JOIM
www.joim.com

Journal Of Investment Management, Vol. 19, No. 2, (2021), pp. 52–88

THE MAGIC FORMULA: VALUE, PROFITABILITY, AND THE CROSS-SECTION OF GLOBAL STOCK RETURNS

Douglas W. Blackburn[a] *and Nusret Cakici*[b]

Buying profitable, undervalued stocks and shorting unprofitable, overvalued stocks yields significant return differentials in North American, Europe, Japan, and Asia. Using data from 1991 to 2016, double sorting stocks into portfolios by gross profits, a measure of profitability, and earnings yield, a measure of value, yields significant abnormal returns for all size groups and in all global regions after controlling for size, book-to-market, momentum, profitability, and investments factors. Abnormal returns persist after accounting for transaction costs, are larger during high sentiment periods and are present across different sectors and countries.

1 Introduction

Profitability and value characteristics describe fundamentally different but complimentary dimensions of a firm and therefore explain different aspects of the cross-section of returns. Both valuation metrics, such as book-to-market equity or earnings-to-price, and profitability metrics, such as net income excluding extraordinary items and gross profits-to-assets, have been shown to be significant determinants of the cross-section of returns with value stocks outperforming growth stocks and the stocks of profitable firms outperforming the stocks of unprofitable firms (Ball and Brown, 1968; Basu, 1983; Fama and French, 1992, 1993, 2012; Asness *et al.*, 2013, 2017; Novy-Marx, 2013, among many others). The set of value stocks, however, includes both profitable and unprofitable firms, and the set of profitable firms includes both relatively cheap and expensive stocks. All else equal, it is reasonable to posit that profitable firms selling at relatively cheap prices should yield higher returns than unprofitable firms selling at similarly cheap prices as the former should be more desirable to investors than the latter.[1] Likewise, all else equal, an unprofitable firm selling at relatively expensive price should yield smaller returns than a profitable firm

[a]JPMorgan Chase, 8181 Communications Parkway, Plano, TX 75024, Email: douglas.w.blackburn@gmail.com.
[b]Gabelli School of Business, Fordham University, 113 W 60th Street, New York City, NY 10023, Email: cakici@fordham.edu.

"What would happen if we decided to only buy shares in good businesses (ones with high returns on capital) but only when they were available at bargain prices (priced to give us high earnings yield)? What would happen? Well, I'll tell you what would happen: We would make a lot of money!" Joel Greenblatt, The Little Book that Still Beats the Market.

selling at the similarly expensive price as the former should be much less desirable than the latter. This suggests that one would do quite well to long profitable, value (cheap) stocks and short unprofitable, growth (expensive) stocks.

This value–profitability paradigm is quantified by Greenblatt (2006, 2010) who heralds the approach as the "magic formula" for beating the market. Because of its straightforward and unambiguous methodology, Greenblatt's magic formula lends itself to rigorous testing, and if effective, also to easy implementation. In this paper, we test Greenblatt's magic formula for its ability to generate positive risk-adjusted abnormal returns and to explain the cross-section of returns across global stock markets. To our knowledge, no rigorous, global study of the value–profitability paradigm currently exists.

Following the magic formula (MF) methodology, we construct a portfolio that is long the quintile portfolio of profitable, value stocks (PV) and is short the quintile portfolio of unprofitable, growth stocks (UG) which we refer to as PV–UG. Profitability is defined as the return on capital (EBIT-to-tangible capital employed) and value is defined using the earnings yield (EBIT-to-enterprise value). The PV–UG portfolio is constructed for the four global regions of North America, Europe, Japan, and Asia using stock return data representing 23 developed markets over the period January 1991 through December 2016 thus providing four distinct samples in which to test MF. Finding similar results across the four geographic regions increases confidence in the robustness of the results, decreases data mining concerns, and demonstrates a degree of global market integration. Furthermore, as finance practitioners have become more globally focused, understanding how the returns of stocks across global markets are similar and different is of primary interest.

We find that the magic formula, as described by Greenblatt (2006), is not very magical. While PV–UG yields positive and significant risk-adjusted returns in Europe, results are insignificant and sometimes negative in North America, Japan, and Asia, suggesting that one would do as well by holding a portfolio of unprofitable growth stocks as they would by holding a portfolio of profitable value stocks. This is quite different from the univariate results reported by Greenblatt for the U.S. market.

Motivated by Novy-Marx (2013), we make one modification to the magic formula. EBIT, the profitability measure used by Greenblatt, is replaced with gross profits, the profitability measure proposed by Novy-Marx.[2] Novy-Marx argues that profitability measures located farther down the balance sheet, such as EBIT, are noisy due to accounting items subtracted from earnings that may not be directly related to expenses incurred to generate revenues. Profitability measures located closer to the top of the income statement, such as gross profits, are therefore cleaner, less noisy measures of a firm's true economic profitability. This claim is not without dissent. Ball *et al.* (2015) challenge this assertion and find that different profitability measures, when deflated by the same value, perform similarly regardless of their location on the income statement. Hence, changing EBIT to gross profits in the magic formula can be considered a test of Novy-Marx's profitability hypothesis.

Our analysis of the improved magic formula (IMF) yields a number of interesting new results. First, the four regional value-weighted PV–UG portfolios formed using IMF yield positive and significant univariate return differentials and risk-adjusted abnormal returns for all global regions. Monthly alphas from the Fama–French four-factor model for North America, Europe, Japan, and Asia are 1.19 (t-stat = 3.91), 0.45 (t-stat = 2.08), 0.53 (t-stat = 3.56) and 0.82 (t-stat = 2.74). Profitable, value stocks significantly outperform unprofitable, growth stocks in all regions. Moreover, we find that the long-only strategy of holding the PV portfolio also yields significant risk-adjusted returns. The positive alphas remain after controlling for trading costs. Our results suggest that gross profits is a more powerful predictor of the cross-section of returns than EBIT, consistent with Novy-Marx's (2013) profitability hypothesis.

Second, the PV–UG portfolios constructed using IMF have negative and significant market betas. Counter to asset pricing theory, PV stocks yield higher average returns and have lower average betas than UG stocks. This finding is related to the low-beta anomaly where low-beta stocks have been shown to provide higher average returns than high beta stocks (Baker *et al.*, 2011; Frazzini and Pedersen, 2014).

Third, we test whether our results are driven by small stocks which tend to be illiquid, have high trading costs, and are often difficult to short. Such a finding would make it challenging to implement any strategy to capture the observed abnormal returns. Results show that PV–UG return differentials are large and statistically significant for both big and small stocks, though spreads are larger within the set of small stocks. This is an interesting result in light of Fama and French (2012) who find returns to value and momentum strategies to be primarily found only in the set of small (microcap) stocks. In contrast, IMF yields positive and significant risk-adjusted returns for both small and big stocks.

Fourth, we test whether size, book-to-market, and momentum are able to explain the relationship between IMF and the cross-section of returns. Results from independent double sorts and Fama–MacBeth regressions suggest that IMF captures a unique dimension of the cross-section of returns. Twenty-five portfolios are constructed by independently double sorting on IMF and market capitalization (size) as well as on IMF and book-to-market equity (B/M). We find significant IMF return differentials for nearly all size and most B/M quintiles. Using the Fama–MacBeth multivariate regression framework, the cross-section of returns is regressed on IMF, size, B/M, and momentum. For all regions, IMF is highly significant. This is true whether the test is performed on the set of all stocks or on the subsets of large-cap and small-cap stocks.

We conduct a series of additional tests to measure the robustness of the IMF methodology. Computing return differentials on an annual basis reveals that the IMF performance is not a result of a particular time period, and we find that PV–UG risk-adjusted returns are not dominated by any one particular sector or country. There is evidence that performance is increased by changing from quintile to decile portfolios as the decile portfolios focus on the more extreme PV and UG stocks, and alpha is robust to the alternative factor models which includes profitability and investments factors. Last, we find that PV–UG return differentials are larger during periods of high sentiment than periods of low sentiment which is driven more so by the short side than the long side. Together, our evidence shows that profitable, value stocks yield significantly greater returns than unprofitable, growth stocks. This result is consistent across all four global regions studied here thus demonstrating that the magic formula is indeed magical in its ability to predict the cross-section of returns globally.

2 Magic formula: The setup

2.1 Data

Stock return and fundamental data for 23 countries are from Datastream for the sample period January 1991 through December 2016 and include both large and small-cap stocks and both active and inactive firms.[3] The 23 countries represent the developed markets as classified by MSCI. This is the same set of countries studied by Fama and French (2012) with the one exception that our set replaces Greece with Israel. Countries are grouped into four regions:

(1) North America—United States and Canada,
(2) Europe—Austria, Belgium, Denmark, Finland, France, Germany, Ireland, Israel, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom,
(3) Japan
(4) Asia—Australia, New Zealand, Hong Kong, and Singapore.

Grouping countries into regions is necessary to ensure portfolios with a sufficiently large number of stocks (Fama and French, 2012, 2017).[4] All returns are denominated in U.S. dollars.[5]

2.2 *Constructing the magic formula*

It has long been the investor's goal to purchase high-quality, profitable companies at cheap prices. This is the objective of Greenblatt's magic formula. There are a number of econometric methods and accounting measures that could be used to achieve this. To alleviate data mining concerns that would certainly arise if we were to systematically produce results for all possible permutations of methodologies and metrics, we specifically study and implement the approach advocated by Greenblatt (2006, 2010) making compromises only when necessary. One such compromise is with the construction of certain accounting measures. Datastream constructs their variables in the attempt to "enhance the comparability of financial data of companies from different countries and industries and across periods of time." As comparability across countries and global regions is of primary importance, we opt to use the various accounting variables as constructed by Datastream even though the variable definitions may differ from those espoused by Greenblatt for the use in the United States.

The MF, as described by Greenblatt, requires two financial ratios. The first is the profitability measure return on capital (ROC) defined as EBIT divided by tangible capital employed, the sum of net working capital and net fixed assets. EBIT and net working capital both come directly from Datastream.[6] For net fixed assets, we use property, plants, and equipment also available from Datastream.

While different profitability measures have been proposed in the literature, such as return on equity or return on assets, Greenblatt argues that ROC is the appropriate profitability metric. EBIT, since it does not include taxes and interest, is preferred over earnings. Taxes and interest, which do not describe the firm's operating efficiency, can vary substantially across firms and would add substantial noise if included. This is particularly important for international studies where tax and interest rates can differ widely from country to country. The choice of deflator, tangible capital employed, measures the total amount of capital required to conduct the firm's business. Tangible capital employed excludes intangibles, specifically goodwill, which is a historic cost related to acquisitions and therefore has no bearing on the current or future operation of the firm. The more commonly used deflator, total assets, includes intangibles making it a noisier measure of the assets required to operate the company.

The measure of value, earnings yield, is defined as EBIT divided by enterprise value—earnings (before interest and tax) relative to the purchase price of the firm.[7] We use the enterprise value data item available from Datastream. It has already been argued that EBIT is preferred to earnings as in the price-to-earnings valuation ratio. The deflator, enterprise value, takes into account both the market value of the equity and the debt financing used to fund projects that generate earnings for the firm.[8] Including both debt and equity in the denominator makes the metric neutral to differences in capital structure. Other valuation ratios, such as the price-to-earnings ratio or the book-to-market ratio, may lead to questionable conclusions when comparing across firms with large differences in capital structure.

For each month t, all N firms within a particular region are ranked from 1 to N based on the ROC observed in month $t - 7$. The firm with the highest ROC is given the rank of 1 and the firm with the lowest ROC is given the rank of N. Lagging by seven months ensures that the data is actually available to be used to form portfolios in month t.[9] Firms are then independently ranked from 1 to N based on earnings yields also observed in month $t - 7$. The firm with the highest earnings yield (value) is assigned the rank of 1 and the lowest earnings yield firm (growth) is given the rank of N. As the final step, the two rankings are added together to form the MF ranking and all stocks are sorted into quintile portfolios according to MF. Stocks in the first quintile are profitable, value stocks (PV) and the stocks in the fifth quintile are unprofitable, growth stocks (UG). The portfolio formed from longing PV and shorting UG is referred to as PV–UG.

Our implementation of the magic formula modestly deviates from Greenblatt's proposed methodology in two ways. Greenblatt recommends picking the top 30 PV stocks each month and then holding the portfolio for one year thus creating a series of overlapping portfolios. We instead use quintile portfolios that are rebalanced every month. This variation of the methodology leads to a more challenging test of the magic formula. Holding quintile portfolios rather than a portfolio of 30 stocks creates less of a distinction between the PV and UG portfolios. Additionally, value strategies often require patience to allow time for the market to realize the quality of undervalued stock. The one-month holding period does not allow much time for the prices of value stocks to reverse. Ultimately, the optimal specification of the magic formula is an empirical issue.

Novy-Marx (2013), in his study of profitability as a predictor of the cross-section of returns, considers a different approach to combining value and profitability. Novy-Marx double sorts all U.S. stocks by gross profits-to-assets (profitability) and book-to-market equity (value) into 5×5 portfolios. Using this approach, the portfolio of profitable, value stocks yields an average excess return of 1.08% while the unprofitable, growth portfolio yields −0.08%. Our study is different as we focus specifically on Greenblatt's MF methodology and its ability to predict the cross-section of returns in global markets.

2.3 *Factor construction*

For each region, we consider four region-specific asset pricing factors: the excess return on the value-weighted market portfolio, small minus big market capitalization (SMB), high minus low book-to-market equity (HML), and the winner minus loser momentum factor (WML). Factors are constructed for each region closely following the methodology used by Fama and French (2012). For month t, size is the market capitalization from month $t - 1$. For B/M, we use the ratio available on Datastream which computes B/M each month using the currently available book value equity and the current month's market value of equity. For each month t, we use B/M from month $t - 7$. Momentum is the cumulative stock return over months $t - 12$ to $t - 2$.

The market factor, Mkt, is the value-weighted portfolio of all stocks within the region less the

Table 1 Descriptive statistics.

