Quality Investing

Robert Novy-Marx

Buying high quality assets without paying premium prices is just as much value investing as buying average quality assets at discount prices. Strategies that exploit the quality dimension of value can be profitable on their own, and accounting for both dimensions of value yields dramatic performance improvements over traditional value strategies. Gross profitability is particularly powerful among popular quality notions, especially among large cap stocks and for long-only investors.

What is quality investing, and how should quality be measured? Quality, unlike value, has no universally accepted definition. Value strategies are easily identified by the fact they hold stocks with low valuations. Quality strategies, in contrast, are generally identified using something more akin to Supreme Court Justice Potter Stewart's obscenity doctrine of "you know it when you see it" (*Jacobellis v. Ohio,* 1964). This paper attempts to identify commonalities across seven of the best know quality strategies. It also looks for differences, running a performance horse race between alternative quality strategies.

Quality is often marketed as an attractive alternative to traditional growth strategies, which performed terribly both during and after the NASDAQ deflation in the early 2000s. Its leading industry proponents include GMO's Jeremy Grantham, whose high quality indicators of "high return, stable return, and low debt" have shaped the design of MSCI's Quality Indices, and Joel Greenblatt, whose "Little Book that Beats the Market" has encouraged a generation of value investors to pay attention to capital productivity, measured by return on invested capital, in addition to valuations.

There has also been increased interest in incorporating academic measures of quality into value strategies. BlackRock, an early adopter (while still Barclays Global Investors) of Sloan's (1996) accruals-based measure of earnings quality, is currently promoting the

Robert Novy-Marx is Lori and Alan S. Zekelman Professor of Business Administration at the Simon Graduate School of Business at the University of Rochester, New York, and a research associate of the National Bureau of Economic Research.

benefits of integrating earnings quality into global equities strategies (Kozlov and Petajisto, 2013). Piotroski and So (2012) argue that strategies formed jointly on valuations and another accounting based measure of financial strength, the Piotroski's (2000) F-score (which uses both Sloan's accruals and aspects of Grantham's quality among its nine subcomponents), have dramatically outperformed traditional value strategies. Societe General has appropriated Piotroski's F-score (without attribution) as the primary screen it employs when constructing its Global Quality Income Index, launched in 2012 (Lapthorne et. al., 2012).

Novy-Marx (2013) finds that a simpler quality measure, gross profitability (revenues minus cost of goods sold, scaled by assets), has as much power predicting stock returns as traditional value metrics. Strategies based on gross profitability are highly negatively correlated with strategies based on price signals, making them particularly attractive to traditional value investors. Novy-Marx's results have influenced the design of both DFA's growth funds and AQR Capital Management's core equity funds.

One common recurring theme is the strong relation between quality and value. The two are quite similar, philosophically. Quality can even be viewed as an alternative implementation of value—buying high quality assets without paying premium prices is just as much value investing as buying average quality assets at a discount. Warren Buffett, Graham's most famous student and the most successful value investor of all time, is fond of saying that it is "far better to buy a wonderful business at a fair price than to buy a fair business at a wonderful price." In fact, Frazzini, Kabiller, and Pedersen (2012) show that the performance of the publicly traded companies held by Berkshire Hathaway, Buffett's primary investment vehicle, can largely be explained by his commitment to buying high quality stocks.

Quality and value strategies are highly dissimilar, however, in the stocks that they actually hold. High quality firms tend to be expensive, while value firms tend to be low quality. Quality strategies are thus short value, while value strategies are short quality. Each of these strategies consequently tends to do well precisely when the other under-performs, making

them exceptionally attractive to run together. With these claims regarding the synergies between quality and value, and the varied and disparate nature of strategies masquerading as "quality," it is natural to ask which quality measures best help investors design successful investment portfolios. This paper attempts to answer this question, by assessing the performance of the best known quality strategies.

It finds that all of the quality measures appear to have some power predicting returns, especially among small cap stocks, and when used in conjunction with value measures. Only gross profitability, however, generates significant excess returns as a stand alone strategy, and has the largest Fama and French (1993) three-factor alpha, especially among large cap stocks. Gross profitability also subsumes most of the power of the other measures.

The paper also shows that long-only investors need to be more careful when designing strategies to exploit quality. For these investors, whose portfolios are dominated by market risk, it is difficult to get large exposures to the attractive opportunities provided by value and quality by running value and quality strategies side by side. Long-only investors can get larger quality and value tilts, and thus achieve higher Sharpe ratios, through the integrated solution of buying only reasonably priced high quality stocks.

Measuring quality

I will be comparing the performance of strategies based on seven of the best known and most widely used notions of quality. These include Graham's quality criteria from his "Intelligent Investor," Grantham's "high return, stable return, and low debt" and Greenblatt's return on invested capital, Sloan's (1996) accruals-based measure of earnings quality and Piotroski's (2000) F-score measure of financial strength, and Novy-Marx's (2013) gross profitability. I also include the low volatility/low beta notion used by "defensive equity" strategies, which look more like a traditional value but are often marketed as high quality. Before comparing the ability of these measures to predict returns, each is discussed in greater detail below, beginning with Graham's quality criteria.

Graham quality

Today Benjamin Graham is primarily associated with value metrics like price-to-earnings or market-to-book, but Graham never advocated just buying cheap stocks. He believed in buying undervalued firms, which means buying high quality firms cheaply. In fact, Graham was just as concerned with the quality of a firm's assets as he was with the price that one had to pay to purchase them. According to Graham, an equity investor should "...apply a set of standards to each [stock] purchase, to make sure that he obtains (1) a minimum of quality in the past performance and current financial position of the company, and also (2) a minimum of quantity in terms of earnings and assets per dollar of price" (Graham 1973, pp. 183). Graham's seven quality and quantity criteria are:

1. "Adequate" enterprise size, as insulation against the "vicissitudes" of the economy;
2. Strong financial condition, measured by current ratios that exceed two and net current assets that exceed long term debt;
3. Earnings stability, measured by 10 consecutive years of positive earnings;
4. A dividend record of uninterrupted payments for at least 20 years;
5. Earnings-per-share growth of at least one-third over the last ten years;
6. Moderate price-to-earnings ratios, which typically should not exceed 15; and
7. Moderate price-to-book ratios, which typically should not exceed 1½.

The first five screens attempt to ensure that one buys only high quality firms, while the last two ensure that one buys them only at reasonable prices.

To turn these criteria into a trading strategy, I create a "Graham score" (G-score) quality index for each stock. This composite of Graham's five quality criteria gets one point if a firm's current ratio exceeds two, one point if net current assets exceed long term debt, one point if it has a ten year history of positive earnings, one point if it has a ten year history

of returning cash to shareholders, and one point if its earnings-per-share are at least a third higher than they were 10 years ago.[1] This results in a score from zero to five, with higher scores signaling higher quality firms. The quality signal employed for stock selection is the rank of a firm's G-score among the applicable universe. For more details of the construction of all the variables employed in this paper, please see Appendix A.

Grantham quality

Grantham's views on quality investing are espoused by his firm, GMO, which argues the merits of quality investing in its 2004 white paper "The Case for Quality—The Danger of Junk." This paper defines quality companies as those that meet the criteria of low leverage, high profitability, and low earnings volatility, and suggests that stocks of firms with these characteristics "have always won over longer holding periods." In a later study, "Profits for the Long Run: Affirming the Case for Quality" (Joyce and Mayer, 2012), GMO shows that since 1965 the least levered firms (lowest 25%) have had average return on equity 5% higher than the most levered firms (highest 25%), and claims that "profitability is the ultimate source of investment returns."

These ideas have been highly influential. MSCI Quality Indices, launched in December 2012, are based on Grantham's basic principles. According to MSCI, their Quality Indices "identify quality growth stocks by calculating a quality score for each security in the eligible equity universe based on three main fundamental variables: high return on equity (ROE), stable year-over-year earnings growth and low financial leverage." The Grantham criteria of "high returns, stable returns, low leverage" also make up half of the score (together with low volatility) used by Russell when constructing their Defensive Indexes, and two of the three criteria (high ROE and low leverage) form the basis of the Dow Jones

[1] This methodology is similar to that employed by Piotroski (2000) to calculate his financial strength F-score, which is investigated in greater detail in later sections. In calculating the G-score I have reduced the required earnings history from 20 to 10 years to get more variation in this component of the measure. I have also relaxed the dividend condition to include net repurchases, because share repurchases have gained popularity as a means for returning cash to shareholders. Graham also preferred large firms, but I have ignored this criterion, as this paper also considers the performance of quality strategies formed entirely within the large and small cap universes.

Quality Index.

Others have argued that the benefits of incorporating quality concerns into equity strategies accrue primarily to value investors. In a recent white paper, "Power Couple: Quality and Value are Strong Drivers of Long-Term Equity Returns" (Mead et. al., 2013), MFS Investment Management studies the performance of strategies based on Grantham's notions of quality, both as a stand-alone investment strategy and in conjunction with value. They conclude that while "... investing in quality without regard for valuation is not a compelling way to drive alpha over time... owning companies that are both high quality and inexpensively valued is... the most compelling way to generate sustainable, long-term performance."

Return on invested capital (ROIC)

Joel Greenblatt's "Little Book that Beats the Market" has been equally influential in getting investors, especially value investors, to pay attention to quality. The logic of Greenblatt's "magic formula investing" is clearly that of combining quality and value, in the spirit of Graham's belief in buying good firms at low prices. Magic formula investing entails ranking firms on the basis of return on invested capital (ROIC) and earnings yield (EY, defined as EBIT-to-enterprise value), respectively, and only buying stocks with the highest combined ranks. In Greenblatt's formula ROIC serves as the quality metric, while EY serves as the value metric. The formula is explicitly intended to ensure that investors are "buying good companies... only at bargain prices" (Greenblatt 2010, p.47).

