Does Size Really Matter?

Jonathan B. Berk

If the size of firms is measured correctly, small firms do not necessarily earn higher returns than larger firms. Yet, this finding is not inconsistent with the empirical fact that firms with small market values earn higher returns. Modern financial theory predicts that when firm size is economically unrelated to return, the relation between firm market value and return will be negative; that is, firms with small market values will have higher expected returns than other firms.

Ever since its discovery by Banz (1981), the size effect has remained an enigma. Although the empirical fact that firms with low market values earn higher returns than firms with high market values is beyond dispute, a theoretical explanation of this regularity has eluded the profession for almost 15 years. Because this empirical regularity is almost universally interpreted as evidence that small firms have higher expected returns than large firms, it has sparked a sometimes-heated debate that on occasion has even spilled over into the popular press (e.g., the *Economist* 1992). Theorists point out that no present theory can explain why small firms earn higher returns. Empiricists, on the other hand, point smugly to the seemingly unequivocal empirical evidence. What is the solution to this enigma?

Both sides of this debate are correct. No theoretical explanation of why small firms earn higher returns exists, because none is needed. If the size of firms is measured correctly, *no* evidence exists that shows small firms earn higher returns than large firms. Yet, this finding is not inconsistent with the empirical fact that firms with low market values earn higher returns than firms with high market values. Modern financial theory predicts that when there is no relation in the economy between firm size and return, the relation between firm market value and return will be negative; that is, firms with small market values will have higher expected returns.

THEORY

The solution to the enigma turns on the difference between a firm's size and the market value of its common equity.[1] Although no formal definition of firm size exists, most people would define it in terms of a firm's current assets. The market value of the firm, in contrast, is equal to the discounted value of the firm's expected future cash flows.[2]

When an investor decides to purchase a firm's stock, he or she bases that decision not on the value of the assets in place but rather on the discounted value of the cash flows those assets are expected to generate. The present value of a stream of cash flows depends on the cash flows' riskiness. All else being equal, riskier cash flows require higher discount rates and so have lower present values than less risky cash flows. Because a company's market value is the discounted expected value of its future cash flows, if the market values of two companies with the same expected cash flows are compared, the riskier company will have a lower market value than the less risky company. Thus, market value is a measure of the firm's discount rate or riskiness.

The correlation between the value of assets in place and the value of expected cash flows is clearly positive—large firms are likely to generate larger cash flows than small firms. Consequently, market value is also a measure of firm size, but firm size is not the only characteristic measured by market value. A company can have a low market value for two reasons: It is small (i.e., its expected cash flows are relatively small), or it is relatively risky (i.e., it has a relatively high discount rate). The second effect alone can completely explain the observed relation between market value and return.

The following thought experiment illustrates the main argument in this article. Imagine an economy in which all firms are the same size (i.e., have the same value of assets in place). The relation between market value and firm size depends on the relation between the firm's expected cash flows and its size. Assume that expected cash flows measure firm size perfectly; that is, assume that all firms, because they have the same size, also have the same expected cash flows. Of course, having the same expected cash flows is not the same as having the same cash flows. Because some firms' cash flows are likely to be riskier than others, discount rates

Jonathan B. Berk is an assistant professor of finance at the School of Business Administration at the University of Washington.

will differ in the economy. Consequently, the discounted value of a riskier firm's cash flows (i.e., its market value) will be lower than the discounted value of a less risky firm's cash flows.

A firm's expected return is defined to be its expected cash flows divided by its market value. The assumption that all firms have the same expected cash flows implies that (risky) firms with low market values have high expected returns and vice versa. Thus, even though I have explicitly assumed that all firms have the same size, the so called "size enigma" exists in this economy—market value is inversely related to return.

Although a risk-based argument is used to justify cross-sectional differences in discount rates, the relation between market value and expected return does not depend on the existence of a relation between risk and return in the economy. To see this point explicitly, consider a similar thought experiment in which stock prices are not influenced by the risk preferences of agents. For the sake of argument, assume that green men on Mars pick stock prices. If all stocks have the same expected cash flow, by definition, the stocks that these green men assign lower (higher) prices will have higher (lower) expected returns. Thus, even in this environment, market value is cross-sectionally inversely related to return. Clearly, the only condition under which market value will not be inversely related to return is if all assets have the same price—risk neutrality. *Any* theory that successfully explains any cross-sectional variation in expected returns must also predict an inverse relation between expected return and market value.

