

Size Matters, if You Control Your Junk

Asness, Clifford S.; Frazzini, Andrea; Israel, Ronen; Moskowitz, Tobias J.; Pedersen, Lasse Heje

Document Version

Final published version

Published in:

Journal of Financial Economics

DOI:

[10.1016/j.jfineco.2018.05.006](https://doi.org/10.1016/j.jfineco.2018.05.006)

Publication date:

2018

License

CC BY

Citation for published version (APA):

Asness, C. S., Frazzini, A., Israel, R., Moskowitz, T. J., & Pedersen, L. H. (2018). Size Matters, if You Control Your Junk. *Journal of Financial Economics*, 129(3), 479-509. <https://doi.org/10.1016/j.jfineco.2018.05.006>

[Link to publication in CBS Research Portal](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us (research.lib@cbs.dk) providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 04. Jul. 2025

Size matters, if you control your junk[☆]Clifford Asness^a, Andrea Frazzini^a, Ronen Israel^a, Tobias J. Moskowitz^{a,b,c,*}, Lasse H. Pedersen^{a,d,e,f}^aAQR Capital Management, LLC, 2 Greenwich Plaza, Greenwich, CT 06830, USA^bYale University, SOM, 165 Whitney Ave., New Haven, CT 06511, USA^cNational Bureau of Economic Research (NBER), MA, USA^dDepartment of Finance, Copenhagen Business School, A4.12, Solbjerg Plads 3, 2000 Frederiksberg, Denmark^eStern School of Business, New York University, 44 West Fourth Street, New York, NY 10012, USA^fCentre for Economic Policy Research (CEPR), London, UK

ARTICLE INFO

Article history:

Received 29 June 2016

Revised 9 February 2017

Accepted 23 March 2017

Available online 15 June 2018

JEL classification:

G11

G12

G14

G15

Keywords:

Size premium

Factor models

Quality

ABSTRACT

The size premium has been accused of having a weak historical record, being meager relative to other factors, varying significantly over time, weakening after its discovery, being concentrated among microcap stocks, residing predominantly in January, relying on price-based measures, and being weak internationally. We find, however, that these challenges disappear when controlling for the quality, or its inverse, junk, of a firm. A significant size premium emerges, which is stable through time, robust to specification, not concentrated in microcaps, more consistent across seasons, and evident for non-price-based measures of size, and these results hold in 30 different industries and 24 international equity markets. The resurrected size effect is on par with anomalies such as value and momentum in terms of economic significance and gives rise to new tests of, and challenges for, existing asset pricing theories.

© 2018 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY license.
(<http://creativecommons.org/licenses/by/4.0/>)

[☆] We thank David Blitz, Mark Carhart, Kent Daniel, Ken French, April Frieda, Jacques Friedman, Nicolae Garleanu, Will Goetzmann, Antti Ilmanen, Sarah Jiang, Don Keim, Bryan Kelly, Christine Kemper, John Liew, Robert Novy-Marx, Scott Richardson, Richard Roll, Laura Serban, Rodney Sullivan, Eric Wu, and seminar participants at Dartmouth College, the FTSE Russell World Investment Forum, the Jacobs Levy Center Conference, the Chicago Booth Asset Pricing Conference, the Q Group 2016 Spring Seminar, London Business School, London School of Economics, and the 2017 American Finance Association meetings (Chicago, Illinois) for helpful comments. We also thank Alex Bennett, Brandon Lee, Xiao Qiao, Kaushik Vasudevan, and Ryan Wei for outstanding research assistance. Moskowitz thanks the Center for Research in Security Prices for financial support when this project began. AQR Capital Management is a global investment manager who may or may not use the ideas in this paper. The views and opinions expressed are those of the authors and do not necessarily reflect the views of AQR Capital Management.

* Corresponding author at: Yale University, SOM, 165 Whitney Ave., New Haven, CT 06511, USA.

E-mail address: Tobias.Moskowitz@yale.edu (T.J. Moskowitz).

1. Introduction

Does size matter? With respect to capital markets, the answer to this question is unclear. Academic research on the relation between firm size and expected returns dates back to at least Banz (1981), who finds that small stocks in the US (those with lower market capitalizations) have higher average returns than large stocks, an effect not accounted for by the higher market beta of small stocks. The relation between firm size and expected returns is important for several reasons. First, the size effect has become a focal point for discussions about market efficiency. Second, a size factor has become one of the main building blocks of current asset pricing models used in the literature and in practice (e.g., Fama and French, 1993, 2016). Third, the size premium implies that small firms face larger costs of capital than large firms, having important implications for

corporate finance, incentives to merge and form conglomerates, and broader industry dynamics. Fourth, the size effect has had a large impact on investment practice (Reinganum, 1983a), including spawning an entire category of investment funds, giving rise to small cap indices, and serving as a cornerstone for money management classification.

We provide new evidence on the size effect and test several competing theories for its existence:

1. *Risk-based theories of frictionless capital markets.*
 - (a) Standard asset pricing models such as the capital asset pricing model (CAPM): If size per se is not a risk, standard models predict that size does not matter when controlling for risk exposures.
 - (b) Size captures time-varying risk premia: Size can be correlated to expected returns only because size is measured by market value, which is influenced by risk premia (Ball, 1978; Berk, 1995a). Riskier firms have higher required returns, leading to lower market value, everything else equal. Hence, any misspecification of risk premia, due perhaps to time-varying risk or risk premia, will be picked up by market prices. According to this theory, size factors based on market prices will mechanically pick up these movements, but size measures other than market value should not predict returns.
 - (c) Theories of growth options: If small firms have more growth options and growth options are risky, then small firms are riskier and have higher required returns (Carlson et al., 2004; Garleanu et al., 2012). Hence, the size effect should be smaller when controlling for measures of risk and growth options.
2. *Behavioral finance theories.* Small firms are more difficult to arbitrage, making these firms more mispriced (Shleifer and Vishny, 1997). If this effect drives the size effect, then small stocks need to be underpriced on average relative to large stocks, and the size effect should be smaller when controlling for measures of mispricing such as investor sentiment, disagreement, and limited arbitrage activity (e.g., trading costs or short-sale costs). Investors could overreact to growth and other quality measures; so according to these theories, the size effect should be smaller when we control for such effects (Barberis et al., 1998; Daniel et al., 1998; Hong and Stein, 1999).
3. *Theories of liquidity and liquidity risk.* Size matters because small firms are less liquid (Amihud and Mendelson, 1986) and face more liquidity risk (Acharya and Pedersen, 2005), both requiring higher expected returns. Hence, the size premium should be related to liquidity level and risk measures, and the size and liquidity premia should be more evident when controlling for other risk exposures, especially those negatively correlated with size.

To test these theories, we consider the size effect controlling for other factors, which can proxy for other sources of risk, growth, mispricing, and liquidity. We focus on the interaction between size and firm quality (or its inverse,

junk). The interaction between size and quality is especially interesting for three reasons. First, quality can be defined as a characteristic of an asset that, all else equal, commands a higher price. As such, size, which is based on market values, should have a strong connection to quality. Second, Novy-Marx (2013) shows that quality, as measured by profitability, has a strong connection to the value effect and helps clean up the relation between value and expected returns. A similar argument can be made for the size effect, where size's relation to average returns can be clearer once we account for quality. Because size also interacts with value (Fama and French, 1993; 2012), sorting out the interactions between these three predictors of returns could help better characterize the cross section of expected returns. Third, quality has a direct theoretical link to some of the theories we aim to test, namely, quality can be related to growth options underlying rational theories for size or sources of overreaction underlying prominent behavioral theories, or both. Measures of quality such as profitability, stability, and growth are intuitively empirical proxies for growth options and have been used as variables related to overreaction by investors.

Supporting these motivations for looking at the size-quality interaction, we find empirically that the key control variable for the size effect is firm quality (measured by profitability, stability, growth, and safety) or its inverse, junk. The interaction between size and quality or junk is far stronger than size's interaction with other factors (beta, value, momentum) and accounting for it produces a more significant size premium that helps test some of the competing theories for size.

Controlling for quality or junk (along with the standard asset pricing factors such as value and momentum) resurrects a strong size effect and helps distinguish among the competing theories. The resurrected strong size effect controlling for other factors can be viewed as a rejection of theory 1(a). We find that non-price-based size measures perform just as well as price-based ones, rejecting theory 1(b). The fact that controlling for quality, including growth, makes the size effect stronger, not weaker, seems inconsistent with the growth theory 1(c), if growth options are more prevalent among the high growth firms as theory predicts. We find that small stocks have higher shorting costs and more disagreement, which, according to behavioral theories, makes them more likely to be overvalued not undervalued, which is inconsistent with behavioral theories 2. Finally, the fact that size matters more when controlling for other factors is consistent with liquidity-based theories 3, in which controlling for these other factors, particularly quality, helps clean up the relation between size, liquidity, and average returns. However, this result does not offer any additional direct evidence in favor of a liquidity story.

Given the importance and prominence of the size effect as the first major challenge to the CAPM, it has naturally come under heavy and often critical scrutiny. Considering almost a century-long sample of US stocks and a broad sample of global stocks in 24 different markets, we confirm the main criticisms of the standard size effect. Consistent with risk-based theories 1, size has a weak

historical record in the US, especially after adjusting for market risk (e.g., the CAPM). It suffers from long periods of poor performance, particularly after its discovery in the 1980s.¹ It has an even weaker record internationally (Crain, 2011 and Bryan, 2014). What little size effect is present seems to be concentrated in difficult to invest in microcap stocks, whose returns occur only in January,² making it arguably fairly insignificant. Based on this evidence, a simple risk-based model such as the CAPM, in which size per se is not a risk factor, seems broadly consistent with the data.

However, we find that measures of size studied by the literature load strongly and consistently negatively on a large variety of what have come to be termed “quality factors.” At a broad level, quality is a characteristic, or set of characteristics, of a security that investors should in theory be, willing to pay a high price for, all else equal. A high-quality firm can be one that is well-managed and has strong economic and accounting performance, such as high profitability and stability of earnings, good growth prospects, and low risk.³

Regardless of the quality metric used, for metrics that vary substantially both qualitatively and in terms of measured correlation, we find a much stronger and more stable size effect when controlling for a firm's quality or its inverse, junk. Controlling for quality, using any of a broad set of measures, reconciles many of the empirical irregularities associated with the size premium that have been shown in the literature. In short, firm size is highly confounded with firm quality, which distorts the relation between size and expected returns. Large firms tend to be high-quality firms on any of the above dimensions or measures, while small

firms tend to be “junk” (i.e., have the opposite characteristics). Given that the literature shows that high-quality stocks, however defined, tend to outperform junk stocks, including when comparing stocks of similar size (Asness et al., 2014; Fama and French, 2016), this means that the basic size effect is fighting a strong quality effect. By going long small stocks and short large stocks, a size-based strategy is long a potential size premium but also short a quality premium, which both understates the actual size effect and introduces additional variation from another factor.

We show that controlling for quality, a significant and more robust size effect emerges. Small quality stocks significantly outperform large quality stocks, and small junk stocks outperform large junk stocks, but the basic size effect suffers from a size-quality composition effect. We show that controlling for quality does even more than simply raise the size premium. Accounting for the influence of quality also explains all of the many empirical irregularities and challenges researchers have identified with the size effect. Controlling for any number of a variety of quality measures with no notable failures (some used in the literature, others novel) not only resuscitates the overall size effect by more than doubling its Sharpe ratio, but also resurrects it in the 1980s and 1990s when it is otherwise conspicuously absent, restores a more linear relation between size and average returns (i.e., no longer concentrated among the tiniest firms), revives the returns to size outside of January while simultaneously diminishing the returns to size in January (making it more uniform across months), and uncovers a larger size effect in almost two dozen other international equity markets (where size has been notably weak). This stronger size effect controlling for other factors, including the market and other risk-based factors, is inconsistent with standard theories of asset pricing.

We next consider theory 1(b) that size matters only because it is measured by market capitalization, which contains market prices. The argument is that any misspecification in the asset pricing model, such as time-varying risk premia, is likely to show up in a cross-sectional relation between any market-based measure containing price and returns. Consistent with this argument, Berk (1995b, 1997) shows that using non-price-based measures of size does not yield a significant relation between size and average returns. Hence, no size effect could exist per se, but rather model misspecification may be showing up in price-based measures. To test this argument, we construct size factors in which stocks are sorted based on measures that do not include the stock price. We construct five new size factors based on book assets, sales, book equity, number of employees, and fixed assets (property, plant, and equipment, PP&E). For each of these size factors, we find a significant alpha when controlling for quality and other factors, rejecting theory 1(b).

Our evidence also does not support the theories of growth options, 1(c). Controlling for growth, which is included among our quality measures, and other risk factors makes the size effect stronger, not weaker. Although we cannot perfectly measure growth options and their risks, any mismeasurement of growth options likely has the op-

¹ See Dichev (1998), Chan et al. (2000), Horowitz et al. (2000), Gompers and Metrick (2001), Van Dijk (2013), Israel and Moskowitz (2013), Mclean and Pontiff (2016), and Chordia et al. (2014). Schwert (2003) suggests that the small firm anomaly disappeared shortly after the initial publication of the papers that discovered it and coincided with an explosion of small-cap-based funds and indices. Gompers and Metrick (2001) argue that institutional investors' continued demand for large stocks in the 1980s and 1990s increased the prices of large companies relative to small companies, which accounts for a large part of the size premium's disappearance over this period.

² Horowitz et al. (2000) find that removing stocks with less than \$5 million in market cap eliminates the small firm premium. Crain (2011) and Bryan (2014) find that the small stock effect is concentrated among the smallest 5% of firms. For more about the January effect, see Keim (1983), Reinganum (1983b), and Roll (1983). Gu (2003) and Easterday et al. (2009) also find that the January effect has declined over time, coinciding with the decline in the small firm premium. Van Dijk (2013) finds the same in a review of the size literature.

³ A variety of quality measures have been proposed in the literature, including profitability (Graham and Dodd, 1934; Novy-Marx, 2013), investment (Fama and French, 2015; Hou and Van Dijk, 2017), growth (Lakonishok et al., 1994; Mohanram, 2005), low asset growth (Cooper et al., 2008), low use of accruals (Sloan, 1996; Richardson et al., 2005), payout (Baker and Wurgler, 2002; Pontiff and Woodgate, 2008), low risk (Ang et al., 2006, 2009; Black et al., 1972; Frazzini and Pedersen, 2014), low leverage (George and Hwang, 2010; Penman et al., 2007), low credit risk (Altman, 1968; Ohlson, 1980; Campbell et al., 2008), and good governance (Gompers et al., 2003; Bebchuk et al., 2009; Core et al., 2006; Cremers and Nair, 2005; Giroud and Mueller, 2011; Johnson et al., 2009); see also Cremers and Ferrell (2014) and Larcker et al. (2015) for reviews. Asness et al. (2014) summarize various measures and dimensions of quality and construct a composite quality index based on a number of proposed variables. They find that higher quality is associated with higher average prices but also with higher expected returns.

posite theoretic effect on the size premium to be consistent with our results. Hence, our overall evidence on size appears to challenge many risk-based theories of frictionless capital markets.

We next turn to the behavioral theories, 2. For these theories to explain the size effect, small stocks would not simply have to be mispriced, but more importantly, small stocks would have to be *underpriced* in the sense of being too cheap relative to large stocks. A key behavioral finance prediction related to certain stocks being cheap versus expensive is limits to arbitrage, in particular, theories of short-selling constraints following Miller (1977). According to these theories, certain stocks are difficult to sell short, implying that these stock prices mostly reflect the opinions of optimists, leading such stocks to become overvalued. Hence, these theories suggest a negative instead of a positive size premium, where small stocks are more likely to be overvalued rather than undervalued, relative to large stocks. Further, this effect should be stronger when differences of opinion among investors are larger, which also tends to go the opposite way of the size effect.

To test the behavioral theories, we examine the levels of short selling, the costs of short selling, the degree of dispersion among analysts' forecasts of earnings, and the degree of mispricing of small cap stocks relative to large cap stocks. To explain the size effect with a limit of arbitrage argument, the costs of short selling should be lower for small cap stocks relative to large cap stocks. We find evidence to the contrary. That is, small stocks have larger shorting costs in each quality group. This finding, while not surprising, implies that a negative small cap expected return should be evident, the opposite of a small cap premium, according to behavioral theories. In addition, we find evidence to support greater differences of opinion among small cap stocks, which, again, should lead to lower returns, not higher returns, for small cap stocks. Finally, when looking at past five-year returns, we find the returns to small cap stocks to be low, not high, providing further evidence that small stocks do not seem to be overvalued as predicted by behavioral theories. When we control for variables that could be related to mispricing or risk premia, such as value, quality, and momentum, the size effect gets stronger, not weaker, which is also inconsistent with a behavioral story for the size premium.

Lastly, we consider the liquidity-based theories, 3. The idea is that size can just be a proxy for illiquidity and liquidity risk, and investors generally require compensation for holding illiquid securities facing the risk of worsening liquidity (here we refer to attempts to measure liquidity itself and a premium for illiquidity, which is more than simply noting that the size effect is largely in microcaps and more difficult to arbitrage). Consistent with these theories, the returns to size seem to be captured by measures of illiquidity suggested by Brennan and Subrahmanyam (1996), Amihud (2002), Hou and Moskowitz (2005), Sadka (2006), and Ibbotson et al. (2013), and measures of liquidity risk (the covariance with changes in liquidity) such as those of Pastor and Stambaugh (2003) and Acharya and Pedersen (2005). Crain (2011) summarizes this evidence.

We study the link between liquidity and size when controlling for quality. We find a difference in bid-ask spreads

between small and large stocks (as in the literature), but, more importantly, we find that this difference is similar across each quality quintile. In other words, small quality stocks are less liquid than large quality stocks and likewise for junk stocks. We also estimate that small stocks have greater market impact costs than large stocks and this difference is similar across quality quintiles. Hence, the return spread between small versus large stocks lines up with the corresponding liquidity spreads across small versus large stocks, after controlling for quality. However, controlling for size, little relation seems to exist between liquidity and quality measures. High-quality small stocks face similar liquidity to junky small stocks (among large stocks it is the same). This is consistent with liquidity-based theories for the size premium, in which size is also correlated with a quality factor that is unrelated to liquidity and so the size-liquidity relation could be partly obscured by quality. We find another sign of illiquidity, that the size factor loads on lagged market returns, consistent with non-synchronous trading for small, illiquid stocks, and that this lagged market exposure is the same whether we control for quality or not. Hence, size seems to be related to both illiquidity (positively) and quality (negatively), but liquidity and quality are not strongly related. Therefore, our results are consistent with the size premium being an illiquidity premium, though they do not offer any additional evidence for their connection.

The size factor is also highly correlated to factors that attempt to capture illiquidity premia more directly. When we regress size on the liquidity risk factor IML (illiquid minus liquid) from Amihud (2014), we find a highly significant loading (t -statistic above 30), consistent with the idea that the size premium is (at least partly) explained by compensation for liquidity risk. The size factor also loads on an alternative liquidity risk factor based on bid-ask spreads and turnover. Controlling for these liquidity risk factors naturally lowers the alpha of the size factor, even after controlling for quality, to a marginally positive effect (t -statistic of 2.03). Size loading on liquidity is similar whether or not we control for quality, indicating that liquidity risk helps explain the size premium, but liquidity risk is relatively unrelated to quality. A theory consistent with these facts is that stocks with higher liquidity risk have higher required returns and, separately, quality stocks have higher expected returns, but these return premia can be driven by different mechanisms. Our results show that quality helps resurrect a premium associated with small stocks, whether that premium is a size effect or a liquidity effect. Because small stocks tend to be junky, the standard size factor underestimates this premium, which is more clearly seen once we control for quality. If purely driven by liquidity risk, then size should have zero alpha when we control for both quality and liquidity risk. Hence, the marginally significant alpha suggests that there is more going on, our attempt to measure liquidity risk contains error, the result itself is just noise (the t -statistic barely passes conventional significance), or the size effect is compensation for the level of both liquidity and liquidity risk, when we capture only the latter. The model of Acharya and Pedersen (2005) suggests that small stocks have higher required return, everything else equal (e.g., for equal qual-

ity), because small stocks are both less liquid on average and face more liquidity risk in the sense that their liquidity deteriorates more when investors value liquidity the most.