Earnings quality

BlackRock has probably been the biggest proponent of incorporating earnings quality signals into value strategies. Sloan (1996) develops the best known and most widely used earnings quality measure. This accruals measure is the difference between cash and accounting earnings, scaled by firm assets. According to Sloan, BlackRock (then BGI) "... was the first place to really pick up on my work" (Businessweek 2007). BGI hired Sloan in 2006, presumably at least in part for his earnings quality expertise. More recently BlackRock

researchers have been promoting the benefits of trading earnings quality in conjunction with value in equity markets around the world, in a paper titled "Global Return Premiums on Earnings Quality, Value, and Size" (Kozlov and Petajisto, 2013). Strategies based on earnings quality are also readily available to long-only investors. The Forensic Accounting ETF (FLAG), for example, is designed to track the Del Vecchio Earnings Quality Index, which "uses financial statement analysis in an attempt to avoid companies with aggressive revenue recognition while investing in companies that have high earnings quality."

Financial strength

Piotroski's (2000) F-score measure of financial strength, another accounting based measure of firm quality, is also commonly employed by professional money managers and widely available on internet stock screeners. Societe General uses the F-score as its primary screen when constructing its Global Quality Income Index, while Morgan Stanley has offered products linked to strategies that combine the F-score with Greenblatt's magic formula (Ng 2009).

The F-score is constructed by summing nine binary variables, and includes elements of both Grantham's quality and Sloan's earnings quality, as well as fundamental momentum (improving earnings) and the equity issuance anomaly. Four of the variables it employs are designed to capture profitability, three to capture liquidity, and two to capture operating efficiency. Each component takes on the value zero, indicating weakness, or one, indicating strength. The F-score thus takes on a value from zero to nine, with higher numbers indicating stronger financial performance. While Piotroski (2000) originally analyzed stand-alone strategies based on the F-score, Piotroski and So (2012) shows that strategies that trade jointly on valuation and the F-score perform even better.

Defensive equity

Defensive equity strategies have become popular over the last five years, partly in response to the market's poor performance in the last quarter of 2008. These strategies promise equity like returns, delivered with less volatility and smaller drawdowns, and are often

marketed as high quality strategies.

Defensive equity strategies typically hold stocks with low volatility and low market betas. Low volatility contributes to these strategies' steady performance, while low market betas generate outperformance in down markets. Because market betas are only weakly correlated with average returns (Black, Jensen, and Scholes, 1972; Black, 1972, 1993; Frazzini and Pedersen, 2013), and high volatility stocks have actually underperformed low volatility stocks (Ang et. al. 2006; Baker, Bradley, and Wurgler, 2011), providers typically claim that these strategies mitigate market risks without sacrificing absolute performance.

Well known defensive indicies include Research Associates' RAFI Low Volatility, Sabrient's Defensive Equity, and Dow Jones' Market Neutral Anti-Beta. Investable products include both traditional mutual funds, such as AQR's and Russell's Defensive Equity Funds, and ETFs, such as Guggenheim Defensive Equity, PowerShares S&P 500 Low Volatility Portfolio, and QuantShares U.S. Market Neutral Anti-Beta Fund.

Defensive equity strategies tilt towards value, and consequently look significantly different than the other strategies considered here. They are included here because some defensive strategies sometimes select stocks using additional themes common to other quality strategies, particularly low leverage and stable return on equity, and they are often marketed to quality oriented investors.

Gross profitability

Novy-Marx (2013) shows that a simple quality metric, gross profits-to-assets, has roughly as much power predicting the relative performance of different stocks as tried-and-true value measures like book-to-price. Buying profitable firms and selling unprofitable firms, where profitability is measured by the difference between a firm's total revenues and the costs of the goods or services it sells, scaled by assets, yields a gross profitability premium. Just as importantly, the performance of strategies based on gross profitability is strongly negatively correlated with value, so profitability strategies not only deliver high average returns, but also provide a valuable hedge to value investors.

Financial economists have long believed that profitability *should* forecast returns, and

puzzled over ROE's poor performance predicting cross sectional differences in average stock performance. This belief that profitability should matter follows from the simplest of economic reasoning. A stock's current price reflects market expectations of its future payouts, discounted at the rate of return investors require to hold it. If two companies have the same expected future profitability (i.e., payoffs), but are priced differently, this must reflect the fact that investors require a higher rate of return for holding the low priced stock (Ball 1978, Berk 1995). That is, simple dividend discounting predicts the value premium. Similarly, if two firms have different expected future profitabilities, and thus different expected future payoffs, but are priced the same, this must reflect the fact that investors require a higher rate of return for holding the stock of the more profitable firm. The same economic reasoning that predicts the value premium thus also predicts a profitability premium, suggesting that the quality and value phenomena are two sides of the same coin.

These arguments for the value and profitability premiums are not predicated on investor rationality. Differences in required rates of return could partially reflect mispricings (a stock is mispriced if and only if investors require the wrong rate of return to hold it). Trading on value and profitability may thus simply be a crude but effective way of exploiting mispricings in the cross section.

Fama and French (2006) use the reasoning of the dividend discount model to motivate their empirical investigation of profitability as a stock return predictor. They find that cross-sectional regressions, which identify primarily off of small cap stocks, suggest that ROE is "related to average returns in the manner expected" (Fama and French 2006), but Fama and French (2008) find that portfolio tests, which better approximate the performance of trading strategies available to investors, "do not provide much basis for the conclusion that, with controls for market cap and B/M, there is a positive relation between average returns and profitability." The surprising fact, from the point of view of the model, is the poor empirical performance of profitability, measured by earnings, predicting returns.

Novy-Marx (2013) argues that gross profitability performs better predicting future stock returns than ROE, the profitability variable most frequently employed in earlier academic

studies, because it is a better proxy for true economic profitability. In particular, the study points to the fact that accountants treat many forms of economic investment (e.g., R&D, advertisement, sales commissions, and human capital development) as expenses, so these activities lower net income, but increase future expected profitability. This makes earnings a poor proxy for true expected economic profitability.

While analysts spend a lot of time thinking about bottom line earnings, and to a lesser extent free cash flow or EBIT, empirically gross profitability, which appears almost at the top of the income statement, is a better predictor of a firm's future stock performance. According to Chi and Fogdall (2012), the co-heads of portfolio management at Dimensional Fund Advisors, "the research breakthrough in this case is not the discovery of expected profitability as a dimension of expected returns per se, something that financial economists have suggested for quite some time... rather, it is the discovery of reasonable proxies for expected profitability, which allow us to use profitability as another dimension of expected returns in the creation of investment solutions."

Quality strategy performance

To compare the performance of these different notions of quality, strategies are constructed by ranking firms on the basis of each of the seven quality metrics.[2] Strategies include both financials (firms with one digit SIC codes of 6) and non-financials. Several of these measures look very different for financial firms and non-financial firms, however, so to avoid strong industry biases financials and non-financials are ranked separately.[3] Strategies are formed as value-weighted portfolios that hold (short) stocks in the top (bottom) 30% by quality rank, using NYSE breaks. Portfolios are rebalanced each year at the end of June, using accounting data from the fiscal year ending in the previous calendar year. The

[2]Data come from CRSP and Compustat. Detailed descriptions of the data items used in the construction of each of the measures is provided in Appendix A.

[3]Financial firms typically have large financial asset bases, but little tangible capital. Financials consequently tend to look low quality when measured using Grantham's notion, earnings quality, or gross profitability, which all are, or have components, scaled by assets. They tend to look high quality when measured using ROIC, which has tangible assets in the denominator, or using the defensive notion of low volatility and low market beta.

Table 1: Value and quality strategy return correlations

Sort variable	B/P	G1	G2	ROIC	EQ	F	D
Graham's G-score (G1)	-0.24						
Grantham's quality (G2)	-0.38	0.47					
ROIC	-0.11	0.29	0.47				
Earnings quality (EQ)	0.14	0.04	0.20	-0.27			
Piotroski's F-score (F)	-0.11	0.18	0.38	0.56	0.06		
Defensive (D)	0.20	0.24	0.57	0.54	0.18	0.40	
Gross profitability (GP)	-0.58	0.42	0.38	0.43	-0.26	0.17	-0.00

sample covers July 1963 to December 2013. Returns are calculated net of estimated transaction costs calculated, as in Novy-Marx and Velikov (2014), using the Hasbrouck (2009) Bayesian-Gibbs sampling procedure for estimating a generalized version of the Roll (1984) effective spread measure. Trading costs are typically modest, on the order of 0.5%/year for large cap strategies and 1.5%/year for small cap strategies, because quality is highly persistent so strategies based on quality turnover infrequently.

Table 1 shows correlations between the returns to the seven quality strategies and a traditional value strategy, which is constructed similarly using book-to-price. The table shows some commonalities across the different measures of quality. Eighteen of the twenty one pair-wise correlations across distinct quality measures are positive. Two of the three negative coefficients are on correlations with Sloan's earnings quality strategy, which is relatively weakly correlated with all the other strategies. This suggests that the earnings quality strategy should not, despite its name, be classified as a quality strategy.