In reality, not all firms are the same size. To apply the above logic to actual markets, firms must have different sizes. Consider sorting these firms into two portfolios based on their size. The first portfolio consists of the smaller half of the firms; the second portfolio consists of the larger half of the firms. Because the assumption is still that firm size is unrelated to return, each portfolio should have the same expected return. Now, repeat the same sort using market value as the criterion. Given the assumptions, any differences between the sorts must result from differences in discount rates. To see why, recall that market value is the discounted value of the firm's expected cash flows. Consider the larger firm size portfolio. The only firms that are in this portfolio but are not in the higher market value portfolio are firms with market values low enough to drop into the lower market value portfolio—that is, firms with large discount rates. Similarly, the only firms in the smaller size portfolio that are not in the lower market value portfolio are small firms with relatively high market values; that is, firms with low discount rates. The difference, then, between the composition of the higher market value portfolio and the larger size portfolio is that firms with high discount rates are replaced with firms with low discount rates. Consequently, the average discount rate in the higher market value portfolio is lower than that in the larger size portfolio.

This finding implies that the expected return of the larger size portfolio exceeds the expected return of the higher market value portfolio. Similarly, the expected return of the lower market value portfolio exceeds the expected return of the smaller size portfolio. Recall, however, that both size portfolios have the same expected return. Thus, the expected return of the lower market value portfolio exceeds the expected return of the higher market value portfolio. Again, the so called "size enigma" is observed, even though in this economy, firm size is unrelated to return.

The same logic can be used to show that, regardless of how many portfolios are constructed, so long as they are well diversified, the expected return of the portfolios should be inversely ranked by a portfolio's average market value. These examples, therefore, demonstrate that the "size enigma" could result solely from the measure of firm size used in the literature—the market value of equity. Consequently, the relation between the size of a firm and its expected return is not necessarily evidence of an enigma.

Consider what will happen if, instead of investigating the relation between expected return and market value, we investigate the relation between expected return and the ratio of expected cash flows to market value. Another name for the ratio of expected cash flows to market value is expected return, so this ratio is tautologically perfectly correlated with expected return. This observation requires no auxiliary assumptions whatsoever. Unfortunately, expected cash flows are unobservable, so no such study has been undertaken; nonetheless, a similar study has been done.

The book value of equity measures depreciated past investment. Because the amount invested is likely to be highly correlated with the expected cash flows of the investment, one would expect the book value of equity and expected cash flow to be highly correlated. Consequently, book equity can be used as a proxy for the expected cash flow.[3] Therefore, the ratio of book equity to market equity is, in principle, a proxy for the ratio of expected cash flows to market value and should therefore be a better measure of expected return than market equity alone. In light of this relation, it is not surprising that the ratio of price to book value (or book to market value of equity) has been found to be a

better predictor of return than market equity alone.

Although the above argument shows that the size enigma *could* potentially result solely from the way firm size is measured, it does not guarantee that this is indeed the case. Consider how an inverse relation between firm size and return would affect the last example. In this case, the size-sorted portfolios would not have the same return; the portfolio with the smaller average size would have a higher expected return than the portfolio with the larger average size. The rest of the logic of the example is unaffected. The lower market value portfolio, therefore, has a higher expected return than the higher market value portfolio for two reasons. As before, the lower (higher) market value portfolio has a higher (lower) expected return than the smaller (larger) size portfolio. Now, however, the smaller size portfolio itself has a higher expected return than the larger size portfolio. The latter effect combines with the former to produce an even stronger inverse relation between market value and return. Consequently, when firm size and return are inversely related, the larger the variability in firm sizes, the more intense the relation between market value and return.

Market value is inversely related to return whether or not firm size itself is inversely related to return. Because all current empirical evidence documents only a relation between market value and return, it cannot answer the question of whether firm size and return are inversely related.