	North America					Europe					Japan					Asia				
	Size	*B*/*M*	ROC	EY	*N*	Size	*B*/*M*	ROC	EY	*N*	Size	*B*/*M*	ROC	EY	*N*	Size	*B*/*M*	ROC	EY	*N*
1991	5.47	0.83	16.27	10.46	1,821	4.50	0.86	19.31	12.20	2,697	6.50	0.37	17.49	5.26	1,260	5.23	1.00	21.16	12.18	248
1992	5.60	0.71	14.67	9.16	1,893	4.45	0.93	14.63	11.54	2,819	5.95	0.47	15.92	6.32	1,582	5.26	0.90	16.86	10.11	279
1993	5.72	0.65	14.10	7.85	2,036	4.44	1.08	13.77	10.31	2,869	5.99	0.62	13.45	6.68	1,756	5.36	0.87	16.10	8.81	320
1994	5.77	0.58	14.34	7.64	2,159	4.73	0.78	12.63	8.98	2,932	6.13	0.53	11.33	5.46	1,807	5.65	0.70	15.07	7.65	343
1995	5.71	0.61	15.29	8.09	2,499	4.79	0.72	14.61	8.89	3,024	6.01	0.56	9.77	4.65	1,883	5.50	0.80	15.76	7.62	382
1996	5.68	0.56	16.22	8.25	2,914	4.88	0.75	15.58	10.02	3,046	6.00	0.61	8.98	5.06	1,979	5.46	0.84	15.23	8.20	482
1997	5.77	0.55	15.56	7.50	3,144	4.91	0.72	14.93	9.46	3,216	5.54	0.65	8.81	4.97	2,047	5.38	0.80	13.74	7.33	612
1998	5.76	0.48	15.45	6.97	3,250	4.86	0.63	14.75	8.84	3,498	5.07	1.03	8.76	5.67	2,080	4.66	0.94	12.87	7.11	681
1999	5.64	0.63	15.18	6.62	3,305	4.74	0.71	15.78	8.59	3,599	5.12	1.26	7.21	5.66	2,385	4.67	1.49	10.65	7.64	720
2000	5.67	0.66	14.04	6.39	3,362	4.79	0.70	14.60	7.94	3,635	4.95	1.20	6.59	5.08	2,717	4.66	1.06	10.02	6.42	774
2001	5.54	0.73	13.78	6.31	3,437	4.51	0.73	13.48	6.93	3,822	4.71	1.35	6.83	5.36	2,794	3.97	1.17	9.61	6.43	1,020
2002	5.51	0.71	11.52	5.31	3,508	4.30	0.82	11.94	6.35	3,919	4.52	1.46	6.45	5.56	2,861	3.54	1.17	5.73	4.21	1,487
2003	5.61	0.79	10.27	5.75	3,512	4.39	1.03	10.88	6.56	3,679	4.59	1.60	6.19	5.51	2,937	3.46	1.25	4.00	3.64	1,809
2004	5.97	0.58	11.29	5.30	3,628	4.83	0.79	10.80	6.62	3,553	5.01	1.27	8.01	6.79	2,984	3.83	0.89	5.51	4.15	1,921
2005	6.08	0.54	13.84	5.66	3,705	5.02	0.68	12.65	6.79	3,537	5.21	1.03	9.85	7.17	3,105	3.91	0.84	8.84	5.56	2,086
2006	6.04	0.60	14.00	5.44	3,952	5.11	0.63	14.89	6.74	3,692	5.28	0.81	10.43	6.46	3,234	4.05	0.84	9.61	5.73	2,266
2007	5.95	0.67	13.56	5.36	4,204	5.24	0.62	17.39	6.48	4,001	5.06	0.89	12.08	7.51	3,315	4.53	0.76	10.95	5.58	2,391
2008	5.59	0.67	12.11	5.03	4,282	4.86	0.71	18.30	6.76	4,213	4.79	1.10	12.54	9.08	3,318	4.18	0.67	11.78	4.52	2,565
2009	5.19	1.08	8.57	5.01	4,265	4.50	1.23	14.77	8.07	4,064	4.71	1.58	9.30	8.93	3,152	3.86	1.39	7.13	4.92	2,558
2010	5.55	0.83	3.91	2.29	4,334	4.79	1.08	9.90	5.56	3,881	4.84	1.48	5.14	5.34	3,098	4.37	0.99	2.59	1.75	2,645
2011	5.71	0.73	8.10	4.36	4,400	4.96	0.98	13.32	6.75	3,760	4.99	1.53	9.03	9.60	3,008	4.57	0.92	7.82	4.86	2,720
2012	5.55	0.76	9.03	4.94	4,460	4.84	1.05	14.76	7.45	3,632	5.00	1.55	8.97	10.58	2,936	4.38	1.14	7.06	4.85	2,769
2013	5.58	0.87	7.87	4.41	4,438	5.05	1.08	12.81	6.94	3,450	5.10	1.50	9.07	10.97	2,910	4.41	1.19	5.21	3.95	2,744
2014	5.81	0.84	7.18	3.62	4,336	5.31	1.05	12.55	5.89	3,331	5.16	1.26	10.26	9.76	2,999	4.55	1.16	5.20	3.89	2,677
2015	5.82	0.83	7.46	3.51	4,251	5.29	1.11	12.53	5.75	3,243	5.16	1.13	10.99	9.31	3,074	4.65	1.09	6.01	3.96	2,664
2016	6.04	1.03	6.58	3.33	3,872	5.46	1.13	13.60	5.62	2,891	5.19	1.12	11.38	9.08	3,061	4.91	1.19	6.67	3.83	2,424
Mean	5.71	0.71	11.93	5.94	3,499	4.83	0.87	14.04	7.77	3,462	5.25	1.08	9.80	6.99	2,626	4.58	1.00	10.05	5.96	1,600

Twenty-three countries are divided into four regions: North America (U.S. and Canada), Europe (Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the U.K.), Japan, and Asia (Australia, New Zealand, Hong Kong, and Singapore). For each year, this table lists the average log market capitalization (size), book-to-market equity (*B*/*M*), return on capital (ROC), earnings yield (EY), and the number of firms for each region. ROC is EBIT divided by the sum of net working capital and net fixed assets. EY is EBIT divided by enterprise value. The last row is the time-series average across all years.

return on the three-month U.S. treasury. For the size and B/M factors, SMB and HML, all stocks are sorted into two size and three B/M groups. Small stocks are those in the bottom 10% of market capitalization and big stocks are those in the top 90%. B/M breakpoints are determined by sorting the big stocks into the bottom 30% (growth), middle 40% (neutral), and top 30% (value). The breakpoints are then applied to the sets of small and big stocks to form six value-weighted portfolios: SV, SN, SG, BV, BN, and BG. B and S denote big and small market capitalization stocks, respectively, and V, N, and G denote value, neutral, and growth, respectively. SMB is the portfolio SV + SN + SG − BV − BN − BG. For B/M, we construct $\text{HML}_S = \text{SV} - \text{SG}$ which is the high minus low B/M portfolio comprised only of the small stocks and $\text{HML}_B = \text{BV} - \text{BG}$ which is the high minus low B/M portfolio of only big stocks. HML is the average of HML_S and HML_B.

The momentum factor, WML, is constructed in a similar fashion as HML. The set of big stocks are sorted into the bottom 30% (losers, L), middle 40% (neutral, N), and top 30% (winners, W) according to momentum. The momentum breakpoints computed from the big stocks are then applied to both the big and small stocks to create six value-weighted portfolios: SL, SN, SW, BL, BN, and BW. Three momentum portfolios are constructed. $\text{WML}_S = \text{SW} - \text{SL}$ is comprised only of small stocks, and $\text{WML}_B = \text{BW} - \text{BL}$ includes only big stocks. WML is the average of WML_S and WML_B.

2.4 Descriptive statistics

Year-by-year descriptive summary statistics are provided in Table 1 for each of the four global regions. On average, the sample includes 11,000 stocks each year distributed across the four regions: 3,499 North American firms, 3,462 European firms, 2,626 Japanese firms, and 1,600 Asian firms. There are large differences in average log market capitalization across regions. North America is by far the largest with average log market cap of 5.71 ($300.65 million), Japan is the next largest with average log market cap 5.25 ($191.03 million). The average log market caps of Europe and Asia are considerably smaller at 4.83 ($125.21 million) and 4.58 ($97.51 million), respectively. There are also substantial cross-region differences in ROC and earnings yield. ROC is on average largest for Europe (14.04%) followed by North America (11.93%), Asia (10.05%), and Japan (9.80%). Europe has the largest earnings yield, 7.77%, followed by Japan (6.99%), Asia (5.96%), and North America (5.94%).

3 Magic formula results

3.1 The not so magical magic formula

Greenblatt (2010) provides year-by-year results from 1988 to 2004 that demonstrate the superiority of his magic formula relative to both the S&P 500 and an equal-weighted portfolio of 1,000 U.S. stocks. Using a sample of U.S. stocks, he reports an average annual return of 30.8% for his magic formula compared to an average annual return of 11.7% for the portfolio of 1,000 stocks and 12.4% for the S&P 500. We test the magic formula using four distinct and well-defined global regions; hence, this can be considered as a set of independent out-of-sample tests of the magic formula. Furthermore, as asset managers have become more globally focused, understanding how MF performs globally is of great interest.

Following the MF methodology, we conduct univariate tests of the magic formula by sorting all stocks in each region equally into quintiles.

Profitable, value (PV) stocks comprise the 1st quintile and unprofitable, growth (UG) stocks comprise the 5th quintile. Both equal-weighted and value-weighted portfolios are formed from the stocks within each quintile. Table 2 lists the results of the univariate sorts for each region. The table includes average monthly portfolio returns, both value- and equal-weighted, as well as the average size, B/M, market share, and number of stocks for each quintile portfolio. The columns labeled EW and VW are the average monthly returns of the equal-weighted and value-weighted portfolios, respectively. Size is the average log market capitalization of the portfolio, B/M is the average book-to-market equity, M-Shr is the average fraction of total market capitalization captured by the stocks in the portfolio, and N is the average number of stocks in each portfolio.

We begin by first describing the characteristics of the PV and UG portfolios. There is a large difference in market capitalization between the two extreme portfolios. PV firms are on average much larger than UG firms. This is best seen from the M-Shr column. Stocks in the PV portfolio account for 30% of the North American market (Panel A) compared to only 4% for the UG stocks. This holds consistently across all regions −20% vs 6% for Europe, 17% vs 11% in Japan, and 19% vs 4% in Asia. If we consider size as a proxy for liquidity, then PV stocks are likely to be more liquid than UG stocks. What is surprising is the difference in B/M across PV and UG portfolios. Since stocks are sorted by their valuation using the earnings yield measure, it would be reasonable to assume that PV stocks would have greater B/M than UG stocks. This is not the case. On average there is little difference in B/M between the portfolios. Average B/M for North America PV is 0.49 as compared to 0.48 for UG. There is some difference in Japan with B/M being a little larger for UG relative to PV, but in Asia, PV has on average greater B/M relative to UG. Last, we see that each portfolio holds an equal number of stocks which range from 320 to 700 depending on the region. As previously discussed, this is a departure from Greenblatt who recommends holding just 30 stocks in the PV portfolio.

Does the magic formula live up to its name? Table 2 clearly shows that it does not. Beginning in North America, Panel A, the PV–UG spread is −0.36% when equal-weighting and is 0.29% when value-weighting. Not only are both values statistically indifferent from zero, the equal-weighted return differential is negative. Similar results are found for Japan and Asia. Japan's equal-weighted and value-weighted return differentials, PV–UG, are −0.03% and 0.11%, respectively, and for Asia they are 0.09% and 0.41%, respectively. All return differentials are statistically insignificant. The one bright spot in Table 2 is found in Europe, Panel B. The equal-weighted PV–UG portfolio yields a statistically significant return of 0.95% per month (t-stat = 4.75) and the value-weighted portfolio yields 0.59% per month (t-stat = 2.14).

Table 2 also lists risk-adjusted returns estimated using three different empirical asset pricing models. The first is Sharpe's (1964) single-factor Capital Asset Pricing Model (CAPM) that uses Mkt as the market portfolio. Next is the Fama–French three-factor model (FF-3) that includes regionally-defined SMB and HML factors in addition to Mkt (Fama and French, 1993). We also use the Fama–French four-factor model (FF-4) that includes the region-specific WML (winners minus losers) factor in addition to the Fama–French three factors (Carhart, 1997). The intercept, alpha, from regressing PV–UG on the different sets of factors is reported in Table 2 along with their Newey–West t-statistics (Newey and West, 1987).

Table 2 The magic formula.

	Panel A: North America						Panel B: Europe					
Quintile	EW	VW	Size	*B/M*	M-Shr	*N*	EW	VW	Size	*B/M*	M-Shr	*N*
1 (PV)	1.50	1.10	6.21	0.49	0.30	700	1.28	0.99	4.75	0.56	0.20	693
2	1.33	0.90	6.58	0.48	0.34	700	1.03	0.90	5.37	0.56	0.33	692
3	1.33	0.84	6.35	0.54	0.22	700	0.85	0.75	5.32	0.62	0.27	692
4	1.32	0.71	5.23	0.59	0.11	700	0.58	0.56	4.52	0.73	0.14	692
5 (UG)	1.86	0.81	4.01	0.48	0.04	700	0.34	0.40	3.48	0.61	0.06	692
PV–UG	−0.36	0.29					0.95	0.59				
	(−0.94)	(0.76)					(4.75)	(2.14)				
	$CAPM_{EW}$	$FF\text{-}3_{EW}$	$FF\text{-}4_{EW}$	$CAPM_{VW}$	$FF\text{-}3_{VW}$	$FF\text{-}4_{VW}$	$CAPM_{EW}$	$FF\text{-}3_{EW}$	$FF\text{-}4_{EW}$	$CAPM_{VW}$	$FF\text{-}3_{VW}$	$FF\text{-}4_{VW}$
Alpha	−0.07	−0.21	−0.18	0.70	0.47	0.46	1.01	0.92	0.67	0.74	0.61	0.24
	(−0.19)	(−0.65)	(−0.56)	(1.93)	(1.77)	(1.63)	(5.43)	(5.61)	(4.16)	(3.07)	(2.68)	(0.87)
	Panel C: Japan						Panel D: Asia					
1 (PV)	0.75	0.47	5.03	0.87	0.17	526	1.52	1.09	4.74	0.73	0.19	320
2	0.61	0.42	5.30	0.92	0.27	525	1.20	0.95	5.15	0.74	0.34	320
3	0.49	0.21	5.29	0.96	0.25	525	1.13	0.87	5.02	0.86	0.32	320
4	0.53	0.37	5.12	1.01	0.21	525	1.00	0.71	4.13	0.77	0.11	320
5 (UG)	0.77	0.36	4.43	1.01	0.11	525	1.42	0.67	3.15	0.67	0.04	320
PV–UG	−0.03	0.11					0.09	0.41				
	(−0.16)	(0.45)					(0.34)	(1.09)				
	$CAPM_{EW}$	$FF\text{-}3_{EW}$	$FF\text{-}4_{EW}$	$CAPM_{VW}$	$FF\text{-}3_{VW}$	$FF\text{-}4_{VW}$	$CAPM_{EW}$	$FF\text{-}3_{EW}$	$FF\text{-}4_{EW}$	$CAPM_{VW}$	$FF\text{-}3_{VW}$	$FF\text{-}4_{VW}$
Alpha	−0.01	0.09	0.00	0.13	0.25	0.25	0.27	0.08	0.12	0.65	0.42	0.66
	(−0.06)	(0.60)	(−0.01)	(0.52)	(1.08)	(1.08)	(1.04)	(0.32)	(0.55)	(1.92)	(1.20)	(2.53)

In each month t and for each region (North America, Europe, Japan, and Asia), stocks are ranked by return on invested capital and by EBIT-to-enterprise value (EBIT/EV). The two rankings are then added together to form a combined rank and the stocks are sorted into quintiles according to the combined ranking using the 20, 40, 60, and 80 percentile breakpoints. Time t returns of the equal-weighted (EW) and value-weighted (VW) portfolios comprising the stocks in each quintile are averaged across all months from January 1991 to December 2016. Size is the average log market capitalization, B/M is the average book-to-market equity, M-Shr is the fraction of the total market capitalization in each quintile, and N is the number of stocks. The return difference between the 1st and 5th quintiles (PV–UG) is reported. Alphas are the risk-adjusted return from regressing the PV–UG portfolio returns, both EW and VW, on the regional market portfolio (CAPM), on the market, SMB, and HML (FF-3), and on the market, SMB, HML, and momentum (FF-4). Newey–West t-statistics are reported in parentheses.

Alpha, when using the equal-weighted European PV–UG is large and highly significant for all three model specifications. Specifically, the FF-4 EW monthly alpha is 0.67% with a t-statistic of 4.16. For the value-weighted results, while alpha remains large and significant when using the CAPM and FF-3, but alpha loses significance when using the FF-4 model. For North America, Japan, and Asia, alpha is nearly always indistinguishable from zero.

The overall results presented in Table 2 are different from the results presented by Greenblatt. The magic formula is not very magical when applied to global stock returns over our 1991–2016 sample period. While results are quite strong for Europe, the strategy is not profitable in North America, Japan, and Asia. Moreover, return differentials are sometimes even negative. These results, however, are subject to the joint hypothesis problem. The failure of Greenblatt's MF may be due to an ineffectual strategy, or it may be a result of our particular implementation of the strategy. For example, rebalancing annually rather than monthly may provide the necessary time for the market to identify the undervalued, profitable stocks and then to act on the information.

