JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 50, No. 6, Dec. 2015, pp. 1237–1267

doi:10.1017/S0022109015000575

Is Momentum an Echo?

Amit Goyal and Sunil Wahal*

Abstract

In the United States, momentum portfolios formed from 12 to 7 months prior to the current month deliver higher future returns than momentum portfolios formed from 6 to 2 months prior, suggesting an "echo" in returns. In 37 countries excluding the United States, there is no robust evidence of such an echo. In portfolios that combine securities in developed and emerging markets, or across three major geographic regions (Americas excluding United States, Asia, and Europe), there is also no evidence of an echo. Any echo in the United States appears to be driven largely by a carryover of short-term reversals from month −2.

I. Introduction

Momentum, the notion that winners continue to win and losers continue to lose, is robust and pervasive. It exists cross sectionally in individual stocks (Jegadeesh and Titman (1993), (2001)), in portfolios (Lewellen (2002)), outside the United States (Rouwenhorst (1998), Griffin, Ji, and Martin (2003), and Chui, Titman, and Wei (2010)), in various asset classes (Asness, Moskowitz, and Pedersen (2013)), and apparently in the time series (Moskowitz, Ooi, and Pedersen (2012)). Unsurprisingly, it has spawned theory, both behavioral (Barberis, Shleifer, and Vishny (1998), Daniel, Hirshleifer, and Subrahmanyam (1998)) and rational (Berk, Green, and Naik (1999), Johnson (2002), and Sagi and Seasholes (2007)). In general, the theory fits one or more of these manifestations of momentum, but not all.

Several authors have worried about the term structure of predictability based on returns over the prior 12 months. Specifically, which months contribute more or less to the profitability of momentum strategies, and why? Jegadeesh (1990) estimates cross-sectional regressions of current-month returns on all prior 12 months'

*Goyal, amit.goyal@unil.ch, University of Lausanne, Swiss Finance Institute, Lausanne, CH-1015, Switzerland; Wahal (corresponding author), sunil.wahal@asu.edu, Arizona State University, Carey School of Business, Tempe, AZ 85287. We thank an anonymous referee, Scott Cederburg (our discussant at the 2013 European Finance Association Annual Meeting), Jennifer Conrad, Kenneth French, Jarrad Harford (the editor), Narasimhan Jegadeesh, Jeffrey Pontiff, and seminar participants at Babson College and the University of Toronto for helpful comments. We especially thank Michael Wolf for help with bootstrap issues. Zhiyi Qian provided excellent research assistance. Goyal thanks Rajna Gibson for her support through her National Centre of Competence in Research (NCCR)–Financial Valuation and Risk Management (FINRISK) project.

returns. His results show a reversal in returns from the prior month (the now well-known 1-month reversal effect), and that the prior 12th month's return has a large positive influence on the current month (sometimes referred to as the "12-month" effect). Heston and Sadka (2008), (2010) estimate similar return-response regressions and find that stocks have high or low returns every year in the same calendar month. The most recent contribution to this area comes from Novy-Marx (2012). His rather striking finding is that portfolios formed from sorting on returns from 12 to 7 months prior (intermediate horizon returns, hereafter denoted IR) predict average future returns better than portfolios formed from sorting on returns from 6 to 2 months prior (recent horizon return portfolios, hereafter denoted RR). The magnitude of the difference in profits from IR versus RR strategies is as much as 0.50% per month, which Novy-Marx argues can make an enormous difference in overcoming transaction costs incurred during implementation. More importantly, this "echo," rather than a continuation in returns, challenges theory and effectively leaves it flat-footed (no model predicts such a term structure) and Novy-Marx himself does not present an explanation for the result.

It is conventional to perform out-of-sample tests to establish the veracity of such provocative results. One way to do this is to examine different sample periods. For example, Jegadeesh and Titman (2001) look for momentum between 1990 and 1998 to establish the out-of-sample reliability of their original results (which are based on data from 1965 to 1989).[1] Novy-Marx (2012) looks at various subperiods and finds that the echo in returns occurs during most (but not all) of his sample period. Another approach is to see if the phenomenon occurs in other markets. Indeed, virtually every major "fact" in cross-sectional asset pricing (such as the value premium or momentum) is established by both U.S. and international evidence. Novy-Marx examines the term structure of momentum returns in commodities, currencies, and international (country) indices. He argues that although a few of these tests are not independent, the fact that the same patterns appear in these indices suggests that the echo is genuine. This apparent robustness, together with the lack of rational or behavioral intuition for why there might be an echo, makes the result puzzling.

As a true out-of-sample test, we examine the term structure of momentum across stocks in 37 countries between 1980 and 2010. We create winner-minus-loser (WML) portfolios defined separately over IR and RR for each country. Future monthly returns and 3-factor alphas for IR portfolios are largely indistinguishable from RR portfolios for all 37 countries. Because the number of stocks and market capitalization of some countries can be small, power can be a concern. We combine firm-countries in a variety of ways, equal or value weighting countries, and pooling securities from various countries into five aggregated portfolios: developed markets, emerging markets, Americas excluding United States, Asia, and Europe. In these more powerful portfolio tests, there is still no robust evidence that IR portfolios deliver returns that are different from those of RR portfolios.

We also construct portfolios independently sorted on RR and IR. Double-sorting allow us to look at returns on portfolio strategies that focus on IR but are

[1]Similarly, Heston and Sadka (2008) confirm the cross-sectional serial correlation patterns in Jegadeesh (1990), extending the sample period by 20 years.

neutral with respect to RR, thereby avoiding any mechanical covariance between strategies. For example, within the loser portfolio constructed from RR, we can look at the WML portfolio constructed from IR. Comparing the above return to its "mirror" (the WML portfolio constructed from RR for the loser portfolio based on IR) allows us to hold one source of variation constant. In the United States, Novy-Marx (2012) finds that the excess returns of IR portfolios are roughly twice that of RR portfolios, even after holding recent and IR constant. We do not see this in international data; there is no systematic evidence that IR portfolios outperform RR portfolios in any of the five aggregated portfolios or in the 37 individual countries.

Both single- and double-sorted portfolio strategies do not explicitly control for stock characteristics. We estimate the Fama–MacBeth (1973) regressions that forecast future monthly returns using IR and RR, controlling for size, book-to-market, and the prior month's return. In developed markets, emerging markets, Americas excluding United States, and Europe, there is no evidence of an echo. In Asia, slopes on IR are higher than those on RR, but this is driven by Japan, which has a large (relative) market capitalization. In individual countries, there are two markets (Japan and China) in which the coefficients on IR are significantly larger than those on RR, but there are also three markets (Belgium, Denmark, and Chile) in which the opposite is true. These likely random exceptions aside, we do not see any evidence that IR are better or worse at forecasting future returns than are RR.

These results raise the obvious question of why the echo is so strong in the United States and not anywhere else. Chui et al. (2010) suggest that the degree of individualism across countries (which they argue is correlated with overconfidence and attribution bias) helps explain cross-country variation in momentum profits. Indeed, the United States has the highest value in Hofstede's individualism index (91); one might, therefore, speculate that this is somehow related to the echo in returns. But the next two countries with the highest individualism index are Australia (90) and the United Kingdom (89). In both these countries, there is no reliable evidence that IR portfolios outperform RR portfolios.

Theory offers no guidance what can or should be considered "intermediate" or "recent." Therefore, we conduct an in-depth examination of the full term structure of returns in the United States. We estimate regressions in the spirit of Jegadeesh (1990) and forecast future monthly returns using each of the past 12 months' returns while also controlling for size and book-to-market ratios. The coefficients from these regressions can be interpreted as standard return responses, describing the impact of each prior month's return on future returns, and are informative for the puzzle. IR portfolios contain the prior 12 month's return response, which is higher than all other months' responses (the 12-month effect described earlier). RR portfolios contain the return response from month −2. On average, the return response of month −2 is negative, which appears to be a spillover from short-term reversals associated with the prior month −1. The return response of all other months (months −3 through −11) are reliably positive, consistent with momentum.

To determine whether these two return response effects account for the outperformance of IR portfolios in the United States, we aggregate coefficients in ways that correspond to all possible definitions of intermediate and recent.

The definitions are obviously not independent, so we take seriously Leamer's (1978), (1983) monition regarding inadvertent data snooping biases.[2] We employ a bootstrap procedure designed by Romano and Wolf (2005), which is a generalization of White's (2000) reality-check method. The procedure asks whether a specific trading strategy (i.e., a specific aggregation of coefficients from the above regressions) generates returns that beat a benchmark, while explicitly recognizing that a researcher could examine many such correlated strategies. In many definitions of IR and RR (although not the Novy-Marx (2012) definition), return responses of IR are higher than those of RR, even when assessing statistical significance using the bootstrap. But this difference is driven by the negative return response of month −2. Indeed, if one excludes coefficients from month −2, there are no differences average coefficients in the United States using any definition of IR versus RR. The importance of these months is also evident in regressions that exclude the United States. Excluding months −2 and −12, there are no differences in IR and RR coefficients in developed markets, emerging markets, Americas excluding United States, and Asia. If anything, in Europe, return responses of RR (excluding month −2) are higher than those of IR (excluding month −12).

Aggregated coefficients from regressions are informative but are not trading strategies per se. Therefore, we also create IR portfolios using months −12 to −8, and RR portfolios using months −7 to −3 (thereby excluding the month −2 return in portfolio formation) in the United States. The IR portfolios have an inherent and deliberate bias because they still include the 12-month effect. The difference in future 3-factor alphas between the two strategies is only 0.24% (t-statistic = 0.98). Using double-sorted portfolios, a similar difference in future 3-factor alphas is 0.23% (t-statistic = 1.20). Thus, even the evidence from portfolio strategies suggests that month −2 has an inordinate influence on the difference in returns between IR and RR portfolios.

If, as it seems, there is no real echo in returns and the apparent echo is just a manifestation of short-term reversals, there is no real challenge to theories of momentum. Instead, the challenge is passed on to theories of short-horizon reversals. There are two such candidates. Models such as Campbell, Grossman, and Wang (1993) argue that reversals are due to liquidity effects (Jegadeesh and Titman (1995a) similarly argue for microstructure biases and inventory effects). Although not outside the realm of possibility, it is hard to believe that these liquidity effects last for almost 2 months. A second candidate explanation is overreaction by investors (Jegadeesh and Titman (1995b)). It is analytically hard (or impossible) to reject overreaction stories because the theory does not identify specific horizons: Overreaction can occur over 2 months just as much as 1 month. Regardless, there does not appear to be a difference in returns of portfolios

[2]Leamer ((1983), p. 43) complains of the "fumes which leak from our computing centers" and argues that it is critical that the fragility of results from trading strategies be studied in a systematic way to establish or alter beliefs. The danger of falsely rejecting the null is of course not unique to finance or the social sciences. For an interesting account of such inadvertent inferences in the physical sciences, see Mlodinow (2008). Leamer (1978), (1983) and Lo and MacKinlay (1990) also emphasize that the effects of data snooping need not be willful; chance has a role to play when researchers and investors engage in completely appropriate specification searches.

constructed from various combinations of IR and RR: The enigma remains short-horizon reversals.

The remainder of the paper is organized as follows: Section II describes our data and explains sampling procedures. Section III describes out-of-sample tests outside the United States. Section IV pursues explanations for the U.S. results and includes bootstrapping exercises. Section V concludes.

II. Data and Sample

We obtain time series of market and accounting information for a broad cross section of firms (excluding the United States) from Datastream. We start with an unconstrained universe of all firms in 45 countries in the MSCI (formerly Morgan Stanley Capital International) All Country Index between 1980 and 2010. MSCI classifies 23 of these countries as developed markets and the remaining 22 as emerging markets. The universe includes live as well as dead stocks, ensuring that the data are free of survivorship bias. We apply the following sequence of filters that are derived from the extensive data investigations by Ince and Porter (2006), Griffin, Kelly, and Nardari (2010), and Hou, Karolyi, and Kho (2011):

i) We require that stocks have data from Datastream and Worldscope. The former is the source for market data whereas the latter contains necessary accounting data.

ii) We retain only issues that are listed as equity and require that they be from the firm's primary exchange. The latter, along with another Datastream supplied field, removes duplications.

iii) We eliminate all nonlocal firms with a GEOG code different from the local market. This eliminates foreign listings and therefore double-counting of securities.

iv) When a security dies, the post-death local currency returns are marked as 0, which we eliminate.

v) We employ a text search algorithm to eliminate securities that are not common stock. This eliminates preferred stock, trusts, warrants, rights, real estate investment trusts, closed-end funds, exchange-traded funds, and global and American depositary receipts that are not caught by the above screens.

vi) We compute returns using the return index (which includes dividends) supplied by Datastream. Because both return indexes and market capitalization are provided in local currency, we convert them to U.S. dollar equivalents using the conversion function built into Datastream. We perform some simple checks to ensure that local and U.S. dollar returns are consistent and computed correctly.

vii) We set returns to missing for a stock when it rises by 300% or more during one month and drops by 50% or more the following month (or falls and subsequently rises). We also treat as missing returns above 1,000% per month.

viii) Each month and for each country, we winsorize returns, market capitalization, and book-to-market ratios using the 1st and 99th percentiles.

For most of our tests, we require that a country-month have at least 100 stocks. This ensures that country-specific portfolios are reasonably well diversified. The restriction removes the following seven countries from our single country tests: Ireland, Colombia, Czech Republic, Egypt, Hungary, Morocco, and Poland. These countries are, however, still included in aggregate regional tests described later.

For the U.S. sample, we use all New York Stock Exchange (NYSE), American Stock Exchange (AMEX), and NASDAQ stocks in the Center for Research in Security Prices (CRSP) database. Book values are from Compustat and are calculated following the procedure described in Fama and French (1993).[3] The rest of the data come from CRSP. The book-to-market ratio is calculated as the ratio of the most recently available (assumed to be available 6 months after the fiscal year-end) book value of equity divided by the current market capitalization. Book-to-market ratio is winsorized each month using the 0.5 and the 99.5 percentiles. We include only common stocks with share codes 10 or 11 on CRSP.

Table 1 provides a brief overview of the sample. Countries are ordered alphabetically within developed and emerging markets. Data availability varies across countries. We report the start date for the data series for which momentum portfolios can be constructed for each country (we discuss portfolio formation procedure in Section II.B). Not surprisingly, the available time series is longer for developed markets. For major developed markets (e.g., Australia, Canada, the United Kingdom, Japan), there is a full time series of 30 years. We report the average and aggregate market capitalization as of the end of 2010 in millions and billions of U.S. dollars, respectively. Canada, France, Germany, Japan, the United Kingdom, China, and India each has market capitalization in excess of $1 trillion at the end of 2010 (with Switzerland and Brazil close behind). At the end of 2010, these seven countries had a joint market capitalization of around $12 trillion, roughly two-thirds that of the United States. These seven countries together constitute about 23% of the global stock market capitalization. As a check on the coverage of our data, we compare the aggregate market capitalization for each country at the end of each year with statistics reported by the World Federation of Exchanges (www.world-exchanges.org). For the years for which these comparisons are feasible, the firms in our sample cover between 80% and 99% of the aggregate market capitalization of each country. We also report the average number of stocks per country, where the average is computed over the entire (monthly) time series. Despite the fact that we require a minimum of 100 securities per month, there is still considerable variation in the average number of stocks per country.