EVIDENCE

If small firms indeed earn higher returns, on average, than large firms, two other relationships should be observed in the market:

- Other measures of firm size besides market value should be inversely related to expected return, and
- the intensity of the relation between market value and return should depend on the variability in firm size within the data set used.

Are Other Firm-Size Measures Related to Return?

The standard approach that has been used to show that market value and return are inversely related is to sort stocks into portfolios based on market value. For example, Reinganum (1983) sorted stocks by market value and assigned the bottom 10 percent to the first portfolio, the next 10 percent to the second portfolio, and so on up to the 10th portfolio, which consisted of the 10 percent of stocks with the highest market values. The average return of each portfolio was then compared. Reinganum found that the returns of the portfolios were ranked, inversely, in the average market value of the portfolio. This approach is repeated here, first using market value and then using two other measures of firm size—the book value of assets (BVA) and the total value of annual sales (Sales).

The data set that was used is all NYSE stocks from 1967 through 1987. Return and market value quotations were obtained from the CRSP database, and BVA and Sales data were obtained from the Compustat database. In June of each year, using each size measure individually, the stocks were sorted into 10 portfolios of decreasing size. The return of each portfolio over the following year (i.e., July through June) was then recorded. This procedure was repeated each June in the sample period so that the portfolios were rebalanced 20 times during the study.

Table 1 presents the payoff that resulted from investing $1 in each portfolio in July 1966, reinvesting the proceeds at the end of the following June in the newly formed portfolio, and repeating this procedure until July 1987. These results are plotted in Figure 1. With only two exceptions, the returns of the portfolios sorted on market value monotonically increase as the average market value of the portfolio decreases.

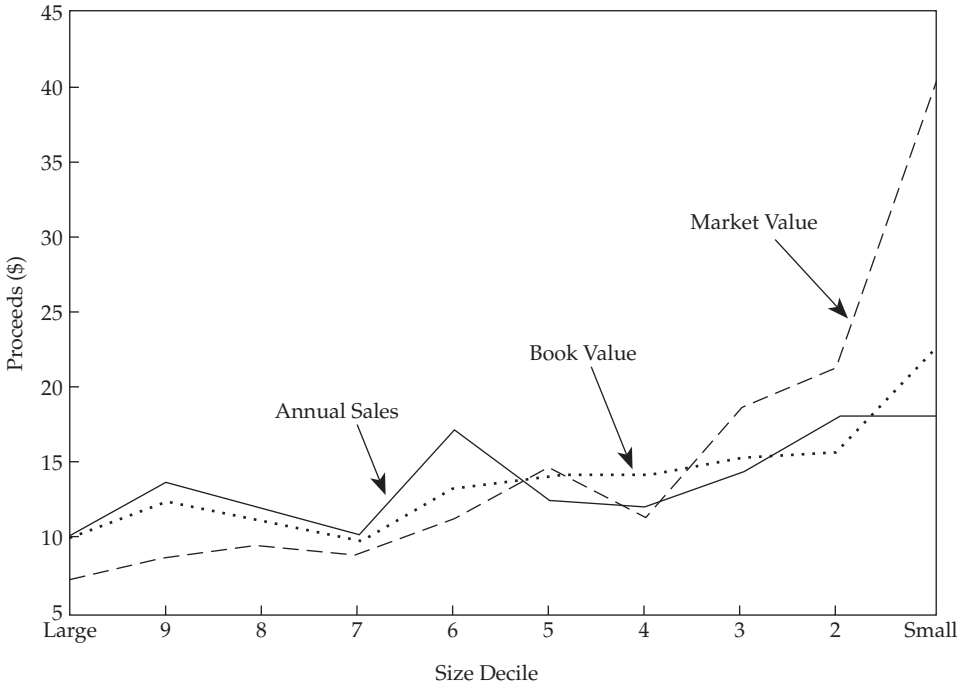
The results for the portfolios sorted using the other two size measures (BVA and Sales) are quite different: A clear pattern is not evident in the returns of these portfolios. When stocks are sorted using BVA, the largest portfolio does not have the lowest return. Although in all three cases the smallest portfolio has a higher return than the largest portfolio, the difference is much greater for the market value sorts than for the other two. For example, the difference in the payoff between the lowest