Our results resurrect the size premium, putting it on a more equal footing with other anomalies such as value and momentum in terms of its efficacy. We motivate and evaluate our analysis in the context of theory, interpreting our results through the lens of rational, behavioral, and liquidity-based theories. We motivate why quality in particular is an interesting characteristic to interact with size in addressing these theories, but we also conclude that the interaction between size and quality presents a set of new stylized facts that seem to challenge standard frictionless asset pricing theories, are not easily explained by existing behavioral theories, and seem most consistent with theories of liquidity and liquidity risk.

The paper proceeds as follows. Section 2 describes the data and reviews the evidence on the size effect, highlighting the challenges to the size premium identified in the literature. As a first test of the competing theories, Section 3 shows that the size effect is resurrected when controlling for a firm's quality or junk. Section 4 tests further distinguishing predictions of the risk-based, behavioral, and liquidity-based theories. Section 5 concludes.

2. Data and preliminaries: reexamining the size effect

We detail the data used in this study and reexamine the evidence of the basic size effect by replicating some of the challenges identified in the literature using an updated sample.

2.1. Data

We examine long-short equity portfolios commonly used in the literature based on firm size. For US equities, we obtain stock returns and accounting data from the union of the Center for Research in Security Prices (CRSP) tapes and the Compustat North America database. Our US equity data include all available common stocks on the merged CRSP and Compustat data (sharecode 10 or 11) between July 1926 and December 2012, our longest historical sample. We include delisting returns when available in CRSP.

For size portfolios, we primarily use the Fama and French SMB (small minus big) factor and a set of value-weighted decile portfolios based on market capitalization sorts, obtained from Ken French's webpage (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). The decile portfolios are formed by ranking stocks every June by their market capitalization (price times shares outstanding) and forming deciles based on NYSE breakpoints, in which the value-weighted average return of each decile is computed monthly from July to June of the following year. The size factor, SMB, is the average return on three small portfolios minus the average return on three big portfolios formed by ranking stocks independently by their market cap and their book-to-market equity ratio (BE/ME) every June and forming two size

portfolios using the NYSE median size and three book-to-market portfolios using 30%, 40%, and 30% breakpoints, value, middle, and growth respectively. The intersection of these groups forms six size and BE/ME portfolios split by small and large (e.g., small value, small middle, small growth and large value, large middle, and large growth), in which SMB is the equal-weighted average of the three small portfolios minus the equal-weighted average of the three large portfolios.

In addition to SMB, the value factor, HML (high minus low), is formed from the equal-weighted average return of the two value portfolios minus the two growth portfolios, $HML = \frac{1}{2} (\text{small value} + \text{big value}) - \frac{1}{2} (\text{small growth} + \text{big growth})$. Fama and French (1993) also add the market factor, RMRF, which is the value-weighted index of all CRSP-listed securities minus the one-month Treasury bill rate.

Ken French's website also provides a momentum factor, which is a long-short portfolio constructed in a similar manner, in which six value-weighted portfolios formed on size and prior returns (the cumulative return from months $t-12$ to $t-2$) are used. The portfolios are the intersections of two portfolios formed on size and three portfolios formed on prior returns. The momentum factor, UMD (up minus down) is constructed as $UMD = \frac{1}{2} (\text{small up} + \text{big up}) - \frac{1}{2} (\text{small down} + \text{big down})$.

Ken French's website also provides a short-term reversal factor, STREV, which is formed in the same way as the momentum factor except using past returns from just the most recent month $t-1$ instead of $t-12$ to $t-2$, and with stocks sorted by the negative of their past one-month return.

We also form SMB and value-weighted size decile portfolios using non-price-based measures of size, as suggested by Berk (1995b, 1997), in lieu of a firm's market capitalization to rank stocks. Using the same methodology, we form five sets of non-price size portfolios based on book value of assets, book value of equity, sales, PP&E, and number of employees.

We also form SMB portfolios within each of 30 industries used by Fama and French (1997) and available on Ken French's website. We construct SMB in a similar fashion within each industry so that we obtain 30 SMB industry-neutral portfolios.

For shorting costs and levels of shorting, we use data from Markit Data Explorers, which provides data on more than \$15 trillion of global securities in the lending programs of more than 20 thousand institutional funds. We focus on the supply of lendable shares divided by the market cap for each stock, expressed as a percentage, and the daily cost of borrow score (DCBS), a number from one to ten indicating the fee for each stock. Our sample uses data from January 2010 to October 2016, when coverage is highest for our universe of stocks.

We use a variety of quality or junk measures proposed in the literature that pertain to different notions of quality: profitability, growth, safety, payout, credit, and investment. We use some alone and some in composites. We use as different determinants of quality the profitability and investment measures of Fama and French (2016); the profitability, growth, safety, and payout measures of Asness et al. (2014) as well as their composite index of

quality, which is an average of these measures; a revised composite index of Asness et al. (2017) that excludes payout and drops accruals from the growth composite; a Safe-Minus-Risky (SMR) factor based off of the betting-against-beta (BAB) factor from Frazzini and Pedersen (2014), where for simplicity we will refer to this factor throughout the paper as simply BAB;⁴ and credit ratings of corporate debt.⁵ For each of these measures, we form a portfolio that is long high quality and short low quality (or junk), formed in a manner similar to the methodology used by Fama and French (1993), in which stocks are ranked by size and quality measures independently into two size and three quality groups and the intersection of the groups forms six portfolios, with the resulting portfolio equally long the two quality portfolios and short the two junk portfolios. The only exception to these factors being constructed as long the top third and short the bottom third (neutral to a size split) is the credit portfolio, CRED, which goes long the equity of firms with A-rated or better debt and short the equity of firms with C-rated or lower debt. The results are robust across all of these measures.

With international data, we form all of the described portfolios and factors in each of 23 other developed equity markets following a similar methodology. Our international equity data include all available common stocks on the Compustat Global database for 23 developed markets from January 1983 to December 2012.⁶ The 23 markets correspond to the union of all countries belonging to the MSCI World Developed Index as of December 2012. We assign individual stocks to the corresponding market based on the location of the primary exchange. For companies traded in multiple markets we use the primary trading vehicle identified by Compustat.

Global portfolio construction closely follows Fama and French (2012), Asness and Frazzini (2012), and Asness et al. (2013). The portfolios are country neutral in the sense that we form long-short portfolios within each country and then compute a global factor by weighting each country's long-short portfolio by the country's total (lagged) market capitalization. The global market factor, RMRF, is the value-weighted return of all available stocks across all markets minus the one-month US Treasury bill rate. The size and value factors are constructed using six value-weighted portfolios formed on size and book-to-market sorts just like in the US. However, while for the US the size breakpoint is the median NYSE market equity, for the international sample the size breakpoint is the 80th percentile by country to roughly match the US size portfolios. Because some countries have a small cross section of stocks in the early years of our sample, we use conditional sorts that

first sort on size and then on book-to-market to ensure we have enough securities in each.

Portfolios are value weighted and rebalanced every calendar month. We require a firm, to be included in any of our tests, to have a non-negative book value and non-missing price at fiscal year-end as well as in June of calendar year t .⁷ All portfolio returns are in US dollars and excess returns are relative to the one-month US Treasury bill rate.

2.2. Reexamining the evidence on the size effect

Table 1 replicates the evidence on the size effect from the literature, including data outside of the sample periods from the original studies. Columns 1–3 report results for SMB; columns 4–6 report the difference in returns between deciles 1 and 10 (a more extreme difference in size than SMB and also unadjusted through bivariate sorts for book-to-price). Table 1 reports the mean, standard deviation, and t -statistic of the size premium over the longest historical sample period from July 1926 to December 2012. SMB yields a premium of 23 basis points (bps) per month that is statistically significant at the 5% level (t -statistic = 2.27). The decile spread returns also yield a positive return of 55bps per month, with a t -statistic of 2.32. This result highlights that the size effect is significant but relatively weak compared with other anomalies such as value and momentum, whose t -statistics over the same period are 3.7 and 4.6, respectively (using the Fama and French factors HML and UMD), indicating stronger and more reliable return premia.⁸ The Sharpe ratio generated for size-based portfolios is roughly 60% of the Sharpe ratio for value portfolios and half the Sharpe ratio that is generated from momentum (all without an attempt to adjust for other risk factors, including market beta).

The table also separates the returns to size into the month of January versus the months of February through December. The returns to SMB are enormous in January at 2.3% for the month and the 1–10 spread in size decile returns is even larger at 6.8% in January. However, from February through December, SMB delivers a negligible 4bps and the 1–10 portfolio spread is –1bp, both of which

⁴ Our SMR factor is a slightly different version of the BAB strategy from Frazzini and Pedersen (2014) constructed following the methodology of Fama and French (1993). For consistency with the other portfolios used in this paper, we rank stocks based on their ex-ante beta and form a portfolio that is long low beta and short high beta stocks based on the intersection of six size- and beta-sorted portfolios.

⁵ The Asness, Frazzini, and Pedersen composite index can be downloaded at <https://www.aqr.com/library/data-sets/quality-minus-junk-factors-monthly>.

⁶ Common stocks are identified by a Compustat issue code (TPCI) of 0. We also drop stocks traded on over-the-counter (OTC) exchanges.

⁷ To obtain shareholders' equity we use stockholders' equity (SEQ). If it is not available, we use the sum of common equity (CEQ) and preferred stock (PSTK). If both SEQ and CEQ are unavailable, we proxy shareholders' equity by total assets (TA) minus the sum of total liabilities (LT) and minority interest (MIB). To obtain book equity, we subtract from shareholders' equity the preferred stock value (PSTKR, PSTKL, or PSTK depending on availability). Finally, to compute book value per share (B), we divide by common shares outstanding (CSHPRI). If CSHPRI is missing, we compute company-level total shares outstanding by summing issue-level shares (CSHOI) at fiscal year-end for securities with an earnings participation flag in the security pricing file.

⁸ Harvey et al. (2014) note that t -statistics greater than 3.0 are likely required to pass the 5% significance test in the presence of the data mining that has taken place by researchers pouring over the same return series, but we note two caveats. On the one hand, the size effect is simple and was discovered several decades ago so it could be less subject to data mining. On the other hand, our resurrection of size is based on the interaction with quality and, as Harvey et al. (2014) point out, conditional strategies could have different, potentially higher, statistical cutoffs to judge significance in light of potential data mining.

Table 1

The size effect across sample periods and markets.

The table reports summary statistics on the size premium over time. Two zero-cost portfolios are used to capture the returns to size: the small minus big (SMB) stock factor of Fama and French (1993), obtained from Ken French's website (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html), and the return spread between size-sorted value-weighted decile portfolios. The annualized mean and standard deviation of the returns are reported on these spread portfolios, as well as the *t*-statistic of the mean, over the longest historical sample period (from July 1926 to December 2012), for January and February–December separately over the longest sample period, for the same sample period as the Banz (1981) study (January 1936 to December 1975), over the period before and after the Banz (1981) study treated as one period ignoring the discontinuity, over the period when the betting-against-beta (BAB) strategy of Frazzini and Pedersen (2014) is available (January 1931 to December 2012), when the Fama and French (2016) new five-factor model is available (July 1963 to December 2012), when returns to credit portfolios are available (July 1987 to December 2012), when the quality variables of Asness et al. (2014) are available (July 1957 to December 2012), and over three ex post selected subperiods when the size effect is strongest (July 1957 to December 1979, golden age), weakest (January 1980 to December 1999, embarrassment), and recently resurged (January 2000 to December 2012, resurrection). Also reported are SMB returns internationally across 23 other markets, including globally (excluding the US), Europe, North America, and the Pacific. The measure of a stock's size is its market capitalization (share price times shares outstanding) from June of the previous year.

Sample Period	Years	SMB			1 - 10 decile spread		
		Mean	Standard deviation	<i>t</i> -statistic	Mean	Standard deviation	<i>t</i> -statistic
Longest sample	1926 - 2012	0.23%	3.26%	2.27	0.55%	7.69%	2.32
January		2.30%	3.26%	6.50	6.83%	8.41%	7.49
February–December		0.04%	3.19%	0.41	−0.01%	7.37%	−0.06
Banz (1981)	1936 - 1975	0.16%	2.83%	1.22	0.61%	7.34%	1.82
Pre- and Post-Banz (1981)	1926 - 1935; 1976 - 2012	0.29%	3.59%	1.92	0.50%	7.99%	1.49
BAB sample	1931 - 2012	0.29%	3.28%	2.78	0.67%	7.74%	2.73
Fama and French five-factor sample	1963 - 2012	0.25%	3.13%	1.95	0.33%	4.89%	1.66
Credit sample	1987 - 2012	0.14%	3.31%	0.74	0.16%	4.89%	0.56
Quality sample	1957 - 2012	0.22%	3.01%	1.93	0.33%	4.72%	1.80
Golden age	1957 - 1979	0.35%	2.87%	2.00	0.68%	4.80%	2.35
Embarrassment	1980 - 1999	−0.04%	2.66%	−0.23	−0.40%	4.12%	−1.49
Resurrection	2000 - 2012	0.42%	3.67%	1.41	0.82%	5.31%	1.92
Global ex US	1986 - 2012	0.13%	2.50%	0.91			
Europe	1991 - 2012	−0.08%	2.50%	−0.60			
North America	1986 - 2012	0.09%	2.02%	0.77			
Pacific	1992 - 2012	−0.31%	3.38%	−1.63			

are statistically and economically zero. Hence, what reliable positive premium exists for size appears to solely reside in January and is completely absent the rest of the year. This result illustrates one of the biggest challenges to the interpretation of the size effect, i.e., all of the returns to size come from small stocks in January.

Table 1 reports results over the original sample period studied by Banz (1981) from 1936 to 1975 as well as the out-of-sample period from the original Banz (1981) study: 1926–1935 and 1976–2012. As the table indicates, SMB is insignificant over Banz's original sample period and the 1–10 decile spread is marginally significant (*t*-statistic of 1.82), though the mean returns are similar to the full period results. The results from Banz (1981) over the same time period for similar decile portfolios are stronger than what we find here, which is likely due to data errors being fixed by CRSP after publication of Banz (1981).⁹ The out-of-sample evidence from Banz (1981) is a bit stronger for SMB but weaker for the decile spread returns. Overall, the original size effect is weaker than originally found, consistent with Israel and Moskowitz (2013).

The basic size effect has also experienced significant variation over time, including over relatively long peri-

ods (a possible consequence of having a low Sharpe ratio). Table 1 reports summary statistics over the periods for which quality measures are available. The Frazzini and Pedersen (2014) BAB measure is available beginning in January 1931, the Fama and French (2016) profitability and investment factors are available from July 1963, credit from July 1987, and the quality measures used by Asness et al. (2014) from July 1957. We refer to the period from July 1957 to December 2012 as the “quality sample,” and we break this period up into three subperiods: (1) from July 1957 to December 1979, shortly before the discovery and publication of the size effect, which we term the “golden age” because the late 1970s was when most researchers were looking at the size effect, which happened to coincide with when its performance was highest,¹⁰ (2) from January 1980 to December 1999, which we call the “embarrassment” because this is when the size effect appears to have vanished, promptly after being discovered and published, and (3) from January 2000 to December 2012, which we term the “resurrection” as the size effect appears to be revitalized during this period. Table 1 highlights these patterns as the size effect seems to have disappeared in the 1980s and 1990s following its discovery but also appears to have made a comeback in the last 13 years of the sample.

⁹ Over time, CRSP has fixed many data errors, which are more common among the smallest firms, and these could have contributed positively to the returns of size. One such error was a delisting bias as noted by Shumway (1997), who shows that many studies focusing on small stocks had inflated returns due to mistreatment of the delisting returns to these stocks.

¹⁰ McLean and Pontiff (2016) claim that many anomalies provide their best returns over the sample period in which they were originally discovered.

Finally, [Table 1](#) reports results for the size premium outside of the US. We report results for an SMB portfolio, constructed in the same manner as for the US, in each of 23 other equity markets, with the data availability ranging from 1986 to 2012. We compute only SMB returns because forming decile portfolios in some of these markets results in too few stocks in some of the deciles. We then average the SMB returns across countries globally excluding the US, and in Europe, North America, and the Pacific, separately, with countries weighted by their lagged total market capitalization. The average returns to SMB outside of the US are weaker, averaging 13bps with an insignificant t -statistic of 0.91. Over the same time period from 1986 to 2012 SMB in the US also averaged only 12bps with a t -statistic of 0.70. Hence, the weak international results could be due to the sample period.

To further show the size premium, the first row of Panel A of [Table 2](#) reports time series regression results of SMB on the market portfolio, RMRF, over the July 1957 to December 2012 time period. The intercept or alpha from the regression is 12bps per month with a t -statistic of 1.12, which is insignificantly different from zero, suggesting that the CAPM explains much of the modest returns to size that existed in [Table 1](#). Next, the lagged return on the market from the previous month is added to capture delayed price responses of stocks, particularly small stocks, to marketwide news [following the results and implications of [Lo and MacKinlay \(1988\)](#) and [Hou and Moskowitz \(2005\)](#) and in the spirit of [Asness et al. \(2001\)](#) to account for non-synchronous price responses due to liquidity differences and lead-lag effects]. SMB has a significantly positive coefficient on the lagged market return, which further pushes its alpha down to an insignificant 7bps. Next reported are results that add HML and UMD to capture value and momentum exposure. The alpha is 14bps with a t -statistic of 1.23. In the presence of the market and other factors (value and momentum), no reliable size premium is evident.

Overall, a weak size effect exists, with substantial variation over time and across season, and meager evidence outside of the US.

3. Resurrecting size by controlling for junk

In this section, we show that accounting for the quality or junk of the stock helps to resurrect the size effect. The strong returns to size when controlling for junk present a challenge to the asset pricing theories.

3.1. The size effect controlling for quality: regression analysis

We first consider the magnitude and significance of the size effect in a regression setting. [Table 2](#) shows the results from adding a quality factor to the regression of size on standard factors in the literature. In this case, we add the [Asness et al. \(2014\)](#) quality minus junk (QMJ) factor, which is a long-short portfolio created from a composite measure of quality, that is long quality stocks and short junk stocks. SMB loads significantly negatively on quality, driving the SMB alpha from 14 to 49bps per month, which is almost

five standard errors from zero (t -statistic = 4.89). The addition of a quality factor to the regression not only raises significantly the average return to size but also increases the precision of the SMB premium, as quality explains a substantial fraction of the variation in SMB's returns, evidenced by the R -squared rising from 15 to 37%.

We show that various other factors or portfolios formed from other measures of quality or junk give similar results on resurrecting the size effect. The QMJ factor constructed by [Asness et al. \(2014\)](#) combines many measures designed to capture quality or junk by looking at variables that proxy for a variety of attributes. We take each component separately (profitability, growth, safety, and payout) as a quality factor and repeat the regression for SMB using each subcomponent. Despite the different measures, in each case the loading on quality, no matter how defined, is significantly negative and the SMB alpha is significantly positive (and more reliable). Using profitability to define quality, the SMB alpha jumps to 42bps, which is almost four standard errors from zero. Controlling for safety or payout as measures of quality yields very similar numbers (35 and 44bps alphas). The weakest, based on realized average return, quality measure is growth, yet even here there is a marginally significant 20bps size premium after adjusting for growth, and SMB loads significantly negatively on this measure of quality, too. [Table 2](#) then uses the latest version of the [Asness et al. \(2017\)](#) quality composite, which excludes payout and drops accruals from the growth composite. As the table shows, the results are unchanged to various perturbations of the quality factor.