3.2 *A much more magical, magic formula*

Recent research on profitability may help resurrect the magic formula. Novy-Marx (2013) argues that profitability measures located farther down the income statement are "polluted." Hence, the cleanest measure of the true economic profitability must be one found at the top of the income statement—gross profits (GP). This suggests that we may be able to resurrect Greenblatt's MF by using GP instead of EBIT as the measure of profitability.[10]

Novy-Marx's claim is not without challenge. Ball *et al.* (2015) argue that Novy-Marx's conclusion is founded on the flawed comparison between GP-to-assets and earnings-to-book equity. Since both the profitability measures and their deflators differ across the two ratios, it is unclear whether the superiority of GP-to-assets is due to the choice of profitability measure, GP vs earnings, or the choice of deflator, assets vs book equity. Ball *et al.* find that GP performs as well as other profitability measures regardless of where they appear on the income statement when the same deflator is used.

In this section, we test a variation of MF referred to as the improved magic formula (IMF). The only difference between MF and IMF is the replacement of EBIT with GP in both the profitability and value measures. Instead of EBIT-to-enterprise value, we now use GP-to-enterprise value as the valuation measure, and instead if EBIT-to-tangible capital employed, we now use GP-to-tangible capital employed. Consistent with the initial construction methodology, we use GP from month $t-7$ to ensure its availability at time t. All other aspects of the MF methodology remain the same.

Table 3 lists the results for IMF. Looking first at the portfolio characteristics, we find no clear relation between the market capitalizations of the PV and UG portfolios. For North America and Europe, the average market share (M-Shr) of PV is nearly the same as UG—12% vs 9% for North America and 13% vs 11% for Europe. In Japan, M-Shr is much larger for UG, but in Asia, M-Shr is much larger for PV. This is different relative to MF where we found the UG portfolios to be comprised of much smaller stocks than the PV portfolios. Similar to our result for MF, we do not find any consistent relation between the PV and UG portfolios with B/M. B/M is a little larger for PV than UG for North America, Japan, and Asia, but UG has a larger average B/M in Europe.

Unlike with MF, IMF yields significant return differentials between the PV and UG portfolios.

Table 3 Univariate sorts—The improved magic formula (IMF).

	Panel A: North America						Panel B: Europe					
Quintile	EW	VW	Size	B/M	M-Shr	N	EW	VW	Size	B/M	M-Shr	N
1 (PV)	1.97	1.17	5.25	0.57	0.12	700	1.26	1.08	4.23	0.62	0.13	667
2	1.66	1.10	5.96	0.50	0.31	700	0.96	0.89	5.03	0.58	0.26	667
3	1.42	0.82	6.24	0.48	0.28	700	0.84	0.74	5.17	0.59	0.27	667
4	1.16	0.83	6.24	0.51	0.20	700	0.77	0.76	5.01	0.64	0.23	667
5 (UG)	1.17	0.30	4.45	0.51	0.09	700	0.39	0.43	4.12	0.66	0.11	666
PV–UG	0.79	0.88					0.87	0.65				
	(2.08)	(2.74)					(6.75)	(3.25)				
	$CAPM_{EW}$	$FF\text{-}3_{EW}$	$FF\text{-}4_{EW}$	$CAPM_{VW}$	$FF\text{-}3_{VW}$	$FF\text{-}4_{VW}$	$CAPM_{EW}$	$FF\text{-}3_{EW}$	$FF\text{-}4_{EW}$	$CAPM_{VW}$	$FF\text{-}3_{VW}$	$FF\text{-}4_{VW}$
Alpha	1.00	0.96	0.95	1.19	1.16	1.19	0.90	0.87	0.75	0.75	0.62	0.45
	(2.66)	(2.50)	(2.39)	(4.01)	(4.05)	(3.91)	(7.25)	(7.22)	(6.60)	(4.01)	(3.59)	(2.08)
	Panel C: Japan						Panel D: Asia					
1 (PV)	0.84	0.91	4.67	0.96	0.10	517	1.66	1.02	4.39	0.77	0.15	315
2	0.77	0.46	4.89	0.99	0.16	517	1.26	1.08	4.81	0.76	0.28	314
3	0.58	0.27	5.08	0.96	0.19	517	1.14	0.86	5.14	0.79	0.34	314
4	0.54	0.34	5.37	0.94	0.25	517	0.93	0.73	4.67	0.80	0.19	314
5 (UG)	0.44	0.25	5.35	0.87	0.31	517	1.34	0.40	3.21	0.68	0.04	314
PV–UG	0.40	0.66					0.32	0.62				
	(3.30)	(3.98)					(1.00)	(1.50)				
	$CAPM_{EW}$	$FF\text{-}3_{EW}$	$FF\text{-}4_{EW}$	$CAPM_{VW}$	$FF\text{-}3_{VW}$	$FF\text{-}4_{VW}$	$CAPM_{EW}$	$FF\text{-}3_{EW}$	$FF\text{-}4_{EW}$	$CAPM_{VW}$	$FF\text{-}3_{VW}$	$FF\text{-}4_{VW}$
Alpha	0.41	0.44	0.36	0.66	0.58	0.53	0.59	0.30	0.25	1.03	0.75	0.82
	(3.49)	(3.54)	(3.15)	(4.15)	(3.74)	(3.56)	(2.08)	(1.12)	(0.99)	(3.02)	(2.24)	(2.74)

In each month t and for each region (North America, Europe, Japan, and Asia), stocks are ranked by gross profits-to-tangible capital employed and by gross profits-to-enterprise value. The two rankings are then summed to form a combined rank. Stocks are sorted into quintiles according to the combined ranking using the 20, 40, 60, and 80 percentile breakpoints. Time t returns of the equal-weighted (EW) and value-weighted (VW) portfolios comprising the stocks in each quintile are averaged across all months from January 1991 to December 2016. Size is the average log market capitalization, B/M is the average book-to-market equity, M-Shr is the fraction of the total market capitalization in each quintile, and N is the number of stocks. The return difference between the 1st and 5th quintiles (PV–UG) is reported. Alphas are the risk-adjusted return from regressing the PV–UG portfolio returns, both EW and VW, on the regional market portfolio (CAPM), on the market, SMB, and HML (FF), and on the market, SMB, HML, and momentum (FFC). Newey–West t-statistics are reported in parentheses.

Whether equal-weighted or value-weighted, the PV–UG return differentials are positive, economically large, and statistically significant for all countries except Asia. The value-weighted monthly PV–UG return differentials for North America, Europe, Japan, and Asia are 0.88% (t-stat = 2.74), 0.65% (t-stat = 3.25), 0.66% (t-stat = 3.98), and 0.62% (t-stat 1.50), respectively. While the differential is not significant in Asia, the magnitude of the spread is on par with Europe and Japan. The lack of significance is due to the higher volatility we observe in the return data. Surprisingly, for all regions except Europe, the PV–UG spreads are larger when value-weighting than when equal-weighting.

Alphas from the CAPM, FF-3, and FF-4 models for both the equal-weighted and value-weighted PV–UG portfolios are listed at the bottom of each panel in Table 3. We observe that alphas are significant for nearly all model specifications and for both value- and equal-weighted portfolios. For North America, Europe, Japan, and Asia, the value-weighted monthly alphas from the FF-4 model are 1.19% (t-stat = 3.91), 0.45% (t-stat = 2.08), 0.53% (t-stat = 3.56), and 0.82% (t-stat = 2.74), respectively. As a comparison, Petajisto (2013) reports average U.S. mutual fund alphas over the 1990–2009 period of 0.31% annually (0.02% monthly), and Joenväärä *et al.* (2016) find that U.S. hedge funds earn an annual alpha of 4.00% (0.33% monthly) over the period 1994–2012. The simple strategy of sorting by profitability and value greatly outperforms the average actively managed mutual fund and hedge fund, and the consistency in results across models and across regions provides great confidence in the ability of IMF to generate returns above what is expected as compensation for risk.

The alphas reveal another interesting and surprising result—alphas when value-weighting tend to be larger than the alphas when equal-weighting. This is the case for all regions except Europe and it implies that the observed risk-adjusted returns are not a result of a small firm effect. Instead, large-cap stocks contribute significantly to the alpha. This is different from what is generally found in the empirical asset pricing literature. For example, Fama and French (2012) find that the value premium and momentum returns are only significant within the set of small-capitalization stocks. We investigate the relation between size and IMF in greater detail in the next section.

It is of great interest of long-only portfolios managers whether the significant alphas reported in Table 3 for the PV–UG portfolios come from longing PV or shorting UG. We report alphas computed using FF-4 model for all equal- and value-weighted quintile portfolios in Table 4. The PV (quintile 1) portfolios yield positive and significant alphas for all regions when using both equal- and value-weights. Equal-weighted alphas are much larger that the value-weighted alphas with the exception of Japan where the magnitudes are similar.

On the short side, alphas from UG (quintile 5), on the other hand, are less consistent. For North America, alpha is negative and significant when value-weighting but not when equal-weighting, but in Europe, the reverse is true. UG alphas in Japan are both insignificant, and for Asia, alphas when equal-weighting are positive and significant while alphas when value-weighting are negative and significant.

Regarding the debate on whether GP has greater predictive power than profitability measures located farther down the income statement. When using EBIT, MF fails to generate significant abnormal returns that are consistent across global regions; however, when we use GP instead of EBIT, the results are indeed magical. Alphas using GP are economically large and statistically significant, and the results are consistent across all four global regions. Hence, in the context of the magic formula, Novy-Marx's (2013) gross profits profitability measure, located at the top of the income statement, outperforms EBIT which is located farther down the income statement. It is important to note that the comparison is fair since only the one change was made in the methodology. Specifically, we use the same deflators for both the profitability and value measures when comparing MF to IMF.

3.3 Turnover and trading costs

The IMF is useful only if the alpha generated is larger than the cost associated with the implementation of the strategy. Trading costs, however, are difficult to evaluate. First, while the explicit transaction costs may be somewhat easy to measure, implicit costs, such as price impact, are not. Second, the methodology implemented here is not optimized to minimize trading costs, and it is well known that including trading costs into the portfolio construction methodology leads to significant reductions in cost (Novy-Marx and Velikov, 2015). Instead of attempting to compute a trading cost, which depends on a variety of factors including the size of the portfolio, we report turnover computed as the percentage of the portfolio that is bought and sold each month averaged across all months from 1991 through 2016. This is a valuable metric as it readily allows for the estimation of trading costs. For example, if turnover is 20% and the portfolio manager's average roundtrip trading cost is 1%, then the strategy is profitable so long as alpha is greater than 20 basis points (bps).

Table 4 reports turnover for each quintile when value-weighting and equal-weighting. An initial glance of the numbers reveals turnover to be quite small. This is likely due to the use of accounting data which is updated on an annual basis. Turnover tends to be modestly larger when value-weighting than when equal-weighting. For the purpose of discussion, we assume an average roundtrip trading cost for the entire portfolio of 1%. Looking first at North America, the IMF long-only strategy of buying the PV (quintile 1) portfolio has a cost of 9 bps and 17.8 bps for equal- and value-weighting, respectively. This is compared to the Fama–French four-factor (FF-4) model alpha of are 81 bps and 27 bps, respectively. Therefore, at a 1% trading cost, the long-only IMF survives by a healthy margin of 72 bps when equal-weighting and 9 bps when value-weighting.

Conducting the same analysis for the other three regions yields similar results. Assuming 1% trading costs in Europe requires a long-only alpha of 8.6 bps and 14.5 bps when equal- and value-weighting, respectively. Both costs are smaller than the report FF-4 alphas of 49 bps and 23 bps. Likewise in Japan, the long-only portfolio yields an alpha of 35 bps compared to the average cost of 7.5 bps when equal-weighting and an alpha of 42 bps compared to the average cost of 16.9 bps. Last, for Asia, the equal-weighted long-only strategy has an average cost of 12 bps relative to an alpha of 83 bps, and the value-weighted strategy has an average cost of 18 bps relative to an alpha of 31 bps.

Another way of looking at this is to compute the breakeven average trading cost by dividing alpha by the turnover. A portfolio manager able to trade below the breakeven trading cost should, on average, find the IMF to be profitable. For North America, IMF remains profitable so long as the trading costs are below 9% and 1.5% when equal-weighting and value-weighting, respectively. Across all regions, the breakeven trading cost ranges from 1.5% (North America) to 2.5% (Japan) when value-weighting and from 4.7% (Japan) to 9% (North America) when equal-weighting.

Table 4 Alphas and turnover by quintile.

	Panel A: North America						Panel B: Europe					
	EW FF-4		VW FF-4		Turnover (%)		EW FF-4		VW FF-4		Turnover (%)	
Quintile	Alpha	*t*-Stat	Alpha	*t*-Stat	EW	VW	Alpha	*t*-Stat	Alpha	*t*-Stat	EW	VW
1 (PV)	0.81	6.63	0.27	2.35	9.01	17.83	0.49	5.68	0.23	2.24	8.67	14.54
2	0.54	5.91	0.25	3.37	15.75	15.88	0.26	4.23	0.23	3.27	14.30	15.91
3	0.28	3.73	0.03	0.35	15.84	17.78	0.16	2.69	0.14	1.77	14.61	16.65
4	−0.07	−0.49	−0.14	−1.27	12.23	16.28	0.07	1.30	−0.06	−0.60	12.07	14.72
5 (UG)	−0.14	−0.34	−0.91	−3.94	8.23	18.69	−0.26	−2.52	−0.22	−1.47	8.56	16.21
PV–UG	0.95	2.39	1.19	3.91			0.75	6.60	0.45	2.08		
	Panel C: Japan						Panel D: Asia					
1 (PV)	0.35	4.60	0.42	3.74	7.52	16.85	0.83	6.87	0.31	2.07	11.94	17.99
2	0.28	5.24	0.15	1.74	14.57	18.65	0.41	4.05	0.24	1.76	18.68	20.22
3	0.11	2.12	0.14	1.17	16.79	19.90	0.21	2.86	−0.07	−0.57	17.23	18.73
4	0.08	1.58	0.05	0.70	14.24	15.63	0.04	0.41	−0.23	−1.94	14.00	19.44
5 (UG)	−0.01	−0.15	−0.10	−1.21	7.28	10.22	0.57	2.29	−0.51	−2.08	10.06	23.42
PV–UG	0.36	3.15	0.52	3.56			0.25	0.99	0.82	2.74		

This table lists alphas for equal- and value-weighted quintile portfolios constructed using the improved magic formula methodology. In each month t and for each region (North America, Europe, Japan, and Asia), stocks are ranked by gross profits-to-tangible capital employed and by gross profits-to-enterprise value. The two rankings are then summed to form a combined rank. Stocks are sorted into quintiles according to the combined ranking using the 20, 40, 60, and 80 percentile breakpoints. Alphas are the risk-adjusted returns from regressing the quintile portfolio returns on the regional market portfolio, SMB, HML, and momentum (FFC). t-Statistics are Newey–West. Turnover is the roundtrip percentage of the portfolio that is bought and sold each month averaged across all months from 1991 through 2016.

To help put this into context, Frazzini *et al.* (2012) study trading costs in the developed markets using proprietary trade data from a large money manager. They find that size, value, and momentum survive trading costs, but short-term reversals do not. Trading costs are substantially decreased when costs are used during the portfolio construction step. For example, the cost to implement the HML strategy drops from 2.28% per year to only 57 bps per year after accounting for trading costs. As the IMF strategy has some similarities with HML, their result provides additional confidence in the viability of IMF across developed markets. Similarly, studying the U.S. markets, Novy-Marx and Velikov (2015) find that low turnover strategies survive transaction costs. The evidence from the literature in addition to our turnover results provides confidence in the profitability of the IMF.

3.4 Sensitivity to risk factors

Table 5 lists the estimated factor loadings from regressing the value-weighted PV–UG portfolios on the FF-4 factors. The alphas presented in Table 5 are the same as those listed in Table 3. The estimate coefficients inform us on the general characteristics of the PV–UG portfolios. Loadings on HML are positive for all regions

Table 5 The improved magic formula—Factor loadings.