The table shows mixed results on the quality strategies' relation to value. The strategies based on Graham's and Grantham's notions of quality, and especially that based on gross profitability, tilt towards growth. The strategies based on ROIC and the accounting notions of quality are only weakly correlated with traditional value. The defensive strategy, as is well known by practitioners, tilts strongly toward value. Interestingly, the defensive strategy, despite this value tilt, still covaries positively with all of the quality strategies except gross profitability. It covaries particularly strongly with the strategy based on Grantham's

Table 2: Quality strategy performance

		Three-factor model regression results			
Sort variable	$E[r^e]$	α	β_{MKT}	β_{SMB}	β_{HML}
		Panel A: Traditional value			
Book-to-price	3.49	-1.54	-0.00	0.27	0.96
	[2.33]	[-2.27]	[-0.31]	[14.5]	[47.5]
		Panel B: Quality strategies			
Graham's G-score	-0.08	1.69	-0.10	-0.16	-0.15
	[-0.08]	[1.93]	[-6.07]	[-6.70]	[-5.57]
Grantham's quality	-0.55	4.84	-0.30	-0.47	-0.48
	[-0.37]	[4.90]	[-15.7]	[-17.0]	[-16.4]
ROIC	2.17	4.66	-0.15	-0.49	-0.01
	[1.16]	[2.78]	[-4.42]	[-10.4]	[-0.27]
Earnings quality	1.17	2.13	-0.10	-0.13	0.01
	[0.99]	[1.83]	[-4.43]	[-3.99]	[0.39]
Piotroski's F-score	2.24	4.33	-0.14	-0.29	-0.08
	[1.69]	[3.57]	[-5.88]	[-8.53]	[-2.14]
Defensive	-1.55	3.45	-0.66	-0.86	0.39
	[-0.52]	[2.09]	[-20.4]	[-18.9]	[7.85]
Gross profitability	2.70	5.21	-0.08	-0.02	-0.46
	[2.15]	[4.65]	[-3.56]	[-0.80]	[-13.7]

Notes: This table shows returns (percent per year), in excess of those on T-bills, to long/short strategies formed by sorting stocks on the basis of book-to-price, or one of seven quality metrics. The table also shows three-factor model alphas and factor loadings.

notion of quality, which uses low earnings volatility as one of its components.

Table 2 shows the performance of the seven quality strategies, as well as that of traditional value. Of the seven quality strategies only gross profitability generates a significant excess return, and the gross profitability spread of 2.7%/year is only three-quarters as large as the 3.5%/year value spread. Three of the quality strategies, those based on Graham's and Grantham's notions of quality and the defensive strategy, actually generate negative spreads, though these are all insignificant.

All of the quality strategies look better when evaluated against the Fama and French three-factor model. All of the strategies have negative market loadings, ranging from -0.08 on gross profitability to -0.66 on defensive equity. All of the strategies also tilt toward large

caps, with SMB loadings ranging from -0.02 on gross profitability to -0.86 on defensive equity. Four of the strategies, those based on Graham's and Grantham's notions of quality, that based on Piotroski's F-score, and gross profitability, have significant growth tilts, while the defensive strategy has a significant value tilt. As a result of the negative Fama and French factor loadings four of the strategies, those based on Grantham's notion of quality, ROIC, Piotroski's F-score, and gross profitability, have highly significant three-factor alphas in excess of 4%/year. The three-factor alpha on the gross profitability strategy is 5.21%/year, with a t-stat in excess of four.

Spanning Tests

This section compares the performance of the seven quality strategies head-to-head, through a series of spanning tests. These tests essentially ask which of the strategies generate significant alpha relative to the others, by regressing the returns of a test strategy, taken from the quality strategies, onto the returns of the Fama and French factors and a potential explanatory strategy, also taken from the quality strategies. Significant abnormal returns suggest an investor already trading the Fama and French factors and the explanatory strategy could realize significant gains by starting to trade the test strategy. Insignificant abnormal returns suggest that the investor has little to gain by starting to trade the test strategy.

Table 3 shows that all of the strategies generally generate positive abnormal returns when evaluated against the Fama an French factors and another quality strategy, but these abnormal returns are typically modest and statistically insignificant. Only two of the strategies consistently generate significant abnormal returns relative to all the others. The second and seventh rows of the table show that the strategies based on Grantham's notion of quality and gross profitability generate significant alpha relative to all the others. The second and seventh columns also show that these two strategies basically subsume all the others. The insignificant alphas in these columns show that an investor trading either of these two strategies has little to gain from starting to trade any of the five other strategies.

Table 3: Spanning tests

	Explanatory strategy (x)						
Test strategy (y)	G1	G2	ROIC	EQ	F	D	GP
Graham's G-score (G1)		0.29	1.06	1.42	1.33	1.36	-0.23
		[0.34]	[1.20]	[1.60]	[1.49]	[1.53]	[-0.28]
Grantham's quality (G2)	4.06		3.82	4.41	4.13	3.95	3.07
	[4.17]		[3.96]	[4.38]	[4.11]	[4.25]	[3.16]
ROIC	3.21	1.32		4.58	1.58	2.83	-0.70
	[1.90]	[0.80]		[3.12]	[1.02]	[1.76]	[-0.50]
Earnings quality (EQ)	1.50	1.14	2.62		1.69	1.57	2.86
	[1.27]	[0.95]	[2.57]		[1.43]	[1.34]	[2.50]
Piotroski's F-score (F)	3.32	2.69	2.21	3.48		3.16	2.37
	[2.71]	[2.17]	[2.00]	[2.85]		[2.59]	[1.94]
Defensive (D)	2.23	-0.72	0.98	2.44	1.68		0.54
	[1.33]	[-0.46]	[0.62]	[1.48]	[1.01]		[0.33]
Gross profitability (GP)	4.26	3.34	3.56	5.35	4.40	4.62	
	[4.06]	[3.03]	[3.80]	[4.88]	[3.87]	[4.13]	

Notes: This table shows alphas (percent per year) from five-factor time-series regressions of the form

$$y = \alpha + \beta_{\text{MKT}} MKT + \beta_{\text{SMB}} SMB + \beta_{\text{HML}} HML + \beta_{\text{UMD}} UMD + \beta_x x + \epsilon,$$

where y and x are the returns to a test strategy and an explanatory strategy, and in each case these are both taken from the seven quality strategies.

Quality performance within the large and small cap universes

Several of the quality strategies, especially defensive equity and that based on Grantham's notion of quality, have strong size biases that are well known by practitioners. In these cases the quality metrics work in part by picking stocks across capitalization universes, which raises concerns regarding the power that these metrics have predicting performance within a given universe. This section addresses this issue by analyzing strategies constructed within the the Russell 1000 and Russell 2000 universes, which make up time series averages of 86.2% and 11.7% of total market capitalization over the sample, respectively.

Table 4 shows the performance of the seven quality strategies, as well as traditional value, constructed entirely within the Russell 1000 and Russell 2000. Panel A shows that among large caps none of the strategies, including value, generated significant excess re-

Table 4: Value and quality strategy performance by size

Sort variable	$E[r^e]$	Three-factor model regression results			
		α	β_{MKT}	β_{SMB}	β_{HML}
		Panel A: Large cap results (Russell 1000)			
Book-to-price	2.06	-2.29	-0.01	0.15	0.91
	[1.43]	[-3.34]	[-1.04]	[8.12]	[44.2]
Graham's G-score	-1.33	1.18	-0.13	-0.15	-0.30
	[-1.29]	[1.26]	[-6.91]	[-5.80]	[-10.6]
Grantham's quality	-0.75	3.29	-0.22	-0.28	-0.42
	[-0.65]	[3.78]	[-13.1]	[-11.6]	[-16.2]
ROIC	0.66	2.62	-0.01	-0.09	-0.36
	[0.52]	[2.17]	[-0.59]	[-2.80]	[-10.1]
Earnings quality	1.27	1.85	-0.08	-0.07	0.02
	[1.18]	[1.72]	[-3.68]	[-2.29]	[0.66]
Piotroski's F-score	1.36	2.67	-0.11	-0.11	-0.07
	[1.18]	[2.36]	[-4.93]	[-3.53]	[-2.16]
Defensive	-2.43	1.02	-0.60	-0.50	0.39
	[-0.98]	[0.67]	[-20.1]	[-11.8]	[8.52]
Gross profitability	1.95	4.99	-0.10	0.00	-0.57
	[1.49]	[4.60]	[-4.70]	[0.13]	[-17.6]
		Panel B: Small cap results (Russell 2000)			
Book-to-price	4.56	1.18	-0.00	-0.11	0.87
	[3.09]	[1.62]	[-0.25]	[-5.43]	[39.8]
Graham's G-score	2.75	4.10	-0.13	-0.12	-0.04
	[2.53]	[3.92]	[-6.42]	[-4.10]	[-1.43]
Grantham's quality	-0.20	3.58	-0.23	-0.31	-0.33
	[-0.17]	[3.93]	[-12.8]	[-12.3]	[-12.2]
ROIC	1.81	2.73	-0.06	-0.22	0.03
	[1.35]	[2.08]	[-2.18]	[-6.02]	[0.73]
Earnings quality	2.02	1.59	-0.01	-0.03	0.14
	[2.45]	[1.96]	[-0.71]	[-1.45]	[5.79]
Piotroski's F-score	2.11	3.68	-0.14	-0.27	0.03
	[1.72]	[3.38]	[-6.69]	[-8.91]	[1.02]
Defensive	-1.50	3.44	-0.64	-0.76	0.29
	[-0.56]	[2.36]	[-22.3]	[-18.8]	[6.70]
Gross profitability	3.32	3.85	0.02	0.05	-0.19
	[2.99]	[3.53]	[0.90]	[1.81]	[-5.87]

turns. Four strategies, those based on Grantham's notion of quality, ROIC, Piotroski's F-score, and gross profitability, generate three factor alphas among large cap stocks, though these alphas, with the exception of that on gross profitability, are 1-2%/year lower than those observed on the strategies constructed using the entire cross-section.

Panel B shows stronger results within the small cap universe. Among small caps, in addition to gross profitability, the strategies based on Graham's notion of quality and Sloan's earnings quality also generate significant excess returns. Among small caps all the strategies generate significant three-factor alpha, though this significance is only marginal in the case of earnings quality. These significant three-factor alphas come despite generally attenuated size and value loadings, because all the strategies generate higher returns in the small cap universe.