Table 1. Proceeds of $1 Invested from 1967 to 1987, by Size Decile and Size Variable Used in Sort

Size Deciles	MVA	BVA	Sales
Largest	$ 7.30	$ 9.95	$10.17
9	8.81	12.53	13.82
8	9.62	11.33	12.28
7	9.06	9.83	10.22
6	11.41	13.39	17.38
5	14.86	14.34	12.57
4	11.53	14.42	12.24
3	18.93	15.55	14.59
2	21.63	16.04	18.33
Smallest	40.51	22.99	18.44

Note: MVA = Market value of assets. BVA and Sales are the reported numbers for firms with fiscal year ends from June to May.

Figure 1. Proceeds of $1 Invested from 1967 to 1987, by Size Decile and Size Variable Used in Sort



and highest market value portfolios is $33, and the difference in the payoff between the smallest and largest BVA portfolios is only $13. This difference is so small for the two nonmarket measures that a formal regression test, undertaken in Berk (1995b), revealed that only the plot of the portfolios formed using market value (Figure 1) has a statistically significant upward slope. The other two plots are not significantly different from a flat line; that is, no significant relation between firm size and return was found for NYSE stocks.[4]

What Happens to Market Value with Firm Size Controlled?

Although in the previous test no statistically significant relation between BVA or Sales and average return was detected, the smallest portfolios did have higher returns than the largest portfolios. Is this result really attributable to statistical noise? A second test can resolve this skepticism: If firm size and return are related, then in a comparison of firms of similar sizes, the relation between market value and return should diminish.

For this analysis, firms were sorted into five size-based categories. The first consisted of the smallest 20 percent of firms, the second category consisted of the next smallest 20 percent, and so on up to the fifth category, which contained the largest 20 percent. By definition, the variability of firm size within each category is less than the variability in the whole sample. The relation between market value and return among firms of similar size can thus be studied by examining this relation within a category. If part of this relation is attributable to the relation between firm size and return, then when the relation between market value and return in the whole market and within a category is compared, the relation in the whole market should be more intense than that within the size category.

Panel A of Tables 2 and 3 presents the return on a $1 investment in 25 portfolios formed by first sorting stocks into five categories by a size variable other than market value. Within each category, stocks are then sorted into five portfolios based on their market values; that is, the lowest market value portfolio within a size category contains the 20 percent of stocks with the lowest market values within that category. The dispersion in the return within any nonmarket size category is shown in the rows of Panel A in each table. The bottom rows of both Panel A sections provide the average return of each market value quintile across the categories. The bottom rows, therefore, provide a measure of the overall effect of market value among firms of similar sizes. In every case, lower market value portfolios are associated with higher returns. Furthermore, the average dispersion in the payoffs within each market value category appears to be no different from the dispersion in the market as a whole (last column, Panel B, Tables 2 and 3), and

Table 2. Proceeds of $1 Invested from 1967 to 1987 in Portfolios Formed on MVA and BVA, by Sort Order

Panel A. Sort order: BVA and then MVA

	MVA Quintiles					
BVA Quintiles	Highest	4	3	2	Lowest	All
Highest	$7.38	$ 7.54	$ 9.41	$13.17	$19.33	$11.37
4	8.83	7.66	9.65	9.30	17.70	10.63
3	6.57	10.53	11.65	24.06	18.57	14.28
2	7.18	12.99	14.12	17.35	31.95	16.72
Lowest	8.31	10.96	15.58	18.92	40.71	18.90
All	7.66	9.93	12.08	16.56	25.65	

Panel B. Sort order: MVA and then BVA

	BVA Quintiles					
MVA Quintiles	Highest	4	3	2	Lowest	All
Highest	$ 8.58	$ 7.13	$ 7.02	$10.33	$ 6.60	$ 7.93
4	14.57	12.80	7.14	5.86	7.57	9.59
3	17.46	10.02	16.42	11.63	12.68	13.64
2	16.84	21.47	16.29	11.51	10.10	15.24
Lowest	30.22	25.20	21.91	22.95	19.96	24.05
All	17.53	15.32	13.76	12.46	11.38	

Note: The column (row) labeled "All" is the average across the cells in each respective row (column).

