Global Factor Premiums

Guido Baltussen*

Erasmus School of Economics, Erasmus University Rotterdam and Robeco Institutional Asset Management

Laurens Swinkels

Erasmus School of Economics, Erasmus University Rotterdam and Robeco Institutional Asset Management

Pim Van Vliet

Robeco Institutional Asset Management

Abstract

We examine 24 global factor premiums across equity, bond, commodity and currency markets via replication and out-of-sample evidence between 1800 and 2016. Replication yields ambiguous evidence within a unified testing framework that accounts for p-hacking. Out-of-sample tests reveal strong and robust presence of the large majority of global factor premiums, with limited out-of-sample decay of the premiums. We find global factor premiums to be generally unrelated to market, downside, or macroeconomic risks in the 217 years of data. These results reveal significant global factor premiums that present a challenge to traditional asset pricing theories.

Forthcoming Journal of Financial Economics

This Version: December 2020

Keywords: factor premium, p-hacking, return anomalies, predictability, seasonality.

JEL classification: C11, C12, F31, G11, G12, G15, N20

*Corresponding author: Guido Baltussen, Finance Group, Erasmus School of Economics, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3000 DR Rotterdam, the Netherlands. Tel: +31 - 10 224 1438. Email: baltussen@ese.eur.nl. We would like to thank an anonymous referee for his great contributions in improving the paper. We also thank David Blitz, Andrea Frazzini, William Goetzmann, Winfried Hallerbach, Matthias Hanauer, Matti Keloharju, Ralph Koijen, Peter Koudijs, Martin Martens, Lasse Pedersen, Gertjan Verdickt, Jeroen van Zundert and seminar participants from the Research in Behavioral Finance Conference (Amsterdam), Hanken School of Economics, Maastricht University, Oslo Business School, Southwestern University of Finance and Economics, Technical University Munich, and Vaasa University for valuable discussions and useful feedback. The views expressed in this study are not necessarily shared by Robeco Institutional Asset Management. We welcome comments, including references to related papers we have inadvertently overlooked.

Electronic copy available at: https://ssrn.com/abstract=3325720

I. Introduction

In this paper we study global factor premiums over a deep sample spanning 217 years across equity index (but not single securities), bond, currency, and commodity markets. We focus on the six main global factor premiums that have received considerable attention in the literature: time-series momentum (henceforth 'trend'), cross-sectional momentum (henceforth 'momentum'), value, carry, return seasonality and betting-against-beta (henceforth 'BAB'). Several recent influential studies document the significant presence of these factor premiums across the major asset classes. Moskowitz, Ooi, and Pedersen (2012) show the presence of a trend premium; Asness, Moskowitz, and Pedersen (2013) reveal momentum and value premiums; Koijen, Moskowitz, Pedersen, and Vrugt (2018) find a carry premium; Keloharju, Linnainmaa, and Nyberg (2016) discover a return seasonality premium, and Frazzini and Pedersen (2014) document a BAB premium. The sample periods in these studies on average start in 1980 and extend earlier empirical asset pricing studies which often focus on a single asset class, usually U.S. equities.[1]

We work from the idea that these published factor premiums could be influenced by 'p-hacking' (see Harvey, 2017). Scientists are faced with statistical limitations of empirical tests, typically have to choose between several degrees of freedom in research (in areas like data manipulation, statistical method, aggregation schemes, etc.), but at the same time tend to have an incentive to publish. Consequently, published findings might reflect a combination of a type I error in testing (i.e. falsely discovering predictability), a multiple hypothesis testing bias, or a publication bias.[2] As a case in a point, Harvey, Liu, and Zhu

[1] Several papers show the presence of factor premiums for individual asset classes. For equities, 'value', 'momentum', and 'low-risk' are well documented (e.g. Fama and French 1993, Jegadeesh and Titman, 1993, Blitz and Van Vliet, 2007). For currencies, Froot and Thaler (1990) and Barroso and Santa-Clara (2015) document a 'carry' factor, Menkhoff et al. (2012) document a 'momentum' factor, while Abuaf and Jorion (1990) and Menkhoff et al. (2017) document a 'value' factor. For commodities, Erb and Harvey (2006) document the 'momentum' factor. For bonds, Fama (1984) documents the term premium, also referred to as 'carry', and Ilmanen (1995) documents 'value' and 'carry' factors.

[2] As Harvey (2017) notes: "*Given the competition for top journal space, there is an incentive to produce "significant" results. With the combination of unreported tests, lack of adjustment for multiple tests, and direct and indirect p-hacking, many of the results being published will fail to hold up in the future.*" P-hacking is not limited to financial economics. For example, Begley and Ellis (2012) show that out of 53 studies on pre-clinical cancer only 11% could be replicated. An open science collaboration (2015) shows that out of 97

Electronic copy available at: https://ssrn.com/abstract=3325720

(2016) find evidence supporting a strong publication bias in the top finance journals, with many of the more than 300 documented stock-level anomalies becoming questionable after analyzing these in a rigorous testing framework that accounts for a multiple hypotheses testing bias.[3]

The first objective of this study is to examine global factor premiums from the perspective of p-hacking in a robust and rigorous way. We take as starting point the six main factor premiums across the four major asset classes (equity indices, government bonds, commodities and currencies), hence resulting in a total of 24 global return factors.[4] Figure 1, Panel A shows the Sharpe ratios of each global factor premium as reported in their original publications, as well as the 5% significance cutoff (the grey-colored dashed line). In general, the studies show evidence on the global factor premiums at the conventional 5% significance level.

INSERT FIGURE 1 HERE

Next, we turn to methods proposed against p-hacking. First, the black-colored dashed line in Figure 1 shows the 3.00 t-value cutoff advocated by Harvey, Liu, and Zhu (2016) and others.[5] Second, we apply the Bayesian perspective on p-values advocated by Harvey (2017), using a symmetric and descending minimum Bayes factor with 'Perhaps' prior

significant psychological studies only 36 could be replicated. In behavioral economics, Camerer et al. (2016) find that out of 18 laboratory studies only 11 can be replicated.

[3] In a similar spirit, Hou, Xue, and Zhang (2020) find that 64% (85%) of almost 450 documented anomalies have t-statistics below two (three) when the importance of small and micro capitalization stocks is reduced. Chordia, Goyal, and Saretto (2019) show with a data mining approach that of about 2.1 million possible trading strategies only a small group survives after correcting for a multiple hypothesis testing bias, or as they state after using the "proper statistical hurdles". Moreover, the few surviving trading strategies seem to have no apparent theoretical underpinning. Further, Linnainmaa and Roberts (2018) show that when tested out-of-sample, many equity anomalies are weak and those that persist do so typically at about half of their original size.

[4] In this paper we do not include individual stocks or other company-level securities.

[5] Harvey, Liu, and Zhu's (2016) suggestion of increasing the t-statistic to 3 is thus very similar to the proposal of Benjamin et al. (2018), who propose to redefine statistical significance across disciplines from the usual arbitrary p-value of 0.05 to an equally arbitrary, but stricter p-value of 0.005, thereby essentially increasing the t-statistic 'hurdle' from 1.96 to 2.81.

Electronic copy available at: https://ssrn.com/abstract=3325720

odds ratio 4:1. The numbers above each bar in Figure 1 indicate these 'Bayesianized' p-values. Using these approaches, we find fewer significant global factor premiums.

Further, most of the studies have differences in, amongst others, testing methodologies, investment universes and sample periods, choices that introduce degrees of freedom to the researcher. To mitigate the impact of such degrees of freedom, we perform a replication as defined by Welch (2019). We re-examine the global factor premiums using uniform choices on testing methodology and investment universe, while we use each asset class-factor specific 'in-sample' period as used in the original paper. Note that for uniformity this now includes testing return seasonality for bonds and currencies, which was not done in the original publication. Figure 1, Panel B shows the results of the replicating exercise. We find Sharpe ratios to remain substantial, but statistical evidence to become ambiguous.

One downside of the Bayesian approach mentioned above is the subjectivity required in the formulation of prior odds, subjectivity that typically has a substantial impact on the Bayesian p-value. As an alternative, we introduce a novel perspective on p-hacking: the 'break-even' prior odds, or those prior odds at which the Bayesian p-value would equal the chosen confidence level. The break-even prior odds remove the need to specify the prior odds, and allow for an interpretation of prior odds that would be required to just accept the alternative. Applying this concept to the replication exercise reveals that one does not need to be very skeptical to disregard the empirical evidence provided in the literature.

The second objective of this study is to provide rigorous out-of-sample evidence on the global factor premiums. To this end, we construct a deep and largely unexplored historical global database on the global return factors in the four major asset classes. This consists of 'pre-sample' data starting in 1800, such that we have an extensive new sample that includes on average more than 150 years of new and independent data to conduct further

Electronic copy available at: https://ssrn.com/abstract=3325720

analyses. If the global factor premiums discovered in the existing literature were the result of p-hacking, we would expect them to disappear for this new sample.

Our pre-sample findings reveal consistent and ubiquitous evidence for the large majority of global factor premiums, as summarized in Figure 1, Panel C. In terms of economic significance, the Sharpe ratios in the pre-sample period are substantial, with an average of 0.40 across the 24 asset class-factor combinations. Remarkably, we see no significant 'out-of-sample' decay of factor premiums, as the average in-sample Sharpe ratio in Figure 1, Panel B is 0.41. Further, the break-even prior odds reveal that one needs to be extremely skeptical to disregard the pre-sample evidence on global factor premiums. As main exceptions, we find currency value and the BAB effect outside equity markets to be weak.[6] Return seasonality in government bonds and currencies are significant factor premiums and are an extension of the empirical asset pricing literature. Figure 1, Panel D reveals a similar picture over our full sample period, which spans 217 years of data from 1800 to 2016. Further, we show these results to be generally robust across time and testing choices.

The third objective of this study is to gain further insights into the economic explanations of the global factor premiums using our full sample period of 217 years of data. We first test the uniqueness of the factor premiums. Our findings reveal that most factors are largely uncorrelated and do not span each other. Noteworthy, value factor premiums become generally higher after controlling for other factors. These findings suggest that it is difficult to formulate a single uniform explanation based on a common global component.

Next, we test if the global factor premiums can be reconciled with market, downside, or macroeconomic risks. The theory suggests that expected returns can vary due to market

[6] These findings are in line with Frazzini and Pedersen (2014), who report strong results for BAB in equities but weak evidence for BAB in currencies, the cross-section of country bonds, and commodities. Frazzini and Pedersen (2014) do report strong evidence for BAB applied to bonds with different maturities *within* a single country (i.e. across the curve), a finding we confirm over an extended sample in Online Appendix D.

Electronic copy available at: https://ssrn.com/abstract=3325720

risk, downside risk (Bawa and Lindenberg, 1977) or macroeconomic risk (Chen, Roll, and Ross, 1986, Fama and French, 1989, Ferson and Harvey, 1991).[7] The typically studied recent period might be biased, given that they were quite favorable historically (no major wars, growing global prosperity and only a few large recessions or periods of social unrest), which severely limits the number of 'bad states'. Our extensive multi-century sample includes a substantial number of bad states; with, for example, many observations on bear markets (43 years) and recessions (74 years). It therefore allows us to more deeply examine risk-based explanations. Across several tests we find no supporting evidence for these explanations, with global return factors bearing basically no relationship to market, downside or macroeconomic risks.

The remainder of this paper is structured as follows: Section II describes the results of the original studies and our replication exercises from an economic and p-hacking perspective. It further introduces the novel Bayesian concept of break-even prior odds. In Section III, we introduce our historical dataset and and evaluate global factor returns out-of-sample since 1800. Section IV tests economic explanations of global factor premiums using the full depth of our sample. Section V presents our conclusions. The appendices contain more details about the data and portfolio construction methods, as well as several additional results.

II. Replicating and p-hacking insights on the recent sample

We start our study by summarizing the main results and testing choices made in the original studies, followed by replication tests. We incorporate economic, traditional

[7] Asness, Moskowitz and Pedersen (2013) show that global macroeconomic variables are generally not related to value and momentum returns, while Keloharju, Linnainmaa, and Nyberg (2016) find that seasonality returns are hard to reconcile with macroeconomic risks. Koijen et al. (2018) report that carry return drawdowns are more likely to occur during global recession periods.

Electronic copy available at: https://ssrn.com/abstract=3325720

statistical, and Bayesian perspectives that account for p-hacking. Further, we introduce a novel perspective on p-hacking: the concept of break-even prior odds.

A. Original studies

We base our paper on five key studies that document global factor premiums across the major asset classes: trend, momentum, value, carry, return seasonality, and BAB. Table 1 summarizes the main findings of, and testing choices made by these studies for the four asset classes that we consider: international equity indices, government bond country indices, commodities, and currencies.[8] The table further contains the definitions used for each factor within each asset class, the reported statistical and economic significance, the sample period, the portfolio construction method, and the number of assets within each asset class.

INSERT TABLE 1 HERE

As becomes clear from Table 1, these studies report evidence supporting the global factor premiums across the four major asset classes we study. The Sharpe ratios for each of the 22 global return factors are positive (return seasonality is not tested in bonds and currencies), with many above 0.30, while t-values are above 1.96 for 14 out of 22 factors. In particular, carry and trend have high t-values for each individual asset class, corresponding to high conventional levels of significance. Value is relatively weaker, with t-values below 1.96 for three out of four asset classes.

In addition, although the studies have many similarities, several degrees of freedom in testing are accommodated differently (driven by sample selection choices, data limitations, etc.). For example, the sample start dates for bonds range from January 1982

[8] Note that we choose to focus on global markets, and not include individual stocks or other company-level securities. Further, we ignore asset classes that are included in one or two of the key studies, but lack deeper historical data, such as credit and option markets.

Electronic copy available at: https://ssrn.com/abstract=3325720

to July 1989. Further, although the portfolio construction method for trend strategies is required to be different from the other factors, for the cross-sectional factors there are also two different methods: forming equally weighted portfolios for the top and bottom terciles and comparing the returns between these two portfolios, or taking a position proportional to the rank of the asset in the cross-section, with long positions for assets above the median, and short positions for those below the median. Moreover, the investment universe within each asset classes varies across studies.

B. Replication

Next, we replicate these original studies over the sample period used in the original studies in a unified testing framework in which we limit degrees of freedom. More specifically, we utilize (i) a uniform cross-section of assets, and (ii) a uniform factor construction method for the cross-sectional factors. Note that we follow Welch's (2019) classification of replication and do not aim to exactly reproduce the results of the original studies, but instead perform a looser replication that uses similar data and testing choices across factors. Our examination differs from the original studies in aspects like data sources, included assets, and portfolio construction methodology, but all within variations that are expected to be unimportant and reasonable. Consequently, our results will not be identical to the original studies but are expected to differ, but only slightly, from the original studies.

Within each asset class we construct factor portfolios on each of the six factors, with the definitions of factors and their motivations mostly following the original studies (the Appendix contains detail on all measures, sample and other testing choices). In summary, trend and momentum are defined as the 12-month-minus-1-month excess return. Value is defined as dividend yield for equity indices, real yield for bonds, five-year reversal in spot prices for commodities, and absolute and relative purchasing power parity for

Electronic copy available at: https://ssrn.com/abstract=3325720

currencies. Carry is defined as the implied yield on each instrument. Return seasonality is defined as the return on an asset in a certain month over the prior 20 years. BAB is long the low beta assets and short the high beta assets with positions neutralized for the ex-ante beta, with beta measured relative to the global asset class portfolio. Novy-Marx and Velikov (2018) report that the size of the BAB factor premium in the equity market is partially driven by a large weight to micro-cap stocks. This is less of a concern for our study, as our assets are at the market level and hence not at the individual stock level.

We construct factor investment portfolios at the end of every month in the spirit of the original papers. For the trend factor, which is directional in nature, we go long (short) markets in each asset class when the trend measure is positive (negative), following Moskowitz, Ooi, and Pedersen (2012). For the other factors, which are all cross-sectional in nature, we rank the markets in each investment universe based on the factor measure and take a position equal to the rank minus its cross-sectional average. This procedure is similar to that used by Asness, Moskowitz, and Pedersen (2013), Frazzini and Pedersen (2014), and Koijen et al. (2018). Further, we scale all positions and factors to a constant ex-ante volatility or by the ex-ante beta (the latter only in the case of BAB). All returns we consider are in excess of local financing rates and expressed in U.S. dollars. Our replication dataset is a collection from Bloomberg, Datastream and the OECD website. The Appendix and Online Appendix A describe the data, factor construction, and sample choices in extensive detail.

With these uniform factor definitions and portfolio construction rules, we replicate the global factor premiums documented in the literature for the original 'in-sample' periods.[9]

[9] A remark is in order about our choice of starting dates, as several factors have been studied before with earlier starting dates within the cross-section of stocks. For example, the value factor for U.S. stocks as documented in Fama and French (1993) starts in 1963, compared to 1972 for value within equity indices in Asness, Moskowitz, and Pedersen (2013). Arguably, this raises the question of whether the sample periods of the original papers are the best proxy for the in-sample period of each asset class-factor combination. As factors across asset classes are generally little related (see for example Section IV), we choose to focus on

Electronic copy available at: https://ssrn.com/abstract=3325720

Note that for uniformity this now also includes testing return seasonality for bonds and currencies, which was not done in the original publication.

INSERT TABLE 2 HERE

The results of the replication exercise are displayed in Table 2. We find that the Sharpe ratios are lower, but remain substantial, as compared to the original studies. The average Sharpe ratio across all 24 asset class-factor combinations is 0.41, and 11 of the 24 factor premiums are significant at the conventional 5% level. The seasonal strategy applied to bond and currency markets, which was not reported in the original paper, is not significant during the period January 1975 to December 2011, the period used by Keloharju, Linnainmaa, and Nyberg (2016) for equity indices and commodities. Moreover, six of the global factor premiums are significant using a t-value of 3.00 as cut-off. We further verify that results are robust and stand up to several variations in testing settings presented in Table 1, as shown in Online Appendix Table B.1.

