Journal of Financial Economics 135 (2020) 213–230

Contents lists available at ScienceDirect

Journal of Financial Economics

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

Anomalies across the globe: Once public, no longer existent?☆

Heiko Jacobs[a,*], Sebastian Müller[b]

[a] University of Duisburg-Essen, Campus Essen, Essen 45117, Germany
[b] German Graduate School of Management and Law, Bildungscampus 2, Heilbronn 74076, Germany

ARTICLE INFO

Article history:
Received 23 January 2017
Revised 1 October 2018
Accepted 17 October 2018
Available online 14 June 2019

JEL classification:
G12
G14
G15
G40

Keywords:
Return predictability
International stock markets
Arbitrage
Publication impact
Anomalies

ABSTRACT

Motivated by McLean and Pontiff (2016), we study the pre- and post-publication return predictability of 241 cross-sectional anomalies in 39 stock markets. We find, based on more than two million anomaly country-months, that the United States is the only country with a reliable post-publication decline in long-short returns. Collectively, our meta-analysis of return predictors suggests that barriers to arbitrage trading can create segmented markets and that anomalies tend to represent mispricing instead of data mining.

1. Introduction

In an intriguing paper, McLean and Pontiff (2016) study 97 return predictors in the US stock market and find that long-short returns shrink significantly post-publication. Their results are consistent with the idea that arbitrageurs bet against mispricing shown in academic publications, which results in lower strategy returns. In this study, we explore post-publication effects of 241 anomalies in the US and 38 international stock markets.

International stock markets are economically important. Based on World Bank data, non-US countries account on average for about 58% of the world market capitalization and for almost 73% of global gross domestic product during our sample period from January 1980 to December 2015. International stock markets are also scientifically important. Existing asset pricing tests tend to focus on the US stock market, part of which leads (Karolyi, 2016, p. 2075) to conclude that "there is a large and persistent US (home) bias in academic research in Finance."

☆ This study was partly conducted while Heiko Jacobs was at the University of Mannheim. Financial support from Acatis (research award) is gratefully acknowledged. We thank Jeffrey Pontiff (the referee), Tim Adam, Itzhak Ben-David, Sreedhar Bharath, Thomas Conlon, Michael Fraikin, Maximilian Germann, Andreas Hoepner, Michael Kurz, Philippe Masset, Paulo Maio, Julien Penasse, Kalle Rinne, Laszlo Sandor, Philip Valta, Martin Weber, and seminar participants at the 44th Annual Meeting of the European Finance Association (Mannheim, Germany), the 20th Annual Conference of the Swiss Society for Financial Market Research (Zurich, Switzerland), the 24th Annual Meeting of the German Finance Association (Ulm, Germany), the Frontiers of Factor Investing Conference (Lancester, UK), the House of Energy Markets and Finance, Acatis, Deka Investments, Invesco Quantitative Strategies, Robeco, Aalto University, the German Graduate School of Management and Law, the University of Bern, University College Dublin, the University of Duisburg-Essen, the Humboldt University of Berlin, the University of Luxembourg, the University of Mannheim, and the Technical University of Munich for valuable comments.

* Corresponding author.
E-mail addresses: heiko.jacobs@uni-due.de (H. Jacobs), sebastian.mueller@ggs.de (S. Müller).

https://doi.org/10.1016/j.jfineco.2019.06.004

International out-of-sample tests can help to provide novel insights to enrich or challenge understanding of price formation.

Our analysis leads to three key conclusions. First, the US stock market is the only market with a statistically significant, economically meaningful, and robust post-publication decline in anomaly profitability. Second, the discrepancy between the US and international markets partly remains even after conditioning on easily tradable stocks and anomalies with similar in-sample profitability, suggesting that structural cross-country barriers to investment management create segmented markets. Third, our findings are consistent with the notion that abnormal anomaly returns are largely attributable to mispricing instead of to data mining.

We find surprisingly large differences between post-publication effects in the US stock market and international markets. Our baseline US estimates, both qualitatively and quantitatively, are in line with McLean and Pontiff (2016). For instance, for equally weighted (value-weighted) anomaly returns, we estimate a 38% (37%) post-sample and a highly significant 62% (66%) post-publication decline. The same empirical framework shows that none of the 38 international markets in our sample yields a reliable post-publication decline in anomaly returns. Post-publication differences are stable across anomaly universes, including a subset of return predictors not studied in McLean and Pontiff (2016). General time effects, differences in local risk factor exposures, or database issues can explain at best some of the return difference, leading us to conclude that a reliable post-publication decline exists in the US market only.

To better understand this difference, we condition on a subset of seemingly easily exploitable anomalies. We focus on return predictors that are constructed from large stocks only and/or that yield similarly high in-sample returns in both the US and the economically most important international markets. This matching procedure is intended to control for firm-level arbitrage costs and to make anomalies more comparable across countries. Consistent with intensified betting against exploitable mispricing post-publication, these specifications are the only ones that sometimes indicate a measurable post-publication drop also in major international markets. Nevertheless, most estimates still point to a meaningful difference between the magnitude of the drop in the US and global markets. We conclude that, with the partial exception of the largest stocks, there appear to be investment barriers that result in segmented markets.

Unconditionally, anomalies exist across the globe, and their magnitude in the US market is similar to their magnitude in major international markets. In-sample returns for the average return predictor are higher (and post-publication returns are lower) in the US than in global markets. At the same time, the vast majority of anomalies were originally found for the US market. The in-sample return difference could be attributable to data mining in US anomalies. Alternatively, it could at least partly reflect the asymmetric academic effort put into identifying variables that truly and reliably predict the cross section of returns in the US stock market versus non-US stock markets. As we are not able to empirically discriminate between these (and other) competing explanations, comparing in-sample returns arguably reveals little about data snooping.

To establish a proxy for the upper bound of data mining, we quantify the typical post-sample or post-publication anomaly return in major international markets. These returns are out-of-sample, both with respect to the time period and the stock universe, and are thus completely free of any data snooping. Relative to US in-sample returns, their magnitude could be smaller due the home bias in identifying reliable return predictors. Overall, we find that anomalies in global stock markets in the out-of-sample periods have about two-thirds of the size of the original US in-sample anomalies. This result suggests that the average anomaly is not merely a product of chance.

The out-of-sample existence of anomalies as well as the observation that international return predictability stays the same (or even goes up) after the end of the original sample also lends credibility to the idea that the post-sample and the post-publication drop in the US tend to be a consequence of arbitrage trading instead of data mining.[1] In this respect, our insights partly differ from McLean and Pontiff (2016), who argue that the US return decay relative to in-sample returns is the result of both data mining and arbitrage trading. They find that anomaly returns are 26% (58%) lower post-sample (post-publication). They argue that these estimates translate into an upper (lower) bound of 26% (58% − 26% = 32%) for the impact of statistical biases (arbitrage trading). Our evidence also contradicts Linnainmaa and Roberts (2018), who study 36 accounting-based US anomalies in periods both prior to and later than the original sample. The average capital asset pricing model (CAPM) alpha decreases by about 60% when moving either backward or forward in time, which Linnainmaa and Roberts (2018) attribute to data mining. Our findings are most consistent with Chen and Zimmermann (2018), who estimate that US bias-adjusted returns are only 12% smaller than in-sample returns.

In our data, further evidence that anomalies are real stems from the cross-sectional and time series variation in the magnitude of anomalies within markets. Data-snooping bias should not be influenced by arbitrage costs. We find, however, that anomalies disproportionately based on stocks with high arbitrage costs, such as high idiosyncratic risk, yield higher returns than anomalies with low limits to arbitrage. This finding is observable both in-sample and post-publication, and it is observable both in the US and in major international markets. Overall, these results are consistent with the behavioral finance view (e.g., Barberis and Thaler, 2003), which argues that anomalies can exist due to a combination of behavioral biases and impediments to arbitrage.

[1] Further support for this conjecture comes from the fact that the US post-sample and post-publication drop is also observable among value-weighted (instead of equally weighted) anomaly portfolio returns, which do not stand in the focus of most original studies. For instance, McLean and Pontiff (2016) report that 78 of the 79 published anomaly papers relied on in their analysis present equally weighted results as their primary finding.

Our analysis contributes to several streams of the literature. First, it extends McLean and Pontiff (2016) and supports their idea of price pressure due to arbitrage trading in anomalies. Existing work provides both supporting and contradicting evidence for this conjecture. For instance, (Israel and Moskowitz, 2013, p. 275) "find little evidence that size, value, and momentum returns are significantly affected by changes in trading costs or institutional and hedge fund ownership over time." In contrast, Hanson and Sunderam (2014, p. 1238) conclude that the "increase in capital has resulted in lower strategy returns" for momentum and value. On the one hand, Edelen et al. (2016) and Lewellen (2011) show that institutional investors do not bet against anomalies or even trade on the opposite side. On the other hand, Calluzzo et al. (2019) and Green et al. (2011) argue that sophisticated practitioners exploit US anomalies following publication.

Second, we contribute to the long-standing debate about anomaly data snooping (e.g., Lo and MacKinlay, 1990; Fama, 1998; Schwert, 2003). Recent work has produced conflicting findings. Chordia et al. (2017), Harvey (2017), Hou et al. (2019), and Linnainmaa and Roberts (2018) highlight the danger of widespread p-hacking in US return predictability. Other studies including Bartram and Grinblatt (2018a), Engelberg et al. (2018), Lu et al. (2018), Wahal (2019), and Yan and Zheng (2017) argue that anomalies are, at least to a large extent, real. Our results provide support for this view.

Third, our findings add to the large literature on international financial market segmentation. Asset pricing tests in Bekaert and Harvey (1995), Bekaert et al. (2011, 2014), Foerster and Karolyi (1999), Froot and Dabora (1999), Griffin (2002), Hau (2011), Rapach et al. (2013), and other papers often yield different results with respect to market integration. We provide evidence for seemingly strong geographic stock market segmentation, which appears to have significant effects on the formation of prices.

Fourth, we contribute to the literature on the meta-analysis of market anomalies. This work aims at developing a better understanding of when, where, and why anomalies tend to work (or not to work). Most of the literature focuses on the US market.[2] We add to Bartram and Grinblatt (2018b), Fama and French (2017), Griffin et al. (2010), Jacobs (2016), Jacobs and Müller (2018), Tobek and Hronec (2018a), and other emerging work that highlights the importance of an international perspective.

The rest of the paper is organized as follows. Section 2 describes our data, our empirical approach, and the unconditional profitability of return predictors across the globe. Section 3 reports the impact of anomaly publication on return predictability and explores potential reasons for the differences between the US and international markets. Section 4 discusses the role of limits to arbitrage. Section 5 concludes.

[2] Selected references include Chordia et al. (2014), Engelberg et al. (2018), Green et al. (2013, 2017), Hou et al. (2015, 2019), Harvey et al. (2016), Jacobs (2015), Keloharju et al. (2016), Lewellen (2015), Novy-Marx and Velikov (2016), Stambaugh et al. (2012, 2014, 2015), and Stambaugh and Yuan (2017).

2. Empirical approach and unconditional return predicability

2.1. Data

We obtain daily stock market data for the US from the Center for Research in Security Prices (CRSP) and for all other countries from Datastream. We gather accounting data in the case of the US market (all international markets) from Compustat (Worldscope). Finally, we collect analyst earnings forecasts and recommendations for all markets from Institutional Brokers' Estimate System (I/B/E/S).

We follow previous work in cleaning the Datastream data. The major screens are as follows. We require stocks to have non-missing identifier, return, and market capitalization data and to be traded in the home country of the firm. We use the generic industry and firm name screens proposed in Griffin et al. (2010) to exclude non-common equity. We identify delisted firms following the method proposed in Ince and Porter (2006) and by checking the Worldscope inactive date. In an attempt to eliminate remaining data errors, we screen returns as proposed in Hou et al. (2011). In addition, we winsorize return and market capitalization data at the 0.1% and the 99.9% level. To assure that our findings are not driven by the smallest and most illiquid stocks, we require stocks to have a one-month lagged market capitalization of at least $10 million.