	Panel A: PV–UG				
	Alpha	β_{MKT}	β_{SMB}	β_{HML}	β_{WML}
North America	1.19	−0.40	−0.27	0.13	−0.03
	(3.91)	(−4.33)	(−2.11)	(0.81)	(−0.37)
Europe	0.45	−0.19	−0.11	0.36	0.14
	(2.08)	(−4.13)	(−1.34)	(3.01)	(2.09)
Japan	0.52	−0.07	0.28	0.19	0.13
	(3.56)	(−1.97)	(5.38)	(3.21)	(3.23)
Asia	0.82	−0.53	−0.85	0.00	−0.05
	(2.74)	(−9.21)	(−7.98)	(0.02)	(−0.64)

Stocks are sorted into quintile portfolios according to the IMF methodology as in Table 3. In Panel A, the hedge portfolio (PV–UG) that longs the portfolio of PV (profitable, value) stocks and shorts the portfolio of UG (unprofitable, growth) stocks is regressed against the Fama–French–Carhart four factors. In Panel B, the portfolio comprising 50% Mkt and 50% PV–UG is regressed on the Fama–French–Carhart factors. In both panels, the estimated coefficients are reported. Alpha is the intercept term. MKT is the loading on the value-weighted market portfolio less the risk-free rate. SMB is loading on the small minus big market capitalization portfolio. HML is estimated coefficient on the high minus low market-to-book equity portfolio, and WML is coefficient on the winner minus loser momentum portfolio. All factors are constructed at the regional level. Data is from January 1991 to December 2016. Values in parentheses are Newey–West t-statistics.

but the loadings on SMB tend to be negative with the exception of Japan. The relation with momentum is ambiguous with some coefficients being positive while others negative. This suggests that the PV–UG portfolios constructed using IMF have properties similar to large-cap, value portfolios.

The most significant results from Table 5 are the negative coefficients on the regional value-weighted market portfolios. The PV–UG portfolios have negative market beta. The market beta for North America is −0.40 which is highly significant with a t-statistic of −4.33, and for Europe, Japan, and Asia, the market betas are −0.19 (t-stat = −4.13), −0.07 (t-stat = −1.97), and −0.53 (t-stat = −9.21). Profitable, value stocks have smaller betas but yield larger returns relative to unprofitable, growth stocks. This contradicts fundamental asset pricing theory which associates high expected returns with high risk.

This result is related to the literature on the low-beta, low-volatility anomaly. The PV portfolios yield large average returns and have smaller betas than the UG portfolios which yield smaller average returns but have larger betas. Baker *et al.* (2011) find evidence of the low-volatility, low-beta anomaly over the long sample period extending from 1968 through 2008, and they find that the result is robust to risk measure and stock size. Frazzini and Pedersen (2014) show that the low-beta anomaly not only exists in the U.S. stock market, but also in international equity markets, equity indices, commodities, country bonds, and foreign exchange. They theorize that the anomaly is due to leverage aversion. Whereas classic asset pricing theory asserts that investors with low aversion to risk will leverage the portfolio with maximum Sharpe ratio to achieve higher expected returns; the leverage aversion theory argues that investors will avoid leverage and will instead tilt their portfolios toward high beta stocks and away from low-beta stocks to achieve high expected returns. The shift in demand leads to an increase in expected returns for the less desirable low-beta stocks and a decrease in expected returns for high beta stocks. Our results add a new dimension to this theory. Investors who overweight high beta stocks may also be overweighting expensive, unprofitable stocks at the expense of holding cheap, profitable stocks.

3.5 Dissecting the magic formula

As both profitability and value have been shown to individually explain the cross-section of returns, it seems pertinent to ask whether it is profitability or value that is driving the performance of IMF. To do this, we look at both components separately. All stocks within each region are sorted into quintile portfolios according to the profitability and value measures, GP-to-tangible capital employed and GP-to-enterprise value. For profitability, we long the value-weighted quintile portfolio of profitable stocks (P) and short the quintile portfolio of unprofitable stocks (U) to form the P–U portfolio. Similarly, for value we long the quintile portfolio of stocks that are relatively undervalued (V) and short the quintile of overvalued stocks (G) which we denote as the V–G portfolio. The portfolios P–U and V–G are then regressed on the four factors of the FF-4 model. Factor sensitivities are provided in Table 6 with profitability results in Panel (A) and value results in Panel (B).

The P–U portfolio, Panel A, yields economically and statistically significant alphas for all regions. Depending on the region, alphas range from 0.55% (Europe) to 1.16% (North America). The large alphas are consistent with the profitability results found by Novy-Marx (2013) for the U.S. market. Market betas as well as the factor loadings on SMB and HML all tend to be negative with large t-statistics. This implies that the P–U portfolio behaves similar to a large-cap, growth portfolio while being counter-cyclical to the market.

From Table 6, Panel (B), value yields smaller and less significant alphas relative to profitability. Alphas are larger and significant for North America and Asia, but are much smaller and have lower t-statistics for Europe and Japan. Market betas again tend to be negative and significant. What is particularly interesting are the loadings on SMB and HML. Whereas the SMB and HML factor sensitivities tend to be negative for P–U, they are mostly positive for V–G. Hence, when the profitability and value measures are combined to construct the PV–UG portfolio, the sensitivities to HML and SMB offset each other making the PV–UG less sensitive to both factors. At the same time, since market betas are negative for both P–U and V–G, combining the two measures strengthens the negative sensitivity of PV–UG to the market as seen by the larger t-statistics on the PV–UG market beta in Table 5 (except for Japan).

Take North America as an example. The market beta for profitability and value are -0.31 (t-stat $= -3.50$) and -0.30 (t-stat $= -3.34$). The market beta for the PV–UG from Table 5 is -0.40 with a larger t-statistic of -4.33. Profitability and value are both highly sensitive to HML but with opposite signs, -0.46 (t-stat $= -4.04$) for profitability and a sensitivity of 0.68 (t-stat $= 4.52$) for value. The diversification effect leads to a PV–UG portfolio that is less sensitive to HML. From Table 5, the factor sensitivity of PV–UG to HML is 0.13 (t-stat $= 0.81$), which is approximately the average of the sensitivities of the two individual

Table 6 Factor loadings for value and profitability portfolios.

	Alpha	β_{MKT}	β_{SMB}	β_{HML}	β_{WML}
Panel A: Gross profits-to-tangible capital employed (P–U)					
North America	1.16	−0.31	−0.73	−0.46	−0.04
	(4.04)	(−3.50)	(−7.33)	(−4.04)	(−0.66)
Europe	0.55	−0.07	−0.43	−0.46	0.07
	(2.78)	(−2.13)	(−6.04)	(−5.04)	(1.05)
Japan	0.60	0.01	0.04	−0.56	0.12
	(3.29)	(0.24)	(0.62)	(−4.91)	(1.73)
Asia	0.78	−0.44	−1.09	−0.22	−0.12
	(2.50)	(−7.89)	(−8.79)	(−1.50)	(−1.29)
Panel B: Gross profits-to-enterprise value (V–G)					
North America	0.94	−0.30	0.16	0.68	0.13
	(3.46)	(−3.34)	(1.08)	(4.52)	(1.92)
Europe	0.36	−0.06	0.09	0.62	0.20
	(1.91)	(−1.32)	(0.78)	(5.81)	(2.80)
Japan	0.30	−0.11	0.49	0.57	0.07
	(1.88)	(−3.38)	(6.14)	(7.12)	(1.70)
Asia	1.02	−0.49	−0.53	0.32	−0.06
	(4.01)	(−8.37)	(−5.28)	(3.04)	(−0.80)

Stocks are sorted into quintile portfolios according to the gross profits-to-tangible capital employed (Panel A) and also by gross profits-to-enterprise value (Panel B). For both sorts, the hedge portfolio that longs the value-weighted quintile five portfolio and shorts the value-weighted quintile one portfolio is regressed against the Fama–French–Carhart four factors. The estimated coefficients are reported. Alpha is the intercept term. MKT is the loading on the value-weighted market portfolio less the risk-free rate. SMB is loading on the small minus big market capitalization portfolio. HML is estimated coefficient on the high minus low market-to-book equity portfolio, and WML is coefficient on the winner minus loser momentum portfolio. All factors are constructed at the regional level. Data is from January 1991 to December 2016. Values in parentheses are Newey–West t-statistics.

characteristics. We see a similar result with SMB. The coefficient on SMB for P–U and V–G are −0.73 and 0.16, respectively. From Table 5, the loading on SMB for PV–UG is −0.27 which is a reduction in exposure relative to profitability and it is the opposite sign relative to value.

In addition to the changes in risk characteristics, combining the two characteristics also pays in terms of risk-adjusted returns. Comparing alphas in Tables 5 and 6, for all regions except Asia, PV–UG provides greater risk-adjusted returns than the value portfolio but provides similar alphas relative to profitability. Altogether, the PV–UG portfolio maintains the large alpha from sorting on profitability but profitability is highly sensitive to HML and SMB. Combining profitability with value, which also has positive alphas, changes the risk characteristics by decreasing the sensitivity to both SMB and HML. PV–UG behaves like a large-cap, value portfolio with less sensitivity to the size factor (SMB) relative to the profitability portfolio and less sensitive to the value factor (HML) relative to the value portfolio.

4 Double sorts

4.1 Double sort on IMF and size

A concern with the magic formula is whether it is actually tradeable. Fama and French (2012), for example, find significant returns to value and momentum internationally, but the results are primarily focused in the set of small stocks where trading costs and illiquidity likely make strategies based on the results unprofitable. Novy-Marx (2013) considers a strategy similar to MF using the sample of large and highly liquid Fortune 500 stocks and finds a statistically significant 0.37% alpha. We now ask whether the large risk-adjusted returns are driven by small, illiquid stocks or whether they are also obtainable by trading large-cap, liquid stocks.

All stocks within each region are independently double sorted into 25 value-weighted portfolios according to market capitalization and IMF. Portfolios are rebalanced each month and the monthly average excess returns are provided in Table 7. We additionally provide the 1–5 portfolio average return for IMF, the 5–1 average portfolio return for size. The row labeled as "Mean" is the column average. All t-statistics are Newey–West.

The 1–5 IMF return differentials are significant for nearly all size groups across all regions. Most importantly, the average spreads are significant for the set of big stocks. For the 5th quintile size group, big stocks, IMF spreads are 0.56%, 0.51%, 0.43%, and 0.18% for North America, Europe, Japan, and Asia, respectively, which are significant for all regions except Asia. Fama and French (2012) find that value and momentum returns decrease with size. We find the similar result with IMF. With return differentials for the smallest quintile for North America, Europe, Japan, and Asia of 0.88%, 0.81%, 0.13%, and 0.73%, respectively, spreads are larger within the set of small stocks for all regions except Japan.

Risk-adjusted alphas are also found to be significant for big stocks as well as for small stocks. All stocks within each region are divided equally into big and small subsamples according to market capitalization. Quintile portfolios and the long–short portfolios PV–UG_B and PV–UG_S are constructed from the stocks within each set. Table 8 lists the FF-4 alphas. In nearly all cases, the alphas for both PV–UG_B and PV–UG_S are positive and significant. Similarly, the long-only alphas for the small and big PV portfolios are also found to be positive and significant in all but one case.

Finding significant alphas for the more liquid large-cap stocks provides some evidence that the IMF is profitable. To further investigate this, Table 8 also lists the percent turnover for each quintile portfolio. The turnover for the big stock portfolios is considerably smaller than for the small stock portfolios. Assuming a 1% roundtrip trading cost, the long–short trading cost of implementing IMF for big stocks in North America is 33 bps relative to an alpha of 72 bps. Alternatively, the big stock strategy in North America is profitable so long as trading costs are below 2.18% (breakeven trading cost). For Europe, costs are about 28 bps vs an alpha of 38 bps. Similar results are found for Japan and Asia. The breakeven trading cost for big stocks in Europe is 1.36%, suggesting that as long as costs remain below 1.36%, the IMF long–short strategy is profitable. Similar results are also found for Japan and Asia.[11]

Table 7 Double sort—The improved magic formula vs size.

	Panel A: North America							Panel B: Europe						
	Improved Magic formula							Improved magic formula						
Size	1 (PV)	2	3	4	5 (UG)	1–5	*t*-Stat	1 (PV)	2	3	4	5 (UG)	1–5	*t*-Stat
1 (Small)	1.53	1.38	1.23	1.07	0.65	0.88	(3.62)	0.94	0.67	0.62	0.42	0.13	0.81	(5.99)
2	1.12	1.14	1.05	0.99	0.53	0.58	(2.55)	0.97	0.85	0.70	0.61	0.33	0.65	(4.45)
3	1.02	1.06	1.01	0.80	0.62	0.40	(1.76)	0.79	0.74	0.64	0.61	0.51	0.28	(2.07)
4	0.97	1.04	0.96	0.83	0.36	0.60	(2.63)	0.84	0.65	0.70	0.62	0.55	0.29	(1.79)
5 (Big)	0.81	0.67	0.58	0.57	0.25	0.56	(2.33)	0.76	0.65	0.54	0.44	0.25	0.51	(2.61)
Mean	1.09	1.06	0.96	0.85	0.48	0.61	(2.94)	0.86	0.71	0.64	0.54	0.35	0.51	(4.35)
5–1	0.72	0.71	0.65	0.50	0.40			0.18	0.02	0.08	−0.02	−0.12		
	(2.75)	(2.66)	(2.45)	(2.10)	(1.51)			(0.88)	(0.11)	(0.31)	(−0.10)	(−0.57)		
	Panel C: Japan							Panel D: Asia						
1 (Small)	0.55	0.50	0.51	0.47	0.42	0.13	(1.41)	1.31	0.93	1.19	0.72	0.58	0.73	(2.93)
2	0.35	0.34	0.19	0.16	0.08	0.27	(2.47)	0.83	0.65	0.86	0.66	0.15	0.67	(2.61)
3	0.32	0.21	0.06	0.18	−0.13	0.45	(3.02)	1.03	0.75	0.55	0.47	0.25	0.78	(2.81)
4	0.36	0.25	0.13	0.07	0.02	0.34	(2.40)	1.02	0.84	0.51	0.50	0.47	0.55	(2.05)
5 (Big)	0.47	0.04	0.13	0.10	0.04	0.43	(2.30)	0.74	0.80	0.64	0.64	0.56	0.18	(0.58)
Mean	0.41	0.27	0.21	0.20	0.09	0.32	(3.53)	0.99	0.79	0.75	0.60	0.40	0.58	(3.17)
5–1	0.08	0.47	0.38	0.37	0.38			0.57	0.13	0.55	0.08	0.02		
	(0.26)	(1.36)	(1.32)	(1.14)	(1.19)			(1.88)	(0.41)	(1.48)	(0.26)	(0.05)		

This table reports the value-weighted returns of portfolios formed by double sorting by the improved magic formula (IMF) and size. Stocks are first independently ranked by gross income to tangible capital employed and by gross income to enterprise value. The sum of the two rankings is IMF. All stocks are then independently double sorted by size and IMF. For both measures, we use the 20, 40, 60, and 80 percentile breakpoints. For IMF, quintile 1 is comprised of profitable, value (PV) stocks, while quintile 5 is comprised of unprofitable, growth (UG) stocks. For size, quintile 1 is the portfolio of small stocks and quintile 5 is the portfolio of big stocks. The 5–1 size and the 1–5 IMF return differentials are provided. Newey–West *t*-statistics are provided in parentheses. The row labeled Mean lists the returns obtained from averaging the five quintile portfolios in each column.

Table 8 Alphas and turnover for big and small cap value-weighted portfolios.