Table 5 shows results of spanning tests, like those performed in Table 3, for the large and small cap quality strategies. Here the results are striking. Panel A shows that in the large cap universe, which accounts for almost 90% of total market capitalization, only gross profitability generates consistently significant abnormal returns relative to the Fama and French factors and the other notions of quality. The relatively weak performance of the strategy based on Grantham's notion of quality results from mitigating the size bias that arises from sorting on Grantham quality metric. In Table 3, which shows results for strategies formed using the entire cross section of stocks, the Grantham strategy's large alpha was driven disproportionately by the short side, which tilted strongly to small stocks where the effects were stronger. Panel B shows that in the small caps three strategies generate consistently significant abnormal returns relative to the Fama and French factors and the other notions of quality, the strategy based on Graham's notion of quality (first row), which had almost no power in the whole cross-section, the strategy based on Grantham's notion of quality, and gross profitability (last row). Graham's quality does little to explain the performance of the other small cap quality strategies (first column), while gross profitability subsumes the power of all the other strategies except for earnings quality, to which it is negatively correlated (last column).

Table 5: Spanning tests, by universe

	Explanatory strategy (x)						
Test strategy (y)	G1	G2	ROIC	EQ	F	D	GP
	Panel A: Large cap results (Russell 1000)						
Graham's G-score (G1)		-0.19	0.30	1.00	1.09	1.03	-1.00
		[-0.21]	[0.34]	[1.04]	[1.13]	[1.07]	[-1.20]
Grantham's quality (G2)	2.32		1.88	2.81	2.59	2.67	1.09
	[2.88]		[2.39]	[3.20]	[2.93]	[3.31]	[1.36]
ROIC	2.00	0.75		3.15	2.09	2.49	-0.83
	[1.75]	[0.68]		[2.93]	[1.74]	[2.03]	[-0.89]
Earnings quality (EQ)	1.12	1.52	2.26		1.36	1.19	2.30
	[1.02]	[1.39]	[2.36]		[1.25]	[1.11]	[2.14]
Piotroski's F-score (F)	1.74	1.36	1.17	1.84		1.69	1.29
	[1.52]	[1.19]	[1.05]	[1.63]		[1.48]	[1.12]
Defensive (D)	0.07	-1.75	-0.27	0.52	0.05		-1.05
	[0.04]	[-1.23]	[-0.17]	[0.34]	[0.03]		[-0.68]
Gross profitability (GP)	3.87	2.97	3.00	4.80	4.35	4.47	
	[4.10]	[2.99]	[3.63]	[4.48]	[3.95]	[4.12]	
	Panel B: Small cap results (Russell 2000)						
Graham's G-score (G1)		3.41	3.95	4.50	4.02	4.19	3.12
		[3.26]	[3.74]	[4.22]	[3.77]	[3.99]	[3.03]
Grantham's quality (G2)	2.37		2.11	3.67	2.04	3.06	1.65
	[2.60]		[2.82]	[4.04]	[2.53]	[3.47]	[1.99]
ROIC	1.95	0.07		3.86	0.67	2.49	-0.84
	[1.46]	[0.07]		[3.21]	[0.63]	[2.00]	[-0.92]
Earnings quality (EQ)	1.52	1.94	2.07		1.46	1.36	2.23
	[1.82]	[2.39]	[2.78]		[1.77]	[1.66]	[2.80]
Piotroski's F-score (F)	2.48	0.99	1.50	3.18		2.65	1.38
	[2.22]	[1.03]	[1.72]	[2.87]		[2.57]	[1.34]
Defensive (D)	0.17	-0.39	0.11	1.60	-0.12		0.17
	[0.12]	[-0.28]	[0.08]	[1.13]	[-0.09]		[0.12]
Gross profitability (GP)	2.81	2.31	2.39	4.75	3.04	4.01	
	[2.62]	[2.34]	[3.21]	[4.48]	[2.94]	[3.68]	

Notes: This table shows alphas (percent per year) from five-factor time-series regressions of the form

$$y = \alpha + \beta_{\text{MKT}} MKT + \beta_{\text{SMB}} SMB + \beta_{\text{HML}} HML + \beta_{\text{UMD}} UMD + \beta_{x} x + \epsilon,$$

where y and x are the returns to a test strategy and an explanatory strategy, and in each case these are both taken from the seven quality strategies.

Long-only investors

Long-only investors face a fundamentally different investment problem than long/short investors. Unconstrained investors can control their risk through leverage, which separates the opportunity and exposure decisions, allowing them to concentrate solely on finding opportunities that provide the highest reward-to-risk ratio. Long-only investors do not have this luxury. They cannot separate the opportunity and exposure decisions, so must evaluate risk and reward jointly, and may rationally choose to pass up an investment with a higher risk/reward tradeoff for an investment that allows them to get greater exposure to another attractive opportunity with a lower risk/reward tradeoff.

This is not merely a theoretical concern for long-only quality investors. Much, and sometimes all, of the observed quality strategy three-factor alpha in Table 2 came from negative loadings on the Fama and French factors. In these cases the strategies do not raise expected returns, but simply provided an attractive hedge for market investors with small cap and value tilts. The value of this hedge is only large if the small cap and value tilts contribute significantly to the investors' portfolio risk.

Most of a well diversified, long-only equity investor's risk comes from the market, however, not from tracking error relative to the market. Well diversified, long-only strategies can only achieve modest size and value loadings. With small tracking errors even relatively tracking error reductions have little impact on portfolio volatility. Adding quality to a value strategy can thus improve the strategy's information ratio while simultaneously reducing the strategy's Sharpe ratio. If adding quality reduces the tracking error more than it reduces expected active return, then it improves the information ratio, but if there is little associated reduction in portfolio volatility it lowers the Sharpe ratio.

Combining value and quality portfolios

Table 6 shows the stand-alone performance of the long-only quality and traditional value strategies, which are just the long sides of the long/short strategies considered in Table 2. All of the quality strategies generate positive CAPM alphas, ranging from 0.03%/year for

Table 6: Long-only quality strategy performance

			Three-factor model regression results			
Sort variable	$E[r^e]$	α_{CAPM}	α	β_{MKT}	β_{SMB}	β_{HML}
			Panel A: Value strategy			
Book-to-price	8.70	3.36	-0.72	1.01	0.18	0.68
	[3.77]	[3.15]	[-1.34]	[96.8]	[12.4]	[42.7]
			Panel B: Quality strategies			
Graham's G-score	5.66	0.03	0.26	0.98	-0.07	-0.02
	[2.61]	[0.14]	[1.34]	[262.0]	[-12.5]	[-2.92]
Grantham's quality	5.74	0.25	1.39	0.94	-0.12	-0.16
	[2.67]	[0.57]	[4.02]	[138.8]	[-12.9]	[-15.9]
ROIC	6.11	0.26	1.53	0.98	-0.08	-0.20
	[2.65]	[0.49]	[3.32]	[109.1]	[-6.12]	[-14.8]
Earnings quality	6.13	0.22	0.78	1.00	-0.02	-0.10
	[2.57]	[0.30]	[1.06]	[69.1]	[-0.84]	[-4.37]
Piotroski's F-score	6.22	0.77	1.34	0.94	-0.08	-0.08
	[2.91]	[1.74]	[3.15]	[112.6]	[-6.69]	[-5.96]
Defensive	5.21	0.99	0.14	0.80	-0.16	0.22
	[2.97]	[1.44]	[0.25]	[72.7]	[-10.5]	[13.0]
Gross profitability	7.10	1.44	2.74	0.94	-0.06	-0.22
	[3.11]	[2.03]	[4.17]	[73.4]	[-3.08]	[-11.1]

Graham's strategy, to 1.44%/year for gross profitability. Only gross profitability's CAPM alphas is statistically significant at the 5% level. All of the quality strategies do, however, with the exception of the earnings quality and defensive strategies, generate highly significant Fama and French three-factor alphas.

Table 7 shows the performance of strategies that run quality side-by-side with value. The table shows the difficulties long-only investors face exploiting quality's benefits. The table shows information ratio gains, but Sharpe ratio losses. In every case, except for the earnings quality and defensive strategies, combining value and quality yields active returns relative to the market that have a better risk/reward trade-off than that provided by value alone (last line of Panel C). This results, however, from a dramatic decrease in the CAPM tracking error, coupled with a smaller decrease in the average active returns. Market exposure plus moderate exposure to the attractive opportunity provided by the pure value

Table 7: Combining value and quality portfolios

	Specification							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: Portfolio weights (%)							
Book-to-price	100	50	50	50	50	50	50	50
Graham's G-score		50						
Grantham's quality			50					
ROIC				50				
Earnings quality					50			
Piotroski's F-score						50		
Defensive							50	
Gross profitability								50
	Panel B: Portfolio performance							
Average annual return	8.70	7.18	7.22	7.40	7.41	7.46	6.95	7.90
Volatility	16.2	15.2	15.0	15.4	15.7	15.0	13.9	15.2
Sharpe ratio	0.54	0.47	0.48	0.48	0.47	0.50	0.50	0.52
	Panel C: Benchmarked performance							
CAPM alpha	3.36	1.69	1.80	1.81	1.79	2.06	2.18	2.40
Tracking error	7.46	3.63	3.22	3.31	4.25	3.77	5.13	3.52
Information ratio	0.45	0.47	0.56	0.55	0.42	0.55	0.42	0.68

tilt turns out to be superior, in a Sharpe ratio sense, to market exposure plus a minimal exposure to the more attractive joint value and quality tilt.[4]

Selecting stocks using value and quality

Greenblatt (2006, ROIC), Piotroski and So (2012, the F-score) and Novy-Marx (2013, gross profitability) all consider strategies that combine value and quality to select stocks. These strategies, instead or running value and profitability side-by-side, select stocks that look attractive on both the value and quality dimensions. Trading quality with value in this manner yields similar information ratio improvements to trading quality along side value, but generates these improvements primarily by increasing rewards, as opposed to reducing