Some of our tests aggregate countries into five different regions as follows:

i) Developed markets: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Hong Kong, Ireland, Israel, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, and the United Kingdom.

[3]Book values from Compustat are supplemented with hand-collected values from Moody's, whenever available (see Davis, Fama, and French (2000) for the exact description of these data). These are available on Kenneth French's Web site at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

TABLE 1

Market Statistics, Returns, and Alphas from Monthly Sorts on Returns over Prior 11 Months (−12 to −2): International Sample (1980–2010)

In Table 1 we sort all stocks with valid data in a country into quintiles based on prior returns over months −12 to −2 (we require at least 100 stocks per month to do the sorts). We report average monthly value-weighted returns of winner-minus-loser quintiles. We calculate alphas only if there are at least 60 months of return data. All returns are in U.S. dollars and are reported in percentages per month (numbers in parentheses are their *t*-statistics). Average (aggregate) market capitalization is at the end of 2010 in U.S. millions (billions) of dollars. For regions, country-level data are value weighted based on the prior month's country market capitalization. CAPM refers to capital asset pricing model. The sample period is 1980–2010.

	Start Date	Average No. of Stocks	Average Market Cap	Aggregate Market Cap	Portfolio Returns		CAPM Alpha		3-Factor Alpha	
Developed Markets										
Australia	1980–2001	742	503	844	2.43	(4.50)	2.51	(4.63)	2.27	(4.31)
Austria	1980–2001	70	1,399	116	2.59	(1.93)	—	—	—	—
Belgium	1980–2001	108	958	172	1.29	(1.89)	1.48	(2.50)	1.55	(2.55)
Canada	1980–2001	831	1,390	1,404	2.17	(4.80)	2.36	(5.31)	2.48	(5.02)
Denmark	1980–2001	150	938	163	1.41	(3.26)	1.50	(3.44)	1.43	(3.20)
Finland	1987–2001	72	1,567	191	0.97	(1.05)	1.11	(1.21)	0.63	(0.67)
France	1980–2001	527	1,729	1,442	0.98	(2.13)	1.14	(2.53)	1.40	(3.08)
Germany	1980–2001	506	1,140	1,120	1.46	(3.50)	1.64	(3.99)	1.75	(4.19)
Greece	1980–2001	158	220	56	1.35	(2.09)	1.42	(2.22)	1.50	(2.33)
Hong Kong	1980–2001	139	3,961	784	0.68	(1.28)	0.92	(1.75)	0.98	(1.84)
Israel	1980–2001	351	228	126	1.36	(2.47)	1.50	(2.65)	1.97	(2.04)
Italy	1980–2001	175	1,688	467	0.54	(1.24)	0.63	(1.48)	0.52	(1.23)
Japan	1980–2001	1,781	1,440	3,316	0.21	(0.59)	0.23	(0.64)	0.10	(0.28)
Netherlands	1980–2001	136	3,198	352	0.41	(0.75)	0.74	(1.42)	0.81	(1.48)
New Zealand	1980–2001	73	317	38	1.51	(1.65)	1.60	(1.79)	1.53	(1.68)
Norway	1980–2001	115	1,024	191	1.36	(2.23)	1.53	(2.54)	1.45	(2.40)
Portugal	1988–2001	63	1,269	75	0.80	(0.85)	0.57	(0.61)	0.37	(0.39)
Singapore	1980–2001	236	734	354	0.14	(0.29)	0.40	(0.85)	0.45	(0.96)
Spain	1980–2001	96	4,688	609	0.65	(1.34)	0.77	(1.59)	0.80	(1.68)
Sweden	1982–2001	172	980	459	1.06	(0.94)	1.52	(1.42)	1.59	(1.46)
Switzerland	1980–2001	183	4,228	994	1.06	(2.54)	1.21	(2.88)	1.15	(2.64)
United Kingdom	1980–2001	1,305	1,684	2,278	1.48	(3.68)	1.61	(4.06)	1.91	(4.30)
Emerging Markets										
Brazil	1990–2001	150	3,127	935	0.85	(0.95)	1.21	(1.35)	1.21	(1.35)
Chile	1989–2007	130	1,783	312	0.83	(1.79)	0.87	(1.85)	0.99	(2.12)
China	1991–2001	326	2,234	1,934	−0.07	(−0.16)	−0.06	(−0.14)	0.53	(1.35)
India	1981–2001	1,609	338	1,112	1.36	(2.21)	1.52	(2.27)	1.56	(2.25)
Indonesia	1990–2004	176	896	339	−0.13	(−0.16)	0.20	(0.26)	0.90	(1.19)
Korea	1980–2001	850	452	804	0.52	(0.93)	0.57	(0.95)	0.97	(1.57)
Malaysia	1980–2001	447	422	334	0.02	(0.04)	0.30	(0.55)	0.47	(0.87)
Mexico	1988–2001	77	3,393	383	0.97	(1.81)	1.11	(2.09)	1.15	(2.04)
Peru	1991–2001	77	606	69	0.26	(0.34)	0.26	(0.33)	0.78	(0.95)
Philippines	1980–2001	123	658	144	−0.28	(−0.31)	−0.08	(−0.10)	−0.04	(−0.05)
Russia	1995–2009	59	2,165	794	−2.14	(−1.11)	—	—	—	—
South Africa	1980–2001	280	1,641	509	1.78	(3.30)	1.71	(2.93)	1.84	(3.06)
Taiwan	1987–2009	351	935	712	0.01	(0.03)	0.04	(0.08)	0.15	(0.29)
Thailand	1987–2001	277	457	232	0.79	(1.25)	1.00	(1.63)	1.46	(2.24)
Turkey	1988–2001	162	898	287	−0.55	(−0.75)	−0.39	(−0.53)	−0.34	(−0.50)
Developed	1980–2001	8,004	1,322	15,603	0.77	(2.75)	0.85	(3.08)	0.81	(2.79)
Emerging	1980–2001	5,244	786	8,346	0.47	(1.98)	0.56	(2.41)	0.52	(2.18)
Americas	1980–2001	1,221	1,522	2,289	1.93	(5.23)	2.36	(4.57)	2.29	(4.28)
Asia	1980–2001	6,313	870	10,066	0.27	(0.87)	0.28	(0.91)	0.26	(0.83)
Europe	1980–2001	4,037	1,565	9,670	1.19	(3.85)	1.36	(4.52)	1.55	(5.12)

ii) Emerging markets: Brazil, Chile, China, Colombia, Czech Republic, Egypt, Hungary, India, Indonesia, Korea, Malaysia, Mexico, Morocco, Peru, Philippines, Poland, Russia, South Africa, Taiwan, Thailand, and Turkey.

iii) Americas excluding United States: Brazil, Canada, Chile, Columbia, Mexico, and Peru.

iv) Asia: China, Hong Kong, India, Indonesia, Japan, Korea, Malaysia, Philippines, Singapore, Taiwan, and Thailand.

v) Europe: Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Netherlands, Norway, Poland, Portugal, Russia, Spain, Sweden, Switzerland, Turkey, and the United Kingdom.

The last five rows of Table 1 show descriptive statistics on the five regions. All five regions are well diversified with at least 1,000 stocks per month on average.

A. Factors

To estimate alphas, we also need factors, which are constructed as follows. For a country to have a nonmissing factor for a particular month, we require at least five stocks per portfolio used to create the factor. For example, typical double-sorted factors that use a 3×2 sort require at least 30 stocks (six portfolios multiplied by five stocks). This is a minimal requirement and most factors are well diversified. To establish size breakpoints, we sort all stocks in a country into two groups, small and big, based on their market capitalization as of June of year t. Big stocks are those in the top 75% of market cap, and small stocks are those in the bottom 25% (the breakpoint for big corresponds to that used by MSCI). Book-to-market ratios are computed by dividing book value for the fiscal year ending in year $t-1$ by market capitalization at the end of December of year $t-1$. These values are used to generate breakpoints corresponding to low, medium, and high portfolios. The low and high portfolios correspond to the 30th and 70th percentiles, respectively, with medium being between those two. For each size and value portfolio, we compute monthly value-weighted returns in U.S. dollars from July of year t through June of year $t+1$. We compute country-specific factor returns as the difference in average returns across size groups.

In constructing factors to evaluate regional aggregated portfolios, we follow approaches that correspond directly to the method of portfolio construction. For instance, when we value weight countries to generate a value-weighted developed markets portfolio, we also value weight country-specific factors (using each country's total market capitalization at the end of the previous month). Similarly, in specifications where we pool stocks in developed markets, we generate breakpoints and factors using all firms in the entire region. This ensures consistency between the construction of portfolios and the factors used to evaluate them.

The U.S. factors are obtained directly from Kenneth French's Web site (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

B. Conventional Momentum Sorts

Table 1 also shows momentum profits for individual countries as well as regions. There are several ways to construct momentum portfolios. We start our analysis with the simple and conventional approach of using prior 11-month returns, skipping the most recent month, and holding these portfolios for 1 month.

For all countries except the United States, we sort all stocks in a country-month with valid returns separately into quintiles based on month -12 to -2 returns.[4] We require at least 100 stocks per month to do the sorts and report

[4]We use quintiles to ensure that portfolios are well populated. We also conduct tests using deciles whenever possible to ensure the robustness of our results. Details are reported in Section III.D.

only value-weighted returns. For aggregated portfolios, we value weight country portfolios. In addition to average monthly returns, we also estimate capital asset pricing model (CAPM) and 3-factor alphas. These are based on time-series regressions of these monthly returns on local market, small-minus-big (SMB), and high-minus-low (HML) factors (Fama and French (1993)). We calculate alphas only if there are at least 60 months of return data. In countries (Austria and Russia) where we are unable to calculate these alphas, we report only portfolio returns.

Table 1 shows WML portfolio returns, CAPM alphas, and 3-factor alphas for these momentum portfolios along with *t*-statistics in parentheses. Momentum portfolio returns are statistically significant in 12 of 22 developed markets and in 2 of 15 emerging markets. The 3-factor alphas are positive in 12 of 22 developed markets and in 5 of 15 emerging markets. As is well known, momentum portfolios generate no profits in Japan. In general, the cross-country variation in momentum lines up well with the results in Chui, Titman, and Wei (2010), despite the fact that they use a different portfolio-formation approach.[5] This is especially true in what one may consider "major" countries. Momentum returns are notably high in countries with large market capitalizations such as Australia, Canada, France, Germany, and the United Kingdom. In aggregated portfolios, momentum returns are large in developed and emerging markets as well as in Americas excluding United States and in Europe. In Asia, momentum portfolio returns are small and statistically insignificant, largely due to the above-noted influence of Japan.

III. Results

We now change our procedures to sort on different horizons of past returns. For all excluding United States countries, we require that all stocks in a country-month have valid intermediate (months −12 to −7) and recent (months −6 to −2) returns, as well as market capitalization and prices in the prior month. Using these stocks, we compute return breakpoints for each country-month, assigning stocks into quintiles.

We require at least 100 stocks per country-month to construct return quintiles; this is a minimum, so in major developed markets the quintiles contain many more securities. For instance, in the United Kingdom, on average, each quintile contains approximately 250 stocks. For the United States, we modify the process slightly by excluding stocks under $1, using deciles instead of quintiles, and establishing breakpoints using only NYSE stocks. We report statistics on portfolio returns for the United States over the sample period 1980–2010 to facilitate comparison with the international sample, as well as over the sample period 1927–2010 to facilitate comparison with prior studies.

Our tests are conducted on portfolios that combine securities within a country as well as aggregated portfolios across countries. Aggregated portfolios alleviate power concerns and offer more concise interpretation. Therefore, we report results based on aggregated portfolios and results based on specific countries in separate tables.

[5]Chui et al. (2010) assign stocks to terciles (not quintiles) based on the prior 6-month return and hold securities for 6 months (not 1 month).

Within a country, securities are always value weighted. There are four ways in which countries can be aggregated. Following Fama and French (1998), we equal weight or value weight (using total country market capitalization) across countries (labeled "EW" and "VW" in the tables). Value weighting countries has obvious advantages, except in some circumstances where one or two large countries can dominate the overall portfolio. For example, Japan has an extremely large weight in Asia (the average weight is 75% in our sample period, although it declines to 33% by the end of the 2010). In such circumstances, equal-weighted portfolios provide additional information. Both these weighting schemes ensure that all countries are represented in the aggregated portfolio. A third approach is to pool all securities within a region (labeled "pooled" in the tables). The resulting WML portfolio need not contain securities from all countries at a point in time but is still diversified. This is the approach used by Fama and French (2012), who combine 23 developed markets into four regions. To the extent that a particular country has a high representation in a pooled portfolio, country effects could be important. Therefore, we also generate pooled portfolios but sort based on stock returns in excess of the local country return (labeled "pooled ex-country" in the tables). All four approaches produce broadly diversified portfolios, albeit in different ways.

A. Single-Sorted Portfolios

1. Portfolio Returns

Table 2 shows WML portfolio returns for IR and RR momentum portfolios along with t-statistics in parentheses. RR(WML) denotes the hedge portfolio based on RR sorts and IR(WML) denotes the hedge portfolio based on IR sorts. We also report the difference in returns between these two hedge portfolios (RR(WML) − IR(WML), labeled "difference" in the tables).

In the United States, between 1980 and 2010, momentum portfolios formed on RR underperform portfolios formed on IR by 0.62% per month with a t-statistic of 1.71. The equivalent return over the full sample period is 0.51% (t-statistic = 2.11), which is close to the numbers reported by Novy-Marx (2012) (0.54% with a t-statistic of 2.21).

Outside the United States, Table 2 shows there are no statistically significant differences in IR versus RR portfolio returns in developed markets, emerging markets, Americas excluding United States, Asia, or Europe. This is true regardless of whether one considers equal-weighted or value-weighted aggregate portfolios. Even using the two pooled portfolio construction approaches, there is no difference in future returns between IR and RR portfolios.

Table 3 shows that there are no statistically significant differences in returns in 35 of 37 countries. The two exceptions are Australia and Israel, where IR portfolios underperform RR portfolios by 1.33% and 1.27% per month, respectively. In addition to looking at the aforementioned exceptions, it is informative to focus on countries with large market capitalizations. In major developed countries, there is no difference in returns for IR versus RR portfolios for Canada, France, Germany, or the United Kingdom. In Japan, the IR portfolio outperforms the RR portfolio by 0.57% per month but the t-statistic is only 1.50. In contrast,

TABLE 2

Returns and 3-Factor Alphas from Monthly Single Sorts on Recent (−2 to −6) or Intermediate (−7 to −12) Returns: International Sample (1980–2010)

In Table 2 we sort stocks into quintiles based on prior recent months (−2 to −6, RR) or intermediate months (−7 to −12, IR) returns separately for each country, requiring at least 100 stocks per month per country. Portfolios for each county are value weighted. Country portfolios are aggregated as follows: EW refers to equal weighting of country returns and VW refers to value weighting (using total country market capitalization of last month) of country returns. Pooled refers to pooling all securities within a region. Pooled ex-country also refers to pooling all securities within a region but we use returns in excess of local country returns for sorting. We require at least 1,000 stocks per month per region for pooled and pooled ex-country portfolios; the portfolios are value weighted. This table reports statistics on returns on winner-minus-loser (WML) quintiles for each region. We report average monthly returns and monthly 3-factor alphas of these portfolios. The 3-factor alphas are based on time-series regressions of monthly returns on local region-specific market, SMB, and HML factors (Fama and French (1993)). We calculate alphas only if there are at least 60 months of return data. In each case, regional factors are constructed analogously to the aggregation method. All returns are in U.S. dollars and are reported in percentages per month (numbers in parenthesis are their *t*-statistics). The overall sample period is 1980–2010. We follow the same procedure for the United States with the following exceptions. We use deciles, the return breakpoints are based only on New York Stock Exchange stocks, and we include only stocks with a share price greater than $1 at the end of the portfolio-formation period.