Panel A of [Table 2](#) switches to using the two additional [Fama and French \(2016\)](#) factors from their five-factor model as quality proxies: RMW (robust minus weak) profitability factor and CMA (conservative minus aggressive) investment factor, which have been suggested as measures of firm quality ([Novy-Marx, 2013](#); [Fama and French, 2016](#)). These two factors are similar to subcomponents of QMJ, but with different specific formations and different creators. Thus, they are not an independent test but a robustness check. Intuitively, both profitability and investment are characteristics that should differ among high versus low quality firms. [Fama and French \(2016\)](#) offer three separate versions of their factors from sorting on combinations of size and profitability and investment. We show the results for the 2×3 versions of their factors from Kenneth French's website, which are nearly identical to using their 2×2 and $2 \times 2 \times 2 \times 2$ factor specifications. The factor returns are available from July 1963, so we first report the regression of SMB on the market, its lag, HML, and UMD over this period for reference. The SMB alpha is an insignificant 16bps (t -statistic = 1.31). Adding the [Fama and French \(2016\)](#) profitability and investment factors, SMB loads significantly negatively on both, which doubles the SMB alpha to a significant 33bps per month (t -statistic of 2.82). Hence, using [Fama and French's \(2016\)](#) two new factors as measures of quality also resurrects the size effect.

We use, as another measure of quality, a further robustness test, and an additional out-of-sample test, the stand-alone BAB factor from [Frazzini and Pedersen \(2014\)](#), which

Table 2

Size premium controlling for junk.

The table reports regression results for the size premium (small minus big, SMB) on the Fama and French factors that include the market (RMRF), its lagged return, high minus low (HML), and up minus down (UMD) and controlling for various measures of quality or junk. Panel A reports results from regressions that add various measures of quality or junk: the quality composite factor (QMJ) from [Asness et al. \(2014, 2017\)](#) and their four dimensions of quality related to profitability, growth, safety, and payout; the [Fama and French \(2016\)](#) five-factor model that includes the factors RMW and CMA, representing profitability and investment, respectively, the [Frazzini and Pedersen \(2014\)](#) betting-against-beta (BAB) factor, which is long low beta stocks and short high beta stocks, and the equity return difference between firms with A-rated debt and higher and firms with C-rated debt and lower (CRED). The sample period for QMJ and its components is July 1957 to December 2012; for the [Fama and French \(2016\)](#) factors July 1963 to December 2012; for the [Frazzini and Pedersen \(2014\)](#) BAB factors January 1931 to December 2012; and for the credit sample July 1987 to December 2012. Panel B reports multivariate regressions of quality or junk, and Panel C reports results using QMJ over the subperiods of the golden age (July 1957 to December 1979), embarrassment (January 1980 to December 1999), and resurrection (January 2000 to December 2012).

Panel A: Controlling for quality/junk																	
$SMB_t = \alpha + \beta RMRF_t + \beta_{-1}RMRF_{t-1} + hHML_t + mUMD_t + qQ_t^* + \varepsilon_t$																	
Sample	α	t(α)	β	t(β)	β_{-1}	t(β_{-1})	h	t(h)	m	t(m)	q	t(q)	RMW	t(r)	CMA	t(c)	R ²
Quality sample	0.0012	1.12	0.21	8.30													0.09
Quality sample	0.0007	0.63	0.20	7.96	0.13	5.09											0.13
Quality sample	0.0014	1.23	0.17	6.36	0.13	5.42	−0.16	−3.96	0.00	0.13							0.15
Q* = QMJ (2014)	0.0049	4.89	−0.04	−1.42	0.10	4.82	−0.24	−6.75	0.06	2.70	−0.74	−15.09					0.37
Q* = Profit	0.0042	3.95	0.06	2.36	0.11	5.07	−0.33	−8.04	0.03	1.24	−0.67	−10.98					0.28
Q* = Growth	0.0020	1.80	0.17	6.57	0.13	5.50	−0.27	−5.39	0.01	0.27	−0.26	−3.68					0.17
Q* = Safety	0.0035	3.53	−0.03	−1.12	0.10	4.82	0.20	4.61	0.05	1.98	−0.87	−14.94					0.36
Q* = Payout	0.0044	4.60	−0.12	−4.28	0.09	4.35	−0.28	−7.93	0.08	3.63	−0.70	−16.86					0.40
Q* = QMJ (2017)	0.0051	4.88	−0.03	−0.97	0.11	5.19	−0.39	−9.82	0.05	2.25	−0.74	−13.76					0.34
Fama and French sample	0.0016	1.31	0.17	6.13	0.14	5.33	−0.17	−3.87	0.01	0.52							0.16
Q* = RMW, CMA	0.0033	2.82	0.11	4.04	0.14	5.63	−0.09	−1.52	0.04	1.57			−0.54	−9.74	−0.15	−1.81	0.28
BAB sample	0.0007	0.72	0.19	10.09	0.13	7.54	0.03	1.09	−0.01	−0.28							0.17
Q* = BAB	0.0023	2.50	−0.13	−4.77	0.14	8.85	0.01	0.24	0.07	3.39	−0.42	−14.85					0.33
Credit sample	0.0005	0.27	0.11	2.77	0.13	3.39	−0.31	−5.23	0.04	1.15							0.17
Q* = Cred	0.0035	2.12	0.04	1.13	0.08	2.10	−0.28	−5.02	0.07	2.15	−0.12	−7.82					0.31

Panel B: Multiple measures of quality/junk																					
$SMB_t = \alpha + \beta RMRF_t + \beta_{-1}RMRF_{t-1} + hHML_t + mUMD_t + rRMW_t + cCMA_t + qQMJ_t + bBAB_t + dCred_t + \varepsilon_t$																					
Sample	α	t(α)	β	t(β)	β_{-1}	t(β_{-1})	h	t(h)	m	t(m)	r	t(r)	c	t(c)	q	t(q)	b	t(b)	d	t(d)	R ²
Fama and French sample	0.0047	4.36	−0.16	−4.69	0.10	4.62	−0.18	−3.06	0.11	4.29	0.08	0.96	0.09	1.15	−0.64	−6.64	−0.24	−5.61			0.41
Credit sample	0.0047	3.12	−0.28	−5.39	0.04	1.25	−0.17	−2.09	0.18	5.46	0.00	0.02	0.12	1.14	−0.43	−3.00	−0.30	−5.36	−0.06	−3.81	0.50

Panel C: Controlling for quality/junk over time													
$SMB_t = \alpha + \beta RMRF_t + \beta_{-1}RMRF_{t-1} + hHML_t + mUMD_t + qQ_t^* + \varepsilon_t$													
Subperiod	α	t(α)	β	t(β)	β_{-1}	t(β_{-1})	h	t(h)	m	t(m)	q	t(q)	R ²
Q* = QMJ													
Golden age	0.0025	1.52	0.27	7.19	0.15	4.10	0.07	0.95	−0.09	−1.83			0.24
	0.0057	4.00	0.07	1.96	0.14	4.70	−0.24	−3.73	−0.06	−1.39	−0.97	−10.73	0.48
Embarrassment	−0.0011	−0.64	0.04	0.97	0.18	5.05	−0.24	−3.56	−0.08	−1.63			0.18
	0.0050	3.06	−0.14	−3.43	0.15	4.85	−0.42	−6.84	−0.06	−1.34	−0.83	−9.08	0.40
Resurrection	0.0054	2.06	0.25	4.25	0.10	1.75	−0.34	−4.46	0.14	3.00			0.25
	0.0089	4.04	−0.17	−2.43	−0.03	−0.59	−0.18	−2.68	0.17	4.43	−0.84	−8.40	0.49

is available from January 1931. While BAB is part of the QMJ composite, it is the only variable available earlier than our quality period start in 1957. That, and the motivation for BAB, coming from Black (1972, 1992), makes it an interesting factor to highlight alone. The BAB factor is long low beta or safe stocks and short high beta risky stocks and hence can be viewed as a quality measure. Panel A of Table 2 further reports regressions of SMB on the market, its lag, HML, and UMD with and without the BAB factor over the 1931–2012 sample period. Absent the BAB factor, the SMB alpha is only 7 bps with a *t*-statistic of 0.72 from 1931 to 2012, indicating no size premium. However, adding just BAB as a single quality factor to the regression bumps up the SMB alpha to a significant 23 bps (*t*-statistic = 2.50). SMB loads significantly negatively on BAB (coefficient of -0.42 with a *t*-statistic of -14.85), indicating that even this very simple measure of quality is strongly and reliably negatively related to size and resurrects the otherwise absent size premium.¹¹

The last two rows of Panel A use another measure of quality that is novel to the literature: the equity returns between firms with A-rated or higher debt minus the equity returns of firms with C-rated or below debt, in which the market capitalization-weighted average of returns is computed for each group. This factor, CRED, captures the equity return difference between firms with high credit-worthy debt minus low-rated debt. Because credit ratings are available for enough firms only beginning in July 1987, the sample period is limited. Consistent with other measures of quality, CRED has positive average returns over the sample period, exhibiting a 0.63 annual Sharpe ratio, which is consistent with the performance of other quality portfolios over this period. The correlation between CRED and QMJ is 0.53. As the table shows, even over this very short time period, this novel measure of quality resurrects the size effect, too. The SMB alpha over this period without controlling for quality is an insignificant 5 bps (*t*-statistic of 0.27). But, controlling for quality using CRED raises the SMB alpha to a significant 35 bps (*t*-statistic of 2.12), and there is a strong negative loading of SMB on this unique quality factor (-0.12 coefficient with a *t*-statistic of -7.82).

Panel B of Table 2 examines multiple measures of quality simultaneously by running a regression of SMB on all of the quality factors. It reports results for the Fama and French (2016) RMW and CMA, Asness et al. (2014) QMJ, and Frazzini and Pedersen (2014) BAB factors simultaneously over the common period July 1963 to December 2012. The negative coefficients on RMW and CMA are soaked up by the very strong negative loadings on the QMJ composite measure and the BAB factor, indicating that they pick up the information in the Fama and French (2016) profitability and investment factors. This regression

is then repeated, adding the credit factor, CRED, to the list of quality factors over the common, shorter sample period July 1987 to December 2012. Even over this shorter time, the Fama and French (2016) profitability and investment factors are still subsumed by the other factors, but QMJ, BAB, and CRED all have significant negative loadings with respect to SMB, suggesting that each captures different aspects of quality that are consistently inversely related to size.

Table OA1 in the Online Appendix reports the correlation matrix of the Fama and French factors, UMD, and the various quality portfolios. SMB is consistently negatively correlated with every quality factor, ranging from -0.18 (CMA) to -0.54 (QMJ). In addition, the quality factors themselves are generally positively correlated with each other, though they seem to capture distinct aspects of quality. Given that quality and size are strongly negatively correlated, we examine the correlation among the different quality factors after hedging out size, reported at the bottom of Table OA1. The correlations are still largely positive, albeit smaller after removing the common size component from returns.

Overall, the results indicate that all of the measures of quality are negatively related to size and are helpful in resurrecting the size premium, even over different sample periods, with the results not particularly sensitive to any particular measure of quality.

Another way to see the size and quality interaction is to look at optimal portfolio weights from a Sharpe ratio maximizing portfolio. Table OA2 in the Online Appendix reports in-sample optimal portfolio weights of SMB with and without a quality factor (using QMJ) in the investment opportunity set and does the same for quality (using QMJ) with and without SMB in the opportunity set. The ex post optimal portfolio weights are a reflection of the regression alphas from before. The optimal portfolios' in-sample Sharpe ratio and correlation to SMB are also reported. Both SMB and QMJ by themselves improve the efficient frontier in the presence of the market, but as the table shows, an optimizer wants to put positive weight on both to maximize Sharpe ratio. The analysis is then repeated, adding the value and momentum factors HML and UMD. Here, both SMB and QMJ again receive positive weight, but are given even larger weight each in the presence of the other. An optimal portfolio wants more SMB when QMJ is also present and vice versa. With all factors present, simply adding SMB to the Fama and French factors only slightly increases the Sharpe ratio from an ex post 1.08 to 1.11. But, adding SMB to the Fama and French factors plus QMJ increases the Sharpe ratio from 1.36 to 1.57. Thus, the marginal impact on the optimal portfolio from adding SMB is small without QMJ present, but it is substantially larger in the presence of QMJ. Essentially, choosing the weight on SMB is not very helpful unless also being exposed to quality. Absent quality, tilting toward SMB means also tilting strongly against quality, with the two effects almost canceling each other out. In all, this evidence suggests that both size and quality are valuable factors not spanned by each other or a rotation of the other factors.

Fig. 1 shows the impact of controlling for quality or junk on the size effect by examining SMB hedged with re-

¹¹ For the period from January 1931 to June 1957, which is completely out of sample from the period when all of the QMJ variables are measured, we find that BAB increases the size alpha from 3 to 16 bps, though the alpha is statistically insignificant given the smaller sample size. The coefficient on BAB is -0.35 with a *t*-statistic of -4.99 , indicating that size and BAB are strongly related in the out-of-sample period as well and BAB helps resurrect the size premium as a result.

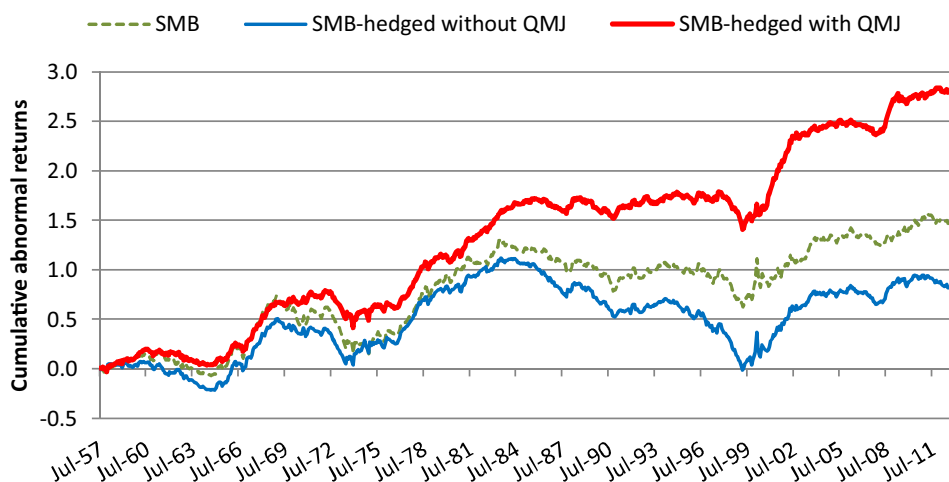


Fig. 1. Cumulative abnormal returns of small minus big (SMB) with and without controlling for junk. The figure plots the cumulative sum of returns over time of SMB; SMB hedged with the market, its lagged value, high minus low (HML), and up minus down (UMD); and SMB hedged with the market, its lagged value, HML, UMD, and the quality factor QMJ. The cumulative returns use the sample estimates of the betas on all factors from the entire period.

spect to the market, its lagged value, and HML and UMD factors and using the composite QMJ factor. The figure plots the cumulative sum of returns over time of SMB unhedged, SMB hedged with the market, its lagged value, HML, and UMD, and SMB hedged with all of those factors plus QMJ. The plot uses the full sample estimates of the betas from July 1957 to December 2012 to estimate the hedged returns to SMB. As Fig. 1 shows, hedging SMB for exposure to the market, value, and momentum factors reduces its returns, but hedging SMB with respect to quality or junk significantly improves returns.

Fig. 2 reports results across 30 different industries. We form SMB portfolios (long the smallest half of firms and short the largest half of firms) within each of 30 industries available from Ken French's data library. We then examine whether the improvement in SMB after controlling for quality or junk is similar within each industry. Though not 30 completely independent tests, this provides 30 different samples of firms from which we can test the robustness of the results.

We compute the alpha of SMB within each industry relative to the market, its lagged value, HML and UMD. We then repeat this computation using the same factors plus QMJ and compare the difference in alphas within each industry. Panel A of Fig. 2 shows the improvement in SMB alpha after controlling for QMJ for each of the 30 industries. The results are remarkably consistent. Every single industry shows positive improvement in SMB's returns after controlling for quality or junk, and for most industries the improvement is significant (with significance harder to achieve in a much smaller sample of firms within a single industry).

Panel B plots the betas of each SMB portfolio on QMJ, which are all negative and are the mirror image of the improvement in alphas in Panel A. These results indicate that the relation between size and quality or junk is robust. Not a single industry fails to find a strong negative relation be-

tween size and quality, and as a result, the size premium, controlling for quality, is consistently stronger within every single industry. Small stocks are, on average, junky stocks and this holds for every industry. Controlling for quality improves the size effect within every industry.¹²

Quality, however it is measured and across a variety of specifications, simultaneously resurrects the return premium to size as well as explains much of its variation, transforming it from a small and insignificant effect to an economically and statistically large one and doing so consistently.

3.2. Variation in the size premium over time

Fig. 1 anticipates the main result in this subsection as casual perusal of the plots in the figure shows a far more consistent size premium when hedged for quality exposure. More formally, we test the stability of the size premium with and without controlling for quality in Panel C of Table 2 by rerunning the regressions of SMB on the market, its lag, HML, and UMD with and without a quality factor over the three subsample periods (golden age, embarrassment, and resurrection) corresponding to the periods over which the unconditional basic size premium varies substantially (and chosen ex post precisely for this purpose). The golden age from July 1957 to December 1979

¹² To summarize these results, we also create an industry-neutral size factor by averaging across all of the industry SMB portfolios (equal-weighted across industries) and regress the returns on the market, lagged market, and industry-neutral versions of the HML, UMD, and QMJ factors all constructed in the same manner. Regressing SMB industry-neutral on these other industry-neutral factors, excluding QMJ, we find an insignificant alpha of 1 bp with a *t*-statistic of 0.13. Adding the QMJ industry-neutral factor to the regression resurrects SMB's alpha to 38 bps with a highly significant *t*-statistic of 5.00. The coefficient of industry-neutral SMB on QMJ is -0.69 with a *t*-statistic of -15.64 . Hence, industry-neutral versions of all the factors show the same pattern, where the size premium is resurrected by controlling for quality even when all industry variation is removed from all of the factors.

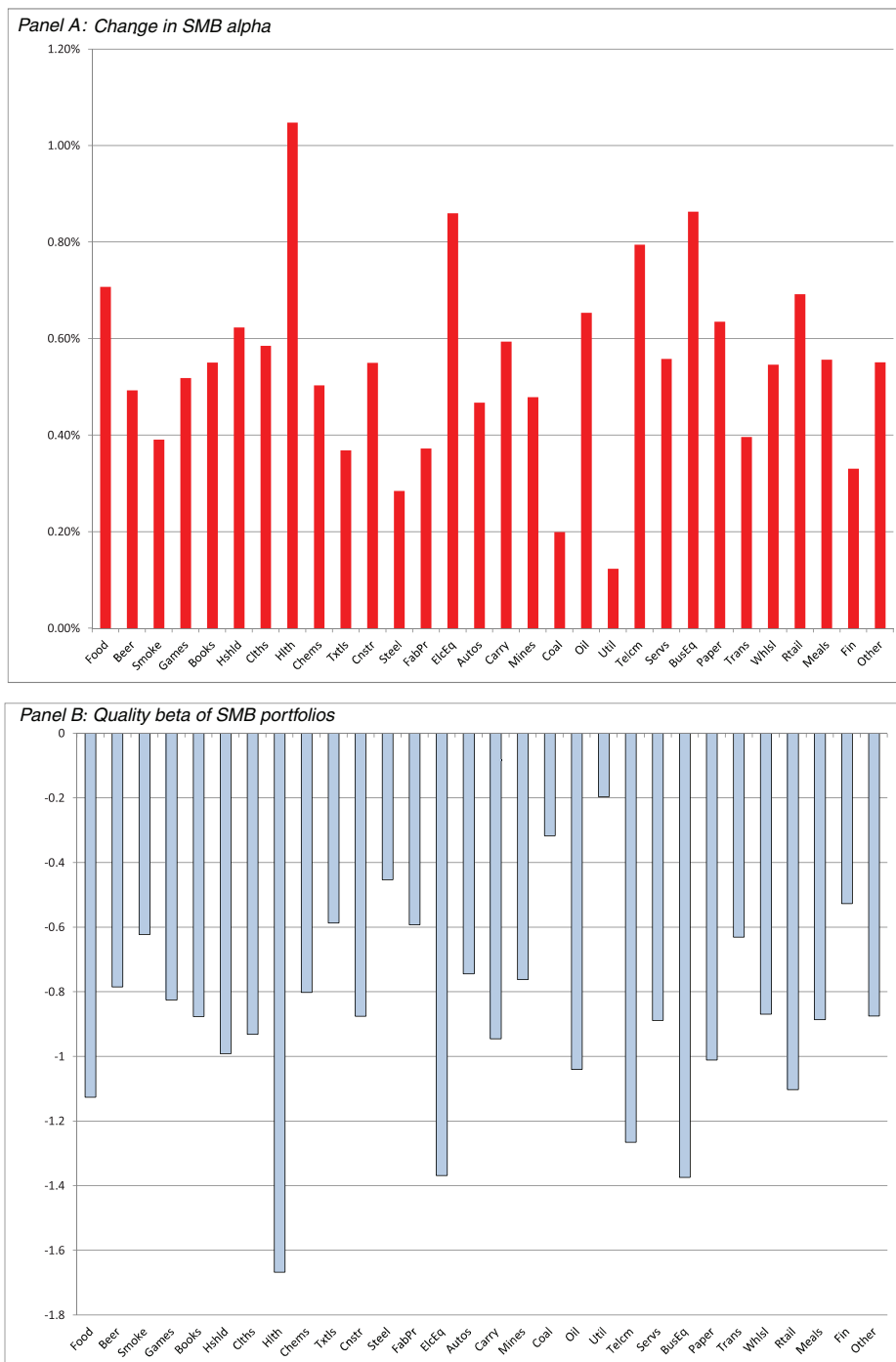


Fig. 2. Intra-industry evidence of small minus big (SMB) premia controlling for quality. Panel A plots the improvement in SMB alphas [relative to the Fama and French market return (RMRF), its one-month lag, high minus low (HML), and up minus down (UMD) factors] after controlling for quality using the Asness et al. (2014) quality factor, QMJ, within the 30 industries defined in Ken French's webpage (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). The difference in SMB alphas between the Fama and French factors and the Fama and French factors augmented with the quality factor is plotted by industry. Panel B plots the betas of each SMB portfolio on quality.

has a more positive size premium of about 25 bps when adjusting for the market, its lag, HML and UMD, though the t -statistic is only 1.52. Adding a quality factor, however, makes the age more golden, as it more than doubles the alpha to 57 bps with a t -statistic of 4.00.

