Journal of Banking & Finance 35 (2011) 3263–3274

Contents lists available at ScienceDirect

Journal of Banking & Finance

journal homepage: www.elsevier.com/locate/jbf

Is size dead? A review of the size effect in equity returns

Mathijs A. van Dijk *

Rotterdam School of Management, Erasmus University, P. O. Box 1738, 3000 DR Rotterdam, The Netherlands

ARTICLE INFO

Article history:
Received 13 February 2011
Accepted 11 May 2011
Available online 17 May 2011

JEL classification:
G12
G15
G31

Keywords:
Size effect
Cross-section of equity returns
CAPM
Anomalies

ABSTRACT

Beginning with Banz (1981), I review 30 years of research on the size effect in equity returns. Since Fama and French (1992), there has been a vigorous, ongoing debate on whether the size premium is a compensation for systematic risk. Since the late 1990s, research on the size effect has been characterized by two developments that are seemingly contradictory. At last, theoretical models have emerged in which the size effect arises endogenously as a result of systematic risk. However, recent empirical studies assert that the size effect has disappeared after the early 1980s. In this review, I address this disconnect between recent theoretical and empirical research.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

In this paper, I examine the evidence on the validity and persistence of the size effect in the cross-section of equity returns and the debate about the explanations for the effect. I also assess the implications for academic research and corporate finance. My review starts with a survey of the empirical studies up to and including Fama and French (1992). Subsequent studies focus on explaining the size effect. Since the late 1990s, a remarkable paradox has developed in research on the size effect. The objections against the lack of theory behind risk-based explanations for the size effect have at last resulted in the development of several theoretical models in which the size effect arises endogenously as a result of systematic risk. On the other hand, empirical research suggests that the size effect has disappeared since the early 1980s. This paradox has brought a virtual halt to research on the size effect.

I argue that there are two reasons why we need further empirical and theoretical research on the size effect. First, because it is premature to conclude that the size effect has gone away. Stock returns are very noisy and standard errors around estimates of the size premium are large, so it is not easy to tell whether the size effect is larger or smaller than it used to be. (In fact, I show that the US size premium was substantial in recent years.) The international evidence is also inconclusive. Second, because although the theoretical explanations offered for the size effect are potentially valuable, whether and how existing models can be reconciled with known patterns in the returns on small and large stocks is not clear. In particular, we need more empirical and theoretical research to evaluate the extent to which theories based on firm-level investment decisions, stock market liquidity, and investor behavior can explain the size effect.

The paper is structured as follows. I present an overview of the empirical evidence on the size effect in US and international stock markets in Section 2. In Section 3, I discuss a number of objections to the methods used in the empirical literature. In Section 4, I examine alternative explanations for the size effect. I assess the current state of the empirical and theoretical literature and discuss implications and directions for further research in Section 5. Section 6 concludes.

2. Empirical evidence on the size effect

In this section, I survey empirical studies on the size premium in US stock returns up to and including Fama and French (1992). I also present an overview of the international evidence on the size effect.

2.1. US evidence on the size effect

Table 1 summarizes the evidence on the size effect in the US equity market presented by studies in the 1980s and early 1990s. Banz (1981) wrote what may be the first empirical paper that

* Tel.: +31 10 408 1124; fax: +31 10 408 9017.
E-mail address: madijk@rsm.nl

0378-4266/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbankfin.2011.05.009

Table 1
Evidence on the size effect in the US. This table presents an overview of the empirical evidence on the size effect in the US up to and including Fama and French (1992). The results of all studies presented in the table are based on univariate sorting procedures on the basis of the market value of individual stocks. The columns present the estimated size premium (in % per month), the sample period, the number of securities analyzed, the number of size portfolios into which the securities are sorted, the ratio of the average market value of the firms in the largest size portfolio to the average market value of the firms in the smallest size portfolio, and the average monthly return and beta estimates of the firms in the largest and the smallest size portfolio. The estimates of the size premium are based on size-sorted portfolio returns that are not adjusted for risk. The returns on the smallest and largest size portfolios cannot be directly compared across studies, as some studies compute returns in excess of the risk-free rate or the return on a market portfolio.

	Size premium (% p.m.)	Test period	# Securities	# Portfolios	MV largest/ smallest	Return smallest (% p.m.)	Return largest (% p.m.)	Risk (β) smallest	Risk (β) largest
Banz (1981)	0.40	1936–1975	NYSE	5	NA	NA	NA	NA	NA
Reinganum (1981)	1.77	1963–1977	566	10	212	1.05	−0.72	1.00	0.82
Brown et al. (1983b)	1.85	1962–1978	566	10	NA	1.22	−0.63	NA	NA
Keim (1983)	2.52	1963–1979	1500–2400	10	248	1.72	−0.80	1.47	0.98
Lamoureux and Sanger (1989)	2.00	1973–1985	7659 Nasdaq	20	449	3.00	1.00	0.69	1.00
Lamoureux and Sanger (1989)	1.70	1973–1985	4170 NYSE/Amex	20	1519	2.50	0.80	0.95	0.91
Fama and French (1992)	0.63	1962–1989	NYSE/Amex/Nasdaq	10	296	1.52	0.89	1.44	0.92

presents evidence of a size effect in US stock returns. He analyzes all common stocks listed on the NYSE between 1936 and 1975. Banz reports that stocks in the quintile portfolio with the smallest market capitalization earn a risk-adjusted return that is 0.40% per month higher than the remaining firms. Fama–MacBeth (1973) regressions show a negative and significant relation between returns and market value. The size effect is not linear and is most pronounced for the smallest firms in the sample. Banz conjectures that many investors do not want to hold small stocks because of insufficient information, leading to higher returns on these stocks. This argument is reminiscent of the investor recognition hypothesis developed by Merton (1987).

Reinganum (1981) analyzes the size effect in a sample of 566 NYSE and Amex firms over the period 1963–1977. He finds that the smallest size decile outperforms the largest by 1.77% per month. Brown et al. (1983b) re-examine the size effect using the Reinganum data and find that there is an approximately linear relation between the average daily return on 10 size-based portfolios and the logarithm of the average market capitalization. Keim (1983) reports a size premium of no less than 2.5% per month in a broader sample of NYSE and Amex firms over the period 1963–1979. Keim shows that small firms have higher betas than large firms, but the difference cannot fully explain the return differential. Based on 20 size-sorted portfolios using a very large sample of firms, Lamoureux and Sanger (1989) find a size premium of 2.0% per month for Nasdaq stocks and of 1.7% for NYSE/Amex stocks over the period 1973–1985. They document that small firms have a lower beta than large firms on Nasdaq.

Although various important contributions appeared in the decade after the original work by Banz (1981), research on the size effect only really took off after the appearance of Fama and French (1992). They examine the size and book-to-market anomalies uncovered by earlier studies and make the case that the empirical shortcomings of the capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965) are simply too important to be ignored. Using a sample of NYSE, Amex, and Nasdaq stocks over the period 1963–1990, Fama and French find that the smallest size decile outperforms the largest by 0.63% per month. When they subdivide each size decile into 10 beta-sorted portfolios, they find no relation between beta and returns. Fama–MacBeth regressions confirm that beta does not help to explain the cross-section of returns, but both size and book-to-market equity have significant explanatory power. The flat relation between beta and returns has become known as the conjecture that "beta is dead." Subsequent research focuses on explaining the apparent breakdown of the CAPM and the causes of the size and book-to-market effects.

2.2. International evidence on the size effect

I am interested in examining the size premium in international equity returns for several reasons. First, because understanding the size effect in various countries makes is important for corporate finance and investment decisions in those countries. Second, because the strength of the size effect can depend on market characteristics such as the trading mechanism, the type of investors, and market efficiency in general. Third, because the finding that the size effect exists in different markets and in different time periods would make for a strong argument against data mining concerns (see, e.g., Lo and MacKinlay, 1990; Black, 1993).

Since the late 1980s, a large number of studies have examined the magnitude of the size effect in an international context. Table 2 presents the most recent estimates that are available on the monthly size premium for 19 individual countries and two groups of countries (emerging markets and Europe). All studies included in the table use univariate sorting procedures. Most studies adjust betas for non-synchronous trading using either the Scholes and Williams (1977) or the Dimson (1979) method. With the exception of the results for New Zealand and Spain, reported portfolio returns are not adjusted for risk.

Table 2 suggests that the international evidence on the size premium is remarkably consistent. Small firms outperform large firms in 18 of the 19 countries investigated, and also in a sample of emerging markets and in Europe. The monthly size premium in these countries ranges from 0.13% for the Netherlands to 5.06% for Australia. In 14 out of 19 countries, the size premium lies between 0.4% and 1.2% per month. Clearly, the reported size effects outside the US are substantial. These findings seem to indicate that data mining is not an important issue in research on the size effect. Numerous independent studies that use different data sets confirm the size effect originally found in US data.

