That is, the finite sample breakdown point of an estimator at a given sample S is the smallest fraction of corruption that can cause the estimator to take on values that are arbitrarily far from the values obtained using the uncorrupted sample (i.e., $\theta(S)$).[6]

For the LS estimator the breakdown point equals

$$\mathrm{bp}(\theta,\ S) = \frac{1}{N}$$

which means that a single observation can cause $\theta(S')$ to take on any value no matter how large a sample is used. It is common to express the breakdown point as a percent of the sample as N becomes large. Thus, it can also be said that LS has a breakdown point of 0 percent.

The LTS estimate has a high breakdown point and is robust against outliers and leverage points. It is defined by

$$\hat{\lambda} = \text{argmin} \sum_{i=1}^{q} (\epsilon^2)_{[i]}$$

where $\epsilon^2_{[1]} \leq \epsilon^2_{[2]} \leq \cdots \epsilon^2_{[N]}$ are the squared residuals listed in order of increasing magnitude, $\hat{\lambda}$ is a parameter vector of length p, and $q \leq N$. In words, the LTS estimator fits q of the observations very well and ignores (trims) the rest. Note that since the LTS trimming depends only on the magnitude of the residuals, it generally will not trim the same number of observations from the upper and lower tails of the distribution. This is in contrast to techniques such as the trimmed mean and the "Winsorized" mean that automatically trim an equal number of observations from each tail.

When the scalar q over which the sum is taken is set equal to $N/2 + (p + 1)/2$ the LTS estimator reaches the highest possible breakdown point for any regression equivariant estimator. When q is defined this way, the breakdown point of the LTS estimator is given by $([(N - p)/2] + 1)/N$, which is approximately 50 percent in the usual case where p is small relative to N. The number q depends explicitly on the sample size, so it is usually more convenient to work with the *trimmed proportion* α, where

$$q = [N(1 - \alpha)] + [\alpha(p + 1)].$$

Note that for the LTS estimator, the trimmed proportion is approximately equal to the breakdown point. If α is set equal to 0 percent then the LS estimator is being used; if α is set equal to 50 percent then the LTS estimator with the highest possible breakdown point is being used.

[6] The breakdown point is a *global* measure of robustness. An alternative measure that focuses on the bias for small neighborhoods is the influence function. Both of these measures of bias have limitations. See Maroni and Yohai (1991) and Martin, Yohai, and Zamar (1989) for a discussion of bias and robustness in a regression context.

Although the LTS estimator was originally developed as a way to deal with "corrupted data," we are using it for a quite different purpose. We believe that the extensive data screening done by CRSP and COMPUSTAT probably eliminates most of the errors, so the LS estimator probably would provide a better single estimate of the mean return associated with a particular factor. This is because the LS estimates are unbiased, regardless of the distribution of the residuals, whereas the LTS estimates generally will be biased estimates of the mean if the residuals are not symmetric. However, by comparing the LTS results with the LS results and by allowing the trimming proportion to vary from 50 percent down to 1 percent, we can get a feel for the fraction of observations that are driving the LS relations. Moreover, closer examination of the observations that are trimmed from the LTS regressions could lead to a better understanding of the economic factors that drive the LS results.

The robust Fama-MacBeth procedure is conducted by running a cross-sectional regression each month using the LTS estimator. This gives a time series of robust slope coefficients (i.e., $\hat{\lambda}_t$ $t = 1, \ldots, T$). We then compute the time-series average of the coefficients and compare the results to those obtained using LS. It should be noted that comparison of LS and robust results could also be useful in the study of other issues in finance that are beyond the scope of this article, such as the long-run performance of IPOs.

C. *Influential Months*

As discussed above, the estimated risk premium is the time series average of the slope coefficients from the monthly regressions. The FM procedure ignores the information from each individual regression that could be used to estimate the standard errors of the coefficients, and uses only their variability across months in calculating the associated t-statistics. This procedure imposes no stationarity assumption on the intercepts or residual volatility across months, and does not require an estimate of the cross-correlation of the residuals within each month.

The use of a standard t-statistic does implicitly assume that the monthly coefficients are independent draws from a stationary normal distribution. The violation of normality can mean that the significance levels calculated from the t-distribution are incorrect, and if the distribution of the slopes has excess kurtosis the significance levels will generally be overstated. In spite of this fact, with a sample size of 330 months it is unlikely that deviations from normality will have an important effect on the calculated significance levels. This is because t-statistics tend to be robust to deviations from normality as a consequence of the central limit theorem. For completeness, in the next section we use bootstrap tests to evaluate whether deviations from normality have any important effect on the statistical significance of the mean slope coefficients. Another concern over the use of the t-statistic in this application is that it is not robust to persistent serial correlation in the monthly slopes.[7] Accordingly, in

[7] See Lehmann (1986) and Cressie (1980) for a complete discussion.

Section II we use the procedure of Newey and West (1987) to calculate adjusted t-statistics and compare them to the standard t-statistics.

Although it is important to address the distributional issues associated with the t-statistics, we believe that the most interesting question to ask about the averages of the slope coefficients is whether they are driven by a small subset of months. This question can be answered by comparing the averages of the monthly LS coefficients to trimmed averages that ignore months whose coefficients take on extreme values. Also, if it is shown that the monthly LS coefficients are themselves driven by the returns on a small subset of the firms in each month, then we can ask whether this effect is uniform across months. This question can be answered by comparing the average *difference* between the LS and LTS slopes to the trimmed average difference that ignores months where the difference is large.

When looking at the time series means, the extreme months are identified using the same least trimmed squares approach that we use in the monthly regressions. For a given trimmed proportion α, we simply sort the sample of monthly slope coefficients and then examine all possible averages that trim x percent of the slopes from the bottom of the distribution and α–x from the top. The least trimmed squares estimate of the average slope is the value that minimizes the sum of the squared residuals, where the sum is taken over all of the months that are not trimmed.

Thus, in the next section we have two related definitions of what constitutes an "influential month." Investigating the months with extreme LS slope coefficients may help in understanding why small firms do well on average. Investigating the months with extreme differences between LS and LTS (along with investigating the specific firms trimmed in the LTS regressions) may help to understand why a few small firms may have extreme positive returns.

II. Regression Results

In this section we present the results of the cross-sectional regressions using a robust FM procedure and compare them to the results from LS. We also present a series of diagnostics designed to provide a statistical characterization of the influential observations, and to set the stage for the discussion of possible economic explanations contained in Section III. In the final subsection, we test for the presence of influential months as described above. As one might expect, this leads to a discussion of the January effect (Keim (1983)) and its relation to our results.