Table 3. Proceeds of $1 Invested from 1967 to 1987 in Portfolios Formed on MVA and Sales, by Sort Order

Panel A. Sort order: Sales and then MVA

	MVA Quintiles					
Sales Quintiles	Highest	4	3	2	Lowest	All
Largest	$ 7.42	$ 7.96	$10.59	$15.56	$17.66	$11.84
4	6.50	7.58	9.91	13.60	19.12	11.34
3	6.68	10.87	15.03	20.23	27.54	16.07
2	5.36	8.52	12.40	17.30	35.95	15.91
Smallest	10.60	10.40	16.32	18.38	29.93	17.13
All	7.31	9.06	12.85	17.01	26.04	

Panel B. Sort order: MVA and then Sales

	Sales Quintiles					
MVA Quintiles	Largest	4	3	2	Smallest	All
Highest	$ 8.78	$ 8.67	$ 9.94	$ 7.88	$ 4.87	$ 8.02
4	17.48	10.56	7.13	8.66	5.45	9.86
3	17.98	18.39	11.70	10.39	10.32	13.75
2	21.03	15.55	16.14	12.39	11.95	15.41
Lowest	35.06	23.56	18.26	28.68	17.23	24.56
All	20.07	15.34	12.63	13.60	9.96	

Note: The column (row) labeled "All" is the average across the cells in each respective row (column).

this dispersion is still much greater than any other size variable.[5] Thus, the relation between market value and return is as strong within firms of similar sizes as it is in the whole economy. We can thus conclude that no part of the relation between market value and return is attributable to a relation between firm size and return.

If the order of the sorts is reversed (Panel B of Tables 2 and 3), the results differ dramatically. If firms are first sorted into market value categories, firm size appears to be *positively* related to average return. Within each market value (MV) category, larger size is associated with higher, not lower, average returns. Although at first glance surprising, on further reflection, this result is consistent with the hypothesis that firm size and average return are unrelated. To see why, recall that the market value of a firm is influenced by two effects: Larger firms have high market values, and firms with high discount rates have lower market values. Consider, then, the set of firms in the lowest market value quintile. These firms consist of small firms and slightly larger firms that are in this quintile because they have high discount rates. In fact, the only way

a larger firm can get into this quintile is if it has a high discount rate. Consequently, *within this quintile*, the larger the firm, the higher the discount rate. Similarly, the only way the smallest firm in the largest market value quintile can remain in that quintile is if it has a low discount rate. Thus, the smaller the firm within this quintile, the lower the discount rate. When firm size itself is unrelated to the discount rate, then within any market value quintile, larger firm size should be associated with larger average returns.

This evidence shows that the size enigma is clearly completely consistent with modern financial theory. The reason firms with low market values earn higher returns is that they have higher discount rates than other firms, presumably because they are riskier. This finding, however, does not imply that *small firms* are riskier. As Tables 2 and 3 show, it is easy to construct portfolios of large firms that outperform portfolios of small firms. For example, the lowest MV firms in the largest Sales category outperform the highest MV firms in the smallest Sales category (see Panel A, Table 3). Yet, the average market value of the former portfolio far exceeds the average market value of the latter portfolio.

IMPLICATIONS FOR PORTFOLIO MANAGERS

The theoretical arguments and accompanying empirical evidence presented here have two important implications for portfolio managers. First, because the size effect is theoretically predicted, it should exist. Thus, the argument of, for example, Fouse (1989) that the relation between market value and return is simply an artifact of the time period studied (and is unlikely to recur) is fallacious. Because firms with low market values are more likely to have higher discount rates, they must have higher expected returns, on average. Therefore, over a long enough time period, a well-diversified portfolio of firms with low market values must earn higher returns than a well-diversified portfolio of firms with high market values.

The second implication is that, although the relation between market value and return should theoretically exist, the idea that somehow this extra return comes as a "free lunch" is misguided. Firms with low relative market values earn higher returns than others because they are riskier. Riskier firms not only have higher upside potential, they also have a higher probability of doing poorly. Thus, the empirical evidence that low market value portfolios underperform high market value portfolios in some time periods should not come as a surprise (see Reinganum 1992).