	Returns						3-Factor Alphas					
	RR(WML)		IR(WML)		Difference		RR(WML)		IR(WML)		Difference	
United States												
1927–2010	0.62	(2.66)	1.13	(5.38)	−0.51	(−2.11)	1.11	(5.25)	1.51	(7.78)	−0.40	(−1.67)
1980–2010	0.56	(1.56)	1.18	(4.08)	−0.62	(−1.71)	0.84	(2.35)	1.47	(5.19)	−0.63	(−1.74)
Developed Markets												
EW	0.88	(4.01)	0.79	(4.37)	0.09	(0.42)	0.91	(4.00)	1.08	(5.33)	−0.09	(−0.39)
VW	0.43	(1.57)	0.78	(3.70)	−0.35	(−1.28)	0.33	(1.14)	0.93	(4.22)	−0.60	(−2.06)
Pooled	0.55	(1.59)	0.91	(3.13)	−0.36	(−0.94)	0.47	(1.28)	0.77	(2.51)	−0.30	(−0.74)
Pooled ex-country	0.45	(1.52)	0.83	(3.62)	−0.38	(−1.25)	0.54	(1.70)	0.92	(3.77)	−0.39	(−1.18)
Emerging Markets												
EW	0.28	(1.08)	0.43	(2.16)	−0.11	(−0.35)	0.57	(2.25)	0.60	(2.83)	−0.03	(−0.10)
VW	0.28	(1.16)	0.54	(2.76)	−0.22	(−0.76)	0.42	(1.82)	0.61	(2.90)	−0.19	(−0.68)
Pooled	0.48	(0.88)	0.59	(1.17)	−0.10	(−0.16)	0.60	(1.08)	0.59	(1.19)	0.01	(0.01)
Pooled ex-country	0.22	(0.60)	0.15	(0.45)	0.08	(0.17)	0.39	(1.03)	0.21	(0.62)	0.19	(0.40)
Americas Excluding the United States												
EW	0.99	(3.42)	0.80	(2.74)	0.21	(0.58)	1.19	(3.72)	0.97	(3.15)	0.22	(0.55)
VW	1.42	(4.21)	1.22	(3.83)	0.24	(0.60)	1.56	(3.17)	1.53	(3.37)	0.02	(0.04)
Pooled	1.06	(2.31)	1.17	(2.95)	−0.11	(−0.21)	1.12	(1.31)	1.15	(1.56)	−0.03	(−0.03)
Pooled ex-country	1.45	(3.37)	0.71	(1.93)	0.75	(1.49)	1.83	(2.28)	0.95	(1.39)	0.88	(0.97)
Asia												
EW	0.12	(0.43)	0.35	(1.80)	−0.23	(−0.82)	0.30	(1.11)	0.56	(2.84)	−0.26	(−0.87)
VW	0.07	(0.22)	0.51	(2.11)	−0.44	(−1.31)	−0.05	(−0.16)	0.60	(2.42)	−0.65	(−1.87)
Pooled	−0.12	(−0.27)	0.72	(1.79)	−0.84	(−1.67)	−0.09	(−0.18)	0.92	(2.26)	−1.01	(−1.98)
Pooled ex-country	0.19	(0.51)	0.36	(1.16)	−0.16	(−0.39)	0.35	(0.93)	0.36	(1.15)	−0.01	(−0.02)
Europe												
EW	0.84	(3.55)	0.87	(4.26)	−0.03	(−0.11)	1.00	(4.24)	1.09	(5.33)	−0.09	(−0.37)
VW	0.74	(2.66)	0.92	(3.91)	−0.19	(−0.70)	1.00	(3.63)	1.10	(4.59)	−0.10	(−0.37)
Pooled	0.73	(2.39)	1.00	(3.75)	−0.27	(−0.88)	1.03	(3.32)	1.23	(4.48)	−0.20	(−0.64)
Pooled ex-country	0.93	(3.43)	0.92	(3.96)	0.01	(0.04)	1.19	(4.32)	1.12	(4.74)	0.07	(0.25)

in Australia, the IR portfolio underperforms the RR portfolio by 1.33% per month (*t*-statistic = 2.72). Given the multitude of numbers and the role of chance, it is not surprising to find a few IR portfolios exhibiting statistically different performance from that of RR portfolios. Our reading of this evidence is that, apart from a few exceptions, IR portfolios do not out- (or under-) perform RR portfolios.

2. Portfolio Alphas

For each aggregated portfolio, we calculate factor model alphas based on analogously constructed market, SMB, and HML factors (Fama and French (1993)). We use the 1-month Treasury bill rate as the risk-free asset and calculate alphas only if we have at least 60 time-series observations.

TABLE 3

Returns and 3-Factor Alphas from Monthly Single Sorts on Recent (−2 to −6) or Intermediate (−7 to −12) Returns: Individual Country Results (1980–2010)

In Table 3 we sort stocks into quintiles based on prior recent months (−2 to −6, RR) or intermediate months (−7 to −12, IR) returns separately for each country, requiring at least 100 stocks per month per country. Portfolios for each county are value weighted. This table reports statistics on returns on winner-minus-loser (WML) quintiles for each country. We report average monthly returns and monthly 3-factor alphas of these portfolios. The 3-factor alphas are based on time-series regressions of monthly returns on local region-specific market, SMB, and HML factors (Fama and French (1993)). We calculate alphas only if there are at least 60 months of return data. All returns are in U.S. dollars and are reported in percentages per month (numbers in parentheses are their *t*-statistics). The overall sample period is 1980–2010.

	Returns						3-Factor Alphas					
	RR(WML)		IR(WML)		Difference		RR(WML)		IR(WML)		Difference	
Developed Markets												
Australia	2.72	(5.54)	1.40	(3.53)	1.33	(2.72)	2.51	(5.13)	1.65	(4.04)	0.86	(1.71)
Austria	0.20	(0.23)	2.40	(2.35)	−2.20	(−1.76)	—	—	—	—	—	—
Belgium	1.35	(2.30)	0.68	(1.20)	0.67	(1.12)	1.58	(2.96)	0.87	(1.64)	0.71	(1.17)
Canada	1.68	(4.02)	1.45	(3.90)	0.23	(0.50)	2.02	(4.37)	1.61	(3.99)	0.41	(0.80)
Denmark	1.70	(4.02)	0.91	(2.27)	0.80	(1.45)	1.64	(3.79)	0.93	(2.25)	0.71	(1.29)
Finland	1.45	(1.63)	1.60	(1.92)	−0.15	(−0.16)	1.34	(1.49)	1.29	(1.54)	0.05	(0.05)
France	0.60	(1.42)	1.03	(2.87)	−0.43	(−0.87)	0.81	(1.96)	1.13	(3.13)	−0.33	(−0.66)
Germany	0.85	(2.05)	1.18	(3.23)	−0.33	(−0.70)	1.03	(2.44)	1.39	(3.77)	−0.36	(−0.75)
Greece	0.86	(1.24)	0.34	(0.59)	0.52	(0.63)	1.04	(1.52)	0.55	(0.99)	0.48	(0.59)
Hong Kong	0.45	(0.73)	0.33	(0.60)	0.12	(0.14)	0.67	(1.10)	0.65	(1.17)	0.02	(0.02)
Israel	1.45	(2.88)	0.18	(0.40)	1.27	(2.10)	1.43	(1.68)	1.40	(1.79)	0.03	(0.04)
Italy	0.61	(1.63)	0.76	(2.09)	−0.15	(−0.37)	0.56	(1.53)	0.74	(2.08)	−0.18	(−0.45)
Japan	−0.02	(−0.04)	0.55	(1.90)	−0.57	(−1.50)	−0.20	(−0.54)	0.58	(2.01)	−0.77	(−1.99)
Netherlands	0.64	(1.19)	0.79	(1.68)	−0.15	(−0.24)	0.89	(1.67)	1.08	(2.25)	−0.19	(−0.29)
New Zealand	0.51	(0.67)	1.06	(1.39)	−0.54	(−0.60)	0.45	(0.57)	0.98	(1.27)	−0.53	(−0.58)
Norway	1.26	(1.99)	1.74	(2.88)	−0.49	(−0.66)	1.29	(2.01)	1.93	(3.19)	−0.64	(−0.86)
Portugal	0.26	(0.28)	0.69	(0.81)	−0.43	(−0.38)	—	—	—	—	—	—
Singapore	0.23	(0.49)	0.25	(0.61)	−0.01	(−0.02)	0.34	(0.73)	0.57	(1.46)	−0.23	(−0.40)
Spain	0.62	(1.34)	0.63	(1.54)	−0.01	(−0.02)	0.82	(1.80)	0.75	(1.86)	0.07	(0.12)
Sweden	1.14	(1.07)	0.46	(0.64)	0.68	(0.61)	1.51	(1.45)	0.64	(0.90)	0.86	(0.77)
Switzerland	0.54	(1.35)	0.91	(2.39)	−0.37	(−0.89)	0.71	(1.69)	1.04	(2.61)	−0.33	(−0.75)
United Kingdom	0.75	(2.02)	1.00	(3.14)	−0.25	(−0.68)	1.10	(2.66)	1.23	(3.33)	−0.13	(−0.31)
Emerging Markets												
Brazil	−0.45	(−0.89)	0.28	(0.52)	−0.72	(−1.06)	−0.22	(−0.44)	0.36	(0.67)	−0.57	(−0.82)
Chile	1.19	(2.60)	0.21	(0.45)	0.98	(1.66)	1.07	(2.49)	0.65	(1.58)	0.41	(0.81)
China	0.11	(0.27)	0.56	(1.45)	−0.45	(−0.92)	0.28	(0.67)	1.08	(3.02)	−0.80	(−1.64)
India	1.59	(2.82)	1.16	(2.16)	0.43	(0.75)	1.74	(2.73)	1.27	(2.11)	0.48	(0.76)
Indonesia	0.11	(0.12)	0.19	(0.30)	−0.08	(−0.08)	1.20	(1.51)	0.76	(1.26)	0.43	(0.47)
Korea	−0.01	(−0.03)	0.42	(0.84)	−0.43	(−0.70)	0.33	(0.57)	0.95	(1.73)	−0.62	(−0.93)
Malaysia	0.27	(0.55)	0.12	(0.30)	0.15	(0.29)	0.70	(1.41)	0.35	(0.91)	0.34	(0.64)
Mexico	0.19	(0.36)	1.07	(2.20)	−0.89	(−1.34)	0.19	(0.35)	1.32	(2.57)	−1.13	(−1.59)
Peru	0.45	(0.59)	−0.04	(−0.05)	0.49	(0.51)	0.04	(0.06)	0.41	(0.52)	−0.36	(−0.36)
Philippines	−0.32	(−0.35)	−0.19	(−0.30)	−0.14	(−0.14)	−0.08	(−0.09)	0.01	(0.02)	−0.09	(−0.09)
Russia	0.44	(0.36)	−1.65	(−1.20)	2.09	(1.21)	—	—	—	—	—	—
South Africa	2.00	(4.06)	1.62	(3.39)	0.37	(0.63)	2.01	(3.68)	1.65	(3.11)	0.35	(0.52)
Taiwan	−0.44	(−0.91)	0.55	(1.16)	−1.00	(−1.42)	−0.25	(−0.47)	0.57	(1.20)	−0.81	(−1.16)
Thailand	0.43	(0.63)	0.42	(0.77)	0.01	(0.01)	1.20	(1.79)	0.94	(1.62)	0.26	(0.36)
Turkey	−0.97	(−1.29)	0.86	(1.24)	−1.83	(−1.71)	−0.96	(−1.27)	0.86	(1.33)	−1.83	(−1.70)

Consistent with the extensive literature on momentum in the United States, factor models have very little effect on the profits of momentum strategies. For instance, the difference in 3-factor alphas between RR(WML) and IR(WML) momentum portfolios for the United States is −0.63% per month with a *t*-statistic of −1.74 (compared to −0.62% for simple portfolio return differences).

In aggregated portfolios, there are similarly small differences between returns and alphas. In each of the five sets of aggregated portfolios, there is no statistically significant difference in alphas between IR and RR portfolios, with two exceptions. The return difference is −0.60% (*t*-statistic = −2.06) for the value-weighted developed markets portfolio and −1.01% (*t*-statistic = −1.98) for the pooled Asia portfolio.

Table 3 shows that there are no statistically significant differences in alphas between IR and RR value-weighted portfolios in 36 of the 37 countries (Japan is the exception). Focusing again on major developed and emerging markets, there is no difference in alphas between IR and RR portfolios for Canada, France, Germany, the United Kingdom, Brazil, China, or India.

As with portfolio returns, Japan shows some evidence of an echo, with a difference in 3-factor alphas of -0.77% (t-statistic $= -1.99$). Given its large market capitalization, Japan has a large weight in developed markets and Asian portfolios. Indeed, if we remove Japan from the sample of developed and Asian countries, there is no difference in future returns of IR versus RR portfolios in the aforementioned two exceptions. Specifically, the return difference between IR and RR portfolios becomes -0.15% (t-statistic $= -0.60$) for value-weighted developed markets portfolio and 0.02% (t-statistic $= 0.03$) for the pooled Asia portfolio. Our reading is that generally, IR portfolios do not out- (or under-) perform RR portfolios even on a risk-adjusted basis.

B. Double-Sorted Portfolios

Another way to analyze the incremental effects of IR versus RR on momentum is via double-sorted portfolios. We sort all stocks into winner and loser groups based on RR, and independently sort the same stocks into winner and loser groups based on IR. The intersection of these two independent sorts informs us about the marginal contribution of IR versus RR. In other words, we can examine the effects of IRs on momentum portfolios while holding RR (roughly) constant, and vice versa.

This approach is not without difficulties in implementation, especially in excluding U.S. data. Given the paucity of stocks in certain countries, we cannot use double-sorted quintiles in many countries; the portfolios would simply have too few securities to have any economic meaning. To alleviate this problem, we sort stocks into terciles rather than quintiles.[6] These 3×3 sorts generate nine portfolios, on average containing 45 securities each per country. For the United States, we use quintiles instead of deciles. All other procedures for establishing breakpoints, assigning stocks to groups, and calculating portfolio returns are identical to those for single-sorted portfolios.