During the embarrassment period, from 1980 to 1999, when we know SMB did not do well, the size premium has a negative alpha of –11 bps. However, controlling for quality using QMJ restores SMB's positive alpha over this period to a sizable 50 bps (t -statistic of 3.06), which is indistinguishable from SMB's alpha over the golden age period. Hence, the embarrassing period is no longer embarrassing and the golden age is not relatively golden once we control for quality. Controlling for quality or junk fully explains the seemingly very different performance of the basic size premium over these two different periods. Finally, the resurrection period has positive SMB alphas but the alpha is larger once we control for quality. Like the other two subperiods, the alpha of SMB controlling for quality is of similar magnitude and highly significant. Hence, accounting for quality, the premium for size is positive and more stable through time.

Because quality seems to explain SMB's performance variation through time, determining whether the variation in basic size's performance is driven by variation in the quality premium or variation in size's exposure to quality would be of interest. Panel A of Fig. 3 plots the average quality premium and SMB's beta to quality over the sample and subsample periods (golden age, embarrassment, and resurrection). Panel A shows that SMB's beta to quality is stable over time and consistently negative, but the returns to quality vary, being lower during the golden age (hence, SMB without a quality adjustment looks relatively stronger in this period) and higher in the embarrassment period (hence, why SMB looks worse during this time).

Panel B of Fig. 3 plots the five-year moving averages of SMB's returns, SMB's alpha with respect to the market, its lag, HML, UMD, and quality and the product of SMB's beta on QMJ times the average return on QMJ. As the plot shows, time variation in SMB's alpha is largely determined by its exposure to, and the returns of, quality. Panel C of Fig. 3 plots SMB's quality beta and the quality premium separately over time and shows that most of the variation comes from the quality premium. Although some variation exists in SMB's quality exposure, it is consistently negative. Hence, the realized quality premium drives most of the variation in the size premium through time.¹³ Size has a very stable negative exposure to quality, but the returns to quality, and not size per se, vary over time in a manner that has confounded previous interpretations of time variation in the basic size premium.

3.3. The size effect controlling for quality: 25 size-quality portfolios

We form size and quality or junk portfolios to look more closely at the size-quality interaction. We form 25 portfolios based on five independent sorts on size (market cap) and five sorts on quality or junk (using QMJ) and group stocks into five quintiles using size and independently five quintiles using quality or junk. The intersection of each of the five categories for each variable determines the composition of the 25 portfolios. Because these are independent sorts with strongly (negatively) correlated sorting variables, the number of firms in each of the 25 portfolios is quite different. The value-weighted average monthly returns in excess of the monthly T-bill rate and their t -statistics are then computed over the sample period from July 1957 to December 2012 for each portfolio.

To get a sense of the intersection between size and quality or junk, Fig. 4 examines the size distribution of stocks within the lowest and highest 20% of quality or junk stocks. Panel A plots the fraction of the number of stocks over time within each of the five independent size quintiles, for the 20% of stocks with the lowest quality ranking (junk). Panel B does the same from the universe of the 20% highest quality stocks. As Panel A shows, junk stocks are mostly small stocks. As Panel B shows, among the highest quality stocks the average size is larger, but plenty of small stocks are represented among the quality group (and the distribution among the various sizes is considerably more even among high-quality stocks than among junk stocks). While junk is more correlated with small stocks and quality is more associated with big stocks, there are plenty of large, junky stocks and plenty of small, quality stocks that enable us to examine the interactions between size and quality, controlling for the other.¹⁴

Fig. 5 plots the reverse exercise of looking at the distribution of quality or junk among the 20% smallest and largest stocks, separately. Panel A shows the distribution of quality among the smallest quintile of stocks, with quality or junk characteristics fairly evenly distributed among the smallest stocks, though a slight tilt toward more junk and less quality. Panel B reports the quality or junk distribution among the largest quintile of stocks and shows that there is a stronger tilt toward high quality and away from junk stocks.

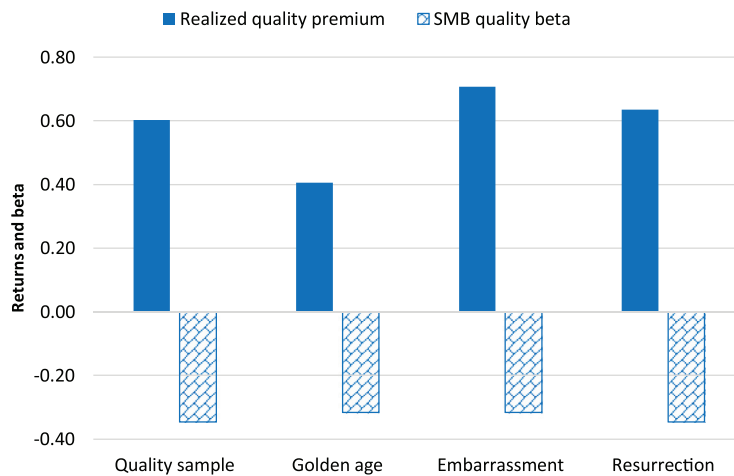
Table 3 reports summary statistics of the 25 size-junk portfolios. The average monthly returns in excess of the Treasury bill rate are reported for each portfolio (along with their t -statistics).¹⁵ Moving across the columns of Table 3, a significant size effect is revealed, as the smallest stocks outperform the largest stocks consistently across the size quintiles. The 25 size-quality sorts represent another

¹³ Fig. OA1 in the Online Appendix plots the year-to-year autocorrelation of portfolio weights based on quality characteristics of stocks for the four different measures of quality: profitability, growth, safety, and payout. For each year, we take the set of stocks that exist at year t and $t-1$ and compare the autocorrelation of their portfolio weights, which are based on quality rankings from year $t-1$ to t . The plot shows tremendous persistence to the quality measures, especially for the first three, which helps explain why size's exposure to quality is stable through time.

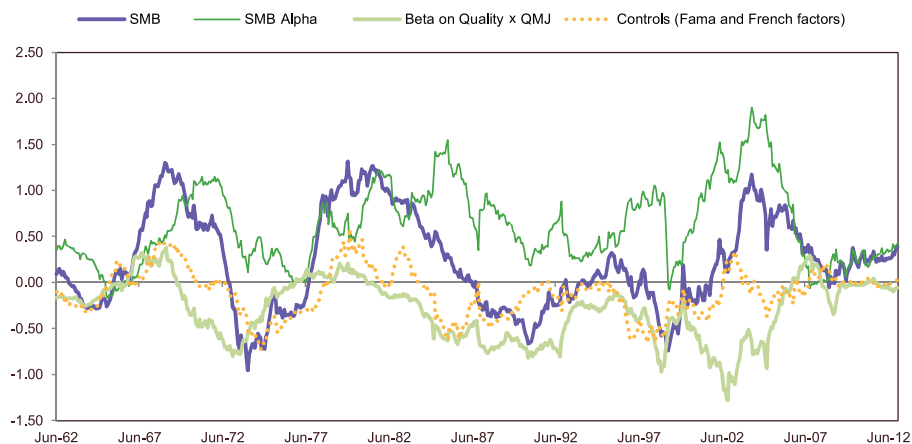
¹⁴ Also, our comments largely refer to the bulk of the sample occurring after the initial period (after about 1960–1965). The very early part of the sample shows a somewhat more even distribution of size amongst both high-quality and junk stocks.

¹⁵ Although the portfolios in Table 6 are constructed from independent sorts on quality and size, dependent-sorted portfolios based on five quintiles of quality and then, within those quintiles, another set of quintile portfolios based on size, yield similar results.

Panel A: Subperiod quality premium and quality beta



Panel B: Five-year moving averages of SMB, SMB alpha, and SMB quality beta x quality premium



Panel C: Five-year moving averages of quality beta, quality premium, and beta x premium for SMB

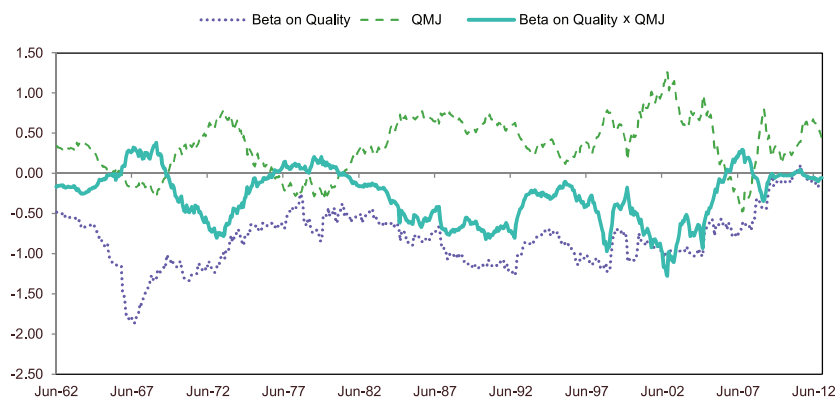


Fig. 3. Time variation in the quality premium and quality beta on small minus big (SMB). Panel A plots the average realized return to long-short portfolios based on quality as well as the quality portfolio's SMB beta over the sample period July 1957 to December 2012 (the quality sample), as well as the Golden age, embarrassment, and resurrection subperiods. Panel B plots the five-year moving averages of SMB returns, alpha with respect to the Fama and French factors augmented with up minus down (UMD) and the quality factor QMJ from Asness et al. (2014), and the product of SMB's beta on quality times the moving average return on quality over time. Panel C plots the five-year moving average of SMB's quality beta and the quality premium separately and their product.

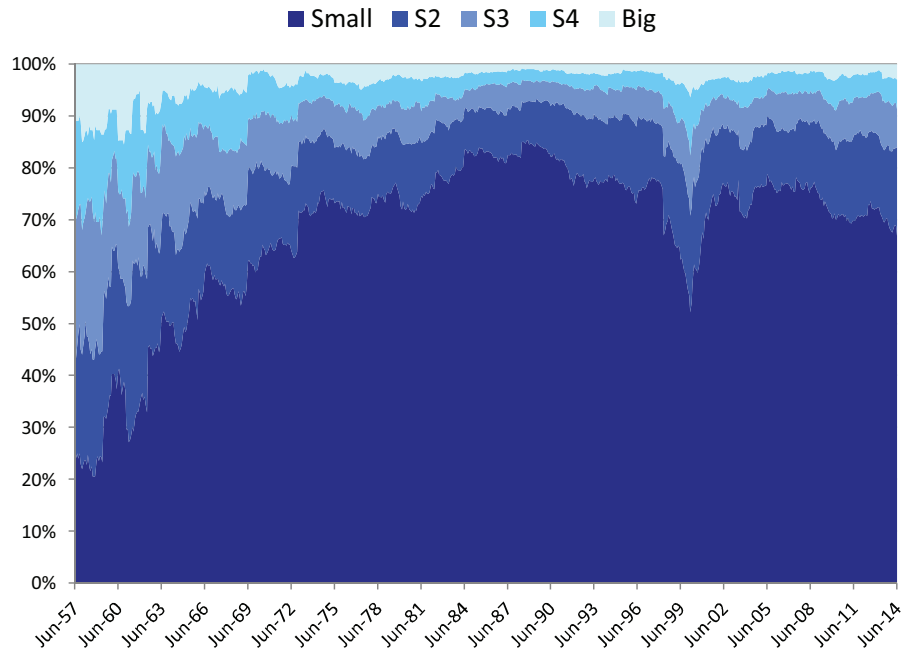
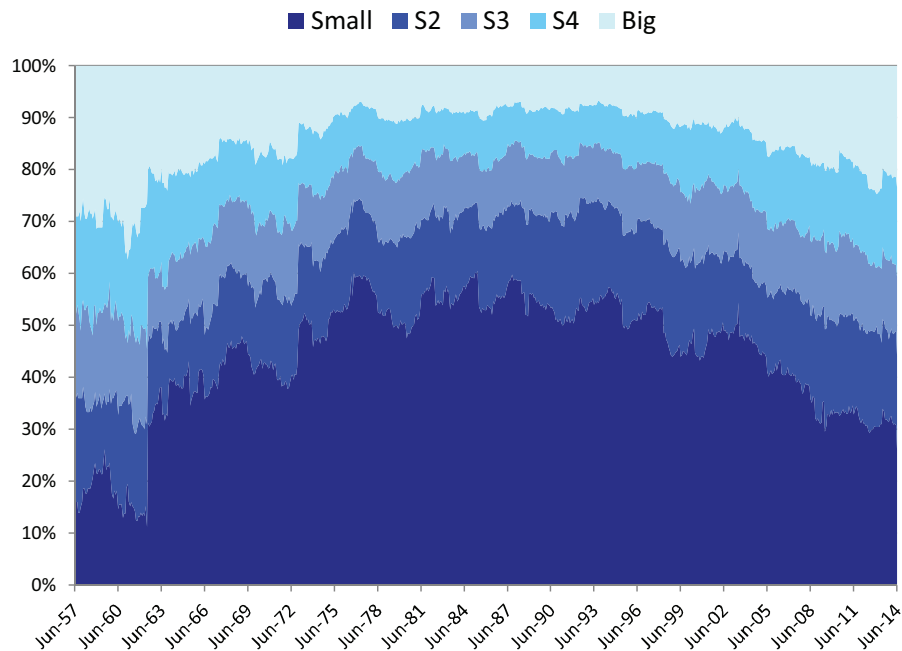
Panel A. Size distribution among low-quality (junk) stocks*Panel B. Size distribution among high-quality stocks*

Fig. 4. Distribution of size among junk and quality stocks. Panel A plots the fraction of the number of stocks over time across five size categories based on size quintiles (S1 (small), S2, S3, S4, and S5 (big)) that make up the 20% of stocks with the lowest quality and highest junk ranking (junk). Panel B plots the fraction of the number of stocks over time across the five size groups that make up the 20% of stocks with the highest quality and lowest junk ranking (quality).

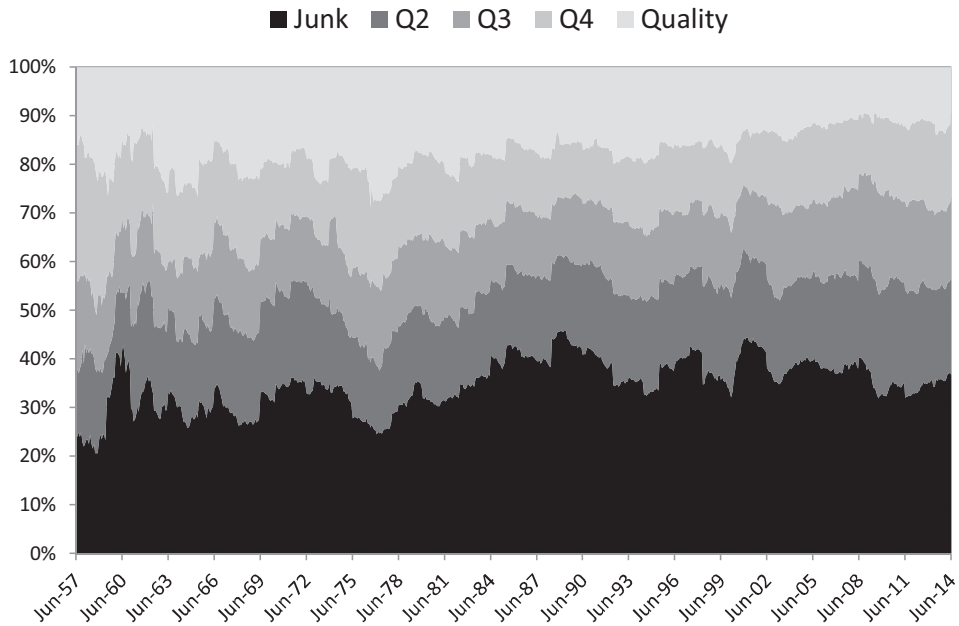
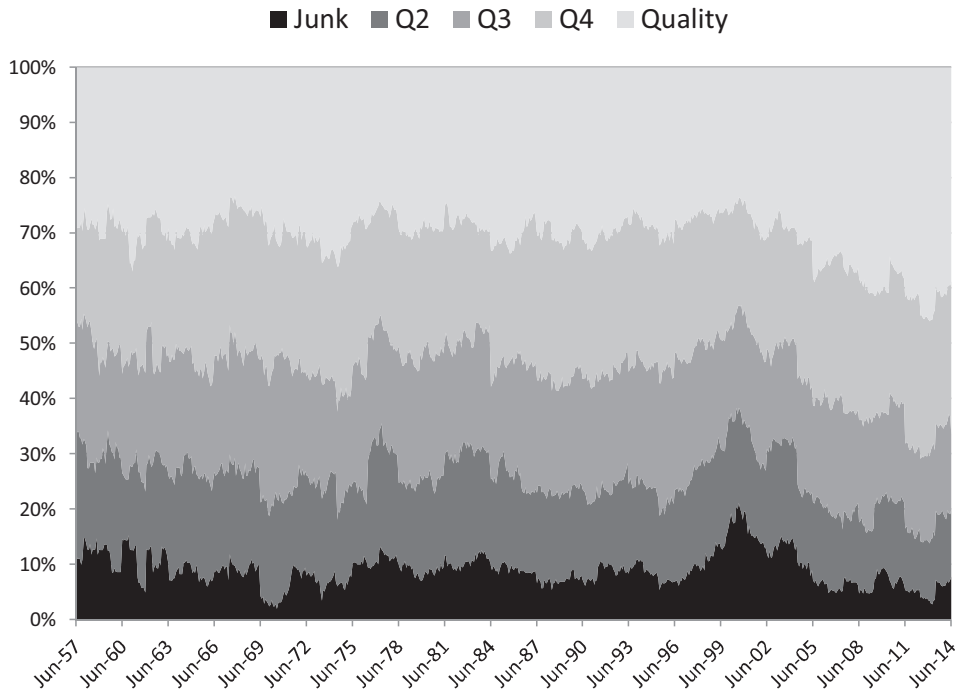
Panel A: *Quality distribution among smallest stocks*Panel B: *Quality distribution among biggest stocks*

Fig. 5. Distribution of quality or junk among large and small stocks. Panel A plots the fraction of the number of stocks over time across five quality groups based on quality quintiles (Q1 (junk), Q2, Q3, Q4, and Q5 (quality)) that make up the 20% of smallest stocks. Panel B plots the fraction of the number of stocks over time across the five quality groups that make up the 20% of largest stocks.