The baseline sample period runs from January 1980 to December 2015. The start date aims at balancing the trade-off between maximizing the length of the time series and maximizing the number of implementable anomalies and available countries. Most international stock markets have limited stock market data in earlier years, and accounting data are generally not available before 1980. The baseline sample period assures that meaningful cross-country comparisons can be made.

2.2. Selecting and implementing anomalies

Our goal is to base our analysis on a reasonably representative universe of all cross-sectional anomalies published in the literature, provided that the return predictors can be implemented for (at least some) international markets. Some subjectivity is necessary in the selection of return predictors. We thus first implement a broad baseline data set following the criteria described below. We then verify that the qualitative nature of our main findings is robust to plausible modifications in the anomaly universe.

We consider only return predictors for which at least five valid estimates for in-sample, post-sample, and post-publication returns can be computed for (at least some) international markets. In-sample returns are defined as the returns during max(first month of the original anomaly sample period, January 1980) and the end of the original sample period. Post-sample returns are defined as the returns following the last month of the original sample period and preceding the month of the publication in a peer-reviewed journal. Post-publication returns start in the publication month. To have a clearly defined publication date, we do not consider anomalies from working papers.

We start by replicating 80 of the 97 anomalies relied on in McLean and Pontiff (2016). Some return predictors such as beta (Fama and MacBeth, 1973) or firm size (Banz, 1981) have to be excluded because they have in-sample periods that end before the start of our sample period. Other anomalies based on, for instance, corporate governance proxies or short interest cannot be computed due to a lack of data availability for international markets. A few further return predictors (such as dividend omissions) have missing or too few international observations during the original sample period.

To these 80 anomalies used in McLean and Pontiff (2016), we add a second set of 161 cross-sectional return phenomena. About two-thirds of these additional predictors are directly taken from other recent work on the meta-analysis of market anomalies. We assume that published and internationally implementable anomalies considered in Green et al. (2013, 2017), Hou et al. (2019), and Harvey et al. (2016) are, with a number of exceptions, suitable for our analysis.[3] Finally, we perform an additional literature review to identify further peer-reviewed papers proposing statistically significant cross-sectional anomalies not considered in the meta studies mentioned above.

The following additional selection rules are worth mentioning. By focusing on monthly returns, we neglect high-frequency trading strategies. Some of the identified predictors, such as changes in analyst recommendations (Jegadeesh et al., 2004), are not truly cross-sectional but still appear to be relevant for quantitative arbitrageurs. We follow McLean and Pontiff (2016) in including some return predictors that the original studies do not necessarily interpret as anomalies or mispricing, such as liquidity-based variables. Further following McLean and Pontiff (2016), we regard interactions of return predictors with other firm variables as distinct anomalies, provided that the original paper considers this interaction to be an important and novel finding. Examples are enhanced momentum strategies, which are based on interactions of past returns and stock characteristics, such as in Lee and Swaminathan (2000), Hong et al. (2000), and Da et al. (2014).

In sum, and as the Online Appendix shows in more detail, our baseline data set consists of 241 anomalies based on 161 papers, most of which have been published in the top finance or accounting journals. Publication years range between 1984 and 2015. The median (mean) publication year is 2005 (2004).

We do not intend to exactly replicate the original studies. Sometimes, this would be impossible due to, for instance, limited global data availability or data base changes even for the US over time. The original papers also differ in their methodologies [e.g., Fama and MacBeth, 1973 slope coefficients versus long-short portfolio returns; weighting schemes], timing conventions (e.g., yearly or quarterly accounting data), data screens (e.g., the treatment of small firms, certain industries, or tails of the distribution), control variables, and further dimensions (such as raw returns versus alphas relative to different factor models).

In this light, we aim at capturing the intent of the study under consideration while simultaneously forming a common framework for all anomalies. We do not use different investment universes or different methodologies for different anomalies. Instead, we always rely on all eligible firms with non-missing data to create country-neutral quintile-based long-short portfolios in each month. Separately for each (country, month, anomaly) combination, we compute the return of a portfolio that goes long (short) in the presumably 20% most underpriced (most overpriced) stocks. To assure diversification, we condition (with few exceptions for some firm event-based anomalies) on country-level anomaly months with at least 25 eligible firms. With respect to accounting data, we use yearly updated values and the conservative timing convention of Fama and French (1993) to avoid look-ahead bias and to assure comparability across countries. Close to 15% of the anomalies are based on binary indicators and thus do not qualify for a quintile-based sorting procedure. In these cases, we go long (or short) the event firms and offset the position with the portfolio of non-event firms.

This common basis is intended to reasonably reflect the real-life arbitrage process. The typical quantitative arbitrageur could be more likely to consistently apply his own data screens and investment guidelines instead of exactly following the many different approaches described in the respective academic papers. From an academic point of view, our framework is also motivated by Richardson et al. (2010) who criticize that “to date very few papers have made a serious attempt to bring some structure to the anomaly literature” (p. 422). Similarly, Subrahmanyam (2010, p. 28) highlights in his literature review on return predictors that “disparate methodologies are used by different researchers and there usually is little attempt to demonstrate robustness across methods. This is another reason why the picture remains murky and suggests a need for clarifying studies.”

Unless noted otherwise, returns are expressed in local currency and account for dividends as well as for capital actions. We consider both equally weighted and value-weighted portfolio returns, as both approaches have their merits. Value-weighting returns gives a better estimate of how economically important an anomaly is, as portfolio returns are dominated by larger firms. Weighting returns equally can reflect better how widespread a return phenomenon is. McLean and Pontiff (2016) follow the baseline approach used in the original anomaly studies, almost all of which rely on equally weighted returns.

2.3. Unconditional profitability of anomalies

Our initial stock market universe consists of all countries classified as developed or emerging stock markets by Morgan Stanley Capital International (MSCI) at least at some point during our sample period. In our baseline analysis, we use only the 39 stock markets that have at least 20 thousand eligible anomaly months, as defined below. This arbitrarily chosen cutoff ensures meaningful comparisons

[3] The most important exceptions are as follows. We do not include accounting anomalies based on quarterly (instead of yearly) data. We do not include a return predictor if it is not reliably statistically significant at least at the 5% level in the original study. We do not include an anomaly if it differs from another anomaly only in the length of the holding period.

Table 1
Anomalies on a country-by-country basis: descriptive statistics.
In column "MSCI group", DM, EM, and FM denote developed, emerging, and frontier markets as classified by Morgan Stanley Capital International (MSCI). "Percent macap" denotes the fraction of the total stock market capitalization in the sample. "Start year" is max[1980, first country year with non-missing anomaly data]. "N firms" quantifies the total number of stocks that meet the data screens outlined in Section 2.1. "Equally weighted" ("value-weighted") denotes the average unconditional monthly long-short anomaly portfolio return with equally weighted (value-weighted) stocks. *t*-statistics are reported in parentheses. Standard errors are clustered by month. Two-tailed statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Country	MSCI	N anomaly	*N*	Percent	Start	*N*	Equally weighted		Value-weighted	
	group	months	anomalies	macap	year	firms	Return	*t*-stat	Return	*t*-stat
Australia	DM	71,861	214	1.8	1980	2,504	0.735***	(11.51)	0.591***	(8.99)
Austria	DM	44,897	157	0.2	1986	166	0.408***	(5.83)	0.263***	(3.72)
Belgium	DM	54,025	174	0.6	1980	221	0.480***	(7.26)	0.303***	(3.94)
Brazil	EM	23,317	113	0.8	1994	246	0.429***	(3.87)	0.255**	(2.07)
Canada	DM	78,468	224	2.4	1980	2,857	0.555***	(7.66)	0.431***	(5.34)
Chile	EM	42,566	162	0.4	1989	251	0.317***	(4.95)	0.294***	(4.26)
China	EM	37,128	157	4.7	1992	2,814	0.215***	(3.18)	0.170**	(2.31)
Denmark	DM	61,678	197	0.4	1982	298	0.547***	(8.71)	0.462***	(6.20)
Finland	DM	42,903	162	0.4	1988	196	0.435***	(4.75)	0.348***	(2.83)
France	DM	81,970	227	3.9	1980	1,512	0.506***	(8.63)	0.341***	(5.85)
Germany	DM	80,274	224	3.4	1980	1,300	0.514***	(7.73)	0.410***	(6.25)
Greece	EM,DM	48,014	176	0.2	1988	394	0.462***	(4.35)	0.569***	(4.17)
Hong Kong	DM	54,950	182	1.3	1982	204	0.289***	(3.00)	0.270***	(3.02)
India	EM	46,975	182	1.7	1990	3,360	0.579***	(6.97)	0.428***	(4.10)
Indonesia	EM	46,319	175	0.4	1990	539	0.413***	(3.36)	0.392***	(2.74)
Ireland	DM	25,045	102	0.2	1987	98	0.487***	(3.94)	0.386***	(2.85)
Israel	EM,DM	33,540	133	0.3	1986	674	0.504***	(6.83)	0.448***	(4.63)
Italy	DM	69,272	210	1.4	1980	512	0.429***	(7.22)	0.293***	(4.72)
Japan	DM	87,644	237	12.5	1980	4,786	0.219***	(4.87)	0.188***	(3.82)
Malaysia	EM	66,948	207	0.7	1984	1,131	0.416***	(4.92)	0.345***	(4.37)
Mexico	EM	42,869	166	0.6	1989	219	0.418***	(4.85)	0.386***	(4.91)
Netherlands	DM	66,373	199	1.1	1980	254	0.556***	(7.84)	0.272***	(3.42)
New Zealand	DM	31,144	127	0.1	1988	234	0.626***	(8.95)	0.336***	(4.44)
Norway	DM	56,060	190	0.4	1982	399	0.523***	(5.78)	0.414***	(3.97)
Pakistan	EM,FM	33,737	144	0.1	1992	274	0.408***	(3.37)	0.461***	(4.10)
Philippines	EM	37,389	151	0.2	1990	253	0.344***	(2.65)	0.287**	(2.11)
Poland	EM	24,166	120	0.2	1995	697	0.528***	(6.48)	0.370***	(3.81)
Portugal	EM,DM	34,899	125	0.1	1988	137	0.533***	(5.93)	0.479***	(5.51)
Singapore	DM	62,825	196	0.6	1983	889	0.476***	(5.67)	0.359***	(4.48)
South Africa	EM	64,398	198	0.9	1980	758	0.727***	(12.95)	0.568***	(8.26)
South Korea	EM	63,595	205	1.5	1984	2,606	0.548***	(5.66)	0.395***	(4.66)
Spain	DM	57,547	195	1.5	1987	239	0.367***	(4.34)	0.375***	(4.35)
Sweden	DM	60,850	202	1.0	1982	792	0.642***	(6.09)	0.435***	(4.01)
Switzerland	DM	70,058	210	2.4	1980	412	0.428***	(7.72)	0.304***	(5.15)
Taiwan	EM	49,685	187	1.5	1987	2,097	0.288***	(4.26)	0.178**	(2.41)
Thailand	EM	54,010	193	0.4	1987	812	0.373***	(2.94)	0.370***	(3.10)
Turkey	EM	44,749	170	0.4	1988	422	0.230***	(3.00)	0.118	(1.19)
UK	DM	88,919	238	6.8	1980	3,260	0.552***	(11.98)	0.365***	(6.76)
US	DM	99,214	241	42.5	1980	20,026	0.559***	(9.65)	0.359***	(6.43)

between the US and international markets. Using alternative cutoffs does not affect our insights. The final sample consists of about 2.14 million anomaly months.