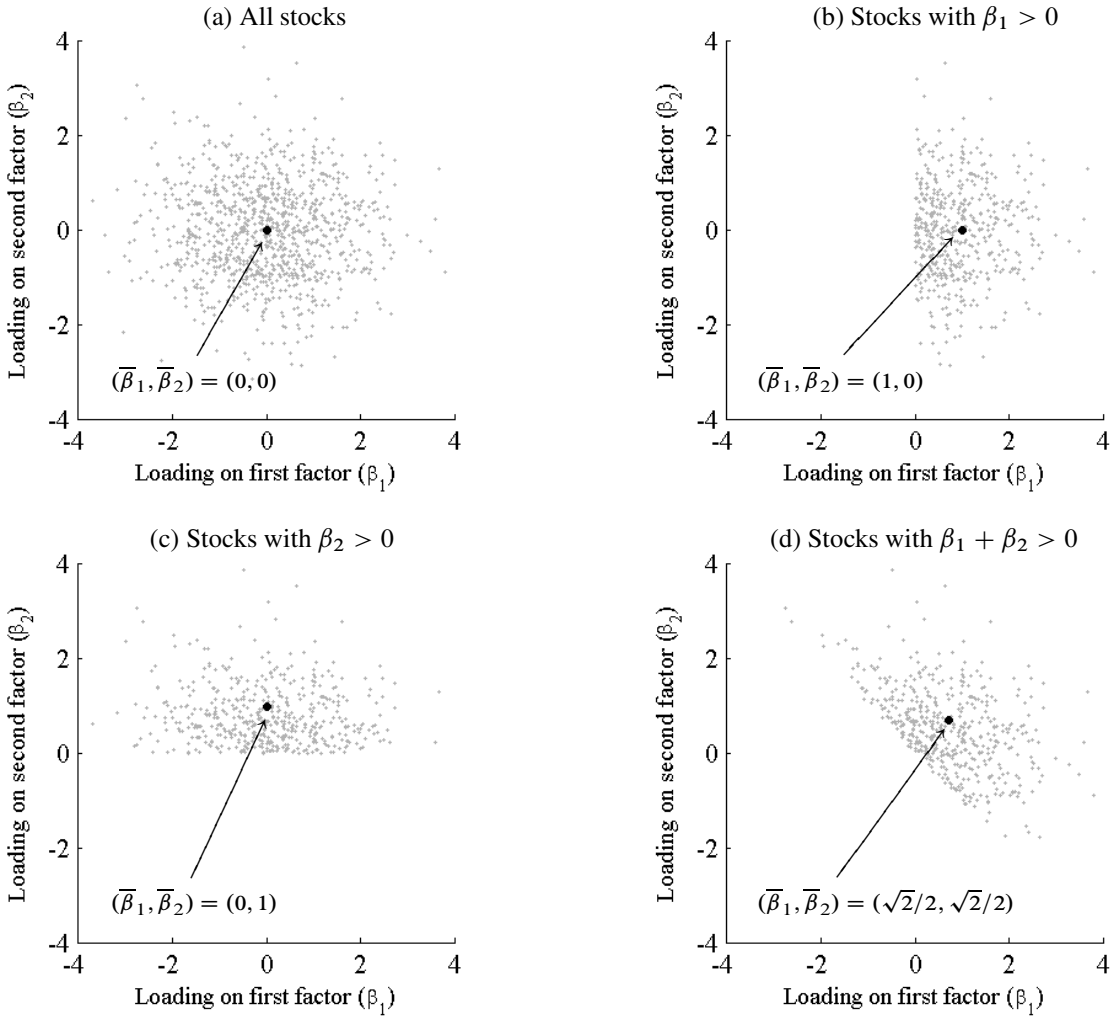
[4]The 50/50 mixes are considered here because in every case except gross profitability the long-only ex-post mean-variance efficient portfolio of value and each of the quality strategies is fully invested in value.

risk. For example, Novy-Marx (2013) finds that a long/short strategy based on combined book-to-price and gross profitability ranks that trades only the 500 largest non-financial firms earned excess returns of 7.4%/year from July 1963 to December 2011. This is significantly more than the 3.2 and 3.8%/year earned on similarly constructed strategies trading on gross profitability or book-to-price alone, or the 3.5%/year one would have gotten trading the pure strategies side-by-side. This is because the the strategy that selects stocks on the basis of both profitability and valuations achieves larger exposures to profitability and value than the strategy that runs profitability next to value. If a stock has moderately high loadings on both factors, then the stock is more attractive to an investor attempting to get exposures to both factors than a stock with a very high loading on one factor but a low loading on the other. A 50/50 combination of pure factor strategies will nevertheless ignore this stock, because it does not achieve a sufficiently high exposure to either factor individually.

The increase in tilts that results from buying stocks with high combined exposures can be further illustrated with a simple example. Panel (a) of Figure 1 shows stocks' loadings on two factors, assuming the loadings on each factor are normally distributed with mean zero, and uncorrelated across stocks. Panel (b) shows that the portfolio that holds only stocks with positive loadings on the first factor has a positive loading on the first factor (arbitrarily scaled to one), but a zero loading on the second factor. Panel (c) shows similar results for the portfolio that holds only stocks with positive loadings on the second factor. Panel (d) shows that a portfolio designed to tilt toward both factors (i.e., the portfolio that holds stocks for which the sum of the loadings on the two factors in positive) achieves loadings of only $\sqrt{2}/2$ on each factor, 71% of the loading one could achieve on either factor individually. Fama and French (2013) emphasize the "variable attenuation" that occurs when one goes from selecting stocks on the basis of a single predictive variable to selecting stocks on the basis of multiple predictors. In their words, "getting the average return benefits of an additional variable involves losing some of the gains from the variables already in the mix."

The integrated solution, which selects stocks on the basis of the combined factor sig-

Figure 1: Factor loadings from single and combined characteristic sorts



nals, achieves significantly higher factor loadings, however, than the portfolio solution. An investor that puts half of her money into each of two uncorrelated pure factor strategies only tilts half way towards each. These tilts are much smaller than the 0.71 loadings on each factor achieved by the combined sort. By selecting stocks directly to maximize both exposures, an investor is able to achieve combined factor loadings that are 40% higher.

The advantage of the integrated solution is even more pronounced when the univariate

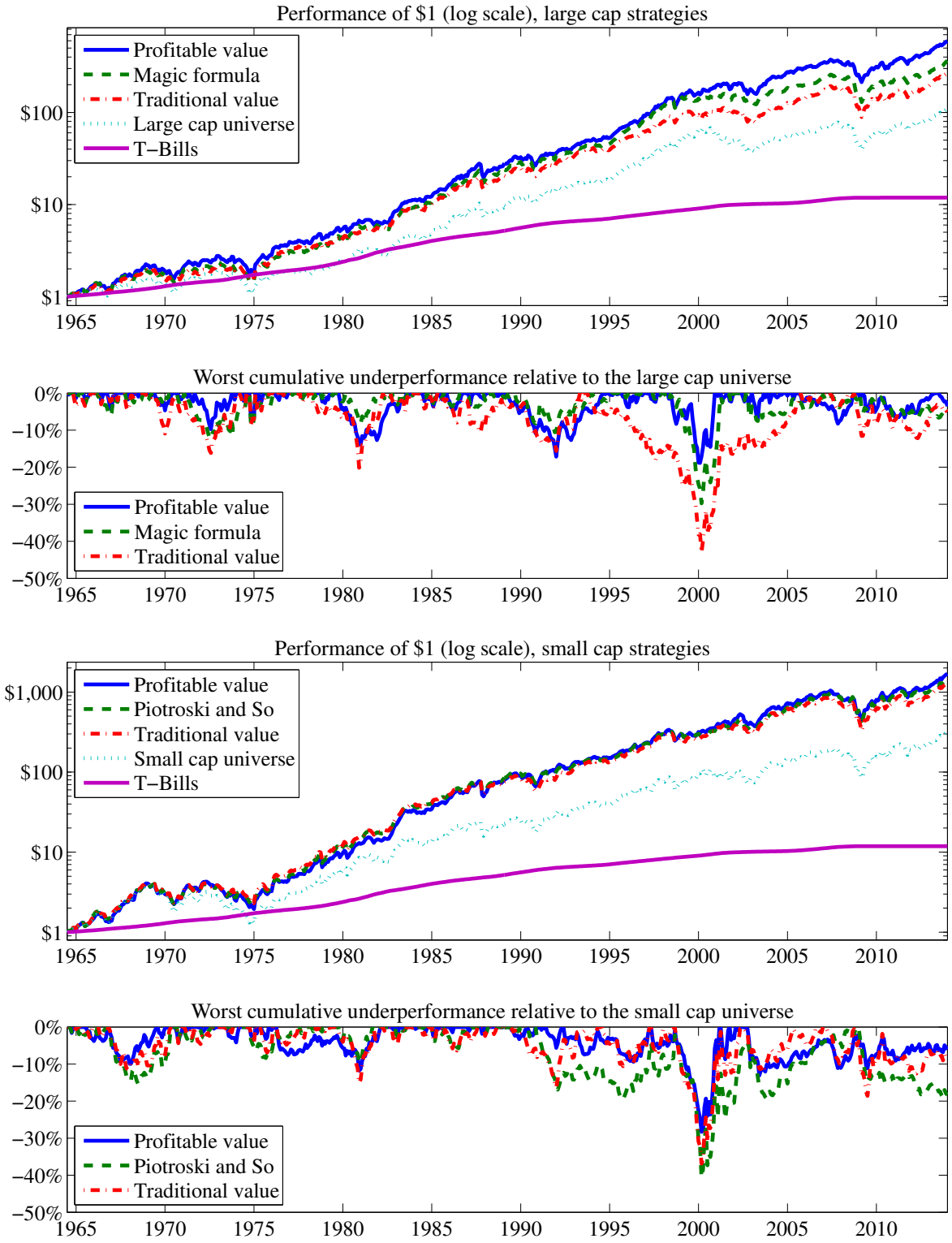
factor loadings are negatively correlated, such as value and profitability. At first glance this may be somewhat surprising. High profitability stocks tend to have high valuations, so Fama and French's (2013) variable attenuation problem is particularly acute when adding profitability metrics to value signals. High profitability value stocks are not nearly as cheap as pure value stocks, so including profitability considerations results in a bigger reduction in value exposures. This attenuation in the value exposure is, however, more than offset by gains in the profitability exposure. Portfolios selected purely on the basis of value signals are significantly short profitability, so incorporating profitability considerations yields disproportionately large gains in the profitability dimension. Appendix B analyzes these effects in more detail, deriving predicted gains in factor loadings obtained by selecting stocks on the basis of value and quality jointly, relative to those obtained by holding equal positions in pure value and quality strategies, when the value and quality characteristics are noisy signals of the true value and quality factor loadings.