A. Cross-Sectional Regressions

Table I contains the cross-sectional regression results. The average size slopes from the univariate regressions are shown in the first line of Table I, and the average size slopes from the bivariate regressions that also include book-to-market are shown as the first number in each pair of numbers in the third line of Table I. As shown in the first column of the first line and as

Table I

Average Slopes from Monthly Cross-Sectional Regressions

The sample includes all New York Stock Exchange (NYSE), American Stock Exchange (AMEX), and Nasdaq firms in the intersection of Center for Research in Security Prices (CRSP) and COMPUSTAT from July 1963 through December 1990. Monthly cross-section regressions are run for the following three specifications:

$$\tilde{R}_{j,t} = \lambda_{0,t} + \lambda_{1,t} \ln(ME)_{j,t} + \epsilon_{j,t} \qquad \text{for} \quad j = 1, 2, \ldots, N_t$$

$$\tilde{R}_{j,t} = \lambda_{0,t} + \lambda_{2,t} \ln(BE/ME)_{j,t} + \epsilon_{j,t} \qquad \text{for} \quad j = 1, 2, \ldots, N_t$$

and

$$\tilde{R}_{j,t} = \lambda_{0,t} + \lambda_{1,t} \ln(ME)_{j,t} + \lambda_{2,t} \ln(BE/ME)_{j,t} + \epsilon_{j,t} \qquad \text{for} \quad j = 1, 2, \ldots, N_t$$

where N_t is the number of firms in month t, $\tilde{R}_{j,t}$ is the realized return on stock j in month t, $\ln(ME)_{j,t}$ is the natural log of market equity for firm j for month t, and $\ln(BE/ME)_{j,t}$ is the natural log of the book-to-market ratio for firm j for month t. The table reports time series averages of the monthly slope coefficients ($\bar{\lambda}$) over the 330 months in the sample. The t-statistics (shown in parentheses) are the time-series average of the slope coefficients divided by the time-series standard error, and then multiplied by the square root of 330.

Separate sets of regressions are run using least squares and then least trimmed squares with different values for the trimmed proportion α. As an example, $\alpha = 5\%$ means that $100 - \alpha = 95\%$ of the observations are used to estimate the coefficients.

Independent Variable	Least Squares	Least Trimmed Squares			
		$\alpha = 50\%$	$\alpha = 25\%$	$\alpha = 5\%$	$\alpha = 1\%$
ln(ME)	−0.12	0.30	0.37	0.33	0.14
	(−2.02)	(7.32)	(8.87)	(6.50)	(2.63)
ln(BE/ME)	0.47	0.07	0.11	0.35	0.48
	(5.41)	(0.81)	(1.24)	(4.24)	(5.73)
ln(ME), ln(BE/ME)	−0.08, 0.35	0.34, 0.37	0.32, 0.32	0.37, 0.57	0.17, 0.50
	(−1.41), (4.35)	(7.98), (4.35)	(7.46), (3.90)	(7.71), (7.70)	(2.94), (6.39)

documented by Fama and French (1992), the LS objective yields a negative average slope on size of −0.12, with a t-statistic of −2.02. In contrast, the LTS objective with 50 percent trimming yields a *positive* average slope of 0.30 with a t-statistic of 7.32. Moreover, the positive average slope on size obtained from LTS remains remarkably consistent as the trimmed proportion is reduced from 50 percent to 25 percent to 5 percent. Even when only 1 percent of the observations are trimmed each month, the average coefficient remains significantly positive. As shown in the third line of the table, the average slopes on size in the bivariate regressions are nearly the same as they are in the corresponding univariate size regressions.

With 5 percent trimming, the LTS univariate regression slope on size is positive for about 75 percent of the months in the sample. Essentially, the LTS objective fits a line to a subset of the data where the observations are more highly concentrated. Intuitively then, the LTS results indicate that in most months the "typical" large firm has a *higher* return than the "typical" small firm. Note that we are not saying that the LS results are "wrong." Our point is that the LTS results and their comparison to the LS results provide a richer characterization of the relations, namely that the higher average return on small firms is driven by a small subset of firms each month that experience extreme returns.

The fact that the LTS slopes are similar using either 50 percent or 95 percent of the data implies that the tendency for the extreme observations to have positive returns is for the most part explained by just 5 percent of the observations. This means that we may be able to understand the economic factors underlying the size effect by examining the unique features of the firms whose returns are trimmed in the 5 percent LTS regressions. Because they seem to provide a clean separation between those firms that drive the size effect and those that do not, in the remainder of the article we focus mostly on the 5 percent LTS regressions and their comparison to the LS regressions.

Turning to the average slopes on book-to-market, the second line of Table I shows that LS and LTS yield different results in the univariate regressions. On the other hand, the third line of the table shows that LS and LTS yield similar slopes on book-to-market in the bivariate regressions. Thus, rather than implying that the estimated risk premium on book-to-market is driven by a subset of the observations, the anomalous results from the univariate regressions are an indication that size is an important omitted factor.

It is not surprising that there should be some interaction between size and book-to-market, because their average correlation is -0.27. Recall that the LTS size results indicate *most* small firms have *lower* returns than larger firms. Accordingly, after trimming the extreme small firm returns, the remaining small firms, which also tend to have higher book-to-market, will pull the LTS line down. Of course, this argument relies on essentially the same observations being trimmed in all three regressions, which will not always be the case with LTS, because the trimming is driven by the distribution of x variables as well as the distribution of the y variables. In the next subsection however, we provide evidence that in this application the trimmed observations are primarily extreme returns (outliers) as opposed to extreme values of the explanatory variables (leverage points), so most of the trimmed observations are common across the three regressions.

B. Influential Observations

In this subsection, we examine in detail the influential observations that drive the differences between the LS and LTS results reported in Table I. Recall that even with non-Gaussian residuals, the LS and LTS estimates should yield similar results provided the errors are symmetric. Thus, the disagreement between the estimates highlighted in Table I suggests skewness in the cross-sectional regressions that is a function of the regressors. This skewness is particularly evident from the scatter plot of return on size in Figure 1 of the introduction; most of the trimmed observations plot above the line and are on the small firm side of the plot. The impact of this skewness is

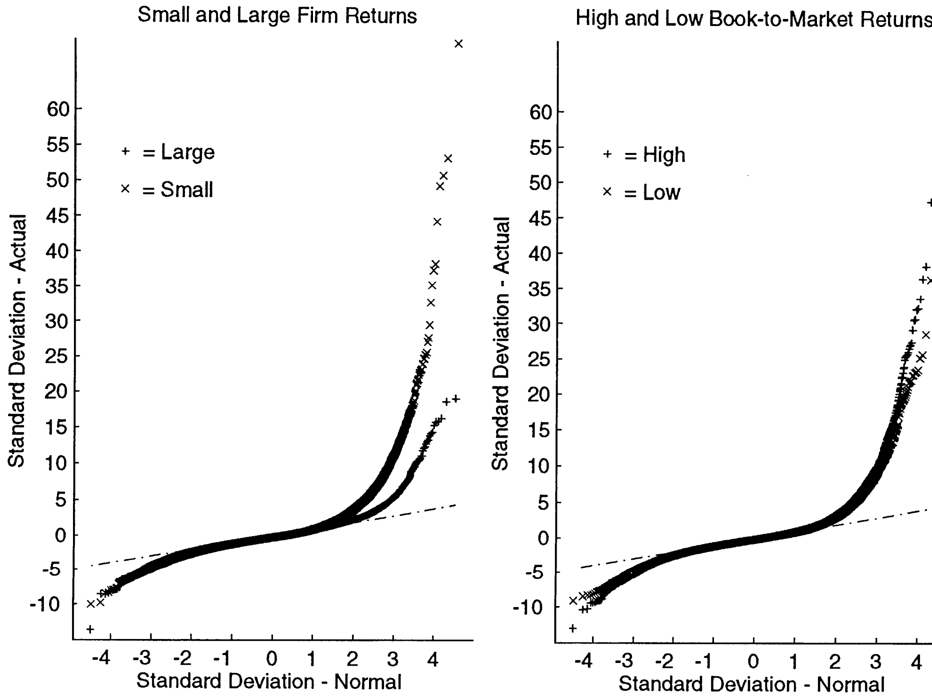

Figure 2. Q-Q plots for small and large firm returns and high and low book-to-market firm returns. Small and large (high and low book-to-market) firms are defined as the top and bottom quintiles of firm size (book-to-market) for each month. For each group, returns are standardized by subtracting the group mean and dividing by 0.7413 times the inter-quartile range (an estimate of the standard deviation). These standardized returns are pooled across months, ordered, and plotted on the vertical axis against the quartiles of the standard normal cumulative distribution function on the horizontal axis. The symbol "x" is used to represent small and low book-to-market firms, and a "+" is used to represent large and high book-to-market firms.