However, there are a number of important caveats to this conclusion. First, it is hard to judge whether small firms also outperform large firms on a risk-adjusted basis. Roughly half of the studies included in Table 2 report a measure of the systematic risk of the size-sorted portfolios, but the other studies make no attempt at all to adjust for risk. Almost none of the international studies performs a formal cross-sectional test to investigate whether firm size can explain cross-sectional variation in stock returns. Fama–MacBeth regression results reported by Hou et al. (forthcoming) show no reliable relation between stock returns and firm size in

Table 2
International evidence on the size effect. This table presents an overview of the results of empirical studies on the size effect in international equity markets. *Sources*: Australia: Beedles (1992) [CRIF is the Center for Research in Finance]; Belgium: Hawawini et al. (1989); Canada: Elfakhani et al. (1998); China: Drew et al. (2003); Emerging markets: Rouwenhorst (1999); Europe: Annaert et al. (2002); Finland: Wahlroos and Berglund (1986); France: Louvet and Taramasco (1991); Germany: Stehle (1997); Ireland: Coghlan (1988); Japan: Chan et al. (1991); Korea: Kim et al. (1992); Mexico: Herrera and Lockwood (1994); Netherlands: Doeswijk (1997); New Zealand: Gillan (1990); Singapore: Wong et al. (1990); Spain: Rubio (1988); Switzerland: Corniolay and Pasquier (1991); Taiwan: Ma and Shaw (1990); Turkey: Aksu and Onder (2003); United Kingdom: Bagella et al. (2000).

Country	Size premium (% p.m.)	Test period	# Securities	# Portfolios	MV largest/smallest	Return smallest (% p.m.)	Return largest (% p.m.)	Risk (β) smallest	Risk (β) largest
Australia	5.06	1974–1987	CRIF data	10	267	6.82	0.76	1.43	1.04
Belgium[a]	0.52	1969–1983	170	5	188	1.17	0.65	1.01	0.98
Canada	0.98	1975–1992	694	5	178	2.00	1.02	0.58	0.60
China	0.92	1993–2000	44–701	2	NA	NA	NA	NA	NA
Emerging markets	0.69	1975–1997	1705	3	NA	2.42	1.73	NA	NA
Europe	1.45	1974–2000	2866	10	196	2.64	1.19	1.06	0.94
Finland	0.76	1970–1981	50	10	133	1.65	0.89	0.52	0.95
France[a]	0.90	1977–1988	529–460	5	NA	1.20	0.30	NA	NA
Germany	0.49	1954–1990	All FSE	9	NA	1.54	1.05	0.80	1.08
Ireland	0.47	1977–1986	40	5	NA	3.10	2.63	NA	NA
Japan	0.97	1971–1988	1570	4	57	2.44	1.47	1.10	0.81
Korea[a]	−0.40	1984–1988	NA	10	62	NA	NA	NA	NA
Mexico	4.16	1987–1992	100	3	37	5.80	1.64	1.31	0.79
Netherlands	0.13[b]	1973–1995	145–183	5	NA	NA	NA	NA	NA
New Zealand[a]	0.51[c]	1977–1984	100	5	60	0.69	0.18	0.90	0.99
Singapore[a]	0.41	1975–1985	NA	3	23	NA	NA	NA	NA
Spain	0.56[c]	1963–1982	160	10	228	0.58	0.02	NA	NA
Switzerland[a]	0.52	1973–1988	153	6	99	0.94	0.42	NA	NA
Taiwan[a]	0.57	1979–1986	53–72	5	17	0.47	−0.10	0.55	0.72
Turkey	3.42	1993–1997	106–173	4	NA	NA	NA	NA	NA
United Kingdom	1.18	1971–1997	541	10	1820	2.17	0.99	NA	NA

[a] Information taken from Hawawini and Keim (1995, 2000).
[b] Not statistically significant.
[c] Based on risk-adjusted returns.

a sample of 26,000 individual stocks from 48 countries. However, their regressions include no less than seven other firm characteristics, such as the book-to-market, cash flow/price, and dividend/price ratios. They do not present results for individual countries.

Second, the sample composition of several studies raises doubts about the reliability of the results. Papers that study 10 years of data or less (Ireland, Korea, Mexico, New Zealand, Taiwan, Turkey), fewer than 100 securities (Finland, Ireland, Taiwan), or sort stocks into three portfolios or less (China, Mexico, Singapore) are unlikely to yield a reliable estimate of the size premium. Also, the fact that some of the studies have not been published in reputable academic journals may give rise to concerns about possible inaccuracies in the sample selection and the application of the methods used. These concerns are reinforced when the reported size effects are extraordinarily large (e.g., 5.1% per month for Australia; 4.2% for Mexico; 3.4% for Turkey).

Third, few of the international studies perform a thorough analysis of the robustness of their results to, among other things, sample selection, return measurement interval, market index choice, extreme returns, and delisting bias. An exception is Chan et al. (1991), who show that the performance of the size variable in cross-sectional tests based on a sample of Japanese stocks is highly dependent on model specification and time period.

Fourth, there is another robustness issue that is important in cross-country studies. Should the size of a firm be measured relative to the average size of firms in its country? Annaert et al. (2002) report a significant size effect in a sample of stocks from 15 European countries. However, when they measure size relative to the average size of the firms within the same country, the size effect is no longer statistically significant. Although using absolute firm size makes it hard to distinguish the size effect in stock returns from a country effect, subtracting the mean firm size in a country from the size of an individual firm ignores the fact that the largest firms from a small country might be relatively small in a European context. These firms should earn relatively high returns if the size effect holds and European markets are integrated.[1] Rouwenhorst (1999) also reports a substantial size effect based on portfolios constructed by sorting stocks from 20 countries on the basis of firm size. But the size premium is positive in only 12 of the individual countries in his sample. Heston et al. (1999) report a significant size effect in their sample of stocks from 12 European countries that is almost entirely due to within-country variation in size. Similarly, Barry et al. (2001) only find evidence of a size effect in 35 emerging markets when they measure size relative to the local market.

3. Critique on the methods of empirical studies

In this section, I assess the most prominent criticisms of various studies' methods and their bearing on the reliability of the empirical evidence on the size effect.

3.1. The pitfalls of sorting methods

Lo and MacKinlay (1990) examine the extent to which the use of attribute-sorted portfolios in the empirical asset pricing literature influences classical statistical inference. Although sorting stocks into portfolios reduces the measurement error and enhances the power of the tests, grouping securities by some characteristic that is empirically motivated can lead to incorrect rejections of the null hypothesis that the asset pricing model is true. The argument is as follows. Asset pricing tests focus on the magnitude of the abnormal returns (alphas). Tests based on combinations of alphas for portfolios of securities can be more powerful. However, estimated alphas are equal to the sum of true alphas and measurement error. If researchers base the choice of the characteristic on which the securities are grouped only on

[1] Note that measuring size relative to the country average diminishes the dispersion in the explanatory variable, which reduces the power of the test. I thank Ken French for pointing this out.

an empirical analysis of the (same) data, then there is no way of knowing whether any resulting cross-sectional relation between the alpha of a portfolio and the characteristic is due to a relation between the characteristic and the true alphas or a relation between the characteristic and the measurement error. Lo and MacKinlay show that the type I error of such statistical tests can be up to 100% when the significance level used is 5%. This result does not necessarily imply that the size effect is spurious, since there may be a relation between firm size and true alphas. However, statistical tests should take account of the bias described by Lo and MacKinlay, although this is a very complex problem.

Berk (2000) criticizes the approach of sorting stocks into size portfolios first and then testing the explanatory power of a certain asset pricing model within each size decile. Berk shows that this technique is biased toward rejecting whatever asset pricing model is examined in the second sorting step. The intuition is that in the first step, by picking a variable that is empirically known to have a relation to stock returns, the return variation across groups is relatively large. Hence, the variation within groups is small and the statistical power to reject the null of a flat beta-return relation is low. Berk mentions Daniel and Titman (1997) as a prominent example of this approach, but Daniel and Titman (1999) dispute that Berk's critique applies to their tests.

3.2. Size picks up any omitted risk factor

Although variables related to a firm's stock price can be used to detect flaws in asset pricing tests, they also contain information about the cross-section of expected returns. Berk (1995) formalizes this argument. He shows that if the asset pricing model is incorrect (or if the empirical specification is incorrect), then firm size (measured by market capitalization) will always be inversely related with the part of return not explained by the model. The intuition is that of two firms with the same size (in the sense that end-of-period cash flows are equal), the firm with the riskier cash flows has a lower market value, and by definition a higher expected return. Hence, if there is an omitted risk factor, market value will show up significantly in cross-sectional tests. Even if the significance of firm size stems from an omitted risk factor and not from misspecification, it yields no economic insight about the nature of this factor.

This argument does not rule out that there is a priced risk factor related to size. In an unpublished manuscript, Berk (1996) investigates whether there is a relation between returns and firm size measured by several non-market-related variables, such as annual sales. Both multivariate sorts and Fama–MacBeth regressions show a negative relation between returns and market value, but no relation between returns and non-market measures of firm size.

3.3. Mismeasurement of the market portfolio

The CAPM implies that the market portfolio is mean–variance efficient and that there is a linear relation between expected returns and betas. Since the market portfolio is unobservable, empirical tests of the CAPM are inevitably flawed. This argument forms the basis of the critique by Roll (1977) and Ross (1977). Roll argues that a correct test requires the inclusion of every individual asset in the market portfolio. Stambaugh (1982) suggests that this problem is not severe, because inferences about the CAPM are similar for market portfolios that include bonds, real estate, and consumer durables in addition to common stocks.