	Panel A: North America						Panel B: Europe					
	Small		Big		Turnover (%)		Small		Big		Turnover (%)	
Quintile	Alpha	*t*-Stat	Alpha	*t*-Stat	Small	Big	Alpha	*t*-Stat	Alpha	*t*-Stat	Small	Big
1 (PV)	0.67	4.29	0.22	2.05	27.13	17.54	0.48	4.46	0.26	2.82	23.89	13.93
2	0.51	4.17	0.22	3.34	33.11	17.41	0.19	2.19	0.21	2.84	28.65	17.87
3	0.05	0.33	0.10	1.21	34.76	20.96	0.05	0.53	0.22	2.59	29.56	18.24
4	−0.39	−1.35	−0.12	−1.20	34.62	18.23	−0.11	−1.30	−0.12	−1.04	26.94	16.25
5 (UG)	−0.69	−1.98	−0.49	−3.30	31.28	15.38	−0.60	−4.18	−0.13	−0.96	26.57	13.80
PV–UG	1.36	3.72	0.72	3.15			1.08	7.46	0.38	1.99		
	Panel C: Japan						Panel D: Asia					
1 (PV)	0.21	2.37	0.45	3.49	20.32	16.12	0.73	4.67	0.27	1.61	30.36	20.98
2	0.23	3.44	0.03	0.30	27.14	19.92	0.50	3.19	0.29	2.03	37.51	25.79
3	0.13	1.74	0.25	2.00	30.82	22.06	0.20	1.38	0.00	0.00	38.44	24.98
4	0.09	1.17	0.01	0.16	29.87	16.72	−0.17	−0.72	−0.18	−1.51	37.12	21.43
5 (UG)	−0.03	−0.28	−0.12	−1.38	22.64	10.47	0.04	0.15	−0.31	−1.71	34.36	21.84
PV–UG	0.24	1.68	0.57	3.29			0.69	2.33	0.58	2.05		

This table lists alphas for value-weighted quintile portfolios constructed using the improved magic formula methodology for large- and small-cap portfolios. Stocks are equally divided into small- and large-cap groups. Then, in each month t and for each region (North America, Europe, Japan, and Asia), stocks are ranked by gross profits-to-tangible capital employed and by gross profits-to-enterprise value. The two rankings are then summed to form a combined rank. Stocks are sorted into quintiles according to the combined ranking using the 20, 40, 60, and 80 percentile breakpoints. Alphas are the risk-adjusted returns from regressing the quintile portfolio returns on the regional market portfolio, SMB, HML, and momentum (FFC). t-Statistics are Newey–West. Turnover is the roundtrip percentage of the portfolio that is bought and sold each month averaged across all months from 1991 through 2016.

4.2 Double sorts on IMF and book-to-market equity

Value is one of the two fundamental inputs into IMF, it is possible that conditioning on the widely used book-to-market equity measure of value may subsume the IMF performance results. To check for this possibility, stocks are independently double sorted into 25 portfolios by book-to-market equity and IMF. Table 9 reports the average monthly returns for the 25 portfolios.

Starting in Panel A of Table 9, the 5–1 book-to-market spreads in North America tend to be small and insignificant while IMF leads to larger spreads with larger t-statistics. Even when controlling for variation in B/M, IMF return differentials tend to be economically large, ranging from 0.31% to 0.82%, though not always significant. Results are similar for Europe and Asia. For Europe, all IMF return differentials are economically large, ranging from 0.37% to 0.87% and are all significant—at the 10% level for the 5th B/M quintile and at the 5% level or higher for all other B/M quintiles. Meanwhile, the B/M spreads are significant for three of the five IMF quintiles—marginally significant for IMF quintile 4 and

Table 9 Double sort—Improved magic formula vs book-to-market.

	Panel A: North America							Panel B: Europe						
	Improved magic formula							Improved magic formula						
B/M	1 (PV)	2	3	4	5 (UG)	1–5	t-stat	1 (PV)	2	3	4	5 (UG)	1–5	t-Stat
1 (Low)	0.72	0.72	0.53	0.62	0.34	0.39	(0.94)	0.55	0.56	0.57	0.14	−0.32	0.87	(3.09)
2	0.91	0.72	0.61	0.47	0.35	0.56	(1.92)	0.81	0.53	0.50	0.16	0.12	0.69	(3.25)
3	1.01	0.75	0.86	0.87	0.19	0.82	(2.90)	0.88	0.67	0.86	0.63	0.42	0.46	(2.62)
4	0.72	0.98	0.83	0.65	0.41	0.31	(1.39)	0.91	0.81	0.76	0.57	0.45	0.46	(2.35)
5 (High)	1.38	0.82	0.74	0.88	0.62	0.76	(3.21)	1.10	0.91	0.78	0.81	0.73	0.37	(1.81)
Mean	0.95	0.80	0.72	0.70	0.38	0.57	(2.62)	0.85	0.69	0.70	0.46	0.28	0.57	(4.05)
5–1	0.66	0.11	0.21	0.26	0.28			0.55	0.35	0.21	0.67	1.05		
	(2.79)	(0.40)	(0.80)	(0.74)	(0.72)			(2.21)	(1.09)	(0.67)	(1.85)	(3.40)		
	Panel C: Japan							Panel D: Asia						
1 (Low)	0.43	−0.18	−0.22	−0.30	−0.29	0.72	(2.29)	0.60	0.66	0.60	0.55	−0.05	0.65	(1.14)
2	0.25	−0.07	0.14	−0.30	−0.13	0.38	(1.76)	1.11	0.59	0.64	0.19	−0.08	1.19	(3.16)
3	0.35	0.13	0.22	0.20	0.33	0.02	(0.11)	0.85	0.95	0.19	0.33	0.32	0.53	(1.72)
4	0.61	0.44	0.38	0.49	0.15	0.46	(2.27)	0.81	0.93	0.46	0.68	0.30	0.51	(1.90)
5 (High)	0.71	0.70	0.61	0.53	0.35	0.36	(2.27)	1.36	1.27	1.09	1.07	1.06	0.30	(1.18)
Mean	0.47	0.20	0.23	0.13	0.08	0.39	(2.89)	0.94	0.88	0.59	0.56	0.31	0.64	(2.99)
5–1	0.29	0.88	0.83	0.83	0.65			0.76	0.61	0.49	0.52	1.11		
	(0.83)	(2.52)	(2.49)	(2.51)	(2.02)			(2.52)	(1.75)	(1.16)	(1.69)	(1.99)		

This table reports the value-weighted returns of portfolios formed by double sorting by the improved magic formula (IMF) and book-to-market equity (B/M). Stocks are first independently ranked by gross income to tangible capital employed and by gross income to enterprise value. The sum of the two rankings is IMF. All stocks are then independently double sorted by B/M and IMF. For both measures, we use the 20, 40, 60, and 80 percentile breakpoints. For IMF, quintile 1 is comprised of profitable, value (PV) stocks, while quintile 5 is comprised of unprofitable, growth (UG) stocks. For B/M, quintile 1 is the portfolio of growth stocks and quintile 5 is the portfolio of value stocks. The 5–1 B/M and the 1-5 IMF hedge portfolios are provided. Newey–West t-statistics are provided in parentheses. The row labeled Mean lists the returns obtained from averaging the five quintile portfolios in each column.

Table 10 Fama–MacBeth regression.

	Panel A: All stocks				Panel B: Small stocks				Panel C: Big stocks			
	NA	EUR	Japan	Asia	NA	EUR	Japan	Asia	NA	EUR	Japan	Asia
log(IMF)	−0.34	−0.34	−0.13	−0.29	−0.41	−0.32	−0.08	−0.37	−0.29	−0.33	−0.22	−0.44
	(−2.94)	(−7.16)	(−2.13)	(−2.77)	(−3.34)	(−6.00)	(−1.29)	(−2.86)	(−2.49)	(−6.16)	(−2.90)	(−4.45)
log(Size)	−0.22	0.00	−0.09	−0.29	−0.64	−0.32	−0.58	−1.30	−0.06	0.07	0.06	0.12
	(−4.02)	(0.07)	(−1.72)	(−5.26)	(−7.65)	(−7.58)	(−8.28)	(−10.88)	(−1.30)	(2.17)	(1.08)	(2.32)
log(B/M)	0.37	0.56	0.37	0.68	0.40	0.66	0.35	0.91	0.28	0.41	0.49	0.56
	(4.02)	(7.45)	(4.46)	(6.37)	(4.22)	(9.57)	(4.21)	(7.04)	(2.91)	(4.50)	(4.94)	(5.90)
MOM	0.15	0.75	−0.37	0.40	0.16	0.81	−0.77	0.67	0.32	0.88	−0.01	0.48
	(0.81)	(3.13)	(−1.27)	(1.92)	(0.92)	(3.32)	(−2.28)	(2.24)	(1.30)	(3.16)	(−0.01)	(2.56)
Constant	5.51	3.63	2.11	4.69	7.67	4.51	3.49	8.15	3.94	3.06	1.93	3.23
	(7.02)	(6.91)	(3.36)	(7.39)	(10.23)	(8.23)	(5.42)	(10.13)	(3.79)	(4.65)	(2.59)	(4.46)
Avg R^2	3.12	2.26	4.47	3.25	2.58	1.80	3.55	3.89	4.64	3.42	5.72	4.23

For each month t from January 1991 to December 2016, returns of all firms in the regions of North America (NA), Europe (EUR), Japan, and Asia are regressed on the log of the more magical, magic formula (IMF), log market capitalization (Size), log book-to-market equity (BM), and momentum (MOM). This table shows the cross-sectional regression coefficients averaged across all months. Avg R^2 is the time-series average R-square. Newey–West t-statistics are provided below each estimate. IMF is the sum of the rankings obtained by sorting stocks by gross profits-to-tangible capital employed and by gross profits-to-enterprise value using income statement data from time $t-7$. Size is measured using market prices from month $t-1$. B/M is computed using the book value from month $t-7$ and market prices from month $t-1$, and MOM is the cumulative return over $t-12$ to $t-2$. Results labeled as "Big Stocks" and "Small Stocks" refer to the 50% of stocks with the largest and smallest market capitalizations, respectively.

higher levels of significance for IMF quintiles 1 and 5. In Asia, IMF return differentials range from 0.30% to 1.19% with three spreads being significant at the 10% level or higher. B/M 5–1 returns are also large and often significant at the 10% level. Japan is a little different in that both IMF and B/M return differentials tend to be highly significant.

The results presented here are interesting. IMF sorts stocks according to both value and profitability. Though B/M and GP-to-enterprise value are both measures of value, our previously reported results in Table 3 do not show a strong or consistent relationship between B/M and IMF quintiles. That we find significant return differentials for both IMF and B/M is further evidence that IMF and B/M capture different dimensions of the cross-section of returns.

4.3 *Fama–MacBeth regressions*

To further show the differences between IMF and the FF-4 factors, we use the Fama and MacBeth (1973) multivariate regression framework to determine whether IMF explains the cross-section of returns in addition to size, B/M, and momentum. For each month t, we regress the cross-section of returns for each region on a set of four characteristics. These include the log market capitalization (size) measured at time $t-1$, the log book-to-market equity (B/M) from time $t-7$, and momentum (MOM) computed as the cumulative return over $t-12$ to $t-2$. We additionally add the log of IMF—the sum of the rankings on the profitability and value measures. Table 10 presents results for the set of all stocks. Results are also provided for the 50% of stocks with the smallest market capitalization (small stocks) and the 50% of stocks with the largest market capitalization (big stocks).

Table 10 shows that IMF is an important determinant of the cross-section returns. The results are consistent across all regions as well as for the sets of all stocks, small-cap stocks, and large-cap stocks. The negative coefficient on log IMF shows that PV stocks yield higher returns than UG stocks which is consistent with all previous results. The estimated coefficient on log IMF is significant for all cases except for small stocks in Japan.

Estimated coefficients on size, B/M, and momentum are generally as expected. With the exception of Japan, momentum tends to have a positive coefficient and is statistically significant for the set of all stocks, small stocks, and big stocks in Europe and Asia. Momentum is not significant in North America, regardless of the size subsamples. B/M is positive and highly significant in all cases showing that value stocks tend to outperform growth stocks. Size yields mixed results. The size coefficient is negative and highly significant within the set of small stocks; however, the coefficient tends to be positive, much closer to zero, and sometimes insignificant within the set of big stocks. Overall, IMF is significant with the expected sign while accounting for other characteristics widely known to explain the cross-section of returns. This provides additional confidents that IMF captures a unique dimension of returns not captured by the FF-4 factors.

5 Additional tests

In this section, we conduct a battery of tests to demonstrate the robustness of the IMF results.

5.1 *Year-by-year analysis*

We are interested to know if the reported IMF results are primarily driven by some specific time

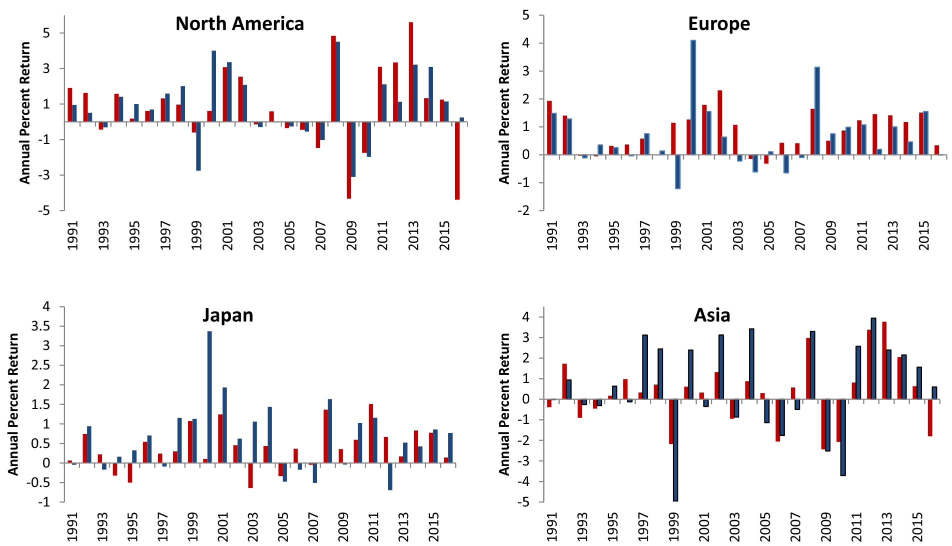

Figure 1 Year-by-year long–short return differentials.

This figure plots the average long–short return differentials for value-weighted portfolios (blue bars) and equal-weighted portfolios (red bars) constructed using the Improved magic formula methodology for each year 1991–2016. In each month *t* and for each region (North America, Europe, Japan, and Asia), stocks are ranked by gross profits-to-tangible capital employed and by gross profits-to-enterprise value. Stocks are sorted into quintiles according to the combined ranking using the 20, 40, 60, and 80 percentile breakpoints.

period, such as the 2008 financial crisis or the subsequent recovery period. Figure 1 plots the average long–short quintile portfolio return differentials for each year 1991–2016 where the red bars represent equal-weighting and the blue bars represent value-weighting. For North America, the IMF yields positive return differentials 17 out of 26 years for both equal- and value-weighted portfolios. IMF fairs even better in Europe and Japan with 21 (18) out of 26 years of positive return differentials when equal (value)-weighting, and Asia yields positive return differentials for 17 (14) years when using equal-weights (value-weights). Annual performance appears to be correlated across regions.[12] IMF yields negative returns for all regions during mid-2000s, roughly 2003 through 2007; however, returns tend to be quite large around 2000–2002.

It is interesting to observe return differentials around crisis periods. It is during such periods when the differences between high- and low-quality firms are likely to be most apparent. IMF may perform particularly well in such periods. For North America, Europe, and Japan, we indeed find large positive return differentials in the years immediately following the bursting of the technology bubble, 2000–2002. Return differentials in 2008, the year of the global financial crisis, are also consistently large across all regions. Differentials in 2009, however, are large and negative for North America and Asia and are only modestly positive for Europe and Japan. While there is strong performance when crisis strikes, it is clear that crisis periods do not explain the strong IMF performance.