One could argue that another way to limit sensitivity to p-hacking is to combine information across asset classes. An objection against such a multi-asset aggregation of evidence is that the factor premiums tend to be lowly correlated across asset classes, as we will show in Section IV. Nevertheless, we also construct equal-volatility-weighted global factor portfolios per asset class and across asset classes ('Multi Asset') by targeting each market, then asset class portfolio, and then multi-asset portfolio at a 10% ex-ante annual volatility and applying equal weighting. Following this procedure, the Sharpe ratios of the six multi-asset combinations vary between 0.38 and 1.07, and the t-values

starting dates reported in the original papers for each specific asset class, as we presume that using these dates is closest related to possible data mining.

Electronic copy available at: https://ssrn.com/abstract=3325720

are above 1.96 for each factor. A t-value threshold of three is not reached for value, seasonality and BAB, leaving three of the six multi-asset factors significant at this cutoff.

C. Statistically accounting for p-hacking

Next, we examine traditional p-values (which do not correct for multiple hypotheses testing bias), Bayesianized p-values advocated by Harvey (2017), and the break-even prior odds. The critique on frequentist p-values is that they indicate how likely it is to observe the data under the assumption that the null hypothesis is true. However, we are typically more interested in knowing the probability that the null hypothesis is true given that we observed the data. This question can be answered by a Bayesian approach. However, a full-blown Bayesian analysis, in which we specify priors on all possible hypotheses and calculate posterior probabilities given the observed data for inference, can be challenging. An alternative to such full-blown Bayesian analysis is to use the Minimum Bayes Factor (MBF; Edwards, Lindman, and Savage, 1963). The Bayes Factor connects the prior odds (before having seen the data) with the posterior odds (after having seen the data). The MBF gives maximum advantage to the alternative hypothesis, and therefore represents the maximum amount the data can move the posterior odds away from the prior odds. The MBF is global in the sense that it accepts all possible alternative hypotheses as potential candidates.

A less 'aggressive' approach in rejecting the null hypothesis would pose some structure on the relation between the null and alternative hypothesis. Harvey (2017) argues that for many finance applications, it seems reasonable to assume that the null hypothesis has the highest prior probability, and that prior probabilities of alternative hypotheses decrease symmetrically around the null. This implies that the same deviation below or above the null hypothesis is assumed to be equally likely. The assumption of additional structure gives less advantage to the alternative compared to the MBF, where the prior alternative exactly coincides with the observed data, and hence is the more conservative

Electronic copy available at: https://ssrn.com/abstract=3325720

choice in the context of this study (i.e. putting a higher hurdle on rejecting the null hypothesis of zero factor premiums). The MBF results more often in rejections of the null hypothesis. A disadvantage of using the symmetric and descending MBF (SD-MBF) is that when prior probabilities are not symmetric and descending around the null hypothesis, they might not reject the null hypothesis quickly enough, and hence result in a lower testing power. For example, even if there is a competing theory implying a positive parameter instead of zero for the standard theory, the SD-MBF assumes that negative deviations from zero are equally likely as positive deviations, unjustly increasing the Bayesian p-value.

Assuming the SD prior probabilities around the null hypothesis, we obtain an SD-MBF (Bayarri and Berger 1998)[10], defined as:

$$\text{SD-MBF} = -\exp(1) \times \text{p-value} \times \ln(\text{p-value}) \tag{1}$$

Consequently, we can transform the frequentist p-value to a Bayesian p-value using a level of prior odds of the null being true relative to an alternative:

$$\text{Bayesian p-value} = \text{SD-MBF} \times \text{prior odds} / (1+\text{SD-MBF} \times \text{prior odds}) \tag{2}$$

For example, for a situation in which the null and alternative are equally likely (i.e. even prior odds), the frequentist p-value of 5% is transformed into a Bayesian p-value of 29%. When the alternative is a 'long shot' with prior odds 99-to-1, the frequentist p-value of 5% will be transformed into a Bayesian one of 98%. In other words, if we are skeptical about the alternative, more convincing data (a lower frequentist p-value) is required to change our minds than if the alternative is more easily conceivable.[11]

[10] The MBF is $\exp(-Z^2/2)$ for a normal approximated z-score Z. This means that for a frequentist p-value of 5% the MBF equals 0.15, while this is 0.41 for the SD-MBF. The latter, which we use throughout our paper, is clearly more conservative as it leads to higher Bayesian p-values.

[11] As also stressed by Harvey (2017), the Bayes factor we employ is in principle a first hurdle to filter effects that are highly unlikely to be true. As a next step one could argue that a more conservative Bayes factor is

Electronic copy available at: https://ssrn.com/abstract=3325720

An important difficulty in the Bayesian approach is the assumption on the level of prior odds. This choice is quite subjective, while it typically has a material impact on the posterior p-value. As an alternative, we can think in terms of the break-even prior odds, or that level of prior odds at which the Bayesian p-value equals the chosen significance level α. When we rewrite the equation above, we obtain the break-even prior odds as a function of α and the frequentist p-value (represented by the SD-MBF):

$$\text{Break-even prior odds} = \alpha / ((1 - \alpha) * \text{SD-MBF}) \tag{3}$$

Supposing we found a frequentist p-value of 0.1%, and want the α to be 5%, this implies a break-even prior odds ratio of 2.8:1. If we are *a priori* more skeptical about the alternative than this, the frequentist p-value of 0.1% (t-statistic of 3.30) does not provide enough evidence to change our minds.

INSERT TABLE 3 HERE

Table 3 shows the results of applying these Bayesian concepts to our replication exercise. For each global return factor, we report the frequentist p-values and the Bayesian p-values with a prior odds ratio of 4:1; prior odds classified by Harvey (2017) as 'perhaps' (see also Figure 1 for the results of both the original and replication study). We find that only four of the 24 factor premiums, and three of their six multi-asset combinations have Bayesian p-values below 5%. Next, turning to the break-even odds, we find (obviously) the same number of global return factors to have prior odds above 4:1. Turning to more conservative prior odds, we find that only two exceed 99 – the prior odds labeled as a 'long shot' by Harvey (2017). These results imply that one does not need to be very skeptical to disregard the empirical evidence.

needed, or even an explicit alternative. However, Bayarri, Benjamin, Berger, and Sellke (2016) show that the SD-MBF is generally very close to the full Bayes factor when p-values are low (as in our case).

Electronic copy available at: https://ssrn.com/abstract=3325720

III. Global factor premiums: evidence since 1800

Next, in order to more thoroughly examine the evidence for global return factors we study a deep, independent and largely unexplored historical sample that was not used for their original discovery.[12] Using freshly uncovered historical data for cross-validation is advocated by, for example, Arnott, Harvey, and Markowitz (2019) to safeguard against data mining.

A. Historical database construction

We compile our data from several sources in order to obtain a reliable and historically extensive dataset. Our sample covers 217 years of data from 31 December 1799 through 31 December 2016. We obtain the most recent historical data on financial market prices and macroeconomic series from Bloomberg, Datastream and the OECD, and splice these before inception with data from (in order of preference): Global Financial Data or monthly commodity futures data from Chicago Board of Trade (CBOT) annual reports (1877-1962) obtained from TwoCenturies.com, the Jordà-Schularick-Taylor Macrohistory Database[13], and/or Jeremy Siegel's website.[14] Our dataset construction for each asset and measure is described in extensive detail in Online Appendix A. In short, we use futures data for our return series where available and splice these with excess index-level returns for all asset classes except commodities before the availability of futures data. For commodities, we only use futures data, as spot returns can materially differ from futures returns for this asset class. Global Financial Data has been used in various studies over the last years, in order to, amongst others, study equity risk premiums and their dynamics, but also

[12] This study is not the first to utilize deep historical samples to study asset pricing. For equity and bond premiums, Siegel (1992) gives evidence stretching back to 1800, Goetzmann (1993) to 1695, and Golez and Koudijs (2018) go even further back to 1629. Others study premiums of a single factor. Hurst, Ooi, and Pederson (2017) find a persistent trend premium going back to 1880, Geczy and Samonov (2017) study momentum in a 215-year sample, while Goetzmann and Huang (2018) and Geczy and Samonov (2016) show that stock-level momentum worked in imperial Russia in the period 1865-1914 and the U.S. pre-CRSP period. Further, Doskov and Swinkels (2015) and Taylor (2002) find that carry and value premiums are present in currency markets since 1900.

[13] http://macrohistory.net. See Jordà, Schularick and Taylor (2017) for more information.

[14] http://www.jeremysiegel.com/. The webpage is no longer active, but is archived here.

Electronic copy available at: https://ssrn.com/abstract=3325720

volatilities and predictability (see Albuquerque et al. 2015; Barro and Ursúa, 2008, 2017; Danielsson, Valenzuela, and Zer, 2018; Goetzmann and Kim, 2018; Goetzmann, Li, and Rouwenhorst, 2005; Hjalmarsson, 2010; Muir, 2017; Schwert, 2011, and Zhang and Jacobsen, 2013). In this paper we utilize these databases to test for the existence of global factor premiums. Tables A.1-A.3 in the Online Appendix summarize the markets included in our sample, the excess return series, and the start dates of the factor measures per market.[15] On average, our sample includes more than 150 years of new and independent data per asset class-factor series.

Even though we (and the data vendors) have paid close attention to data quality, the deep historical data tends to be of lesser quality compared to the more recent data, as digital archives and the use of indices with strong requirements on data processes did not exist. Instead, data was maintained typically by exchanges, statistical agencies, newspapers and investor annuals, often in manual writing. Potential data quality issues that could be at work include (manual) misprints and other measurement errors, but also indices consisting of just a handful of stocks, the use of old data (for example, bond yield based on padded values if no yield for a particular months was found), the use of time-averaged prices over a month (often average of the lowest and highest monthly prices), and the timing of dividends or coupons sometimes being unknown but assigned at quarter or year ends.

Lesser data quality could influence our tests in a number of ways. On the one hand, it could create random measurement errors in our data, thereby, biasing our results towards the null hypothesis that a global factor premium does not exist. On the other hand, if biases in the data correlate with factor premiums, they could create spurious results. For example, Schwert (1990) shows that the use of average of high and low prices

[15] Note we have a limited number of assets per asset class in roughly the first 50 years of our sample period, making it more difficult to detect the existence of global factor premiums. Even though the average returns need not necessarily be affected, the variation around the average is probably higher due to limited diversification benefits in the factor portfolios.

Electronic copy available at: https://ssrn.com/abstract=3325720

over a month generates an artificial AR(1) process in the return series. Further, measurement errors could cause prices to be spuriously inflated, trigger potential reversal (value) profits. And, assignment of dividends or coupons to year ends could generate spurious seasonality in returns centered in these periods.

To construct a high-quality dataset, we have taken the following steps. First, we have checked and corrected each data series on potential data errors as outlined in detail in the Appendix, Section A.III. Second, we have verified our data sources, where possible, against each other and found that average returns and volatilities are generally of comparable magnitudes across databases. The same applies for subsamples statistics, as shown in Table A.1 in the Online Appendix. Figures A.1-A.4 in the Online Appendix show the behavior of each return series we use over time with their sample inclusion. Third, we apply a number of conservative screens on our data series and remove data points when they do not pass these screens.

These screens include (i) a 'zero return screen' – leaving out data series with more than one zero or missing spot return observations in the past 12 months, (ii) a 'return interpolation screen' – leaving out identical returns one month to month, and (iii) a 'stale return screen' – leaving out observation which do not have nine or more differentiating returns over the past 12 months. The first screen filters for data historically available at a non-monthly frequency, reduced liquidity, and also captures non-tradability due to currency pegs, as assets with lower liquidity or no trades are more likely to have zero returns. Lesmond, Ogden, and Trzcinka (1999) show that the number of periods with zero returns is an efficient proxy for liquidity. The second screen filters an unlikely return pattern, exactly identical consecutive monthly returns, which indicate return interpolation. The third screen filters returns not updated at the monthly frequency. To this end, we remove an asset at each point in time when over the past 12 months there are less than nine unique monthly returns when rounded to five basis points. We have

Electronic copy available at: https://ssrn.com/abstract=3325720

simulated that such a pattern is unlikely under a normal or the in-sample return distribution for all asset classes in our universe.[16] Further, we always skip a month between our trend or momentum signals and investing, which removes possible spurious autocorrelation at the monthly frequency. Please note that these screens mitigate data quality concerns, but could also bias factor premium estimates downwards if they remove correct data points. In the robustness section we discuss the impact of these various screens, as well as other robustness tests to data quality.

Due to problems with exceptionally high levels of data uncertainty, we next exclude hyperinflation episodes from our sample by excluding assets from countries with ex-ante levels of monthly inflation over 50%. This definition follows Cagan (1956). We only start investing 12 months after the hyperinflation period has ended, as this is real-time available information. This definition of hyperinflation affects Germany during the 1920-1926 period; France in 1920-1921; France, Italy, and Japan during the post-World War II years (1946-1949/1950 and 1943-1948 for Italy); and South Korea in 1950-1956.

B. Global factor premiums since 1800

The premiums of each global return factor, in terms of Sharpe ratios, over the 'pre-sample' period are displayed in Panel A of Table 4. In this deep historical sample, we find a strong presence of most global return factors. In terms of economic significance, we find an average Sharpe ratio of 0.40, while in terms of statistical significance we find that 18 of the 24 t-values are above 1.96 and 17 are above 3.00. Note that one can wonder whether the t-value cut-off of 3.00 is required, as we conduct a pure out-of-sample study for which multiple testing becomes less of a concern.

[16] More specifically, we have randomly drawn 10,000 observations (with replacement) from the normal distribution with mean and volatility equal to the equal-weighted average asset class values over the 1800-1971 period (see Online Appendix Table A.1), or from the empirical in-sample distribution, and examined the occurrence of this screen. For bonds the screen is triggered for less than 2% of the observations, while for all other asset classes the number is below 0.5%.

Electronic copy available at: https://ssrn.com/abstract=3325720

INSERT TABLE 4 HERE

Panel B of Table 4 exploits the full power of our sample by studying the full sample results, which spans 217 years of data from 1800 until 2016.[17] (Note that this includes a limited number of observations on 'post-sample' data until 2016.) In general, results are stronger as measured by t-values because of the additional three decades with positive (although not always statistically significant) returns. The average Sharpe ratio across all global return factors equals 0.40, and 20 (17) of the 24 global factor premiums have t-values above 1.96 (3.00). Over both the new sample and the full sample, trend, momentum, carry, and seasonality are generally the strongest. For each of the four asset classes, these factors have Sharpe ratios above 0.20 and t-values well above three, except for commodity carry which has a t-value of 2.94. Return seasonality in government bonds and currencies are strong factor premiums and are an extension of the empirical asset pricing literature. For value, we find statistically and economically significant positive returns for three out of four asset classes, with a t-value of 3.21 for equity value, 2.41 for bond value and 2.76 for commodity value. Value for currencies is weak with t-values in the pre-sample (full sample) of -0.02 (1.03). The results for BAB reveal that the effect is strongly present for equity indices (and within the U.S. bond term structure, see Online Appendix D), but not significant for country bond indices, commodities or currencies.[18]

[17] By including the post-sample we maximize the power of our tests that utilize the full sample period. Because of a relatively limited number of observations we do not analyze the post-sample period separately.

[18] Our results for the BAB factor generally align with Frazzini and Pedersen (2014) over the replication sample period. Table 2 reveals that, over the replication sample, BAB has an annualized Sharpe ratio of 0.34 for equities, 0.57 for country bonds, 0.30 for commodities and 0.13 for currencies. This compares to 0.51, 0.14, 0.11, and 0.22, respectively, documented in Frazzini and Pedersen (2014). Furthermore, Frazzini and Pedersen (2014) find substantially stronger results for BAB applied to bonds across the curve within a single country (U.S.). In a separate U.S. bond sample covering various maturities, they find a Sharpe ratio of 0.81. We examine BAB on the U.S. bond curve (bond data of various maturities is generally lacking deep history outside the U.S.) in Online Appendix D across 1-year to 30-year U.S. government bonds between 1922 and 2016. Over this period we find a Sharpe ratio of 0.37 over the Frazzini and Pedersen (2014) period (January 1952–March 2012), and 0.43 over the 1922-1951 pre-sample period, as shown in Table D.2 in Online Appendix D. These numbers are similar when using 1-year or 5-year (instead of 3-year) beta estimation windows.

Electronic copy available at: https://ssrn.com/abstract=3325720

McLean and Pontiff (2016) and Linnainmaa and Roberts (2018), amongst others, find that many (equity) return factors display significant weaker performance when tested in out-of-sample periods. Panel C of Table 4 formally tests out-of-sample performance decay in global factor premiums, by showing the difference in Sharpe ratios between the pre-sample and in-sample periods and the t-statistics of a test on equal factor premiums in both periods. We find very limited out-of-sample decay of the global factor premiums, with the average Sharpe ratio dropping to 0.40 from 0.41. At the 5% significance level two factors (bond BAB and commodity BAB) display a significant decay in factor premiums, while also one factor (seasonality in bonds) displays a significant increase. Similarly, one multi-asset combination (BAB) displays a significant decay in factor premiums, while also one (seasonality) displays a significant increase. In other words, we find very little evidence for out-of-sample performance decay in global factor premiums.

INSERT TABLE 5 HERE

Table 5 shows the results of applying the Bayesian concepts to the new sample (Panel A) and the full sample (Panel B). In Table 3, we report for each global return factor the frequentist p-values, the Bayesian p-values with a prior odds ratio of 4:1, and the break-even prior odds that lead to a posterior probability of 5% for the null hypothesis. Over the full sample period (i.e. utilizing the full statistical power of our data) we find that 15 of the 24 global return factors have Bayesian p-values below 5%. Exceptions are bond momentum, value in all asset classes, commodity carry, and BAB outside equities. In addition, the prior odds ratios are mostly above 99-to-1. This is even more true for the multi-asset combinations, for which the break-even prior odds ratios are generally above 99-to-1, except for value with break-even prior odds of 59.52. These results imply that one

Electronic copy available at: https://ssrn.com/abstract=3325720

generally needs to be extremely skeptical in order not to reject the null hypothesis that global factor premiums are zero.