Table 3

Size and junk double sorts.

Panel A reports results from time series regression tests of 25 portfolios sorted on size (market cap) and quality or junk as defined by Asness et al. (2014). The 25 portfolios are formed from independent sorts of stocks into five quintiles using size and quality or junk. The average returns in excess of the monthly T-bill rate and their *t*-statistics are reported over the sample period from July 1957 to December 2012. Panel B reports summary statistics on Fama and French's SMB factor as well as a small minus big (SMB) factor adjusted for quality (SMBQ), which is an average of the "small" minus "big" returns within each quality or junk quintile, averaged equally across the five quality or junk groups. Reported are the annualized means and Sharpe ratios of SMB, SMBQ, market return (RMRF), high minus low (HML), up minus down (UMD), and quality composite portfolio from Asness et al. (2014) (QMJ), along with their correlations with all of the other factors.

Panel A: Time series regression tests							
Portfolio	Small	2	3	4	Big	Small - Big	
Excess returns (%)							
Junk	0.35	0.42	0.44	0.40	0.12	0.23	
2	0.84	0.73	0.68	0.56	0.37	0.46	
3	0.87	0.77	0.74	0.59	0.34	0.54	
4	0.89	0.86	0.76	0.77	0.47	0.42	
Quality	0.97	0.89	0.83	0.78	0.53	0.44	
Quality - Junk	0.62	0.47	0.39	0.38	0.42		
t-statistics							
Junk	1.18	1.46	1.65	1.59	0.50	1.21	
2	3.40	3.12	3.18	2.80	1.99	3.02	
3	3.80	3.56	3.74	3.19	1.93	3.44	
4	4.15	4.09	3.95	4.10	2.72	2.82	
Quality	4.55	4.23	4.11	4.08	3.23	2.87	
Quality - Junk	4.78	3.82	3.37	3.19	2.78		
Panel B: Summary statistics							
	Annual Mean (%)	Sharpe	Correlation with:				
			SMB	RMRF	HML	UMD	QMJ
SMBQ	5.0	0.39	0.85	0.15	0.02	−0.18	−0.31
SMB	2.7	0.26	1.00	0.31	−0.22	−0.02	−0.48
RMRF	5.8	0.38		1.00	−0.29	−0.14	−0.54
HML	4.5	0.47			1.00	−0.17	0.06
UMD	8.7	0.60				1.00	0.22
QMJ	5.5	0.59					1.00

way to control for quality or junk when considering the size effect, although the relation between size and quality is so strong that these double-sorts are only a partial control. Each row represents the relation between size and average returns within (e.g., controlling for) a quality quintile. The only exception to the significant negative relation between size and average returns is among the junkiest, lowest-quality stocks, although the difference between the smallest and largest quintiles is still a healthy 23 bps per month (the *t*-statistic is insignificant). The rest of the quality or junk quintiles exhibit a very strong size effect and a clear relation between size and average returns.

The reverse is also true, that is, controlling for size, there is a clear quality premium. In every size quintile, quality outperforms junk and the relation is fairly stable. Hence, quality or junk and the size effect are not the same thing, though they are (negatively) related.

The results in Table 3 provide further insight into our earlier findings. Controlling for quality resurrects size in many places where it was previously and seemingly absent. As Table 3 highlights, the junk stocks could be the most interesting set of firms, where among them the re-

lation between returns and size breaks down. These junk stocks have on average poor and very volatile returns and, hence, help explain many of the empirical challenges to the size effect.

Another effective way to look at the size premium independent of quality is to create a quality-hedged size portfolio. Taking an equally-weighted average of the five small minus big portfolios within each quality quintile, we find a return spread of 42 bps per month with a *t*-statistic of 3.18. This average size portfolio, which we refer to as small minus big adjusted for quality or SMBQ, represents the average size premium controlling for quality, although the control is imperfect, as SMBQ still has a negative correlation with quality but a far less strong one than SMB without the quality control. In Panel B of Table 3, we report the annualized mean and Sharpe ratio of SMBQ as well as the Fama and French factor SMB, which does not control for quality in its construction. SMBQ has nearly twice the mean return as SMB and has a 50% higher Sharpe ratio, 0.39 for SMBQ versus 0.26 for SMB. For comparison, we also report the means and Sharpe ratios of the market (RMRF), HML, UMD, and QMJ. SMB has the weakest

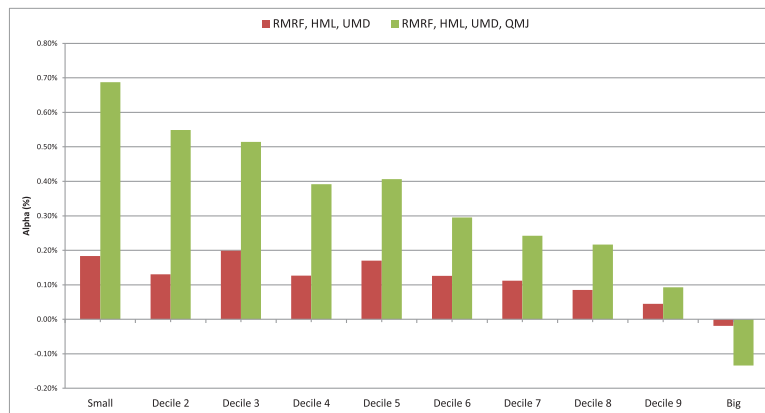


Fig. 6. Size decile alphas. The alphas of each size decile are plotted with respect to the Fama and French market return (RMRF), RMRF lagged a month, high minus low (HML), and up minus down (UMD) factors and those same factors augmented with the quality factor, QMJ, from Asness et al. (2014).

performance among all factors, but this is because SMB is short a very significant quality premium. Reducing SMB's exposure to quality as SMBQ does (though not eliminating it) puts its performance on a more equal footing with the other factors.

Looking at the correlations among the factors reveals that SMBQ is highly correlated with SMB (0.85) and less negatively correlated with QMJ (-0.31) than SMB is (-0.48). The -0.31 correlation shows, however, that even SMBQ is still negatively correlated to quality, indicating that making the size portfolio completely quality-neutral is difficult, unless specifically hedged from a regression. The other correlations are interesting, too, as SMB has a strong negative value tilt (-0.22) and SMBQ is value-neutral (correlation of 0.02). For these reasons, to examine the independent variation of size, considering alphas from multivariate regressions on all these factors, as in Table 2, is more intuitive.

3.4. Is the size premium concentrated in extreme stocks?

Fig. 6 examines the returns to size more finely by looking across size-sorted decile portfolios. From this analysis, we can address whether the size premium is concentrated in the extremes and if the relation between size and average returns is monotonic.

Fig. 6 plots the alphas of each size decile with respect to the Fama and French factors RMRF, RMRF lagged a month, HML, and UMD and these same factors augmented with a quality factor. All regressions are run over the sample period from July 1957 to December 2012. As the figure shows, the alphas adjusted for the Fama and French factors are barely higher for the smallest decile of stocks (Decile 1) compared with the largest (Decile 10) and are essentially flat across Deciles 2 through 9 and exhibit no reliable pattern. In short, no consistent relation exists between size and average returns across the deciles in terms of market or Fama and French-adjusted alphas. This finding is con-

sistent with claims in the literature that the size-return relation is not linear. However, when adding quality as a factor, not only is the return difference between the smallest and largest size deciles magnified, but perhaps more interesting, an almost linear monotonic relation between the size deciles and their alphas also emerges. Moving from small to big stocks, the alphas steadily decline and become negative for the largest stocks. Hence, controlling for quality or junk restores a more consistent relation between size and average returns.

The fact that quality resurrects the size premium can in part be related to it restoring more decile-monotonicity as well, if a larger absolute premium can reduce the influence of noise on each portfolio. However, it did not have to work out this way. A quality factor could have just as easily raised the returns on all size deciles equally without improving an increasing relation, or it could have added more to the larger deciles or to random deciles and reduced monotonicity. The fact that when size increases proportionately less alpha is evident when controlling for quality suggests that quality or junk exposure is related to size in a linear way and that controlling for quality or junk therefore restores a tighter linear relation between size and average returns, which current asset pricing models (e.g., Fama and French, 2015, 2016) assume.

The relation between size and quality is also stable through time. Panel A of Fig. 7 plots 10-year rolling beta estimates of SMB on QMJ over the sample period (July 1967 to December 2012) and shows that the betas are always negative and range from -0.40 to -1.25 . Panel B of Fig. 7 plots the rolling ten-year betas of each size decile. Again, the time series variation in the betas is relatively small, but more interesting, the ordered relation between size and quality or junk is extremely stable though time, as smaller size deciles consistently have more negative quality betas and the effect is very stable throughout the sample. Few periods exist in which betas with respect to quality are not ordered almost perfectly by size, a remarkable

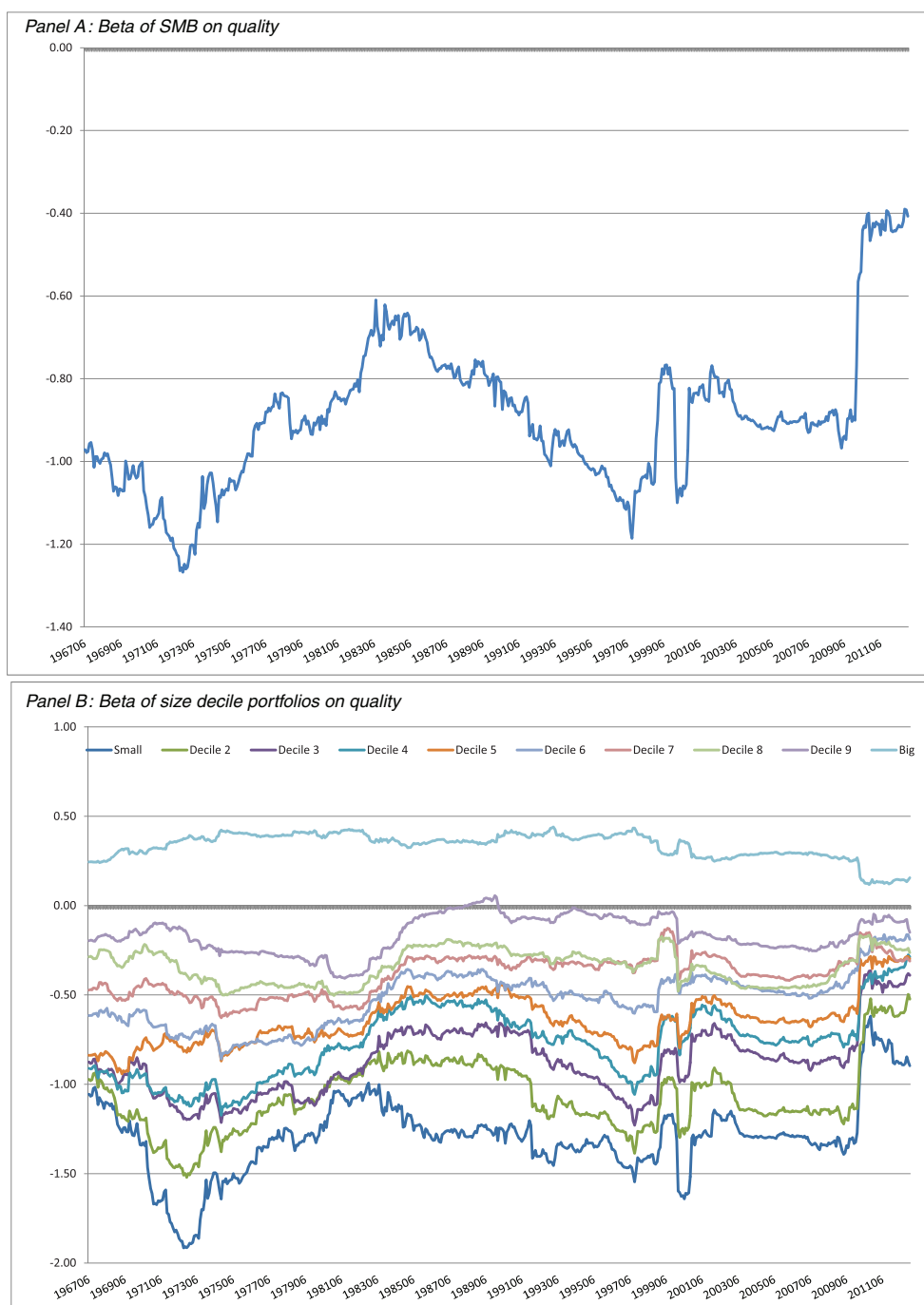


Fig. 7. Rolling beta estimates of size portfolios on quality. Panel A plots the ten-year rolling beta estimates of small minus big (SMB) on quality, using the Asness et al. (2014) composite quality factor QMJ, estimated from a model that also includes RMRF, its lag, high minus low (HML), and up minus down (UMD), using the monthly returns over the preceding 120 months. Panel B plots the rolling quality betas of each size decile portfolio.

feat considering the estimation error inherent in beta estimates. Repeating the same exercise for other measures of quality or junk using the various measures of Asness et al. (2014), Fama and French (2016), or Frazzini and Pedersen (2014) yields similar results.

3.5. Seasonality in size: the January effect

Prior research has shown that the size effect mostly resides in January (Keim, 1983; Roll, 1983; Reinganum, 1983b). Table 1 showed that all of the returns to size are

Table 4

Seasonal patterns and the size premium.

The table reports regression results for the size premium (SMB) on the factors RMRF, its lagged value, HML, and UMD and the composite quality factor from Asness et al. (2014), where the alphas are estimated for the months of January and non-January separately using dummy variables for those months. Also reported is the difference between January and other months, along with a t -statistic on that difference in the last column. Results are reported over four sample periods: the full quality sample period (July 1957 to December 2012), and the golden age (July 1957 to December 1979), embarrassment (January 1980 to December 1999), and resurrection (January 2000 to December 2012) subperiods for the size premium.

Period	$SMB_t = \alpha_{Non-Jan.} + \alpha_{Jan.} + \beta RMRF_t + \beta_{-1} RMRF_{t-1} + hHML_t + mUMD_t + qQMJ_t + \varepsilon_t$															January	
	$\alpha_{Non-Jan.}$	$t(\alpha)$	$\alpha_{Jan.}$	$t(\alpha)$	β	$t(\beta)$	β_{-1}	$t(\beta_{-1})$	h	$t(h)$	m	$t(m)$	q	$t(q)$	R^2	difference	$t(\text{difference})$
Quality sample	−0.0004	−0.32	0.0209	5.59	0.16	6.21	0.13	5.29	−0.19	−4.68	0.02	0.90			0.18	0.0213	5.46
	0.0038	3.62	0.0157	4.74	−0.03	−1.28	0.10	4.77	−0.26	−7.10	0.07	3.08	−0.71	−14.37	0.38	0.0119	3.42
Golden age	−0.0001	−0.08	0.0354	6.34	0.25	6.95	0.14	4.02	−0.10	−1.41	−0.03	−0.67			0.34	0.0355	6.13
	0.0033	2.42	0.0359	7.61	0.05	1.55	0.14	4.75	−0.38	−6.02	−0.01	−0.21	−0.94	−11.27	0.55	0.0326	6.67
Embarrassment	−0.0016	−0.89	0.0045	0.79	0.03	0.79	0.18	5.01	−0.25	−3.67	−0.07	−1.46			0.19	0.0061	1.04
	0.0058	3.35	−0.0013	−0.27	−0.14	−3.42	0.15	4.87	−0.42	−6.81	−0.06	−1.51	−0.86	−9.12	0.40	−0.0071	−1.37
Resurrection	0.0041	1.50	0.0180	1.98	0.27	4.44	0.09	1.55	−0.33	−4.22	0.15	3.22			0.26	0.0139	1.45
	0.0091	3.86	0.0069	0.90	−0.18	−2.40	−0.03	−0.58	−0.18	−2.68	0.17	4.33	−0.84	−8.19	0.49	−0.0022	−0.27

concentrated in January, with no evidence of any size effect (economically or statistically) outside of January.

Table 4 reexamines the seasonality in the size premium after controlling for quality. Reported are results from regressions of the returns to SMB on a January dummy, a non-January dummy (February through December), and the Fama and French factors RMRF, RMRF lagged, HML, and UMD over the full sample period. Confirming earlier results and those in the literature, there is a large January size premium (2.09% with a t -statistic of 5.59), no evidence of any size effect outside of January (−0.04% with a t -statistic of −0.32), and the Fama and French factors do not capture much of this seasonality. The last column of the table reports a statistical test for the difference in January versus non-January alphas, which easily rejects the null that January has the same returns as the rest of the year.

Table 4 also reports results when adding a quality factor (QMJ) to the regression. Controlling for quality delivers a positive and significant size premium outside of January of 38 bps (t -statistic = 3.62). Controlling for quality mitigates the very large size premium in January, dropping it from 2.09 to 1.57%. While a large January premium still remains, the premium for size is now present throughout the year and the difference between the January and non-January alpha, which is still significant, is now halved.

Table 4 then repeats this exercise over the subsample periods golden age, embarrassment, and resurrection. In every subperiod, quality resurrects the size effect outside of January, delivering a consistent premium of at least 33 bps (golden age) and as much as 91 bps (resurrection). Outside of January, the returns to size controlling for quality are larger (almost twice as large) during the embarrassment period than they are during the supposed golden age for size. Hence, regarding February to December, the notion of a golden age period for size and an embarrassment period for size is backward. As Table 4 shows, this is due to the quality premium confounding the performance of size over these periods. A major reason the golden age for size exists is because a size portfolio is long junk and short quality, with junk greatly outperforming quality in January

months over this period. Hence, failure to control for quality or junk obscures the size effect. Consistent with this notion, quality also diminishes the size premium in January, when it is insignificant in the last two subperiods and insignificantly different from the returns in February to December over these subperiods.

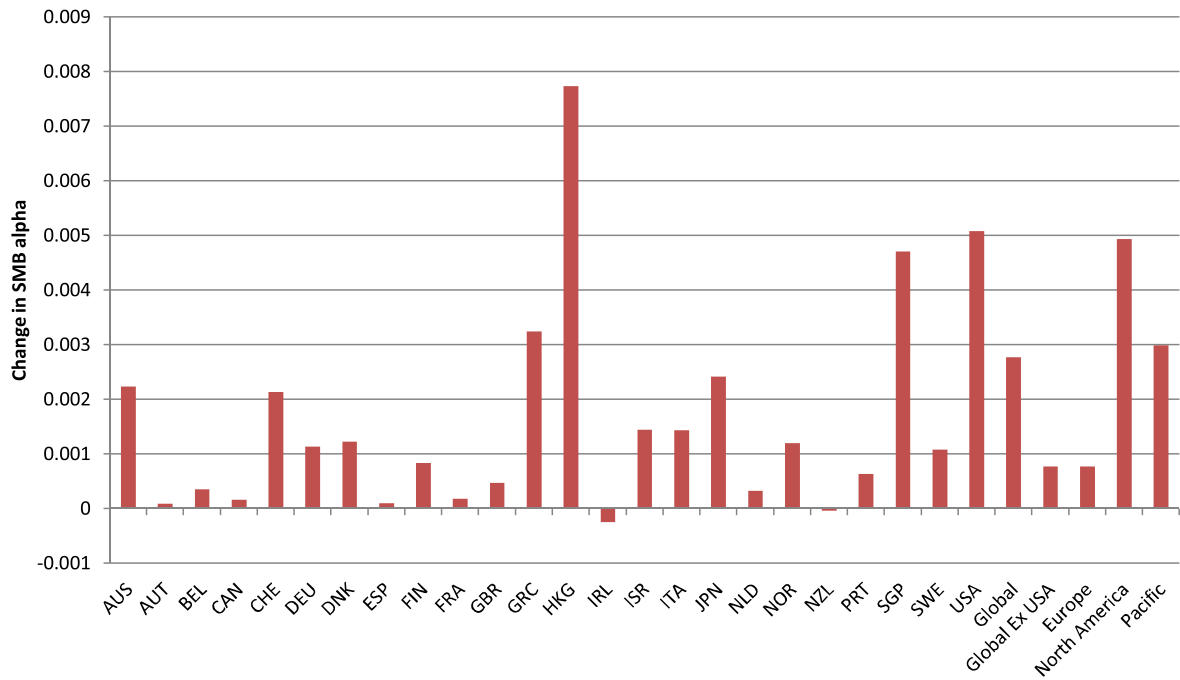
These results suggest that quality or junk also helps explain the strong seasonality associated with size-based strategies. The strong performance of junk stocks in January drives a significant fraction of the apparently high returns to size in January, while depressing the returns to size outside of January. Controlling for quality or junk reduces this seasonal component substantially and shows a strong size premium throughout the year.¹⁶

3.6. International evidence

Finally, we examine the size effect in 23 other countries to perform out of sample tests on the role of quality in reviving the size effect. We form SMB portfolios within each international equity market following the same procedure as above. Similarly, we form a quality factor based on the Asness et al. (2014) QMJ factor construction in each of these markets following their procedure. Fig. 8 reports the change in SMB's alpha for each country after controlling for quality by regressing SMB in each country on the local stock market index, the market lagged, HML, and UMD factors constructed within that market. The same regression is repeated including the quality factor in that market as a regressor, and the difference in SMB alphas between the two regressions, with and without controlling for quality, are plotted country by country in Fig. 8. Panel A shows

¹⁶ This evidence also contradicts theories for the seasonality of size related to turn-of-the-year price pressure due to tax-loss selling, window dressing by institutions, or cash infusion of investors. Controlling for quality, the turn-of-the-year evidence on size is much weaker and the size effect is more evenly distributed throughout the year. Hence, while price pressure effects in December and January could still be present, they are less about small versus large stocks than previously thought.