5.2 *Sentiment and mispricing*

An important question is whether the reported large alphas are a result of an unidentified risk factor or whether the strategy has identified a mispricing related to a behavioral bias. We perform two tests to differentiate between the two possibilities. The evidence presented here suggests that IMF returns are related to mispricing.

Baker *et al.* (2006) construct a composite sentiment index from a collection of sentiment indexes proposed for the U.S. stock market.[13] They find that stocks which are difficult to valuate are more susceptible to sentiment than easy to value stocks. For example, small, unprofitable, extreme-growth, and distressed stocks are more sensitive to sentiment than big, profitable, low-volatility stocks. Since the sentiment index is derived specifically for the U.S. market, we focus our attention only on North America. Figure 2 plots average returns during for periods of high and low sentiment for the five IMF quintiles. We find that the unprofitable, growth (UG) portfolio is sensitive to sentiment, while the profitable value (PV) portfolio is not. As a result, larger average PV–UG returns may be obtained during high sentiment months than low sentiment months due to the sensitivity of the UG portfolio to sentiment.

Our second test is based on the Stambaugh and Yuan (2016) four-factor model which includes two mispricing factors, MGMT and PERF, constructed from a set of U.S. equity return anomalies. The two factors roughly represent mispricing related to management decisions and performance, respectively.[14] We regress IMF returns on the four factors to identify whether IMF is related to one or both of the mispricing factors. The analysis focuses on the North American region only as the factors are specific to the U.S. and U.S. equity anomalies. Regression results, provided in Table 11 for the PV, UG, and PV–UG portfolios show a significant relation between IMF and both mispricing factors. Consistent with the sentiment results, coefficients on the mispricing factors are larger for UG than for PV, suggesting that greater mispricing on the short side than the long side.

5.3 Sector and country analysis

Institutional investors may be concerned that IMF performance is isolated to particular countries or sectors, or that IMF methodology yields portfolios that are highly concentrated in particular

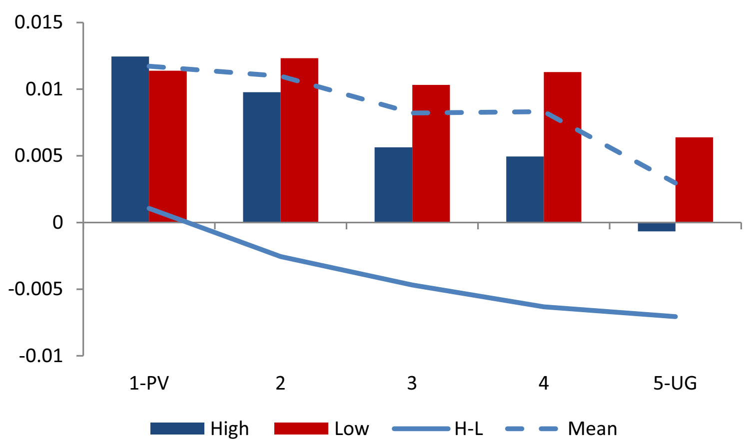

Figure 2 High and low sentiment returns by quintile—North America.

Using the Baker and Wurgler (2006) sentiment index, we compute the average returns of the IMF quintile returns over high and low sentiment periods. The dark blue and red bars are the average returns over periods when sentiment is positive and negative, respectively. The solid blue line is the difference between high and low sentiment average returns and the dashed line is the average return quintile return over the sample period. The sample period extends from January 1991 through September 2015 as this is when the sentiment index stops.

Table 11 Stambaugh and Yuan model—North America.

	Equal-weighted			Value-weighted		
	PV_{EW}	UG_{EW}	$PV–UG_{EW}$	PV_{VW}	UG_{VW}	$PV–UG_{VW}$
Constant	0.98	0.77	0.21	0.03	−0.35	0.38
	(6.09)	(1.66)	(0.51)	(0.27)	(−1.45)	(1.46)
Mkt-Rf	0.86	0.85	0.01	0.94	1.01	−0.07
	(21.42)	(7.25)	(0.13)	(33.42)	(12.25)	(−0.76)
SMB	0.78	0.46	0.32	0.17	0.12	0.05
	(11.80)	(4.38)	(3.28)	(3.89)	(1.45)	(0.58)
Mgmt	−0.02	−0.76	0.74	0.26	−0.42	0.68
	(−0.31)	(−3.64)	(4.05)	(3.28)	(−3.26)	(5.30)
Perf	−0.17	−0.26	0.09	0.12	−0.14	0.26
	(−3.86)	(−2.86)	(0.97)	(3.07)	(−1.80)	(3.01)

The North America IMF portfolios PV, UG, and PV–UG are regressed on the Stambaugh and Yuan (2016) four factors. Mkt-Rf is the value-weighted portfolio of U.S. stocks. SMB is the Stambaugh and Yuan small minus big factor. Mgmt and Perf are factors derived from anomalies observed in the U.S. stock market and are intended to capture mispricing related to management decisions and performance, respectively. Newey–West *t*-statistics are provided in parentheses.

Table 12 The improved magic formula by sector.

	Oil/gas	Basic materials	Industrials	Consumer goods	Healthcare	Consumer services	Financials	Technology
Panel A: Four-factor sector alphas for PV–UG value-weighted portfolio								
North America	0.95	0.59	0.30	0.53	0.87	0.18	0.49	0.29
	(3.09)	(1.13)	(1.84)	(2.08)	(2.16)	(0.75)	(0.86)	(0.72)
Europe	0.72	1.34	0.26	0.68	0.79	0.19	1.19	−0.14
	(1.48)	(3.01)	(1.44)	(3.19)	(2.10)	(0.82)	(3.61)	(−0.36)
Japan	N/A	0.61	0.32	0.01	0.02	0.62	0.37	0.14
		(2.68)	(1.77)	(0.02)	(0.05)	(2.40)	(0.67)	(0.43)
Asia	0.04	0.25	0.96	1.03	1.08	1.10	0.25	N/A
	(0.05)	(0.64)	(2.53)	(2.93)	(1.39)	(3.54)	(0.49)	
Panel B: Relative portfolio weights (%)								
North America	−8.04	−11.07	0.80	5.12	−2.56	11.66	−1.04	5.12
Europe	−2.75	−5.09	0.91	1.26	−2.39	5.38	−1.42	4.10
Japan	−0.37	−8.02	−18.13	6.58	0.16	22.03	−1.19	−1.05
Asia	−3.59	−9.77	1.51	7.22	−0.69	5.07	−2.84	3.08

Panel A shows the Fama–French four-factor model value-weighted alphas for PV–UG sector-specific improved magic formula (IMF) Portfolios for the regions of North America, Europe, Japan, and Asia. N/A indicates that the sector has too few observations for reliable statistics. Panel B reports the degree to which each sector is under- or over-weighted in the PV portfolio for each region as measured by the difference in the average portfolio weight of each sector in the regional value-weighted PV portfolio and the value-weight of the sector in the region.

countries or sectors. We address these issues in this section.

In Table 12 we apply the IMF methodology to stocks in eight different sectors. Region-specific Fama–French four-factor alphas are reported for the value-weighted PV–UG portfolio.[15] Panel A of Table 12 shows that the high reported IMF alphas are not driven by a particular sector. For all regions except Japan, there are at least three sectors yielding significant alphas. IMF does perform particularly well in the consumer services sector. Sector-neutral alphas are computed by taking the arithmetic average of the alphas across all sectors and tend to be large with values of 53, 63, 30, and 67 bps for North America, Europe, Japan, and Asia, respectively.

Panel B shows the degree to which each sector is on average over- or under-weighted in the value-weighted PV IMF portfolio. The relative portfolio weights are computed as each sector's value-weight in the PV IMF portfolio less the sector's value-weight in the region. Figure 3 plots the time-series of relative sector value-weights for each region. Not surprising, consumer services is the most over-weighted sector in North America, Europe, and Japan over-weighting the sector by 11.66%, 5.38%, and 22.03%, respectively, while Consumer Goods is the most over-weighted sector in Asia. The most under-weighted sectors are Basic Materials for North America (−11.07%), Europe (−5.09%), and Asia (−9.77%), and Industrials for Japan (−18.13%).

Table 13 lists the FF-4 alphas for each country included in the sample.[16] For North America, both Canada and the U.S. yield economically large PV–UG alphas when equal- and value-weighting. The alphas are statistically significant in both Canada and the U.S. when value-weighting and in the U.S. when equal-weighting. Taking the arithmetic average of the alphas across the two North American countries yields averages of 76 bps and 94 bps when country-specific alphas are equal- and value-weighted, respectively. While economically large, they are smaller than the regional alphas presented in Table 3 of 95 bps and 119 bps when equal- and value-weighting, respectively.

PV–UG alphas for the 15 European countries also tend to be economically large with 11 countries having equally-weighted alphas greater than 50 bps. Alphas are weaker when value-weighting. t-Statistics, however, are small relative to the regional results presented in Table 3 due to the small sample of stocks available in each individual country. The arithmetic average alphas across all European countries are 60 bps and 49 bps when using the equal- and value-weighted alphas, respectively. This is compared to the regional alphas of 75 bps and 45 bps as listed in Table 3.

Last, results for the four countries in the Asia region show greater dispersion. The alphas for Australia and New Zealand are statistically insignificant, while alphas for Hong Kong and Singapore are economically large and statistically significant. The arithmetic average alphas are 27 bps and 64 bps when using the equal- and value-weighted alphas, respectively, which are similar to the alphas reported in Table 3 of 12 bps and 66 bps.

The last column in Table 13 reports the average difference in each country's value-weight in the IMF regional value-weighted PV portfolio and the country's value-weight in the region. Figure 4 plots the time-series of relative value-weights for each country. For North America, the IMF PV portfolio over-weights the U.S. by 11.98% and under-weights Canada by 11.98% in order to capture the larger value-weighted PV alpha in the U.S. The European PV value-weighted portfolio over-weights the U.K. by 14.61% and under-weights France by 6.34%. All other European countries are within 3% of their value-weight in the region. Last, the Asian PV value-weighted

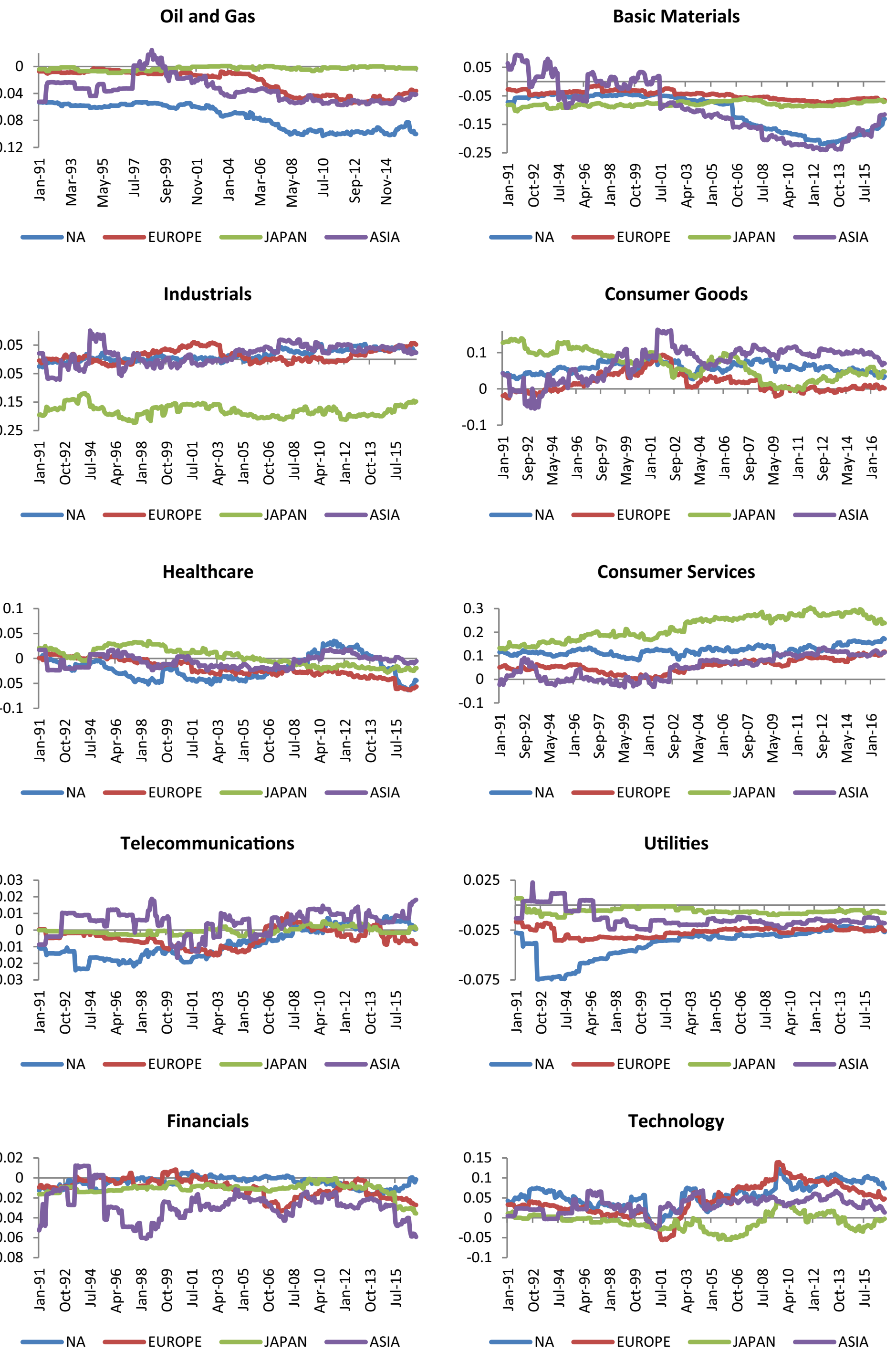

Figure 3 Relative portfolio weights by sector.

The time-series of relative portfolio weights for the IMF value-weighted PV (long-only) portfolio are plotted for each sector in the regions of North America (NA), Europe, Japan, and Asia. The relative portfolio weight is the value-weight of each sector in the region's IMF PV portfolio minus the value-weight of the sector in the region.

Table 13 The improved magic formula by country.