C. Historical investability of global factor premiums

Our results show that global factor premiums have robustly existed in over 217 years of data, which includes more than 150 years of new and independent data. A related question is to which extent the documented global factor premiums can be attributed to investment frictions faced by investors, such as cross-border investment restrictions, leverage constraints, practical or legal boundaries to shorting, and transaction costs.[19] This assumption of frictionless trading has been challenged in the literature, especially for stock-level factor premiums, which require high amounts of trading in illiquid stocks. For example, Korajczyk and Sadka (2004) and Avramov, Chordia and Goyal (2006) examine the impact of frictions on the stock-level momentum or short-term reversal factors, and Novy-Marx and Velikov (2015) show that most stock-level factor premiums survive after costs when using cost-mitigation techniques.

It is commonly assumed that investment frictions were higher in the 19th century than in the 21th century. Although data to assess the impact of investment frictions is generally lacking, indications exist that it was neither impossible nor extremely expensive to trade in the markets we examine. Well-developed international markets for stocks, bonds, currencies, and commodities existed with speculation in the assets we research (Jobst, 2009; Accominotti, 2016; Accominotti and Chambers, 2016). Trading in many assets was feasible at limited transaction costs (e.g. Koudijs, 2016), and asset markets in the 19th century seem in many dimensions similar to today's: active and liquid derivatives markets existed with the ability to go short, with a well-developed system of trading on

[19] Exceptions of asset pricing models that deal with partial segmentation are, amongst others, Black (1974), Stulz (1981), and Karolyi and Wu (2018).

Electronic copy available at: https://ssrn.com/abstract=3325720

margin, market-making and arbitrage mechanism (resulting in the financial integration of prices of identical assets that traded simultaneously in multiple markets), and a centralized exchange with clearing mechanism (Harrison, 1998; Koudijs, 2016; Neal, 1985, 1987). Online Appendix E provides further detail on the investability frictions in historical markets. For a more in-depth discussion on this history of financial markets, see Poitras (2009), Chambers and Dimson (2016), and Zinkina, Andreev, and Mosakova (2017).

Consequently, our results do not necessarily imply that the global factor premiums could have been profitably exploited. Related, we do not examine optimal factor investment strategies, nor aspects linked to arbitrage or tradability (such as transaction costs, legal controls, capital mobility, etc.). Amongst others, the use of liquid instruments, smart trade rules, and multiple factor strategies typically reduce implementation costs significantly (see Novy-Marx and Velikov, 2015, Fitzgibbons et al. 2017). Further, investors did not need to have frictionless access to all instruments in all markets in order to have profited from global factor premiums. For example, even a long-only investor with access to a limited number of markets could have postponed the buying of equities, bonds, currencies or commodities if a market was in a negative trend, overvalued, or in a poorly returning season.

D. Robustness analyses – subperiods, factor construction and data quality

An additional manner to mitigate the influence of p-hacking is to examine the robustness of the global factor premiums to common degrees of freedom for the researcher (e.g. sample periods, factor construction or other methodological choices). Further, we have made several choices to construct a dataset of good quality, of which we next assess the robustness. These include data filters and our sample selection choice.

Electronic copy available at: https://ssrn.com/abstract=3325720

Subperiods

We start our robustness analyses with subperiod results. To reduce the degrees of freedom (that might lead to picking convenient subperiods), we consider all 10-year rolling subperiods, and report the frequency of positive Sharpe ratios in Table 6. The global return factors with significant full sample Sharpe ratios display consistent premiums over time, with frequencies typically significantly above 50%. Exceptions are equity value, currency value, and BAB outside equities. Further, the multi-asset combinations have a positive Sharpe ratio in at least 74% (BAB) to 100% (trend) of the rolling 10-year periods between 1800 and 2016, all significantly greater than 50%. Online Appendix C, Table C.1 shows the results for four fixed economically motivated subperiods over our typical pre-sample period: 1800-1869, 1870-1914, 1915-1945, and 1946-1971. These periods represent the rise of capital markets, globally integrated markets, the global period of war and unrest, and the post-war era and boom period, respectively. Between 1870-1914, for example, financial markets were globally well-developed and integrated with little capital restrictions. Over this and later subperiods, we find strong evidence for global factor premiums, with especially trend, momentum, carry and seasonality being generally consistent over subperiods. Notable exceptions are bond momentum (which is only significant in the 1870-1914 subperiod), and commodity carry (which is significant in the 1946-1971 subperiod, and as shown in Table 2 also in the replicating sample). Value and BAB are generally less consistent, witnessing no significant factor premiums pre-1870, and mainly equity value and equity BAB offering consistent factor premiums afterwards. Bond value is significant post-1946, commodity value is only significant in the 1915-1945 subperiod, currency value and bond BAB are not significant at the 5% level in any of the subperiods (including the replicating sample), commodity BAB is only significant in the replicating sample, and currency BAB being not significant pre-1915 (and in the replicating sample). Overall, the general consistency of performance over time further

Electronic copy available at: https://ssrn.com/abstract=3325720

strengthens our empirical evidence for the existence of most of the global factor premiums we examine. Further, the robustness over subperiods is a further safeguard against possible data biases, as such biases are most likely larger in earlier parts of our sample.

INSERT TABLE 6 HERE

Factor construction

Online Appendix B, Table B.2 further considers methodological variations relative to our 'baseline strategies' examined so far, as made by several of the original studies. We focus on the full sample results, but we have verified that results are similar for the pre-sample period (see Table B.1 in Online Appendix B for robustness in the replication sample). First, instead of using the portfolio weights that depend on the cross-sectional ranking, we take equally weighted long and short positions in the top and bottom tercile within each asset class (the portfolio construction choice employed by Keloharju, Linnainmaa, and Nyberg, 2016).[20] Second, we do not scale the positions using past volatility or beta (in the case of BAB), but rather take simple equal notional positions in each asset. This means that more weight is given to the riskiest assets within an asset class. The resulting differences are small in economic and statistical terms for each variation, except for BAB when not scaling positions by beta (in line with its expectation).

Data quality controls

In order to build a high-quality dataset, we have applied several data filters, of which we next examine the impact. The results are summarized in Table 7. We focus on the pre-sample period, as data quality is a larger concern over the earlier sample period (we report the results over the 1800-2016 period in Online Appendix C, Table C.2). First,

[20] Novy-Marx and Velikov (2018) argue that the cross-sectional ranking method biases portfolio weights to equal weighting, which in the case of individual stocks leads to larger weights in economically less relevant micro-cap stocks. This is less of a concern for our study, as we focus on assets at the market level instead of individual stocks.

Electronic copy available at: https://ssrn.com/abstract=3325720

we remove the zero-return liquidity screen ('No liquidity screen'), as this is not commonly employed in the original studies underlying this paper. Overall, removing this screen tends to increase the Sharpe ratios, most notably for value in currencies. Second, we apply only the zero return screen ('Zero return screen'), with results being very similar to the baseline. Third, we test robustness with respect to outliers (coming, for example, from measurement errors) by trimming asset returns at their 1st and 99th percentiles ('+ Trimmed returns'), yielding generally stronger factor premiums. Fourth, as a potential spurious assignment of dividends or coupons would typically be to year ends, we also show robustness of the global factor premiums by reporting Sharpe ratios by quarter. The findings align with those in Table 4, with factor premiums being generally significant over most quarters, and not originating solely in the fourth quarter ('Q4'), nor any of the other quarters. Further, if there are any spurious effects of capital distributions, we would expect them to be captured by the seasonality factor. Consequently, we report spanning alphas over the seasonality (and all other factors) in Table 8 in the next section. These spanning alphas generally align with the results in Table 4. Overall, the findings that most global factor premiums are significantly different from zero are robust with regard to the above variations, with Sharpe ratios generally being close to the baseline results and most of the tests yielding significant Sharpe ratios.

INSERT TABLE 7 HERE

In addition to our data filters, we conduct a number of additional tests that control the impact of potential data biases on specific factors premiums: trend, momentum, value and seasonality. First, we include a one-month or three-month implementation lag on the factor signals, which removes any impact from the use of the average of high and low prices over a month on trend or momentum, and spuriously inflated prices on value. Note that this lag is on top of the one-month lag for trend and momentum in the baseline

Electronic copy available at: https://ssrn.com/abstract=3325720

results. Table C.3 in Online Appendix C shows trend, momentum and value factor premiums remain generally significant (with currency value remaining weak) when imposing these implementation lags. Second, we exclude capital distributions such as dividends and coupons from returns for the seasonality factor, which removes the impact of possible spurious assignment of payouts on this factor. Seasonality factor premiums are robust to this control, with premiums in spot returns dropping over the pre-sample period for equities, bonds and currencies, but remaining significantly positive for all but currencies. Note that these results also reveal that a substantial part of seasonality factor premiums are driven by seasonal patterns in capital distributions.

In summary, our conclusions on the hypotheses are robust for different data filters and controls, do not critically depend on one specific historical period, nor on specific quarters during the year. This leaves us to conclude that our findings are unlikely to be a spurious result driven by lower historical data quality.

Survivorship bias

In this study, we choose to focus on the main markets in each asset class based on what we can assess today, thereby ignoring the smaller markets. Most of these smaller markets were generally of lesser importance for investors, and hence this choice prevents us from finding factor premiums that would have been of small importance economically. Nevertheless, some of these markets might have been more relevant historically, possibly creating a look-ahead or survivorship bias, as described in Banz and Breen (1986) and Brown, Goetzmann, and Ross (1995). For example, Belgium, Austria (-Hungary) and Russia used to be important financial centers (representing over 10% of the world equity market capitalization in 1900) that have become of substantially smaller importance nowadays (see, for example, Goetzmann and Huang, 2018).[21] In the years after the

[21] Argentina is commonly believed to have been an economy of comparable size to large European markets at the beginning of the 1900s. By contrast, Dimson, Marsh, and Staunton (2008) claim that several Latin

Electronic copy available at: https://ssrn.com/abstract=3325720

communist revolution in Russia in 1917, stock prices in Russia dropped by close to 100%. That said, it is not directly clear in which direction a survivorship bias would influence our results, as factor strategies take dynamic long and short positions over time. Goetzmann and Kim (2018) analyze the equity indices in Global Financial Data for survivorship bias in a long-only setting that is arguably more exposed to survivorship bias (i.e. crashes and rebounds) and show that this is at best a minor concern. Further, our results are generally robust across ten-year subperiods (including the recent subperiods), which suggest little impact of a potential survivorship bias. To have a more formal assessment of the impact of a potential survivorship bias we rerun our baseline test with additional markets on which we could obtain data coverage; Austria (1923-2016), Belgium (1899-2016) and Russia (1865-1930) equity market returns, Austria bond returns (1860-2016), and Austrian currency returns (1861-1998). The results present in the last column of Table 7, Panel B are very similar to the baseline results, with the average Sharpe ratio increasing by 0.01, and Sharpe ratios differences ranging between -0.08 and 0.05. Overall, these small differences do not affect the rejection of the null hypothesis.[22]

IV. Economic explanations of global factor premiums

We have documented robust evidence for the presence of global factor premiums using replication and out-of-sample testing and utilizing various methods that account for p-hacking. In this Section, we turn to insights into potential explanations of these global

American countries combined (including Argentina) constituted less than 1.5% of the equity market capitalization during the 1900s. As our data sources do not cover equity or bond markets in Argentina around the 1900s, we do not include this market in our sample.

[22] A related issue is the possibility of a delisting bias within the database. If large negative returns are not well documented, for example in case of bankruptcy or a default, this tends to overstate the returns of risky assets and understate the returns of less risky assets. For example, the CRSP database contained a delisting bias for many years before it was detected and cleaned by Shumway (1997). This bias was most severe among small risky stocks, thus leading to an overestimation of the size premium (see Shumway and Warther, 1999). In this study we do not test a global size factor, as this is a bottom-up effect not documented across markets. However, a possible delisting bias in general overstates the returns of risky assets thus leading to a potential underestimation of the BAB premium in particular. That said, we observe few delistings in our database, and in such special cases observe extreme negative returns that align with more anecdotical sources (for example, in the case of German and Japan debt restructurings, and Russian equity markets post the communist revolution).

Electronic copy available at: https://ssrn.com/abstract=3325720

factor premiums. In line with, amongst others, Koijen et al. (2018) and Keloharju, Linnainmaa, and Nyberg (2016), we consider common variation, and explanations related to market risk, downside risk, and macroeconomic risks.[23] To this end, we exploit the power of our full sample period, ranging from 1800 to 2016, as a deep sample allows for a more powerful identification of explanations that relate to relatively infrequent observations, such as severe market corrections or macroeconomic risks.

A. Market risk and common variation

The Sharpe ratios for the factors that we have shown in the previous section measure return per unit of risk, but do not account for exposures to traditional asset classes. If the factor returns can be explained by correlation to traditional global markets, they are redundant, even if they have an economically significant Sharpe ratio. In order to adjust for these market exposures, we start by displaying the appraisal ratios, or Jensen's alpha divided by the residual volatility, for each of the factor-asset class combinations, where we adjust the factor return series for exposures to market risks, in Table 8, Panel A. The results from these appraisal ratios are similar to those based on Sharpe ratios, with currency value and BAB outside equities being insignificant. These results indicate that none of the global return factors are strongly correlated to global asset class risks.

We continue in a next step to show the average pairwise monthly return correlations. These can be found in Panel B of Table 8. The average correlation of each factor across asset classes is close to zero, with only BAB (0.03) and trend (0.07) having significant correlations. A similar picture emerges for the average correlation for factors within an asset class, with values ranging from an insignificant 0.01 (commodities) to a significant but small 0.08 (currencies) or 0.09 (equities). Furthermore, most individual correlation

[23] Another explanation offered for many of the global factor premiums is market or funding liquidity risk (see, for example, Asness, Moskowitz and Pedersen, 2013, and Koijen et al. 2018). Due to the limited availability of deep historical data on the measures used in these studies, we do not examine such explanations in this paper.

Electronic copy available at: https://ssrn.com/abstract=3325720

coefficients between each factor-asset class series are also close to zero, as shown in Online Appendix Table B.3 (containing the entire correlation matrix). From this perspective, most of the 24 global return factors are unique drivers of return that share little common variation. The main exceptions are trend and momentum, which correlate positively (0.55 for the multi-asset factor series), while both correlate negatively with value.

To show the added value of each of the factors, we regress its monthly time-series on the 23 other return series and four market factors (as in Panel A). To explore the maximum sample period that is available for each return factor, we impute missing values of independent variables with the projection of an independent variable on the other independent variables, following Berk and Van Binsbergen (2015).[24] The intercept of this regression can be interpreted as an expansion of the mean-variance efficient frontier with respect to the other factors (see De Roon and Nijman, 2001), and is hence a test of the unique added value of each of the 24 global factor premiums. We scale these (annualized) intercepts by their residual volatilities for ease of comparison with previous tables and Panel A. Table 8, Panel C shows the results from these spanning regressions.

In line with the above results, most (scaled) intercepts are significantly positive, with the main exception being BAB. BAB in equities has a significantly positive intercept only at the 10% significance level and BAB loses significance at the multi-asset level.[25] Furthermore, value gains substantially in significance with a multi-asset scaled intercept of 0.54 (t-value = 7.91), compared to 0.29 (t-value = 4.33) in Panel A, and obtaining significant appraisal ratios for all asset classes. In particular, bond value and commodity value now exhibit higher factor premiums with t-statistics well exceeding 3.00, driven by the negative correlation with trend and momentum. Further, currency value becomes

[24] We have verified that alternative approaches for imputing missing values, like replacing missing values by their sample average, or predictive mean matching (Rubin, 1986, Little, 1988), yield very similar results.
[25] These results suggest long positions to most of the global return factors in optimal asset allocations. To check this, we have evaluated optimal long-only mean-variance asset allocations to traditional asset classes and the multi-asset combinations of the global return factors. Our findings reveal high allocation weights to each of the Multi Asset factors except for cross-sectional momentum and BAB. To save space, we do not report these results separately.

Electronic copy available at: https://ssrn.com/abstract=3325720

significant at the 10% level. In other words, the factor premiums generally have higher t-statistics once controlling for other factors.

As trend and momentum are highly correlated, an important question is how these factors span each other. Moskowitz, Ooi, and Pedersen (2012) report that over the period 1985 to 2009 trend factors subsume the momentum factors. Recently, Goyal and Jegadeesh (2018) put these results in a different perspective by arguing that the dominance of trend over momentum is driven by the time-varying net long investment of trend in asset with positive premiums, and when controlling for these effects trend's dominance over (cross-sectional) momentum disappears. An important part of the difference between trend and momentum is a static effect (i.e. being on average long in assets with average positive premiums), which arguably is not a reflection of factor premiums. To control for such static effects, we have added each asset classes' return in our spanning regressions. Confirming Goyal and Jegadeesh (2018), we find relatively little evidence for the dominance of trend over (cross-sectional) momentum, with significant scaled intercepts for trend in all asset classes, and momentum in equities, commodities and multi-asset.[26]

INSERT TABLE 8 HERE

B. Downside risk

A large and growing literature considers whether various return anomalies compensate investors for downside or crash risk. For example, Doskov and Swinkels (2015) show that the currency carry factor is exposed to crash risk, and Koijen et al. (2018)

[26] We note that these results critically depend on including the asset class returns into the spanning regression. When omitting the asset class returns, we find more evidence supporting trend factors subsuming momentum factors (similar to Moskowitz, Ooi, and Pedersen, 2012, over the 1985-2009 period), with scaled intercepts on the cross-sectional momentum in bond, currencies and multi-asset being statistically indistinguishable from zero, and scaled intercepts on trend factors being significantly positive in all asset classes.

Electronic copy available at: https://ssrn.com/abstract=3325720

show that downside risk captures a part of the carry return in fixed income and commodity markets. In this subsection, we consider downside risk explanations via the Downside Risk CAPM (DR CAPM) of Bawa and Lindenberg (1977). When the DR CAPM holds, assets with higher downside betas should have higher expected returns. Another way to interpret downside risk is as a conditional risk factor based on falling markets, also referred to as downstate beta.