However, Roll and Ross (1994) show that OLS estimates of the cross-sectional relation are very sensitive to the choice of the index. Even market portfolio proxies close to the mean–variance efficient frontier can produce zero slopes. Sampling error exacerbates these problems. Kandel and Stambaugh (1995) demonstrate that when the market index used is arbitrarily close to the mean–variance efficient frontier, OLS estimates of the slope of beta in a cross-sectional regression of expected stock returns can be arbitrarily close to zero. And a near perfect linear relation between beta and expected returns can be observed if the market index employed is far from efficient. Kandel and Stambaugh's analysis indicates that the use of generalized least squares (GLS) can reduce this problem. The GLS estimate of the slope is positive as long as the expected return on the market proxy exceeds the expected return on the minimum variance portfolio. But GLS assumes that all covariance parameters are known. The use of feasible GLS in empirical studies introduces further sampling error.

Black (1993) makes the more general point that if researchers use a market portfolio that differs from the true market portfolio, betas are estimated with error. It is likely that stocks that seem to have low betas will on average have higher betas when the true market portfolio would be used. This invalidates the analysis of whether the higher return on small stocks constitutes a premium for extra market risk. Ferguson and Shockley (2003) demonstrate that if the market proxy in empirical tests of the CAPM is equity-only, characteristics correlated with a firm's relative leverage and relative distress (such as firm size) will appear to explain returns.

3.4. Time-variation in risk-loadings and factor premia

The CAPM is a static model, and many empirical tests assume that betas are constant over time. However, the relative risk of a firm's cash flows is likely to fluctuate over time and depend on the business cycle. Conditional versions of the CAPM take this variability into account by making expected returns conditional on the information available to investors at a given point in time.

Jagannathan and Wang (1996) examine whether a conditional version of the CAPM can explain the cross-sectional variation in returns on 100 size-beta sorted portfolios of NYSE and Amex stocks between 1962 and 1990. They show that a conditional CAPM is able to explain roughly 30% of the cross-sectional return variation, compared to only 1% for the static CAPM. A second contribution of Jagannathan and Wang's (1996) paper is the inclusion of a measure of human capital in the market portfolio, in response to the market proxy problem. This specification of the conditional CAPM explains roughly 50% of the cross-sectional variation in average returns and leaves no additional explanatory power for firm size.

Lettau and Ludvigson (2001) analyze whether a conditional version of the Consumption CAPM (CCAPM) captures cross-sectional variation in the returns on 25 portfolios of US stocks sorted on size and book-to-market. They find that this model performs much better than either the unconditional specifications of the CAPM or the CCAPM, and about as well as the Fama–French (1993) three-factor model. Furthermore, the conditional CCAPM eliminates the residual size effect in the CAPM. Santos and Veronesi (2006) test a conditional CAPM in which labor income is the main state variable. Fama–MacBeth regressions on 25 portfolios sorted on size and book-to-market show no evidence of a size effect. Daniel and Titman (2005) argue that asset pricing tests based solely on 25 portfolios sorted on size and book-to-market can yield misleading results. In particular, they show that the ability of the CCAPM of Lettau and Ludvigson (2001) to explain the returns on industry portfolios is limited.

Lewellen and Nagel (2006) question whether the conditional CAPM can explain asset pricing anomalies. They estimate betas over short windows and thus do not need to rely on a proxy for investors' information sets. They focus on the model's ability to explain time-series (instead of cross-sectional) variation in the returns on 25 portfolios sorted on size and book-to-market over the period 1964–2001. Although betas vary considerably over time, they do not vary enough to explain known anomalies.

Pettengill et al. (1995) investigate a different, but related issue. They criticize the Fama and French (1992) methodology for not taking into account that the predictions of the CAPM are based on expected returns and thus imposing the restriction that the beta-return relation is the same in up and down markets. When relaxing this restriction, Pettengill et al. find a significantly positive beta in up markets and a significantly negative beta in down markets. In a follow-up study, Pettengill et al. (2002) show that the cross-sectional return premium associated with firm size is also asymmetric: the size effect is much more pronounced in down markets. They argue that the common assumption that betas are the same in up and down markets leads to an underestimation of the size effect.

4. Explanations for the size effect

The question why small firms earn higher returns than traditional asset pricing models predict has become the subject of a heated debate. Some papers contend that the systematic risk of a stock is driven by multiple risk factors, and firm size is a proxy for the exposure to state variables that describe time-variation in the investment opportunity set. An alternative interpretation is that the size premium is a compensation for trading costs and/or liquidity risk. A third fundamental explanation is embedded in asset pricing models that relax the assumption that investors are fully rational. Other papers argue that the size effect is little more than a statistical fluke. In this section, I provide an overview of the academic debate on the causes of the size effect.

4.1. Risk

Fama and French (1993) construct mimicking portfolios for the underlying risk factors related to size (SMB) and book-to-market (HML). The market, SMB, and HML portfolios capture a substantial part of the time-series variation in the returns on 25 stock portfolios formed on size and book-to-market over the period 1963–1991. Fama and French interpret this as evidence that size and book-to-market proxy for sensitivity to common risk factors in stock returns.

Fama and French (1995) find that size and book-to-market factors exist in earnings. And the SMB factor in returns is related to the SMB factor in earnings. Fama and French (1996) show that their three-factor model also captures the returns on portfolios formed on the basis of other anomalies, such as cash flow to price and the long-term return reversal documented by DeBondt and Thaler (1985). Fama and French argue that the empirical success of the three-factor model indicates that it is an equilibrium pricing model, a three-factor version of Merton's (1973) Intertemporal CAPM (ICAPM) or Ross's (1976) arbitrage pricing theory (APT).

Fama and French do not address the issue what state variables produce variation in earnings and returns related to size and book-to-market. Fama and French (1995, 1996) suggest that one of the state variables is related to financial distress. Chan et al. (1985) find evidence that the default spread and other variables that pick up changes in the economic environment capture the size effect in Fama–MacBeth regressions. Chan and Chen (1991) argue that small firms are generally "fallen angels" that have lost market value due to bad performance. Vassalou and Xing (2004) investigate the relation between the size and book-to-market effects and default risk. The size effect turns out to be only statistically significant within the highest default risk quintile. Petkova (2006) shows that SMB is related to innovations in variables that describe investment opportunities, such as the default spread. Hwang et al. (2010) find that the size and value effects can be explained by a CAPM extended with a credit spread factor, which they interpret as a proxy for the option feature of equity.[2]

Other studies question the conclusion that the size effect can be explained by distress. Dichev (1998) specifies a distress factor by measuring the probability of bankruptcy. Measures of ex ante bankruptcy risk are not associated with higher stock returns. Campbell et al. (2008) show that US firms with a high probability of bankruptcy have a high loading on the SMB factor. However, inconsistent with the conjecture that the size premium is a compensation for distress risk, these firms do not earn higher returns.

The interpretation that firm size proxies for a firm's exposure to an underlying risk factor is controversial. Ferson et al. (1999) use the hypothetical "alpha factor asset pricing model" to show that even if an observed phenomenon is completely unrelated to systematic risk, attribute-sorted portfolios (such as SMB and HML) based on that phenomenon will appear to be useful risk factors.

Daniel and Titman (1997) argue that firm characteristics, rather than factor loadings on the SMB and HML portfolios, determine expected returns. Within portfolios formed on size, there is essentially no relation between returns and loadings on the SMB factor. Expected stock returns thus seem to be related to firm characteristics for reasons that may have nothing to do with the covariance structure of returns.[3] Heston et al. (1999) present similar results for a large sample of European stocks. Davis et al. (2000) indicate that over the extended sample period 1929–1997, the US size effect is too small to accurately distinguish between the risk model and the characteristics model.

Starting with the study of Berk et al. (1999), a small but growing line of research addresses the critique that the risk-based explanations of the size effect are not founded on economic theory that identifies the state variables that drive variation in returns related to firm size. These papers analyze firm-level investment decisions in models in which the relation between firm size and stock returns arises endogenously. In the model of Berk et al., the dynamics of a firm's systematic risk and its expected returns are related to the size and the book-to-market ratio of the firm. The model predicts that time-series and cross-sectional variation in stock returns can be explained by market value, which serves as a proxy for the state variable in the model that describes the relative importance of assets in place and growth opportunities. Simulations based on this model perform reasonably well in generating empirical patterns similar to those detected by Fama and French (1992). Theoretical papers that build on Berk et al. include Gomes et al. (2003) and Carlson et al. (2004). However, empirical research that directly tests whether the patterns observed in the returns of small and large stocks are in line with the predictions of these models is essentially lacking.

[2] An alternative explanation for the size effect is idiosyncratic risk. Malkiel and Xu (1997, 2004) extend the dataset of Fama and French (1992) to the year 2000 and show that incorporating a measure of idiosyncratic risk absorbs the size effect in Fama–MacBeth regressions. Other papers, e.g., Goyal and Santa-Clara (2003), confirm that stock returns are related to idiosyncratic risk, but do not look at the relation with the size effect.