	Equal-weighted alpha			Value-weighted alpha			Relative weights (%)
	PV	UG	PV–UG	PV	UG	PV–UG	PV–UG
Canada	0.42	−0.21	0.63	−0.01	−1.12	1.11	−11.98
	(1.77)	(−0.30)	(1.41)	(−0.00)	(−2.10)	(2.43)	
U.S.	0.84	−0.06	0.90	0.29	−0.48	0.76	11.98
	(6.40)	(−0.40)	(5.20)	(2.35)	(−3.50)	(3.45)	
Austria	0.00	−0.57	0.58	−0.31	−0.77	0.46	−0.90
	(0.01)	(−2.10)	(1.99)	(−1.10)	(−2.50)	(1.37)	
Belgium	0.17	−0.39	0.56	0.10	−0.17	0.27	−1.50
	(0.87)	(−1.70)	(2.53)	(0.42)	(−0.60)	(0.77)	
Denmark	0.39	−0.36	0.75	0.45	−0.37	0.82	−1.01
	(1.46)	(−1.2)	(3.18)	(1.68)	(−0.80)	(1.99)	
Finland	0.25	−0.01	0.26	−0.20	−0.23	0.03	−0.54
	(0.96)	(−0.00)	(1.01)	(−0.50)	(−0.80)	(0.07)	
France	0.48	0.06	0.43	0.22	−0.02	0.23	−6.34
	(3.89)	(0.31)	(2.70)	(1.71)	(−0.00)	(1.04)	
Germany	0.36	−0.17	0.54	0.22	−0.02	0.24	1.42
	(2.09)	(−0.70)	(2.69)	(1.26)	(−0.00)	(0.70)	
Italy	−0.09	−0.40	0.31	0.09	−0.14	0.23	−0.21
	(−0.3)	(−1.20)	(1.44)	(0.30)	(−0.40)	(0.79)	
Ireland	0.70	−0.40	1.11	1.15	−0.49	1.63	1.61
	(2.27)	(−0.70)	(1.65)	(2.12)	(−0.80)	(1.99)	
Netherlands	0.38	−0.14	0.52	0.11	0.17	-0.05	1.33
	(2.10)	(−0.70)	(2.70)	(0.45)	(0.45)	(-0.10)	
Norway	0.28	−0.27	0.56	−0.34	−0.33	-0.01	−1.48
	(1.09)	(−0.70)	(1.59)	(−1.10)	(−0.70)	(-0.00)	
Portugal	0.30	−0.47	0.77	0.40	−0.48	0.88	−0.99
	(0.87)	(−1.10)	(1.74)	(1.17)	(−1.10)	(1.88)	
Spain	0.14	−0.29	0.43	−0.13	−0.44	0.31	−1.80
	(0.46)	(−1.00)	(1.78)	(−0.50)	(−1.30)	(0.83)	
Sweden	0.85	−0.08	0.93	0.79	−0.48	1.27	−2.96
	(3.77)	(−0.20)	(3.01)	(2.93)	(−1.20)	(3.16)	
Switzerland	0.58	−0.03	0.61	0.47	−0.03	0.50	−1.26
	(2.78)	(−0.10)	(2.52)	(2.27)	(−0.10)	(1.78)	
UK	0.67	−0.50	1.17	0.29	−0.15	0.44	14.61
	(3.56)	(−1.80)	(5.39)	(1.63)	(−0.50)	(1.21)	
Australia	1.06	1.18	−0.13	0.36	0.06	0.31	−7.00
	(4.39)	(3.38)	(−0.4)	(1.79)	(0.17)	(0.99)	
New Zealand	0.46	0.89	−0.44	0.68	0.48	0.21	0.74
	(1.42)	(2.46)	(−1.20)	(1.95)	(1.09)	(0.49)	
Hong Kong	0.56	−0.28	0.83	0.31	−0.75	1.06	5.01
	(2.28)	(−0.70)	(2.91)	(1.22)	(−1.8)	(2.42)	
Singapore	0.67	−0.13	0.80	0.59	−0.39	0.98	−2.97
	(2.56)	(−0.40)	(3.25)	(2.06)	(−1.30)	(3.26)	
Japan	0.35	−0.01	0.36	0.42	−0.10	0.53	0.00
	(4.60)	(−0.10)	(3.15)	(3.74)	(−1.20)	(3.56)	

This table reports the Fama–French four-factor alphas and the Newey–West t-statistics for the IMF country portfolios constructed using both equal- and value-weights for the improved magic formula portfolios PV, UG, and PV–UG. The last column reports the difference between each country's value-weight in the regional value-weighted PV portfolio and the country's overall regional value-weight.

portfolio takes a 5.01% larger position in Hong Kong and a 7.00% lower position in Australia.

5.4 Decile portfolios

Given the strong results obtained by cutting the sample into quintile portfolios, it is reasonable to suspect that even larger alphas are obtainable by zeroing in more closely on those stocks with the highest and the lowest IMF ranking. To investigate this, we construct decile portfolios and compute alphas for the PV and the PV–UG portfolios using the CAPM, FF-3, and FF-4 with both equal and value-weights. Results are presented in Table 14. Consistent with the hypothesis, long-only (PV) FF-4 alphas relative to those presented in Table 4 are larger when cutting the sample into decile portfolios rather than quintile portfolios. For North America and Europe, the increase is around 12 bps. Differences in Japan are mild

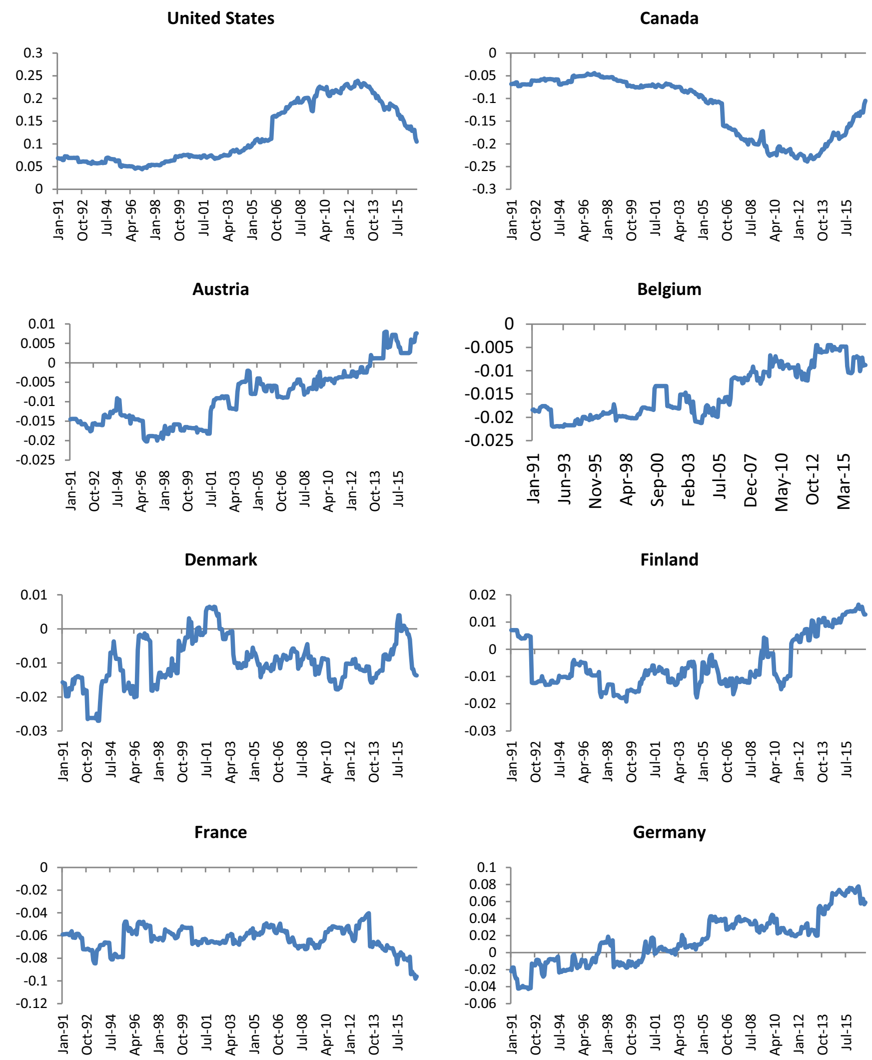

Figure 4 Relative portfolio weights by country.

The time-series of relative portfolio weights for the IMF value-weighted PV (long-only) portfolio are plotted for each country in the regions of North America (NA), Europe, Japan, and Asia. The relative portfolio weight is the value-weight of each country in the region's IMF PV portfolio minus the value-weight of the country in the region.

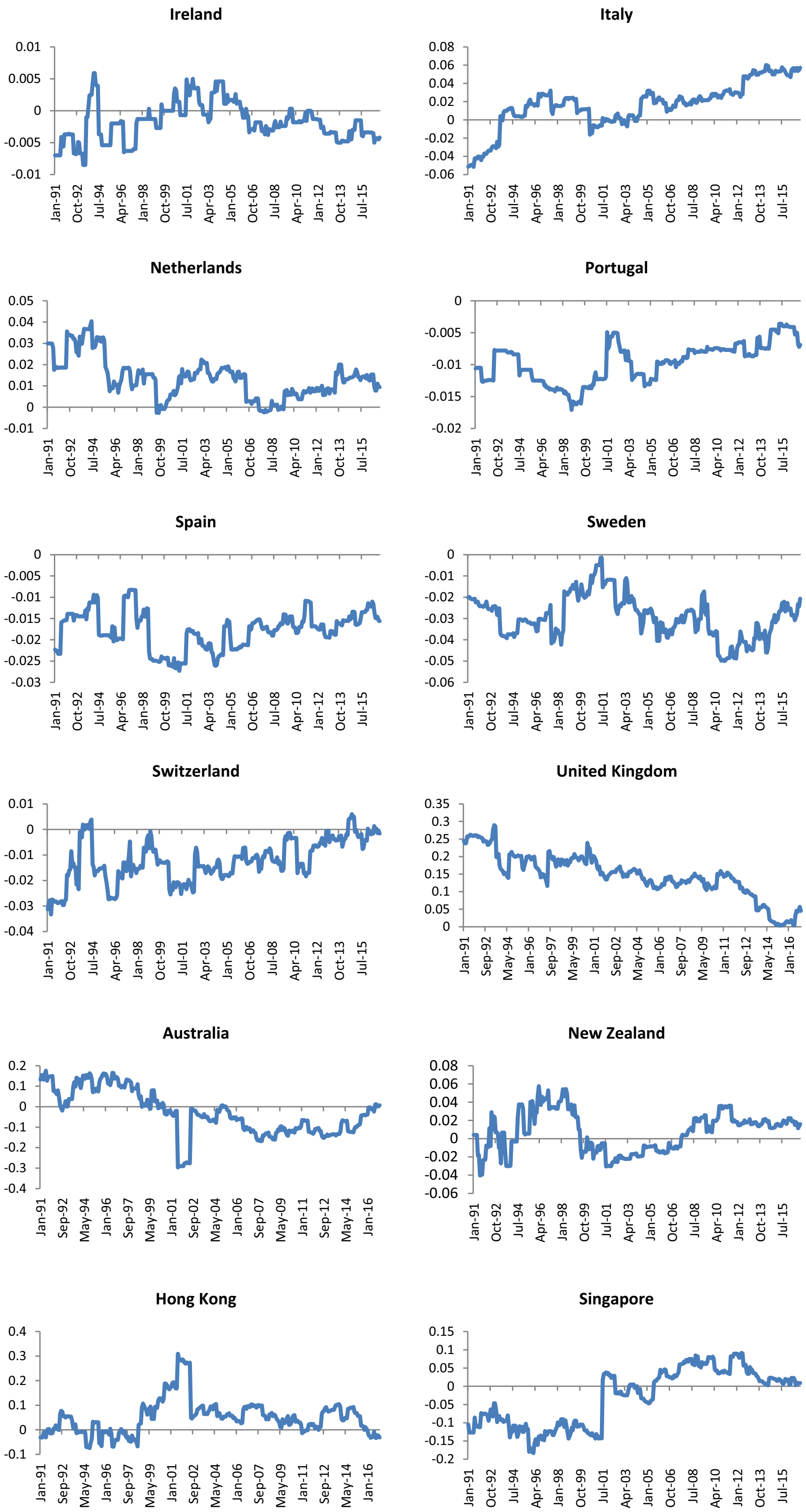

Figure 4 (*Continued*)

Table 14 The improved magic formula—Decile portfolios.

	$CAPM_{EW}$	$FF\text{-}3_{EW}$	$FF\text{-}4_{EW}$	$CAPM_{VW}$	$FF\text{-}3_{VW}$	$FF\text{-}4_{VW}$
Panel A: North America						
Alpha $_{PV}$	1.19	0.89	0.95	0.60	0.39	0.39
	(4.94)	(6.76)	(6.64)	(2.91)	(2.75)	(2.67)
Alpha $_{PV-UG}$	1.04	0.94	0.97	1.65	1.52	1.43
	(2.35)	(2.14)	(2.18)	(4.28)	(4.31)	(4.03)
Panel B: Europe						
Alpha $_{PV}$	0.72	0.63	0.61	0.49	0.41	0.33
	(3.59)	(5.83)	(5.68)	(3.28)	(3.229)	(2.304)
Alpha $_{PV-UG}$	1.16	1.09	0.95	1.17	1.05	0.79
	(7.36)	(7.22)	(6.67)	(5.10)	(4.53)	(2.66)
Panel C: Japan						
Alpha $_{PV}$	0.59	0.44	0.39	0.64	0.59	0.50
	(2.83)	(4.46)	(4.23)	(3.36)	(2.17)	(2.20)
Alpha $_{PV-UG}$	0.54	0.56	0.48	0.68	0.70	0.64
	(3.83)	(3.78)	(3.53)	(2.97)	(2.49)	(2.57)
Panel D: Asia						
Alpha $_{PV}$	0.91	1.01	1.04	0.53	0.67	0.61
	(4.55)	(7.968)	(7.13)	(3.22)	(4.14)	(3.64)
Alpha $_{PV-UG}$	0.35	0.05	0.06	0.97	0.65	0.84
	(1.09)	(0.16)	(0.21)	(2.22)	(1.53)	(2.36)

In each month t and for each region (North America, Europe, Japan, and Asia), stocks are ranked by gross profits-to-tangible capital employed and by gross profits-to-enterprise value. The two rankings are then summed to form a combined rank. Stocks are sorted into deciles according to the combined ranking using the 10, 20, ..., 90 percentile breakpoints. Time t returns of the equal-weighted (EW) and value-weighted (VW) portfolios comprising the stocks in each decile are averaged across all months from January 1991 to December 2016. Alphas are the risk-adjusted return from regressing the PV and PV–UG portfolio returns, both EW and VW, on the regional market portfolio (CAPM), on the market, SMB, and HML (FF-3), and on the market, SMB, HML, and momentum (FF-4). Newey–West t-statistics are reported in parentheses.

at 4 bps to 8 bps depending on whether equal or value-weights are used, but large increases in alpha are observed for Asia in the range of 20 bps to 30 bps.

Comparing the decile PV–UG portfolio alphas listed in Table 13 to the quintile portfolio alphas from Table 3 also show an increase in alpha. The evidence for North America and Europe strongly support the conjecture. Comparing the long–short value-weighted alpha computed using FF-4 shows an increase of 24 basis points in North America and an increase of 34 basis points in Europe when changing from quintile to decile portfolios. Results for Japan and Asia are much milder yielding increases of only 10 bps and 2 bps, respectively.

5.5 *Alternative factor models*

There have been a number of alternative factor models proposed in the recent literature to better explain observed equity return anomalies. Fama and French (2016, 2017) suggest adding two additional factors to their three-factor model—a profitability factor (RMW) and investment factor (CMA). The q-factor model of Hou *et al.* (2015) is a four-factor model which includes a market factor, a size factor, an investment factor and a profitability factor. We test whether these models are able to explain IMF returns.

Table 15 reports alphas obtained from the Fama–French five-factor model (FF-5) and Hou *et al.* (2015) q-factor model (HXZ). For each region we report alphas for each quintile portfolio and the alpha for the PV–UG long–short portfolio. Comparing the FF-5 alphas in Table 15 with the FF-4 alphas in Table 4 shows that replacing the momentum factor with the profitability and investment factor does tend to reduce alpha. For example, the value-weighted PV–UG alpha in North America drops from 1.19% when using the FF-4 factors to 0.73% when using the FF-5 factors—both values are statistically significant. FF-5 alphas are significant for the equal-weighted long-only PV

Table 15 Alternative factor models.