These results suggest that larger value and quality tilts can be achieved sorting on a combined quality and value signal than can be obtained running value and quality strategies side-by-side, which should translate into higher active returns. Table 8 shows the performance of the long-only strategies sorted on the combined quality and value signals, the average book-to-price and quality metric ranks, and confirms this hypothesis. Panel A gives large cap results, and shows that combining valuations with ROIC, Piotroski's F-score, or gross profitability, yields higher returns than using valuation alone. In the case of gross profitability this improvement is almost 2%/year. In every case, with the exception of earnings quality, incorporating quality concerns leads to improvements in the Sharpe ratios and CAPM information ratios, though with the exception of gross profitability these improvements are quite marginal. Panel B gives small cap results, where traditional value already delivered stellar performance over the sample, yielding a CAPM alpha of 5.35%/year, and should consequently be difficult to beat. The table shows that while only incorporating gross profitability concerns actually yielded a strategy that generated higher excess returns (12.3 vs. 11.7%/year), all of the joint quality and value strategies with the exception of

Table 8: Long-only joint quality and value strategy performance

			Three-factor model regression results			
Sort variable	$E[r^e]$	α_{CAPM}	α	β_{MKT}	β_{SMB}	β_{HML}
			Panel A: Large cap results (Russell 1000)			
Traditional value	7.49	2.38	-0.78	0.99	-0.01	0.59
	[3.47]	[2.60]	[-1.76]	[114.1]	[-1.01]	[44.1]
Graham value	6.99	1.89	-0.01	0.96	-0.04	0.36
	[3.34]	[2.54]	[-0.03]	[89.3]	[-2.78]	[22.3]
Grantham value	6.90	2.20	1.01	0.89	-0.13	0.27
	[3.57]	[3.13]	[1.79]	[80.8]	[-8.35]	[16.0]
Magic formula	8.15	2.75	0.94	1.00	-0.01	0.33
	[3.71]	[3.72]	[1.61]	[86.7]	[-0.65]	[19.1]
Sloan value	6.76	1.77	0.09	0.94	-0.07	0.33
	[3.26]	[2.18]	[0.13]	[72.8]	[-3.93]	[17.0]
Piotroski and So	7.83	2.94	1.29	0.92	-0.08	0.33
	[3.80]	[3.41]	[1.79]	[65.1]	[-4.00]	[15.3]
Cheap defensive	6.42	2.35	0.11	0.83	-0.17	0.48
	[3.51]	[2.43]	[0.17]	[65.4]	[-9.44]	[24.6]
Profitable value	9.20	3.68	1.70	1.00	0.08	0.33
	[4.10]	[4.86]	[2.75]	[82.4]	[4.92]	[18.0]
			Panel B: Small cap results (Russell 2000)			
Traditional value	11.7	5.31	-0.74	1.07	0.87	0.79
	[3.96]	[3.19]	[-1.18]	[87.1]	[49.9]	[42.1]
Graham value	11.8	5.81	0.67	0.98	0.78	0.65
	[4.32]	[3.88]	[1.00]	[75.4]	[42.1]	[32.9]
Grantham value	11.4	5.56	1.36	0.95	0.70	0.51
	[4.40]	[4.21]	[2.09]	[74.2]	[38.6]	[26.3]
Magic formula	11.2	4.68	0.05	1.05	0.80	0.55
	[3.83]	[3.07]	[0.06]	[67.0]	[36.2]	[23.0]
Sloan value	11.4	4.76	-0.59	1.09	0.85	0.67
	[3.78]	[2.93]	[-0.76]	[71.6]	[39.7]	[28.8]
Piotroski and So	11.7	5.64	0.75	1.00	0.76	0.62
	[4.22]	[3.70]	[0.91]	[62.5]	[33.3]	[25.2]
Cheap defensive	10.0	5.15	0.63	0.83	0.58	0.61
	[4.43]	[3.99]	[0.92]	[62.2]	[30.6]	[30.2]
Profitable value	12.3	5.73	1.19	1.03	0.89	0.50
	[4.16]	[3.67]	[1.62]	[71.8]	[43.7]	[23.0]

Figure 2: Long-only joint value and quality performance



those incorporating ROIC or earnings quality concerns had higher Sharpe ratios, and higher CAPM information ratios, than traditional value.

The top panel of Figure 2 shows the growth of a dollar invested at the end of June 1963 in T-bills, the Russell 1000, the large cap value strategy, and the two large cap joint value and quality strategies that generated the highest returns, those based on ROIC and gross profitability. The second panel shows the value and joint value and quality strategies' drawdowns relative to the large cap universe (i.e., cumulative under-performance relative to the benchmark). The third and fourth panel show similar results for small caps, using the two small cap joint value and quality strategies generating the highest returns, those based on Piotroski's F-score and gross profitability. The most striking feature of the figure

Table 9: Growth of a dollar, drawdowns, and outperformance frequencies

	Growth of $1 (value at end of sample, $)	Max. drawdown (% cumulative underperformance)	One-year outperformance frequency (%)	Five-year outperformance frequency (%)
Panel A: Large cap strategies (benchmarked to Russell 1000)				
Benchmark (R1000)	111			
Traditional value	269	-43.0	55.7	67.5
Graham value	218	-43.9	58.8	62.3
Grantham value	226	-34.8	57.1	66.8
Magic formula	364	-29.8	69.0	75.3
Sloan value	196	-41.2	58.9	57.4
Piotroski and So	335	-37.7	60.8	71.2
Cheap defensive	187	-52.2	48.8	54.0
Profitable value	595	-18.9	72.2	81.3
Panel B: Small cap strategies (benchmarked to Russell 2000)				
Benchmark (R2000)	269			
Traditional value	1,294	-36.9	65.3	72.2
Graham value	1,561	-37.1	61.3	74.9
Grantham value	1,401	-38.1	62.1	73.9
Magic formula	1,022	-48.6	63.8	64.8
Sloan value	1,055	-27.5	62.0	73.7
Piotroski and So	1,462	-40.6	66.8	75.3
Cheap defensive	870	-55.4	53.5	55.4
Profitable value	1,690	-28.3	68.4	76.9

is the marked improvement in the drawdown performance that results in value strategies by incorporating gross profitability concerns. Table 9 provides end of sample values for a dollar invested in the start of the sample for all of the joint value and quality strategies, as well as the largest drawdowns experienced by each of these over the sample, and the frequencies with which these strategies outperform their benchmarks at one- and five-year horizons.

Conclusion

Quality investing exploits another dimension of value. Value strategies endeavor to acquire productive capacity cheaply. Traditional value strategies do this by buying assets at bargain prices; quality strategies do this by buying uncommonly productive assets. Strategies based on either of value's dimensions generate significant abnormal returns, but the real benefits of value investing accrue to investors that pay attention to both price and quality. Attention to quality, especially measured by gross profitability, helps traditional value investors distinguish bargain stocks (i.e., those that are undervalued) from value traps (i.e., those that are cheap for good reasons). Price signals help quality investors avoid good firms that are already fully priced. Trading on both signals brings the double benefit of increasing expected returns while decreasing volatility and drawdowns. Cheap, profitable firms tend to outperform firms that are just cheap or just profitable. Quality tends to perform best when traditional value suffers large drawdowns, and vice versa, so strategies that trade on both signals generate steadier returns than do strategies that trade on quality or price alone. These benefits are available to long-only investors as well as long/short investors. Accounting for quality also significantly improves the performance of strategies that incorporate momentum as well as price signals.

Several practical considerations make joint quality and value strategies look even more attractive. The signal in gross profitability is extremely persistent—even more persistent than that in valuations—and works well in the large cap universe. Joint quality and value strategies thus have low turnover, and can be implemented using liquid stocks with the ca-

pacity to absorb large trades. The joint profitability and value signal is also less susceptible to industry biases that are uninformative about future stock returns. Both the value and profitability premiums are largely intra-industry phenomena, reducing the informativeness of simple, univariate measures of value and profitability. This is less of a problem for strategies that trade on the combined quality and value signal. Because industry capital intensity is positively correlated with value signals (which have book values in the numerator) but negatively correlated with profitability signals (which have book values in the denominator), systematic industry variation in the value and quality metrics tend to cancel in the joint signal. Joint quality and value strategies can thus be implemented effectively while paying less attention to industry controls.

The basic message is that investors, in general but especially traditional value investors, leave money on the table when they ignore the quality dimension of value. All of the best known notions of quality contribute, at least marginally, to investment performance. Gross profitability generally contributes the most, however, especially among large caps stocks and for long-only investors, and largely subsumes the power of other notions of quality.