In spite of the fact that the relation between market value and return is not evidence of a free lunch in the economy, market value may still be a useful tool for designing portfolios. Given the risk–return trade-off in the economy, many clients are willing to take on increased risk if it is associated with a higher return. Small-capitalization stocks offer one means to pick such stocks. Indeed, the strategy of investing in firms with low market values has grown increasingly popular during the past decade. This strategy, however, is not necessarily the best means to achieve the high returns associated with risky stocks.

The riskiness of a stock does not depend on its size. Consequently, by restricting their investments to low market value stocks, portfolio managers ignore equally lucrative investments in large stocks. For example, $1 invested in 1967 in a portfolio consisting of the lowest market-value quintile produced $24 by 1987 (Panel B, Table 2). The same $1 invested in the lowest market-value quintile of the largest BVA category (i.e., the largest 20 percent of firms) returned $19 by 1987 (Panel A, Table 2). Statistically, these two payoffs are indistinguishable, although the average market capitalization of the latter portfolio far and away exceeds that of the former portfolio. By investing in the latter portfolio, a portfolio manager would have been able to construct a portfolio of only large-capitalization stocks yet still provide the same risk–return trade-off as investing in a portfolio of small-cap stocks. Once the increased transaction costs associated with investing in small-cap stocks are taken into account, the large-stock strategy could be more lucrative than a small-stock strategy.

CONCLUSION

The size enigma results from the measure of firm size used—the market value of common equity. Because this measure is not only a measure of firm size but also a measure of a firm's discount rate, even when firm size is not related to return, market value will be inversely related to return. Other measures of firm size show no evidence of a relation between firm size and return. The conclusion is that the size enigma results from the part of market value that measures the firm's discount rate and not from a relation between the size of firms and returns.[6]

NOTES

1. This section is an intuitive summary of the arguments first published in Berk (1995a). A reader interested in a formal approach should consult that article.
2. Throughout this article, "firm size" is used to denote the common definition of the size of the firm (e.g., value of current assets), not the market value of equity. The market capitalization of a firm will be referred to exclusively as "market value."
3. Berk, Green, and Naik (1996) provide a formal model in which value of book equity is a perfect proxy for the expected cash flow.
4. The small, statistically insignificant relation between the other measures of firm size and return (upward slope of the plots) could quite likely be the result of a data-snooping bias; that is, before this study was undertaken, the strong relation between market value and return was well known.
5. The slight differences in this column between the two tables result from the fact that, because of missing data, the data sets in the two tables are slightly different.
6. This research was supported by a grant from the Financial Research Foundation of Canada. The author would like to thank Paul Halpern, Rob Heinkel, and Dick Michaud for their comments and suggestions. A previous draft of this article is available on the author's World Wide Web home page: http://weber.u.washington.edu/~berk/.

REFERENCES

Banz, Rolf W. 1981. "The Relationship between Return and Market Value of Common Stocks." *Journal of Financial Economics*, vol. 9, no. 1 (March):3–18.

Berk, Jonathan B. 1995a. "A Critique of Size-Related Anomalies." *Review of Financial Studies*, vol. 8, no. 2 (Summer):275–86.

———. 1995b. "An Empirical Re-Examination of the Relation between Firm Size and Return." Working paper, University of Washington. Available at http://weber.u.washington.edu/~berk.

Berk, J.B., R.C. Green, and V. Naik. 1996. "Optimal Investment, Growth Options and Security Returns." Working paper, University of British Columbia. Available at http://weber.u.washington.edu/~berk.

Economist. 1992. "Beta Beaten." (March 7):87.

Fouse, William L. 1989. "The 'Small Stocks' Hoax." *Financial Analysts Journal*, vol. 45, no. 4 (July/August):12–15.

Reinganum, Marc R. 1983. "Portfolio Strategies Based on Market Capitalization." *Journal of Portfolio Management*, vol. 9, no. 2 (Winter):29–36.

———. 1992. "A Revival of the Small-Firm Effect." *Journal of Portfolio Management*, vol. 18, no. 3 (Spring):55–62.