[3] Daniel et al. (2005) point out that even in models with risk neutral investors (i.e., without risk premia), loadings on attribute-sorted portfolios (such as HML and SMB) predict returns. However, in contrast to rational asset pricing models, their model with overconfident investors predicts that characteristics also have at least some explanatory power in the cross-section of stock returns.

4.2. Liquidity

The CAPM and other traditional asset pricing models abstract from the influence of liquidity and other market microstructure issues. However, transaction costs and liquidity risk can potentially account for the size premium. Since the early 1980s several empirical papers have directly or indirectly examined this issue. Stoll and Whaley (1983) find that it is not possible to earn abnormal risk-adjusted returns on small stocks after accounting for transaction costs in a sample of firms listed on NYSE over the period 1960–1979. Schultz (1983) extends the sample to Amex stocks and concludes that the size effect cannot be explained solely by differences in transaction costs between small and large firms.

Amihud and Mendelson (1986) develop a model in which investors require a compensation for expected trading costs. The model predicts that investors with longer holding periods tend to hold securities with larger bid-ask spreads. Because large trading costs are amortized over a longer holding period, the larger the spread, the smaller the compensation required for an additional increase in the spread. The relation between expected return and bid-ask spread should thus be concave. The authors present empirical evidence that supports this prediction. Eleswarapu and Reinganum (1993) criticize the approach of Amihud and Mendelson (1986) for excluding very small stocks from the sample. Their analysis of a much broader data set indicates that cross-sectional variation in the bid-ask spread cannot fully explain the size effect.

Other measures of liquidity are also related to stock returns. Brennan and Subrahmanyam (1996) show that fixed and variable transaction costs, estimated from transactions data by using market microstructure models, are positively and significantly related to returns. Datar et al. (1998) find that that turnover (the number of shares traded over shares outstanding) explains the cross-section of stock returns, even after controlling for size and book-to-market. Neither of these studies explicitly examine the relation between firm size and liquidity. Amihud (2002) uses the ratio of absolute stock return to dollar trading volume as a measure of price impact. Both size and this illiquidity measure are significant in Fama–MacBeth regressions, which suggests that the illiquidity variable does not capture the size effect completely.

Recent studies consider the possibility that liquidity is a priced state variable. If the returns on small stocks are more sensitive to this state variable, then part of the size effect can be related to liquidity risk. Amihud (2002) finds that the returns of small firms are relatively sensitive to time-series variation in market liquidity. Variation in the size premium may thus be related to time-variation in the price of liquidity risk. However, changes in market liquidity account for only a minor part of the time-series variation in returns.

Pastor and Stambaugh (2003) present evidence that systematic liquidity variation is a priced source of risk. They construct a market-wide liquidity factor based on the average across firms of the coefficient of trading volume on day t "signed" by the return on day t in regressions with the return on day $t + 1$ as the dependent variable. Portfolios of small firms have the highest loadings on the liquidity factor, but Pastor and Stambaugh contend that the relation between liquidity risk and firm size is not straightforward. They do not investigate whether size is a significant determinant of expected returns after correcting for liquidity risk.

In the model of Acharya and Pedersen (2005), the expected return on a stock depends on its expected liquidity and on the covariances of its own return and liquidity with the market return and liquidity. They test the model using Amihud's (2002) liquidity measure. Cross-sectional tests show that the model has a higher explanatory power than the CAPM, and that the liquidity risk premia are economically significant. Small stocks have lower average liquidity and higher exposures to the three liquidity risk factors. The liquidity risk factors improve the fit for portfolios of small stocks, but Acharya and Pedersen do not examine whether liquidity risk absorbs the size effect either.

4.3. Investor behavior

Investor behavior is often used to explain the value effect (e.g., the overreaction hypothesis of Lakonishok et al. (1994)), but similar explanations for the size effect are relatively unexplored. Yet, the overreaction interpretation of the value effect might hold water for the size effect. The argument is that value firms are typically firms that have shown poor performance in the past. If investors overextrapolate past performance, the stock price of value firms will be too low, which will result in higher returns when the overreaction is eventually corrected. Papers such as Chan and Chen (1991) indicate that small firms also tend to be firms that have done poorly in the past, but I am not aware of any research that explores whether overextrapolation is a driving force of the size effect.

A second plausible, albeit informal, explanation for the value effect is that investors like growth stocks and dislike value stocks. It seems equally legitimate to argue that investors prefer large stocks over small stocks. Gompers and Metrick (2001) suggest that the growing share of the US equity market held by institutional investors has boosted the demand for large and liquid stocks and thus diminished the relative performance of small stocks over the 1980–1996 period. Daniel and Titman's (1997) finding that size and book-to-market determine expected returns as characteristics rather than as proxies for risk could be interpreted in this light. Lakonishok et al. (1992) argue that agency relations affect portfolio selection by professional money managers. Investments in small stocks are potentially harder to justify to sponsors.

The size effect can also originate from incomplete information about small firms. Merton (1987) predicts that less well-known stocks of firms with smaller investor bases have higher expected returns. Hou and Moskowitz (2005) offer an empirical analysis of the influence of investor recognition on the size effect. As a broad measure for market frictions, the authors propose the average delay with which a firm's stock price reacts to information. Price delay has a significant impact on the cross-section of US stock returns over the period 1963–2001, and captures a substantial part of the size effect. Hou and Moskowitz argue that the results are most consistent with frictions associated with investor recognition.

Daniel et al. (2001) present a theoretical asset pricing model based on the premise that some investors are overconfident about their abilities. These investors overestimate the precision of information signals they receive. This behavior leads to mispricing that is not completely eliminated by fully rational investors, but is corrected over time. Hence, the model implies that proxies for misvaluation, such as fundamental/price ratios and market value, predict future returns. Daniel et al. argue that the empirical success of the Fama–French (1993) three-factor model does not discriminate between rational asset pricing theory and a mispricing theory such as the one they develop. I am not aware of any papers that do provide direct evidence on whether the size effect is consistent with a mispricing theory.

4.4. Statistical fluke

Several studies contend that the empirical finding of a size effect in stock returns is a chance result, driven by data mining, missing or extreme observations, and/or seasonal patterns in stock returns that have little to do with risk or liquidity. Here, I present a brief review of these papers.

4.4.1. Data mining and robustness

Among others, Black (1993), Lo and MacKinlay (1990), and MacKinlay (1995) argue that many researchers have used the same data to uncover the size effect and other asset pricing anomalies. Only the most successful, unusual, and striking results are published. As a result, it is impossible to assess their statistical significance, which depends on the number of attempts made to discover a certain effect. Out-of-sample tests are needed to counter the data mining argument.

Several authors investigate the robustness of the size effect over time. Banz (1981) and Keim (1983) show that the size effect in the US varies considerably over the period 1926–1975. Handa et al. (1989) report different estimates of the size effect in the US in three subperiods of their sample period 1941–1982. The size effect is negative (though not statistically significant) for 1941–1954. Pettengill et al. (2002) show that the size effect differs across bull and bear markets.

Recent papers suggest that the size effect disappeared after the early 1980s. Eleswarapu and Reinganum (1993), Dichev (1998), Chan et al. (2000), Horowitz et al. (2000a,b), and Amihud (2002) find no size premium over their sample periods 1980–1990, 1980–1995, 1984–1998, 1979–1995/1982–1996, and 1980–1997, respectively. Dimson and Marsh (1999) report that small stocks underperformed large stocks by 2.4% between 1983 and 1997. Hirshleifer (2001) contends that the size effect vanished after 1983. Schwert (2003) suggests that the size anomaly disappeared around the time of its discovery because practitioners began to use investment vehicles that tried to exploit the anomaly.

There is some indication that, just as in the US, the size premium varies across different time periods in non-US markets. Dimson and Marsh (1999) show that the size premium reversed in the UK. The size premium was 5.9% per year over the period 1955–1988, while it amounted to −5.6% over the period 1989–1997. Dimson et al. (2002) use large-cap and small-cap indices from Independence International Associates and FTSE International to obtain a crude indication of the sign and magnitude of the size premium in international equity markets. In 18 out of the 19 countries in their sample, the size effect appears to be reversed in the period after which an academic study on the size effect in that country appeared.

A reversal of the size effect in certain periods does not necessarily mean that small firms do not on average earn higher returns than their beta suggests. In fact, a risk-based explanation of the size effect implies that small stocks should be expected to underperform large stocks with some frequency. Bonds also occasionally outperform stocks over prolonged periods of time, but most economists would agree that stocks yield higher expected returns than bonds as a compensation for their higher systematic risk. But if these reversals occur often and/or over extended time periods, we might question the reliability of the empirical findings.

Fig. 1 plots the annual market-weighted return differential between the smallest and the largest size quintile of all NYSE, Amex, and Nasdaq stocks over the period 1927–2010.[4] The average return differential amounts to 6.7% per year over this period. There is a lot of variation in the size effect over time. There are 3 years in which the return on small stocks was over 50% points higher than the return on large stocks, but in many years the difference is modest. In 38 out of 84 years, the size premium is negative. Averaged over the past 30 years, the return difference is only 1.2% per year. Especially in the periods 1946–1957 and 1980–1999, small stocks clearly underperformed large stocks. Although the graph suggests that the size premium has become considerably smaller in the past decades, it is not straightforward to draw inferences about the validity of the size effect, because stock returns are very noisy and standard errors around estimates of the size premium are large. And over the period 2001–2010, the average size premium was no less than 11.3% per year.