to pull the LS fit upward in the region of small stocks, thereby inducing a negative slope in the estimated relation between size and returns.[8]

Figure 2 provides a graphical illustration of the skewness in returns as a function of size and book-to-market for the entire sample. The *q-q* (quantile-quantile) plots in Figure 2 are constructed by separating the observations in each month into quintiles based on the values of the regressors. For each group, returns are standardized by subtracting the group mean and dividing by 0.7413 times the interquartile range. This estimate of the standard deviation of the data generating process is consistent under the null hypothesis that the distribution is normal, and provides a more powerful tool than the sample standard deviation for spotting deviations from normality. This is because the sample standard deviation is heavily affected by a few extreme observations.

[8] Richardson and Smith (1993) develop a GMM procedure to test multivariate normality. They find strong evidence that stock returns and market model residuals are non-normal. Ball, Kothari, and Shanken (1995) document positive skewness in the return distributions of "loser" stocks.

For each group, these standardized returns are pooled across months, ordered, and plotted on the vertical axis against the quantiles of the standard normal cumulative distribution function. The xs in Figure 2 represent the distribution of the returns taken from the smallest quintile of the values of the regressors (small and low book-to-market) and the +s represent the distribution from the largest quintile (large and high book-to-market).

The *q*-*q* plot is used for spotting deviations from normality by comparing the sample points to the line that passes through zero with a slope of one (shown in the figure as alternating dashes and dots). If data plot both above the line to the right and below the line to the left this is evidence that the observations in the tails of the distribution are relatively more extreme than would be expected if the data were normal (excess kurtosis). If the data on one side of the plot are farther from the line than they are on the other side of the plot, this is evidence of skewness. From comparing the *q*-*q* plots for the small and large size quintiles in the left panel of Figure 2, it is readily apparent that the distribution of small firm returns has excess kurtosis and is strongly positively skewed. Look for example, at the points plotted above the values 3 and -3 on the horizontal axis. These are approximately the 0.1 percent and 99.9 percent quantiles of the distribution, and they should be 3 standard deviations away from the mean if the data were normally distributed. The figure shows that, for both small and large firm returns, the observations at the 0.1 percent quantile are about 5 standard deviations below the mean. The large firm returns at the 99.9 percent quantile are about 6 standard deviations above the mean and the small firm returns at the 99.9 percent quantile are about 13 standard deviations above the mean.

Another way to quantify skewness in the data is with the sample skewness coefficient, which is calculated as

$$\frac{n^{1/2}\sum_{i=1}^{n}(R_i-\bar{R})^3}{\left(\sum_{i=1}^{n}(R_i-\bar{R})^2\right)^{3/2}}.$$

This statistic has an expected value of zero under the hypothesis that the data are normally distributed. We calculate this statistic each month for the returns in both the top and bottom size quintiles, and then average the statistics across months. This allows us to use the time series distribution of the skewness coefficient to calculate *t*-statistics in exactly the same way as is done for the slope coefficients from the monthly regressions. The average monthly skewness coefficient for the returns of the small firm quintile is 1.94, which compares to an average skewness coefficient of 0.49 for the returns from the large firm quintile. The *t*-statistic for the difference between these two is 15.2.

These extreme observations are the source of the difference between the LS and LTS estimators. They exert a large influence on the LS estimator because of its zero breakdown point property, but are trimmed from the LTS estimator. The trimming has a larger effect on small firms because the positive observations trimmed tend to be much more extreme than the positive observations trimmed for large firms. Note that the results from Table I show that trimming just 5 percent of the observations changes the average slope on size from negative to strongly positive.

The q-q plot in the right panel of Figure 2 shows that the high book-to-market firm returns exhibit more positive skewness than do low book-to-market firm returns, but the difference between the two groups is less dramatic than the difference between small and large firm returns. The average monthly skewness coefficient for the returns of the firms in the high book-to-market quintile is 1.80, as compared to 1.34 for the returns of the firms in the low book-to-market quintile, and the t-statistic for the difference is 4.3.

On average, high book-to-market firms are somewhat smaller than low book-to-market firms, so the greater skewness in high book-to-market firm returns may simply reflect the excess skewness of small firm returns. To test this possibility, we construct subsamples of firms in each month drawn from the top and bottom book-to-market quintiles, in a way that ensures that the two subsamples have similar firm sizes. To create the subsamples, we start with the smallest firm in the low book-to-market quintile and attempt to match it with a firm from the high book-to-market quintile that is larger than the low book-to-market firm under consideration, but by no more than 1 percent. If no such similarly sized high book-to-market firm can be found, the low book-to-market firm is discarded and we move on to the next larger firm in the low book-to-market sample. When there is a match, both firms are put into the size-matched subsamples. Using these subsamples, the average monthly skewness coefficient for the high book-to-market firms is 1.21, as compared to 1.15 for the low book-to-market firms. The t-statistic for the difference between these two values is only 0.6, which means we are unable to reject the hypothesis that the two samples have the same degree of skewness once we control for size. These results confirm the earlier conjecture that the anomalous results obtained in the univariate book-to-market returns are primarily due to the fact that size is an omitted factor.

As we indicate in the earlier subsection, the interaction between size and book-to-market in the LTS regressions could potentially be somewhat complex because observations that are influential in the univariate regressions may not be influential in the other univariate regression or the bivariate regressions. The first three rows of Table II give an indication of the prevalence of this effect by showing the percent of trimmed observations that are common across the three LTS regressions with $\alpha = 5$ percent. The table compares the univariate size and book-to-market regressions and then compares each univariate regression to the bivariate regressions. These comparisons show that most of each month's trimmed observations are the same across the three regressions. Table II also shows that the trimmed observations have extremely large returns (average absolute month returns of 36 percent). Finally, in spite of the fact that the LTS algorithm is inherently symmetric, Table II shows that positive trimmed observations are generally larger and much more frequent. Both effects are a consequence of the positive skewness of the return distributions.

Table II

Summary Statistics for Trimmed Observations Using LTS with $\alpha = 5\%$

Trimmed observations are from the monthly least trimmed squares (LTS) regressions with a trimming proportion of $\alpha = 5\%$. Each statistic is calculated each month and the values reported in the table are the averages across the 330 months in the sample. The dependent variable in all of the monthly regressions is return. The monthly size regressions use the natural log of the market value of equity as the independent variable, the book-to-market regression uses the natural log of the ratio of book value of equity to market value of equity, and the bivariate regression uses both. The determination of which observations are trimmed can depend on the value of the independent variable (size and/or book-to-market) as well as the dependent variable (return). The first section of the table, labeled "Matching of trimmed observations," shows the percentage of trimmed observations that are common across the two regressions.