Table 1 provides country-by-country information about the number of eligible anomaly months, the number of distinct anomalies, and the start of the sample period. In terms of coverage, the non-US G7 states (Canada, Germany, France, Italy, Japan, UK) and Australia are broadly comparable to the US market. In all these countries, the sample period starts in 1980, and we can rely on at least 210 eligible anomalies accounting for more than 69 thousand anomaly months. Given their economic importance, their large stock markets, and the high degree of data availability, the G7 + Australia sample thus seems well suited to serve as an international benchmark in most of our later tests.

Table 1 also displays the unconditional magnitude of anomalies per country. Equally weighted (value-weighted) pooled anomaly returns are significantly positive at the 1% level for all (all but five) of the 39 countries in our sample. In 36 countries, point estimates for equally weighted portfolios are larger than for value-weighted portfolios. This finding is consistent with the notion that both mispricing and limits to arbitrage tend to be stronger for smaller stocks. With the exception of Japan, the G7 + Australia sample is broadly comparable to the US market with respect to long-short anomaly returns. However, these findings are based on the whole sample period and, thus, mask possible differences in the in-sample or post-publication period. We address this possibility in the following.

Table 2
Anomalies in different event periods: descriptive statistics.
This table shows average pooled long-short anomaly returns in different country universes and different event periods (in-sample, post-sample, post-publication). "N" denotes the number of anomaly months. In-sample (post-sample, post-publication) returns account for roughly 47% (16%, 37%) of the about 2.14 million anomaly months in our global sample. "Large markets" refers to countries with more than 60 thousand anomaly months. "G7 + Australia" refers to Australia, Canada, Germany, France, Italy, Japan, and the United Kingdom. We report results for two different country aggregation techniques. "Pooled" ("composite") means that the unit of observation is a long-short anomaly return at the individual country level (at the aggregated international markets level). *t*-statistics are reported in parentheses. Standard errors are clustered by month (US market, composite international markets) or double-clustered by both country and month (pooled international markets). Two-tailed statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

		International markets (pooled)				International markets (composite)			
Event period	US	All	Developed	Large markets	G7 + Australia	All	Developed	Large markets	G7 + Australia
N	99,214	2,041,067	1,246,763	1,075,133	558,408	92,806	92,750	92,788	92,692
Panel A: Equally weighted returns									
In-sample returns	0.742***	0.413***	0.453***	0.465***	0.476***	0.367***	0.379***	0.374***	0.373***
	(12.52)	(9.16)	(9.05)	(8.76)	(6.94)	(10.90)	(10.27)	(10.12)	(9.72)
Post-sample returns	0.466***	0.498***	0.532***	0.597***	0.562***	0.452***	0.454***	0.477***	0.444***
	(5.23)	(8.65)	(7.45)	(8.40)	(6.60)	(9.50)	(7.96)	(8.65)	(7.68)
Post-publication returns	0.292***	0.523***	0.514***	0.555***	0.498***	0.438***	0.419***	0.446***	0.403***
	(3.69)	(9.08)	(7.98)	(8.89)	(6.29)	(9.97)	(7.94)	(8.97)	(7.29)
Panel B: Value-weighted returns									
In-sample returns	0.489***	0.347***	0.365***	0.372***	0.382***	0.254***	0.255***	0.259***	0.262***
	(8.27)	(8.24)	(7.62)	(7.70)	(6.55)	(5.98)	(5.76)	(5.93)	(5.72)
Post-sample returns	0.310***	0.383***	0.391***	0.431***	0.425***	0.242***	0.275***	0.295***	0.287***
	(3.45)	(6.62)	(5.29)	(5.86)	(5.20)	(4.12)	(4.47)	(4.89)	(4.55)
Post-publication returns	0.163**	0.371***	0.336***	0.370***	0.333***	0.198***	0.199***	0.222***	0.210***
	(2.07)	(6.02)	(5.19)	(5.92)	(4.59)	(4.44)	(4.25)	(4.87)	(4.30)

3. The impact of publication on anomalies across the globe

3.1. Baseline results

We can benchmark the behavior of anomalies in the US market either against individual countries or against aggregated international markets. We use two conceptually different aggregation methods, which turn out to yield similar results. For a given anomaly, we either pool long-short portfolio returns obtained from individual countries or we construct a single composite international long-short portfolio return.

In the pooling approach, the unit of observation is a monthly long-short anomaly return at the country level. Holding the number of anomalies fixed, the weight of a given country remains stable over time. Depending on the country universe, a drawback of this method is that it puts a large weight on minor markets with only few eligible firms. The approach could give a good indication on how widespread post-publication changes in anomaly profitability are.

In the composite approach, the unit of observation is a monthly long-short anomaly return aggregated from all eligible non-US markets. We first calculate returns in US dollars. For a given return predictor and for each country separately, we then construct a long-short portfolio. All stocks contained in these country-level long-short portfolios are then aggregated into a single global portfolio, based on which we compute monthly composite anomaly returns. One the one hand, the method puts a larger weight on larger stock markets, which are of particular economic and academic interest. On the other hand, a drawback of this method in particular with respect to value-weighted returns is that post-publication changes in global anomaly returns could, to some extent, simply reflect changes in country weights over time.[4]

In sum, from an empirical point of view, no optimal way arguably exists of aggregating anomaly returns across countries (pooled or composite) and within countries (equally weighted or value-weighted). In the remainder of the paper, we thus compute all four specifications to test for the robustness of our findings.

To get a first glance at publication effects, Table 2 displays the pooled long-short anomaly returns in the US market and different international samples separately for in-sample, post-sample, and post-publication periods. US publication effects are pronounced. For equally weighted portfolio returns, the profitability drops from 74 basis points (bps) in-sample to 47 bps post-sample, and to 29 bps post-publication. For value-weighted portfolio returns, the corresponding numbers are 49 bps, 31 bps, and 16 bps. In international markets, no clear difference emerges between in-sample, post-sample, and post-publication periods. This finding holds irrespective of whether we rely on the pooling or on the composite approach, whether we look at equally weighted or at value-weighted returns, or whether we consider all international markets, only MSCI developed markets, only large stock markets with more than 60 thousand anomaly months, or the G7 + Australia sample.

[4] In our sample, this effect is mostly attributable to Japan. As Table 1 shows, the country's total market capitalization is by far the largest of all non-US markets. It also exhibits unusually low average anomaly returns. At the same time, the relative weight of Japan in our sample heavily fluctuates over time. For instance, it accounts for more than 40% of world market capitalization in the late 1980s, but for less than 10% in 2015.

In the in-sample period, equally weighted (value-weighted) monthly global anomaly returns range from 37 bps to 48 bps (25 bps–38 bps), which is lower than the corresponding US estimate of 74 bps (49 bps). In the post-sample period, equally weighted (value-weighted) global anomaly returns range from 44 bps to 60 bps (24 bps–43 bps), which is comparable to the US estimate of 47 bps (31 bps). In the post-publication period, equally weighted (value-weighted) global returns range from 42 bps to 56 bps (20 bps–37 bps), which is larger than the corresponding US estimate of 29 bps (16 bps). The fact that anomalies are statistically significant and economically meaningful in out-of-sample firm universes and time periods suggests that data mining is not the major driver behind the return predictability reported in the original US studies.

To more formally test for publication effects in different markets, we run the following regression model at both the country level and the aggregate international sample level:

$$R_{i,t} = \alpha_i + \beta_1 * Post\ Sample\ Dummy_{i,t} + \beta_2 * Post\ Publication\ Dummy_{i,t} + \epsilon_{i,t}. \quad (1)$$

In Eq. (1), $R_{i,t}$ refers to the raw long-short return of anomaly i in month t in a given country. The post-sample dummy is one if t is after the end of the original sample but still pre-publication and zero otherwise. The post-publication dummy is one if t is equal to or larger than the month of the publication and zero otherwise. α_i captures anomaly fixed effects. In the pooled regressions, we instead include fixed effects for (country, anomaly) pairs. Standard errors are clustered by month.

We are interested in measuring the market impact of publication-informed trading. The most natural proxy is thus the regression coefficient of the post-publication dummy. Clearly, the coefficient is only a noisy proxy, for instance because trading against anomalies could have started long before publication (e.g., Penasse, 2018). The coefficient of the post-sample dummy could reflect both (the upper bound of) data mining and arbitrage trading. Again, the latter assumes that at least some sophisticated market participants learn from academic research about mispricing before the official publication. For instance, practitioners could attend academic seminars and read the working paper version or forthcoming publication. Alternatively, information could flow from practitioners to researchers. For instance, academics may write papers about anomalies currently exploited by arbitrageurs.

Our key findings are summarized in Tables 3 and 4. Table 3 reports country-level regression results, which quantify the average change (in bps per month) in anomaly profitability in the post-sample and post-publication period relative to the in-sample period. Table 4 reports aggregated results and provides estimates of implied relative changes (in percent) in monthly anomaly profitability.

In Table 3, our findings for the US stock market are in line with the results in McLean and Pontiff (2016). This is noteworthy as we rely on a shorter sample period, on a broader set of anomalies, and on a common methodological framework instead of exactly following the screens and approaches of the original papers. The US post-sample (post-publication) coefficient for equally weighted returns in regression Eq. (1) is −0.276 (−0.450), which indicates a 27.6 bps (45 bps) drop in monthly anomaly profitability post-sample (post-publication). Both estimates are significant at the 1% level.[5] The average predictor in the US stock market has an in-sample mean return of 72.4 bps. The regression coefficients thus imply that, relative to the in-sample mean, the average anomaly return decreases by 38% post-sample and by 62% post-publication (see Table 4). For value-weighted US portfolio returns, the post-sample coefficient is −0.173 (t-statistic −2.02) and the post-publication coefficient is −0.305 (t-statistic −2.96). Relative to the in-sample mean of 46.4 bps, these estimates imply a decrease of 37% and 66%.

Our estimates for international markets in both Tables 3 and 4 stand in contrast to these findings. With respect to the post-sample dummy in the country-level regressions, only the estimates for Hong Kong and Mexico are (partly) significantly negative. With respect to the post-publication dummy, only the value-weighted results for China are significantly negative. Some estimates for both dummies are significantly positive. This finding holds for both equally weighted and value-weighted returns.

Table 4, which aggregates the international markets, provides a similar message. In the 16 specifications for international markets (four samples, two country aggregation techniques, two portfolio return weighting schemes), the coefficient for the post-publication dummy ranges from about −5 bps to +15 bps (t-statistics −0.66 to 2.13). Regression coefficients for the post-sample dummy are similar. Focusing on the relative change in anomaly profitability in Panels B and D of Table 4, we find the post-publication change to range between −19% and +42%.

In other words, the post-publication drop in anomaly returns exists in the US stock market only. In the following, we explore potential reasons for this finding. For brevity and due to the presumed economic similarity, we contrast the US against the G7 + Australia sample in all remaining tests. Using alternative country groups does not change any insights, as already suggested by Table 4.

3.2. Changes in the anomaly universe

The baseline results rely on the full set of 241 anomalies. These results perhaps do not represent a general phenomenon but could be traced back to a specific group of anomalies. Table 5 shows that this is not the case.

In Panels A to D of Table 5, we distinguish between four anomaly categories as in McLean and Pontiff (2016). Event predictors are anomalies based on corporate events or changes in performance. Market predictors are mainly or exclusively based on lagged financial data such as returns, prices, trading volume, or shares outstanding. Fundamental predictors are based on financial statement data.

[5] Nevertheless, the post-publication coefficient is statistically different from the post-sample coefficient both in the equally weighted portfolio analysis ($p = 0.033$) and the value-weighted analysis ($p = 0.100$). Thus, even under the conservative assumption that the post-sample coefficient captures only statistical biases and no arbitrage trading, there is still a reliable post-publication drop in US anomaly profitability.