Keywords: Value Investing, Quality Investing, Gross Profitability, GARP, Asset Pricing.

Appendix A: Variable Definitions

Variables employed in this paper are constructed primarily from Compustat data, which is assumed to be publically available by the end of June in the calendar year following that in which each firms' fiscal year ends. Detailed definitions, as well as the Compustat data items employed in the construction of these variables, are given below.

- **Book-to-price (B/P):** Book equity scaled by market equity, where market equity is lagged six months in the strategies that do not trade momentum to avoid taking unintentional positions in momentum. Book equity is shareholder equity, plus deferred taxes, minus preferred stock, when available. For the components of shareholder equity, I employ tiered definitions largely consistent with those used by Fama and French (1993) to construct their high minus low factor (HML). Stockholders equity is as given in Compustat (SEQ) if available, or else common equity plus the carrying value of preferred stock (CEQ + PSTX) if available, or else total assets minus total liabilities (AT - LT). Deferred taxes is deferred taxes and investment tax credits (TXDITC) if available, or else deferred taxes and/or investment tax credit (TXDB and/or ITCB). Preferred stock is redemption value (PSTKR) if available, or else liquidating value (PSTKRL) if available, or else carrying value (PSTK).

- **Earnings-to-price (E/P):** Net income (NI) scaled by market equity.

- **Earnings yield (EY):** Earnings before interest and taxes (EBIT) scaled by enterprise value (EV). Enterprise value is market equity, plus long term debt (DLTT), plus debt in current liabilities (DLC), plus preferred stock (as defined above), minus cash and short term investments (CHE).

- **Graham G-score:** The G-score gets one point if current assets (ACT) exceeds twice current liabilities (LCT) , one point if net current assets (WCAP) exceed long term debt (DLTT), one point if net earnings have been positive each of the last ten years, one point if dividends plus buy-backs have been positive each of the last ten years, and one point if current earnings per share are at least 33% higher than 10 years ago.

- **Grantham quality rank:** Average ranks of returns-on-equity (ROE), asset-to-book equity, and the inverse of ROE volatility. ROE is net income-to-book equity. ROE volatility is the standard deviation of ROE over the preceding five years.

- **Return on invested capital (ROIC):** EBIT-to-tangible capital, where tangible capital is property, plant and equipment (PPEGT) plus working capital (WCAP).

- **Sloan's accruals:** Measured as the year-over-year change in current assets (ACT) excluding cash and short term liabilities (CHE), minus the change in long term liabilities (LCT) excluding debt in current liabilities (LCT) and income taxes payable (TXP), minus the depreciation and amortization (DPC). Following Sloan (1996), accruals are scaled by the average of total assets and total assets lagged one year.

- **Piotroski's F-score:** Constructed as the sum of nine binary variables that take the value zero (indicating weakness) or one (indicating strength). The F-score can get gets one point for each of four profitability signals [positive earnings before extraordinary items (IB), positive cash flows from operations (OANCF), increasing returns-on-assets (IB/AT that exceeds that of the previous year), and negative accruals]; one point for each of three liquidity signals [decreasing debt, increasing current ratio, and no equity issuance]; and one point for each of two efficiency signals [increasing gross margins (revenues (REVT) minus cost of goods sold (COGS) scaled by revenues) and increasing asset turnover (revenues scaled by assets)].

- **Gross profits-to-assets (GP/A):** Revenues minus cost of goods sold (REVT - COGS) scaled by total book assets (AT).

Appendix B: Factor loadings from selecting stocks on value and quality

Suppose stocks' exposures to two factors are correlated across stocks and normally distributed. That is, suppose the exposure of stock j to factor $i \in (1,2)$ is distributed $\beta_{ij} \sim N(0, \sigma_\beta^2)$ and $\text{corr}(\beta_{1j}, \beta_{2j}) = \rho$. Suppose further that these loadings are not directly observable, but investors see publicly available signals $S_{ij} = \beta_{ij} + \epsilon_{ij}$, where the

$\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$ are normally distributed noise independent across stocks and factors. Define $z_{ij} \equiv S_{ij}/\sigma_S$ as the z-scores for the signals S_{ij}, where $\sigma_S^2 = \sigma_\beta^2 + \sigma_\epsilon^2$. Then the loading on factor i of a portfolio that holds the fraction α of stocks with the highest signal for that factor is

$$\begin{aligned} E\left[\beta_{ij}|z_{ij} > N^{-1}(1-\alpha)\right] &= \sigma_S E\left[\beta_{ij}/\sigma_S|z_{ij} > N^{-1}(1-\alpha)\right] \\ &= \frac{\sigma_\beta^2}{\sigma_S}\left(\frac{n\big(N^{-1}(\alpha)\big)}{\alpha}\right), \end{aligned}$$

where $n(\cdot)$ and $N(\cdot)$ are the standard normal probability density function and cumulative distribution function, respectively, and the second equality follows from the facts that β_{ij}/σ_S contributes a fraction $\sigma_\beta^2/\sigma_S^2$ of the total variance of z_{ij} and $E\left[z_{ij}|z_{ij} > x\right] = n(-x)/N(-x)$ for any x. The portfolio also has a loading ρ times as large on the other factor, so a 50/50 mix of the portfolios selected on the basis of the individual factor signals had loadings $(1+\rho)/2$ times as large on each factor.

The loading on factor i of a portfolio that holds the fraction α of stocks with the highest combined z-scores is

$$\begin{aligned} E\left[\beta_{ij}\Big|\frac{z_{1j}+z_{2j}}{\sigma_{z_1+z_2}} > N^{-1}(1-\alpha)\right] &= \frac{\sigma_{S_1+S_2}}{2} E\left[\frac{\beta_{1j}+\beta_{2j}}{\sigma_{S_1+S_2}}\Big|\frac{z_{1j}+z_{2j}}{\sigma_{z_1+z_2}} > N^{-1}(1-\alpha)\right] \\ &= \frac{\sigma_{\beta_1+\beta_1}^2}{\sigma_{S_1+S_2}}\left(\frac{n\big(N^{-1}(\alpha)\big)}{\alpha}\right) \\ &= \frac{1+\rho}{\sqrt{2\left(1+\rho\sigma_\beta^2/\sigma_S^2\right)}}\left(\frac{\sigma_\beta^2}{\sigma_S}\right)\frac{n\big(N^{-1}(\alpha)\big)}{\alpha}, \end{aligned}$$

where the first equality follows from symmetry and the second from the fact that $(\beta_{1j} + \beta_{2j})/\sigma_{S_1+S_2}$ contributes a fraction $\sigma_{\beta_1+\beta_2}^2/\sigma_{S_1+S_2}^2$ of the total variance of $z_{1j} + z_{2j}$.

The ratio of the factor loadings of the portfolio that trades on the combined signal to

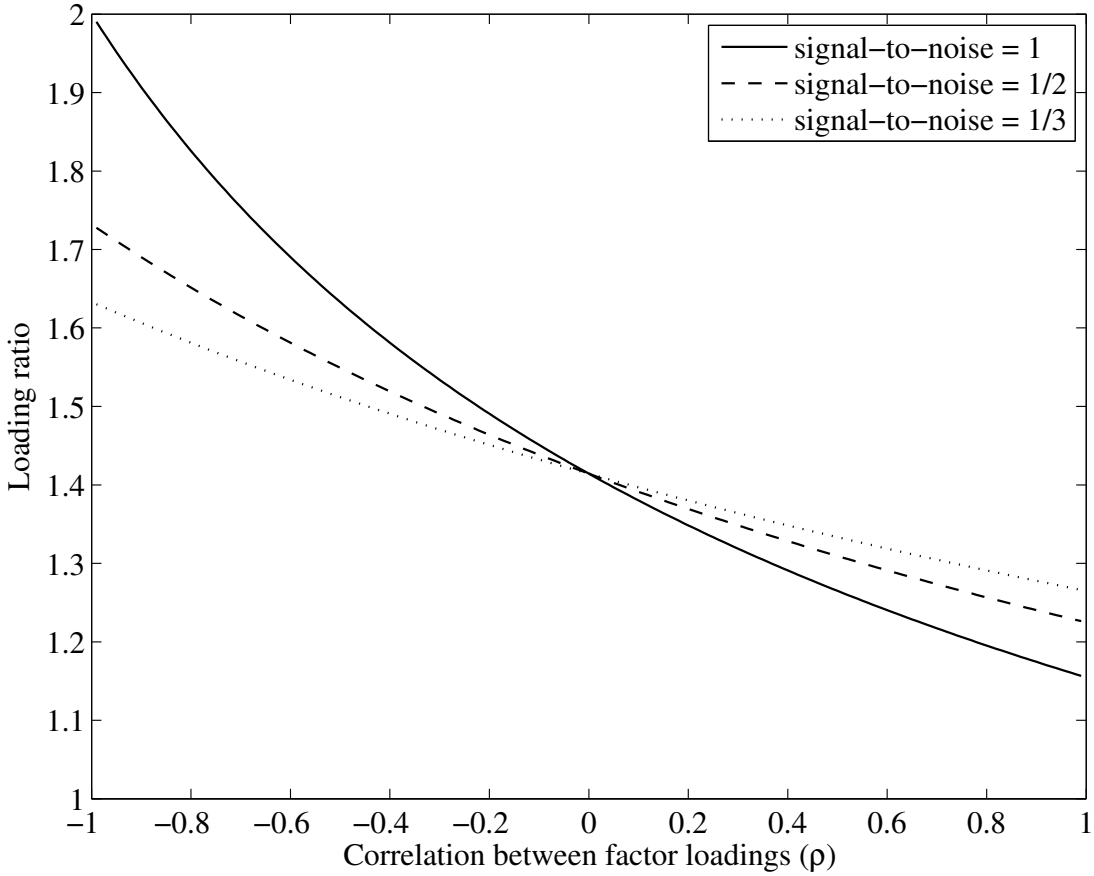


Figure 3: Factor loadings from integrated solution, relative to those from portfolio solution.

the portfolio that holds a 50/50 mix of the portfolios that trade on the pure signals is thus

$$\sqrt{\frac{2}{1+\rho\sigma_\beta^2/\sigma_s^2}}.$$

This is bounded from below be one, so there are always gains to trading on the combined characteristic, but a decreasing function of ρ, so it is less important to trade on the combined characteristic when the factors are more correlated. It also tends toward $\sqrt{2/(1+\rho)}$ when the factor loadings are directly observable (i.e., when $\sigma_\epsilon = 0$, so $S = \beta$), and tends toward $\sqrt{2}$ as the signal becomes uninformative regarding factor loadings (i.e., as $\sigma_\beta/\sigma_\epsilon \to 0$).