Recently, Lettau, Maggiori and Weber (2014) have used the DR CAPM model and find that the carry factor has higher downside beta compared to the regular beta. Degrees of freedom in testing downside beta are the thresholds at which downside returns start to count, and the choice of benchmark return. Typical choices are zero, or lower thresholds deeper into the left tail of the return distribution, with threshold and equity market as benchmark. Furthermore, a common challenge when estimating downside risk exposures and premiums is the general reduction in the number of observations in the left tail (Post and Van Vliet, 2006). Market crashes in particular do not happen very often. For example, Lettau, Maggiori and Weber (2014) use a threshold of -1 standard deviation of the equity market return, which results in 55 monthly observations out of 435 in their 1974-2010 sample (87% of all observations are excluded). In our 217-year sample, we have many more such events, allowing us to rigorously examine the hypothesis that downside risk can explain the premiums on the global return factors. For example, we have 43 years of equity bear markets and 218 downside market states in the case of a threshold of -1 standard deviation of the equity market return. This relatively large number of

Electronic copy available at: https://ssrn.com/abstract=3325720

observations also enables us to study downside risk even further into the left tail of the distribution.

INSERT TABLE 9 HERE

INSERT FIGURE 2 HERE

We apply both the CAPM and the DR CAPM using the settings used in Lettau, Maggiori and Weber (2014). Table 9 summarizes the results. Shown are the CAPM and DR CAPM alphas, its t-values, sorted according to their downside equity market betas. Figure 2 depicts the excess return versus beta (right plot) or downside beta (left plot) of each global return factor. Overall, the average downside beta is very similar to the regular beta, with differences not larger than 0.10. Focusing on individual global return factors, seasonal and momentum for equities and bonds, and value and carry for currencies have the highest increase in downside beta, the latter in line with the results of Koijen et al. (2018) and Lettau, Maggiori and Weber (2014). Still, downside risk explains at best a small fraction of momentum and carry profits, with, for example, a downside beta - regular beta difference of 0.04 for the multi-asset combination, reducing the alpha by a marginal 0.14%. In Online Appendix B, Table B.4 we further show the sensitivity of the downside beta results to using a zero-return cutoff or using equity crash betas (i.e. with the threshold deeper in the left tail). Moreover, Baltussen, Post and Van Vliet (2012) show that the stock-level value premium is exposed to downside risk in both equity and bond markets. Consequently, we also consider conditioning on bond market movements in Table B.4. Overall, these results are similar to the above, with downside risk explaining at best a part of the global factor premiums.

Next, we estimate the risk premiums attached to beta or downside beta using the Fama and MacBeth (1973) approach, and find that the price of beta as priced in the cross-

Electronic copy available at: https://ssrn.com/abstract=3325720

section of global return factors is of the wrong sign (-0.02) and insignificant (t-value = -0.21), as also evident from the flat line in Figure 2. When beta is replaced by downside beta in the Fama and MacBeth regressions, we find an (unreported) insignificant cross-sectional risk premium of 0.40 percent (t-value = 0.60). Moreover, instead of estimating the risk premium, we can infer it from the data as the time-series average equity premium (3.05 percent). Using this risk premium, the downside beta required to bring the average global factor return (3.94%) to zero would be above 1.25. This is a very high downside beta not observed for any of the factors in the full sample (for example, the maximum tail beta show in Online Appendix B is 0.23). Moreover, we do not observe such large downside betas for any of the 24 global return factors when considering any 10-year sample period (unreported for the sake of brevity). Based on these long-run sample results we conclude that downside market risk does not materially explain global factor premiums.

C. Bad states and macroeconomic risk

Risk-based explanations of asset pricing anomalies argue for time-variation in expected returns related to time variation in its risk or risk premium, aspects that can be expected to relate to macroeconomic or market conditions (Chen, Roll, and Ross, 1986, Fama and French, 1989, Ferson and Harvey, 1991).[27] A major advantage of our extended sample (resulting from over 200 years of data) is the presence of many observations within and across various economic and market regimes (e.g. recessions and crisis episodes). A sample based on the past few decades might be biased, given that they were quite exceptional (no major wars, growing global prosperity and only a few large recessions or periods of social unrest), which severely limits the number of 'bad states'. Our extensive

[27] For example, Merton's (1973) Intertemporal Capital Asset Pricing Model stipulates that in a risk-averse economy any variable that affects the set of future investment opportunities or the level of consumption earns a risk premium. Macroeconomic variables should be high up on the list of candidates, since they impact the cash flows of many agents in the real economy simultaneously, typically impact the real investment opportunities available (for example via government stimulus) and covary with risk appetite in the markets. Further, papers by, amongst others, Fama and French (1989) and Ferson and Harvey (1991) argue that expected business conditions are fundamental drivers of time variation in expected risk premiums.

Electronic copy available at: https://ssrn.com/abstract=3325720

multi-century sample includes many bear markets (43 years) and recessions (74 years). It therefore allows us to examine and describe in greater depth to what extent the global factor premiums can be explained with macroeconomic risk.

Many approaches have been used in the literature to examine sensitivity to, and pricing of, macroeconomic risks. Following Griffin, Ji and Martin (2003) and Keloharju, Linnainmaa, and Nyberg (2016), we comprehensively examine macroeconomic risk explanations of the global factor premiums by taking the following approaches. First, we divide periods into 'good' and 'bad' states and examine the pricing of global return factors over each. Second, we examine macroeconomic risk exposures and pricing using an unconditional approach, in the spirit of Chen, Roll and Ross (1986).

To examine the pricing of global return factors over good and bad states, we construct indicators of each market state and compute contemporaneous average annual factor returns for each state, as well as the return difference between the states. The market states, or regimes, that we examine are constructed per calendar year and are:

- Recession versus expansionary periods, where we mark calendar years as recessionary when at least six months of a calendar year are in a recession, and as expansionary otherwise. Our sample has 74 recession and 143 expansion years.
- Global crisis versus non-crisis periods, where we mark calendar years as 'Crisis' ('Non-crisis') when the Rogoff and Reinhart Banking Currency Default Inflation (BCDI) indicator is above (below) 50. Our sample has 52 (165) crisis (non-crisis) years.
- Equity bull versus bear market periods, since one could argue that the equity markets provide a forward-looking assessment of macroeconomic conditions. We mark calendar years as a bull (bear) market when the calendar year global equity return series were positive (negative). Our sample has 43 (174) bear (bull) market years.

INSERT TABLE 10 HERE

Electronic copy available at: https://ssrn.com/abstract=3325720

INSERT TABLE 11 HERE

The results are summarized in Table 10. If the global factor premiums are driven by macroeconomic risks one would expect their returns to be especially strong in bad states. Overall, most global return factors do not display stronger returns during bad states (i.e. recessions, crises and bear markets). The most notable exceptions are trend in commodities, value in equities and seasonality in bonds and commodities, which have significantly higher returns during recession periods. However, these findings do not persist in crisis periods and returns are, except for value in equities, also significant during expansion or bull markets. The last row of Table 10 examines the common effects of macroeconomic states on all global factor premiums jointly, using a panel regression (with asset class/factor-fixed effects and date-corrected standard errors). The results confirm significant global factor returns across all states, and reveal no significant variation across states.[28]

Our second analysis examines exposures to, and unconditional pricing of, macroeconomic factors. To this end, we construct the most widely used Chen, Roll and Ross (1986) factors – log changes in industrial production (MP; as in Chen et al. led by 1

[28] In addition, we have tested the impact of investor sentiment, as in Keloharju, Linnainmaa, and Nyberg (2016). Many studies argue that the significant presence of sentiment-driven investors can cause persistent mispricing and that profits on various return anomalies should be higher following high periods of sentiment if they are a reflection of mispricing (Baker and Wurgler, 2006, Stambaugh, Yu, and Yuan, 2012). Related, Keloharju, Linnainmaa and Nyberg (2018) argue that return seasonalities are likely due to temporary mispricing. We explore a mispricing explanation by reconstructing the text-based market-wide pessimism measure of Garcia (2013) on a daily basis and extend it to span the 1899–2014 period (before 1899 data coverage is sparse). Subsequently, we average the word measures over each month, and divide our sample into high and low sentiment months. In the spirit of Baker and Wurgler (2006) and Keloharju, Linnainmaa, and Nyberg (2016), we define high sentiment months as those in which the value of the market-wide pessimism index for the previous month is below the index's median value for the sample period. The low sentiment months are those that follow above-median pessimism index values. Our findings reveal that global factor premiums are generally significantly weaker in low sentiment states, and stronger in high sentiment states (but significantly positive in both), in line with the results of, amongst others, Stambaugh, Yu, and Yuan (2012), who find higher returns on stock-level return anomalies after periods of high sentiment. More specifically, we find that 13 out of 24 coefficients indicate higher factor premiums after high sentiment months (five out of six for the multi-asset combinations), an effect that is strong and significant in the panel regression.

Electronic copy available at: https://ssrn.com/abstract=3325720

month), term spread (UTS), changes in expected inflation (DEI), and unexpected inflation (UI) – for our global sample using monthly data.[29] We regress the time series of each global return factor on these macroeconomic variables and obtain coefficients and intercepts. Our sample starts in February 1869 due to the availability of deep historical inflation data. Table 11 summarizes the results. If global factor premiums are driven by macroeconomic risk, then they should exhibit significant sensitivity to the factors proposed by Chen, Roll and Ross (1986). Our findings reveal that the global macroeconomic variables are generally not related to global factor returns or subject to the wrong sign, with a couple of noteworthy exceptions. Momentum and carry (value) tend to load positively (negatively) on MP, seasonality tends to load positively on MP and negatively on UTS and UI, and BAB tends to load positively on UTS and negatively on UI. Moreover, all significant global factor premiums of Section III (all except for value in currencies and BAB outside equities) have intercepts that are highly significant and are of similar magnitude to the raw returns over this sample (reported in the column 'Actual'). These results suggest that macroeconomic risks have very limited explanatory power for the global factor premiums.

Next, to examine risk premiums attached to each macroeconomic factor and to what extent they can explain the global factor premiums, we apply the Fama and MacBeth (1973) technique on a monthly frequency with our global return factors as test assets. We combine the premiums with the estimated loadings to decompose the returns on the global return factors into predicted and unexplained components. If the Chen, Roll and Ross factors suffice for explaining global factor premiums, then the difference between the actual and predicted returns (or unexplained) should not be significantly different from

[29] More specifically, we construct these measures for the U.S., U.K., Germany, France and Japan (more details on the data can be found in the Appendix), and average these for a global measure. We have verified that our results are robust if U.S.-specific measures are used instead. Further, Chen, Roll and Ross (1986) show that these macroeconomic variables and the default premium are priced risk factors using the Fama and MacBeth (1973) regression on the cross-section of U.S. size portfolios. Akin to Griffin, Ji and Martin (2003), we omit the default premium, as its historical data availability is limited, especially outside the U.S.

Electronic copy available at: https://ssrn.com/abstract=3325720

zero. The empirical results show that several global return factors have significant expected macroeconomic premiums, but that the unexplained return is generally more substantial and highly significant. The main exceptions are momentum in bonds (where the unexplained returns turn significantly negative), and seasonality in bonds and currencies. However, this result coincides with several negative loadings or negative estimated risk premiums; the risk premiums (unreported) on DEI is insignificant and on UTS and UI significantly negative. When imposing risk premiums to be positive (unreported), this result disappears.

In summary, our tests reveal no supporting evidence of a link between macroeconomic risks and global factor premiums. This finding is consistent with Asness, Moskowitz, and Pedersen (2013) and Koijen et al. (2018), who report that value, momentum, and carry returns cannot be explained by macroeconomic risks in their shorter samples.

V. Conclusion

We examine global factor premiums from three dimensions. First, we replicate existing studies to global factor premiums with a uniform factor construction methodology and investment universe, thus limiting the degrees of freedom for each individual factor. Our results are generally similar to the published results. Next, we examine the impact of p-hacking and introduce the novel Bayesian concept of break-even odds, the prior odds at which the Bayesian p-value would equal the confidence level chosen. These analyses reveal ambiguous replication evidence on the global factor premiums, with on average relatively low prior odds being sufficient to question the replication sample evidence on the global factor premiums. Therefore, besides the replication, an out-of-sample analysis is needed to draw firm conclusions on the existence of global factor premiums.

Second, we build a wide and deep sample on the global factor premiums, covering 217 years of data across the major international asset classes, which includes more than

Electronic copy available at: https://ssrn.com/abstract=3325720

150 years of new and independent data. The new sample evidence reveals that the large majority of global factor premiums are convincingly present from economic, statistical, and p-hacking perspectives. Sharpe ratios are on average around 0.40 and generally persistent over time and robust for variations in factor construction settings or data choices. Further, most t-statistics are well above 3.00, and break-even prior odds imply that one needs to be extremely skeptical to disregard the deep sample evidence provided in this study. Interestingly, the global return factors generally have close to zero decay in performance between the replication sample and out-of-sample period. Factor premiums that are weak (strong) in the original studies are also weak (strong) in the extended sample period.

Third, we examine economic explanations for the global factor premiums. Factors generally do not span each other, while global factor premiums are not significantly explained by downside risk, are consistently present across various macroeconomic states, and cannot be explained by macroeconomic risks. Consequently, our results seem hard to reconcile with risk-based explanations, although we have to be cautious on such an interpretation as risk exposures and especially risk premiums are not directly observable. Instead, we interpret our findings as providing no positive evidence of a relationship between global factor premiums and risk, nor on a unified explanation for the global factor premiums. Consequently, explanations of the documented factor premiums are an important topic for future research.

Our results have strong implications for research in asset pricing. The literature on asset pricing theory and return predictability has often evolved separately by asset class or factor. Most studies focus on a single asset class, market or factor at a time, ignoring the wide presence of a group of return factors across asset classes. Our findings reveal that a theory that aims to model variation in expected returns should consider multiple asset classes and global factor premiums simultaneously. Furthermore, the documented

Electronic copy available at: https://ssrn.com/abstract=3325720

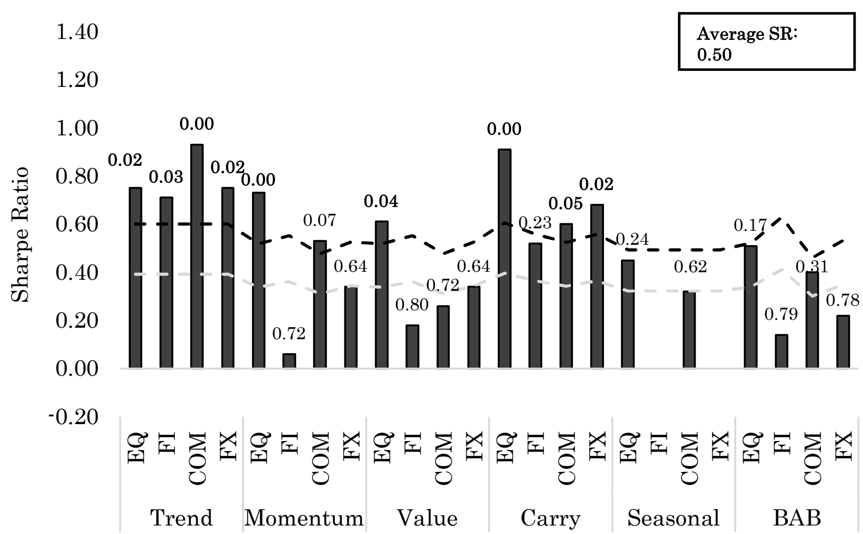
factors are important controls for empirical studies into new asset pricing factors. All in all, our study shows a strong, robust and persistent presence of economically important global factor premiums.

Electronic copy available at: https://ssrn.com/abstract=3325720

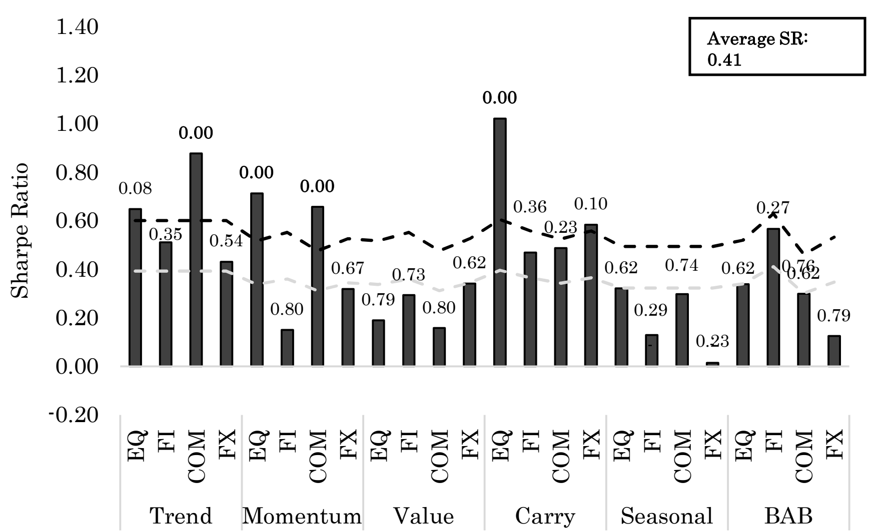
Figures

Figure 1: Global factor returns since 1800. The figure shows the annualized Sharpe ratios for the 24 global return factors. The dashed lines show the cutoff on the Sharpe ratio corresponding to t-values of 1.96 (in light grey) and 3.00 (in black) respectively. The values above each bar are the Bayesian p-values using a 4-to-1 prior odds ratio. Numbers in bold are below the 5% significance level. Panel A shows the originally documented results, Panel B shows the replication results for the asset class-factor specific 'in sample' period, Panel C shows the 'pre-sample' period starting in 1800, and Panel D shows the results over the full 1800-2016 period. The global factors are long-short portfolios applied on international equity indices ("EQ"), 10-year maturity government bond indices ("FI"), commodities ("COM"), and currencies ("FX").