Table 3
Anomalies and publication effects: results at the individual country level.
This table shows the main findings obtained from country-level regressions of pooled long-short anomaly returns on dummies for post-sample and post-publication periods. The post-sample dummy is one if anomaly month *t* is after the end of the original sample but still pre-publication and zero otherwise. The post-publication dummy is one if *t* is equal to or larger than the month of the publication and zero otherwise. All regressions include anomaly fixed effects. *t*-statistics are reported in parentheses. Standard errors are clustered by month. Two-tailed statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Country	Equally weighted long-short returns				Value-weighted long-short returns			
	Post-sample		Post-publication		Post-sample		Post-publication	
	Return	*t*-stat	Return	*t*-stat	Return	*t*-stat	Return	*t*-stat
Australia	0.069	(0.69)	0.051	(0.36)	−0.086	(−0.82)	0.050	(0.32)
Austria	0.135	(1.25)	−0.003	(−0.02)	0.103	(0.85)	−0.083	(−0.53)
Belgium	0.028	(0.33)	−0.068	(−0.53)	0.106	(0.78)	−0.097	(−0.57)
Brazil	0.166	(0.89)	0.551**	(2.24)	0.038	(0.13)	0.205	(0.72)
Canada	0.041	(0.35)	0.045	(0.30)	0.082	(0.62)	−0.083	(−0.49)
Chile	−0.053	(−0.54)	0.047	(0.31)	−0.125	(−1.17)	−0.004	(−0.03)
China	0.081	(0.73)	−0.162	(−1.09)	0.037	(0.31)	−0.275*	(−1.74)
Denmark	0.081	(0.87)	0.157	(1.27)	−0.093	(−0.81)	0.137	(0.91)
Finland	0.097	(0.81)	0.060	(0.33)	−0.142	(−0.76)	−0.242	(−0.93)
France	0.006	(0.06)	0.000	(0.00)	−0.071	(−0.65)	−0.142	(−1.44)
Germany	0.256***	(2.89)	0.253***	(2.98)	0.232*	(1.95)	0.038	(0.38)
Greece	0.494***	(3.29)	0.350	(1.56)	0.589***	(3.18)	0.765**	(2.32)
Hong Kong	−0.227*	(−1.71)	−0.217	(−1.08)	−0.238*	(−1.75)	−0.193	(−1.06)
India	−0.120	(−0.98)	0.168	(0.96)	−0.210	(−1.35)	−0.012	(−0.05)
Indonesia	0.102	(0.48)	0.073	(0.29)	−0.008	(−0.03)	0.009	(0.03)
Ireland	0.204	(0.83)	0.637**	(2.02)	0.037	(0.10)	0.181	(0.47)
Israel	−0.071	(−0.62)	0.056	(0.33)	−0.002	(−0.01)	0.191	(0.81)
Italy	0.208**	(2.32)	0.081	(0.68)	0.171	(1.57)	0.070	(0.58)
Japan	0.066	(0.92)	0.037	(0.37)	0.060	(0.75)	−0.006	(−0.05)
Malaysia	0.276***	(2.65)	0.467***	(2.86)	0.172	(1.61)	0.231	(1.43)
Mexico	−0.228	(−1.60)	0.066	(0.35)	−0.327**	(−2.28)	−0.060	(−0.34)
Netherlands	0.118	(1.12)	0.091	(0.67)	−0.050	(−0.36)	−0.110	(−0.79)
New Zealand	0.114	(1.04)	0.086	(0.56)	0.169	(1.35)	0.004	(0.03)
Norway	0.103	(0.80)	0.238	(1.43)	0.091	(0.67)	−0.025	(−0.14)
Pakistan	−0.129	(−0.80)	0.194	(0.83)	0.016	(0.09)	0.361	(1.55)
Philippines	0.041	(0.20)	0.231	(0.78)	−0.095	(−0.46)	0.160	(0.51)
Poland	−0.092	(−0.64)	0.063	(0.35)	−0.019	(−0.10)	0.046	(0.22)
Portugal	0.204	(1.44)	0.280	(1.45)	−0.060	(−0.40)	0.252	(1.26)
Singapore	0.348***	(2.84)	0.437**	(2.52)	0.256*	(1.95)	0.291*	(1.72)
South Africa	0.176*	(1.69)	−0.005	(−0.05)	0.052	(0.44)	−0.153	(−1.14)
South Korea	0.356**	(2.58)	0.222	(1.22)	0.340**	(2.54)	0.170	(1.03)
Spain	−0.112	(−0.87)	−0.086	(−0.43)	0.040	(0.26)	0.131	(0.64)
Sweden	0.114	(0.96)	0.159	(1.01)	−0.033	(−0.22)	−0.075	(−0.44)
Switzerland	0.134	(1.37)	0.040	(0.39)	−0.048	(−0.35)	−0.075	(−0.69)
Taiwan	0.284**	(2.38)	0.331**	(2.36)	0.297**	(2.31)	0.330**	(2.12)
Thailand	0.237*	(1.70)	0.241	(1.01)	0.208	(1.16)	0.025	(0.11)
Turkey	0.024	(0.16)	0.195	(1.16)	0.173	(0.84)	0.115	(0.48)
UK	0.118	(1.59)	0.114	(1.15)	−0.008	(−0.10)	0.037	(0.29)
US	−0.276***	(−3.43)	−0.450***	(−4.75)	−0.173**	(−2.02)	−0.305***	(−2.96)

Valuation predictors are constructed from both market and fundamental data. In Panels E and F of Table 5, we distinguish between the 80 anomalies also relied on in McLean and Pontiff (2016) and our alternative set of 161 anomalies.[6] In the Online Appendix, we partition the sample in other plausible ways, such as focusing on anomalies published in the top finance and accounting papers or collapsing anomaly returns at the paper level. Inferences do not change.

3.3. Controlling for time effects

The results presented so far are not necessarily related to anomaly publication but could be attributable to general time effects. In Panel A (B) Table 6, we thus add a linear time trend (month fixed effects) to our baseline estimation. The US post-publication dummies remain significant, and findings for our international sample are close to zero.

3.4. Accounting for risk factor exposure

Investors could be primarily concerned with alphas, not with raw long-short returns. US anomaly portfolios could differ from their international counterparts in their local

[6] For the the US, the correlation between a meta anomaly that averages the equally weighted monthly in-sample returns of the original 80 anomalies and an equivalent meta anomaly constructed from the alternative anomalies is 0.72. The corresponding estimate for our main international sample (G7 + Australia, pooled or composite) is 0.56. In absolute terms, our US estimates for the post-publication drop are even larger for the alternative set of anomalies than for the original set. In relative terms, US estimates for both sets of anomalies are similar, as the alternative set has higher in-sample returns.

Table 4
Anomalies and publication effects: aggregate results.
This table shows the main findings obtained from regressions of pooled anomaly returns on dummies for post-sample and post-publication periods. The post-sample dummy is one if anomaly month t is after the end of the original sample but still pre-publication and zero otherwise. The post-publication dummy is one if t is equal to or larger than the month of the publication and zero otherwise. "Large only" refers to countries with more than 60 thousand eligible anomaly months. "G7 + Australia" refers to Australia, Canada, Germany, France, Italy, Japan, and the United Kingdom. We report results for two different country aggregation techniques. "Pooled" ("composite") means that the unit of observation is a long-short anomaly return at the individual country level (at the aggregated international markets level). Panels A and C show regression coefficients. Panels B and D translate these coefficients into relative changes with respect to the magnitude of the average in-sample anomaly return. t-statistics are reported in parentheses. Regressions include anomaly fixed effects or, for the pooled international sample, fixed effects for (country, anomaly) pairs. Standard errors are clustered by month. Two-tailed statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

		International markets (pooled)				International markets (composite)			
Period	US	All	Developed	Large	G7 + Australia	All	Developed	Large	G7 + Australia
N	99,214	2,041,067	1,246,763	1,075,133	558,408	92,806	92,750	92,788	92,692
Panel A: Regression coefficients, equally weighted long-short returns									
Post-sample	−0.276***	0.103**	0.094*	0.153***	0.108*	0.113**	0.108**	0.142***	0.104*
	(−3.43)	(2.41)	(1.81)	(3.07)	(1.95)	(2.59)	(2.09)	(2.85)	(1.95)
Post-publication	−0.450***	0.132*	0.093	0.138**	0.083	0.128**	0.106	0.146**	0.098
	(−4.75)	(1.88)	(1.26)	(2.00)	(1.14)	(2.11)	(1.48)	(2.13)	(1.29)
Panel B: Implied relative changes in anomaly profitability, equally weighted long-short returns									
Mean in-sample return	0.724	0.410	0.450	0.459	0.463	0.340	0.352	0.345	0.343
Post-sample change	−38%	25%	21%	33%	23%	33%	31%	41%	30%
Post-publication change	−62%	32%	21%	30%	18%	38%	30%	42%	29%
Panel C: Regression coefficients, value-weighted long-short returns									
Post-sample	−0.173**	0.047	0.029	0.064	0.053	−0.007	0.033	0.053	0.039
	(−2.02)	(0.94)	(0.47)	(1.08)	(0.82)	(−0.11)	(0.51)	(0.82)	(0.59)
Post-publication	−0.305***	0.044	−0.010	0.022	−0.006	−0.046	−0.032	−0.008	−0.024
	(−2.96)	(0.57)	(−0.12)	(0.29)	(−0.08)	(−0.66)	(−0.45)	(−0.12)	(−0.32)
Panel D: Implied relative changes in anomaly profitability, value-weighted long-short returns									
Mean in-sample return	0.464	0.331	0.356	0.372	0.362	0.245	0.247	0.250	0.249
Post-sample change	−37%	14%	8%	17%	15%	−3%	13%	21%	16%
Post-publication change	−66%	13%	−3%	6%	−2%	−19%	−13%	−3%	−10%

risk factor exposure, which could explain part of the differences in our baseline findings. We test this conjecture in Panels C and D of Table 6. In Panel C, we rely on the CAPM, which already has been public knowledge at the beginning of our sample period. We implement a two-stage regression approach. First, we regress the time series of country-specific anomalies on the local market excess return. We do so separately for the time series of in-sample returns and the time series of post-sample/post-publication returns. Using a single regression does not change insights. Abnormal returns are then defined as the intercept plus the fitted value of the residual. Second, we perform a regression as in Eq. (1). In Panel D, we rely on local Fama and French (1993) alphas instead of the CAPM. In both cases, inferences remain unchanged.

3.5. Database issues

While Datastream and Worldscope have been relied on in many top published finance papers, market coverage could be an issue, especially in earlier years of our sample period.[7] Nevertheless, we do not find convincing evidence for the conjecture that the level or changes in international data availability during our sample period could be the main driver of our findings. First, coverage is likely to be best for large developed markets and large stocks, but we do not find evidence for post-publication effects in countries such as Canada, Japan, Germany, or the United Kingdom. In Section 4 we create anomalies from large stocks only but still find a difference between the US and non-US markets. Second, as shown in the Online Appendix, the findings for non-US markets are not significant in a more recent sample period starting from 1995, when international data appear to be more widely available. Third, taking time fixed effects or time trends into account does not change the qualitative nature of our findings. Fourth, the Online Appendix reports results obtained for the US market when conditioning on stock months with joint availability on both Datastream and Worldscope as well as on CRSP and Compustat. Findings are slightly weaker, but insights do not change. The US remains the only market with a statistically significant, robust, and economically meaningful post-publication drop in anomaly profitability.