Panel A: Change in SMB alpha after controlling for quality



Panel B: Quality beta of SMB within country

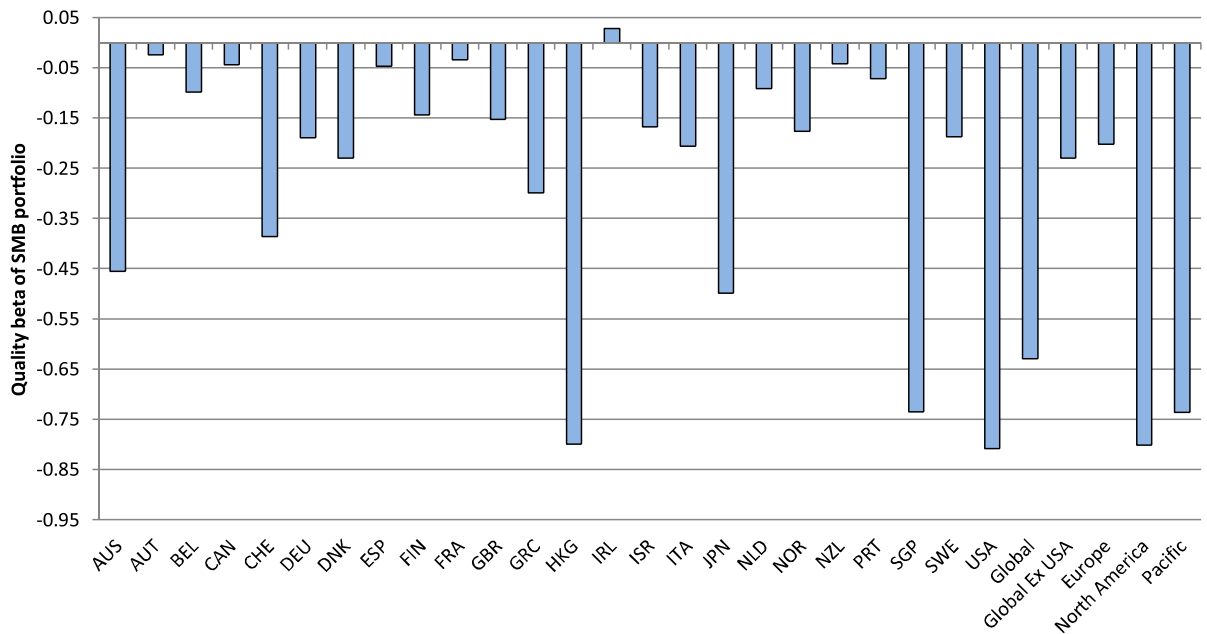


Fig. 8. International evidence of SMB premia controlling for quality. Panel A plots the improvement in SMB alphas [relative to the Fama and French factors market return (RMRF), RMRF lagged a month, high minus low (HML), and up minus down (UMD)] after controlling for quality (using the [Asness et al. \(2014\)](#) composite quality factor (QM)) across 24 countries, as well as five regions: global, global excluding US, Europe, North America, and Pacific. Plotted is the difference in SMB alphas between the Fama and French factors versus the Fama and French factors augmented with the quality factor, by country and region. Panel B plots the betas of each SMB portfolio on quality. The regressions are estimated using rolling five years of data for each country.

a positive increase in SMB alpha for 23 out of 24 countries once we control for quality (the exception being Ireland, where the point estimate is very close to zero and statistically no different from zero). Panel B of Fig. 8 shows that the betas of SMB on quality are also uniformly negative (excluding Ireland) and where we find that 14 out of 24 are statistically significant. The t -statistics range from 0.5 to -17 , and the median t -statistic is -2.2 .

Aggregating across all countries outside of the US, the t -statistic of the beta on quality is -3.32 . These results are remarkably consistent across countries, providing evidence of a negative relation between size and quality across countries and, therefore, a healthy size premium internationally once we control for quality and, hence, a wealth of out-of-sample evidence for our earlier findings.

As a further test of the international results, we replace the international version of QMJ with the international versions of the Fama and French (2016) RMW and CMA factors (available on Ken French's website) to control for quality. Using SMB formed across countries, excluding the US, and regressing it on the international versions of the market, HML, and UMD, we obtain an insignificant alpha of 11 bps with a t -statistic of 0.99. Adding RMW and CMA to the regression to control for quality, SMB's alpha goes up to a significant 29 bps with a t -statistic of 2.38. Regressing SMB adjusted for quality on SMB unadjusted for the Fama and French quality factors, the difference in alphas (of 18 bps) has a t -statistic of 6.36, more than six standard errors from zero.

These results are further testament to the strength of the size-quality interaction, in which the same patterns emerge in 23 other international markets using a variety of quality metrics, including those created by other researchers, such as Fama and French (2016). The size premium appears to be alive and well across all of these markets once we control for quality in any number of ways.

4. Further tests of the theories of the size effect

We turn now to further tests of the theories that have been offered for the size effect, and in particular how the interaction between size and quality/junk may help inform us about various theories.

4.1. Risk-based theories: characteristics and asset pricing tests

The strong returns to size when controlling for quality present a challenge to standard asset pricing theories to the extent that size is not a risk in and of itself. In other words, for size to be priced in a rational model, size would have to proxy for a rational risk factor such as the risk associated with growth options. Hence, the more we control for actual risk factors, the smaller should be the residual return of size. Our results presented in Section 3 test this hypothesis using a variety of risk factors proposed in the literature and what we find is the exact opposite: Controlling for growth and other factors makes the size effect stronger, not weaker.

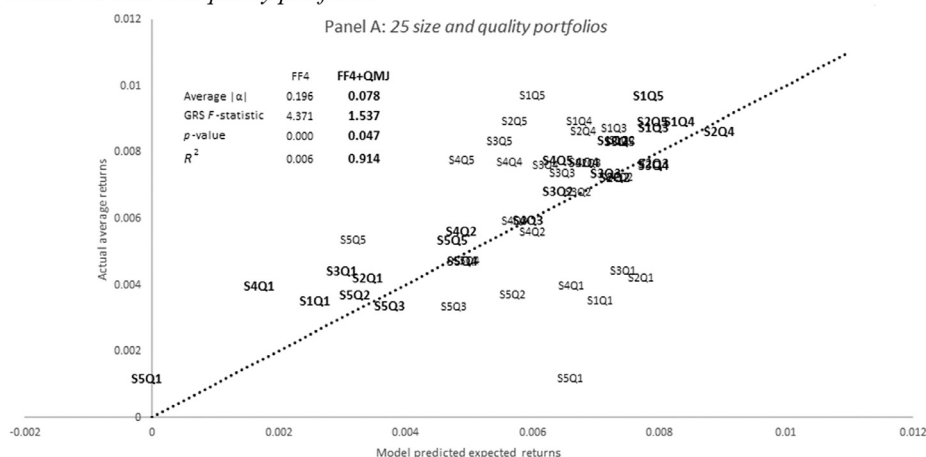
Hence, our results in Section 3 present a rejection of this interpretation of standard asset pricing models, although we acknowledge that size could still proxy for other risk factors (unknown to us, but presumably known to the investors who are setting prices) or measurement errors, e.g., in the measurement of growth options (for which the theoretical literature does not offer clear guidance on the best empirical measures). In this section, we conduct further standard asset pricing tests of size-quality or junk portfolios.

We conduct asset pricing tests using the 25 size-quality/junk portfolios described in Section 3.3 as well as the 25 size-BE/ME portfolios of Fama and French (1996). We examine two different factor models: (1) the Fama and French (1993) factors RMRF, SMB, and HML, augmented with the UMD momentum factor and (2) a factor model that includes those same factors plus the composite quality factor, QMJ. The goal is to see how these models explain the portfolio returns. In other words, here we follow Fama and French (1996) and include size in the risk model even though standard theory does not predict that size is a priced risk. We show that, even when size is included as a risk factor, we can reject the Fama and French model, but including quality significantly improves the fit. The success of the model that includes both size and quality (and the other standard factors) can either be viewed as supportive of a rational model where we are yet to understand why size is a rational risk factor (as well as why some of the other factors are risk-based) or an expression of the fact that we need to include size and quality to explain the cross section of returns, rejecting the standard theories that do not predict such risk factors.

Fig. 9 plots the actual average returns of portfolios versus the predicted expected returns of those portfolios from each asset pricing model. A 45-degree line forced through the origin is plotted to highlight the pricing errors from each asset pricing model for each portfolio, with the distance between the points and the 45-degree line representing the pricing error under each model, assuming that the intercept from the cross-sectional regression between average and model-predicted returns is zero, which is the same as forcing each model to price the equity risk premium in addition to the cross section of returns. Also reported are the average absolute pricing errors or alphas from each model, the Gibbons et al., (1989) (GRS) F -statistic on whether the alphas are jointly zero, its p -value, and the cross-sectional R -squared from a regression of average returns on predicted model expected returns. Panel A plots the results for the 25 size-quality portfolios.

As Fig. 9 shows, the Fama and French plus momentum factors do not explain the returns to size-junk portfolios, leaving an average absolute alpha of almost 20 bps per month with a Gibbons et al., (1989) (GRS) F -statistic of 4.37 that easily rejects the null that the alphas are zero. The cross-sectional R -squared is 0.006, indicating no reliable relation between the predicted returns from the Fama and French factors and the average returns of the size-quality portfolios. The Fama and French size, value, and momentum factors fail to capture the dimension of quality in returns and the size-quality interaction. Adding QMJ to the Fama and French factors, how-

Panel A: 25 size and quality portfolios



Panel B: 25 size and BE/ME portfolios

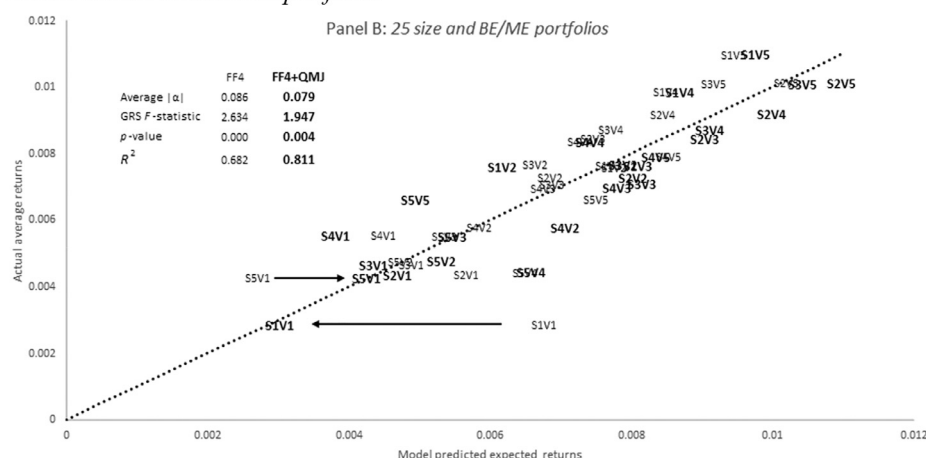


Fig. 9. Asset pricing tests and pricing errors on size, book-to-market equity ratio (BE/ME), and quality portfolios. Plotted are the actual average returns of portfolios versus the predicted expected returns of those portfolios from an asset pricing model. The data points plotted in normal font correspond to the predicted expected returns from the Fama and French (1993) model augmented with a momentum factor (FF4) and the data points plotted in **bold** correspond to the predicted expected returns from the FF4 model plus the quality factor, QMJ. A 45-degree line forced through the origin is plotted to highlight the pricing errors from each asset pricing model for each portfolio, with the distance between the points and the 45-degree line representing the pricing error under each model assuming the intercept from the cross-sectional regression between average and model-predicted returns is zero, which is the same as forcing each model to price the equity risk premium in addition to the cross section of returns. Also reported are the average absolute pricing errors or alphas from each model, the Gibbons et al., (1989) F -statistic on whether the alphas are jointly zero, its p -value, and the cross-sectional R -squared from a regression of average returns on predicted model expected returns. Panel A plots the results for the 25 size-quality portfolios; Panel B, for the 25 size and BE/ME portfolios from Fama and French (1996).

ever, explains the portfolio returns nicely, as the average absolute alpha drops to 7.8bps with an F -statistic of only 1.54 that fails to reject the null. The cross-sectional R -squared here is 0.914, suggesting a tight relation between the model's predicted returns and actual average returns. It is also worth noting that both models contain SMB, yet the first model could not explain the variation in average returns across the size-quality spectrum without the quality factor. This indicates that quality is not subsumed by size and is necessary in explaining these portfolio returns.

Panel B of Fig. 9 shows the results for the 25 size and BE/ME portfolios of Fama and French (1996). This exer-

cise is interesting because the test portfolios are not sorted by quality, hence we can see how well the quality factor prices the classic Fama and French 25 size and BE/ME portfolios. Here, the Fama and French factors do better, but not as well as adding quality. Under the Fama and French factors, an average absolute alpha of 8.6bps with a GRS F -statistic of 2.63 easily rejects the null. The cross-sectional R -squared is 0.682. As is well known, Fama and French (1993, 1996) statistically reject their own model due largely to the poor performance of extremely small growth stocks (and to a lesser extent, large growth stocks). These are highlighted on the graph as S1V1 (smallest size, lowest BE/ME) and S5V1 (largest size, lowest BE/ME).

Adding quality to the regression reduces the pricing errors to an absolute average 7.9 bps with a GRS F -statistic of only 1.95 and a cross-sectional R -squared of 0.811. The quality factor helps improve the fit of the 25 size and BE/ME portfolios. The plot shows precisely how the quality factor is helping, as the two most troublesome portfolios, smallest and largest growth, are captured by their covariance with the quality factor that Fama and French's SMB and HML portfolios fail to capture. The arrows on the graph highlight how these two portfolios are priced under the two different models (one excluding and one including the quality factor). Hence, quality also helps capture the extreme low returns to small growth stocks which lead to rejection of the Fama and French model.

4.2. Risk-based theories: non-price-based size measures

Berk (1995a), using an argument from Ball (1978), shows that because size is typically measured by market capitalization, which contains market prices, any misspecification of the asset pricing model leads to a negative relation between size and average returns. In other words, according to this theory, size is a measure of time-varying risk premiums because size is measured by market value, which is influenced by risk premia. Berk (1995b, 1997) suggests that using non-price-based measures of size is therefore a better way to test the true relation between size and average returns and finds that using non-price-based size measures results in no reliable size premium.

Table 5 reexamines the relation between non-price-based measures of size and average returns. In Panel A, we rank stocks based on the non-price size measures suggested by Berk (1997) plus two others, i.e., book assets, book equity, sales, PP&E, and number of employees. For each non-price size measure, stocks are ranked into deciles every June and the value-weighted returns of each decile are computed over the following year (the exact same procedure we use to form the market cap size deciles). Panel A reports the alphas of the return difference between the smallest and largest decile portfolios using the non-price-based size measures from regressions on the factors RMRF, RMRF lagged, HML, and UMD. No reliable size premium exists for any of the non-price-based size measures over the sample period. The last column of the table reports results for the portfolios sorted on market capitalization for comparison. Here, too, the results are insignificant. Thus, consistent with Berk (1995a, 1995b, and 1997), we find no reliable relation between non-price-based size measures and average returns, but inconsistent with Berk's theoretical argument we also fail to find any relation between price-based size measures and returns.

Panel A then repeats the regressions but adds a quality factor (QMJ) to the regression. Controlling for quality systematically resurrects a size premium among every non-price-based size measure (and the market cap measure, too). The contrast in results across all size measures is striking: Every estimated alpha from the first row is insignificant (ranging from 0 to 17 bps), and every alpha in the second row, which controls for quality, is large, positive, and significant (ranging from 58 to 83 bps with t -statistics of 4.5 to almost 6). Comparing the magnitude

of these alphas with those based on market capitalization (in the last column), we reject the Berk (1995a) conjecture that the non-price-based size deciles deliver a smaller or insignificant size premium. Book assets, sales, book equity, PP&E, and number of employees produce size decile premia of 83, 67, 66, 58, and 68 bps per month, respectively, and the market cap size decile premium is 64 bps over the same period, after controlling for quality.

Table 5 also regresses the non-price-based size portfolios on the same factors plus the market cap-based size portfolio SMB. These regressions test whether the non-price-based size portfolio returns are any different from the price-based size factor, in the presence of the other factors that include quality. Without quality, no reliable alpha is associated with the non-price-based size portfolios. However, controlling for quality, the non-price-based size portfolios deliver positive alphas even after adjusting for market cap-based size via SMB. All five non-price-based size measures produce positive alphas, ranging from 9 to 29 bps, with respect to SMB after controlling for quality, and three out of five are statistically significant. Hence, non-price-based size measures seem to deliver at least as high return premia as market cap-based measures, which is the opposite of the conjecture in Berk (1995a).¹⁷

Panel B of Table 5 examines the relation between non-price-based size measures and returns using double-sorted portfolios based on book assets and size/junk. The analysis is identical to that used for Table 3, where we independently sort stocks based on size and quality or junk into quintiles and calculate returns to the intersection of those sorts to form 25 size-junk portfolios, except here we sort by book assets instead of market cap for our size measure. Panel B reports the average returns in excess of the one-month T-bill rate for each of these 25 portfolios, along with their t -statistics, as well as returns adjusted for the market, HML, and UMD (and their t -statistics). The results show a healthy size premium within each quality category. Controlling for quality, small stocks significantly outperform

¹⁷ Repeating these regressions over the three subperiods (golden age, embarrassment, and resurrection), the non-price-based size measures fail to generate significant alphas on their own. However, once we control for quality, all non-price-based size measures yield very significant alphas. Controlling for quality also makes the non-price-based size premia stable across the different subsamples. In addition, Fig. OA2 in the Online Appendix reports results from the intra-industry exercise we conducted earlier, but using SMB portfolios formed from the non-price based size measures instead of market capitalization. Within each of the 30 industries, we form SMB portfolios based on book assets, sales, book equity, PP&E, and number of employees. We then regress the non-price-based SMB returns on the market, its lagged value, and HML and UMD factors and regress the SMB returns on these same factors plus a quality factor. The difference between the alphas are then plotted industry by industry in Fig. OA2, representing the improvement in SMB performance from controlling for quality or junk. Fig. OA3 plots the change in alpha before and after controlling for quality by country for non-price-based measures of size using book assets, sales, book equity, PP&E, and employees to sort stocks within each country. For the vast majority of countries, there is a significant size premium even for non-price-based measures of size once we control for quality. These results provide even more evidence of a robust size effect internationally as well as a large number of out-of-sample tests for non-price-based size measures that help alleviate any data mining concerns.