	Positive Residual Return	Negative Residual Return	All
Matching of trimmed observations			
Size regression vs. book-to-market regression	93%	86%	92%
Size regression vs. bivariate regression	96%	91%	95%
Book-to-market vs. bivariate regression	94%	88%	93%
Average absolute monthly returns of trimmed observations			
Size regression	37.0%	29.4%	36.1%
Book-to-market regression	37.5%	29.0%	36.3%
Bivariate regression	36.8%	29.4%	35.9%
Trimmed observations as a percentage of all trimmed observations			
Size regression	74.2%	25.8%	100.0%
Book-to-market regression	71.9%	28.1%	100.0%
Bivariate regression	74.8%	25.2%	100.0%
Trimmed observations as a percentage of all observations			
Size regression	3.7%	1.3%	5.0%
Book-to-market regression	3.6%	1.4%	5.0%
Bivariate regression	3.7%	1.3%	5.0%

Table III decomposes the effects of the trimming into cells in the size and book-to-market plane. The table is constructed in a manner similar to that used by Fama and French (1992) in their Table V on page 446. This analysis is a useful complement to the regression results, because it provides a check that the regression relations are monotonic and approximately linear. Each month the firms are sorted into NYSE size quintiles, and then within these groups by book-to-market quintiles. Panel A of the table shows the average return for the firms in each group, and Panel B shows the average returns for each group excluding all observations that are trimmed in the bivariate regressions with $\alpha = 5$ percent. As in Fama and French, these averages are first calculated for each month and then the monthly results are averaged across the 330 months in the sample. Consistent with the LS regression results shown in the third line of Table I, Panel A of Table III shows that average returns are a decreasing function of size and an increasing function of book-

Table III

Returns for Size and Book-to-Market Portfolios

Each month from July 1963 through December 1990, the firms are sorted into New York Stock Exchange (NYSE) size quintiles, and then within these groups into book-to-market quintiles. Panel A of the table shows the average returns using all of the firms in each group, and Panel B shows the average returns excluding all firms in the group that are trimmed in each month's bivariate least trimmed squares (LTS) regressions with $\alpha = 5\%$. As in Fama and French (1992), returns are first averaged within each month and then across the 330 months in the sample.

	Book-to-Market Quintiles				
Size Quintiles	1 (Lowest)	2	3	4	5 (Highest)
Panel A: Mean Monthly Return					
1 (smallest)	0.91	1.14	1.48	1.58	1.76
2	0.81	1.22	1.31	1.46	1.55
3	0.88	1.06	1.21	1.29	1.49
4	0.95	0.89	0.95	1.12	1.45
5 (largest)	0.86	0.91	0.98	0.85	1.06
Panel B: Mean Monthly Return Excluding Trimmed Observations					
1 (smallest)	−1.27	−0.35	−0.04	0.11	−0.02
2	0.19	0.64	0.81	1.04	0.96
3	0.53	0.77	0.91	0.98	1.16
4	0.77	0.80	0.79	0.97	1.23
5 (largest)	0.86	0.87	0.89	0.80	1.00

to-market. Panel B shows that, excluding the trimmed observations, average returns are a *increasing* function of size and an increasing function of book-to-market. This is consistent with the LTS results shown in the third line of Table I.

Table II shows that the LTS trimming eliminates more positive returns than negative returns and that the positive trimmed returns tend to be more extreme. By using the same size and book-to-market portfolios as in Table III, Table IV shows that both of these effects are much more pronounced for the small firms. Table IV also shows that these effects do not seem to vary much across book-to-market portfolios that control for size.

Table V examines to what extent the trimmed observations represent different firms each month. Panel A of Table V uses the firms that are in the sample in as of the final month (December 1990) and shows the percentage of these firms that are trimmed at least once in the 5 percent LTS bivariate regressions during the entire sample period. The results are shown separately for positive and negative trimming, and are not additive because a firm could have an extreme positive return in one month and an extreme negative return in another. The fact that the values in Panel A of Table V are much larger than the corresponding values in Table IV indicates that it is not the same firms being trimmed every month. A more interesting question is whether trimming tends to concentrate in a subset of firms, beyond what would be expected from

Table IV

Positive and Negative Trimmed Returns for Size and Book-to-Market Portfolios

Each month from July 1963 through December 1990, the firms are sorted into New York Stock Exchange (NYSE) size quintiles, and then within these groups into book-to-market quintiles. Panel A of the table shows the percentage of firms in each cell that are trimmed in the bivariate least trimmed squares (LTS) regressions with $\alpha = 5\%$, split between positive and negative residual returns. Panel B shows the average absolute value of the trimmed returns for the firms in the portfolio, again split between positive and negative residual returns. The statistics in both panels are calculated for each month and then averaged across the 330 months in the sample.

	Book-to-Market Quintiles									
	1 (Lowest)		2		3		4		5 (Highest)	
Size Quintiles	pos.	neg.	pos.	neg.	pos.	neg.	pos.	neg.	pos.	neg.
	Panel A: Percent Trimmed									
1 (small)	7.72	2.88	5.32	1.73	5.18	1.34	5.19	1.44	6.21	2.02
2	3.60	1.71	2.65	0.83	2.16	0.60	1.76	0.43	2.56	0.73
3	2.30	1.16	1.51	0.63	1.53	0.45	1.45	0.44	1.55	0.41
4	1.37	0.77	0.66	0.47	0.71	0.24	0.82	0.30	1.14	0.34
5 (large)	0.42	0.34	0.44	0.36	0.43	0.28	0.34	0.21	0.40	0.21
	Panel B: Average Absolute Value of Trimmed Returns									
1 (small)	37.8	32.0	37.2	31.2	38.0	31.2	36.6	30.2	39.2	30.1
2	33.2	31.1	33.9	30.1	32.9	28.2	33.6	30.0	36.9	29.5
3	29.3	28.6	31.6	28.6	30.8	25.9	33.0	27.4	32.6	26.9
4	29.4	29.0	32.2	27.4	31.8	25.1	28.4	24.8	30.8	27.8
5 (large)	25.7	29.7	29.0	25.0	34.9	24.2	27.4	22.3	31.6	23.3

the size and book-to-market differentials shown in Panel A of Table IV. To help answer this question, Panel B of Table V shows the average values from simulations that assume trimming probabilities are independent across months and constant across firms in a given size and book-to-market cell. By comparing Panels A and B of Table V, it is apparent that trimmed returns do concentrate somewhat in a subset of firms, but the effect is not dramatic for positive trimming. On the other hand, the percentages of firms that experienced at least one negative trimmed return are much lower than would be expected if the trimming probabilities were the same for all firms in a particular size/book-to-market cell. This result may be at least partially explained by the fact that the firms that experience large negative returns are less likely to survive to the end of the sample. We provide an analysis of the relation between trimming and delisting in the next section.