	Q1 (PV)		Q2		Q3		Q4		Q5 (UG)		PV–UG	
	Alpha	t-Stat	Alpha	t-Stat	Alpha	t-Stat	Alpha	t-Stat	Alpha	t-Stat	Alpha	t-Stat
Panel A: North America												
$FF5_{EW}$	0.84	5.79	0.55	4.80	0.34	3.34	0.08	0.53	0.25	0.62	0.60	1.62
$FF5_{VW}$	0.05	0.53	0.15	2.08	−0.05	−0.60	−0.03	−0.25	−0.68	−2.86	0.73	2.63
HXZ_{EW}	1.01	5.64	0.72	5.00	0.51	4.12	0.27	1.57	0.55	1.29	0.46	1.13
HXZ_{VW}	0.09	0.84	0.20	2.35	0.02	0.26	0.06	0.44	−0.52	−2.17	0.62	2.11
Panel B: Europe												
$FF5_{EW}$	0.40	3.93	0.17	2.25	0.12	1.42	0.03	0.37	−0.15	−1.26	0.55	4.68
$FF5_{VW}$	0.10	1.08	0.14	2.24	0.08	0.89	−0.17	−1.73	−0.17	−1.08	0.27	1.20
HXZ_{EW}	0.57	3.56	0.28	2.24	0.21	1.64	0.11	0.88	−0.16	−0.92	0.73	6.58
HXZ_{VW}	0.20	2.18	0.17	2.25	0.11	1.16	0.01	0.06	−0.11	−0.82	0.31	1.68
Panel C: Japan												
$FF5_{EW}$	0.38	4.60	0.30	5.18	0.10	1.46	0.06	0.77	−0.02	−0.31	0.40	3.80
$FF5_{VW}$	0.49	4.55	0.11	1.33	0.02	0.23	0.00	−0.04	−0.07	−0.96	0.57	3.95
HXZ_{EW}	0.41	1.87	0.32	1.52	0.15	0.67	0.12	0.58	0.00	0.02	0.40	3.40
HXZ_{VW}	0.52	4.29	0.15	1.61	0.04	0.34	0.11	1.42	−0.02	−0.25	0.55	3.43
Panel D: Asia												
$FF5_{EW}$	0.69	6.19	0.19	1.67	0.05	0.41	0.00	−0.02	0.49	1.88	0.20	0.71
$FF5_{VW}$	0.19	1.40	0.14	0.97	0.10	0.57	−0.29	−2.02	−0.35	−1.30	0.53	1.65
HXZ_{EW}	0.75	4.13	0.32	1.86	0.12	0.67	−0.20	−0.93	0.14	0.43	0.61	2.07
HXZ_{VW}	0.27	1.82	0.20	1.56	0.04	0.28	−0.36	−2.70	−0.75	−2.26	1.03	2.69

This table lists alphas for equal-weighted (EW) and value-weighted (VW) quintile portfolios constructed using the improved magic formula methodology and on the 1–5 long–short portfolio (PV–UG). In each month t and for each region (North America, Europe, Japan, and Asia), stocks are ranked by gross profits-to-tangible capital employed and by gross profits-to-enterprise value. The two rankings are then summed to form a combined rank. Stocks are sorted into quintiles according to the combined ranking using the 20, 40, 60, and 80 percentile breakpoints. Alphas are the risk-adjusted returns on the Fama–French five-factor model (FF5) and on the Hou, Xue, and Zhang (HXZ; 2015) factors. t-statistics are Newey–West.

portfolio in all regions and the PV–UG equal-weighted alphas are significant in Europe and Japan. The value-weighted PV FF-5 alpha is significant in Japan and the PV–UG FF-5 alpha is significant in North America and Japan. The HXZ model yields similar results. The equal-weighted long-only PV alphas are positive and significant for all regions when using the HXZ model and the equal-weighted PV alphas are significant in Europe and Japan. PV–UG alphas are significant in Europe, Japan, and Asia when equal-weighting and are significant in North America, Japan, and Asia when value-weighting. While these alternative factor models do capture more of the alpha than the Fama–French four-factor model, long-only PV and long–short PV–UG alphas remain significant in many cases.

5.6 Book-to-market value measure

As our final test, we ask whether IMF is sensitive to the choice of value metric by substituting the GP-to-enterprise ratio for the more commonly used book-to-market ratio. Table 16 lists the FF-5 alphas for the IMF_{BM} quintile portfolios. Similar to the results presented in Table 15 for IMF, the long-only PV_{BM} portfolio yields positive and significant alphas for all regions when equal-weighting, and value-weighted PV_{BM} alphas are only significant for

Table 16 Magic formula using the book-to-market ratio.

	Panel A: North America				Panel B: Europe			
	EW FF-5		VW FF-5		EW FF-5		VW FF-5	
Quintile	Alpha	*t*-Stat	Alpha	*t*-Stat	Alpha	*t*-Stat	Alpha	*t*-Stat
1 (PV)	1.05	6.46	0.09	0.82	0.59	5.46	0.26	1.59
2	0.61	4.54	0.16	1.60	0.27	3.43	0.17	1.77
3	0.36	2.39	0.08	1.21	0.17	2.32	0.01	0.24
4	0.11	0.76	−0.02	−0.24	−0.06	−0.76	0.01	0.17
5 (UG)	−0.07	−0.26	−0.23	−1.66	−0.42	−3.98	−0.17	−2.01
PV–UG	1.12	5.16	0.32	1.68	1.01	8.77	0.43	2.11
	Panel C: Japan				Panel D: Asia			
1 (PV)	0.45	5.26	0.39	3.79	0.92	7.52	0.20	1.31
2	0.34	4.99	0.16	1.58	0.35	3.06	0.14	1.08
3	0.21	3.34	0.20	1.95	0.12	1.07	0.00	−0.01
4	0.00	−0.02	0.04	0.52	0.12	0.77	−0.06	−0.53
5 (UG)	−0.18	−2.11	−0.12	−1.71	−0.09	−0.42	−0.08	−0.32
PV–UG	0.63	5.67	0.50	4.16	1.02	4.40	0.28	0.91

This table lists alphas for equal- and value-weighted quintile portfolios constructed using the improved magic formula methodology where the book-to-market ratio is used as the value measure instead of GP-to-enterprise value. In each month *t* and for each region (North America, Europe, Japan, and Asia), stocks are ranked by gross profits-to-tangible capital employed and by the book-to-market ratio. The two rankings are then summed to form a combined rank. Stocks are sorted into quintiles according to the combined ranking using the 20, 40, 60, and 80 percentile breakpoints. Alphas are the risk-adjusted returns from regressing the quintile portfolio returns on the Fama–French Five-Factor model. *t*-Statistics are Newey–West.

Japan. Alpha is significant for the long–short PV–UG$_{BM}$ portfolio for all regions when equal-weighting and for Europe and Japan when value-weighting. These results suggest that IMF's performance is quite similar whether book-to-market or GP-to-enterprise value is used as the measure for value.

6 Conclusion

Greenblatt's (2006, 2010) magic formula recognizes and exploits the complimentary characteristics of value and profitability. It is well known that value stocks outperform growth stocks (Fama and French, 1992, 1993), and it has been shown that profitable stocks outperform unprofitable stocks (Novy-Marx, 2013); however, value stocks may be either profitable or unprofitable, and profitable stocks may be either cheap or expensive. It is reasonable to expect that the price of an inexpensive, profitable stock is more likely to trend upward than the price of an inexpensive, unprofitable stock, and in the same way, it is more likely that the price of an expensive, unprofitable stock will fall than the price of an expensive, profitable stock. Combining the two characteristics together yields an intuitively appealing strategy that is missed when focusing on only one of the two characteristic. This paper explicitly tests Greenblatt's magic formula across global markets to determine whether it is indeed magical.

Our sample includes 23 developed markets divided into the 4 regions of North America, Europe, Japan, and Asia, and it extends from January 1991 through December 2016. Our empirical results do not show global support for the magic formula using the measures of profitability and value prescribed by Greenblatt—EBIT-to-tangible capital employed (return on capital) and EBIT-to-enterprise value (earnings yield). Return differentials from portfolios sorted using the magic formula methodology are insignificant as are the risk-adjusted alphas from the Fama–French four-factor model for North America, Japan, and Asia. Return differential are only significant in Europe.

The magic formula, however, is easily resurrected by replacing EBIT with gross profits (GP). Novy-Marx (2013) claims GP to be a cleaner measure of economic profitability than EBIT. Motivated by this, we substitute GP for EBIT and construct the new measures GP-to-tangible capital employed for the profitability measure and GP-to-enterprise value as our valuation measure. Contributing to the debate on the superiority of GP over other profitability measures located farther down the income statement (Novy-Marx, 2013; Ball *et al.*, 2015), we find that this substitution leads to results that are much more magical. Profitable value (PV) stocks significantly outperform unprofitable growth (UG) stocks, and with return differentials (PV–UG) of 0.88%, 0.65%, 0.66%, and 0.62% for North America, Europe, Japan, and Asia, respectively, the results are consistent and strong across all regions. We further demonstrate that the improved magic formula remains profitable after taking into account trading costs.

We provide a number of additional results that support our primary conclusions. While PV–UG yields larger returns within the set of small stocks, spreads are also significant within the quintile of large market cap stocks. There is an increase in alpha when moving from quintile to decile portfolios, the IMF alphas are not isolated to a specific industry, and IMF yields larger return differentials during periods of high sentiment than low sentiment months—primarily due to the short side and not the long side. Double sorts on IMF and book-to-market equity show that IMF captures a different dimension of the cross-section of returns than the book-to-market value measure. Both book-to-market and IMF yield significant return differentials. Using the Fama–Macbeth regression framework, we show that IMF explains the cross-section of returns in addition to size, book-to-market equity, and momentum. This is shown using the sample of all stocks, small stocks, and big stocks. The multivariate regressions confirm that low IMF (profitable, value), small market cap, high book-to-market, and high momentum are predictors of high returns.

It is likely that further changes to the methodology may lead to even stronger results. Some possibilities may be to use different profitability and value measures, double sorting instead of taking the sum of two ranks, or accounting for transaction costs during the construction stage. Overall, our results confirm Greenblatt's value and profitability paradigm and provide strong and consistent evidence that it is globally profitable.

Notes

1 Asness *et al.* (2017) argue that market efficiency directly implies that the highest quality firms should yield higher prices than low-quality firms.

2 This is the only change made to the methodology. Everything else remains exactly the same.

3 Though Greenblatt (2006, 2010) recommends removing firms related to the financial sector from the data set, we opt to keep all financial stocks in our sample.

4 Country-by-country analysis was initially considered. We found, however, too few stocks for many individual countries, and the presentation and exposition of results was found to be too lengthy.

5 Throughout this paper, returns are expressed in U.S. dollars. Issues related to exchange rate risk are assumed away.

6 Datastream calculates EBIT by taking "pre-tax income and adding back interest expense on debt and subtracting interest capitalization." Working capital is the difference between total current assets and total current liabilities.

7 There are numerous other value multiples such as enterprise value-to-sales, price-to-earnings, and book-to-market. In this paper we attempt to stay close to Greenblatt's magic formula by using EBIT-to-enterprise value.

8 Datastream computes enterprise value as the market capitalization at the fiscal year end date plus preferred stock plus minority interest plus total debt less cash.

9 Datastream updates accounting variables each June. By lagging seven months, the same value of each accounting item is held constant over the calendar year. Ideally, instead of lagging accounting data by seven months we would prefer to use the data at the moment it is made available. Unfortunately, the data availability date is not provided in Datastream. We therefore use the methodology that is standard in the literature by lagging the accounting data by seven months—an approach Fama and French (1992) consider to be conservative.

10 For gross profits, we use the Datastream item #01100 which is defined as "the difference between sales or revenues and cost of goods sold and depreciation/depletion, and amortization".

11 Results for equal-weighted portfolios are not shown in order to conserve space. Results, however, are stronger when equal-weighting as alphas tend to be larger and turnover tends to be smaller. These results are available upon request.

12 Average correlation across all pairs of regions is 50%. The high correlation suggests a commonality in IMF returns across regions. This appears to differ from Chui *et al.* (2010) who find that differences in individualism across countries helps explain cross-country differences in momentum profits; however, while their analysis is conducted on countries, ours is conducted at the regional level.

13 The sentiment index is available on Jeffrey Wurgler's website—http://people.stern.nyu.edu/jwurgler/

14 The factors are available on Robert Stambaugh's website—http://finance.wharton.upenn.edu/∼stambaug/

15 Dividing the sample into sectors greatly diminishes the sample size. In two situations, the sample sizes are too small to form quintile portfolios.

16 Results are not computed for Israel due to the small sample size.

References

Asness, Clifford S., Andrea Frazzini, and Lasse Heje Pedersen, (2017). Quality Minus Junk. Working Paper.

Asness, Clifford S., Tobias J. Moskowitz, and Lasse Heje Pedersen, (2013). "Value and Momentum Everywhere," *The Journal of Finance* **68**, 929–985.

Baker, Malcolm, Brendan Bradley, and Jeffrey Wurgler, (2006). "Investor Sentiment and the Cross-section of

Stock Returns," *The Journal of Finance* **61**, 1645–1680.

Baker, Malcolm, Brendan Bradley, and Jeffrey Wurgler, (2011). "Benchmarks as Limits to Arbitrage: Understanding the Low-Volatility Anomaly," *Financial Analysts Journal* **67**, 40–54.

Ball, Ray and Philip Brown, (1968). "An Empirical Evaluation of Accounting Income Numbers," *Journal of Accounting Research* **6**, 159–178.

Ball, Ray, Joseph Gerakos, Juhani T. Linnainmaa, and Valeri V. Nikolaev, (2015). "Deflating Profitability," *Journal of Financial Economics* **117**, 225–248.

Basu, Sanjoy, (1983). "The Relationship between Earnings Yield, Market Value, and Return for NYSE Common Stocks: Further Evidence," *Journal of Financial Economics* **12**, 129–156.

Carhart, Mark M. (1997). "On Persistence in Mutual Fund Performance," *The Journal of Finance* **52**, 57–82.

Chui, A. C. W, Sheridan, T. and John Wei, K. C. (2010). "Individualism and momentum around the world." *The Journal of Finance* **65**, 361–392.

Fama, Eugene F. and Kenneth R. French, (1992). "The Cross-section of Expected Stock Returns," *The Journal of Finance* **47**, 427–465.

Fama, Eugene F., and Kenneth R. French, (1993). "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics* **33**, 3–56.

Fama, Eugene F., and Kenneth R. French, (2012). "Size, Value, and Momentum in International stock Returns," *Journal of Financial Economics* **105**, 457–472.

Fama, Eugene F., and Kenneth R. French, (2016). "Dissecting Anomalies with a Five-Factor Model," *Review of Financial Studies* **29**, 69–103.

Fama, Eugene F., and Kenneth R. French, (2017). "International Tests of a Five-Factor Asset Pricing Model," *Journal of Financial Economics* **123**, 441–463.

Fama, Eugene F., and James D. MacBeth, (1973). "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy* **81**, 607–636.

Frazzini, Andrea, Ronen Israel, and Tobias J. Moskowitz, (2012). "Trading Costs of Asset Pricing Anomalies," Working Paper SSRN #2294498.

Frazzini, Andrea and Lasse Heje Pedersen, (2014). "Betting against Beta," *Journal of Financial Economics* **111**, 1–25.

Greenblatt and Joel, (2006). *The Little Book That Beats the Market*. John Wiley & Sons.

Greenblatt and Joel, (2010). *The Little Book That Still Beats the Market*. Hoboken.

Hou, Kewei, Chen Xue, and Lu Zhang, (2015). "Digesting Anomalies: An Investment Approach," *The Review of Financial Studies* **28**, 650–705.

Joenväärä, Juha, Robert Kosowski, and PekkaTolonen, (2016). Hedge Fund Performance: What Do We Know? Working Paper.

Newey, Whitney K., and Kenneth D. West, (1987). "A Simple, Positive Semi-definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* **55**, 703–708.

Novy-Marx and Robert, (2013). "The Other Side of Value: The Gross Profitability Premium," *Journal of Financial Economics* **108**, 1–28.

Novy-Marx, Robert, and Mihail Velikov, (2015). "A Taxonomy of Anomalies and Their Trading Costs," *Review of Financial Studies* **29**, 104–147.

Petajisto, Antti, (2013). "Active Share and Mutual Fund Performance," *Financial Analyst Journal* **69**, 73–93.

Sharpe, William F. (1964). "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance* **19**, 425–442.

Stambaugh, Robert F., and Yu Yuan, (2016). "Mispricing Factors," *Review of Financial Studies* **30**, 1270–1315.

Keywords: Return predictability; profitability; value; magic formula; cross-section of returns

JEL Classification: F21, F30, G10, G11, G12, G15