Table 5

Size premium for portfolios sorted on non-price-based measures of size.

Panel A reports regression results for the return difference between the smallest size decile portfolio (P1) and the largest size decile portfolio (P10), P1–P10, where portfolios are value-weighted based on sorts using non-priced-based measures of size, book assets, sales, book equity, property, plant, and equipment (PP&E), and number of employees. We also include portfolios sorted on market cap for comparison in the last column. The P1–P10 spread portfolio for each of the non-price-based size measures is constructed in the same manner used for market capitalization-sorted portfolios. We form decile portfolios by sorting stocks each July, based on their June measure of size using each of the non-price-based size measures, and then compute returns to each decile portfolio, in which securities are weighted by their market values, over the following year. Results are reported for regressions of the non-price-based size premia on the market, the lagged market, high minus low (HML), and up minus down (UMD), with and without controlling for quality using the composite factor, QMJ, from Asness et al. (2014) as a regressor. We also report results from regressions that also include small minus big (SMB) as a regressor (market cap-based size portfolio) to examine if the non-market-based size portfolios exhibit any different returns from the market cap-based size portfolios, with and without controlling for junk. For brevity, we report only the estimated alphas and their *t*-statistics. Panel B reports returns and *t*-statistics of double sorted portfolios based on the non-price-based measure of size, book assets, and quality or junk, with stocks sorted independently on book assets and junk into five quintiles and the intersection of the groups forming the 25 portfolios. Raw returns in excess of the one-month Treasury bill rate, as well as risk-adjusted returns or alphas net of the Fama and French market return (RMRF), HML, and UMD factors, are reported. Results are reported over the full sample period over which quality variables are available (July 1957–December 2012).

Panel A: Regressions of P1-P10 using non-price based measures of size

Size measure	Book assets		Sales		Book equity		PP&E		Employees		Market Cap	
	α	$t(\alpha)$	α	$t(\alpha)$	α	$t(\alpha)$	α	$t(\alpha)$	α	$t(\alpha)$	α	$t(\alpha)$
$P1 - P10_t = \alpha + \beta RMRF_t + \beta_{-1} RMRF_{t-1} + hHML_t + mUMD_t + qQMJ_t + \varepsilon_t$												
No control for quality ($q=0$)	0.0017	0.96	0.0002	0.10	0.0004	0.22	0.0008	0.00	0.0000	0.01	0.0004	0.20
Control for Quality ($q \neq 0$)	0.0083	5.98	0.0067	5.52	0.0066	4.98	0.0058	4.57	0.0068	5.78	0.0064	4.21
$P1 - P10_t = \alpha + \beta RMRF_t + \beta_{-1} RMRF_{t-1} + sSMB_t + hHML_t + mUMD_t + qQMJ_t + \varepsilon_t$												
Control for market cap size												
No control for quality ($q=0$)	-0.0002	-0.31	-0.0015	-1.71	-0.0015	-2.14	-0.0070	-0.88	-0.0015	-1.48		
Control for Quality ($q \neq 0$)	0.0025	3.42	0.0020	2.66	0.0009	1.43	0.0010	1.19	0.0029	3.30		

Panel B: Book size and junk double sorted portfolios

Portfolio	Small	2	3	4	Big	Small - Big
Excess Returns (%)						
Junk	0.71	0.45	0.30	0.33	0.23	0.49
2	1.11	0.85	0.67	0.65	0.55	0.56
3	1.22	0.93	0.84	0.77	0.60	0.62
4	1.14	0.94	0.82	0.79	0.67	0.48
Quality	1.07	0.95	0.91	0.83	0.73	0.34
Quality - Junk	0.35	0.50	0.60	0.49	0.50	
t-statistics for excess returns						
Junk	1.97	1.31	0.93	1.08	0.74	1.96
2	3.44	2.96	2.59	2.73	2.47	2.31
3	4.05	3.59	3.61	3.70	3.21	2.63
4	4.18	3.83	3.83	3.98	3.69	2.30
Quality	4.15	4.11	4.33	4.34	4.27	1.84
Quality - Junk	2.06	2.75	3.38	2.74	2.44	
Returns adjusted for RMRF, HML, UMD (%)						
Junk	0.37	0.13	-0.04	-0.08	-0.26	0.63
2	0.75	0.50	0.22	0.13	-0.03	0.78
3	0.82	0.49	0.35	0.25	0.04	0.79
4	0.75	0.53	0.34	0.28	0.13	0.62
Quality	0.68	0.55	0.48	0.38	0.30	0.38
Quality - Junk	0.31	0.42	0.52	0.46	0.56	
t-statistics for returns adjusted for RMRF, HML, UMD						
Junk	1.31	0.54	-0.18	-0.45	-1.70	2.65
2	3.05	2.72	1.50	1.12	-0.41	3.24
3	3.67	3.15	2.87	2.69	0.61	3.40
4	3.86	3.56	3.16	3.29	2.37	3.05
Quality	3.95	4.38	4.92	4.94	5.60	2.03
Quality - Junk	1.82	2.39	3.12	2.89	3.26	

large stocks, even when size is measured without using market prices.

The challenge that the size premium shows up only for market price-based measures of size is met by controlling for quality or junk. Doing so, we find a healthy, and at least

as large (if not larger) size premium associated with portfolios sorted on non-price-based measures of size that is evident in different sample periods, within 30 different industries, and across 23 other international equity markets. This evidence rejects the notion that the size premium is

Table 6

Level and costs of shorting for size and junk.

Reported are pooled averages of shorting activity and shorting costs for the 25 portfolios sorted on size (market cap) and quality or junk from Table 3. We report the daily cost of borrow score (DCBS), a number from 1 to 10 indicating the fee for each stock (Panel A) and the lendable shares defined as the supply of lendable shares divided by the market cap for each stock, expressed as a percentage (Panel B). The data cover the period January 2010 to October 2016.

Portfolio	Small	2	3	4	Big
<i>Panel A: DCBS score</i>					
Junk	4.4	3.0	2.5	2.3	2.3
2	3.0	1.9	1.7	1.6	1.7
3	2.1	1.5	1.3	1.2	1.4
4	1.6	1.3	1.2	1.2	1.3
Quality	1.3	1.1	1.1	1.1	1.0
<i>Panel B: Lendable shares (%)</i>					
Junk	3.1	3.7	3.6	3.1	3.3
2	9.5	12.9	12.8	12.0	11.6
3	15.6	18.9	21.5	22.6	21.0
4	18.0	22.0	23.7	25.0	24.9
Quality	17.3	20.9	22.0	22.6	23.6

driven by the effect of time-varying risk premia on market-based measures of size.

4.3. Behavioral versus risk-based theories: characteristics of size-quality portfolios

Our results challenge standard risk-based theories of asset pricing, but what about the behavioral theories? Behavioral finance suggests that stocks can be mispriced and more so for small stocks with greater limits of arbitrage (e.g., greater implementation costs). However, controlling for the level of mispricing, the behavioral theories do not predict any size effect. Hence, the fact that we find a strong size effect when controlling for value and quality factors is puzzling in light of this reading of the behavioral theories, although we should acknowledge that future research could come up with a new behavioral theory to match these findings.

The size effect is about a particular kind of mispricing in a particular direction. That is, small stocks deliver higher returns than large stocks, or small stocks are cheap relative to large stocks, *ceteris paribus*. The parts of behavioral finance that speak most directly to such a directional mispricing is the literature on short-selling frictions and price-optimism models caused by dispersion in opinion. These theories hypothesize that stocks that are more difficult to sell short and have greater dispersion of opinion about their value are more likely to become overvalued (Miller, 1977).

Table 6, Panel A, shows that small stocks have higher short-selling costs than large stocks over the sample period from January 2010 to October 2016. We compare the daily cost of borrow score (DCBS) from Markit Data Explorers, a number from 1 to 10 indicating the relative fee to short the stock, for our 25 size-quality portfolios, and find the average score for smallcap stocks to be higher than the

average score for large cap stocks.¹⁸ This is especially true for the smallest and junkiest stocks. When controlling for the quality of a stock, the same pattern emerges in which the smallest stocks have the highest shorting costs.

Panel B shows the amount of shorting for small versus large cap stocks, measured by the lendable shares percentage, defined as the supply of lendable shares divided by the market cap of each stock, over the sample period from January 2010 to October 2016. Small stocks on average are shorted less than large stocks, especially the smallest and junkiest stocks. Given this evidence, according to the behavioral theories, smaller stocks should be more overvalued, which is counter to the evidence on the size effect. The smallest, junkiest stocks have the worst expected returns, consistent with them being the most overvalued. For both Panels A and B of Table 6, the junkiest stocks are the most expensive to short and the least shorted, possibly consistent with their expected returns being poor according to the behavioral theory. However, the costs of shorting and amount of shorting seem to go the opposite way for explaining why small stocks have higher expected returns than large stocks.

To further test whether larger shorting constraints lead to smaller stocks becoming more overvalued as in Miller (1977), we examine the past five-year returns for our 25 size-quality portfolios. DeBondt and Thaler (1985) argue that markets exhibit an overreaction behavioral bias as evidenced by the return reversals that occur at five-year horizons. They find that past five-year winners are overvalued and tend to subsequently revert to lower valuations, while past five-year losers are undervalued and tend to subsequently revert to higher valuations. Thus, if small stocks are overvalued, we would expect to see high past five-year returns. Table 7, Panel A, presents contradictory evidence showing that the smaller stocks have experienced the weakest past five-year returns on average. This is especially true for the smallest and junkiest stocks. Here, we find little support that small stocks are overvalued. Moreover, we find that the worst past five-year performance is in the junkiest stocks, particularly the small junky stocks. So, despite facing large shorting costs and few lendable shares, these stocks do not appear to be overvalued, as Miller (1977) predicts.

The price-optimism models starting with Miller (1977) also suggest that when greater disagreement exists about the value of a stock, the stock becomes more overvalued, leading to lower subsequent returns. Following Diether et al. (2002), we use the dispersion in analysts' forecasts of earnings as our measure of disagreement. Panels B and C of Table 7 present two different versions of the measure of dispersion. Panel B is normalized by the absolute value of the mean forecast, as in Diether et al. (2002). Panel C normalizes by the price of the stock to avoid issues of having zero in the denominator which can affect the measure in Panel B. Based on the evidence

¹⁸ We obtain similar results using other measures of shorting costs from Markit Data Explorers. We use other cost measures, such as the average fee of stock borrow transactions from hedge funds, expressed as a percentage (SAF), and the value-weighted average cost, a number from one to five indicating the fee for each stock (VWAF).

Table 7

Past five-year returns and analysts' earnings forecasts dispersion for size and junk.

Reported are average statistics on past five-year returns and dispersion of earnings for the 25 portfolios sorted on size (market cap) and quality or junk from Table 3. We report the past five-year returns as an indication of market overreaction (Panel A) and two measures of dispersion in analysts' earnings forecasts, one normalized by average forecast (Panel B) and one normalized by stock price (Panel C). The data cover the period January 1980 to December 2012.

Portfolio	Small	2	3	4	Big
<i>Panel A: Past five-year returns (%)</i>					
Junk	−7.2	−2.0	4.0	9.3	15.8
2	2.8	9.8	15.4	21.6	28.7
3	11.1	17.5	22.1	27.9	36.4
4	19.3	23.0	27.4	31.5	37.7
Quality	25.4	24.5	28.1	32.2	38.2
<i>Panel B: Earnings dispersion/mean earnings (%)</i>					
Junk	67	71	61	74	80
2	69	55	42	35	26
3	58	41	29	21	14
4	49	31	18	11	8
Quality	46	25	12	7	5
<i>Panel C: Earnings dispersion/price (%)</i>					
Junk	40	13	8	6	4
2	13	4	3	2	1
3	5	2	1	1	1
4	3	1	1	1	0
Quality	2	1	1	0	0

in Panels B and C, smaller and junkier stocks have greater levels of dispersion which should lead to overvaluation of those stocks. While junkier stocks have lower subsequent returns than quality stocks, consistent with them being overvalued according to this theory, small stocks outperform large stocks. Once again, the behavioral theories fail to predict the size effect, in which the overvaluation theories go the wrong way.

4.4. Liquidity-based theories: transaction costs and the level of liquidity

We next consider theories for the pricing of liquidity (Amihud and Mendelson, 1986) and liquidity risk (Acharya and Pedersen, 2005). Per these theories, investors prefer stocks that are more liquid and face less liquidity risk, all else equal (i.e., assuming similar risk exposures). Therefore, more illiquid stocks and stocks with more liquidity risk must offer higher expected returns. Because small stocks are more illiquid and face more liquidity risk, these theories can explain the size effect. Investors who buy small stocks must be compensated for their relatively high transaction costs and for the risk that these transaction costs rise unexpectedly when liquidity is most desired. This compensation is in fact the size premium, according to these theories.

We first consider the level of liquidity across our size and quality portfolios. Table 8 reports two measures of liquidity across the 25 size-junk portfolios, namely the average percentage half bid-ask spread (Panel A) and the market impact cost per dollar traded (Panel B). The market im-

Table 8

Liquidity level: size, junk, and trading costs.

Reported are pooled averages of liquidity and trading cost measures for the 25 portfolios sorted on size (market cap) and quality or junk from Table 3. We report the average half bid-ask spread as a percentage of share price (Panel A), which is half of the bid minus ask price divided by the mid-price, and the market impact cost per dollar traded estimated from Frazzini et al. (2013) assuming a constant fund net asset value (NAV) of \$1 billion plus one half of the effective bid-ask spread, all expressed in basis points (Panel B). The trading cost data cover the period January 2000 to December 2012.

	Small	2	3	4	Big
<i>Panel A: Bid/ask spread (%)</i>					
Junk	3.1	0.9	0.3	0.2	0.1
2	3.4	1.0	0.3	0.2	0.1
3	3.8	1.1	0.3	0.1	0.1
4	3.4	1.2	0.3	0.1	0.1
Quality	2.5	1.2	0.3	0.1	0.1
<i>Panel B: Market impact cost per dollar traded (bps)</i>					
Junk	33.98	20.46	15.50	12.47	6.61
2	35.76	21.10	15.51	12.09	5.70
3	38.15	21.74	15.49	12.08	4.88
4	36.43	22.34	15.56	12.02	4.50
Quality	33.04	22.14	15.58	11.89	4.42

pact costs are estimated from Frazzini et al. (2013), who use live proprietary trading data from a large institutional trader to calibrate a trading cost model. The model computes price impact for a given fund size (net asset value, NAV). For the present calculations, we assume a constant fund NAV of \$1 billion in capital for use in the trading cost model.¹⁹

As Table 8 shows, small firms face larger percentage bid-ask spreads than large firms consistently across the quality spectrum. However, the bid-ask spreads are similar across the quality groups, controlling for size. Hence, liquidity seems to vary strongly with size, but not with quality. Market impact cost per dollar traded portrays a similar picture.

From this evidence, we draw two conclusions. Liquidity can help explain the size effect and is tightly connected to size, and liquidity appears unrelated to quality and not likely to explain the quality premium. Thus, two unrelated effects could be correlated with size: a liquidity effect and a quality effect. Looking at size unconditionally conflates liquidity and quality, just as it does size and quality. As such, any liquidity-related size effect should become stronger when controlling for quality by cleaning up the relation between size and liquidity. Consistent with this notion, the spread in liquidity across small versus large stocks is similar across quality groups, highlighting a clear pattern between size and liquidity within each quality group. From our previous evidence, we know that

¹⁹ The price impact trading cost data used by Frazzini et al. (2014) to calibrate their model and estimate its parameters pertains to stocks that are predominantly in the top 60% of market capitalization, because their execution database does not trade stocks below this threshold. While the model includes a size factor to allow the trading costs to scale with the size of the firm and its average daily trading volume, the functional form of this relation is estimated only from the larger stocks in their database. Therefore, extrapolating estimates of price impact costs for the smallest, particularly micro-cap stocks, could be less accurate.

Table 9

Can liquidity risk explain size controlling for quality?

Panel A reports regression results for the size premium (small minus big, SMB) on the market return (RMRF), high minus low (HML), up minus down (UMD), the quality factor (QMJ), and two proxies for liquidity risk. IML (illiquid minus liquid) is the return of a portfolio that is long illiquid stocks and short liquid stocks, with liquidity measured as in Amihud (2014). LIQ is the decile spread in returns from portfolios sorted on bid-ask spreads. The sample is 1957–2012, and *t*-statistics are shown in parentheses. Panel B reports returns from double sorted portfolios based on illiquidity and quality or junk. Stocks are sorted independently based on illiquidity and quality into quintiles, and the intersection of those groups form 25 illiquidity-junk portfolios. The value-weighted returns of the 25 portfolios are computed and their averages net of the market are reported along with their *t*-statistics. Two measures of illiquidity are used: Amihud's (2002) measure, which forms the basis of the IML factor used in Amihud (2014), and the market impact cost per dollar traded estimated from Frazzini et al. (2013) assuming a constant fund net asset value (NAV) of \$1 billion. Amihud's measure covers the 1957–2012 sample period, and the trading cost data cover the period August 1998 to December 2012.

Panel A: Regressions of SMB on liquidity factors						
alpha	RMRF	HML	UMD	QMJ	IML	LIQ
0.20%	0.17	−0.15	−0.00			
(1.70)	(6.49)	(−3.64)	(−0.10)			
−0.19%	0.26	−0.35	0.07		0.79	
(−2.70)	(16.01)	(−13.07)	(3.97)		(33.47)	
0.65%	−0.09	−0.24	0.06	−0.84		
(6.64)	(−3.58)	(−6.69)	(2.45)	(−18.04)		
0.16%	0.07	−0.38	0.10	−0.56	0.68	
(2.72)	(4.44)	(−17.65)	(7.14)	(−19.55)	(34.94)	
0.12%	0.06	−0.34	0.09	−0.51	0.70	0.10
(2.03)	(3.54)	(−15.53)	(6.82)	(−17.59)	(36.56)	(6.53)
Panel B: Illiquidity and junk double sorted portfolios						
Portfolio	Illiquid	2	3	4	Liquid	Illiquid - Liquid
Adjusted returns to Amihud (2002, 2014) and junk-sorted portfolios (%)						
Junk	−0.62	−0.52	−0.38	−0.41	−0.52	−0.10
2	0.00	−0.05	−0.08	−0.05	−0.25	0.25
3	0.24	0.02	0.11	0.07	−0.11	0.34
4	0.38	0.26	0.22	0.17	−0.03	0.41
Quality	0.20	0.24	0.27	0.22	0.10	0.10
Quality - Junk	0.82	0.76	0.66	0.63	0.62	
t-statistics for adjusted returns to Amihud (2002, 2014) and junk-sorted portfolios						
Junk	−3.54	−3.43	−2.54	−3.03	−3.95	−0.58
2	0.03	−0.49	−0.81	−0.56	−3.31	1.75
3	1.87	0.19	1.29	1.02	−1.72	2.44
4	3.14	2.77	2.69	2.40	−0.48	2.91
Quality	1.77	2.50	3.10	2.91	1.86	0.71
Quality - Junk	5.20	5.38	4.64	4.81	3.90	
Adjusted returns to price impact and junk-sorted portfolios (%)						
Junk	−0.31	−1.46	−1.12	−0.85	−1.18	0.87
2	0.13	−0.19	−0.43	−0.25	−0.60	0.72
3	−0.18	0.06	−0.17	0.03	−0.39	0.22
4	0.24	0.11	0.04	0.18	−0.35	0.59
Quality	0.23	0.12	0.07	0.27	−0.23	0.47
Quality - Junk	0.54	1.58	1.19	1.12	0.94	
t-statistics for adjusted returns to price impact and junk-sorted portfolios						
Junk	−0.54	−4.04	−3.74	−3.20	−4.33	1.56
2	0.36	−0.77	−2.21	−1.44	−4.52	2.04
3	−0.83	0.33	−1.04	0.23	−3.54	1.05
4	1.19	0.66	0.26	1.21	−4.19	2.91
Quality	1.02	0.80	0.45	2.10	−2.77	1.87
Quality - Junk	0.95	4.63	4.45	4.70	3.21	

the spread in returns across small versus large stocks is also cleaned up when controlling for quality and is similar across quality groups.