The following table illustrates the sort procedure and clarifies the exposition of our results. The top row shows IR terciles (IRL through IRW), where "L" and "W" denote loser and winner terciles, respectively. Analogously, the first column shows RR terciles (RRL through RRW). We compute the individual portfolio returns X1 through X9 and then calculate the return differences for the extreme portfolios. For example, we calculate IR(WML)/RRL (the WML portfolio formed from IR, for RR losers) as well as its mirror RR(WML)/IRL (the WML portfolio formed from RR, for IR losers). If momentum is driven by intermediate returns, we expect IR(WML)/RRL > RR(WML)/IRL, or equivalently, that IR(WML)/RRW is greater than its mirror RR(WML)/IRW.[7]

[6] The tercile approach is also used by Chui et al. (2010) in dealing with markets with fewer stocks.

[7] IR(WML)/RRL is computed as X3 − X1, and RR(WML)/IRL is computed as X7 − X1. The difference is therefore X3 − X7. Similarly IR(WML)/RRW is computed as X9 − X7 and RR(WML)/IRW is computed as X9 − X3. The difference is X3 − X7.

	IRL	IR2	IRW	IR(WML)
RRL	X1	X2	X3	IR(WML)/RRL
RR2	X4	X5	X6	X6 − X4
RRW	X7	X8	X9	IR(WML)/RRW
RR(WML)	RR(WML)/IRL	X8 − X2	RR(WML)/IRW	

To conduct inferences we compute differences in these return series (i.e., IR(WML)/RRW − RR(WML)/IRW) and use the time series of return differences to generate *t*-statistics. An added advantage is one of exposition; instead of looking at a large number of pairwise comparisons, we can simply examine average differences in returns/alphas. Table 4 shows these results.

TABLE 4

Returns and 3-Factor Alphas from Monthly Double Sorts on Recent (−2 to −6) and Intermediate (−7 to −12) Returns: International Sample (1980–2010)

In Table 4 we sort all stocks with valid data in a country independently into terciles based on prior recent months (−2 to −6, RR) or intermediate months (−7 to −12, IR) returns. We require at least 100 stocks per month to do the sorts. We calculate value-weighted returns of each of these nine (3 × 3) portfolios. RRL denotes recent return losers and RRW denotes recent return winners; IRL denotes intermediate return losers and IRW denotes intermediate return winners. Winner-minus-loser portfolios are denoted by WML. Then, RR(WML)/IRW denotes winner RR minus loser RR portfolio of stocks in the winner IR tercile portfolio. Similarly, IR(WML)/RRL denotes the winner intermediate return minus loser intermediate return portfolio of stocks in the winner recent return tercile portfolio. This table reports statistics on the difference in portfolio returns of RR(WML)/IRW and IR(WML)/RRW. Country portfolios are aggregated as follows: EW refers to equal weighting of country returns and VW refers to value weighting (using total country market capitalization of last month) of country returns. Pooled refers to pooling all securities within a region. Pooled ex-country also refers to pooling all securities within a region but we use returns in excess of local country returns for sorting. We require at least 1,000 stocks per month per region for pooled and pooled ex-country portfolios; the portfolios are value weighted. We report average monthly returns and monthly 3-factor alphas of these portfolios. The 3-factor alphas are based on time-series regressions of monthly returns on local country-specific market, SMB, and HML factors (Fama and French (1993)). We calculate alphas only if there are at least 60 months of return data. In each case, regional factors are constructed analogously to the aggregation method. All returns are in U.S. dollars and are reported in percentages per month (numbers in parentheses are their *t*-statistics). The overall sample period is 1980–2010. We follow the same procedure for the United States with the following exceptions. We use quintiles, return breakpoints are based only on New York Stock Exchange stocks, and we include only stocks with a share price greater than $1 at the end of the portfolio-formation period.

	Returns		3-Factor Alphas	
United States				
1927–2010	−0.48	(−2.25)	−0.36	(−1.73)
1980–2010	−0.75	(−2.48)	−0.85	(−2.86)
Developed Markets				
EW	−0.06	(−0.37)	−0.18	(−1.05)
VW	−0.42	(−2.12)	−0.58	(−2.77)
Pooled	−0.40	(−1.45)	−0.37	(−1.28)
Pooled ex-country	−0.40	(−1.82)	−0.42	(−1.80)
Emerging Markets				
EW	−0.17	(−0.73)	−0.03	(−0.14)
VW	−0.21	(−0.92)	−0.08	(−0.38)
Pooled	−0.53	(−1.21)	−0.34	(−0.79)
Pooled ex-country	−0.05	(−0.16)	0.08	(0.23)
Americas Excluding the United States				
EW	0.50	(1.73)	0.40	(1.32)
VW	0.46	(1.48)	0.45	(1.01)
Pooled	−0.37	(−0.92)	−0.86	(−1.16)
Pooled ex-country	0.05	(0.13)	−0.56	(−0.82)
Asia				
EW	−0.18	(−0.79)	−0.18	(−0.78)
VW	−0.41	(−1.70)	−0.53	(−2.13)
Pooled	−0.64	(−1.78)	−0.74	(−2.03)
Pooled ex-country	−0.23	(−0.75)	−0.16	(−0.51)
Europe				
EW	−0.32	(−1.82)	−0.33	(−1.79)
VW	−0.38	(−1.87)	−0.29	(−1.40)
Pooled	−0.48	(−2.01)	−0.43	(−1.74)
Pooled ex-country	−0.27	(−1.37)	−0.21	(−0.98)

In the United States between 1980 and 2010, RR portfolios underperform IR portfolios by 0.75% per month with a *t*-statistic of 2.48. The difference in 3-factor alphas is 0.85% (t-statistic = 2.86). In developed markets, inferences are sensitive to specifications. When countries are equal weighted or pooled, there is no statistically significant difference in double-sorted portfolio returns or alphas. If we value weight countries, it appears that IR portfolios deliver higher returns (0.42%) and alphas (0.58%) than do RR portfolios. However, this is driven by Japan. If we remove Japan from the sample of developed countries, there is no difference in future returns of IR versus RR portfolios in the value-weighted developed market portfolio (returns and 3-factor alphas for IR − RR portfolios are only 0.23% and 0.28%, respectively; numbers not reported in Table 4).

Japan's influence is also apparent in the results for Asia. In value-weighted and pooled specifications, portfolio returns and alphas of IR portfolios are higher than RR-based portfolios. When countries in Asia are equal weighted, or when the country effect is removed from the pooled portfolio, the return differences are smaller and no longer statistically significant. Once again, removing Japan yields no statistically significant difference in returns or 3-factor alphas for any aggregation method of Asian portfolios.

In the other regions (emerging markets, Americas excluding United States, and Europe), there is no strong evidence that IR portfolios have higher returns or alphas compared to RR portfolios. In country-specific portfolios in Table 5,

TABLE 5

Returns and 3-Factor Alphas from Monthly Double Sorts on Recent (−2 to −6) and Intermediate (−7 to −12) Returns: Individual Country Results (1980–2010)

In Table 5 we sort all stocks with valid data in a country independently into terciles based on prior recent months (−2 to −6, RR) or intermediate months (−7 to −12, IR) returns. We require at least 100 stocks per month to do the sorts. We calculate value-weighted returns of each of these nine (3 × 3) portfolios. RRL denotes recent return losers and RRW denotes recent return winners; IRL denotes intermediate return losers and IRW denotes intermediate return winners. Winner-minus-loser portfolios are denoted by WML. Then, RR(WML)/IRW denotes winner RR minus loser RR portfolio of stocks in the winner IR tercile portfolio. Similarly, IR(WML)/RRL denotes the winner IR minus loser IR portfolio of stocks in the winner RR tercile portfolio. This table reports statistics on the difference in portfolio returns of RR(WML)/IRW and IR(WML)/RRW. We report average monthly returns and monthly 3-factor alphas of these portfolios. The 3-factor alphas are based on time-series regressions of monthly returns on local country-specific market, SMB, and HML factors (Fama and French (1993)). We calculate alphas only if there are at least 60 months of return data. All returns are in U.S. dollars and are reported in percentages per month (numbers in parentheses are their t-statistics). The overall sample period is 1980–2010.

	Returns		3-Factor Alphas	
Developed Markets				
Australia	0.78	(1.79)	0.89	(2.04)
Austria	−1.47	(−1.05)	—	—
Belgium	−0.47	(−1.06)	−0.35	(−0.79)
Canada	0.42	(1.11)	0.35	(0.86)
Denmark	0.35	(0.78)	0.32	(0.71)
Finland	−0.53	(−0.77)	−0.61	(−0.86)
France	−0.45	(−1.24)	−0.41	(−1.12)
Germany	−0.60	(−1.52)	−0.60	(−1.48)
Greece	0.88	(1.21)	0.77	(1.07)
Hong Kong	0.38	(0.60)	0.21	(0.34)
Israel	1.10	(2.21)	0.46	(0.65)
Italy	−0.67	(−1.80)	−0.73	(−1.99)
Japan	−0.54	(−1.98)	−0.66	(−2.38)
Netherlands	0.01	(0.03)	0.03	(0.05)
New Zealand	−0.16	(−0.19)	−0.20	(−0.23)
Norway	−0.94	(−1.57)	−0.93	(−1.52)
Portugal	−0.10	(−0.11)	—	—
Singapore	−0.37	(−0.84)	−0.37	(−0.83)
Spain	−0.47	(−1.07)	−0.53	(−1.18)
Sweden	0.22	(0.27)	0.25	(0.31)
Switzerland	−1.07	(−2.85)	−0.90	(−2.29)
United Kingdom	−0.23	(−0.81)	−0.06	(−0.20)

(continued on next page)

TABLE 5 (continued)

Returns and 3-Factor Alphas from Monthly Double Sorts on Recent (−2 to −6) and Intermediate (−7 to −12) Returns: Individual Country Results (1980–2010)

	Returns		3-Factor Alphas	
Emerging Markets				
Brazil	−0.66	(−0.90)	−0.56	(−0.75)
Chile	1.14	(2.19)	0.79	(1.73)
China	−0.51	(−1.33)	−0.80	(−2.05)
India	0.06	(0.13)	0.00	(−0.00)
Indonesia	−0.27	(−0.38)	0.18	(0.27)
Korea	−0.07	(−0.15)	−0.07	(−0.14)
Malaysia	0.21	(0.45)	0.34	(0.70)
Mexico	−0.41	(−0.68)	−0.67	(−1.08)
Peru	−0.03	(−0.04)	−0.17	(−0.19)
Philippines	−0.54	(−0.72)	−0.58	(−0.76)
Russia	1.11	(0.71)	—	—
South Africa	−0.02	(−0.04)	−0.14	(−0.25)
Taiwan	−0.40	(−0.75)	−0.09	(−0.15)
Thailand	−0.11	(−0.20)	0.21	(0.36)
Turkey	−1.71	(−2.06)	−1.81	(−2.16)

there are four countries (Japan, Switzerland, China, and Turkey) in which RR portfolios underperform IR portfolios significantly, and three countries (Australia, Israel, and Chile) in which the opposite appears to be true. In the vast majority of countries, the differences are not statistically significant. In countries with large market capitalizations, there are no differences in returns or alphas in Canada, France, Germany, the United Kingdom, Brazil, and India.

C. Fama–MacBeth (1973) Regressions

Another way to determine whether RR or IR are more important for momentum is to estimate regressions of future returns on past returns, while controlling for stock characteristics. We do so by estimating the following monthly cross-sectional Fama–MacBeth (1973) regressions:

$$(1) \quad R_{it} - R_{ft} = \gamma_{0t} + \gamma_{1t}\ln(\mathrm{ME}_{it-1}) + \gamma_{2t}\ln(\mathrm{BM}_{it-1}) + \gamma_{3t}R_{it-1} + \gamma_{4t}R_{it-2:t-6} + \gamma_{5t}R_{it-7:t-12} + \varepsilon_{i,t}.$$

The independent variables include the logarithm of market value of equity in the prior month, the logarithm of the book-to-market ratio in the prior year, the last month's return (month −1), the RR (from months −6 to −2), and the IR (from months −12 to −7). For each cross section, we require 100 stocks in a country-month to estimate the regression. Table 6 shows the average γ coefficients for the return variables (t-statistics generated using the Newey–West (1987) procedure with 3 lags are reported in parentheses) for each region. In addition to the coefficients, we report the difference in the slopes for IR and RR ($\gamma_4 - \gamma_5$) and its t-statistic. The slopes on size and book-to-market ratios follow patterns reported in other papers and are not reported.

The 1-month return reversal (γ_3) is strong across all regions. Consistent with the results in Tables 1 and 2, the coefficients on RR and IR show the importance of momentum, with weaker effects for emerging markets and Asia because of the influence of Japan. Our primary interest is in the difference in the IR and

TABLE 6

Monthly Fama–MacBeth (1973) Regressions: International Sample (1980–2010)

In Table 6 we estimate the following Fama–MacBeth (1973) regressions each month for each country/region:

$$R_{it} - R_{ft} = \gamma_{0t} + \gamma_{1t}\ln(\text{ME}_{it-1}) + \gamma_{2t}\ln(\text{BM}_{it-1}) + \gamma_{3t}R_{it-1} + \gamma_{4t}R_{it-2:t-6} + \gamma_{5t}R_{it-7:t-12} + \varepsilon_{i,t}.$$

ME is market capitalization, BM is book-to-market, and $R_{t-\tau}$ is the return of prior-month τ. We estimate these regressions for each country each month if the number of stocks in a month is greater than 100. Country-specific coefficients are aggregated as follows: EW refers to equal weighting of country coefficients and VW refers to value weighting (using total country market capitalization of last month) of country coefficients. In pooled regressions, we include all securities in a region before estimating regressions. In pooled ex-country, we also include all securities in a region but use the stock return minus the local country return as the dependent variable. We require at least 1,000 stocks per month in pooled regressions. All returns and independent variables are in U.S. dollars. Coefficients on ME and BM are not reported. The difference column shows the differences in coefficients on recent ($R_{t-2:t-6}$) and intermediate ($R_{t-7:t-12}$) returns. All coefficients are multiplied by 100 and t-statistics, reported in parentheses, are based on the Newey–West (1987) procedure with 3 lags. The overall sample period is 1980–2010.