4. The role of arbitrage costs

We explore the impact of impediments to arbitrage on the magnitude and post-publication change of anomalies across the globe. In Section 4.1, we focus on possible differences in arbitrage costs between the US and major international markets. The idea is to compare US and non-US anomalies with similar in-sample returns and similar

[7] The percentage of stocks covered by Datastream and Worldscope could be smaller than the corresponding coverage of CRSP and Compustat for US stocks, and the differences in coverage could be related to differences in stock characteristics such as firm size. Further discussions or comparisons of databases can be found in Dai (2012), Ince and Porter (2006), Jacobs (2016), Karolyi (2016), and Tobek and Hronec (2018b), among others.

Table 5
Anomalies and publication effects: the impact of the return predictor universe.
This table shows variations of the baseline analysis displayed in Table 4. We condition on subsets of anomalies as described in the text. *t*-statistics are reported in parentheses. Standard errors are clustered by month. Two-tailed statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

	Equally weighted returns			Value-weighted returns		
Period	US	G7 + Australia		USA	G7 + Australia	
		Pooled	Composite		Pooled	Composite
Panel A: Event-based anomalies						
Post-sample	-0.238***	0.062	0.059	-0.173**	0.001	-0.036
	(-3.84)	(1.37)	(1.14)	(-2.17)	(0.02)	(-0.49)
Post-publication	-0.443***	0.011	0.044	-0.265***	-0.046	-0.076
	(-6.28)	(0.20)	(0.78)	(-2.78)	(-0.66)	(-1.06)
N	24,883	125,641	22,359	24,883	125,641	22,359
Panel B: Fundamental-based anomalies						
Post-sample	-0.131*	0.208***	0.141**	-0.021	0.185***	0.139
	(-1.91)	(3.46)	(2.30)	(-0.24)	(2.82)	(1.55)
Post-publication	-0.366***	0.269***	0.225***	-0.158*	0.129*	0.139*
	(-4.58)	(4.21)	(3.68)	(-1.77)	(1.96)	(1.79)
N	27,736	156,114	25,649	27,736	156,114	25,649
Panel C: Market-based anomalies						
Post-sample	-0.414**	0.083	0.096	-0.321*	-0.016	0.017
	(-2.29)	(0.86)	(0.91)	(-1.78)	(-0.13)	(0.13)
Post-publication	-0.473***	-0.004	0.035	-0.396**	-0.051	-0.094
	(-2.96)	(-0.04)	(0.26)	(-2.32)	(-0.38)	(-0.65)
N	29,356	179,434	28,525	29,356	179,434	28,525
Panel D: Valuation-based anomalies						
Post-sample	-0.332	0.076	0.142	-0.138	0.050	0.053
	(-1.58)	(0.66)	(1.16)	(-0.56)	(0.34)	(0.33)
Post-publication	-0.540***	0.039	0.077	-0.445***	-0.083	-0.088
	(-3.76)	(0.37)	(0.64)	(-2.80)	(-0.66)	(-0.60)
N	17,239	97,219	16,159	17,239	97,219	16,159
Panel E: Subset of anomalies from (McLean and Pontiff, 2016)						
Post-sample	-0.157*	0.160***	0.121**	-0.065	0.064	0.054
	(-1.94)	(3.11)	(2.29)	(-0.79)	(1.10)	(0.71)
Post-publication	-0.377***	0.157**	0.147*	-0.227**	0.027	0.015
	(-4.03)	(2.17)	(1.92)	(-2.30)	(0.36)	(0.20)
N	33,498	193,360	31,309	33,498	193,360	31,309
Panel F: Alternative set of anomalies						
Post-sample	-0.337***	0.083	0.096	-0.227**	0.048	0.032
	(-3.55)	(1.24)	(1.43)	(-2.17)	(0.60)	(0.38)
Post-publication	-0.490***	0.042	0.070	-0.348***	-0.026	-0.046
	(-4.66)	(0.52)	(0.85)	(-2.92)	(-0.28)	(-0.53)
N	65,716	365,048	61,383	65,716	365,048	61,383

firm-level arbitrage costs. We test whether the difference in post-publication change can be traced back to differences in in-sample anomaly profitability, or firm size, or both. On a broader level, this matching approach helps to determine whether or not cross-country barriers to investment management create segmented markets.

In Section 4.2, we focus on possible differences in arbitrage costs within (as opposed to between) markets. Separately for the US market and global markets, we test whether anomalies constructed from stocks with higher (lower) arbitrage costs yield higher (lower) returns. The behavioral finance view on anomalies suggests that costs and risks can prevent arbitrageurs from aggressively betting against mispricing (e.g., Barberis and Thaler, 2003; Pontiff, 1996; Shleifer and Vishny, 1997. Limits to arbitrage should thus be positively correlated with anomaly returns in-sample (post-publication) provided that at least some arbitrageurs try to bet against mispricing in the pre-publication (post-publication) period. The data mining view on return predictability does not establish a link between arbitrage costs and anomaly returns. The analysis thus helps to get a better understanding of the mechanisms underlying the typical anomaly.

4.1. Differences in arbitrage costs between markets

Many limits to arbitrage are related to firm characteristics. We focus on firm size in our cross-country tests and analyze the impact of other stock-level variables in the within country tests reported in Section 4.2. Firm size is of utmost importance in anomaly asset pricing tests (e.g., Fama and French, 2008; Hou et al., 2019) and is widely regarded as one of the key limitations to arbitrage trading (e.g., Baker and Wurgler, 2006). Firm size is also correlated with other arbitrage costs. For instance, larger stocks tend to exhibit lower total and idiosyncratic risk (e.g., Herskovic et al., 2016; Fama and French, 2008), they are less costly to

Table 6
Anomalies and publication effects: the impact of time effects and asset pricing models.
This table shows variations of the baseline analysis displayed in Table 4. In Panel A, we include a linear time trend that is 1/100 for January 1980 and increases by 1/100 in each sample month. In Panel C (D), we rerun the baseline analysis but rely on long-short returns that are adjusted for their exposure to the local market excess return [local, Fama and French, 1993 factors]. *t*-statistics are reported in parentheses. Standard errors are clustered by month. Two-tailed statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

	Equally weighted returns			Value-weighted returns		
Period	US	G7 + Australia		USA	G7 + Australia	
		Pooled	Composite		Pooled	Composite
Panel A: Linear time trend						
Post-sample	-0.188***	0.042	0.051	-0.161**	0.020	0.027
	(-2.68)	(0.74)	(1.01)	(-2.06)	(0.31)	(0.43)
Post-publication	-0.299***	-0.035	0.003	-0.285**	-0.065	-0.045
	(-2.62)	(-0.39)	(0.04)	(-2.27)	(-0.66)	(-0.49)
Time trend	-0.063	0.055**	0.042	-0.009	0.027	0.010
	(-1.39)	(2.01)	(1.56)	(-0.18)	(0.84)	(0.28)
N	99,214	558,408	92,692	99,214	558,408	92,692
Panel B: Month fixed effects						
Post-sample	-0.119**	0.059	0.058	-0.116*	0.044	0.031
	(-2.18)	(1.34)	(1.35)	(-1.79)	(0.83)	(0.55)
Post-publication	-0.168**	0.056	0.061	-0.189**	0.030	0.026
	(-2.18)	(1.03)	(1.04)	(-2.50)	(0.48)	(0.37)
N	99,214	558,408	92,692	99,214	558,408	92,692
Panel C: Capital Asset Pricing Model (CAPM) alphas						
Post-sample	-0.278***	0.131***	0.122***	-0.180**	0.080	0.040
	(-4.02)	(2.99)	(2.65)	(-2.51)	(1.48)	(0.64)
Post-publication	-0.424***	0.106*	0.115*	-0.280***	0.018	-0.026
	(-5.14)	(1.72)	(1.82)	(-3.19)	(0.26)	(-0.36)
N	99,214	558,408	92,692	99,214	558,408	92,692
Panel D: Three-factor model alphas						
Post-sample	-0.204***	0.125***	0.169***	-0.124*	0.065	0.106*
	(-3.13)	(2.90)	(3.65)	(-1.83)	(1.24)	(1.76)
Post-publication	-0.373***	0.080	0.155***	-0.239***	-0.020	0.037
	(-5.37)	(1.58)	(2.79)	(-3.24)	(-0.35)	(0.56)
N	99,214	553,584	91,692	99,214	553,584	91,692

short (e.g., D'Avolio, 2002), and they are more liquid (e.g., Corwin and Schultz, 2012).[8]

In Table 7, we rerun our analysis with anomalies constructed from three size groups. As a benchmark, Panels A and B of Table 7 replicate our baseline analysis. The market capitalization for the median (mean) firm month in this sample is about \$145 million (\$1.95 billion) for the US market and \$114 million (\$1.39 billion) for the average G7 + Australia country. Following the size cut-offs proposed in Fama and French (2008), Panels C and D (E and F) are based on anomalies constructed solely from large firms with a market capitalization above the 20th percentile (50th percentile) of stocks trading at the New York Stock Exchange. For perspective, the average size cut-offs in 2015 are about \$600 million and \$2.6 billion, respectively.

Table 7 shows clear patterns. For all size groups and both in absolute and relative terms, there is a strong post-publication decline in the US, but no decline in G7 + Australia. With increasing firm size, the difference between the US and non-US markets becomes smaller, but it remains meaningful and significant even for the largest stocks. For instance, Panel E of Table 7 concentrates on equally weighted anomaly returns constructed from firms larger than the NYSE median. The post-publication drop in the US market is 20.6 bps, which implies a 66% decline relative to the average in-sample return of 31 bps. The post-publication change in pooled (composite) stock markets is just −5 bps (+5 bps), both of which are significantly different from the pronounced change in the United States.

Nevertheless, limits to arbitrage have many facets. As shown in Table 2, in-sample returns tend to be higher in the United States than in the average non-US market. All else equal, arbitrageurs could be more likely to trade on anomalies with higher in-sample returns. We thus implement an anomaly matching procedure based on the two following criteria. First, anomalies need to have an in-sample profitability of at least 50 bps per month in both the US and the international sample of interest. Second, the absolute difference between the in-sample profitability in the US and international markets has to be smaller than 25 bps. This matching procedure is run separately for

[8] Arbitrage costs at the firm level are often divided into transaction costs and holding costs (e.g., Pontiff, 2006). Firm size is one of the leading proxies for transaction costs, and idiosyncratic risk relative to the hedge portfolio available to the arbitrageur is a leading holding cost. On average, stocks in our US anomaly portfolios exhibit comparable or higher idiosyncratic risk (relative to standard market or factor models) than stocks in anomaly portfolios of major international markets. This finding is consistent with Bartram et al. (2012) and justifies our focus on firm size in cross-country comparisons.

Table 7

Anomalies and publication effects: the role of firm size.