This ratio is shown in Figure 3 as a function of ρ, the correlation in the true factor

Table 10: Performance of long/short joint value and quality strategies

Quality measure used with B/M	$E[r^e]$	α_{FF4}	Four-factor model regression results				
			α	β_{MKT}	β_{SMB}	β_{VMG}	β_{SMI}
None	3.49	-1.54	0	0	0	1	0
	[2.33]	[-2.27]					
Graham value	2.17	0.58	0.07	-0.07	-0.19	0.69	0.61
	[1.64]	[0.71]	[1.43]	[-5.93]	[-11.4]	[39.8]	[21.1]
Grantham value	2.79	2.35	0.08	-0.08	0.00	0.74	0.52
	[2.13]	[2.63]	[1.60]	[-5.83]	[0.26]	[37.6]	[21.3]
Magic formula	5.74	3.31	0.17	-0.04	0.15	0.67	0.48
	[3.82]	[2.84]	[2.50]	[-2.27]	[5.84]	[29.0]	[23.4]
Sloan value	3.09	-0.21	-0.00	-0.01	0.01	0.76	0.45
	[2.13]	[-0.21]	[-0.04]	[-1.26]	[0.41]	[46.1]	[21.5]
Piotroski and So	3.50	0.59	-0.00	-0.01	-0.03	0.77	0.45
	[2.36]	[0.58]	[-0.04]	[-0.83]	[-1.48]	[35.0]	[16.7]
Cheap defensive	1.32	0.85	0.03	-0.11	0.01	0.67	0.47
	[0.61]	[0.76]	[0.42]	[-6.00]	[0.53]	[32.5]	[26.5]
Profitable value	6.55	3.66	0.09	-0.04	0.26	0.82	0.75
	[4.98]	[3.62]	[1.63]	[-3.16]	[14.0]	[36.0]	[28.6]

Notes: This table shows alphas (percent per year) from four-factor time-series regressions of the form

$$y = \alpha + \beta_{\text{MKT}} MKT + \beta_{\text{SMB}} SMB + \beta_{\text{VMG}} VMG + \beta_{\text{SMI}} SMI + \epsilon,$$

where the test strategy y is formed on the basis of average value and quality ranks, and the the explanatory strategies include similarly constructed pure value and quality strategies, VMG (value-minus-growth) and SMI (superior-minus-inferior), respectively.

loadings, for signal-to-noise ratios ($\sigma_\beta^2/\sigma_\epsilon^2$) of one, 1/2, and 1/3 (solid, dashed, and dotted lines, respectively). The figure shows that the magnitude of the increase in the tilts that can be obtained by selecting stocks on the basis of the combined signals is greater when the true factor loadings are negatively correlated, and that the impact of the correlation is increasing in the quality of the signal. Whenever the true factor correlation in negative, however, the tilts that can be obtained by combining the signals is at least 40% higher than those that can be obtained by combining the pure strategy portfolios.

Table 10 shows empirical results consistent with this prediction. The table shows the performance of long/short strategies formed on the basis of joint value and quality signals,

and these strategies loadings on pure value and quality strategies (controlling for market and SMB loadings). The mean value loading is 0.73, and the mean quality loading is 0.53. The total loadings on value and quality thus average 1.27, 27% higher than the total loadings one would get by trading a 50/50 mix of the pure quality and value strategies side-by-side.

One could somewhat mitigate these differences by holding more concentrated pure factor portfolios. If one holds only half as many positions in the portfolios sorted on the individual signals, on the theory that the 50/50 mix will hold twice as many positions (though of course some positions will be common across the two strategies), the ratio becomes

$$\frac{n\left(N^{-1}(\alpha)\right)}{n(N^{-1}(\alpha/2))}\sqrt{\frac{1}{2\left(1+\rho\sigma_{\beta}^{2}/\sigma_{S}^{2}\right)}}.$$

This ratio is still always bigger than one for factors that have negatively correlated loadings across stocks when portfolios hold less than about 40% of stocks. This suggests that even with highly diversified portfolios investors can obtain larger tilts toward value and profitability selecting stocks on the basis of valuations and profitability metrics jointly than they can with an equally well diversified combination of pure value and profitability strategies. This difference is more pronounced for more concentrated strategies.

References

[1] Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang. 2006. “The Cross-Section Of Volatility and Expected Returns." Journal of Finance, 259–299.

[2] Baker, Malcolm, Brendan Bradley, and Jeffrey Wurgler. 2011. “Benchmarks as Limits to Arbitrage: Understanding the Low-Volatility Anomaly." Financial Analysts Journal, 40–54.

[3] Ball, Ray. 1978. “Anomalies in relationships between securities’ yields and yield-surrogates.” Journal of Financial Economics, 103–126.

[4] Black, Fisher. 1972. “Capital Market Equilibrium with Restricted Borrowing.” Journal of Business, 444–455.

[5] Black, Fisher. 1993. “Beta and Return.” Journal of Portfolio Management, 8–18.

[6] Black, Fischer, Michael C. Jensen, and Myron S. Scholes, “The Capital Asset Pricing Model: Some Empirical Tests," in Studies in the Theory of Capital Markets (Michael C. Jensen, editor), New York: Praeger, 1972, pp. 79–121.

[7] Berk, Jonathan B. 1995. “A critique of size-related anomalies.” Review of Financial Studies, 275–286.

[8] Campbell, John Y., Jens Hilscher, and Jan Szilagyi. 2008. “In Search of Distress Risk,” Journal of Finance, pp. 2899–2939.

[9] Chi, Joseph, and Jed Fogdall. 2012. “Integrated Equity Solutions,” Dimensional Fund Advisors Quarterly Institutional Review, Fourth Quarter.

[10] Frazzini, Andrea, David Kabiller, and Lasse H. Pedersen. 2012. “Buffett’s Alpha.” Working paper.

[11] Frazzini, Andrea, and Lasse H. Pedersen. 2012. “Betting Against Beta.” Journal of Financial Economics.

[12] Fama, Eugene F., and Kenneth R. French. 1993. “Common risk factors in the returns of stocks and bonds.” Journal of Finance, pp. 3–56.

[13] Fama, Eugene F., and Kenneth R. French. 1995. “Size and Book-to-Market Factors in Earnings and Returns.” Journal of Financial Economics, pp. 131–155.

[14] Fama, Eugene F., and Kenneth R. French. 2006. “Profitability, investment and average returns.” Journal of Financial Economics, pp. 491–518.

[15] Fama, Eugene F., and Kenneth R. French. 2008. “Dissecting anomalies.” Journal of Finance, pp. 1653–1678.

[16] Fama, Eugene F., and Kenneth R. French. 2013. “Incremental Variables and the Investment Opportunity Set.” Working paper.

[17] GMO. 2004. “The Case for Quality—The Danger of Junk.” GMO White Paper.

[18] Graham, Benjamin. 1973. The Intelligent Investor (4th Rev. ed.). Harpers & Row, New York, New York.

[19] Greenblatt, Joel. 2010. “The Little Book That Still Beats the Market.” John Wiley & Sons, Hoboken, New Jersey.

[20] Hasbrouck, J. “Trading costs and returns for U.S. equities: estimating effective costs from daily data.” 2009. Journal of Finance 64, pp. 1446-1477.

[21] Joyce, Chuck, and Kimball Mayer. 2012. “Profits for the Long Run: Affirming the Case for Quality.” GMO White Paper.

[22] Lapthorne, Andrew, Rui Antunes, John Carson, Georgios Oikonomou, and Charles Malafosse. 2012. “Global Quality Income Index: The Methodology.” Societe General White paper. http://www.structured-solutions.de/downloads/DE000SLA0SG2_leitfaden.pdf

[23] Mead, Katrina, Jonathan Sage, and Mark Citro. 2013. “Power Couple: Quality and Value are Strong Drivers of Long-Term Equity Returns.” MFS White Paper Series.

[24] Ng, Edmund. 2009. “A “New Magic” Formula—a ‘Combo’ Strategy Inspired By Greenblatt & Piotroski.” Morgan Stanley Research Paper.

[25] Novy-Marx, Robert. 2013. “The Other Side of Value: The Gross Profitability Premium.” Journal of Financial Economics, forthcoming.

[26] Novy-Marx, Robert, and Mihail Velikov. 2014. “Anomalies and their Trading Costs.” Working paper.

[27] Piotroski, Joseph D. 2000. “Value Investing: The Use of Historical Financial Statement Information to Separate Winners from Losers.” Journal of Accounting Research, pp. 1–41.

[28] Piotroski, Joseph D., and So, Eric C. 2012. “Identifying Expectation Errors in Value/Glamour Strategies: A Fundamental Analysis Approach.” Review of Financial Studies, forthcoming.

[29] Roll, R. “A simple implicit measure of the effective bid-ask spread in an efficient market.” 1984. Journal of Finance 39, pp. 1127-1139.