This evidence is consistent with a liquidity-driven story for the size effect, where quality is a separate factor related to size but unrelated to liquidity that confounds the relation between size and liquidity and average returns. While the evidence in Table 8 is consistent with separate liquidity and quality effects related to size, the results do not provide any further direct evidence of a liquidity story for size per se. In addition, the variation in liquidity across

the quality dimension within each size group is negligible, consistent with quality not being driven by liquidity.²⁰

²⁰ The only place where a hint of variation could exist in liquidity measures across quality is among the smallest stocks based on bid-ask spread and among the largest stocks for price impact, in which the junkiest firms appear a tiny bit less liquid than the quality firms. The relation between quality and these liquidity measures, however, is not consistent within a size group. For bid-ask spread, it is the highest quality firms among the smallest quintile of stocks that appear different from the other quality quintiles within that group. For price impact, it is the lowest quality firms among the largest stocks that appear different from the other

4.5. Liquidity-based theories: liquidity risk

Some researchers have found that liquidity risk can help explain the size effect (Pastor and Stambaugh, 2003; Acharya and Pedersen, 2005). Absent controlling for quality, the size effect is fairly weak and, therefore, is not very difficult to explain with other correlated factors such as liquidity risk. However, controlling for quality, the size premium is substantially stronger, raising the question of whether liquidity risk can help explain these much larger returns.

To address this question, we regress the return to the size factor on standard factors, quality, and liquidity factors. A popular liquidity measure is the price impact measure of Amihud (2002), which is used by Acharya and Pedersen (2005) to construct a liquidity risk factor. To easily interpret the alphas, we use the traded version of this liquidity factor, relying on the illiquid minus liquid (IML) factor of Amihud (2014). We also construct a simple liquidity factor that we denote LIQ, which is the equal-weighted average of the decile return spread in portfolios sorted on turnover and on bid-ask spreads. [We do not use the Pastor and Stambaugh (2003) measure because it is not available over the full sample period, though we obtain similar results over the shorter sample period.]

Panel A of Table 9 reports the results of these regressions over the full quality sample, 1957–2012. The first regression shows that when we control only for the market, value, and momentum, the size effect is weak (t -statistic = 1.7). When we control for liquidity risk (in the second regression of the table) based on IML, we find a negative and significant alpha.

Next, the third and fourth regressions show that, when we control for quality, the alpha becomes significantly positive at 65 bps (with a t -statistic of 6.64). However, when we control for liquidity risk using the IML factor, we see a large drop in the alpha from 65 bps to only 16 bps. Controlling for both liquidity risk factors, IML and LIQ, we see a further drop in the alpha (the last regression in the table) to 12 bps, which is marginally statistically significant (t -statistic of 2.03). This marginally positive alpha could suggest that liquidity risk is not the full explanation for the size premium (once we control for quality) or that the liquidity proxies we use are measured with error and perhaps more precise liquidity measures would drive the alpha to zero. In addition, Acharya and Pedersen (2005) argue that investors want compensation for both the average illiquidity level and liquidity risk of an asset. Controlling for liquidity risk only is often a quick way to capture both liquidity levels and risk (because stocks that have high liquidity risk are likely stocks that are also illiquid so the estimated liquidity risk premium can be viewed as the sum of compensation for liquidity level and risk), but it is also possible that the level effect is not fully captured here.

Panel B of Table 9 reports results from double sorted portfolios based on illiquidity and quality or junk, in which

stocks are sorted independently based on illiquidity and quality into quintiles and the returns in excess of the market are reported on the 25 portfolios formed from the interaction between these sorts. Two measures of illiquidity are used: Amihud (2002), which forms the basis of the IML factor in Amihud (2014), and the market impact measure of Frazzini et al. (2013). The former covers the full sample period from 1957 to 2012, and the latter covers the 1998 to 2012 period. For both measures of illiquidity, evidence exists of an illiquidity premium within each quality category that is directionally consistent with the size premium within each quality category from similar double-sorts in Tables 3 and 5. A significant quality premium remains within each illiquidity category, just as within each size category. Hence, this evidence suggests that measures of illiquidity interact with quality or junk much like the size measures do and that size and illiquidity are very much related and possibly picking up the same return effects.

In summary, small stocks have high bid-ask spreads, have high market impact, and face high liquidity risk. These liquidity level and risk effects can help explain most, or perhaps even all, of the size effect, even when controlling for quality.

5. Conclusion

Size matters, and in a much bigger way than previously thought, after controlling for quality or junk. We find that previous evidence on the variability of the size effect is largely due to the volatile performance of small, low-quality junky firms. Controlling for junk, a much stronger and stable size premium emerges that is robust across time (including periods when the size effect seems to fail), monotonic in size and not concentrated in the extremes, prevalent across months of the year, existent even for non-market-price-based measures of size, and present internationally across nearly two dozen countries. These results are robust across a variety of quality measures, sample periods, within industries, and 23 international markets.²¹

Our evidence that shows a resurgence of the size effect after controlling for quality helps distinguish among many theories for why a size premium could exist. The fact that the size premium rises after controlling for other known risk factors, including growth factors, seems at odds with a simple risk-based explanation or rational explanation based on growth options. The fact that non-price-based measures of size deliver a return premium at least

quality quintiles. Moreover, no consistency is evident across the other size quintiles (2 through 4), in which no discernible relation between quality and liquidity is apparent.

²¹ Combining these results, Figures OA4 and OA5 in the Online Appendix show seasonal patterns in size controlling for quality across time, among the smallest stocks, and using non-price-based size measures. Controlling for quality or junk has the effect of smoothing the returns to size, establishing a clear size premium that is no longer concentrated in January and, most importantly, is significant outside of January and no longer concentrated in certain time periods or for certain measures. The behavior of small, junk firms varies substantially and is chiefly responsible for diluting the size effect at certain times, for certain months, and for certain measures and exaggerating it at other times. Controlling for these firms through exposure to a quality factor, the size premium is robust over time, and for different measures of size, even those not based on market prices.

as large as that from market capitalization-based measures when controlling for quality or junk is inconsistent with time-varying risk premia models for the size effect as suggested by Berk (1995a). The fact that measures of limited arbitrage activity such as shorting costs and measures of disagreement such as analyst forecast dispersion suggest that small stocks should be overpriced, implying the opposite of the strong positive size premium we find. Hence, behavioral theories based on limits to arbitrage and overvaluation do not appear to explain our findings. Finally, we show that measures of liquidity and liquidity risk are strongly correlated to and significantly reduce the size premium, even after controlling for quality or junk. This evidence is consistent with liquidity risk-based theories for the size effect. However, a smaller but still significant size premium remains even after controlling for liquidity and liquidity risk, which suggests that either our liquidity measures contain error or liquidity is only a partial explanation for the size effect.

Finally, an interesting avenue for further research is to explore why small, junk stocks in particular do so poorly and why size is strongly negatively correlated with quality. Research by Asness et al. (2014) and Bouchard et al. (2016) seeks to explain the quality premium, with potential implications for its interaction with size.

Our focus in this paper is on the size effect, and our examination of its interaction with quality is aimed at providing further evidence in favor of or against various theories for the size effect. Our results revive the size anomaly, putting it on a more equal footing with other anomalies such as value and momentum in terms of its efficacy, and dismiss several previous explanations and challenges to the size effect. Thus, size, controlling for quality or junk, should be restored as one of the central cross-sectional empirical anomalies for asset pricing theory to explain.

References

- Acharya, V., Pedersen, L.H., 2005. Asset pricing with liquidity risk. *J. Financ. Econ.* 77, 375–410.
- Altman, E.I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *J. Finance* 23, 589–609.
- Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time series effects. *J. Financ. Mark.* 5, 31–56.
- Amihud, Y., 2014. The Pricing of the Illiquidity Factor's Systematic Risk. New York University, Stern School of Business Unpublished working paper.
- Amihud, Y., Mendelson, H., 1986. Asset pricing and the bid-ask spread. *J. Financ. Econ.* 17, 223–249.
- Ang, A., Hodrick, R., Xing, Y., Zhang, X., 2006. The cross-section of volatility and expected returns. *J. Finance* 61, 259–299.
- Ang, A., Hodrick, R., Xing, Y., Zhang, X., 2009. High idiosyncratic volatility and low returns: international and further U.S. evidence. *J. Financ. Econ.* 91, 1–23.
- Asness, C.S., Frazzini, A., 2012. The devil in HML's details. *J. Port Manage.* 39, 49–69.
- Asness, C.S., Frazzini, A., Pedersen, L.H., 2014. Quality minus junk. AQR Capital Management, Greenwich, CT Unpublished working paper.
- Asness, C.S., Frazzini, A., Pedersen, L.H., 2017. Quality Minus Junk. AQR Capital Management, Greenwich, CT Unpublished working paper.
- Asness, C., Krail, R.J., Liew, J.M., 2001. Do hedge funds hedge? *J. Portf. Manage.* 28, 6–19.
- Asness, C.S., Moskowitz, T.J., Pedersen, L.H., 2013. Value and momentum everywhere. *J. Finance* 68, 929–985.
- Baker, M., Wurgler, J., 2002. Market timing and capital structure. *J. Finance* 57, 1–32.
- Ball, R., 1978. Anomalies in relations between securities' yields and yield-surrogates. *J. Financ. Econ.* 6, 103–126.
- Banz, R., 1981. The relationship between return and market value of common stocks. *J. Financ. Econ.* 9, 3–18.
- Barberis, N., Shleifer, A., Vishny, R., 1998. A model of investor sentiment. *J. Financ. Econ.* 49, 307–343.
- Bebchuk, L., Cohen, A., Ferrell, A., 2009. What matters in corporate governance? *Rev. Financ. Stud.* 22, 783–827.
- Berk, J., 1995a. A critique of size-related anomalies. *Rev. Financ. Stud.* 8, 275–286.
- Berk, J., 1995b. An Empirical Reexamination of the Relation Between Firm Size and Return. University of Washington Unpublished working paper.
- Berk, J., 1997. Does size really matter? *Financ. Anal. J.* 53, 12–18.
- Black, F., 1972. Capital market equilibrium with restricted borrowing. *J. Bus.* 45, 444–455.
- Black, F., 1992. Beta and return. *J. Portf. Manage.* 20, 8–18.
- Black, F., Jensen, M.C., Scholes, M., 1972. The capital asset pricing model: some empirical tests. In: Jensen, M.C. (Ed.), *Studies in the Theory of Capital Markets*. Praeger Publishers, New York, pp. 79–121.
- Bouchaud, J.-P., Ciliberti, S., Landier, A., Simon, G., Thesmar, D., 2016. The Excess Returns of “Quality” Stocks: A Behavioral Anomaly. HEC Paris, Paris, France Research paper no. FIN-2016-1134.
- Brennan, M., Subrahmanyam, A., 1996. Market microstructure and asset pricing: on the compensation for illiquidity in stock returns. *J. Financ. Econ.* 49, 441–464.
- Bryan, A., 2014. Does the Small-Cap Premium Exist?. Morningstar Inc., Chicago, IL.
- Campbell, J., Hilscher, J., Szilagyi, J., 2008. In search of distress risk. *J. Finance* 63, 2899–2939.
- Carlson, M., Fisher, A., Giammarino, R., 2004. Corporate investment and asset price dynamics: implications for the cross-section of returns. *J. Finance* 59, 2577–2603.
- Chan, L.K.C., Karceski, J., Lakonishok, J., 2000. New paradigm or same old hype in equity investing? *Financ. Anal. J.* 56, 23–36.
- Chordia, T., Subrahmanyam, A., Tong, Q., 2014. Have capital market anomalies attenuated in the recent era of high liquidity and trading activity? *J. Account. Econ.* 58, 41–58.
- Cooper, M., Gulen, H., Schill, M., 2008. Asset growth and the cross-section of stock returns. *J. Finance* 63, 1609–1651.
- Core, J.E., Guay, W.R., Rusticus, T.O., 2006. Does weak governance cause weak stock returns? An examination of firm operating performance and investors' expectations. *J. Finance* 61, 655–687.
- Crain, M.A., 2011. A Literature Review of the Size Effect. Financial Valuation Group, Tampa, FL Unpublished working paper.
- Cremers, M., Ferrell, A., 2014. Thirty years of shareholder rights and firm value. *J. Finance* 69, 1167–1196.
- Cremers, M., Nair, V., 2005. Governance mechanisms and equity prices. *J. Finance* 60, 2859–2894.
- Daniel, K., Hirshleifer, D., Subrahmanyam, A., 1998. A theory of overconfidence, self-attribution, and security market under and over-reactions. *J. Finance* 53, 1839–1885.
- De Bondt, W.F.M., Thaler, R., 1985. Does the stock market overreact? *J. Finance* 40, 793–805.
- Dichev, I.D., 1998. Is the risk of bankruptcy a systematic risk? *J. Finance* 53, 1131–1148.
- Diether, K., Malloy, C., Scherbina, A., 2002. Differences of opinion and the cross section of stock returns. *J. Finance* 57, 2113–2141.
- Easterday, K., Sen, P., Stephan, J., 2009. The persistence of the small firm/January effect: is it consistent with investors' learning and arbitrage efforts? *Q. Rev. Econ. Finance* 49, 1172–1193.
- Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. *J. Financ. Econ.* 33, 3–56.
- Fama, E.F., French, K.R., 1996. Multifactor explanations of asset pricing anomalies. *J. Finance* 51, 55–84.
- Fama, E.F., French, K.R., 1997. Industry costs of equity. *J. Financ. Econ.* 43, 153–193.
- Fama, E.F., French, K.R., 2012. Size, value, and momentum in international stock returns. *J. Financ. Econ.* 105, 457–472.
- Fama, E.F., French, K.R., 2015. A five-factor asset pricing model. *J. Finance* 116, 1–22.
- Fama, E.F., French, K.R., 2016. Dissecting anomalies with a five-factor model. *Rev. Financ. Stud.* 29, 69–103.
- Frazzini, A., Israel, R., Moskowitz, T.J., 2013. Trading costs of asset pricing anomalies. University of Chicago, Chicago, IL Unpublished working paper.
- Frazzini, A., Pedersen, L.H., 2014. Betting against beta. *J. Financ. Econ.* 111, 1–25.
- Garleanu, N., Panageas, S., Yu, J., 2012. Technological growth and asset pricing. *J. Finance* 67, 1265–1292.

- George, T.J., Hwang, C.Y., 2010. A resolution of the distress risk and leverage puzzles in the cross section of stock returns. *J. Financ. Econ.* 96, 56–79.
- Gibbons, M.R., Ross, S.A., Shanken, J., 1989. A test of the efficiency of a given portfolio. *Econometrica* 57, 1121–1152.
- Giroud, X., Mueller, H.M., 2011. Corporate governance, product market competition, and equity prices. *J. Finance* 66, 563–600.
- Gompers, P., Metrick, A., 2001. Institutional investors and equity prices. *Q. J. Econ.* 116, 229–259.
- Gompers, P., Ishii, J., Metrick, A., 2003. Corporate governance and equity prices. *Q. J. Econ.* 118, 107–156.
- Graham, B., Dodd, D., 1934. *Security Analysis*. McGraw-Hill, New York.
- Gu, A.Y., 2003. The declining January effect: evidences from the U.S. equity markets. *Q. Rev. Econ. Finance* 43, 395–404.
- Harvey, C., Liu, Y., Zhu, H., 2014. . . . and the Cross-Section of Expected Returns. Duke University, Durham, NC Unpublished working paper.
- Hong, H., Stein, H., 1999. A unified theory of underreaction, momentum trading, and overreaction in asset markets. *J. Finance* 54, 2143–2184.
- Horowitz, J.L., Loughran, T., Savin, N.E., 2000. Three analyses of the firm size premium. *J. Empir. Finance* 7, 143–153.
- Hou, K., Moskowitz, T.J., 2005. Market frictions, price delay, and the cross-section of expected returns. *Rev. Financ. Stud.* 18, 981–1020.
- Hou, K., Van Dijk, M.A., 2017. Resurrecting the Size Effect: Firm Size, Profitability Shocks, and Expected Stock Returns. Ohio State University, Columbus, OH Unpublished working paper.
- Ibbotson, R., Chen, Z., Kim, D.Y.J., Hu, W., 2013. Liquidity as an investment style. *Financ. Anal. J.* 69, 30–44.
- Israel, R., Moskowitz, T.J., 2013. The role of shorting, firm size, and time on market anomalies. *J. Financ. Econ.* 108, 275–301.
- Johnson, S., Moorman, T., Sorescu, S., 2009. A reexamination of corporate governance and equity prices. *Rev. Financ. Stud.* 22, 4753–4786.
- Keim, D.B., 1983. Size-related anomalies and stock return seasonality: further empirical evidence. *J. Financ. Econ.* 12, 13–32.
- Lakonishok, J., Shleifer, A., Vishny, R.W., 1994. Contrarian investment, extrapolation and risk. *J. Finance* 49, 1541–1578.
- Larcker, D.F., Reiss, P.C., Xiao, Y., 2015. Corporate Governance Data and Measures Revisited. Stanford Graduate School of Business, Stanford, CA Unpublished working paper no. 3379.
- Lo, A., MacKinlay, C., 1988. Stock market prices do not follow random walks: evidence from a simple specification test. *Rev. Financ. Stud.* 1, 41–66.
- McLean, R.D., Pontiff, J., 2016. Does academic research destroy stock return predictability? *J. Finance* 71, 5–32.
- Mohanram, P.S., 2005. Separating winners from losers among low book-to-market stocks using financial statement analysis. *Rev. Account. Stud.* 10, 133–170.
- Miller, E.M., 1977. Risk, uncertainty, and divergence of opinion. *J. Finance* 32, 1151–1168.
- Novy-Marx, R., 2013. The other side of value: the gross profitability premium. *J. Financ. Econ.* 108, 1–28.
- Ohlson, J., 1980. Financial ratios and the probabilistic prediction of bankruptcy. *J. Account. Res.* 1980, 109–131.
- Pastor, L., Stambaugh, R.F., 2003. Liquidity risk and expected stock returns. *J. Polit. Econ.* 111, 642–685.
- Penman, S., Richardson, S., Tuna, I., 2007. The book-to-price effect in stock returns: accounting for leverage. *J. Account. Res.* 45, 427–467.
- Pontiff, J., Woodgate, W., 2008. Share issuance and cross-sectional returns. *J. Finance* 63, 921–945.
- Reinganum, M., 1983a. Portfolio strategies based on market capitalization. *J. Portf. Manag.* 9, 29–36.
- Reinganum, M., 1983b. The anomalous stock market behavior of small firms in January: empirical tests for tax-loss selling effects. *J. Financ. Econ.* 12, 89–104.
- Richardson, S., Sloan, R.G., Soliman, M., Tuna, I., 2005. Accrual reliability, earnings persistence and stock prices. *J. Account. Econ.* 39, 437–485.
- Roll, R., 1983. The turn-of-the-year effect and the return premia of small firms. *J. Portf. Manag.* 9, 18–28.
- Schwert, G.W., 2003. Anomalies and market efficiency. In: Constantinides, G., Harris, M., Stulz, R. (Eds.), *Handbook of the Economics of Finance*, 1. Elsevier, pp. 939–974.
- Sadka, R., 2006. Momentum and post-earnings-announcement drift anomalies: the role of liquidity risk. *J. Financ. Econ.* 80, 309–349.
- Shleifer, A., Vishny, R.W., 1997. The limits of arbitrage. *J. Finance* 52, 35–55.
- Shumway, T., 1997. The delisting bias in CRSP data. *J. Finance* 52, 327–340.
- Sloan, R.G., 1996. Do stock prices reflect information in accruals and cash flows about future earnings? *Account. Rev.* 71, 289–315.
- Van Dijk, M.A., 2013. Is size dead? A review of the size effect in equity returns. *J. Bank. Finance* 35, 3263–3274.