Fig. 2 displays the annual returns on the US market portfolio in excess of the risk-free rate and the SMB and HML mimicking portfolios for 1927–2010. The SMB portfolio shows a similar return pattern as in Fig. 1. Although the returns on the market and the HML factor are more consistently positive over time, they also exhibit considerable volatility. The graphs do not show unambiguous evidence that the size effect has disappeared.

In short, we simply do not have enough data to conclude that the size effect has disappeared. It remains unclear what caused the remarkable decline in the size premium in the US after the early 1980s (and the remarkable comeback in the 2000s). Furthermore, Elton (1999) argues that realized returns can deviate from expected returns over prolonged periods of time.[5] So the "disappearance" of the size effect could be a temporary phenomenon driven by transitory information surprises that made realized returns on small and large stocks in the 1980s and 1990s deviate from expected returns.

4.4.2. Delisting bias and extreme returns

Shumway and Warther (1999) investigate the implications of the delisting bias in Nasdaq data. CRSP does not record a significant fraction of the returns associated with delistings. Shumway and Warther collect over-the-counter data on delisting returns and propose using a delisting return of –55% for the delisted stocks with missing data. They re-examine the evidence of Lamoureux and Sanger (1989) based on Nasdaq data over the period 1972–1995 and find no evidence that there ever was a size effect on Nasdaq. Wang (2000) uses simulation experiments to argue that the size effect is a spurious inference resulting from survival bias.

Knez and Ready (1997) show that the size effect is driven by the extreme 1% of the observations. They analyze the Fama and French (1992) data with a robust regression technique, least trimmed squares, which trims a proportion of the observations and fits the remaining observations using least squares. Fama and French (1992) trim the 0.5% most extreme book-to-market observations in the Fama–MacBeth regressions, but do not trim the extreme return observations. When Knez and Ready trim the extreme 1% of observations, the Fama–MacBeth regressions do not yield a significantly negative coefficient on firm size. Instead, they find a positive coefficient. They emphasize that the extreme observations should not be considered as outliers. However, the analysis of Knez and Ready does suggest that most small firms actually underperform big firms. The size effect seems to be due to a tiny fraction of the small firms that do extremely well. The authors speculate that this phenomenon is related to the "turtle eggs" effect: most small stocks do not perform well, but this is compensated by a few extremely successful firms. Fama and French (2007) examine the migration of firms across size portfolios and conclude that the size premium stems almost entirely from small stocks that earn extreme positive returns and as a result become big stocks.

4.4.3. Seasonality

The size effect can to a large extent be attributed to the extraordinary performance of small caps in January. Keim (1983) shows that a large part of the size premium is due to a return differential of no less than 15% between small and large stocks in January. Much of this difference originates from the first five trading days.

[4] I thank Ken French for providing the returns on the size quintiles, the market portfolio, and the SMB and HML factors on his website.

[5] Pettengill et al. (1995) also highlight the importance of distinguishing between realized and expected returns.

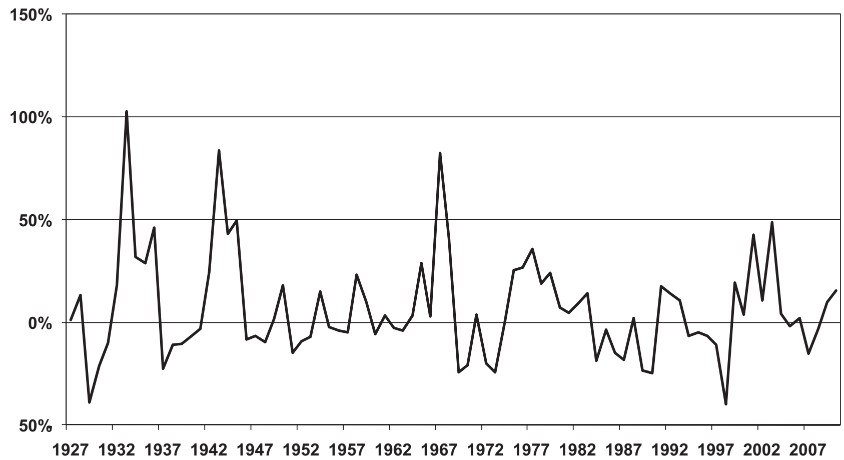

Fig. 1. The size effect in US equity returns 1927–2010. This figure depicts the annual, value-weighted return differential between the smallest and largest size quintiles of all NYSE, Amex, and Nasdaq stocks over the period 1927–2010. Returns are not adjusted for (market) risk.

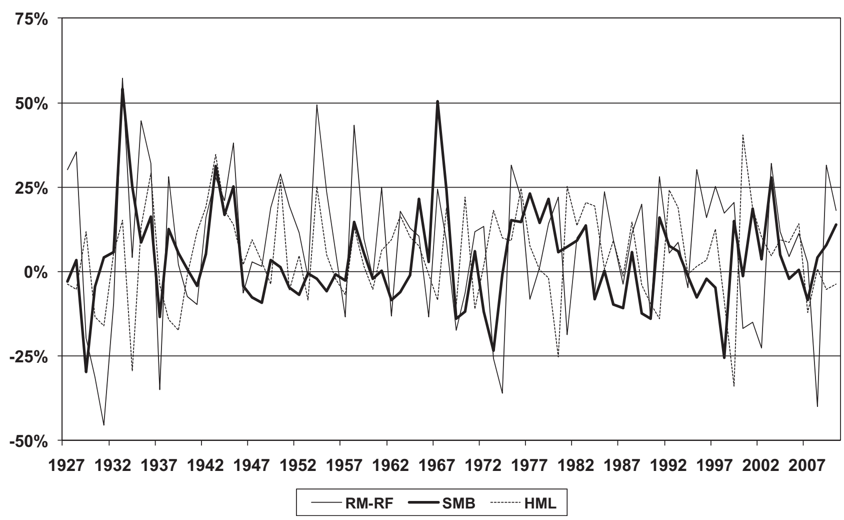

Fig. 2. The returns on the market, SMB, and HML portfolios 1927–2010. This figure shows the annual returns on the US market portfolio (in excess of the risk-free rate) and the SMB and HML portfolios over the period 1927–2010.

Brown et al. (1983b) find a strong size effect in January for NYSE stocks over the period 1967–1979. In a sample of Nasdaq stocks over the period 1972–1985, Lamoureux and Sanger (1989) observe that the smallest size decile outperforms the largest by 10.4% in January. Again, much of this outperformance happens in the first five trading days. Daniel and Titman (1997) find a strong relation between the size effect for NYSE stocks and the January seasonal in small cap returns in the period 1963–1993. Moller and Zinca (2008) document that the January effect has not decreased in strength in recent years, although it has become more concentrated in the beginning of the month.

Fig. 3 shows seasonal patterns in the market-weighted return differential between the smallest and the largest size quintile of all NYSE, Amex, and Nasdaq firms over the period 1927–2010. The graph suggests that the size effect in the US is indeed almost entirely due to higher returns on small stocks in January. The return differential amounts to more than 5% in January (not annualized) and is close to zero in all other months. A closer inspection of the origin of this seasonal effect shows that the strong January effect primarily shows up in the returns on the smallest size quintile, while the returns on the largest size quintile exhibit little seasonal variation.

Many researchers use the tax-loss selling hypothesis to explain the January effect. To take advantage of tax benefits, toward the end of the year individual investors have an incentive to sell stocks that declined in price during the year. After the turn of the year, in the absence of selling pressure, prices recover. This effect can be especially important for portfolios of small stocks, since these are biased toward shares that have experienced large price declines.

Thaler (1987) surveys early research on the January effect and the tax-loss selling hypothesis. Some studies, e.g., Roll (1983), report a negative relation between stock returns in January and the returns over the previous year. This finding is consistent with the tax-loss selling hypothesis. However, international evidence suggests that taxes cannot be the entire explanation. Brown et al. (1983a) expect a July seasonal effect in the return of small stocks in Australia, because the Australian tax year ends in June. However, the Australian size premium is stable over the year. Berges et al. (1984) find evidence of a January seasonal in Canadian stocks returns before the introduction of the capital gains tax in Canada in 1973. Kato and Schallheim (1985) find a strong January-size effect in Japan despite the fact that the tax regime in Japan generates no reason for tax-loss selling in December.