To this point we have not yet addressed the possibility that the influential observations represent errors in the data. Note that a "typical" price error involving a transposition will result in positive skewness. For example, a price incorrectly recorded as $25 instead of $52 will result in a negative 50 percent

Table V

Percentages of Firms Trimmed at Least Once Over the Full Sample

For December 1990, the firms are sorted into New York Stock Exchange (NYSE) size quintiles, and then within these groups into book-to-market quintiles. Panel A of the table shows the percentage of firms in each cell that are trimmed at least once in the July 1963 through December 1990 monthly bivariate least trimmed squares (LTS) regressions with $\alpha = 5\%$. The first number for each cell is the percent of firms that had at least one positive residual trimmed return, and the second number is the percent of firms that had at least one negative residual trimmed return. Panel B shows the average simulated percentages of firms that are trimmed at least once under the null hypothesis that each month's trimming probability is independent of other months and constant across all firms in a particular size/book-to-market cell. In these simulations, the size and book-to-market quintile values for the December 1990 firms are calculated each month exactly as in the actual regressions. The trimming probabilities used in the simulations are equal to the fraction trimmed in the actual regressions for each cell and for each month. The simulations use 1,000 trials, and the values in Panel A that are either greater than 97.5 percent of the simulated values or less than 2.5 percent of the simulated values (5 percent significance) are shown with an asterisk.

	Book-to-Market Quintiles									
	1 (Lowest)		2		3		4		5 (Highest)	
Size Quintiles	pos.	neg.	pos.	neg.	pos.	neg.	pos.	neg.	pos.	neg.
Panel A: Percent Trimmed at Least Once										
1 (small)	80.4	48.4*	71.5*	42.0*	76.1*	45.3*	74.8*	47.6*	78.6*	58.1*
2	64.2	38.9*	58.9*	34.7*	60.0*	35.8*	58.9*	30.5*	66.3*	51.6*
3	50.9*	23.6*	67.3*	21.8*	83.3	37.0*	61.8*	30.9*	50.9*	36.4*
4	50.0*	28.3*	54.3*	23.9*	57.4*	44.7	47.8*	28.3*	63.0	37.0*
5 (large)	51.3*	23.1*	56.4	15.4*	50.0*	42.1	56.4	38.5	46.2*	25.6*
Panel B: Simulated Percent Trimmed at Least Once										
1 (small)	79.0	61.6	76.6	61.3	82.3	66.5	83.0	66.1	83.9	69.7
2	62.7	52.3	66.6	51.1	74.3	59.5	77.6	59.5	79.8	62.6
3	63.4	51.3	77.3	57.2	81.8	60.5	83.3	62.0	81.6	56.8
4	63.5	49.0	70.4	49.4	79.3	56.5	76.4	52.0	73.2	54.2
5 (large)	70.0	55.0	69.4	54.4	73.9	55.9	64.8	47.6	68.8	54.3

return followed by a positive 100 percent return in the next month. It is unlikely that data errors are responsible for the results in our application, because the CRSP data have been thoroughly checked for errors, and the size and book-to-market effects have already received a great deal of scrutiny (see Breen and Korajczyk (1995), Kothari, Shanken, and Sloan (1995)).

C. Influential Months and the January Effect

We now turn to an investigation of the monthly slope estimates that make up the averages of the slopes (λs) reported in Table I. The t-tests for these averages implicitly assume the λs are normal and independently distributed, but an examination of the q-q plots for the slope coefficients (not shown) yields some evidence of excess kurtosis. In principle, this nonnormality could mean that the traditional t-statistic cutoffs are incorrect; however, bootstrap simulations using the demeaned empirical distributions of the slope coefficients indicate that this degree of kurtosis has only a minimal effect on the p-values.

As discussed in Section I the inferences from the standard t-statistics potentially are sensitive to serial dependence in the λs, so we recompute the t-statistics for Table I using the Newey-West correction. The corrected t-statistics for the mean LS slopes on size in the univariate and bivariate regressions fall to -1.42 and -0.926, respectively. The corrected t-statistics for the mean LS slopes on book-to-market are also reduced, but they remain statistically significant as do all of the t-statistics for the averages from the LTS regressions. Strictly speaking, once we allow for the possibility that the average risk premium on size may vary over time, we can no longer be confident that its unconditional mean is positive. This result is consistent with the findings of Brown, Keim, Kleidon, and Marsh (1983) that the size premium seems to disappear in certain subperiods. Of course, it doesn't necessarily imply that the size effect doesn't exist; rather, it means that our tests lack the power to allow us to be sure. Indeed, the estimated coefficients are economically interesting because there is roughly a 0.5 difference in the log of market value of equity between adjacent NYSE size deciles. Accordingly if the unconditional mean monthly slope on size is truly equal to the -0.12 mean shown in Table I, then there is a $12 \times 0.12 \times 0.5 \times 100 = 72$ basis point difference in expected annual returns between adjacent size deciles.

The next question to be addressed is whether there are particular months that have an extraordinary influence on the time series averages. Table VI shows the effect of using least trimmed squares to calculate the averages of the monthly LS slope coefficients excluding either 25 percent (82 of 330) or 5 percent (16 of 330) of the extreme months. Note that a priori there is no reason to expect that trimming would have any particular effect on the averages, and it will have no effect if the distribution of the monthly slope coefficients is symmetric. Table VI shows, however, that this trimming has a dramatic effect on the average size coefficients (but relatively little effect on the average book-to-market coefficients). The average LS size coefficient from the univariate regressions, which is -0.12 using all of the months, becomes $+0.11$ when 25 percent of the months are trimmed and -0.01 when only 5 percent of the months are trimmed. Thus, in most months there is a positive relation between average returns and size, even in the LS regressions. In addition, the negative average of the LS size coefficients can be entirely explained by the 16 months with the most extreme coefficients.

Table VI also shows trimmed mean calculations of the differences between the LS and LTS coefficients. We have shown above that the difference between the LS and LTS size coefficients is driven by large positive returns on a few small firms. Accordingly, the trimmed average of the difference between LS and LTS can indicate whether this phenomenon is much more pronounced in certain months, or whether it is unusually absent in certain months. As shown in Table VI, the trimmed means of the difference between the LS and LTS size

Table VI

Influential Months in the Average Slope Coefficients

Regressions of returns on size and book-to-market are run for each of the 330 months in the sample, using least squares (LS) and least trimmed squares (LTS) with $\alpha = 5\%$ of the individual firm observations trimmed. The averages of the monthly LS slope coefficients across the 330 months in our sample are reported in the "Least Squares" column in Table I and in column (1) below, which is labeled "no trim" because all 330 months are used in the average. Columns (2) and (3), labeled "25% trim" and "5% trim," are the averages of the same monthly LS slope coefficients, where the average is calculated using the LTS criterion to eliminate extreme months. The monthly differences between the coefficients from the LS and LTS regressions are averaged across all months and reported in column (4). Columns (5) and (6) are the averages of the differences between the LS and LTS monthly coefficients, where the average is calculated using the LTS criterion to eliminate extreme months.

	Averages of Monthly LS Slopes			Averages Difference Between LS and LTS Slopes		
	(1) No trim	(2) 25% trim	(3) 5% trim	(4) No trim	(5) 25% trim	(6) 5% trim
Univariate size regressions	−0.12	0.11	−0.01	−0.45	−0.35	−0.41
Univariate book-to-market regressions	0.47	0.46	0.41	0.12	0.07	0.11
Bivariate regressions						
Size coefficient	−0.08	0.09	0.01	−0.45	−0.37	−0.43
Book-to-market coefficient	0.35	0.40	0.40	−0.23	−0.17	−0.22

coefficients are somewhat lower than the mean difference using all months. Thus, the data do seem to indicate that there are some months with an unusually large difference between the LS and LTS slopes.