	R_{t-1}		$R_{t-2:t-6}$		$R_{t-7:t-12}$		Difference	
United States								
1927–2010	−6.31	(−16.40)	0.81	(3.63)	1.08	(6.57)	−0.27	(−1.15)
1980–2010	−2.95	(−7.82)	0.85	(3.30)	0.86	(4.52)	−0.02	(−0.07)
Developed Markets								
EW	−4.20	(−9.40)	1.13	(3.77)	1.08	(4.07)	0.05	(0.18)
VW	−5.04	(−10.14)	0.26	(0.78)	0.76	(2.66)	−0.50	(−1.51)
Pooled	−1.88	(−2.99)	0.90	(2.63)	0.94	(3.05)	−0.04	(−0.11)
Pooled ex-country	−3.27	(−5.96)	0.85	(2.75)	0.78	(2.79)	0.07	(0.18)
Emerging Markets								
EW	−4.18	(−6.32)	−0.25	(−0.42)	−0.17	(−0.44)	−0.09	(−0.15)
VW	−3.56	(−5.62)	0.09	(0.23)	0.01	(0.03)	0.08	(0.17)
Pooled	−1.57	(−1.92)	−0.47	(−0.72)	0.33	(0.70)	−0.80	(−1.38)
Pooled ex-country	−2.53	(−3.35)	−0.67	(−1.46)	0.23	(0.58)	−0.90	(−2.09)
Americas Excluding the United States								
EW	−1.46	(−2.00)	1.88	(4.45)	0.77	(2.25)	1.11	(2.52)
VW	−1.59	(−2.27)	1.80	(4.32)	0.85	(2.49)	0.96	(2.25)
Pooled	−0.60	(−0.62)	0.49	(0.91)	0.18	(0.38)	0.30	(0.66)
Pooled ex-country	−0.40	(−0.44)	0.60	(1.08)	0.20	(0.44)	0.40	(0.85)
Asia								
EW	−5.97	(−9.48)	−0.88	(−1.88)	0.16	(0.46)	−1.04	(−2.15)
VW	−6.25	(−8.88)	−0.87	(−2.13)	0.26	(0.77)	−1.14	(−2.56)
Pooled	−3.41	(−3.44)	−1.17	(−1.87)	0.45	(1.01)	−1.62	(−2.53)
Pooled ex-country	−4.72	(−4.89)	−0.94	(−1.97)	0.06	(0.15)	−1.00	(−2.02)
Europe								
EW	−4.13	(−7.51)	1.44	(4.20)	1.37	(5.03)	0.07	(0.23)
VW	−3.52	(−6.29)	1.44	(4.17)	1.30	(4.94)	0.14	(0.47)
Pooled	−1.24	(−2.01)	1.68	(4.22)	1.12	(3.35)	0.56	(1.57)
Pooled ex-country	−1.59	(−2.96)	1.69	(4.54)	0.99	(3.25)	0.70	(2.08)

RR slopes. In our sample period, in the United States this difference is only 0.02 with a t-statistic of 0.07 (the difference is slightly larger at 0.27, albeit with a t-statistic of 1.15, in the full sample period of 1927–2010). By way of comparison, Novy-Marx (2012) reports a coefficient of 0.25 with a t-statistic of 0.69 for the 1990–2010 sample period.

In developed markets, the differences in slopes are small, except for the value-weighted specification. Similarly, in emerging markets, there is no difference in slopes of IR and RR in equal-weighted, value-weighted, or pooled portfolios. In the pooled ex-country emerging markets portfolios, the slope on IR is reliably higher than on RR.

Portfolios for other geographic regions show similar such diversity. In Americas excluding United States, the coefficients for RR are statistically and significantly higher than those for IR in equal-weighted and value-weighted specifications, but when securities are pooled, there is no difference. In contrast, in Europe, there is no difference in slopes for equal-weighted, value-weighted, and pooled portfolios, but slopes for IR are higher than for RR when the country return is removed (pooled ex-country). In Asia, the slopes for RR are always higher than IR, but that is once again due to Japan. If we remove Japan, we see no statistically significant differences in slopes for RR and IR. Overall, outside of Asia, we see no systematic evidence of slope differences in IR and RR.

D. Robustness

In all of the results thus far, there are differences between equal- and value-weighted portfolios, particularly for the United States. Recognizing that momentum effects are especially concentrated in small stocks, Novy-Marx (2012) also pays special attention to large stocks. He finds that the echo in returns for the United States is in fact stronger in large stocks. Because size effects can also be important in markets outside the United States (Fama and French (2012)), we redo our basic tests for large stocks only. We identify large stocks as those above the 75th percentile in market capitalization for each country. We then calculate returns for portfolios based on IR and RR and independent double-sorted portfolios, and reestimate the Fama–MacBeth (1973) regressions. Even in large stocks, we see no robust evidence that IR are statistically better or worse at predicting future returns than RR. These results are available from the authors.

The majority of our results excluding the United States are based on sorting stocks into quintiles (for single sorts) or terciles (for double sorts). Given the large number of firms in the United States (and following convention), we also use deciles for single sorts in the United States. To ensure that our results excluding the United States are not an artifact of the sorting procedure, we also use decile sorts for countries with an adequate number of securities, namely, Australia, Canada, France, Germany, Japan, India, and Korea. In addition, we use decile sorts for developed markets, emerging markets, and regional portfolios. For France, Germany, India, and Korea, there is no difference in the returns or alphas for IR versus RR portfolios. In Japan, IR portfolio alphas are higher than RR portfolio alphas by 0.90% per month with a t-statistic of 1.86 (the difference in alphas from quintile sorts in Table 3 was 0.77% with a t-statistic of 1.99). Interestingly, in Australia and Canada, IR portfolio alphas are lower than RR portfolio alphas by 1.29% (t-statistic $= 1.90$) and 1.58% (t-statistic $= 2.26$), respectively, compared to 0.86% (t-statistic $= 1.71$) and 0.41% (t-statistic $= 0.80$) from Table 3. Thus, sorting into deciles leaves the results for Japan largely the same but emphasizes the importance of RR for Australia and Canada.

IV. Explanations

The data thus far indicate that there is no robust evidence of an echo in returns outside the United States in various regions or in the 37 individual countries. In the United States, however, IR predict future monthly returns better than does RR. The obvious question then remains: What explains the echo in the United States? As discussed in the Introduction, we do not believe that cross-country differences attributed to individualism-collectivism can account for the results. Several countries that have the highest score on individualism indices (most notably, Australia, Canada, and the United Kingdom) show no evidence of an echo. Similarly, other conjectured variables, such as those describing variation in information uncertainty, level of development, or liquidity, face the same hurdle: In countries where they take on values or ranges similar to the United States, there is no evidence that IR portfolios outperform RR portfolios. One could also consider alternative explanations for momentum such as differences in growth opportunities (Berk et al. (1999), Sagi and Seasholes (2007)). But they too would have to fundamentally vary across countries and somehow link to IR versus RR. This seems implausible.

To reconcile the differences between the United States and non-U.S. countries, we conduct an in-depth examination of the full term structure of returns in the United States. We take this approach because theory provides no guidance; no extant models of momentum offer advice on what should be considered intermediate versus recent.

A. Full Term Structure for the United States

Because IR and RR are computed by compounding prior monthly returns, they obfuscate individual prior months' effects. Modifying Jegadeesh (1990) slightly, a simple way to see the influence of each prior month's return on future return is to estimate monthly cross-sectional regressions of the form:

$$(2)\quad R_{it} - R_{ft} \quad = \quad \gamma_{0t} + \sum_{k=1}^{12} \gamma_{kt} R_{it-k} + \gamma_{13t} \ln(\mathrm{ME}_{it-1}) + \gamma_{14t} \ln(\mathrm{BM}_{it-1}) + \varepsilon_{i,t},$$

where R_{it} is the monthly return for stock i in month t. The modification to the original Jegadeesh specification is the inclusion of size and book-to-market ratios as regressors. The γs have a straightforward interpretation as return responses at various lags k and can be interpreted as excess returns on zero-cost hedge portfolios.

Panel A of Table 7 reports the full term structure regressions in equation (2) for the full sample period from 1927 to 2010. In addition to using simple stock returns (labeled "raw returns"), we also use 3-factor-adjusted returns in which loadings are estimated over the entire sample period.

The pattern of coefficients offers important insights into what drives the difference between IR and RR portfolio returns in the United States. Not surprisingly, and consistent with the well-known short-term reversal effect, the coefficient on the prior month ($k = 1$) is large and negative. Because the prior month is skipped in the formation of all momentum portfolios, this has no effect on the IR versus RR results and we ignore it in the rest of our discussion. In general, the remaining coefficients are positive, indicating momentum. But they show significant variation. Unlike all the other prior returns, the coefficient on $k = 2$ is negative. The t-statistics of the coefficient are -1.26 using raw returns and -2.29 using 3-factor adjusted returns. More importantly, the economic magnitude of the $k = 2$ coefficients stands in stark contrast to ks $= 3$ through 12, which are all reliably positive. Thus, the data show that some of the short-term reversal effect typically associated with the prior month's return ($k = 1$) slips into the next month

TABLE 7

Average Coefficients of Cross-Sectional Regressions of Monthly Returns on Prior 12 Monthly Returns: U.S. Sample (1927–2010)

In Table 7 we estimate the following cross-sectional regressions:

$$R_{it} - R_{ft} - \hat{\beta}_i F_t \quad = \quad \gamma_{0t} + \sum_{k=1}^{12} \gamma_{k,t} R_{i,t-k} + \gamma_{13t} \ln(\text{ME}_{it-1}) + \gamma_{14t} \ln(\text{BM}_{it-1}) + \varepsilon_{i,t}.$$

ME is market capitalization, BM is book-to-market, and $R_{t-\tau}$ is the return of prior-month τ. For raw returns, the dependent variable is the monthly stock return in excess of the risk-free rate. For 3-factor adjusted returns, the dependent variable is the stock return, minus loading on the market, SMB, and HML factors (Fama and French (1993)), multiplied by their respective factor realizations. Loadings are estimated over the entire sample period. The regressions are estimated for each month between Jan. 1927 and Dec. 2010. This table shows average γ coefficients. All coefficients are multiplied by 100, and t-statistics, reported in parentheses, are based on the Newey–West (1987) procedure with 3 lags.

	R_{t-1}	R_{t-2}	R_{t-3}	R_{t-4}	R_{t-5}	R_{t-6}	R_{t-7}	R_{t-8}	R_{t-9}	R_{t-10}	R_{t-11}	R_{t-12}
Panel A. Full Sample												
1927–2010												
Raw	−6.78	−0.34	1.64	0.97	1.27	1.83	1.27	0.63	1.40	1.08	1.41	1.88
	(−17.47)	(−1.26)	(6.09)	(3.71)	(4.95)	(6.79)	(5.54)	(2.74)	(6.34)	(4.55)	(6.28)	(7.73)
3-factor	−7.27	−0.55	1.53	0.94	1.20	1.77	1.14	0.58	1.25	1.07	1.24	1.67
	(−19.75)	(−2.29)	(6.57)	(4.47)	(5.97)	(7.93)	(6.02)	(3.03)	(6.63)	(5.26)	(6.38)	(7.98)
Panel B. Subperiods												
1927–1947												
Raw	−10.96	−2.29	0.62	−0.53	1.07	1.44	1.34	0.42	1.64	1.28	2.01	1.79
	(−11.45)	(−3.21)	(0.77)	(−0.69)	(1.49)	(1.77)	(1.93)	(0.63)	(2.68)	(1.89)	(3.06)	(2.41)
3-factor	−11.51	−2.40	0.81	−0.32	1.24	1.60	1.37	0.39	1.22	1.59	1.54	1.66
	(−12.74)	(−4.01)	(1.24)	(−0.58)	(2.25)	(2.43)	(2.43)	(0.74)	(2.31)	(2.89)	(2.80)	(2.56)
1948–1968												
Raw	−6.42	0.88	2.93	2.16	2.07	2.93	1.21	−0.16	1.03	1.17	1.50	2.58
	(−11.41)	(1.83)	(6.01)	(4.93)	(4.39)	(6.84)	(3.18)	(−0.39)	(2.69)	(2.63)	(3.39)	(6.00)
3-factor	−7.04	0.71	2.44	2.05	1.85	2.54	1.21	0.01	0.88	1.14	1.38	2.19
	(−13.31)	(1.78)	(5.38)	(5.87)	(5.23)	(6.68)	(3.75)	(0.02)	(2.54)	(2.93)	(3.67)	(6.12)
1969–1989												
Raw	−3.41	−0.26	0.78	0.87	0.55	0.78	0.68	0.63	0.93	0.67	0.62	1.06
	(−7.64)	(−1.27)	(4.28)	(4.53)	(2.56)	(3.98)	(3.73)	(3.36)	(4.70)	(3.89)	(4.45)	(6.03)
3-factor	−3.60	−0.35	0.70	0.74	0.44	0.64	0.69	0.77	0.97	0.58	0.56	0.79
	(−8.23)	(−2.02)	(4.74)	(4.67)	(2.77)	(4.03)	(4.63)	(4.74)	(5.63)	(3.90)	(4.05)	(5.53)
1990–2010												
Raw	−0.84	0.15	0.41	0.11	0.23	0.40	0.36	0.32	0.32	0.15	0.26	0.28
	(−4.96)	(1.11)	(3.44)	(1.03)	(2.04)	(2.91)	(3.23)	(3.02)	(2.78)	(1.26)	(2.75)	(2.71)
3-factor	−1.12	−0.00	0.33	0.14	0.19	0.41	0.20	0.15	0.19	0.08	0.17	0.20
	(−6.11)	(−0.04)	(3.07)	(1.39)	(2.25)	(4.39)	(2.67)	(1.93)	(2.44)	(0.82)	(2.16)	(2.58)

($k = 2$) as well. The second insight that emerges from these regressions is that the largest positive coefficient is for the prior 12-month return ($k = 12$). This 12-month effect, first documented in Jegadeesh (1990), shows that the excess returns of momentum portfolios benefit considerably from the "kick" generated by the month -12 return.

Panel B of Table 7 shows similar regressions for four equal subperiods. The negative coefficient on $k = 2$ appears particularly strong in the first and third subperiods (1927–1947 and 1969–1989). In the other two subperiods (1948–1968 and 1990–2010), this coefficient is not reliably different from 0, but the contrast to positive coefficients for ks $= 3$ through 12 is still stark. The high return response associated with the coefficient on $k = 12$ (the 12-month effect) is strong in the first three subperiods. In the last subperiod (1990–2010), there is a noticeable decline in the $k = 12$ coefficient. These results closely match and help explain the Fama–MacBeth (1973) results reported by Novy-Marx (2012). He finds that the slope difference between IR and RR drops to 0.25 in the last subperiod with a *t*-statistic of only 0.69 (compared to 0.58 for the overall sample with a *t*-statistic of 2.47). The decline in the 12-month effect in the last subperiod accounts for some of this degradation in performance between IR and RR strategies.

B. Definitions of IR and RR

The data thus suggest that at least some of the difference in performance of IR and RR strategies in the United States may be associated with the return response to months −12 and −2. The former is well known. Indeed, Novy-Marx (2012) shows that the 12-month effect does not drive his results (by excluding it in the construction of IR portfolios as a robustness check). The latter, however, is not well known and appears to be a carryover from short-term reversals typically thought of as occurring from the prior month.

As we point out earlier, theory does not help us determine what is recent or intermediate. Assuming one uses returns from months −12 to −2, there are in fact 55 ways IR or RR can be defined. Table 8 describes this comprehensive set of feasible definitions that can be compiled from the prior 11 months of returns.

The first two columns in Table 8 show the specific months used to identify IR and RR. The third column shows the difference between the number of months used in definitions of IR and RR. This table is sorted in ascending order by this difference; definitions that use more months in computation of IR appear at the top of the table. For the months listed in the IR and RR definition columns, we calculate the sums and averages of the coefficients from the full sample term structure regression (3) reported in Table 7. The table shows the difference in the sums and averages of these coefficients (IR − RR) with *t*-statistics in parentheses.

An example is useful in clarifying Table 8. Consider the first row in the table. The IR definition is based on months −12 through −3. The RR definition is based only on month −2. The sum of coefficients (from Table 7) from regressions using raw returns for months −12 through −3 is 13.38, and the coefficient for month −2 is −0.34 (these two numbers are not shown separately in the table). The difference in sums is 13.73 with a *t*-statistic of 11.18. Similarly, the average of coefficients from months −12 through −3 is 1.34, and the average coefficient for month −2 is −0.34 (again, these two numbers are not shown separately in the table). The difference in averages is shown in the table and is 1.68 (*t*-statistic = 6.83).