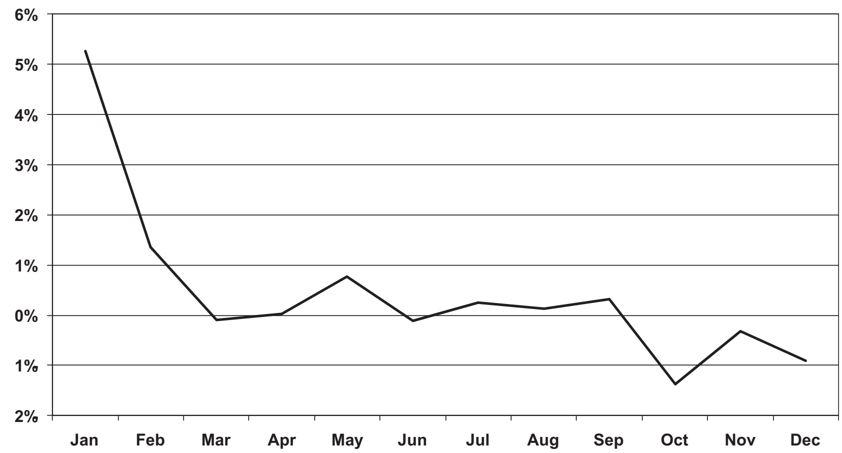

Fig. 3. Seasonal patterns in the size effect in US equity returns 1927–2010. This figure depicts the average market-weighted return differential between the smallest and largest size-quintiles of all NYSE, Amex, and Nasdaq stocks in each month over the period 1927–2010. Returns are not adjusted for (market) risk.

An alternative explanation for the January seasonal is the window dressing hypothesis. To present sound portfolio holdings, institutional investors have an incentive to buy winners (or low-risk stocks) and sell losers at the end of the year. Early in January, they rebalance their portfolios in favor of more speculative securities. Consistent with this hypothesis, Ritter and Chopra (1989) show that high-beta small firms earn higher returns than low-beta small firms, regardless of whether the market return is positive or negative. On the other hand, Sias and Starks (1997) show that the turn-of-the year effect is much stronger for securities dominated by individual investors. Thus, tax-loss selling by individuals is likely to be more important for these stocks than window dressing by institutions. Poterba and Weisbenner (2001) find that changes in the capital gains tax rule affect the pattern of turn-of-the-year returns. Since these changes have no implications for the incentives of institutional investors to engage in window dressing, these authors conclude tax-loss trading contributes to the January effect.

5. Implications for academic research and corporate finance

In this section, I evaluate the current state of empirical and theoretical research on the size effect.

Many of the early empirical studies identify a significant and consistent size premium in US equity returns, but more recent papers report that the effect may not be robust over time. It is remarkable that hardly any research has addressed the question of whether structural or institutional changes can account for the decrease in the size premium since the early 1980s – and for its strong comeback after 2000. The international evidence suggests that a substantial size premium exists in many non-US equity markets. However, systematic cross-sectional asset pricing tests as well as more elaborate robustness checks are required to make a truly compelling case for the existence of a size effect in international stock returns.

The debate on the causes of the size effect is still open. Fama and French (1993, 1995) interpret the evidence as indicating that the systematic risk of a stock is multidimensional, consistent with Merton's (1973) ICAPM. One of the state variables driving expected returns may be related to financial distress. Empirical research corroborates that a portfolio constructed to mimic such a risk factor explains cross-sectional variation in returns that is not explained by the market portfolio. But the empirical evidence on distress as a rationale for the size effect is mixed. And, importantly, this explanation is not backed by a formal economic theory.

Explanations for the size effect that are based on economic theory are essential for our understanding of the effect. Both sorting procedures and cross-sectional tests are vulnerable to data mining. And because market value and firm risk are inversely related, size will pick up any omitted risk factor in asset pricing tests. Ironically, around the publication of the first papers that suggest that the size effect has disappeared in the US, several new research initiatives started to address this deficiency. I identify three promising strands of theoretical literature.

First, models of firm-level investment decisions generate an endogenous relation between firm size and stock returns. Berk et al. (1999) and other studies present evidence that these models can reproduce several well-known features of stock returns. This body of research is still in a relatively early stage and it is not sufficiently clear to what extent these models can explain patterns uncovered by empirical research on the size effect.

Second, asset pricing models predict that stock returns not only depend on transaction costs, but also on liquidity risk. The available evidence indicates that liquidity is an important factor in asset pricing. However, most studies do not explicitly examine whether the size effect can be explained by liquidity factors, and the few studies that do present mixed evidence. How the size effect and liquidity interact is an important area for future research.

Third, the size effect can be rooted in the behavior of less than fully rational investors, in the sense that either these investors may prefer assets with specific characteristics or that size (market value) proxies for the mispricing that their behavioral biases cause. I am aware of only one paper that formalizes these arguments. In the model of Daniel et al. (2001), investors' overconfidence creates an endogenous relation between size and future returns.

The January effect presents a challenge to each of these three potential explanations for the size effect. It is not obvious why theories based on firm-level investment decisions, market liquidity, or investor behavior would predict a January seasonal in the returns on small stocks. For example, Lamoureux and Sanger (1989) show that transaction costs exhibit hardly any seasonal pattern and Sun and Tong (2010) find no evidence of increased volatility in January. Although several studies examine the January seasonal in relation to the size effect, I am not convinced that our current understanding of this phenomenon is complete. We need to know a great deal more about the January effect in international equity markets, and especially about the causes of seasonal variation in small-cap returns.

There are several other potentially valuable explanations for the size effect. First, the idea that idiosyncratic risk or investor recognition can account for the size effect merits further analysis. Second, the explanation based on the turtle eggs hypothesis is still embryonic. A third, and to my knowledge completely unexplored, avenue for future research is suggested by Graham and Harvey in their 2001 survey among CFOs: "... the fact that the practice of corporate finance differs based on firm size could be an underlying cause of size-related asset pricing anomalies."

Research that criticizes the methods used in empirical studies on the size effect has yielded two key insights. First, measuring the market portfolio is an important issue in asset pricing tests. It is surprising that few studies use market proxies that include other assets than stocks or apply GLS estimation to address this problem. Better-quality data on broad market proxies would be helpful for academic research in this area. The lack of readily available data also hampers applications of the CAPM, for example in estimating the cost of capital. Applying the CAPM with a market proxy that is equity-only will lead to systematic errors. Second, risk-loadings are time-varying. But although conditional asset pricing models are intuitively appealing, whether they can explain asset pricing anomalies is subject to debate. It seems unlikely that conditional versions of the CAPM will soon be used in event studies and cost of capital estimations, so the performance of unconditional models remains a relevant research topic.

In short, I find that the empirical evidence for the size effect is consistent at first sight, but fragile at closer inspection. I believe that more empirical research is needed to establish the validity of the size effect in both US and international stock returns. At the same time, there is little consensus about the origin of the size effect. Promising new theoretical research has been developed that relates firm size to firm-level investment decisions, liquidity, and investor behavior. These strands of the literature are still in a relatively early stage. Any theoretical explanation should address the issue why the size effect is especially pronounced in January. And more empirical research is needed to test the predictions of these theories.

Where does that leave academics and practitioners? Despite the ongoing debate about the merits of the Fama–French (1993) three-factor model, it is increasingly used in event studies, mutual fund performance evaluation, and consultancy reports on the estimation of a firm's cost of equity capital (see, e.g., Ibbotson Associates and Duff & Phelps, 2006).[6] Fama and French (2004) rationalize these applications by underscoring that researchers should want to measure the premia that investors require to hold stocks, whether these premia capture the exposure to unknown state variable risks or are the result of irrational investor behavior. Likewise, in event studies and performance evaluation, researchers should want to account for known patterns in returns, whether they are sample specific or permanent.

I hesitate to recommend the application of an empirically inspired asset pricing model while there is ambiguity about the robustness and the causes of the size effect it incorporates. I recognize the rationale of using the Fama–French (1993) three-factor model to evaluate whether portfolio managers achieve higher returns than expected on the basis of common investment styles. And researchers who identify new anomalies, other patterns in returns, or profitable trading strategies should show that these are not different manifestations of the size effect. But particularly for estimating a firm's cost of equity capital, the three-factor model seems barely less crude than using the average stock return of that firm (or its industry) in the past – especially since the estimates it produces are notoriously imprecise (see, e.g., Fama and French, 1997).

[6] Survey evidence presented by Graham and Harvey (2001) and Brounen et al. (2004) indicates that most companies in the US and in Europe still use the CAPM for the computation of their cost of capital.

6. Conclusion

Banz (1981) concludes his study on the size effect in US stock returns as follows: "It is not known whether size *per se* is responsible for the effect or whether size is just a proxy for one or more true unknown factors correlated with size" (his emphasis). Since the publication of Banz's 1981 paper, a huge body of literature has developed on the size effect. The size effect has been investigated empirically for many countries, and numerous studies have attempted to explain the anomaly. Academic research on the size effect has produced a wealth of knowledge on the behavior of asset prices, the application of statistical methods, and potential causes of the effect. But we still lack persuasive answers to the puzzle that Banz raised.

This survey identifies two seemingly contradictory developments in research on the size effect that both started in the late 1990s. Researchers have formulated theoretical models based on firm-level investment decisions, stock market liquidity, and investor behavior in which the size effect arises endogenously, but empirical studies have declared the size effect to be "dead" after the early 1980s.

I argue that the conclusion that the size effect has gone away is premature. In fact, I point out that the size premium in the US has been positive and large in recent years. We need more empirical research to examine the robustness of the size effect on US and international equity markets. In particular, we know little about the remarkable shifts in the size premium in the past few decades. At the same time, theories that potentially provide explanations for the size effect have neither been sufficiently developed nor systematically tested.