Given the data in Table VI, it is logical to ask what is special about the months that are being trimmed. Not surprisingly, one answer to this question is that they tend to be Januarys. In fact, Januarys represent 20–30 percent of the months that are trimmed in calculating either the average size coefficients or the average difference between the LS and LTS size coefficients, for both 25 percent and 5 percent trimming. Table VII directly illustrates the dramatic difference between Januarys and other months by showing the average LS and LTS size coefficients from the univariate regressions, calculated separately for Januarys and for the other months. The table shows that there is a strong January effect in both the LS and LTS size coefficients. In addition, the difference between the LS and LTS size coefficients is more pronounced in January.

This pattern in the LS size coefficients was first discovered by Keim (1983), who found that January accounted for more than fifty percent of the size effect for the period 1963 to 1979. The LS results in Table VII merely show that Januarys more than explain the size effect for the period examined by Fama and French (1992). The new result contained in Table VII is that the January effect in small firm returns is in part due to an increase in the positive skewness of small firm returns. That is, in January there is a greater tendency

Table VII

Average Size Slopes for January and Other Months

The table reports the averages of the monthly slope coefficients from univariate cross-sectional regressions of return on size, where size is the natural logarithm of the market value of equity. Five separate regressions are run each month, one using least squares (LS) and the others using least trimmed squares (LTS) with trimmed proportions of 50%, 25%, 5%, and 1%. The averages are calculated separately for January months in the sample and for all other months. Averages of the slopes on size for all 330 months in the sample are the same as those reported in the first line of Table I. The t-statistics for the averages are shown in parentheses.

	Least Squares	Least Trimmed Squares			
		$\alpha = 50\%$	$\alpha = 25\%$	$\alpha = 5\%$	$\alpha = 1\%$
January	−1.65	−0.56	−0.85	−0.98	−1.37
	(−6.96)	(−2.62)	(−3.90)	(−4.37)	(−5.64)
Other months	0.03	0.38	0.48	0.44	0.28
	(0.66)	(10.1)	(12.9)	(10.6)	(6.10)
All months	−0.12	0.30	0.37	0.33	0.14
	(−2.02)	(7.32)	(8.87)	(6.50)	(2.63)

for a few small firms to have large positive returns. Of course, the average LTS coefficients are still strongly negative in January, so this result explains only part of the January effect.

III. Economic Interpretation of Influential Observations

Our initial motivation for using the robust FM procedure was to isolate the influential observations to help uncover why size and book-to-market appear to be useful for explaining cross-sectional variation in returns. The fact that the size effect disappears when the 1 percent most extreme observations are trimmed each month provides a puzzle whose solution may cast light on the fundamental mechanisms that determine risk premia. Perhaps even more importantly, understanding what is special about the few small firms that experience large positive returns may yield insights into how firms grow. In this section, we discuss three possible hypotheses that might explain these influential observations. We refer to these hypotheses as: the price effect, turtle eggs, and takeover activity. We provide some preliminary tests of these hypotheses, but more extensive tests are clearly warranted. Our hope in this article is to provide a promising direction for future research in this area.

A. The Price Effect

Very low price firms have large bid-ask spreads as a percent of price. It is well known that this can induce a bias in average returns that is particularly severe when using equally-weighted averages.[9] To provide a simple and ex-

[9] See Roll (1983) and Blume and Stambaugh (1983). Fama and French (1993) use value-weighted average returns to attenuate this effect.

treme example, suppose that a subset of small firms bounce between a bid price of $¼ and an ask price of $⅜. Random arrivals of buys and sells will result in one fourth of these firms having positive returns of 50 percent and one fourth having negative returns of −33 percent. If there is no change in the quotes for these firms, the mean monthly return for this subset of firms measured using transactions prices will be 0.5 (0 percent) +0.25 (50 percent) +0.25 (−33 percent) = 4.25 percent. Since there was no change in the quotes, any investor who bought these firms at the beginning of the month and sold at the end earned a return of −33 percent. To test whether our results are driven by firms with low prices, we run the regressions reported in Table I with firms that have a price level of 5 dollars or less deleted from the sample. The resulting average size coefficients have somewhat smaller magnitudes when these low price firms are excluded, but the mean of the LS coefficients is still significantly positive and the means of the LTS coefficients with α ranging from 50 percent to 1 percent are all significantly negative. Thus, the differences between LS and LTS regressions are not explained by low price firms.

B. Turtle Eggs

The skewness inherent in the return distributions for small firms may provide insight into the mechanisms that drive returns. The turtle eggs hypothesis is motivated by the observation that while a few small firms burst forward each month, most seem to languish (at least relative to larger firms). The name follows from the fact that mother sea turtles lay many eggs, but few will hatch and fewer still will make it to the ocean. Those that do make it will have been the fortunate beneficiaries of proper tides and weather and a lack of predators. In the same way, relatively few small firms ultimately become large. Those that do are often the beneficiaries of a product discovery, a single large contract, or some other fortunate confluence of favorable events partly beyond their control. When such events occur, the firm leaps forward and the owners earn a large positive return. It is important to realize that following these events the firm is now larger, so the future return distribution changes. Indeed, returns are a measure of percentage change in value, so even if another event occurs with the same dollar impact, the percentage change will be smaller. This explains why the effect should be predominately associated with smaller firms.

We do recognize that although some of the small firms in the market may be "waiting to hatch," clearly not all firms will fit the metaphor. Some small firms, for example regional utility companies, have been small for a long time and will almost certainly remain small in the future. This suggests that information apart from size, such as age of the firm, could be useful in predicting which firms will be more likely to experience large returns. Indeed, Dunne, Roberts, and Samuelson (1989) find that young manufacturing plants are more likely to grow quickly and also are more likely to fail than older plants of similar size. In Table VIII, we test the relation between age and the probability of extreme returns, controlling for size and book-to-market.

Table VIII

Percentages of Positive and Negative Trimmed Returns for Young and Old Firms

Each month from January 1981 through December 1990, the firms are sorted into New York Stock Exchange (NYSE) size quintiles, and then within these groups into book-to-market quintiles. The firms are then further divided within each cell into young (added to the Center for Research in Security Prices (CRSP) data after January 1973) and old (in the CRSP data as of January 1973). For each of these subgroups, we calculate the percentage of returns trimmed in the bivariate least trimmed squares (LTS) regressions with $\alpha = 5\%$. These percentages are calculated separately for firms with positive residual returns from the regressions and for firms with negative residual returns. The values reported in the table are the time series averages of these percentages. Statistical significance for the difference in percentage of positive (negative) trimmed residual returns between young and old firms in the same size/book-to-market category is established using a t-test based on the monthly differences. Differences between young and old firms that are significant at the 0.05 level (two-tailed) are shown with an asterisk beside the percentage for young firms. Months for which there are no young firms in a particular category are excluded from the calculation of both the time series averages and the t-statistics.