In this table, we rerun our baseline analysis but modify the eligible firm universe when constructing anomalies. To facilitate comparison, Panels A and B show our baseline findings from Table 4. To test for differences between the US market and international markets, we pool the samples and report the coefficient on the interaction effect between a US market dummy and the post-sample and post-publication dummies. In Panels C and D (E and F), we re-create all anomalies with stocks having a one-month lagged market capitalization larger than the 20th (50th) percentile of stocks trading at the NYSE in that month. This criterion is imposed both on US stocks and on international stocks. In all panels, t-statistics are reported in parentheses. Standard errors are clustered by month. Two-tailed statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Period	US	G7 + Australia Pooled	Difference to US	G7 + Australia Composite	Difference to US
Panel A: All firms, equally weighted returns					
Mean in-sample return	0.724	0.463		0.343	
N	99,214	558,408		92,692	
Post-sample	−0.276***	0.108*	−0.384***	0.104*	−0.380***
	(−3.43)	(1.95)	(−5.86)	(1.95)	(−5.17)
Post-publication	−0.450***	0.083	−0.533***	0.098	−0.548***
	(−4.75)	(1.14)	(−6.97)	(1.29)	(−6.78)
Panel B: All firms, value-weighted returns					
Mean in-sample return	0.464	0.362		0.249	
N	99,214	558,408		92,692	
Post-sample	−0.173**	0.053	−0.226***	0.039	−0.212**
	(−2.02)	(0.82)	(−3.17)	(0.59)	(−2.45)
Post-publication	−0.305***	−0.006	−0.299***	−0.024	−0.281***
	(−2.96)	(−0.08)	(−3.70)	(−0.32)	(−2.92)
Panel C: Firms larger than 20th NYSE size percentile, equally weighted returns					
Mean in-sample return	0.479	0.303		0.271	
N	97,094	474,450		90,358	
Post-sample	−0.245***	0.041	−0.286***	0.098	−0.343***
	(−3.17)	−0.81	(−4.34)	(1.43)	(−4.44)
Post-publication	−0.352***	−0.030	−0.322***	0.077	−0.429***
	(−4.14)	(−0.47)	(−4.64)	(0.84)	(−4.72)
Panel D: Firms larger than 20th NYSE size percentile, value-weighted returns					
Mean in-sample return	0.355	0.272		0.227	
N	97,094	474,450		90,358	
Post-sample	−0.155*	0.011	−0.166**	0.047	−0.202**
	(−1.96)	(0.22)	(−2.37)	(0.61)	(−2.35)
Post-publication	−0.242***	−0.039	−0.202***	0.003	−0.245**
	(−2.59)	(−0.59)	(−2.70)	(0.04)	(−2.46)
Panel E: Firms larger than 50th NYSE size percentile, equally weighted returns					
Mean in-sample return	0.311	0.214		0.215	
N	93,086	382,246		88,020	
Post-sample	−0.119	−0.010	−0.108	0.052	−0.171**
	(−1.59)	(−0.22)	(−1.62)	(0.74)	(−2.24)
Post-publication	−0.206**	−0.054	−0.152**	0.046	−0.251***
	(−2.42)	(−0.87)	(−2.22)	(0.53)	(−2.90)
Panel F: Firms larger than 50th NYSE size percentile, value-weighted returns					
Mean in-sample return	0.275	0.190		0.184	
N	93,086	382,246		88,020	
Post-sample	−0.100	−0.032	−0.067	0.045	−0.144*
	(−1.32)	(−0.68)	(−0.97)	(0.57)	(−1.71)
Post-publication	−0.182**	−0.055	−0.127*	−0.016	−0.166*
	(−1.99)	(−0.85)	(−1.69)	(−0.19)	(−1.70)

equally weighted and value-weighted returns as well as separately for the two international markets aggregation schemes. In Panels A and B of Table 8, we run the analysis without further restrictions on firm size. In Panels C and D (E and F), we also condition on stocks with a lagged market capitalization larger than the 20th percentile (50th percentile) of NYSE stocks, as in Table 7.

Table 8 provides several insights. First, in all tests, the US market shows a reliable and large post-publication drop, both in absolute terms (change in bps per month) and in relative terms (percentage change). Second, also in the international sample, the coefficient on the post-publication dummy is now negative. However, this effect is statistically reliable in only three of the 12 specifications. Consistent with the arbitrage cost argument, the post-publication decrease in non-US markets is most pronounced in Panel G, which represents historically highly profitable anomalies constructed from a value-weighted portfolio of big international stocks. Third, with the partial exception of the largest stocks, the absolute post-publication drop in the US remains statistically reliably stronger than in the major non-US markets. In relative terms, the difference remains economically large in all samples. For instance, in the two specifications of Panel

Table 8
Anomalies and publication effects: matching approach.
This table explores whether differences in publication effects between the US market and major international markets still exist after conditioning on profitable anomalies with comparable in-sample returns. We condition on anomalies with an average in-sample profitability of at least 50 basis points (bps) per month in both the US and the international sample of interest. In addition, the absolute difference between the average monthly in-sample profitability in both samples has to be smaller than 25 bps. This matching procedure is run separately for equally weighted and value-weighted portfolio returns as well as for the two international markets aggregation schemes (pooled, composite). The eligible firm universe in Panels A and B corresponds to the universe in our baseline analysis (see Table 4). To compute differences between the US market and international markets, we pool the samples and report the coefficient on the interaction effect between a US market dummy and the post-sample and post-publication dummies. In Panels C and D (E and F), we condition on firms with a one-month lagged market capitalization larger than the 20th percentile (50th percentile) of NYSE stocks, as in Table 7. In all panels, *t*-statistics are reported in parentheses. Standard errors are clustered by month. Two-tailed statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Period	US	G7 + Australia Pooled	Difference to US	US	G7 + Australia Composite	Difference to US
Panel A: Matched in-sample profitability, equally weighted returns						
Matched strategies	29	29		22	22	
Mean in-sample return	1.143	1.089		0.857	0.813	
N	11,804	68,812		8,953	8,185	
Post-sample	−0.538***	−0.118	−0.420***	−0.245*	−0.159	−0.085
	(−2.63)	(−0.83)	(−2.69)	(−1.87)	(−1.51)	(−0.66)
Post-publication	−0.672***	−0.120	−0.552***	−0.428***	−0.183*	−0.245*
	(−3.28)	(−0.72)	(−3.54)	(−3.40)	(−1.78)	(−1.84)
Panel B: Matched in-sample profitability, value-weighted returns						
Matched strategies	26	26		18	18	
Mean in-sample return	0.803	0.878		0.708	0.654	
N	10,908	64,182		7,443	6,905	
Post-sample	−0.476*	−0.089	−0.387*	−0.580***	−0.263	−0.317
	(−1.73)	(−0.44)	(−1.86)	(−2.72)	(−1.49)	(−1.52)
Post-publication	−0.624**	−0.177	−0.447*	−0.493**	−0.159	−0.334*
	(−2.08)	(−0.76)	(−1.94)	(−2.47)	(−0.94)	(−1.70)
Panel C: Matched in-sample profitability, firms larger than 20th NYSE size percentile, equally weighted returns						
Matched strategies	30	30		18	18	
Mean in-sample return	0.937	0.912		0.757	0.748	
N	12,489	66,091		7,552	7,032	
Post-sample	−0.584***	−0.075	−0.509***	−0.477**	0.083	−0.560**
	(−2.79)	(−0.50)	(−2.94)	(−2.06)	(0.48)	(−2.48)
Post-publication	−0.716***	−0.228	−0.488***	−0.569***	−0.118	−0.451**
	(−2.90)	(−1.19)	(−2.60)	(−4.30)	(−0.74)	(−2.42)
Panel D: Matched in-sample profitability, firms larger than 20th NYSE size percentile, value-weighted returns						
Matched strategies	16	16		20	20	
Mean in-sample return	0.726	0.751		0.700	0.680	
N	6,634	34,188		8,195	7,362	
Post-sample	−0.125	−0.263	0.138	−0.539***	−0.260	−0.279
	(−0.47)	(−1.36)	(0.60)	(−2.70)	(−1.48)	(−1.29)
Post-publication	−0.690**	−0.318	−0.372*	−0.448***	−0.220	−0.228
	(−2.48)	(−1.26)	(−1.70)	(−2.62)	(−1.42)	(−1.16)
Panel E: Matched in-sample profitability, firms larger than 50th NYSE size percentile, equally weighted returns						
Matched strategies	19	19		16	16	
Mean in-sample return	0.746	0.790		0.708	0.716	
N	7,672	33,258		6,521	6,136	
Post-sample	−0.398	−0.202	−0.197	−0.527***	−0.097	−0.430**
	(−1.33)	(−1.24)	(−0.70)	(−2.79)	(−0.60)	(−2.37)
Post-publication	−0.615*	−0.199	−0.416*	−0.538***	−0.275*	−0.263
	(−1.90)	(−0.86)	(−1.82)	(−3.21)	(−1.66)	(−1.60)
Panel F: Matched in-sample profitability, firms larger than 50th NYSE size percentile, value-weighted returns						
Matched strategies	16	16		17	17	
Mean in-sample return	0.692	0.697		0.705	0.684	
N	6,650	29,345		7,090	6,550	
Post-sample	−0.514*	−0.438*	−0.321	−0.402*	−0.338	−0.064
	(−1.96)	(−1.73)	(−1.54)	(−1.78)	(−1.48)	(−0.29)
Post-publication	−0.750***	−0.137	−0.435**	−0.667***	−0.386**	−0.281
	(−2.70)	(−0.67)	(−2.33)	(−3.27)	(−2.01)	(−1.37)

E, which shows equally weighted portfolio returns constructed from stocks larger than the NYSE median, the post-publication drop in the US (international markets) is estimated to be 76% to 82% (25% to 38%).

In sum, arbitrage costs, as proxied for by firm size and in-sample anomaly profitability, can explain some, but not all, of the cross-county differences. Our findings thus point to investment barriers between the US and international

Table 9
Anomalies and publication effects: the role of limits to arbitrage.
This table explores whether limits to arbitrage are associated with higher long-short anomaly returns across the globe, both in-sample and post-publication. The dependent variable in the regressions is the monthly long-short anomaly return in one of our three samples (US anomalies, pooled or composite international anomalies). The independent variables reflect average in-sample characteristics of the stocks contained in the long or short leg of the anomaly. To quantify average firm characteristics, we first rank all eligible stocks on a country-by-country and month-by-month basis on the variable of interest (e.g., idiosyncratic risk). Ranks are standardized to range from zero to one. We then compute the average rank of all stocks in the long and short leg of a given anomaly in a given month. Finally, we average these ranks for all available in-sample months of this given anomaly. To illustrate, the independent variable in the US market regressions in Panel A is the average idiosyncratic risk rank of the stocks in a given anomaly portfolio during the in-sample period of this anomaly. To ease interpretation, the firm characteristics used as independent variables are standardized to have a mean of zero and a variance of one. In the bottom two rows of each panel, we test whether the sum of the limits to arbitrage variable ("Arbitrage") plus the interaction ("Pub x Arbitrage") between the publication dummy ("Pub") and the variable is statistically different from zero. In all panels, standard errors are clustered by month. *t*-statistics are reported in parentheses. Two-tailed statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

	USA	G7 + Australia Pooled	G7 + Australia Composite
Panel A: Idiosyncratic risk, equally weighted returns			
Post-publication	−0.372***	0.023	0.025
	(−5.06)	(0.41)	(0.44)
Pub x Idiosyncratic risk	−0.136***	−0.030	−0.043
	(−3.52)	(−0.83)	(−1.19)
Idiosyncratic risk	0.241***	0.199***	0.117***
	(8.27)	(7.99)	(4.72)
Constant	0.677***	0.384***	0.383***
	(11.48)	(8.65)	(10.09)
Arbitrage + (Pub x Arbitrage)	0.105	0.169	0.074
p-value	0.001***	0.000***	0.014**
Panel B: Idiosyncratic risk, value-weighted returns			
Post-publication	−0.274***	−0.038	−0.048
	(−3.52)	(−0.61)	(−0.87)
Pub x Idiosyncratic risk	−0.172***	0.016	−0.016
	(−3.62)	(0.38)	(−0.35)
Idiosyncratic risk	0.218***	0.175***	0.092***
	(6.75)	(5.98)	(2.82)
Constant	0.443***	0.488***	0.263***
	(7.63)	(11.31)	(6.15)
Arbitrage + (Pub x Arbitrage)	0.046	0.191	0.077
p-value	0.242	0.000***	0.029**
Panel C: Market capitalization, equally weighted returns			
Post-publication	−0.360***	0.025	0.021
	(−4.87)	(0.44)	(0.37)
Pub x Firm size	0.059**	0.024	0.062**
	(1.97)	(0.98)	(2.20)
Firm size	−0.162***	−0.141***	−0.095***
	(−8.87)	(−8.55)	(−5.02)
Constant	0.674***	0.486***	0.384***
	(11.38)	(11.29)	(10.08)
Arbitrage + (Pub x Arbitrage)	−0.103	−0.118	−0.033
p-value	0.003***	0.000***	0.139
Panel D: Market capitalization, value-weighted returns			
Post-publication	−0.257***	−0.035	−0.052
	(−3.32)	(−0.57)	(−0.94)
Pub x Firm size	0.124***	0.008	0.071**
	(3.49)	(0.30)	(2.06)
Firm size	−0.193***	−0.141***	−0.097***
	(−8.89)	(−8.23)	(−4.26)
Constant	0.436***	0.382***	0.264***
	(7.45)	(8.58)	(6.09)
Arbitrage + (Pub x Arbitrage)	−0.069	−0.133	−0.026
p-value	0.011**	0.000***	0.313