Panel A: Original documentation

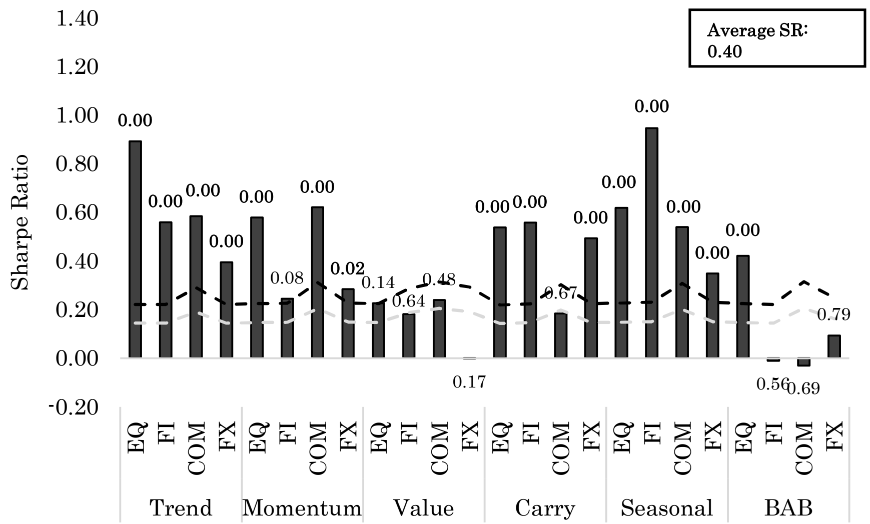


Panel B: Replicating factors

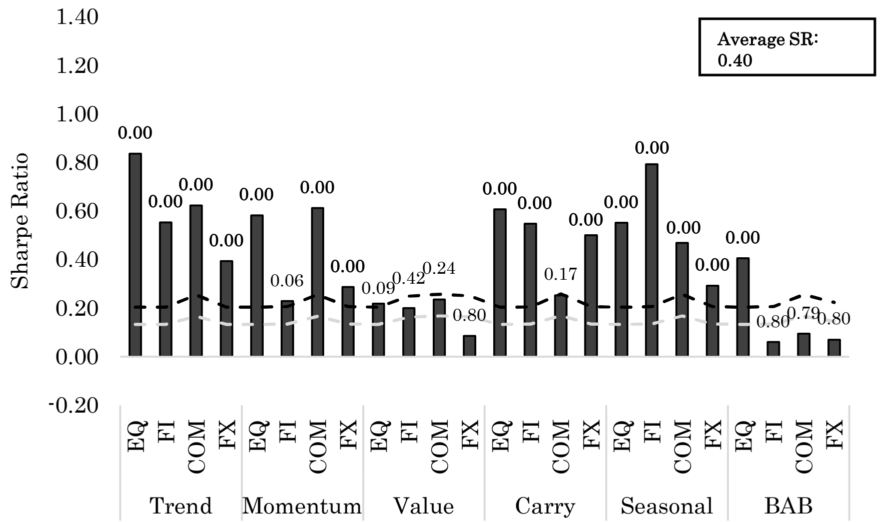


Electronic copy available at: https://ssrn.com/abstract=3325720

Panel C: Pre-sample evidence

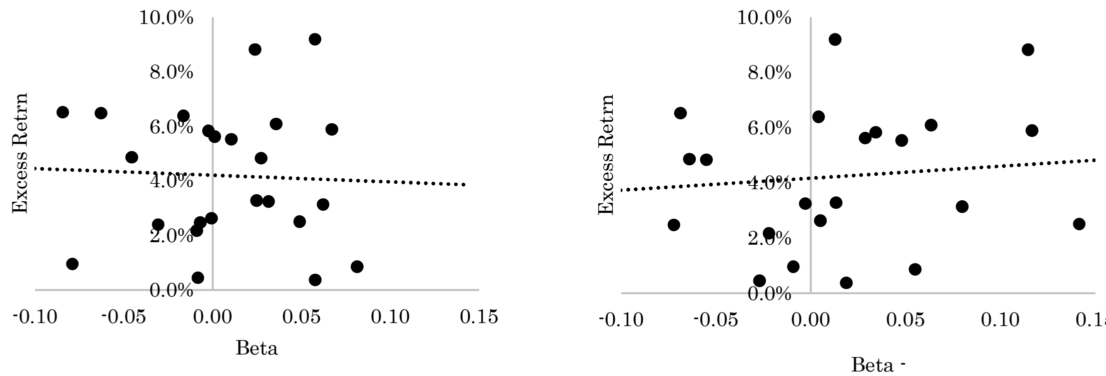


Panel D: Full sample evidence; 1800-2016



Electronic copy available at: https://ssrn.com/abstract=3325720

Figure 2: Downside risk and global factor premiums. The figure shows regular and downside betas versus the average return on the global factors. The left panel shows the regular beta ("Beta") and the right panel shows the downside beta ("Beta-"). The sample runs from 1800 until 2016 and is at the monthly frequency.



Electronic copy available at: https://ssrn.com/abstract=3325720

Tables

Table 1: Original studies on global factor premiums

The table contains the factor definitions ('Definition'), t-statistics ('t-stat'), Sharpe ratios ('Sharpe'), start dates ('Start') and end dates ('End'), format YYYYMM, portfolio construction methods ('Portfolio'), and maximum number of assets used ('# Assets') for each of the original studies on global factor premiums: Moskowitz, Ooi, and Pedersen (2012) for time-series momentum ('Trend'), Asness, Moskowitz, and Pedersen (2013) for cross-sectional momentum ('Momentum') and value ('Value'), Koijen, Moskowitz, Pedersen, and Vrugt (2018) for carry ('Carry'), Keloharju, Linnainmaa, and Nyberg (2016) for return seasonal ('Seasonal'), and Frazzini and Pedersen (2014) for betting-against-beta ('BAB'). The asset classes refer to country equity indices ('Equities') and country bonds ('Bonds'), commodities ('Commodities'), and currencies ('FX'). Abbreviations for the factor definitions are: 12M(-1M): past 12-month returns (minus the last month), BE/ME: book value of equity divided by the market value of equity, 5Y Rev: 5-year reversal, 5Y Rel. PPP: 5-year relative purchasing power parity, Impl DY-rf: futures-implied excess dividend yield, Rate diff: interest rate differential, Slope+Roll: bond yield-rf+curve rolldown, Basis: futures-implied convenience yield, 20Y,M: average monthly return over the past 20-year period, Beta: market beta using past year of daily data for volatilities and past five-year data for correlation. Abbreviations for the portfolio construction are: < >0: volatility-weighted portfolio position depending on the sign of the past 12-month return, P1-P3: equally-weighted top tercile portfolio minus the bottom tercile portfolio, Rank: using weights proportional to the asset's rank. T-statistics or Sharpe ratios are those reported as main result, with numbers in italics calculated using information provided in the original study.

Factor	Asset	Definition	t-stat	Sharpe	Start	End	Portfolio	# Assets
Trend	Equities	12M	3.77	*0.75*	198501	200912	<> 0	9
	Bonds	12M	3.53	*0.71*	198501	200912	<> 0	13
	Commodities	12M	4.66	*0.93*	198501	200912	<> 0	24
	FX	12M	3.41	*0.68*	198501	200912	<> 0	10
Momentum	Equities	12M-1M	4.14	0.73	197801	201107	P1-P3	18
	Bonds	12M-1M	0.35	0.06	198201	201107	P1-P3	10
	Commodities	12M-1M	3.29	0.53	197201	201107	P1-P3	27
	FX	12M-1M	1.90	0.34	197901	201107	P1-P3	10
Value	Equities	BE/ME	3.45	0.61	197801	201107	P1-P3	18
	Bonds	5Y Rev.	0.97	0.18	198201	201107	P1-P3	10
	Commodities	5Y Rev.	1.61	0.26	197201	201107	P1-P3	27
	FX	5Y Rel. PPP	1.89	0.34	197901	201107	P1-P3	10
Carry	Equities	Impl. DY-rf	*4.50*	0.91	198803	201209	Rank	13
	Bonds	Slope+Roll	*2.79*	0.52	198311	201209	Rank	10
	Commodities	Basis	*3.43*	0.60	198001	201209	Rank	24
	FX	Rate diff.	*3.65*	0.68	198311	201209	Rank	20
Seasonal	Equities	20Y, M	2.76	*0.45*	197501	201112	P1-P3	15
	Bonds	20Y, M	-	-	-	-	-	-
	Commodities	20Y, M	1.97	*0.32*	197501	201112	P1-P3	24
	FX	20Y, M	-	-	-	-	-	-
BAB	Equities	Beta	2.93	0.51	197901	201203	Rank	13
	Bonds	Beta	0.67	0.14	198907	201203	Rank	9
	Commodities	Beta	0.72	0.11	196911	201203	Rank	25
	FX	Beta	1.23	0.22	198007	201203	Rank	10

Electronic copy available at: https://ssrn.com/abstract=3325720

Table 2: Historical performance of global return factors: replicating sample

The table summarizes the historical performance of the global return factors. Shown are per factor per asset class the historical annualized Sharpe ratio. The sample has asset class-factor specific start and end dates based on the key papers in the literature displayed in Table1, with the Multi-Asset period covering the earliest start date and latest end date per factor as displayed in Table 1. As seasonality in bonds and currencies are not tested in the original paper, we use the period January 1975 to December 2011, the period employed for seasonality in equity indices and commodities. The return series are at the monthly frequency. Covered are equity indices ("Equities"), 10-year maturity government bond indices ("Bonds"), commodities ("Commodities"), currencies ("FX"), and their equally-weighted combination across the four asset classes ("Multi Asset"). Numbers in parentheses indicate t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

Sharpe ratio	Trend	Momentum	Value	Carry	Seasonality	BAB
Equities	0.65***	0.71***	0.19	1.02***	0.32**	0.34*
	(3.24)	(4.13)	(1.11)	(5.06)	(1.96)	(1.95)
Bonds	0.51**	0.15	0.29	0.47**	0.13	0.57***
	(2.55)	(0.82)	(1.59)	(2.52)	(0.79)	(2.71)
Commodities	0.88***	0.66***	0.16	0.49***	0.30*	0.30*
	(4.39)	(4.13)	(0.99)	(2.79)	(1.82)	(1.95)
FX	0.43**	0.32*	0.34*	0.58***	0.01	0.13
	(2.15)	(1.82)	(1.95)	(3.14)	(0.09)	(0.70)
Multi Asset	1.04***	0.88***	0.40**	1.07***	0.38**	0.46***
	(5.21)	(5.52)	(2.53)	(6.15)	(2.28)	(2.98)

Electronic copy available at: https://ssrn.com/abstract=3325720

Table 3: Statistical perspectives on global return factors: replicating sample

The table summarizes various statistical perspectives on the historical performance of the global return factors. Shown per factor per asset class are the historical frequentist p-value ("p-value"), Bayesianized p-value using a 4-to-1 prior odds ratio ("Bayesian-p") and break-even prior odds at a 5% confidence level ("BE-odds") of its performance. The sample has asset class-factor specific start and end dates based on the key papers in the literature displayed in Table1, with the Multi-Asset period covering the earliest start date and latest end date per factor as displayed in Table 1. As seasonality in bonds and currencies are not tested in the original paper, we use the period January 1975 to December 2011, the period employed for seasonality in equity indices and commodities. The return series are at the monthly frequency. Covered are equity indices ("Equities"), 10-year maturity government bond indices ("Bonds"), commodities ("Commodities"), currencies ("FX"), and their equally-weighted combination across the four asset classes ("Multi Asset").

		Trend	Momentum	Value	Carry	Seasonality	BAB
Equities	p-value	0.00	0.00	0.27	0.00	0.05	0.05
	Bayesian-p	0.08	0.00	0.79	0.00	0.62	0.62
	BE-odds	2.18	47.39	0.05	2877.94	0.12	0.12
Bonds	p-value	0.01	0.41	0.11	0.01	0.43	0.01
	Bayesian-p	0.35	0.80	0.73	0.36	0.80	0.27
	BE-odds	0.36	0.05	0.07	0.34	0.05	0.51
Commodities	p-value	0.00	0.00	0.32	0.01	0.07	0.05
	Bayesian-p	0.00	0.00	0.80	0.23	0.67	0.62
	BE-odds	132.66	47.38	0.05	0.63	0.09	0.11
FX	p-value	0.03	0.07	0.05	0.00	0.93	0.48
	Bayesian-p	0.54	0.67	0.62	0.10	0.43	0.79
	BE-odds	0.16	0.09	0.11	1.63	0.25	0.05
Multi Asset	p-value	0.00	0.00	0.01	0.00	0.02	0.00
	Bayesian-p	0.00	0.00	0.36	0.00	0.48	0.16
	BE-odds	>9,999	>9,999	0.34	>9,999	0.21	1.03

Electronic copy available at: https://ssrn.com/abstract=3325720

Table 4: Historical performance of global return factors: 1800 - 2016

The table summarizes the historical performance of the global return factors. Shown per factor per asset class are the historical annualized Sharpe ratio. Panel A shows the results for the 'pre-sample' period, Panel B shows the results for the full sample period, and Panel C shows the performance decay between the pre-sample and replication periods. The pre-sample period starts as early as January 1800 and ends at the asset class-factor specific start date of the sample used in the original studies, see Tables 1 and 2. The full sample period spans January 1800 till December 2016. The sample is at the monthly frequency. Covered are equity indices ("Equities"), 10-year maturity government bond indices ("Bonds"), commodities ("Commodities"), currencies ("FX"), and their equally-weighted combination across the four asset classes ("Multi Asset"). Numbers in parentheses indicate t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

Panel A: Pre-sample; starting in 1800

Sharpe ratio	Trend	Momentum	Value	Carry	Seasonality	BAB
Equities	0.89***	0.58***	0.23***	0.54***	0.62***	0.42***
	(12.15)	(7.72)	(3.02)	(7.38)	(8.17)	(5.64)
Bonds	0.56***	0.24***	0.18*	0.56***	0.95***	-0.01
	(7.62)	(3.25)	(1.90)	(7.48)	(12.33)	(-0.15)
Commodities	0.58***	0.62***	0.24**	0.18*	0.54***	-0.03
	(6.04)	(5.99)	(2.29)	(1.81)	(5.27)	(-0.30)
FX	0.40***	0.28***	0.00	0.49***	0.35***	0.09
	(5.38)	(3.75)	(-0.02)	(6.60)	(4.54)	(1.13)
Multi Asset	1.03***	0.70***	0.24***	0.88***	1.15***	0.25***
	(14.07)	(9.14)	(3.09)	(11.80)	(15.18)	(3.20)

Panel B: Full sample; 1800-2016

Sharpe ratio	Trend	Momentum	Value	Carry	Seasonality	BAB
Equities	0.84***	0.58***	0.22***	0.61***	0.55***	0.41***
	(12.33)	(8.58)	(3.21)	(8.93)	(8.12)	(5.97)
Bonds	0.55***	0.23***	0.20**	0.55***	0.79***	0.06
	(8.15)	(3.33)	(2.41)	(7.99)	(11.53)	(0.88)
Commodities	0.62***	0.61***	0.24***	0.25***	0.47***	0.09
	(7.35)	(7.19)	(2.76)	(2.94)	(5.48)	(1.11)
FX	0.39***	0.29***	0.09	0.50***	0.29***	0.07
	(5.79)	(4.18)	(1.03)	(7.29)	(4.26)	(0.94)
Multi Asset	1.01***	0.72***	0.28***	0.91***	0.99***	0.30***
	(14.95)	(10.55)	(4.19)	(13.47)	(14.65)	(4.41)

Panel C: Pre-sample -/- replication sample decay

Sharpe ratio	Trend	Momentum	Value	Carry	Seasonality	BAB
Equities	0.24	-0.13	0.04	-0.48*	0.30*	0.08
	(0.48)	(-0.59)	(0.22)	(-1.86)	(1.69)	(-0.22)
Bonds	0.05	0.09	-0.11	0.09	0.82***	-0.58**
	(0.32)	(0.57)	(-0.27)	(0.30)	(4.68)	(-2.14)
Commodities	-0.29*	-0.04	0.08	-0.30*	0.24	-0.33**
	(-1.62)	(-0.79)	(0.29)	(-1.64)	(1.10)	(-2.13)
FX	-0.04	-0.04	-0.34	-0.09	0.33*	-0.03
	(0.37)	(0.14)	(-1.50)	(-0.14)	(1.89)	(-0.43)
Multi Asset	-0.01	-0.18*	-0.17	-0.20	0.77***	-0.21**
	(-0.46)	(-1.60)	(-1.20)	(-1.51)	(4.33)	(-2.21)

Electronic copy available at: https://ssrn.com/abstract=3325720

Table 5: Statistical perspectives on global return factors: 1800 - 2016

The table summarizes various statistical perspectives on the historical performance of the global return factors. Shown per factor per asset class are the historical frequentist p-value ("p-value"), Bayesian p-value using a 4-to-1 prior odds ratio ("Bayesian-p") and break-even prior odds at a 5% confidence level ("BE-odds") of its performance. Panel A shows the results for the 'pre-sample' period, Panel B shows the results for the full sample period. The pre-sample period starts as early as January 1800 and ends at the asset class-factor specific start date of the sample used in the original studies, see Tables 1 and 2. The full sample period spans January 1800 till December 2016. The sample is at the monthly frequency. Covered are equity indices ("Equities"), 10-year maturity government bond indices ("Bonds"), commodities ("Commodities"), currencies ("FX"), and their equally-weighted combination across the four asset classes ("Multi Asset").