		Book-to-Market Quintiles									
		1 (Lowest)		2		3		4		5 (Highest)	
Size Quintiles		pos.	neg.	pos.	neg.	pos.	neg.	pos.	neg.	pos.	neg.
1	young	8.11*	3.99*	5.92*	2.83*	6.24*	2.54*	5.40*	2.41*	6.15*	3.09*
(Smallest)	old	5.86	2.02	3.58	1.00	2.94	0.76	2.89	0.69	3.89	1.52
2	young	2.34*	1.73*	1.59	0.89	1.42	1.18	2.14	0.87	1.20	0.90
	old	1.54	0.61	1.19	0.43	1.14	0.27	1.14	0.27	1.24	0.55
3	young	1.27	1.41*	2.48*	1.76*	1.79	0.74	1.55	0.00*	0.54	0.00*
	old	0.84	0.45	0.55	0.23	0.72	0.17	0.64	0.29	0.58	0.30
4	young	1.82*	1.61*	0.67	0.64	0.73	0.93	0.12	0.00*	0.00*	0.00*
	old	0.69	0.35	0.41	0.15	0.55	0.06	0.27	0.10	0.45	0.21
5	young	0.29	0.44	0.29	0.39	0.69	0.00	0.00*	0.00	0.00*	0.00*
(Largest)	old	0.27	0.25	0.34	0.04	0.36	0.06	0.20	0.10	0.25	0.16

The number of months on the CRSP tapes gives a crude measure of age, but since the Nasdaq firms were not covered before 1973, we can't just use this to compare the age of Nasdaq firms with those from the NYSE and AMEX. In addition, the truncation of the ages in the sample as of the end of 1972 means even comparisons among Nasdaq firms in the 1970s would be unreliable. We control for these two problems by examining only the last 10 years of the data (1981–1990), and by making the somewhat arbitrary classification of firms as either young and old based on whether they were covered on CRSP as of January 1973. Table VIII separates the data into the same size and book-to-market cells as in Tables III and IV and then further separates the data in each cell into old and young firms based on the above definition. The table reports the percentage of positive and negative trimmed returns (in the bivariate LTS regression with a 5 percent trimming proportion) for both the young and old firms within each cell. The table shows that a larger fraction of young firms experience large (trimmed) returns, which is consistent with the turtle eggs story. Even for these young firms, however, the fraction of trimmed returns declines with size. Moreover, the old firms in the smallest size quintile are more likely to be trimmed than the young firms in the second size quintile, so size is still the more powerful predictor of extreme returns.

Another prediction of the turtle eggs hypothesis is that the large positive returns should be associated with announcements of positive events affecting the firm, such as the awarding of a large contract or the development of a new product. This hypothesis could be tested directly by looking at public announcement data, but we leave such tests to future research.

C. Takeover Activity

Another hypothesis concerning extreme returns, which would be consistent with the data presented thus far, is that small young firms are more likely takeover targets. In this case, the extreme large positive return results when the firm is put "in play." Contrary to the turtle eggs hypothesis, the returns realized in a takeover are related to the potential synergy gain resulting from combination with the acquiring firm, rather than resulting from a dramatic change in the fortunes of the firm on its own.

While it is undoubtedly true that some of the firms with extreme positive returns are takeover targets, the important question is to what degree this can explain the overall effect. To test this, Table IX examines whether the firms with trimmed returns are more likely to subsequently be taken over or go bankrupt. The table again sorts the firms each month into size quintiles, and then within size quintiles into book-to-market quintiles. The table further separates firms into subgroups based on whether they are trimmed in the bivariate LTS regressions with $\alpha = 5$ percent, and if so whether their residual return is positive or negative. For each subgroup, we calculate the percentage of firms that were subsequently taken over and the percentage of firms that were subsequently delisted or liquidated. These calculations use the CRSP delisting codes as of December 1993 (we consider codes 200–399 to be takeovers and codes 400–799 to be delistings or liquidations). The percentages are calculated each month and then averaged across the 330 months in the sample.

The truncation of the data as of December 1993 will mean that the percentages calculated for the later months will be biased downward, but under the null hypothesis that there is no difference in the probabilities of takeover or bankruptcy among the groups, the differences between the percentages each month should still be zero. Accordingly, we test the significance of the differences between the trimmed firms and the nontrimmed firms using t-statistics based on the monthly differences in the percentages.

Recall from Table III that most of the trimmed observations are in the small firm quintile, which is shown in the first three lines of Table IX. For small firms with extreme positive returns, the percentage subsequently taken over

Table IX

Percentages of Takeovers and Bankruptcies for Trimmed and Nontrimmed Returns

Each month from July 1963 through December 1990, the firms are sorted into New York Stock Exchange (NYSE) size quintiles, and then within these groups by book-to-market quintiles. These 25 groups are further subdivided into firms that have positive residual returns and are trimmed in the bivariate least trimmed squares (LTS) regressions with $\alpha = 5\%$, firms with negative residual returns that are trimmed, and firms whose returns are not trimmed. For each of these subgroups, we calculate the percentages that were subsequently taken over (columns labeled "T.O.") and the percentages that were subsequently liquidated or delisted (columns labeled "Del.") through December 1993. The values reported in the table are the time series averages of these percentages. Statistical significance for the difference between the firms with positive (negative) trimmed returns and the firms whose returns are not trimmed is established using a t-test based on the monthly differences. Trimmed groups that are significantly different from the nontrimmed group at the 5% level (two-tailed) are marked with an asterisk. Months for which there are no positive (negative) trimmed returns in a particular category are excluded from the calculation of both the time series averages and t-tests for the positive (negative) trimmed returns in that category.

		Book-to-Market Quintiles									
		1 (Lowest)		2		3		4		5 (Highest)	
Size Quintiles	Trimming	T.O.	Del.	T.O.	Del.	T.O.	Del.	T.O.	Del.	T.O.	Del.
1	positive	33.9	33.8*	38.1	27.4*	40.1	29.6*	42.1	29.9*	35.2	39.1*
(Small)	negative	28.3*	47.5*	28.4*	45.3*	26.1*	48.5*	31.9*	47.2*	26.0*	54.2*
	not trimmed	35.1	27.7	39.8	21.8	41.6	21.4	42.4	23.3	37.0	30.9
2	positive	50.1*	14.1	53.3*	9.5	54.7*	9.1	59.8*	6.8	55.0	8.9*
	negative	42.8	22.9*	52.5*	13.6*	44.2	13.1	36.1	20.5*	40.3*	13.4*
	not trimmed	41.7	10.4	40.5	7.1	40.8	6.1	43.9	6.4	51.0	5.6
3	positive	44.8	8.5*	54.2*	6.0	49.8	5.5	50.8*	4.3	55.6	10.7*
	negative	42.6	13.8*	48.5	15.0*	50.7	7.8	39.4	13.4*	56.9	11.8
	not trimmed	39.4	3.8	39.3	3.8	39.9	3.7	39.5	2.7	48.5	4.0
4	positive	46.8*	3.6	52.1*	3.5	51.0*	2.7	50.7*	5.7	37.2	7.2
	negative	41.4	3.3	52.0*	6.4	38.9	7.4	40.0	15.0	18.9	8.1
	not trimmed	28.3	1.1	26.1	2.2	31.5	1.0	27.2	1.6	31.9	2.9
5	positive	35.0*	0.0	53.8*	0.0	51.5*	0.0	29.2	4.2	33.8	0.0
(Large)	negative	34.7*	5.6	32.9	0.0	32.7	3.8	45.0*	7.5	17.4	2.2
	not trimmed	14.4	0.2	19.8	0.1	21.4	0.3	19.8	0.3	16.8	0.1

(shown in the columns labeled "T.O.") is no larger than for small firms that are not trimmed. On the other hand, small firms with extreme negative returns are more likely to be taken over than other small firms. This could indicate that some of the dramatic negative returns occur when investors suddenly realize that the firm has poor management, in which case a subsequent takeover could be more likely. The important conclusion to be drawn from these data is that takeovers do not explain the positive skewness in small firm returns. If anything the reverse is true; that is, controlling for takeover activity would make the tendency for small firms to have large positive returns appear even more extreme.