There are good economic reasons to consider both sums and averages of coefficients. The sum of the coefficients corresponds to the IR and RR coefficients in Novy-Marx's (2012) Fama–MacBeth (1973) regressions (Novy-Marx (2012), Tab. 1). The first (perhaps obvious) observation from Table 8 is that the difference in the sums of coefficients follows an almost monotonic pattern. There are large positive differences when the IR definition uses more months than the RR definition (at the top of the table) and large negative differences when the reverse is true (at the bottom of the table). This pattern is, of course, mechanical. Suppose that all individual month coefficients have the same impact on future returns so that they contribute equally to momentum. Under this null of no term structure, if an IR definition employs more months than a RR definition, the IR portfolios are expected to deliver higher future returns than are the RR portfolios.

TABLE 8

Differences in Sums and Averages of Coefficients from the Fama–MacBeth (1973) Return Regressions: U.S. Sample (1927–2010)

In Table 8, columns labeled IR and RR show the months contained in intermediate and recent return definitions. The adjacent column shows the differences in the number of months between IR and RR. For each definition, we calculate the sum and average of the coefficients from the Fama–MacBeth (1973) regressions provided in Table 7. Table 8 shows the differences in the sum (or averages) of these coefficients. Conventional *t*-statistics for a two-sided test of difference in coefficients appear in parentheses. The returns and conventional *t*-statistics are bootstrapped using the stepwise multiple testing method of Romano and Wolf (2005). *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

			Difference in Sum of Coefficients				Difference in Average of Coefficients			
		Months	Raw		3-Factor		Raw		3-Factor	
IR	RR	Diff.	Diff.	*t*-Stat.	Diff.	*t*-Stat.	Diff.	*t*-Stat.	Diff.	*t*-Stat.
[12, 3]	[2]	−9	13.73***	(11.18)***	12.90***	(12.78)***	1.68***	(6.83)***	1.78***	(8.25)***
[12, 4]	[2]	−8	12.09***	(11.02)***	11.37***	(12.48)***	1.65***	(6.53)***	1.75***	(7.90)***
[12, 4]	[3, 2]	−7	10.45***	(10.27)***	9.84***	(11.44)***	0.66**	(3.35)**	0.71***	(4.15)***
[12, 5]	[2]	−7	11.12***	(11.54)***	10.42***	(12.87)***	1.69***	(6.62)***	1.78***	(7.94)***
[12, 5]	[3, 2]	−6	9.48***	(10.41)***	8.89***	(11.41)***	0.70**	(3.45)***	0.74***	(4.20)***
[12, 6]	[2]	−6	9.85***	(11.78)***	9.21***	(12.87)***	1.70***	(6.64)***	1.78***	(7.93)***
[12, 5]	[4, 2]	−5	8.51***	(9.79)***	7.94***	(10.55)***	0.59**	(3.38)**	0.59***	(3.89)***
[12, 6]	[3, 2]	−5	8.21***	(10.22)***	7.68***	(11.01)***	0.71**	(3.45)***	0.75***	(4.17)***
[12, 7]	[2]	−5	8.02***	(11.05)***	7.46***	(11.98)***	1.62***	(6.22)***	1.70***	(7.42)***
[12, 6]	[4, 2]	−4	7.24***	(9.10)***	6.74***	(9.71)***	0.60**	(3.32)**	0.59***	(3.82)***
[12, 7]	[3, 2]	−4	6.38***	(8.87)***	5.92***	(9.55)***	0.63**	(2.98)**	0.66***	(3.61)***
[12, 8]	[2]	−4	6.75***	(10.49)***	6.32***	(11.41)***	1.62***	(6.09)***	1.70***	(7.28)***
[12, 6]	[5, 2]	−3	5.97***	(7.47)***	5.53***	(8.02)***	0.47**	(2.92)**	0.45**	(3.34)**
[12, 7]	[4, 2]	−3	5.41***	(7.28)***	4.98***	(7.81)***	0.52**	(2.77)**	0.51**	(3.17)**
[12, 8]	[3, 2]	−3	5.11***	(7.67)***	4.79***	(8.37)***	0.63**	(2.86)**	0.66***	(3.50)***
[12, 9]	[2]	−3	6.11***	(11.10)***	5.75***	(11.91)***	1.79***	(6.57)***	1.85***	(7.70)***
[12, 7]	[5, 2]	−2	4.14***	(5.29)***	3.77***	(5.74)***	0.39*	(2.30)*	0.37*	(2.59)**
[12, 8]	[4, 2]	−2	4.14***	(5.77)***	3.84***	(6.32)***	0.52**	(2.64)**	0.51**	(3.05)**
[12, 9]	[3, 2]	−2	4.47***	(7.32)***	4.22***	(8.00)***	0.79***	(3.44)**	0.81***	(4.09)***
[12, 10]	[2]	−2	4.72***	(9.78)***	4.50***	(10.65)***	1.80***	(6.37)***	1.86***	(7.47)***
[12, 7]	[6, 2]	−1	2.31*	(2.73)**	2.01*	(2.89)**	0.21	(1.27)	0.17	(1.31)
[12, 8]	[5, 2]	−1	2.86**	(3.69)***	2.63**	(4.09)***	0.40*	(2.18)*	0.37*	(2.47)**
[12, 9]	[4, 2]	−1	3.50**	(5.04)***	3.27***	(5.57)***	0.69**	(3.25)**	0.66***	(3.71)***
[12, 10]	[3, 2]	−1	3.08**	(5.42)***	2.96***	(6.04)***	0.81***	(3.33)**	0.82***	(3.93)***
[12, 11]	[2]	−1	3.63***	(8.81)***	3.43***	(9.47)***	1.99***	(6.60)***	1.99***	(7.54)***
[12, 8]	[6, 2]	0	1.04	(1.19)	0.87	(1.24)	0.21	(1.19)	0.18	(1.24)
[12, 9]	[5, 2]	0	2.23*	(2.87)**	2.06*	(3.24)**	0.56**	(2.87)**	0.52**	(3.24)**
[12, 10]	[4, 2]	0	2.11	(3.13)**	2.02*	(3.55)***	0.70**	(3.13)**	0.67***	(3.55)***
[12, 11]	[3, 2]	0	1.99	(3.78)***	1.90*	(4.22)***	1.00***	(3.78)***	0.95***	(4.22)***
[12]	[2]	0	2.23*	(6.79)***	2.20**	(7.56)***	2.23***	(6.79)***	2.20***	(7.56)***
[12, 8]	[7, 2]	1	−0.24	(−0.25)	−0.26	(−0.34)	0.17	(1.05)	0.15	(1.10)
[12, 9]	[6, 2]	1	0.41	(0.45)	0.31	(0.43)	0.37*	(1.97)*	0.32*	(2.12)*
[12, 10]	[5, 2]	1	0.84	(1.08)	0.81	(1.28)	0.57**	(2.75)**	0.53**	(3.08)**
[12, 11]	[4, 2]	1	1.02	(1.57)	0.95	(1.76)	0.89***	(3.63)***	0.80***	(3.88)***
[12]	[3, 2]	1	0.59	(1.24)	0.67	(1.66)	1.23***	(4.21)***	1.17***	(4.58)***
[12, 9]	[7, 2]	2	−0.87	(−0.87)	−0.83	(−1.04)	0.34*	(1.85)	0.29*	(2.01)*
[12, 10]	[6, 2]	2	−0.99	(−1.10)	−0.95	(−1.31)	0.38*	(1.90)*	0.34*	(2.04)*
[12, 11]	[5, 2]	2	−0.25	(−0.32)	−0.25	(−0.41)	0.76***	(3.31)**	0.66***	(3.47)***
[12]	[4, 2]	2	−0.39	(−0.62)	−0.27	(−0.53)	1.13***	(4.04)***	1.01***	(4.25)***
[12, 9]	[8, 2]	3	−1.50	(−1.38)	−1.40	(−1.62)	0.40*	(2.33)*	0.36*	(2.55)**
[12, 10]	[7, 2]	3	−2.27	(−2.24)	−2.09	(−2.56)	0.35*	(1.78)	0.31*	(1.92)*
[12, 11]	[6, 2]	3	−2.07	(−2.29)	−2.01	(−2.79)	0.57**	(2.55)**	0.46**	(2.52)**
[12]	[5, 2]	3	−1.66	(−2.19)	−1.48	(−2.42)	1.00***	(3.74)***	0.87***	(3.86)***
[12, 10]	[8, 2]	4	−2.90	(−2.59)	−2.65	(−2.96)	0.42*	(2.20)*	0.37*	(2.38)**
[12, 11]	[7, 2]	4	−3.35	(−3.28)	−3.15	(−3.86)	0.54**	(2.46)*	0.44**	(2.42)**
[12]	[6, 2]	4	−3.48	(−3.88)	−3.24	(−4.51)	0.81***	(3.10)**	0.68***	(3.08)**
[12, 10]	[9, 2]	5	−4.29	(−3.62)	−3.91	(−4.07)	0.37*	(2.07)*	0.33*	(2.24)*
[12, 11]	[8, 2]	5	−3.98	(−3.50)	−3.72	(−4.10)	0.61**	(2.83)**	0.50**	(2.81)**
[12]	[7, 2]	5	−4.76	(−4.67)	−4.38	(−5.37)	0.77***	(3.03)**	0.65***	(3.01)**
[12, 11]	[9, 2]	6	−5.38	(−4.43)	−4.97	(−5.08)	0.56**	(2.72)**	0.46**	(2.68)**
[12]	[8, 2]	6	−5.39	(−4.72)	−4.95	(−5.43)	0.84***	(3.33)**	0.72***	(3.33)**
[12, 11]	[10, 2]	7	−6.46	(−5.07)	−6.03	(−5.87)	0.56**	(2.84)**	0.45**	(2.72)**
[12]	[9, 2]	7	−6.79	(−5.55)	−6.20	(−6.27)	0.80***	(3.23)**	0.68***	(3.20)**
[12]	[10, 2]	8	−7.87	(−6.08)	−7.26	(−6.94)	0.80***	(3.31)**	0.67***	(3.22)**
[12]	[11, 2]	9	−9.28***	(−6.90)***	−8.49***	(−7.75)***	0.77***	(3.31)**	0.64***	(3.21)**

Difference in sums of coefficients would, thus, indicate an echo for definitions at the top of Table 8 and a reverse echo for definitions at the bottom of the table. These mechanical effects do not reflect a true echo. One way around this issue is to consider the differences in averages of coefficients, rather than their sums. The averages control for differences in number of months and are shown in the last column of the raw and 3-factor adjusted return portion of the table. Unsurprisingly, the monotonicity disappears. However, it is notable that averages of coefficients for all IR definitions are still larger than those for RR definitions. In the 55 combinations displayed in the table, there are still 49 to 51 definitions (depending on the specification) in which the average coefficients for IR months are larger than those for RR months at the 95% significance level.

C. Bootstrapping Correlated Strategies

The above comparisons are complicated by the fact that the definitions are correlated. This necessitates that statistical inference of differences in coefficients account for dependencies between rows. White's (2000) reality check bootstrap recognizes dependencies between different rules to construct trading strategies but does so in the context of the best trading strategy. Specifically, White's approach asks whether the best of a multitude of strategies beats its benchmark, after controlling for dependencies between all possible strategies. Our interest is not in the best strategy/definition, but in a specific definition (the Novy-Marx (2012) specification) and more generally with all possible IR and RR definitions. Romano and Wolf (2005) generalize the intuition of White's reality check bootstrap and develop a method that allows one to assess the performance of any specific strategy, while recognizing dependencies among potential strategies. This is accomplished by controlling the familywise error rate (FWE), which is the probability of incorrectly identifying at least one strategy as beating the benchmark when, in truth, it does not.

The details of our implementation of the Romano and Wolf (2005) procedure are provided in the Appendix. Briefly, the procedure starts by defining a test statistic; in our case, the difference in the sum or average of coefficients of IR and RR definitions or their studentized versions, the *t*-statistic of those coefficient differences.[8] Definitions are then organized in descending order using this test statistic and critical values for the statistic of each definition are constructed by bootstrapping the joint confidence interval. White (2000) and Romano and Wolf recommend a bootstrap procedure that is applicable to time-series data; accordingly, we use Politis and Romano's (1994) stationary bootstrap. The one-sided null hypothesis is that the test statistic is less than 0. Because the statistics are ordered, the procedure is continued until no further hypotheses are rejected. This allows one to reject (or fail to reject) the null hypothesis for each definition.

We implement the procedure for significance at the 10%, 5%, and 1% levels. In Table 8, next to each conventional *t*-statistic, we insert asterisks to indicate whether the difference in sum/average of coefficients rejects the one-sided null

[8]There are power advantages to using studentized versions of the test statistic. See Romano and Wolf ((2005), App. C).

hypothesis of less than 0 based on a 10% (*), 5% (**), or 1% (***) level of bootstrapped significance.[9]

If one considers the sum of the coefficients equivalent to the Novy-Marx (2012) specification (the row with IR $[12, 7]$ and RR $[6, 2]$ in Table 8), the difference in sums of coefficients is statistically significant at the 5% level even when assessing statistical significance using the bootstrap. If we adjust for the mechanical effect of unequal number of months in IR and RR strategies by using averages instead of sums, the difference is no longer statistically significant, even at the 10% level. This is true regardless of whether assessing statistical significance using conventional t-statistics or the Romano–Wolf (2005) bootstrap. Thus, it seems that Novy-Marx's (2012) effect derives from the fact that his IR strategy uses 7 months of portfolio-formation returns while the RR strategy uses only 6 months.

At the same time, Table 8 shows that there are many definitions for which even the difference of average IR and RR coefficients is positive and statistically significant. For instance, looking at IR constructed from months -12 to -6, and RR constructed from months -5 to -2, the difference in average coefficients using raw returns is 0.47, which is significant at the 5% level using the bootstrap. Indeed, as we noted before, differences in the averages of coefficients for IR and RR are positive for all 55 definitions in Table 8. It seems that even though Novy-Marx's (2012) definition of IR and RR does not lead to statistically significant evidence of an echo, there are many other definitions of IR and RR that do provide this evidence, even after accounting for data-snooping biases.

However, there is an obvious and noticeable commonality in all these definitions. Every RR definition includes month -2, which as we show in Table 7 has a return response coefficient that is either negative or indistinguishable from 0. Symmetrically, every IR definition includes month -12, which we know has a return response coefficient that is large and positive. This means that averages of coefficients from RR are pushed down by the inclusion of month -2, and averages of coefficients from RR are pushed up by the inclusion of month -12.

We formally investigate the influence of these 2 months in Table 9. This table is constructed and formatted the same as Table 8 but excludes month -2 in averaging RR coefficients. There are 45 possible ways IR or RR can be defined using months -3 to -12. We still include month -12 in computing averages of IR coefficients, deliberately biasing inferences in favor of IR definitions. Excluding month -2, virtually none of the IR definitions have higher average coefficients than those of RR definitions, using either conventional t-statistics or bootstrapped levels of significance. A prime example is when IR are constructed using months -12 to -8, and RR use months -7 to -3 (the row with IR $[12, 8]$ and RR $[7, 3]$ in the table). Despite the fact that the IR coefficients include month -12, the differences in the averages of coefficients are -0.12 using raw returns and -0.16 using 3-factor adjusted returns. The corresponding conventional t-statistics are only -0.68 and -1.18, and the bootstrap fails to reject the null hypothesis of equality. The few exceptions in Table 9, where difference in averages of coefficients are

[9]Because the null hypothesis is one-sided, negative differences of coefficients in Table 8 do not show bootstrapped statistical significance (in other words, our implementation of the bootstrap does not show statistical significance for the reverse echo).