Panel A: Pre-sample; starting in 1800

		Trend	Momentum	Value	Carry	Seasonality	BAB
Equities	p-value	0.00	0.00	0.00	0.00	0.00	0.00
	Bayesian-p	0.00	0.00	0.14	0.00	0.00	0.00
	BE-odds	>9,999	>9,999	1.15	>9,999	>9,999	>9,999
Bonds	p-value	0.00	0.00	0.06	0.00	0.00	0.88
	Bayesian-p	0.00	0.08	0.64	0.00	0.00	0.56
	BE-odds	>9,999	2.21	0.11	>9,999	>9,999	0.15
Commodities	p-value	0.00	0.00	0.02	0.07	0.00	0.77
	Bayesian-p	0.00	0.00	0.48	0.67	0.00	0.69
	BE-odds	>9,999	>9,999	0.21	0.09	7,905.80	0.09
FX	p-value	0.00	0.00	0.98	0.00	0.00	0.26
	Bayesian-p	0.00	0.02	0.17	0.00	0.00	0.79
	BE-odds	>9,999	11.44	0.92	>9,999	259.54	0.05
Multi Asset	p-value	0.00	0.00	0.00	0.00	0.00	0.00
	Bayesian-p	0.00	0.00	0.12	0.00	0.00	0.09
	BE-odds	>9,999	>9,999	1.41	>9,999	>9,999	1.91

Panel B: Full sample; 1800-2016

		Trend	Momentum	Value	Carry	Seasonality	BAB
Equities	p-value	0.00	0.00	0.00	0.00	0.00	0.00
	Bayesian-p	0.00	0.00	0.09	0.00	0.00	0.00
	BE-odds	>9,999	>9,999	2.01	>9,999	>9,999	>9,999
Bonds	p-value	0.00	0.00	0.02	0.00	0.00	0.38
	Bayesian-p	0.00	0.06	0.42	0.00	0.00	0.80
	BE-odds	>9,999	2.86	0.26	>9,999	>9,999	0.05
Commodities	p-value	0.00	0.00	0.01	0.00	0.00	0.27
	Bayesian-p	0.00	0.00	0.24	0.17	0.00	0.79
	BE-odds	>9,999	>9,999	0.59	0.94	>9,999	0.05
FX	p-value	0.00	0.00	0.30	0.00	0.00	0.35
	Bayesian-p	0.00	0.00	0.80	0.00	0.00	0.80
	BE-odds	>9,999	58.19	0.05	>9,999	80.63	0.05
Multi Asset	p-value	0.00	0.00	0.00	0.00	0.00	0.00
	Bayesian-p	0.00	0.00	0.00	0.00	0.00	0.00
	BE-odds	>9,999	>9,999	59.52	>9,999	>9,999	148.30

Electronic copy available at: https://ssrn.com/abstract=3325720

Table 6: Sub-period performances of global return factors

The table summarizes the historical sub-period performance of the return factors. Shown per factor per asset class is the percentage of rolling 10-year period with positive Sharpe ratios ("$P(SR_{10y}>0)$"). The sample starts in January 1800 and ends December 2016 and is at the monthly frequency. Covered are equity indices ("Equities"), 10-year maturity government bond indices ("Bonds"), commodities ("Commodities"), currencies ("FX"), and their equally-weighted combination across the four asset classes ("Multi Asset"). Numbers in italics indicate p-values (based on Newey-West to account for 10-year overlapping observations). Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

$P(SR_{10y}>0)$	Trend	Momentum	Value	Carry	Seasonality	BAB
Equity	99%***	92%***	60%	88%***	88%***	72%***
	0.00	*0.00*	*0.17*	*0.00*	*0.00*	*0.01*
Bond	93%***	80%***	69%**	96%***	88%***	46%
	0.00	*0.00*	*0.03*	*0.00*	*0.00*	*0.68*
Commodities	97%***	100%***	67%*	71%**	93%***	48%
	0.00	*0.00*	*0.09*	*0.05*	*0.00*	*0.81*
FX	82%***	84%***	60%	83%***	76%***	42%
	0.00	*0.00*	*0.49*	*0.00*	*0.00*	*0.37*
Multi Asset	100%***	93%***	76%***	92%***	96%***	74%***
	0.00	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*

Electronic copy available at: https://ssrn.com/abstract=3325720

Table 7: Robustness to data quality of global return factors: pre-sample period

The table summarizes the robustness of the historical performance of the global return factors over the pre-sample period to screens and controls on data quality. We consider the following data filters in Panel A: no liquidity screen ("No liquidity screen"), only the zero return liquidity screen ("Zero return screen"), the combination of the zero return, the return interpolation and stale return screens ("Baseline"), and the addition of trimming asset returns at their 1st and 99th percentiles on the Baseline ("+ Trimmed returns"). Panel B shows the global factor returns by quarter ("Q1" to "Q4"), and the results when adding additional markets to the sample ("Survivorship test"). Shown are the historical annualized Sharpe ratios for each factor per asset class and for the equally-weighted multi-asset combination across the four asset classes. The sample starts as early as January 1800 and ends at the asset class-factor specific start date of the sample used in the original studies, see Tables 1 and 2, and is at the monthly frequency. Numbers in parentheses indicate t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

Panel A:

Factor	Asset Class	No liquidity screen		Zero return screen		Baseline		+ Trimmed returns	
Trend	Eq.	0.89***	(12.06)	0.90***	(12.21)	0.90***	(12.15)	0.93***	(12.68)
Trend	Bo.	0.67***	(9.06)	0.55***	(7.50)	0.55***	(7.62)	0.61***	(8.31)
Trend	Co.	0.57***	(5.87)	0.58***	(6.04)	0.58***	(6.04)	0.60***	(6.20)
Trend	FX	0.56***	(7.68)	0.44***	(5.96)	0.44***	(5.38)	0.43***	(5.80)
Momentum	Eq.	0.50***	(6.67)	0.58***	(7.74)	0.58***	(7.72)	0.62***	(8.22)
Momentum	Bo.	0.34***	(4.58)	0.26***	(3.53)	0.26***	(3.25)	0.30***	(3.94)
Momentum	Co.	0.59***	(5.70)	0.62***	(5.99)	0.62***	(5.99)	0.65***	(6.31)
Momentum	FX	0.49***	(6.50)	0.26***	(3.52)	0.26***	(3.75)	0.36***	(4.72)
Value	Eq.	0.20***	(2.66)	0.23***	(3.07)	0.23***	(3.02)	0.23***	(3.10)
Value	Bo.	0.36***	(3.81)	0.28***	(2.97)	0.28*	(1.90)	0.27***	(2.83)
Value	Co.	0.17*	(1.65)	0.24**	(2.29)	0.24**	(2.29)	0.25**	(2.42)
Value	FX	0.51***	(5.29)	0.00	(0.01)	0.00	(-0.02)	-0.02	(-0.17)
Carry	Eq.	0.58***	(7.94)	0.53***	(7.29)	0.53***	(7.38)	0.57***	(7.88)
Carry	Bo.	0.60***	(8.18)	0.53***	(7.06)	0.53***	(7.48)	0.71***	(9.50)
Carry	Co.	0.31***	(3.13)	0.18*	(1.81)	0.18*	(1.81)	0.18*	(1.82)
Carry	FX	0.82***	(11.12)	0.54***	(7.21)	0.54***	(6.60)	0.57***	(7.65)
Seasonality	Eq.	0.57***	(7.53)	0.61***	(8.06)	0.61***	(8.17)	0.64***	(8.46)
Seasonality	Bo.	1.41***	(18.60)	1.07***	(13.96)	1.07***	(12.33)	1.28***	(16.62)
Seasonality	Co.	0.51***	(5.01)	0.54***	(5.27)	0.54***	(5.27)	0.54***	(5.30)
Seasonality	FX	0.56***	(7.39)	0.39***	(5.07)	0.39***	(4.54)	0.37***	(4.84)
BAB	Eq.	0.37***	(4.97)	0.43***	(5.74)	0.43***	(5.64)	0.44***	(5.93)
BAB	Bo.	-0.09	(-1.19)	-0.05	(-0.73)	-0.05	(-0.15)	-0.05	(-0.68)
BAB	Co.	0.07	(0.65)	-0.03	(-0.30)	-0.03	(-0.30)	-0.03	(-0.33)
BAB	FX	0.05	(0.66)	0.11	(1.36)	0.11	(1.13)	0.14	(1.63)
Trend	MA	1.17***	(15.90)	1.06***	(14.43)	1.06***	(14.07)	1.10***	(14.95)
Momentum	MA	0.83***	(10.88)	0.70***	(9.12)	0.70***	(9.14)	0.79***	(10.30)
Value	MA	0.47***	(6.10)	0.29***	(3.80)	0.29***	(3.09)	0.26***	(3.43)
Carry	MA	1.18***	(15.77)	0.90***	(12.10)	0.90***	(11.80)	1.01***	(13.53)
Seasonality	MA	1.47***	(19.50)	1.22***	(16.09)	1.22***	(15.18)	1.32***	(17.43)
BAB	MA	0.19**	(2.43)	0.23***	(2.95)	0.23***	(3.20)	0.25***	(3.28)

Electronic copy available at: https://ssrn.com/abstract=3325720

Panel B:

Factor	Asset Class	Q1		Q2		Q3		Q4		Survivorship test	
Trend	Eq.	0.82***	(5.57)	1.25***	(8.49)	0.90***	(6.09)	0.67***	(4.57)	0.90***	(12.22)
Trend	Bo.	0.86***	(5.86)	0.51***	(3.45)	0.31**	(2.14)	0.53***	(3.60)	0.56***	(7.60)
Trend	Co.	0.57***	(2.92)	0.39**	(2.00)	0.60***	(3.09)	0.82***	(4.24)	0.58***	(6.04)
Trend	FX	0.42***	(2.85)	0.49***	(3.34)	0.29**	(2.00)	0.38**	(2.57)	0.43***	(5.79)
Momentum	Eq.	0.64***	(4.27)	0.58***	(3.86)	0.50***	(3.32)	0.60***	(3.97)	0.56***	(7.49)
Momentum	Bo.	0.44***	(2.89)	-0.02	(-0.14)	0.24	(1.59)	0.31**	(2.04)	0.23***	(3.04)
Momentum	Co.	0.95***	(4.57)	0.28	(1.34)	0.51**	(2.47)	0.82***	(3.96)	0.62***	(5.99)
Momentum	FX	0.21	(1.41)	0.63***	(4.18)	0.00	(-0.01)	0.43***	(2.84)	0.28***	(3.74)
Value	Eq.	0.10	(0.68)	0.37**	(2.44)	0.11	(0.73)	0.34**	(2.25)	0.29***	(3.82)
Value	Bo.	0.35*	(1.83)	0.23	(1.21)	-0.06	(-0.29)	0.28	(1.48)	0.15	(1.57)
Value	Co.	0.11	(0.51)	0.31	(1.48)	0.39*	(1.86)	0.09	(0.44)	0.24**	(2.29)
Value	FX	-0.10	(-0.53)	-0.29	(-1.48)	0.35*	(1.80)	0.00	(0.01)	0.00	(-0.04)
Carry	Eq.	0.50***	(3.46)	0.67***	(4.58)	0.54***	(3.72)	0.44***	(3.03)	0.62***	(8.44)
Carry	Bo.	0.59***	(3.97)	0.34**	(2.25)	0.78***	(5.21)	0.58***	(3.87)	0.56***	(7.57)
Carry	Co.	0.38*	(1.86)	0.16	(0.82)	-0.16	(-0.79)	0.50**	(2.45)	0.18*	(1.81)
Carry	FX	0.68***	(4.52)	0.41***	(2.75)	0.26*	(1.74)	0.72***	(4.81)	0.51***	(6.81)
Seasonality	Eq.	0.67***	(4.45)	0.47***	(3.09)	0.91***	(6.01)	0.44***	(2.93)	0.66***	(8.68)
Seasonality	Bo.	0.62***	(4.01)	0.81***	(5.25)	1.21***	(7.89)	1.25***	(8.16)	0.97***	(12.61)
Seasonality	Co.	0.63***	(3.08)	0.58***	(2.84)	0.40*	(1.93)	0.55***	(2.69)	0.54***	(5.27)
Seasonality	FX	0.60***	(3.94)	0.36**	(2.37)	0.25	(1.59)	0.25*	(1.64)	0.34***	(4.41)
BAB	Eq.	0.73***	(4.91)	0.31**	(2.09)	0.25*	(1.66)	0.35***	(2.32)	0.37***	(4.92)
BAB	Bo.	-0.04	(-0.29)	0.10	(0.66)	-0.06	(-0.40)	-0.06	(-0.40)	-0.02	(-0.28)
BAB	Co.	0.23	(1.08)	-0.22	(-1.04)	-0.26	(-1.25)	0.21	(0.99)	-0.03	(-0.30)
BAB	FX	0.04	(0.25)	0.17	(1.02)	0.16	(0.98)	-0.03	(-0.18)	0.12	(1.39)
Trend	MA	1.14***	(7.77)	1.18***	(7.99)	0.85***	(5.78)	0.98***	(6.65)	1.06***	(14.40)
Momentum	MA	0.85***	(5.57)	0.71***	(4.63)	0.41***	(2.70)	0.88***	(5.77)	0.68***	(8.92)
Value	MA	0.11	(0.73)	0.35**	(2.30)	0.18	(1.20)	0.32**	(2.12)	0.25***	(3.26)
Carry	MA	1.06***	(7.09)	0.73***	(4.91)	0.72***	(4.83)	1.02***	(6.86)	0.93***	(12.53)
Seasonality	MA	1.16***	(7.64)	1.02***	(6.75)	1.30***	(8.57)	1.11***	(7.33)	1.16***	(15.40)
BAB	MA	0.47***	(3.05)	0.16	(1.04)	0.13	(0.86)	0.21	(1.36)	0.23***	(3.00)

Electronic copy available at: https://ssrn.com/abstract=3325720

Table 8: Market risk, common variation and global return factors

The table summarizes the market risk and common variation of the global return factors. Panel A shows per factor per asset class the historical annualized Jensen's divided by its residual volatility (i.e. appraisal ratio). relative to global excess equity, bond, currency and commodity market returns. Panel B contains the average pairwise monthly return correlations across the four asset classes ("Factor"), and per asset class or their equally weighted multi-asset aggregation ("Asset class"). Panel C shows the results of spanning tests for each factor return series per asset class on all other factor return series and the four market factors. Shown is the (annualized) intercept divided by its residual volatility. The sample starts in January 1800 and ends December 2016 and is at the monthly frequency. Numbers in parentheses indicate t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

Panel A: Jensen's alpha

Appraisal Ratio	Trend	Momentum	Value	Carry	Seasonality	BAB
Equities	0.81***	0.58***	0.23***	0.60***	0.54***	0.38***
	(11.95)	(8.54)	(3.37)	(8.79)	(7.88)	(5.58)
Bonds	0.58***	0.23***	0.20**	0.53***	0.77***	0.05
	(8.61)	(3.36)	(2.44)	(7.73)	(11.20)	(0.71)
Commodities	0.65***	0.61***	0.24***	0.25***	0.48***	0.04
	(7.64)	(7.14)	(2.79)	(2.89)	(5.56)	(0.46)
FX	0.38***	0.27***	0.10	0.49***	0.29***	0.03
	(5.56)	(3.87)	(1.16)	(7.09)	(4.21)	(0.43)
Multi Asset	1.02***	0.71***	0.29***	0.89***	0.98***	0.25***
	(14.99)	(10.40)	(4.33)	(13.15)	(14.38)	(3.70)

Panel B: Average pairwise correlations

	Trend	Momentum	Value	Carry	Seasonality	BAB
Factor	0.07***	0.03	0.01	0.01	0.01	0.03*
	Equities	Bonds	Commodities	FX	Multi Asset	
Asset class	0.09***	0.03	0.01	0.08***	0.07***	

Panel C: Spanning tests

Intercept / residual volatility	Trend	Momentum	Value	Carry	Seasonality	BAB
Equities	0.38***	0.14**	0.18***	0.35***	0.40***	0.13*
	(5.56)	(2.02)	(2.59)	(5.13)	(5.84)	(1.85)
Bonds	0.33***	-0.08	0.32***	0.23***	0.67***	-0.05
	(4.90)	(-1.23)	(3.88)	(3.37)	(9.81)	(-0.66)
Commodities	0.35***	0.39***	0.53***	0.08	0.37***	0.03
	(4.16)	(4.61)	(6.15)	(0.97)	(4.35)	(0.31)
FX	0.24***	-0.03	0.15*	0.23***	0.15**	-0.01
	(3.60)	(-0.38)	(1.85)	(3.37)	(2.24)	(-0.15)
Multi Asset	0.62***	0.18***	0.54***	0.49***	0.64***	0.06
	(9.19)	(2.63)	(7.91)	(7.19)	(9.38)	(0.88)

Electronic copy available at: https://ssrn.com/abstract=3325720

Table 9: Downside risk and global return factors

The table shows the results of the downside risk analysis of the global return factors. Shown are the regular CAPM beta (β), downside beta (β^-), (annualized) CAPM alpha (α; in percent), and DR-CAPM alpha (α^-; in percent) for each factor per asset class and their equally-weighted multi-asset aggregation The downside beta is calculated versus the excess return of the equity market portfolio and uses -1 standard deviation as the threshold. The table is sorted on the difference between downside beta (β^-) and regular CAPM beta (β). The sample starts in January 1800 and ends December 2016 and is at the monthly frequency. Numbers in parentheses indicate t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