(*continued on next page*)

Table 9 (*continued*)

	USA	G7 + Australia Pooled	G7 + Australia Composite
Panel E: Amihud (2002) illiquidity, equally weighted returns			
Post-publication	−0.374***	0.012	0.035
	(−5.23)	(0.22)	(0.59)
Pub x Illiquidity	−0.023	−0.079***	−0.035
	(−0.85)	(−3.66)	(−1.56)
Illiquidity	0.146***	0.146***	0.098***
	(8.27)	(9.24)	(6.86)
Constant	0.682***	0.491***	0.379***
	(11.44)	(11.28)	(9.99)
Arbitrage + (Pub x Arbitrage)	0.123	0.068	0.063
p-value	0.000***	0.000***	0.002***
Panel F: Amihud (2002) illiquidity, value-weighted returns			
Post-publication	−0.271***	−0.048	−0.044
	(−3.61)	(−0.79)	(−0.77)
Pub x Illiquidity	−0.105***	−0.054**	−0.057*
	(−3.07)	(−2.24)	(−1.91)
Illiquidity	0.183***	0.145***	0.090***
	(8.52)	(8.81)	(4.85)
Constant	0.445***	0.387***	0.259***
	(7.55)	(8.60)	(6.00)
Arbitrage + (Pub x Arbitrage)	0.078	0.091	0.033
p-value	0.006***	0.000***	0.158
Panel G: Dollar trading volume, equally weighted returns			
Post-publication	−0.377***	0.048	0.032
	(−5.29)	(0.84)	(0.55)
Pub x Dollar trading volume	0.012	0.009	0.036*
	(0.39)	(0.43)	(1.74)
Dollar trading volume	−0.109***	−0.142***	−0.086***
	(−6.11)	(−8.63)	(−6.19)
Constant	0.683***	0.479***	0.379***
	(11.45)	(11.23)	(9.98)
Arbitrage + (Pub x Arbitrage)	−0.097	−0.133	−0.050
p-value	0.001***	0.000***	0.006***
Panel H: Dollar trading volume,value-weighted returns			
Post-publication	−0.271***	−0.018	−0.045
	(−3.63)	(−0.29)	(−0.80)
Pub x Dollar trading volume	0.090**	0.029	0.061**
	(2.34)	(1.29)	(2.20)
Dollar trading volume	−0.160***	−0.139***	−0.085***
	(−7.39)	(−8.62)	(−4.67)
Constant	0.444***	0.375***	0.259***
	(7.53)	(8.44)	(5.97)
Arbitrage + (Pub x Arbitrage)	−0.071	−0.110	−0.024
p-value	0.035**	0.000***	0.255
Panel I: Bid-ask spread, equally weighted returns			
Post-publication	−0.357***	0.003	0.016
	(−4.73)	(0.06)	(0.29)
Pub x Bid-ask spread	−0.096***	−0.153***	−0.106***
	(−2.81)	(−6.48)	(−4.36)
Bid-ask spread	0.182***	0.198***	0.109***
	(8.69)	(9.23)	(4.92)
Constant	0.669***	0.496***	0.386***
	(11.33)	(11.31)	(10.05)
Arbitrage + (Pub x Arbitrage)	0.086	0.045	0.003
p-value	0.002***	0.001***	0.818
Panel J: Bid-ask spread, value-weighted returns			
Post-publication	−0.254***	−0.058	−0.056
	(−3.20)	(−0.96)	(−1.02)
Pub x Bid-ask spread	−0.142***	−0.103***	−0.083**
	(−3.56)	(−3.77)	(−2.45)
Bid-ask spread	0.202***	0.176***	0.094***
	(8.32)	(7.05)	(3.20)
Constant	0.432***	0.392***	0.266***
	(7.41)	(8.65)	(6.12)
Arbitrage + (Pub x Arbitrage)	0.060	0.073	0.011
p-value	0.054*	0.000***	0.540

(*continued on next page*)

Table 9 (*continued*)

	USA	G7 + Australia Pooled	G7 + Australia Composite
Panel K: Composite limits to arbitrage proxy, equally weighted returns			
Post-publication	−0.364***	0.026	0.029
	(−4.98)	(0.46)	(0.49)
Pub x Composite limits to arbitrage	−0.067**	−0.049*	−0.058**
	(−2.35)	(−1.92)	(−2.16)
Composite limits to arbitrage	0.176***	0.174***	0.110***
	(9.18)	(9.33)	(6.10)
Constant	0.675***	0.486***	0.380***
	(11.39)	(11.27)	(10.06)
Arbitrage + (Pub x Arbitrage)	0.109	0.125	0.052
p-value	0.012**	0.000***	0.135
Panel L: Composite limits to arbitrage proxy, value-weighted returns			
Post-publication	−0.262***	−0.035	−0.046
	(−3.42)	(−0.57)	(−0.82)
Pub x Composite limits to arbitrage	−0.133***	−0.021	−0.061*
	(−3.76)	(−0.71)	(−1.80)
Composite limits to arbitrage	0.201***	0.164***	0.100***
	(8.85)	(8.15)	(4.41)
Constant	0.438***	0.382***	0.260***
	(7.49)	(8.59)	(6.07)
Arbitrage + (Pub x Arbitrage)	0.068	0.143	0.039
p-value	0.000***	0.000***	0.019**

markets, which appear to be strong enough to leave discernable traces in price formation. While identifying the precise nature of these barriers is beyond the scope of our paper, the literature suggests several possibilities. Short selling restrictions represent an obvious constraint, and short selling is generally more difficult internationally than in the US (e.g., Boehmer et al., 2017; Gagnon and Karolyi, 2010). There can also be frictions related to investment mandates and benchmarking (e.g., Froot and Dabora, 1999), to the degree of market liberalization (e.g., Henry, 2003), and to investor protection rights (e.g., La Porta et al., 1997). Other factors that can deter arbitrage capital include political or currency risk, the operational or institutional market framework, and regulatory restrictions.[9] The US stock market arguably also has lower information costs. For instance, data are more readily available, fewer cross-firm differences exist in accounting practices, backtesting periods can be easily extended, and most academic asset pricing papers cover the US market only.

In the Online Appendix, we report an additional test on the role of investment barriers. We run several cross-country regressions of the average post-publication return change on a dummy quantifying short selling restrictions at the country level (Bris et al., 2007; McLean et al., 2009). When controlling for time effects, the post-publication change is about 15 bps more negative in countries in which short selling is feasible. This statistically significant finding is also consistent with the idea that market frictions are at least partly responsible for the cross-country differences in post-publication returns.

4.2. Differences in arbitrage costs within markets

If return predictability is the result of mispricing, then anomalies concentrated in stocks that are costlier to arbitrage should generate higher long-short returns (e.g., McLean and Pontiff, 2016). To test this hypothesis, we consider five widely used arbitrage cost variables: (1) idiosyncratic risk computed as the standard deviation of the residual obtained from rolling regressions of a stock's daily excess return on a country-specific Fama and French (1993) model over the previous 12 months, (2) lagged firm size, (3) illiquidity as computed in Amihud (2002), (4) dollar trading volume, and (5) bid-ask spreads as estimated in Corwin and Schultz (2012). We also compute an aggregate limits to arbitrage index as the first principal component of the five individual proxies.

To measure arbitrage costs of a given anomaly, we begin by ranking all eligible stocks separately for each country month on the variable of interest (e.g., idiosyncratic risk). Ranks are standardized to range from zero to one. For each anomaly month, we then compute the average rank of all stocks in the long and short leg. Following McLean and Pontiff (2016), we average these ranks for all available in-sample months of the anomaly. We then regress anomaly returns on the publication dummy, the arbitrage cost variable of interest, and the interaction term. The coefficient on the arbitrage cost variable quantifies the impact of limits to arbitrage during the in-sample period. We expect a positive (negative) coefficient for idiosyncratic risk, illiquidity, bid-ask spreads, and the composite index (firm size, dollar volume). The impact during the post-publication period is given by the sum of the coefficient on the arbitrage cost variable and the

[9] For instance, since 2008 (2012), Japan [European Union (EU) countries] have required the public disclosure of large short positions (e.g., Boehmer et al., 2018; Jones et al., 2016). The Alternative Investment Fund Managers Directive of the EU Parliament and the European Council also imposes restrictions on hedge fund leverage, transparency standards, remuneration structures, redemption policies, liquidity management, and more. The application of the directive is generally designated for EU and non-EU investment fund managers provided that they manage EU funds (Council of European Union, 2011)

coefficient on the interaction term. Both the in-sample and the post-publication effects are reported in Table 9.

As we have three anomaly universes (US market, pooled G7 + Australia, composite G7 + Australia), two stock-weighting schemes (equally weighted, value-weighted), and six proxies for arbitrage costs, we estimate 36 specifications in total. With respect to the in-sample returns, all coefficients on the arbitrage costs are statistically significant with the predicted sign. With respect to the post-publication returns, 28 of the 36 coefficients are statistically significant with the predicted sign. Our findings are also economically meaningful. To illustrate, in Panel A, an anomaly based on stocks whose idiosyncratic risk is one standard deviation higher than on average is estimated to have an in-sample US market return that is 24 bps higher than on average. The corresponding estimate for pooled (composite) international markets is 20 bps (12 bps). Post-publication, the anomaly is estimated to generate a return that is 11 bps, 17 bps, and 8 bps higher than the return of the average anomaly. These findings lend credibility to the idea that anomalies tend to reflect mispricing rather than statistical biases.

5. Conclusion

We implement country-level versions of 241 cross-sectional return anomalies published in peer-reviewed finance and accounting journals. The resulting more than two million anomaly country-months allow us to conclude that, unconditionally, long-short return predictability is similar across most developed stock markets. Anomalies in international markets are strong both post-sample and post-publication, and the magnitude of anomalies across the globe is related to arbitrage costs. These findings suggest that data mining is not the key reason for the return predictability originally found in US markets.

In our analysis, only the US market shows a robust and significant post-publication decline in long-short returns. Even after focusing on a matched sample of anomalies, our estimates often still point to economically substantial and statistically significant differences between the return dynamics in the US and international markets. This finding points to cross-country barriers to investment management, which can create segmented markets.

Our findings suggest different directions for future research. A thorough analysis of geographic differences in capital devoted to quantitative arbitrage strategies is needed. Even among non-US markets, cross-country variation exists in anomaly profitability that warrants further investigation.