TABLE 9

Differences in Sums and Averages of Coefficients from the Fama–MacBeth (1973) Return Regressions Excluding Month −2: U.S. Sample (1927–2010)

In Table 9, columns labeled IR and RR show the months contained in intermediate and recent return strategies. The adjacent column shows the differences in the number of months between IR and RR strategies. For each strategy, we calculate the sum and average of the coefficients from the Fama–MacBeth (1973) regressions provided in Table 7. Table 9 shows the differences in the sum (or averages) of these coefficients, calculated as IR minus RR. Conventional *t*-statistics for a two-sided test of whether IR coefficients are greater than RR coefficients appear in parentheses. The returns and conventional *t*-statistics are bootstrapped using the stepwise multiple testing method of Romano and Wolf (2005). *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

			Difference in Sum of Coefficients				Difference in Average of Coefficients			
		Months	Raw		3-Factor		Raw		3-Factor	
IR	RR	Diff.	Diff.	*t*-Stat.	Diff.	*t*-Stat.	Diff.	*t*-Stat.	Diff.	*t*-Stat.
[12, 4]	[3, 3]	−8	10.10***	(9.61)***	9.29***	(10.62)***	−0.34	(−1.46)	−0.33	(−1.63)
[12, 5]	[3, 3]	−7	9.13***	(9.85)***	8.34***	(10.70)***	−0.29	(−1.23)	−0.30	(−1.43)
[12, 5]	[4, 3]	−6	8.16***	(9.44)***	7.40***	(10.03)***	0.04	(0.22)	0.00	(−0.03)
[12, 6]	[3, 3]	−6	7.86***	(9.74)***	7.14***	(10.36)***	−0.28	(−1.16)	−0.29	(−1.39)
[12, 6]	[4, 3]	−5	6.89***	(8.89)***	6.19***	(9.28)***	0.05	(0.26)	0.00	(−0.01)
[12, 7]	[3, 3]	−5	6.04***	(8.62)***	5.38***	(9.08)***	−0.36	(−1.45)	−0.38	(−1.78)
[12, 6]	[5, 3]	−4	5.62***	(7.37)***	4.98***	(7.64)***	0.06	(0.36)	0.01	(0.07)
[12, 7]	[4, 3]	−4	5.07***	(7.27)***	4.43***	(7.51)***	−0.03	(−0.12)	−0.09	(−0.50)
[12, 8]	[3, 3]	−4	4.76***	(7.62)***	4.24***	(8.04)***	−0.36	(−1.39)	−0.38	(−1.72)
[12, 7]	[5, 3]	−3	3.79***	(5.26)***	3.23***	(5.37)***	−0.01	(−0.08)	−0.08	(−0.51)
[12, 8]	[4, 3]	−3	3.79***	(5.83)***	3.29***	(6.05)***	−0.02	(−0.11)	−0.08	(−0.47)
[12, 9]	[3, 3]	−3	4.13***	(7.56)***	3.67***	(7.94)***	−0.20	(−0.73)	−0.23	(−1.01)
[12, 7]	[6, 3]	−2	1.97	(2.55)**	1.47	(2.35)*	−0.15	(−0.87)	−0.21	(−1.54)
[12, 8]	[5, 3]	−2	2.52**	(3.61)***	2.09**	(3.64)***	−0.01	(−0.07)	−0.07	(−0.47)
[12, 9]	[4, 3]	−2	3.16**	(5.20)***	2.72***	(5.39)***	0.14	(0.59)	0.06	(0.32)
[12, 10]	[3, 3]	−2	2.73**	(5.71)***	2.42**	(5.98)***	−0.18	(−0.65)	−0.21	(−0.90)
[12, 8]	[6, 3]	−1	0.69	(0.89)	0.33	(0.53)	−0.15	(−0.80)	−0.21	(−1.42)
[12, 9]	[5, 3]	−1	1.89	(2.76)**	1.52	(2.75)**	0.15	(0.71)	0.07	(0.44)
[12, 10]	[4, 3]	−1	1.76	(3.12)**	1.47	(3.12)**	0.15	(0.63)	0.08	(0.38)
[12, 11]	[3, 3]	−1	1.65	(4.06)***	1.36	(3.93)***	0.01	(0.02)	−0.09	(−0.35)
[12, 8]	[7, 3]	0	−0.58	(−0.68)	−0.81	(−1.18)	−0.12	(−0.68)	−0.16	(−1.18)
[12, 9]	[6, 3]	0	0.06	(0.08)	−0.24	(−0.39)	0.02	(0.08)	−0.06	(−0.39)
[12, 10]	[5, 3]	0	0.49	(0.74)	0.27	(0.50)	0.16	(0.74)	0.09	(0.50)
[12, 11]	[4, 3]	0	0.68	(1.30)	0.41	(0.94)	0.34	(1.30)	0.20	(0.94)
[12]	[3, 3]	0	0.24	(0.75)	0.13	(0.46)	0.24	(0.75)	0.13	(0.46)
[12, 9]	[7, 3]	1	−1.21	(−1.36)	−1.38	(−1.95)	0.05	(0.24)	−0.02	(−0.10)
[12, 10]	[6, 3]	1	−1.34	(−1.70)	−1.49	(−2.40)	0.03	(0.14)	−0.04	(−0.26)
[12, 11]	[5, 3]	1	−0.59	(−0.92)	−0.80	(−1.55)	0.35	(1.46)	0.22	(1.11)
[12]	[4, 3]	1	−0.73	(−1.52)	−0.82	(−2.07)	0.58*	(1.96)	0.42	(1.70)
[12, 9]	[8, 3]	2	−1.84	(−1.87)	−1.95	(−2.52)	0.17	(0.97)	0.11	(0.77)
[12, 10]	[7, 3]	2	−2.61	(−2.90)	−2.63	(−3.68)	0.06	(0.30)	0.00	(0.00)
[12, 11]	[6, 3]	2	−2.42	(−3.10)	−2.56	(−4.17)	0.22	(0.94)	0.08	(0.44)
[12]	[5, 3]	2	−2.00	(−3.21)	−2.03	(−4.07)	0.59*	(2.12)	0.43	(1.87)
[12, 10]	[8, 3]	3	−3.24	(−3.22)	−3.20	(−4.01)	0.19	(0.96)	0.13	(0.79)
[12, 11]	[7, 3]	3	−3.69	(−4.11)	−3.69	(−5.18)	0.25	(1.11)	0.13	(0.70)
[12]	[6, 3]	3	−3.83	(−4.99)	−3.78	(−6.24)	0.45	(1.70)	0.30	(1.34)
[12, 10]	[9, 3]	4	−4.64	(−4.33)	−4.45	(−5.17)	0.17	(0.93)	0.12	(0.77)
[12, 11]	[8, 3]	4	−4.32	(−4.24)	−4.26	(−5.29)	0.38	(1.72)	0.25	(1.41)
[12]	[7, 3]	4	−5.10	(−5.72)	−4.92	(−6.95)	0.48	(1.86)	0.34	(1.57)
[12, 11]	[9, 3]	5	−5.72	(−5.23)	−5.52	(−6.28)	0.36	(1.71)	0.24	(1.41)
[12]	[8, 3]	5	−5.73	(−5.63)	−5.49	(−6.81)	0.61*	(2.39)*	0.47	(2.17)
[12, 11]	[10, 3]	6	−6.80	(−5.90)	−6.58	(−7.08)	0.38	(1.93)	0.26	(1.57)
[12]	[9, 3]	6	−7.13	(−6.48)	−6.74	(−7.60)	0.59*	(2.38)*	0.46	(2.16)
[12]	[10, 3]	7	−8.21	(−7.01)	−7.81	(−8.24)	0.62*	(2.56)*	0.48	(2.30)*
[12]	[11, 3]	8	−9.62***	(−7.86)***	−9.04***	(−9.05)***	0.60*	(2.60)**	0.47	(2.35)*

statistically significant, all share the common theme of including month −12 in IR definitions and, therefore, benefit from the 12-month effect.

D. Portfolio Strategies and Reconciling the Evidence

The above analysis of regression coefficients strongly suggests that the superior performance of IR strategies in the United States is likely driven by the underperformance of RR strategies, which is really due to the inclusion of month -2 in RR portfolios (the inclusion of month -12 in IR portfolios also has a role to play). If this is indeed the case, then two additional tests are informative: one for the U.S. evidence and a second for the importance of months -2 and -12, in general.

First, if one constructs RR strategies excluding month -2, the difference in future portfolio returns between IR and RR portfolios should not be different from 0. To check this, we reconstruct the single- and double-sorted portfolios in Tables 2 and 4 but define months -12 to -8 as intermediate, and months -7 to -3 as recent. Excluding month -2 also means that both IR and RR strategies now contain 5 months in the portfolio-formation period. An added advantage of portfolio analysis over averaging regression coefficients is that portfolios readily deliver economic significance, which is not straightforward to infer from differences in averages of regression coefficients.

Excluding month -2, in single-sorted portfolios, the future return difference between value-weighted IR and RR portfolios is 0.34% per month with a t-statistic of 1.40. The difference in 3-factor alphas is only 0.23% (t-statistic $= 0.98$). In double-sorted portfolios, the difference in returns of value-weighted portfolios is 0.37% with a t-statistic of 1.91 and the difference in 3-factor alphas is 0.23% with a t-statistic of 1.20. Clearly, even with the experiment being biased in favor of IR strategies, there is no statistically significant difference in returns between IR and RR portfolios, implying that the influence of month -2 is economically large.

Second, one could estimate the Fama–MacBeth (1973) regressions for non-U.S. markets but with return lags that mitigate the negative effect of month -2 on RR and the strong positive effect of month -12 on IR. Accordingly, we estimate the following three additional specifications for each region:

$$(3)\quad \begin{aligned} R_{it} - R_{ft} &= \gamma_{0t} + \gamma_{1t}\ln(\text{ME}_{it-1}) + \gamma_{2t}\ln(\text{BM}_{it-1}) + \gamma_{3t}R_{it-1} \\ &\quad + \gamma_{4t}R_{it-2:t-6} + \gamma_{5t}R_{it-7:t-11} + \gamma_{6t}R_{it-12} + \varepsilon_{it}, \\ R_{it} - R_{ft} &= \gamma_{0t} + \gamma_{1t}\ln(\text{ME}_{it-1}) + \gamma_{2t}\ln(\text{BM}_{it-1}) + \gamma_{3t}R_{it-1} \\ &\quad + \gamma_{4t}R_{it-2} + \gamma_{5t}R_{it-3:t-6} + \gamma_{6t}R_{it-7:t-12} + \varepsilon_{it}, \\ R_{it} - R_{ft} &= \gamma_{0t} + \gamma_{1t}\ln(\text{ME}_{it-1}) + \gamma_{2t}\ln(\text{BM}_{it-1}) + \gamma_{3t}R_{it-1} \\ &\quad + \gamma_{4t}R_{it-2} + \gamma_{5t}R_{it-3:t-6} + \gamma_{6t}R_{it-7:t-11} + \gamma_{7t}R_{it-12} + \varepsilon_{it}. \end{aligned}$$

Each regression contains a slightly different definition of IR and RR. Rather than inundate the reader with each of the individual slopes, we report only the differences in slopes of interest $(R_{t-2:t-6} - R_{t-7:t-11})$, $(R_{t-3:t-6} - R_{t-7:t-12})$, and $(R_{t-3:t-6} - R_{t-7:t-11})$, along with their t-statistics in Table 10.

The results are informative. The first pair, $(R_{t-2:t-6} - R_{t-7:t-11})$, takes out the month -12 return, thereby biasing inferences in favor of RR. Table 10 shows that in Americas excluding United States, RR slopes are larger than IR slopes. In developed markets, emerging markets, and Europe, the differences are, in general, statistically insignificant. In Asia, the evidence in favor of IR is also weaker (excluding Japan weakens the results even further). The second pair, $(R_{t-3:t-6} - R_{t-7:t-12})$, takes out the month -2 return, biasing in favor of IR. There is some

TABLE 10

Differences in Coefficients from the Fama–MacBeth (1973) Regressions: International Sample (1980–2010)

In Table 10 we estimate the following three versions of the Fama–MacBeth (1973) regressions each month for each country/region:

$$R_{it} = \gamma_{0t} + \gamma_{1t}\ln(\text{ME}_{it-1}) + \gamma_{2t}\ln(\text{BM}_{it-1}) + \gamma_{3t}R_{it-1} + \gamma_{4t}R_{it-2:t-6} + \gamma_{5t}R_{it-7:t-11} + \gamma_{6t}R_{it-12} + \varepsilon_{it},$$

$$R_{it} = \gamma_{0t} + \gamma_{1t}\ln(\text{ME}_{it-1}) + \gamma_{2t}\ln(\text{BM}_{it-1}) + \gamma_{3t}R_{it-1} + \gamma_{4t}R_{it-2} + \gamma_{5t}R_{it-3:t-6} + \gamma_{6t}R_{it-7:t-12} + \varepsilon_{it},$$

$$R_{it} = \gamma_{0t} + \gamma_{1t}\ln(\text{ME}_{it-1}) + \gamma_{2t}\ln(\text{BM}_{it-1}) + \gamma_{3t}R_{it-1} + \gamma_{4t}R_{it-2} + \gamma_{5t}R_{it-3:t-6} + \gamma_{6t}R_{it-7:t-11} + \gamma_{7t}R_{it-12} + \varepsilon_{it}.$$

ME is market capitalization, BM is book-to-market, and $R_{t-\tau}$ is the return of prior month τ. We estimate these regressions for each country each month if the number of stocks in a month is greater than 100. Country-specific coefficients are aggregated as follows: EW refers to equal weighting of country coefficients and VW refers to value weighting (using total country market capitalization of last month) of country coefficients. In pooled regressions, we include all securities in a region before estimating regressions. In pooled ex-country, we also include all securities in a region but use the stock return minus the country's return as the dependent variable. We require at least 1,000 stocks per month in pooled regressions. Coefficients on ME and BM are not reported. The difference column shows the differences in coefficients on recent and intermediate returns, which are the following pairs corresponding to each of the three models: $(R_{t-2:t-6}, R_{t-7:t-11})$, $(R_{t-3:t-6}, R_{t-7:t-12})$, and $(R_{t-3:t-6}, R_{t-7:t-11})$. All returns and independent variables are in U.S. dollars. All coefficients are multiplied by 100, and *t*-statistics, reported in parentheses, are based on the Newey–West (1987) procedure with 3 lags. The overall sample period is 1980–2010.