Factor	Asset Class	β^-	β	$\beta^- - \beta$	α	t-stat	α^-	t-stat
Momentum	Bonds	0.11	0.04	0.07	2.33***	(3.10)	2.06***	(2.72)
Value	FX	-0.02	-0.07	0.05	1.34	(1.43)	1.05	(1.12)
Seasonal	Bonds	0.08	0.03	0.05	8.73***	(11.42)	8.52***	(11.12)
Seasonal	Equities	0.04	0.01	0.04	5.47***	(8.00)	5.34***	(7.81)
Momentum	Equities	0.03	0.00	0.03	5.86***	(8.61)	5.75***	(8.45)
Carry	FX	0.03	0.00	0.03	5.62***	(7.29)	5.51***	(7.15)
Carry	Bonds	0.09	0.06	0.03	5.64***	(7.67)	5.53***	(7.52)
Carry	Equities	0.06	0.04	0.02	5.94***	(8.66)	5.85***	(8.52)
BAB	Equities	-0.05	-0.07	0.02	2.91***	(5.55)	2.86***	(5.45)
Trend	Commodities	0.10	0.08	0.02	6.94***	(7.84)	6.88***	(7.77)
Momentum	Commodities	0.00	-0.01	0.01	6.47***	(7.28)	6.43***	(7.24)
Value	Equities	-0.02	-0.01	-0.01	3.25***	(4.67)	3.29***	(4.73)
Carry	Commodities	0.04	0.05	-0.01	4.80***	(4.93)	4.84***	(4.97)
Trend	Equities	0.09	0.10	-0.01	8.98***	(12.06)	9.02***	(12.11)
BAB	Bonds	0.01	0.02	-0.01	0.49	(0.94)	0.53	(1.02)
Momentum	FX	-0.01	0.01	-0.01	3.21***	(4.09)	3.26***	(4.15)
BAB	Commodities	-0.03	-0.01	-0.02	0.46	(0.59)	0.55	(0.71)
Seasonal	Commodities	-0.07	-0.05	-0.02	5.09***	(5.74)	5.20***	(5.86)
Trend	Bonds	-0.05	-0.02	-0.02	6.72***	(8.46)	6.80***	(8.56)
Seasonal	FX	0.00	0.03	-0.03	3.08***	(4.04)	3.18***	(4.17)
Value	Commodities	-0.07	-0.03	-0.05	2.51***	(2.80)	2.67***	(2.98)
BAB	FX	-0.03	0.02	-0.05	0.16	(0.40)	0.29	(0.72)
Trend	FX	0.05	0.11	-0.06	4.75***	(5.69)	4.98***	(5.96)
Value	Bonds	-0.10	-0.03	-0.08	2.55**	(2.56)	3.00***	(2.99)
Momentum	Multi asset	0.07	0.03	0.04	7.86***	(10.44)	7.71***	(10.23)
Carry	Multi asset	0.09	0.05	0.04	9.74***	(13.23)	9.60***	(13.03)
Seasonal	Multi asset	0.04	0.01	0.02	10.62***	(14.52)	10.54***	(14.41)
BAB	Multi asset	0.06	0.08	-0.02	1.95***	(3.87)	2.03***	(4.01)
Value	Multi asset	-0.07	-0.04	-0.03	4.14***	(5.55)	4.27***	(5.72)
Trend	Multi asset	-0.06	-0.01	-0.05	12.79***	(15.01)	12.97***	(15.22)

Electronic copy available at: https://ssrn.com/abstract=3325720

Table 10: Global return factors in 'good' and 'bad' states

The table summarizes the historical performance of global return factors across various 'good' and 'bad' states based on macroeconomic and market sub-periods. Sub-periods examined are at the annual frequency and include recession versus non-recession, global crisis versus non-crisis, and bear and bull equity markets. Shown are historical (annualized) return per macroeconomic state for each factor-asset class combination and the equally-weighted multi-asset combinations across the four asset classes. The column "Dif." contains the differential factor returns between bad and good states. The final row ("Panel") contains the aggregate impact estimated using panel regressions across all global return factors. The panel regression include index fixed-effects and standard errors clustered by date. The sample starts in January 1800 and ends December 2016 and is at the monthly frequency. Bold numbers indicate significance at the 5% level.

		Recession/expansion			Crisis/non-crisis			Bear/bull market		
Factor	Asset Class	Rec.	Exp.	Dif.	Crisis	Non-Crisis	Dif.	Bear	Bull	Dif.
Trend	Eq.	**8.30**	**9.66**	-1.36	**9.70**	**9.04**	0.66	**8.98**	**9.27**	-0.29
Trend	Bo.	**6.90**	**6.28**	0.62	**5.03**	**6.95**	-1.92	**8.54**	**5.79**	2.75
Trend	Co.	**8.88**	**4.98**	**3.90**	**4.78**	**7.30**	-2.51	**8.05**	**6.06**	1.99
Trend	FX	**4.18**	**5.17**	-0.99	**5.88**	**4.50**	1.39	**5.54**	**4.59**	0.95
Momentum	Eq.	**6.33**	**5.57**	0.76	**6.55**	**5.60**	0.94	**5.47**	**5.95**	-0.48
Momentum	Bo.	2.29	**2.63**	-0.34	-0.59	**3.49**	**-4.08**	-0.09	**3.39**	**-3.48**
Momentum	Co.	**6.47**	**6.33**	0.14	**4.70**	**7.14**	-2.44	**7.66**	**6.01**	1.65
Momentum	FX	2.31	**3.79**	-1.48	**3.38**	**3.25**	0.13	3.01	**3.38**	-0.37
Value	Eq.	**5.24**	0.59	**4.65**	2.30	**2.14**	0.16	**4.31**	1.45	2.86
Value	Bo.	1.70	**2.89**	-1.19	1.75	**2.70**	-0.95	2.73	**2.31**	0.42
Value	Co.	2.05	**2.76**	-0.71	**5.73**	1.02	**4.71**	3.33	**2.22**	1.11
Value	FX	0.29	1.43	-1.14	0.41	1.21	-0.80	-1.72	1.78	-3.50
Carry	Eq.	**7.88**	**5.16**	2.72	2.51	**7.21**	**-4.70**	**5.34**	**6.34**	-1.00
Carry	Bo.	**7.67**	**4.95**	2.72	**3.78**	**6.56**	-2.77	**4.56**	**6.34**	-1.78
Carry	Co.	**3.79**	1.84	1.95	0.56	**3.58**	-3.02	**4.12**	**2.18**	1.94
Carry	FX	**4.15**	**6.39**	-2.24	1.46	**6.90**	**-5.44**	**4.95**	**5.85**	-0.90
Seasonality	Eq.	**6.42**	**5.07**	1.35	**3.53**	**6.16**	-2.63	**5.86**	**5.42**	0.44
Seasonality	Bo.	**10.91**	**7.71**	**3.20**	**5.33**	**9.93**	**-4.59**	**6.93**	**9.46**	-2.53
Seasonality	Co.	**7.86**	**2.91**	**4.94**	**4.24**	**5.14**	-0.91	**9.18**	**3.57**	**5.61**
Seasonality	FX	**2.99**	**3.39**	-0.39	2.66	**3.43**	-0.77	**5.37**	**2.53**	2.84
BAB	Eq.	**2.92**	**3.25**	-0.33	**2.54**	**3.32**	-0.78	1.11	**3.83**	**-2.72**
BAB	Bo.	0.27	0.55	-0.29	1.10	0.25	0.85	-0.08	0.64	-0.71
BAB	Co.	-1.78	**2.61**	**-4.39**	0.34	1.10	-0.76	-1.53	1.57	-3.10
BAB	FX	-0.69	**0.99**	**-1.68**	1.65	-0.03	1.68	-0.68	0.73	-1.41
Trend	MA	**13.17**	**12.51**	0.65	**12.42**	**12.83**	-0.42	**14.45**	**12.15**	2.30
Momentum	MA	**7.94**	**7.95**	-0.01	**6.58**	**8.38**	-1.81	**6.53**	**8.43**	-1.90
Value	MA	**4.28**	**2.37**	1.91	**4.54**	**2.55**	1.99	**3.36**	**2.91**	0.44
Carry	MA	**11.57**	**9.07**	2.50	**4.30**	**11.69**	**-7.39**	**9.01**	**10.23**	-1.22
Seasonality	MA	**13.40**	**9.31**	**4.09**	**7.63**	**11.67**	**-4.05**	**12.32**	**10.16**	2.17
BAB	MA	0.60	**3.09**	**-2.48**	**2.73**	**2.09**	0.64	-0.24	**3.08**	**-3.32**
Panel		**4.86**	**5.17**	-0.31	**4.77**	**5.27**	-0.50	**5.65**	**4.91**	0.74

Electronic copy available at: https://ssrn.com/abstract=3325720

Table 11: Macroeconomic risks and global return factors

The table summarizes the explanatory power of macroeconomic risk for the global return factors using methods outlined in Griffin, Ji and Martin (2003). We explain returns to each global return factor, regressing their returns on global macroeconomic variables. The macroeconomic variables are as in Chen, Roll and Ross (1986): industrial production growth (MP), term premium (UTS), change in expected inflation (DEI), and unexpected inflation (UI). The coefficients and annualized intercept ("Interc. (ann.)") of the regression are shown in the table. We combine the resulting loadings against macroeconomic risks with estimates of risk premiums of these risks (estimated using Fama and MacBeth on all global return factors) to get the predicted return originating from an unconditional macroeconomic risk model ("Predicted"). The table further contains the historical average annual return ("Actual") and the differences with predicted returns (i.e. the unexplained return; "Diff."). The sample starts in February 1869 and ends December 2016 and is at the monthly frequency. Bold numbers indicate significance at the 5% level.

Factor	Asset Class	MP	UTS	DEI	UI	Interc. (ann.)	Actual	Predicted	Diff.
Trend	Eq.	0.01	0.01	-0.11	-0.01	**9.74**	**10.21**	**0.89**	**9.32**
Trend	Bo.	**-0.03**	-0.07	**0.49**	-0.01	**9.27**	**7.87**	**1.53**	**6.34**
Trend	Co.	0.00	0.07	**0.48**	-0.01	**6.00**	**6.52**	**1.46**	**5.05**
Trend	FX	-0.01	0.13	-0.36	-0.02	**4.59**	**5.49**	-0.57	**6.06**
Momentum	Eq.	0.01	-0.03	0.14	0.00	**7.21**	**7.54**	**1.64**	**5.90**
Momentum	Bo.	**0.05**	**-0.15**	0.13	-0.02	**2.57**	**3.77**	**6.61**	**-2.83**
Momentum	Co.	**0.03**	0.01	0.17	-0.01	**5.13**	**6.39**	**3.03**	**3.32**
Momentum	FX	-0.01	0.10	0.04	-0.01	**3.44**	**3.93**	-0.40	**4.35**
Value	Eq.	-0.02	-0.09	0.07	0.00	**4.56**	**3.23**	0.28	**2.95**
Value	Bo.	**-0.05**	0.11	**-0.54**	0.00	**3.58**	**2.41**	**-5.19**	**7.70**
Value	Co.	-0.01	0.04	**-0.51**	-0.01	**2.24**	**2.47**	-0.75	**3.18**
Value	FX	0.01	-0.08	0.13	0.00	1.04	0.96	**1.63**	-0.68
Carry	Eq.	**0.03**	**-0.14**	-0.16	-0.02	**6.49**	**6.83**	**4.61**	**2.22**
Carry	Bo.	**0.03**	-0.02	-0.04	-0.03	**5.95**	**7.34**	**4.39**	**2.94**
Carry	Co.	-0.01	0.06	0.05	-0.01	**2.41**	**2.63**	0.29	**2.33**
Carry	FX	0.01	-0.09	0.30	0.00	**5.80**	**5.57**	**2.69**	**2.88**
Seasonality	Eq.	0.00	**-0.15**	-0.19	-0.03	**6.85**	**6.17**	**3.85**	**2.32**
Seasonality	Bo.	**0.05**	**-0.25**	-0.32	-0.03	**8.66**	**9.25**	**7.61**	1.68
Seasonality	Co.	0.00	-0.03	-0.17	0.01	**5.27**	**4.86**	**-1.00**	**5.87**
Seasonality	FX	0.00	-0.07	0.12	-0.02	**2.42**	**2.34**	**3.63**	-1.30
BAB	Eq.	0.01	**0.16**	0.08	-0.03	**2.82**	**4.71**	**1.77**	**2.94**
BAB	Bo.	**-0.02**	0.06	-0.12	-0.01	0.60	0.21	**-1.62**	**1.83**
BAB	Co.	0.01	0.06	-0.03	-0.01	-0.12	0.86	**1.15**	-0.33
BAB	FX	0.00	0.05	-0.09	**-0.03**	-0.42	0.56	**2.39**	**-2.14**
Trend	MA	-0.02	0.07	0.19	-0.02	**14.70**	**14.92**	**1.31**	**13.61**
Momentum	MA	**0.04**	-0.03	0.23	-0.02	**9.02**	**10.68**	**5.40**	**5.28**
Value	MA	**-0.03**	-0.01	-0.35	-0.01	**5.73**	**4.47**	**-2.06**	**6.54**
Carry	MA	**0.03**	-0.10	0.07	-0.03	**10.25**	**11.15**	**6.18**	**4.97**
Seasonality	MA	**0.02**	**-0.25**	-0.26	**-0.04**	**11.51**	**11.25**	**7.25**	**4.00**
BAB	MA	0.00	**0.17**	-0.08	**-0.03**	1.63	**3.16**	**1.31**	**1.85**

Electronic copy available at: https://ssrn.com/abstract=3325720

Appendix: Factor definitions, portfolio and dataset construction

A.I: Factor definitions

Time-series momentum ('Trend'). Moskowitz, Ooi, and Pedersen (2012) use a 12-month trend measure for their sample of liquid futures contracts after 1985. We skip the last month on trend investing, as this safeguards against potential liquidity issues that might be especially prominent in the earlier parts of our sample.

Cross-sectional momentum ('Momentum'). We follow Asness, Moskowitz, and Pedersen (2013) and use the 12-month-minus-1-month excess return as momentum measure. We skip the last month on cross-sectional momentum investing, as this safeguards against potential liquidity issues and data biases that might be at work in the earlier parts of our sample.

Value ('Value'). For equity, we use the dividend-to-price ratio (D/P), or dividend yield, defined as the past 12-month dividend payment divided by the current price. Other studies typically consider book-to-market value ratios (e.g. Asness, Moskowitz, and Pedersen, 2013), but these are not readily available historically. For government bonds, we use the real yield, which is defined as the 10-year bond yield over the past 1-year inflation, as in Asness, Moskowitz, and Pedersen (2013). For currencies, we use an equally weighted combination of absolute and relative purchasing power parity (PPP), as both are tested in previous studies.[30] Absolute PPP follows Rogoff (1996) and Taylor (2002). Relative PPP is a 5-year reversal of the spot rate corrected for inflation differences defined

[30] Menkhoff, Sarno, Schmeling, Schrimpf (2017) describe more sophisticated currency value measures such as productivity, the quality of export goods, net foreign assets, and output gaps, that we ignore due to limited historical data availability.

Electronic copy available at: https://ssrn.com/abstract=3325720

as in Asness, Moskowitz, and Pedersen (2013). For commodities, we use the 5-year reversal in spot prices, as defined in Asness, Moskowitz, and Pedersen (2013).

Carry ('Carry'). We use the carry definitions as in Koijen, Moskowitz, Pedersen, and Vrugt (2018). For equity we use the excess implied dividend yield priced into the futures versus spot contract. This effectively captures the implied excess dividend yield of an equity index for the month ahead. We splice these series before the existence of equity futures by means of the following method: we regress the monthly dividend yield implicit in the total versus price return indices on month dummies using the past five years of data to predict the dividend yield for the month ahead, and subsequently subtract the risk-free rate. The average correlation between predicted and actual carry numbers over the period that both are available is 54%. For government bonds, we take the slope of the yield curve defined as the 10-year yield minus the short rate and omit the quantitatively smaller roll-down on the interest rate curve because of data limitations. For currencies, we use the short-term yield differential, which we infer from forwards and before their availability from short-term money market rates, and for commodities the slope of the futures curve.

Return seasonality ('Seasonality'). Our definition follows Keloharju, Linnainmaa, and Nyberg (2016): we use the return on an asset in a certain month over the prior 20 years (requiring at least 12-months of data). Assets that did relatively well over a particular month in the past are likely to do relatively well in the same month going forward. For example, in January the monthly equity index seasonal will buy those equity markets that had the best relative performance in January during the past 20 years and shorts those with lowest relative January performance.

Betting-against-beta ('BAB'). This factor postulates that low-beta securities outperform high-beta securities on a beta-adjusted basis (Frazzini and Pedersen, 2014). We test this factor by estimating the betas over a 36-month period (requiring at least 12-

Electronic copy available at: https://ssrn.com/abstract=3325720

months of data) relative to the global asset class level portfolio return (we refer to Appendix A.III for details about the construction of the global asset class portfolio). The position sizes of each short and long leg are chosen such that the ex-ante betas are the same, so that the excess return contains as little market effect as possible.

A.II Portfolio construction procedure

After obtaining the factor measures per market in each investment universe we construct factor investment portfolios at the end of every month in the following manner. For the trend factor, which is directional in nature, we go long (short) markets in each asset class when the trend measure is positive (negative), following Moskowitz, Ooi, and Pedersen (2012). For the other factors, which are all cross-sectional in nature, we rank the markets in each investment universe based on the factor measure and take a position equal to the rank minus its cross-sectional average (requiring a minimum of two markets to be present). This procedure is similar to that used by Asness, Moskowitz, and Pedersen (2013), Frazzini and Pedersen (2014), and Koijen, Moskowitz, Pedersen, and Vrugt (2018). (By contrast, Keloharju, Linnainmaa, and Nyberg (2016) construct their return seasonality strategy via long the top quintile and short the bottom quintile at each point in time.) Consequently, positions for all factors, except trend, add up to zero at each point in time:

$$\boldsymbol{w_t^i = z_t} \cdot \left(\boldsymbol{Rank(S_t^i)} - \frac{\boldsymbol{N_t + 1}}{\boldsymbol{2}}\right),$$

with $\boldsymbol{w_t^i}$ the weight of asset i at time t, $\boldsymbol{S_t^i}$ the factor signal, $\boldsymbol{N_t}$ the number of assets in the cross-section, and $\boldsymbol{z_t}$ a scaling factor to ensure that the portfolio sums to zero.

Next, we size positions in each market in each asset class by its simple 3-year rolling volatility estimate or beta estimate (for BAB only), in the same spirit as Asness,

Electronic copy available at: https://ssrn.com/abstract=3325720

Moskowitz, and Pedersen (2013), Frazzini and Pedersen (2014), and Moskowitz, Ooi, and Pedersen (2012), but fitted to our sample frequency (i.e. monthly data). To prevent undue impact from extremely low volatility estimates (and hence keep the factor strategy robust from an investor perspective), especially in the earlier part of our sample, we floor each volatility (beta) estimate at the maximum of the 10% quantile of volatility (beta) estimates per asset class or 2.5% (0.25), whichever is greater.