References

Amihud, Y., 2002. Illiquidity and stock returns: cross section and time series effects. J. Financ. Mark. 5, 31–56.
Baker, M., Wurgler, J., 2006. Investor sentiment and the cross section of stock returns. J. Financ. 61, 1645–1680.
Banz, R.W., 1981. The relationship between returns and market value of common stock. J. Financ. Econ. 9, 3–18.
Barberis, N., Thaler, R., 2003. A survey of behavioral finance.. In: Constantinides, G.M., Harris, M., Stulz, R. (Eds.), Handbook of the Economics of Finance. Elsevier Science B.V., Amsterdam. 1051–1121
Bartram, S.M., Brown, G., Stulz, R., 2012. Why are US stocks more volatile? J. Financ. 67, 1329–1370.
Bartram, S.M., Grinblatt, M., 2018. Agnostic fundamental analysis works. J. Financ. Econ. 128, 125–147.
Bartram, S.M., Grinblatt, M., 2018. Global Market Inefficiencies. Warwick Business School, Coventry, UK, and University of California, Los Angeles, CA.
Bekaert, G., Fratzscher, M., Mehl, A., 2014. Global crises and equity market contagion. J. Financ. 69, 2597–2649.
Bekaert, G., Harvey, C.R., 1995. Time-varying world market integration. J. Financ. 50, 403–444.
Bekaert, G., Harvey, C.R., Lundblad, C.T., Siegel, S., 2011. What segments equity markets? Rev. Financ. Stud. 24, 3841–3890.
Boehmer, E., Duong, T.X., Huszr, Z.R., 2018. Short covering trades. J. Financ. Quant. Anal. 53, 723–748.
Boehmer, E., Huszar, Z.R., Wang, Y., Wang, X., 2017. Are Shorts Equally Informed? A Global Perspective. Singapore Management University, Singapore. Unpublished working paper.
Bris, A., Goetzmann, W.N., Zhu, N., 2007. Efficiency and the bear: short sales and markets around the world. J. Financ. 62, 1029–1079.
Calluzzo, P., Moneta, F., Topaloglu, S., 2019. When anomalies are publicized broadly, do institutions trade accordingly? Manag. Sci. (forthcoming).
Chen, A.Y., Zimmermann, T., 2018. Publication Bias and the Cross Section of Stock Returns. Board of Governors of the Federal Reserve System, Washington, DC. Unpublished working paper.
Chordia, T., Goyal, A., Saretto, A., 2017. p-hacking: Evidence from Two Million Trading Strategies. Emory University, Atlanta, GA. Unpublished working paper.
Chordia, T., Subrahmanyam, A., Tong, Q., 2014. Have capital market anomalies attenuated in the recent era of high liquidity and trading activity? J. Account. Econ. 58, 41–58.
Corwin, S.A., Schultz, P., 2012. A simple way to estimate bid-ask spreads from daily high and low prices. J. Financ. 67, 719–759.
Council of European Union, 2011. Directive 2011/61/EU of the European Parliament and of the Council of 8 June 2011 on alternative investment fund managers and amending directives 2003/41/EC and 2009/65/EC and regulations (EC) no. 1060/2009 and (EU) no. 1095/2010. Off. J. Eur. Union OJ L 174, 1–73.
Da, Z., Gurun, U., Warachka, M., 2014. Frog in the pan: continuous information and momentum. Rev. Financ. Stud. 27, 2171–2218.
Dai, R., 2012. International Accounting Databases on WRDS: Comparative Analysis. Wharton School of the University of Pennsylvania, Wharton Research Data Services, Pennsylvania, PA. Unpublished working paper.
D'Avolio, G., 2002. The market for borrowing stock. J. Financ. Econ. 66, 271–306.
Edelen, R., Ince, O., Kadlec, G.B., 2016. Institutional investors and stock return anomalies. J. Financ. Econ. 119, 472–488.
Engelberg, J., McLean, R.D., Pontiff, J., 2018. Anomalies and news. J. Financ. 73, 1971–2001.
Fama, E., French, K.R., 2008. Dissecting anomalies. J. Financ. 63, 1653–1678.
Fama, E.F., 1998. Market efficiency, long-term returns, and behavioral finance. J. Financ. Econ. 49, 283–306.
Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. J. Financ. Econ. 33, 3–56.
Fama, E.F., French, K.R., 2017. International tests of a five-factor asset pricing model. J. Financial Econ. 123, 441–463.
Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: empirical tests. J. Political Econ. 81, 607–636.
Foerster, S., Karolyi, A., 1999. The effect of market segmentation and investor recognition on asset prices: evidence from foreign stock listing in the US. J. Financ. 54, 981–1013.
Froot, K.A., Dabora, E.M., 1999. How are stock prices affected by the location of trade? J. Financ. Econ. 53, 189–216.
Gagnon, L., Karolyi, G.A., 2010. Multi-market trading and arbitrage. J. Financ. Econ. 97, 53–80.
Green, J., Hand, J.R., Zhang, F., 2017. The characteristics that provide independent information about average US monthly stock returns. Rev. Financ. Stud. 30, 4389–4436.
Green, J., Hand, J.R., Zhang, X.F., 2013. The supraview of return predictive signals. Rev. Account. Stud. 18, 692–730.
Green, J., Hand, J.R.M., Soliman, M.T., 2011. Going, going, gone? The apparent demise of the accruals anomaly. Manag. Sci. 57, 797–816.
Griffin, J.M., 2002. Are the Fama and French factors global or country-specific? Rev. Financ. Stud. 15, 783–803.
Griffin, J.M., Kelly, P.J., Nardari, F., 2010. Do market efficiency measures yield correct inferences? A comparison of developed and emerging markets. Rev. Financ. Stud. 23, 3225–3277.

Hanson, S.G., Sunderan, A., 2014. The growth and limits of arbitrage: evidence from short interest. Rev. Financ. Stud. 27, 1238–1286.
Harvey, C.R., 2017. Presidential address: the scientific outlook in financial economics. J. Financ. 72, 1399–1440.
Harvey, C.R., Liu, Y., Zhu, H., 2016. ...and the cross-section of expected returns. Rev. Financ. Stud. 29, 5–68.
Hau, H., 2011. Global versus local asset pricing: a new test of market integration. Rev. Financ. Stud. 24, 3891–3940.
Henry, P.B., 2003. Capital account liberalization, the cost of capital, and economic growth. Am. Econ. Rev. 93, 91–106.
Herskovic, B., Kelly, B., Lustig, H., Nieuwerburgh, S.V., 2016. The common factor in idiosyncratic volatility: quantitative asset pricing implications. J. Financ. Econ. 119, 249–283.
Hong, H.G., Lim, T., Stein, J.C., 2000. Bad news travels slowly: size, analyst coverage, and the profitability of momentum strategies. J. Financ. 55, 265–295.
Hou, K., Karolyi, G.A., Kho, B.-C., 2011. What factors drive global stock returns? Rev. Financ. Stud. 24, 2527–2574.
Hou, K., Xue, C., Zhang, L., 2015. Digesting anomalies: an investment approach. Rev. Financ. Stud. 28, 650–705.
Hou, K., Xue, C., Zhang, L., 2019. Replicating anomalies. Rev. Financ. Stud. (forthcoming).
Ince, O.S., Porter, R.B., 2006. Individual equity return data from Thomson Datastream: handle with care!. J. Financ. Res. 29, 463–479.
Israel, R., Moskowitz, T.J., 2013. The role of shorting, firm size, and time on market anomalies. J. Financ. Econ. 108, 275–301.
Jacobs, H., 2015. What explains the dynamics of 100 anomalies? J. Bank. Financ. 57, 65–85.
Jacobs, H., 2016. Market maturity and mispricing. J. Financ. Econ. 122, 270–287.
Jacobs, H., Müller, S., 2018. ... and Nothing Else Matters? On the Dimensionality and Predictability of International Stock Returns. University of Duisburg-Essen, Essen, Germany. Unpublished working paper.
Jegadeesh, N., Kim, J., Krische, S.D., Lee, C.M., 2004. Analyzing the analysts: when do recommendations add value? J. Financ. 59, 1083–1124.
Jones, C.M., Reed, A.V., Waller, W., 2016. Revealing shorts: an examination of large short position disclosures. Rev. Financ. Stud. 29, 3278–3320.
Karolyi, G.A., 2016. Home bias, an academic puzzle. Rev. Financ. 20, 2049–2078.
Keloharju, M., Linnainmaa, J.T., Nyberg, P., 2016. Return seasonalities. J. Financ. 71, 1557–1590.
La Porta, R., de Silanes, F.L., Shleifer, A., Vishny, R.W., 1997. Legal determinants of external finance. J. Financ. 52, 1131–1150.
Lee, C.M.C., Swaminathan, B., 2000. Price momentum and trading volume. J. Financ. 55, 2017–2069.
Lewellen, J., 2011. Institutional investors and the limits of arbitrage. J. Financ. Econ. 102, 62–82.
Lewellen, J., 2015. The cross section of expected stock returns. Crit. Financ. Rev. 4, 1–44.
Linnainmaa, J.T., Roberts, M.R., 2018. The history of the cross section of stock returns. Rev. Financ. Stud. 31, 2606–2649.
Lo, A.W., MacKinlay, A.C., 1990. Data-snooping biases in tests of financial asset pricing models. Rev. Financ. Stud. 3, 431–467.
Lu, X., Stambaugh, R.F., Yuan, Y., 2018. Anomalies Abroad: Beyond Data Mining. Shanghai Jiao Tong University, Shanghai, China. Unpublished working paper.
McLean, R.D., Pontiff, J., 2016. Does academic research destroy stock return predictability? J. Financ. 71, 5–32.
McLean, R.D., Pontiff, J., Watanabe, A., 2009. Share issuance and cross-sectional returns: international evidence. J. Financ. Econ. 94, 1–19.
Novy-Marx, R., Velikov, M., 2016. A taxonomy of anomalies and their trading costs. Rev. Financ. Stud. 29, 104–147.
Penasse, J., 2018. Understanding Alpha Decay. University of Luxembourg, Luxembourg City, Luxembourg. Unpublished working paper.
Pontiff, J., 1996. Costly arbitrage: evidence from closed-end funds. Q. J. Econ. 111, 1135–1151.
Pontiff, J., 2006. Costly arbitrage and the myth of idiosyncratic risk. J. Account. Econ. 42, 35–52.
Rapach, D.E., Strauss, J.K., Zhou, G., 2013. International stock return predictability: what is the role of the United States? J. Financ. 68, 1633–1662.
Richardson, S., Tuna, I., Wysocki, P., 2010. Accounting anomalies and fundamental analysis: a review of recent research advances. J. Account. Econ. 50, 410–454.
Schwert, W.G., 2003. Anomalies and market efficiency. In: Constantinides, G.M., Harris, M., Stulz, R. (Eds.), Handbook of the Economics of Finance. Elsevier Science B.V., Amsterdam, pp. 937–972.
Shleifer, A., Vishny, R.W., 1997. The limits of arbitrage. J. Financ. 52, 35–55.
Stambaugh, R.F., Yu, J., Yuan, Y., 2012. The short of it: investor sentiment and anomalies. J. Financ. Econ. 104, 288–302.
Stambaugh, R.F., Yu, J., Yuan, Y., 2014. The long of it: odds that investor sentiment spuriously predicts anomaly returns. J. Financ. Econ. 114, 613–619.
Stambaugh, R.F., Yu, J., Yuan, Y., 2015. Arbitrage asymmetry and the idiosyncratic volatility puzzle. J. Financ. 70, 1903–1948.
Stambaugh, R.F., Yuan, Y., 2017. Mispricing factors. Rev. Financ. Stud. 30, 1270–1315.
Subrahmanyam, A., 2010. The cross section of expected stock returns: what have we learnt from the past 25 years of research? Eur. Financ. Manag. 16, 27–42.
Tobek, O., Hronec, M., 2018. Does it Pay to Follow Anomalies Research? Machine Learning Approach with International Evidence. University of Cambridge, Cambridge, UK. Unpublished working paper.
Tobek, O., Hronec, M., 2018. Does the Source of Fundamental Data Matter?. University of Cambridge, Cambridge, UK. Unpublished working paper.
Wahal, S., 2019. The profitability and investment premium: pre-1963 evidence. J. Financ. Econ. 31, 362–377.
Yan, X., Zheng, L., 2017. Fundamental analysis and the cross section of stock returns: a data-mining approach. Rev. Financ. Stud. 30, 1382–1423.