The percentages of firms that were subsequently delisted or liquidated are shown in the columns labeled "Del." in Table IX. Again focusing on the small firms shown in the first three lines of the table, it is no surprise that more of the small firms with large negative returns ultimately fail. What is surprising is that this also seems to be true for small firms with large positive returns. This is probably due to the fact that small firms are far from homogeneous. Some of the small firms are eggs that hatch into turtles, but many of these newly hatched turtles are eaten before they reach the ocean. Other small firms are just rocks on the beach, with no chance of hatching, but with no danger of being eaten.

IV. Conclusions

In this article, we extend the Fama and French (1992) monthly cross-sectional regressions to include a robust regression estimator that is not sensitive to outliers or leverage points. The estimator, called least trimmed squares, trims a fraction of the observations and fits the remaining data using least squares. By comparing the results from the standard least squares regressions to the results from least trimmed squares regressions with various trimming proportions, we show that the negative relation between firm size and average returns is driven by a few extreme positive returns in each month. In fact, when only 1 percent of each month's observations are trimmed, there is a significant positive relation between firm size and average returns.

We show that the difference between the results obtained using our robust Fama-MacBeth procedure and those obtained using standard least squares regressions is due to positive skewness in the return distributions that is particularly evident for small young firms. Our results suggest that investors who hold small firms anticipate a few major successes and many minor disappointments. That is, they lay many "turtle eggs" hoping a few will hatch and make it to the ocean. The idea that such firms should carry a risk premium is sensible and quite similar to the explanation given by Fama and French (1992). Our results show that small firms tend to realize this successful transition in "bunches." Risk averse investors should demand a premium to hold these small firm stocks if these transition events are correlated with changes in the discounted value of the investors' future labor income.

Our robust regression techniques provide a new way of isolating the periods (months) in which large numbers of small firms achieve dramatic success. These are just the months where the slope from the LTS regression differs from the slope from the LS regression. Examining the characteristics of the economy surrounding these months and the characteristics of the firms that succeed may yield new insights into the factors that drive risk premia, and a better understanding of how firms grow.

Although our results do not yield any specific trading strategies, they do have two implications for portfolio managers. The first is that the tendency for small firms to experience extreme positive returns increases the potential rewards from successful fundamental analysis. The second implication follows from the concentration of the small firms' successes in relatively few months. Portfolio managers able to understand and predict these events could achieve superior returns from a portfolio strategy that periodically switches between large and small firms. Of course, both of these implications are subject to the caveat that small firms have high transactions costs. Knez and Ready (1996) show that effective spreads are on the order of 5 percent for NYSE firms in the lowest size quintile. It would be necessary to have reasonably reliable predictions of the effects documented in this article in order to design trading strategies that would generate profits after transactions costs.

REFERENCES

Ball, Ray, 1978, Anomalies in relationships between securities' yields and yield-surrogates, *Journal of Financial Economics* 6, 103–126.

Ball, Ray, S. P. Kothari, and Jay Shanken, 1995, Problems in measuring portfolio performance an application to contrarian investment strategies, *Journal of Financial Economics* 38, 79–107.

Banz, Rolf W., 1981, The relationship between return and market value of common stocks, *Journal of Financial Economics* 9, 3–18.

Berk, Jonathan, 1995, A critique of size-related anomalies, *Review of Financial Studies* 8, 275–286.

Blume, Marshal E., and Robert F. Stambaugh, 1983, Biases in computed returns, *Journal of Financial Economics* 12, 387–404.

Breen, William J., and Robert A. Korajczyk, 1995, On selection biases in book-to-market based tests of asset pricing models, Working paper, Northwestern University.

Brown, Philip, Donald B. Keim, Alan W. Kleidon, and Terry A. Marsh, 1983, Stock return seasonalities and the tax-loss selling hypothesis, *Journal of Financial Economics* 12, 105–127.

Cressie, Noel, 1980, Relaxing assumptions in the one sample t-test, *Australian Journal of Statistics* 22, 143–153.

Dunne, Timothy, Mark J. Roberts, and Larry Samuelson, 1989, The growth and failure of U.S. manufacturing plants, *Quarterly Journal of Economics* 104, 672–698.

Fama, Eugene F., 1976, *Foundations of Finance* (Basic Books, New York).

Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–466.

Fama, Eugene F., and Kenneth R. French, 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.

Frankel, Richard, and Charles M. C. Lee, 1995, Accounting valuation, market expectation, and the book-to-market effect, NYSE Working paper 95–103.

Hampel, Frank, Elvezio Ronchetti, Peter J. Rousseeuw, and Werner Stahel, 1986, *Robust Statistics: The Approach Based on Influence Functions* (Wiley, New York).

Huber, Peter J., 1991, Between robustness and diagnostics, in Werner Stahel and Sanford Weisberg, Eds.: *Directions in Robust Statistics and Diagnostics* (Springer-Verlag, New York), pp. 121–130.

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

Knez, Peter J., and Mark J. Ready, 1996, Estimating the profits from trading strategies, *Review of Financial Studiess* 9, 1121–1163.

Kothari, S. P., Jay Shanken, and Richard G. Sloan, 1995, Another look at the cross-section of expected stock returns, *Journal of Finance* 50, 185–224.

Lakonishok, Josef, and Alan C. Shapiro, 1986, Systematic risk, total risk, and size as determinants of stock market returns, *Journal of Banking and Finance* 10, 115–132.

Lehmann, Edward L., 1986, *Testing Statistical Hypotheses* (Wadsworth & Brooks/Cole, California).

Maroni, Ricardo A., and Victor J. Yohai, 1991, Recent results on bias-robust regression estimates, in: Werner Stahel and Sanford Weisberg, Eds.: *Directions in Robust Statistics and Diagnostics* (Springer-Verlag, New York) pp. 121–130.

Martin, R. Douglas, Victor J. Yohai, and Ruben H. Zamar, 1989, Min-max bias robust regression, *The Annals of Statistics* 17, 1608–1630.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, heteroskedastic and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Richardson, Matthew, and Tom Smith, 1993, A test for multivariate normality in stock returns, *Journal of Business* 66, 295–321.

Roll, Richard, 1983, On computing mean returns and the small firm premium, *Journal of Financial Economics* 12, 371–386.

Rousseeuw, Peter J., 1984, Least median of squares regression, *Journal of the American Statistical Association* 79, 871–880.

Rousseeuw, Peter J., and Annick M. LeRoy, 1987, *Robust regression and outlier detection* (Wiley, New York).

Shanken, Jay, 1992, On the estimation of beta-pricing models, *Review of Financial Studies* 5, 1–33.

Stoll, Hans R., and Robert E. Whaley, 1983, Transaction costs and the small firm effect, *Journal of Financial Economics* 12, 57–79.