New research could break the current deadlock in an area of the academic finance literature that has important implications for our understanding of asset pricing and our approach to carrying out event studies, performance evaluation, and cost of capital estimation.

Acknowledgements

I thank Ike Mathur (the editor), an anonymous referee, Jonathan Berk, Long Chen, Karl Diether, Rüdiger Fahlenbrach, Ken French, David Hirshleifer, Kewei Hou, Kees Koedijk, Geert Rouwenhorst, Pieter van Oijen, and Christian Wulff for helpful suggestions and discussion. I am grateful for the hospitality of the Department of Finance at the Fisher College of Business (Ohio State University) where some of the work on this paper was performed. I thank Sandra Sizer for editorial assistance and PricewaterhouseCoopers for financial support. Any errors, misrepresentations, and omissions are my own.

References

Acharya, V.V., Pedersen, L.H., 2005. Asset pricing with liquidity risk. Journal of Financial Economics 77, 375–410.
Aksu, M.H., Onder, T., 2003. The Size and Book-to-market Effects and their Role as Risk Proxies in the Istanbul Stock Exchange. SSRN Working Paper.
Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. Journal of Financial Markets 5, 31–56.
Amihud, Y., Mendelson, H., 1986. Asset pricing and the bid-ask spread. Journal of Financial Economics 17, 223–249.
Annaert, J., Van Holle, F., Crombez, J., Spinel, B., 2002. Value and Size Effect: Now You See It, Now You Don't. SSRN Working Paper.
Bagella, M., Becchetti, L., Carpentieri, A., 2000. The first shall be last: size and value strategy premia at the London stock exchange. Journal of Banking and Finance 24, 893–919.

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

Barry, C.B., Goldreyer, E., Lockwood, L., Rodriguez, M., 2001. Robustness of Size and Value Effects in Emerging Equity Markets, 1985–2000. SSRN Working Paper.

Beedles, W.L., 1992. Small firm equity cost: evidence from Australia. Journal of Small Business Management 57, 65.

Berges, A., McConnell, J.J., Schlarbaum, G.G., 1984. The turn-of-the-year in Canada. Journal of Finance 39, 185–192.

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

Berk, J.B., 1996. An Empirical Re-Examination of the Relation between firm Size and Return. Working Paper, University of Washington.

Berk, J.B., 2000. Sorting out sorts. Journal of Finance 55, 407–427.

Berk, J.B., Green, R.C., Naik, V., 1999. Optimal investment, growth options, and security returns. Journal of Finance 54, 1553–1607.

Black, F., 1993. Beta and return. Journal of Portfolio Management 8, 18.

Brennan, M.J., Subrahmanyam, A., 1996. Market microstructure and asset pricing: on the compensation for illiquidity in stock returns. Journal of Financial Economics 41, 441–464.

Brounen, D., de Jong, A., Koedijk, C.G., 2004. Corporate finance in Europe: confronting theory with practice. Financial Management 33, 71–101.

Brown, P., Keim, D.B., Kleidon, A.W., Marsh, T.A., 1983a. Stock return seasonalities and the tax-loss selling hypothesis: analysis of the arguments and Australian evidence. Journal of Financial Economics 12, 105–127.

Brown, P., Kleidon, A.W., Marsh, T.A., 1983b. New evidence on the nature of size related anomalies in stock prices. Journal of Financial Economics 12, 33–56.

Campbell, J.Y., Hilscher, J., Szilagyi, J., 2008. In search of distress risk. Journal of Finance 63, 2899–2939.

Carlson, M., Fisher, A., Giammarino, R., 2004. Corporate investment and asset price dynamics: implications for the cross-section of returns. Journal of Finance 59, 2577–2603.

Chan, L.K.C., Chen, N.-F., 1991. Structural and return characteristics of small and large firms. Journal of Finance 46, 1467–1484.

Chan, L.K.C., Chen, N.-F., Hsieh, D.A., 1985. An exploratory investigation of the firm size effect. Journal of Financial Economics 14, 451–471.

Chan, L.K.C., Hamao, Y., Lakonishok, J., 1991. Fundamentals and stock returns in Japan. Journal of Finance 46, 1739–1764.

Chan, L.K.C., Karceski, J., Lakonishok, J., 2000. New paradigm or same old hype in equity investing? Financial Analysts Journal 56, 23–36.

Coghlan, H., 1988. Small firms versus large on the Irish stock exchange: an analysis of the performances. Irish Business and Administrative Research 9, 10–20.

Corniolay, C., Pasquier, J., 1991. CAPM, Risk premium seasonality and the size anomaly: the Swiss case. Journal of the French Finance Association 12, 23–44.

Daniel, K., Titman, S., 1997. Evidence on the characteristics of cross-sectional variation in stock returns. Journal of Finance 52, 1–33.

Daniel, K., Titman, S., 1999. Sorting out Sorting out Sorts. Working Paper, Northwestern University.

Daniel, K., Titman, S., 2005. Testing Factor-Model Explanations of Market Anomalies. Working Paper, Northwestern University.

Daniel, K., Hirshleifer, D., Subrahmanyam, A., 2001. Overconfidence, arbitrage, and equilibrium asset pricing. Journal of Finance 56, 921–965.

Daniel, K., Hirshleifer, D., and Subrahmanyam, A., 2005. Investor Psychology and Tests of Factor Pricing Models. Working Paper, Northwestern University.

Datar, V.T., Naik, N.Y., Radcliffe, R., 1998. Illiquidity and stock returns: An alternative test. Journal of Financial Markets 1, 203–219.

Davis, J.L., Fama, E.F., French, K.R., 2000. Characteristics, covariances, and average returns: 1929 to 1997. Journal of Finance 55, 389–406.

DeBondt, W.F.M., Thaler, R.H., 1985. Does the stock market overreact? Journal of Finance 40, 793–805.

Dichev, I.D., 1998. Is the risk of bankruptcy a systematic risk? Journal of Finance 53, 1131–1147.

Dimson, E., 1979. Risk measurement when shares are subject to infrequent trading. Journal of Financial Economics 7, 197–226.

Dimson, E., Marsh, P., 1999. Murphy's law and market anomalies. Journal of Portfolio Management 25, 53–69.

Dimson, E., Marsh, P., Staunton, M., 2002. Triumph of the Optimists: 101 Years of Global Investment Returns. Princeton University Press.

Doeswijk, R.Q., 1997. Contrarian investment in the Dutch stock market. De Economist 145, 198–573.

Drew, M.E., Naughton, T., Veeraraghavan, M., 2003. Firm size, book-to-market equity and security returns: evidence from the Shanghai Stock Exchange. Working Paper, Queensland University of Technology.

Eleswarapu, V.R., Reinganum, M.R., 1993. The seasonal behavior of the liquidity premium in asset pricing. Journal of Financial Economics 34, 373–386.

Elfakhani, S., Lockwood, L.J., Zaher, T.S., 1998. Small firm and value effects in the Canadian Stock Market. Journal of Financial Research 21, 277–292.

Elton, E.J., 1999. Expected return, realized return, and asset pricing tests. Journal of Finance 54, 1199–1220.

Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. Journal of Finance 47, 427–465.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–65.

Fama, E.F., French, K.R., 1995. Size and book-to-market factors in earnings and returns. Journal of Finance 50, 131–155.

Fama, E.F., French, K.R., 1996. Multifactor interpretations of asset pricing anomalies. Journal of Finance 51, 55–84.

Fama, E.F., French, K.R., 1997. Industry costs of equity. Journal of Financial Economics 43, 153–193.

Fama, E.F., French, K.R., 2004. The CAPM: theory and evidence. Journal of Economic Perspectives 18, 25–46.

Fama, E.F., French, K.R., 2007. Migration. CRSP Working Paper, 614.

Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: empirical tests. Journal of Political Economy 81, 607–636.

Ferguson, M.F., Shockley, R.L., 2003. Equilibrium anomalies. Journal of Finance 58, 2549–2580.

Ferson, W.E., Sarkissian, S., Simin, T., 1999. The alpha factor asset pricing model: a parable. Journal of Financial Markets 2, 49–68.

Gillan, F.L., 1990. An investigation into CAPM anomalies in New Zealand: the small firm and price-earnings ratio effects. Asia Pacific Journal of Management 7, 63–78.

Gomes, J., Kogan, L., Zhang, L., 2003. Equilibrium cross section of returns. Journal of Political Economy 111, 693–732.

Gompers, P.A., Metrick, A., 2001. Institutional investors and equity prices. Quarterly Journal of Economics 116, 229–259.

Goyal, A., Santa-Clara, P., 2003. Idiosyncratic risk matters! Journal of Finance 58, 975–1007.

Graham, J.R., Harvey, C.R., 2001. The theory and practice of corporate finance: evidence from the field. Journal of Financial Economics 60, 187–243.

Handa, P., Kothari, S.P., Wasley, C., 1989. The relation between the return interval and betas: implications for the size effect. Journal of Financial Economics 23, 79–100.

Hawawini, G., Keim, D.B., 1995. On the predictability of common stock returns: world-wide evidence. In: Jarrow, R., Maksimovic, V., Ziemba, W.T. (Eds.), Handbooks in Operations Research and Management Science. North Holland, Amsterdam.