	$R_{t-2:t-6} - R_{t-7:t-11}$		$R_{t-3:t-6} - R_{t-7:t-12}$		$(R_{t-3:t-6} - R_{t-7:t-11})$	
Developed Markets						
EW	0.29	(1.06)	0.31	(1.13)	0.55	(2.08)
VW	−0.22	(−0.69)	−0.13	(−0.42)	0.14	(0.45)
Pooled	0.03	(0.09)	0.17	(0.47)	0.24	(0.68)
Pooled ex-country	0.30	(0.94)	0.27	(0.84)	0.50	(1.59)
Emerging Markets						
EW	−0.04	(−0.07)	0.35	(0.69)	0.40	(0.77)
VW	0.06	(0.13)	0.52	(1.27)	0.49	(1.16)
Pooled	−0.99	(−1.62)	−0.49	(−0.88)	−0.66	(−1.11)
Pooled ex-country	−1.04	(−2.18)	−0.72	(−1.53)	−0.81	(−1.61)
Americas Excluding the United States						
EW	1.17	(2.64)	1.16	(2.42)	1.21	(2.50)
VW	0.91	(2.16)	0.91	(1.98)	0.87	(1.90)
Pooled	0.17	(0.38)	0.23	(0.47)	0.11	(0.24)
Pooled ex-country	0.31	(0.69)	0.36	(0.73)	0.29	(0.61)
Asia						
EW	−0.83	(−1.72)	−0.60	(−1.32)	−0.39	(−0.85)
VW	−0.88	(−2.00)	−0.61	(−1.48)	−0.36	(−0.90)
Pooled	−1.53	(−2.45)	−1.23	(−2.04)	−1.18	(−1.99)
Pooled ex-country	−0.94	(−1.88)	−0.65	(−1.35)	−0.58	(−1.18)
Europe						
EW	0.44	(1.46)	0.47	(1.59)	0.83	(2.83)
VW	0.53	(1.73)	0.54	(1.84)	0.93	(3.11)
Pooled	0.66	(1.86)	0.79	(2.25)	0.89	(2.54)
Pooled ex-country	0.87	(2.67)	0.90	(2.63)	1.07	(3.20)

evidence that RR slopes are higher than IR slopes in Americas excluding United States (the difference is statistically significant for equal-weighted and value-weighted portfolios) and in Europe (the difference is statistically significant for pooled and pooled ex-country portfolios). There is some weak evidence to the contrary in Asia (IR slopes are statistically and significantly higher than RR slopes in pooled portfolios), but excluding Japan (results not reported) eliminates the difference. There is no robust evidence of differences in slopes for these definitions of RR and IR in developed and emerging markets portfolios. Finally, the last pair, $(R_{t-3:t-6} - R_{t-7:t-11})$, allows us to examine slope differences after removing the effects of both months -2 and -12. Most of the slope differences are positive but with modest *t*-statistics, with two notable exceptions. In Asia, the slope differences are negative (but with *t*-statistics below 2), again driven by the influence of Japan. In Europe, the slope differences are positive, indicating that RR are more informative for future returns, with large t-statistics. Overall, removing the effects of months -2 and -12 in portfolios excluding the United States shows no evidence of an echo.

V. Conclusions

Novy-Marx (2012) provides startling evidence that in the United States, momentum portfolios formed on the basis of returns from 12 to 7 months prior to portfolio formation have substantially higher profits than momentum portfolios formed based on returns from 6 to 2 months prior. These results, reminiscent of an echo, appear remarkably robust, present in different subperiods, in different groups of stocks and industries, and in size and book-to-market portfolios. Similar effects appear to exist in currencies, commodities, and international equity indices. What does one make of this? It is hard to imagine a story that could generate such an effect on prices; even the age-old "relative strength" trading strategies of Wall-Street lore have nothing to say about such an effect. For financial economists, the challenge to theory is enormous. No existing theory, whether behavioral or rational, predicts an echo in returns.

We conduct out-of-sample and in-sample tests to assess whether this predictability is genuine. We look for the same echo in regions and in 37 individual countries. We do not find convincing evidence of an echo outside the United States. Within the United States, an examination of the full term structure of returns suggests that the superior performance of strategies based on IR is driven by the presence of reversal (or at least no continuation) from month -2 returns. The true puzzle is not why IR forecast future returns better than RR, but why return reversals from month -1 also extend (somewhat) to month -2.

Appendix. Romano–Wolf (2005) Procedure

Denote the returns on M strategies by the matrix $R_{t,m}$, where $t = 1, \ldots, T$, and $m = 1, \ldots, M$. The "returns" in our setting are differences of sums or averages of the Fama–MacBeth (1973) coefficients on intermediate returns (IR) and recent returns (RR).[10] Consider average return, $\bar{R}_m = \sum_{t=1}^{T} R_{t,m} \big/ T$, as the basic test statistic. This statistic tests a univariate parameter θ_m. This parameter is defined in such a way that $\theta_m \leq 0$ under the null hypothesis (e.g., IR do not beat RR under strategy m using sums of coefficients as in Table 8). Thus, for a given strategy m, the individual testing problem is:

$$H_m\colon \theta_m \leq 0 \quad \text{versus} \quad H_m'\colon \theta_m > 0.$$

A multiple testing method yields a decision concerning each testing problem by either rejecting H_m or not. A commonly used method for controlling for the familywise error rate (FWE) is the Bonferroni method. The Bonferroni method, however, is conservative and results in low power because it does not account for the dependence structure in correlated strategies. White (2000) provides the first solution to this problem using bootstrap reality

[10]Because we use estimated parameters in the construction of our test statistic, it is not obvious whether these parameters have to be reestimated in each simulation draw. However, White (2000) and Romano and Wolf (2005) show that it is not necessary to reestimate the parameters for first-order consistency.

check (BRC). The BRC estimates the asymptotic distribution of $\max_{1 \leq m \leq M} \bar{R}_m$ accounting for the dependence structure of the individual test statistics. It addresses the question of whether the strategy that appears best in the observed data really beats the benchmark. However, it does not attempt to identify specific strategies as beating the benchmark or as many outperforming strategies as possible. The solution to this problem is proposed by Romano and Wolf (2005). The algorithm to conduct the test is described below.

0) Relabel the strategies according to the size of the individual test statistics (in this case, the average return), from largest to smallest. Thus, $\bar{R}_1 \geq \bar{R}_2 \geq \ldots \geq \bar{R}_M$.

1) Construct a rectangular joint confidence region for the vector $(\theta_1, \ldots, \theta_M)'$ with nominal joint coverage probability $1 - \alpha$. The confidence region is of the form $[\bar{R}_1 - c_1, \infty) \times \ldots \times [\bar{R}_M - c_1, \infty)$, with a common critical value c_1.

   a) Estimate the critical value c_1 using bootstrap (procedure described below).

   b) Reject all the null hypotheses H_m, where $\bar{R}_m$ is bigger than c_1.

   c) If no null hypothesis is rejected, stop. Otherwise, count the number M_1 of strategies where the null is rejected in this step.

2) $M - M_1$ strategies remain. Repeat step 1. Construct a rectangular joint confidence region for the vector $(\theta_{M_1+1}, \ldots, \theta_M)'$ with nominal joint coverage probability $1 - \alpha$. The confidence region is of the form $[\bar{R}_{M_1+1} - c_2, \infty) \times \ldots \times [\bar{R}_M - c_2, \infty)$, with a common critical value c_2.

   a) Estimate the critical value c_2 using bootstrap (procedure described below).

   b) Reject all the null hypotheses H_m, where $\bar{R}_m$ is bigger than c_2.

   c) If no null hypothesis is rejected, stop. Otherwise, count the number M_2 of strategies where the null is rejected in this step.

3) Repeat the stepwise process until no further hypotheses are rejected.

The critical values in the above steps are computed using the Politis and Romano (1994) stationary bootstrap with smoothing parameter of 0.2 to generate $B = 10{,}000$ bootstrap data matrices $R^{*b}_{t,m}$, where the asterisk denotes simulated data and b denotes the bth bootstrap iteration. For each bootstrap iteration b, we compute the test statistics, $\bar{R}^{*b}_1, \ldots, \bar{R}^{*b}_M$ and compute $\max^{*b}_j = \max_{m \in \{\text{Remaining strategies}\}} \left(\bar{R}^{*b}_m - \bar{R}_m\right)$ for the jth step described in the previous paragraph. Finally, c_j is computed as the $1 - \alpha$ empirical percentile of the B values $\max^{*1}_j, \ldots, \max^{*B}_j$.

The algorithm is extended for a studentized test statistic defined as $z_m = \bar{R}_m / \hat{\sigma}_m$, where $\hat{\sigma}_m$ is the standard error of $\bar{R}_m$. For example, null hypothesis m in step j is rejected if $\bar{R}_m - \hat{\sigma}_m \times d_j > 0$, where d_j is the critical value in the jth step. We modify the computation of critical values as follows. For each bootstrap iteration b, we compute the test statistics $\bar{R}^{*b}_1, \ldots, \bar{R}^{*b}_M$ and the corresponding standard errors $\hat{\sigma}^{*b}_1, \ldots, \hat{\sigma}^{*b}_M$. Then, we compute the maximum test statistic $\max^{*b}_j = \max_{m \in \{\text{Remaining strategies}\}} (\bar{R}^{*b}_m - \bar{R}_m) / \hat{\sigma}^{*b}_m$ for the jth step. Finally, d_j is computed as the $1 - \alpha$ empirical percentile of the B values $\max^{*1}_j, \ldots, \max^{*B}_j$. The main advantage of a studentized test statistic is that of improved power (see Romano and Wolf (2005), App. C).

Finally, we can use alphas and their z-statistics, in place of returns and their z-statistics, as the objects of interest. As mentioned before, White's (2000) procedure can be viewed as the first step of the above algorithm.[11] By continuing after the first step in the Romano and Wolf (2005), more false hypotheses can be rejected.

[11] Another approach to multiple testing is that of false discovery rate (FDR). For example, Barras, Scaillet, and Wermers (2010) use this approach to identify the fraction of skilled mutual fund managers. FDR is defined as an error rate, or more precisely the expected value of the fraction of falsely rejected null hypotheses, that the researcher is willing to tolerate. However, our main interest is in precisely identifying which null hypotheses are rejected, which is not possible with FDR methods.

References

Asness, C. S.; T. J. Moskowitz; and L. H. Pedersen. "Value and Momentum Everywhere." *Journal of Finance*, 68 (2013), 929–985.

Barberis, N.; A. Shleifer; and R. Vishny. "A Model of Investor Sentiment." *Journal of Financial Economics*, 49 (1998), 307–343.

Barras, L.; O. Scaillet; and R. Wermers. "False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas." *Journal of Finance*, 55 (2010), 179–216.

Berk, J.; R. Green; and V. Naik. "Optimal Investment, Growth Options and Security Returns." *Journal of Finance*, 54 (1999), 1553–1608.

Campbell, J. Y.; S. J. Grossman; and J. Wang. "Trading Volume and Serial Correlation in Stock Returns." *Quarterly Journal of Economics*, 108 (1993), 905–939.

Chui, A. C. W.; S. Titman; and K. C. J. Wei. "Individualism and Momentum around the World." *Journal of Finance*, 65 (2010), 361–392.

Daniel, K.; D. Hirshleifer; and A. Subrahmanyam. "Investor Psychology and Security Market Under- and Overreactions." *Journal of Finance*, 53 (1998), 1839–1885.

Davis, J. L.; E. F. Fama; and K. R. French. "Characteristics, Covariances, and Average Returns: 1929 to 1997." *Journal of Finance*, 55 (2000), 389–406.

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

Fama, E. F., and K. R. French. "Value versus Growth: The International Evidence." *Journal of Finance*, 53 (1998), 1957–1999.

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

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

Griffin, J. M.; S. Ji; and S. Martin. "Momentum Investing and Business Cycle Risk: Evidence from Pole to Pole." *Journal of Finance*, 58 (2003), 2515–2547.

Griffin, J. M.; P. J. Kelly; and F. Nardari. "Do Market Efficiency Measures Yield Correct Inferences? A Comparison of Developed and Emerging Markets." *Review of Financial Studies*, 23 (2010), 3225–3277.

Heston, S. L., and R. Sadka. "Seasonality in the Cross-Section of Stock Returns." *Journal of Financial Economics*, 87 (2008), 418–445.

Heston, S. L., and R. Sadka. "Seasonality in the Cross Section of Stock Returns: The International Evidence." *Journal of Financial and Quantitative Analysis*, 45 (2010), 1133–1160.

Hou, K.; G. A. Karolyi; and B.-C. Kho. "What Factors Drive Global Stock Returns?" *Review of Financial Studies*, 24 (2011), 2527–2574.

Ince, O. S., and R. B. Porter. "Individual Equity Return Data from Thomson Datastream: Handle with Care!" *Journal of Financial Research*, 29 (2006), 463–479.

Jegadeesh, N. "Evidence of Predictable Behavior of Security Returns." *Journal of Finance*, 45 (1990), 881–898.

Jegadeesh, N., and S. Titman. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *Journal of Finance*, 48 (1993), 65–91.

Jegadeesh, N., and S. Titman. "Short-Horizon Return Reversals and the Bid–Ask Spread." *Journal of Financial Intermediation*, 4 (1995a), 116–132.

Jegadeesh, N., and S. Titman. "Overreaction, Delayed Reaction, and Contrarian Profits." *Review of Financial Studies*, 4 (1995b), 973–993.

Jegadeesh, N., and S. Titman. "Profitability of Momentum Strategies: An Evaluation of Alternative Explanations." *Journal of Finance*, 56 (2001), 699–720.

Johnson, T. C. "Rational Momentum Effects." *Journal of Finance*, 57 (2002), 585–608.

Leamer, E. E. *Specification Searches: Ad Hoc Inference with Non-Experimental Data*. New York: Wiley (1978).

Leamer, E. E. "Let's Take the Con Out of Econometrics." *American Economic Review*, 73 (1983), 31–43.

Lewellen, J. "Momentum and Autocorrelation in Stock Returns." *Review of Financial Studies*, 15 (2002), 533–563.

Lo, A. W., and A. C. MacKinlay. "Data-Snooping Biases in Tests of Financial Asset Pricing Models." *Review of Financial Studies*, 3 (1990), 431–468.

Mlodinow, L. *The Drunkard's Walk: How Randomness Rules Our Lives*. New York, NY: Pantheon Books (2008).

Moskowitz, T. J.; Y. H. Ooi; and L. H. Pedersen. "Time Series Momentum." *Journal of Financial Economics*, 104 (2012), 228–250.

Newey, W. K., and K. D. West. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, 55 (1987), 703–708.
Novy-Marx, R. "Is Momentum Really Momentum?" *Journal of Financial Economics*, 103 (2012), 429–453.
Politis, D., and J. P. Romano. "The Stationary Bootstrap." *Journal of the American Statistical Association*, 89 (1994), 1303–1313.
Romano, J. P., and M. Wolf. "Stepwise Multiple Testing as Formalized Data Snooping." *Econometrica*, 73 (2005), 1237–1282.
Rouwenhorst, G. K. "International Momentum Strategies." *Journal of Finance*, 53 (1998), 267–284.
Sagi, J. S., and M. S. Seasholes. "Firm-Specific Attributes and the Cross-Section of Momentum." *Journal of Financial Economics*, 84 (2007), 389–434.
White, H. L. "A Reality Check for Data Snooping." *Econometrica*, 68 (2000), 1097–1126.