Hawawini, G., Keim, D.B., 2000. The cross-section of common stock returns: a review of the evidence and some new findings. In: Keim, D.B., Ziemba, W.T. (Eds.), Security Market Imperfections in World-Wide Equity Markets. Cambridge University Press.

Hawawini, G., Michel, P., Corhay, A., 1989. A look at the validity of the CAPM in light of equity market anomalies: the case of Belgian common stocks. In: Taylor, S. (Ed.), A Reappraisal of the Efficiency of Financial Markets, NATO ASI Series. Springer Verlag.

Herrera, M.J., Lockwood, L.J., 1994. The size effect in the Mexican stock market. Journal of Banking and Finance 18, 621–632.

Heston, S.L., Rouwenhorst, K.G., Wessels, R.E., 1999. The role of beta and size in the cross-section of European stock returns. European Financial Management 5, 9–27.

Hirshleifer, D., 2001. Investor psychology and asset pricing. Journal of Finance 56, 1533–1597.

Horowitz, J.L., Loughran, T., Savin, N.E., 2000a. Three analyses of the firm size premium. Journal of Empirical Finance 7, 143–153.

Horowitz, J.L., Loughran, T., Savin, N.E., 2000b. The disappearing size effect. Research in Economics 54, 83–100.

Hou, K., Moskowitz, T.J., 2005. Market frictions, price delay, and the cross-section of expected returns. Review of Financial Studies 18, 981–1020.

Hou, K., Karolyi, G.A., Kho, B.C., forthcoming. What fundamental factors drive global stock returns? Review of Financial Studies.

Hwang, Y.-S., Min, H.-G., McDonald, J.A., Kim, H., Kim, B.-H., 2010. Using the credit spread as an option-risk factor: size and value effects in CAPM. Journal of Banking and Finance 34, 2995–3009.

Ibbotson Associates and Duff & Phelps, 2006. Risk Premium Report.

Jagannathan, R., Wang, Z., 1996. The Conditional CAPM and the cross-section of expected returns. Journal of Finance 51, 3–53.

Kandel, S., Stambaugh, R.F., 1995. Portfolio inefficiency and the cross-section of expected returns. Journal of Finance 50, 157–184.

Kato, K., Schallheim, J.S., 1985. Seasonal and size anomalies in the Japanese stock market. Journal of Financial and Quantitative Analysis 20, 243–260.

Keim, D.B., 1983. Size-related anomalies and stock return seasonality: further empirical evidence. Journal of Financial Economics 12, 13–32.

Kim, Y.G., Chung, K.H., Pyun, C.S., 1992. Size, price-earnings ratio, and seasonal anomalies in the Korean stock market. In: Rhee, S.G., Chang, R.P. (Eds.), Pacific-Basin Capital Markets Research, vol. III. Elsevier, pp. 303–319.

Knez, P.J., Ready, M.J., 1997. On the robustness of size and book-to-market in cross-sectional regressions. Journal of Finance 52, 1355–1382.

Lakonishok, J., Shleifer, A., Vishny, R.W., 1992. The structure and performance of the money management industry. Brookings Papers on Economic Activity 339, 391.

Lakonishok, J., Shleifer, A., Vishny, R.W., 1994. Contrarian investment, extrapolation, and risk. Journal of Finance 49, 1541–1578.

Lamoureux, C.G., Sanger, G.C., 1989. Firm size and turn-of-the-year effects in the OTC/Nasdaq market. Journal of Finance 44, 1219–1245.

Lettau, M., Ludvigson, S., 2001. Resurrecting the (C)CAPM: a cross-sectional test when risk premia are time-varying. Journal of Political Economy 109, 1238–1287.

Lewellen, J., Nagel, S., 2006. The conditional CAPM does not explain asset-pricing anomalies. Journal of Financial Economics 82, 289–314.

Lintner, J., 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics 47, 13–37.

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

Louvet, P., Taramasco, O., 1991. L'effet jour-de-la-semaine à la Bourse de Paris: Un effet transactionnel. Journal de la Société Statistique de Paris 2, 50–76.
Ma, T., Shaw, T.Y., 1990. The relationships between market value, P/E ratio, and trading volume and the stock return of Taiwan Stock Exchange. In: Rhee, S.G., Chang, R.P. (Eds.), Pacific-Basin Capital Markets Research, vol. I. Elsevier, pp. 313–335.
MacKinlay, A.C., 1995. Multifactor models do not explain deviations from the CAPM. Journal of Financial Economics 38, 3–28.
Malkiel, B.G., Xu, Y., 1997. Risk and return revisited. Journal of Portfolio Management 9, 14.
Malkiel, B.G., Xu, Y., 2004. Idiosyncratic Risk and Security Returns. Working Paper, Princeton University.
Merton, R.C., 1973. An intertemporal capital asset pricing model. Econometrica 41, 867–887.
Merton, R.C., 1987. A simple model of capital market equilibrium with incomplete information. Journal of Finance 42, 483–510.
Moller, N., Zinca, S., 2008. The evolution of the January effect. Journal of Banking and Finance 32, 447–457.
Pastor, L., Stambaugh, R.F., 2003. Liquidity risk and expected stock returns. Journal of Political Economy 111, 642–685.
Petkova, R., 2006. Do the Fama–French factors proxy for innovations in predictive variables? Journal of Finance 61, 581–612.
Pettengill, G.N., Sundaram, S., Mathur, I., 1995. The conditional relation between beta and returns. Journal of Financial and Quantitative Analysis 30, 101–116.
Pettengill, G.N., Sundaram, S., Mathur, I., 2002. Payment for risk: constant beta vs. dual-beta models. The Financial Review 37, 123–136.
Poterba, J.M., Weisbenner, S.J., 2001. Capital gains tax rules, tax-loss trading, and turn-of-the-year returns. Journal of Finance 61, 353–368.
Reinganum, M.R., 1981. Misspecification of asset pricing: empirical anomalies based on earnings' yields and market values. Journal of Financial Economics 9, 19–46.
Ritter, J.R., Chopra, N., 1989. Portfolio rebalancing and the turn-of-the-year effect. Journal of Finance 44, 149–165.
Roll, R., 1977. A critique of the asset pricing theory's test. Part I: On past and potential testability of the theory. Journal of Financial Economics 4, 129–176.
Roll, R., 1983. Vas ist das? The turn-of-the-year effect and the return premia of small firms. Journal of Portfolio Management 9, 18–28.
Roll, R., Ross, S., 1994. On the cross-sectional relation between expected returns and betas. Journal of Finance 49, 101–121.
Ross, S., 1976. The arbitrage theory of capital asset pricing. Journal of Economic Theory 13, 341–360.
Ross, S., 1977. The capital asset pricing model (CAPM), short-sale restrictions and related issues. Journal of Finance 32, 177–183.
Rouwenhorst, K.G., 1999. Local return factors and turnover in emerging stock markets. Journal of Finance 54, 1439–1464.
Rubio, G., 1988. Further international evidence on asset pricing: the case of the Spanish capital market. Journal of Banking and Finance 12, 221–242.
Santos, T., Veronesi, P., 2006. Labor income and predictable stock returns. Review of Financial Studies 19, 1–44.
Scholes, M., Williams, J., 1977. Estimating betas from nonsynchronous data. Journal of Financial Economics 5, 309–327.
Schultz, P., 1983. Transaction costs and the small firm effect: a comment. Journal of Financial Economics 12, 81–88.
Schwert, G.W., 2003. Anomalies and market efficiency. In: Constantinides, G.M., Harris, M., Stulz, R.M. (Eds.), Handbook of the Economics of Finance. Amsterdam, North Holland.
Sharpe, W.F., 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. Journal of Finance 19, 425–442.
Shumway, T., Warther, V.A., 1999. The delisting bias in CRSP's Nasdaq data and its implications for the size effect. Journal of Finance 54, 2361–2379.
Sias, R.W., Starks, L.T., 1997. Institutions and individuals at the turn-of-the-year. Journal of Finance 52, 1543–1562.
Stambaugh, R.F., 1982. On the exclusion of assets from tests of the two-parameter model: a sensitivity analysis. Journal of Financial Economics 10, 237–268.
Stehle, R., 1997. Der Size-Effekt am Deutschen Aktienmarkt. Zeitschrift für Bankrecht und Bankwirtschaft 9, 237–260.
Stoll, H.R., Whaley, R.E., 1983. Transaction costs and the small firm effect. Journal of Financial Economics 12, 57–79.
Sun, Q., Tong, W.H.S., 2010. Risk and the January effect. Journal of Banking and Finance 34, 965–974.
Thaler, R.H., 1987. Anomalies: the January effect. Journal of Economic Perspectives 1, 197–201.
Vassalou, M., Xing, Y., 2004. Default risk in equity returns. Journal of Finance 59, 831–868.
Wahlroos, B., Berglund, T., 1986. Anomalies and equilibrium returns in a small stock market. Journal of Business Research 14, 423–440.
Wang, X., 2000. Size effect, book-to-market effect, and survival. Journal of Multinational Financial Management 10, 257–273.
Wong, P.L., Neoh, S.K., Lee, K.H., 1990. Seasonality in the Malaysian stock market. Asia Pacific Journal of Management 7, 43–62.