We subsequently sum the product of position, sizes, and market returns across markets within an asset class for each date to generate the return on the factor strategy per asset class. We then adjust the position sizes of each of the factor strategies per asset class using a 10-year rolling window such that each factor strategy-asset class combination targets an ex-ante volatility of 10% per annum (an adjustment that implicitly accounts for non-perfect correlations between markets). We floor this estimate at 2.5% to prevent extreme leverage. This approach takes an ex-ante view of portfolio construction, as available in real time. However, our results are not materially different if we simply scale by in-sample ex-post volatility, as in Koijen, Moskowitz, Pedersen and Vrugt (2018).

We rebalance the portfolios each month based on the signals and volatility estimates. This methodology results in balanced long-short portfolios that are per factor comparable across asset classes, and which will facilitate combining multiple asset classes per factor. Subsequently, we construct 'Multi Asset' factor portfolios by taking an equally weighted average of factor portfolios within each asset class and applying a scaling factor equal to the square root of the number of factor series present.[31]

[31] Note that we thereby implicitly assume factor series are uncorrelated across an asset class, an assumption we make for the sake of simplicity. We realize that this assumption is sometimes at odds with the data, but we have verified that this choice does not materially impact our conclusions.

Electronic copy available at: https://ssrn.com/abstract=3325720

A.III Dataset construction

We have compiled our data from several sources in order to obtain a reliable and historically extensive dataset. Our sample covers 217 years of data from 31 December 1799 through 31 December 2016. We obtain the most recent historical data on financial market prices and macroeconomic series from Bloomberg, Datastream and the OECD website, and splice these before inception with data from (in order of preference): Global Financial Data or monthly commodity futures data from Chicago Board of Trade (CBOT) annual reports (1877-1962) obtained from TwoCenturies.com, the Jordà-Schularick-Taylor Macrohistory Database, and/or Jeremy Siegel's website. Below we outline the data sources and the construction of each series we use in detail.

Equities: we source price and return data of equity futures and indices from Bloomberg, Datastream and Global Financial Data. Our primary source is the futures from Bloomberg, with gaps filled in by Datastream data, and spliced before futures inception with index-level data, as in Baltussen, Da, and Van Bekkum (2019). Next, we backfill these data with equity index level data downloaded from Global Financial Data. We obtain dividend yields from the same sources. For carry, we use the spot, front futures, and second futures prices. Before we have data on futures, we reconstruct the monthly 'implied carry' as if these markets had listed futures using the regression methodology on the difference between total return and price indices as mentioned in the previous section. The markets we consider are spread around the globe and cover the major developed markets with substantial data history. Online Appendix Table A.4 shows the construction of each equity series.

Bonds: we source bond futures price and return data from Bloomberg, splice these with bond index-level data from Datastream, backfilled before inception with Global Financial Data. From the same sources we obtain yields, and inflation data, the latter extended where possible with data from Macrohistory.net. We apply a two-months lag to

Electronic copy available at: https://ssrn.com/abstract=3325720

inflation numbers to mimic their real-time availability. The markets we consider are the major developed bond markets around the globe. Online Appendix Table A.5 shows the construction of each bond series.

Currencies: forward and spot prices are primarily from Datastream, spliced with Bloomberg data and Global Financial Data. Before the availability of forward rates, we use short rates from Bloomberg, Datastream, Global Financial Data, and for the U.S., data from Jeremy Siegel. Purchasing power parity data is obtained from the OECD website, and before 1971 with data from Macrohistory.net. We include the major developed currency markets (or 'G10') in our sample (being USD, GBP, EUR (before 1999: BEF, DEM, ESP, FRF, ITL, and NLG), JPY, CHF, CAD, AUD, NZD, SEK, and NOK), all versus the USD. Online Appendix Table A.6 shows the construction of each currency series.

Commodities: we source commodity futures price and return data from Bloomberg, spliced with monthly commodity futures data from Chicago Board of Trade (CBOT) annual reports (1877-1962) obtained from TwoCenturies.com. For the commodity value measure, we use commodity spot prices from Bloomberg and Datastream, spliced with spot data from Global Financial Data and TwoCenturies.com. For carry, we use the front futures and second futures prices. We use the main contracts based on their general usage and liquidity. Due to restrictions on tradability, we exclude gold as a speculative asset during the currency gold standard up to the end of the Bretton Woods system (which was effectively a gold standard). Online Appendix Table A.7 shows the construction of each commodity series.

Financing rates: our main measure for the financing rates are short-term LIBOR rates (sourced from Bloomberg and Datastream), spliced with (in order of usage) Eurodollar rates from Datastream, short-term Treasury bill rates and commercial paper yields from Global Financial Data, short rates from Macrohistory.net and for the U.S. with data from

Electronic copy available at: https://ssrn.com/abstract=3325720

Jeremy Siegel. When all are unavailable, we splice with lagged Treasury-bill returns. Online Appendix Table A.8 describes the sources per market in detail.

Global asset class portfolios: for equity indices we use the market-value weighted portfolio of equities, spliced before its data existence in 1926 with an equal-weighted portfolio across all equity markets included in the sample. For bonds, we use a GDP-weighted global bond portfolio. For currencies, we use an equal-weighted portfolio of all currencies included in our sample versus the US dollar, and for commodities we also use an equal-weighted portfolio of all commodity futures included in our sample.

Economics: we construct our global recession data from a splicing of the OECD G7 recession indicator from the OECD website (1960-2016), the NBER US recession indicator from the NBER website (1864-1959), and the contraction of real-GDP from Global Financial Data (1800-1863). Inflation data is described above. We obtain the historical data on crisis periods from Carmen Reinhart and Kenneth Rogoff, using their Banks, Currency, Default, Inflation (BCDI) index, which starts in 1800.[32] International macroeconomic data are from GFD and the OECD. The Chen, Roll and Ross (1986) UTS factor is constructed as the yield on the (more than) 10-year maturity government bond minus the 3-month T-bill rate. The factor MP is the log difference in industrial production led by 1-month. Expected inflation and unexpected inflation are calculated following Fama and Gibbons (1984).

Data quality: the deep historical data tends to be of lesser quality compared to the more recent data, as digital archives and the use of indices with strong requirements on data processes did not exist. Instead, data was maintained typically by exchanges, statistical agencies, newspapers and investor annuals, often in manual writing. Potential data quality issues that could be at work include:

[32] http://www.reinhartandrogoff.com/data/

Electronic copy available at: https://ssrn.com/abstract=3325720

- Misprints and other measurement errors. This could cause prices to be spuriously inflated, trigger potential reversal (value) profits.
- Reported prices in our databases are not necessarily transaction prices, but bid prices, ask prices, average prices of the day or month, or average of daily or monthly high and low prices. The use of bid or ask prices creates artificial short-term reversal effects, while the use of average prices over a month creates an artificial AR(1) process (see Working, 1960, Schwert, 1990). Working (1960) shows that such time averaging does not induce autocorrelation beyond a one-month horizon, and therefore does not preclude testing for momentum effects, provided that one skips a month between the end of the formation period and the beginning of the holding period.
- Missing data, which have sometimes been solved by interpolating, or padding, prices or returns known at a lower frequency to the monthly frequency. For example, bond prices are sometimes constructed based on interpolated yields.
- The timing of equity dividends and bond coupons were not always known historically. As a solution, to construct return series, they have sometimes been distributed to fixed points over the year, often year ends. For equities this can result in high returns during 'assigned dividend' months, while returns may be artificially low on the actual ex-dividend months (as prices may drop to reflect the dividend payment). For bonds this poses an issue when bond prices are dirty prices (meaning the coupon is embedded in the price), causing artificial drops in bond prices after coupon payments. This could especially affect the seasonality factor (as returns are artificially inflated at a seasonal cycle).
- Equity indices consisting historically of a handful of stocks, or being equal- or price-weighted. Consequently, indices are less diversified, or more tilted to smaller assets.

Electronic copy available at: https://ssrn.com/abstract=3325720

Data cleaning procedure: we have taken the following steps to check the quality of each data series and clean for obvious measurement errors. First, we run a studentized residuals outlier test on each series to capture potential data errors, and visually checked each series on jumps and outliers. Potential outliers were manually verified (by comparing the other data sources where possible and by searching for reasons for large price moves), and when due to a data error corrected. The correction in pre-sample data points include:

- German equities have a misprint in the total return index on 1923-01-31, which we replaced by the total return based on the price return and dividend yield in this month.
- Dutch government bond returns have a misprint on 1964-11-30, which we replaced by the approximation based on changes in yields times duration.
- Several systematic corrections for dividend yield, financing rate and bond yields, as some yield switch between decimals and percentage notations.

Next, we compared the statistics of each series where possible with comparable series from other databases (like the global historical financial markets database of Dimson, Marsh, and Staunton (starting in 1900) and the Jordà-Schularick-Taylor Macrohistory Database (starting in 1870), as well as specific data series for the United States (data of Schwert starting in 1802) and the United Kingdom (data from John Turner)). We find that asset class premiums are generally of comparable magnitudes across databases, and over our pre-sample period (and sub-periods) at roughly similar levels as over our in-sample period (i.e. 1981-2011). Further, we checked each series on gaps, the level and dynamics in the first- and second-order autocorrelations. Series with odds patterns were removed. These include equity returns in South Korea before 1965 and dividend yields in Switzerland between 1918 and 1939.

Further, we applied a number of ex-ante screens on our data series and remove data points when they do not pass these screens (including a hyperinflation screen), as outlined in Section III.A of the paper. These screens reduce the impact of data issues such as

Electronic copy available at: https://ssrn.com/abstract=3325720

missing monthly data, reduced liquidity or non-tradability due to currency pegs ('zero return screen'), the possibility that prices or returns known at a lower frequency have been interpolated to the monthly frequency ('return interpolation screen'), and the possibility that returns are stale or update infrequently ('stale return screen'). These screens eliminate 13% of the pre-1980 equity observations, 46% of the bond observations, 10% of the commodity observations, and 42% of the currency observations. The large bulk of this is due to the zero-return screen, eliminating missing bond returns and observations subject to currency pegs.

Summary statistics: Online Appendix A contains more details about our dataset. Table A.1 summarizes the key statistics of the return series per asset over the full sample period, the in-sample period as proxied by approximately the earliest start date and latest end date in Table 1 (1972-2012), and pre-sample period (1800-1971). Shown are the start dates of each series (after the three data screens), the average (annualized) return and volatility. Table A.2 show the start dates of each factor series (after data screens), while Table A.3 shows the start dates of each individual market return (before data screens), value and carry series. Figures A.1 to A.4 shows the historical monthly returns on each return series used in the main text of the paper, with in shaded grey the included sample periods.

Electronic copy available at: https://ssrn.com/abstract=3325720

References

Abuaf, N., Jorion, P., 1990. Purchasing power parity in the long run. Journal of Finance 45(1), 157-174.

Accominotti, O., 2016. Foreign exchange markets and currency speculation: historical perspectives. In: Chambers, D., Dimson, E. (Eds.) Financial market history: Reflections on the past for investors today. CFA Research Institute Foundation, Chapter 4.

Accominotti, O. Chambers, D., 2016. If you're so smart: John Maynard Keynes and currency speculation in the interwar years. Journal of Economic History 76(2), 342-386.

Albuquerque, R., Eichenbaum, M., Papanikolaou, D., Rebelo, S., 2015. Long-run bulls and bears. Journal of Monetary Economics 76, S21-S36.

Arnott, R., Harvey, C., Markowitz, H., 2019. A backtesting protocol in the era of machine learning. Journal of Financial Data Science 1(1), 64-74.

Asness, C., Moskowitz, T., Pedersen, L., 2013. Value and momentum everywhere. Journal of Finance 68(3), 929-985.

Avramov, D., Chordia, T., Goyal, A., 2006. Liquidity and autocorrelations in individual stock returns. Journal of Finance 61(5), 2365-2394.

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. Journal of Finance 61, 1645-1680.

Baltussen, G., Da, Z., Van Bekkum, S., 2019. Indexing and stock market serial dependence around the world. Journal of Financial Economics 132(1), 26-48.

Baltussen, G., Post, G., Van Vliet, P., 2012. Downside risk aversion, fixed-income exposure, and the value premium puzzle. Journal of Banking and Finance 36(12), 3382-3398.

Banz, R., Breen, W., 1986. Sample-dependent results using accounting and market data: some evidence. Journal of Finance 41(4), 779-793.

Barro, R., Ursúa, J., 2008. Macroeconomic crises since 1870. Brookings Papers on Economic Activity 39(1), 255-350.

Barroso, P., Santa-Clara, P., 2015. Beyond the carry trade: optimal currency portfolios. Journal of Financial and Quantitative Analysis 50(5), 1037-1056.

Bayarri, M., Benjamin, D., Berger, J., Sellke, T., 2016. Rejection odds and rejection ratios: a proposal for statistical practice in testing hypotheses. Journal of Mathematical Psychology 72, 90-103.

Bayarri, M., Berger, J., 1998. Quantifying surprise in the data and model verification. In: Bernardo, J., Berger, J., David, A., Smith, A. (Eds.), Bayesian Statistics 6, Oxford University Press: Oxford, U.K.

Bawa, V., Lindenberg, E., 1977. Capital market equilibrium in a mean-lower partial moment framework. Journal of Financial Economics 5(2), 189-200.

Electronic copy available at: https://ssrn.com/abstract=3325720

Begley, C., Ellis, L., 2012. Raise standards for preclinical cancer research. Nature 483, 531–533.

Benjamin, D., Berger, J., Johannesson, M., Nosek, B., Wagenmakers, E., Berk, R., Bollen, K., Brembs, B., Brown, L., Camerer, C., Cesarini, D., Chambers, C., Clyde, M., Cook, T., De Boeck, P., Dienes, Z., Dreber, A., Easwaran, K., Efferson, C., Fehr, E., Fidler, F., Field, A., Forster, M., George, E., Gonzalez, R., Goodman, S., Green, E., Green, D., Greenwald, A., Hadfield, J., Hedges, L., Held, L., Ho, T., Hoijtink, H., Hruschka, D., Imai, K., Imbens, G., Ioannidis, J., Jeon, M., Jones, J., Kirchler, M., Laibson, D., List, J., Little, R., Lupia, A., Machery, E., Maxwell, S., McCarthy, M., Moore, D., Morgan, S., Munafó, M., Nakagawa, S., Nyhan, B., Parker, T., Pericchi, L., Perugini, M., Rouder, J., Rousseau, J., Savalei, V., Schönbrodt, F., Sellke, T., Sinclair, B., Tingley, D., Van Zandt, T., Vazire, S., Watts, D., Winship, C., Wolpert, R., Xie, Y., Young, C., Zinman, J., Johnson, V., 2018. Redefine statistical significance. Nature Human Behaviour 2, 6-10.

Berk, J., and Van Binsbergen, J., 2015. Measuring skill in the mutual fund industry. Journal of Financial Economics 118(1), 1-20.

Black , F., 1974. International capital market equilibrium with investment barriers. Journal of Financial Econometrics 1, 337-352.

Blitz, D., van Vliet, P., 2007. The volatility effect. Journal of Portfolio Management 34(1), 102-113.

Brown, S., Goetzmann, W., Ross, S., 1995. Survival. Journal of Finance 50(3), 853-873.

Cagan, P., 1956. The monetary dynamics of hyperinflation. In: Friedman, M. (Ed.) Studies in the quantity theory of money, 25-117. Chicago: University of Chicago Press, Chicago.

Camerer, C., Dreber, A., Forsell, E., Ho, T., Huber, J., Johannesson, M., ... Heikensten, E., 2016. Evaluating replicability of laboratory experiments in economics. Science 351(6280), 1433-1436.

Chambers, D., Dimson, E., 2016. Financial market history: reflections on the past for investors today. CFA Research Institute Foundation.

Chen, N.-F., Roll., R., Ross, S., 1986. Economic forces and the stock market, Journal of Business 59(3), 383-403.

Chordia, T., Goyal, A., Saretto, A., 2019. Anomalies and false rejections, Review of Financial Studies, in press.

Danielsson, J., Valenzuela, M., Zer, I., 2018. Learning from history: volatility and financial crises. Review of Financial Studies 31(7), 2774-2805.

De Roon, F., Nijman, T., 2001. Testing for mean-variance spanning: a survey. Journal of Empirical Finance 8, 111-155.

Dimson, E., Marsh, P., Staunton, M., 2008. The worldwide equity premium: a smaller puzzle. In: Mehra, R. (Ed), Handbook of the equity risk premium. Elsevier, Chapter 11.

Doskov, N., Swinkels, L., 2015. Empirical evidence on the currency carry trade, 1900–2012. Journal of International Money and Finance 51, 370-389.

Electronic copy available at: https://ssrn.com/abstract=3325720

Edwards, W., Lindman, H., and Savage, L., 1963. Bayesian statistical inference for psychological research. Psychological Review 70, 193-242.

Erb, C., Harvey, C., 2006. The strategic and tactical value of commodity futures. Financial Analysts Journal 62(2), 69-97.

Fama, E., 1984. Term premiums in bond returns. Journal of Financial Economics 13(4), 529-546.

Fama, E., French, K., 1989. Business conditions and expected returns on stocks and bonds. Journal of Financial Economics 25, 23-49.

Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33(1), 3-56.

Fama, E., Gibbons, M., 1984. A comparison of inflation forecasts. Journal of Monetary Economics 13, 327-348.

Fama, E., MacBeth, J., 1973. Risk, return, and equilibrium: empirical tests. Journal of Political Economy 81(3), 607–636.

Ferson, W., Harvey, C., 1991. The variation of economic risk premiums. Journal of Political Economy 99, 385-415.

Fitzgibbons, S., Friedman, J., Pomorski, L., Serban, L., 2017. Long-only style investing: don't just mix, integrate. Journal of Investing 26(4), 153-164.

Frazzini, A., Pedersen, L., 2014. Betting against beta. Journal of Financial Economics 111(1), 1-25.

Froot, K., Thaler, R., 1990. Anomalies: foreign exchange. Journal of Economic Perspectives 4(3), 179-192.

Garcia, D., 2013. Sentiment during recessions. Journal of Finance 68(3), 1267-1299.

Geczy, C., Samonov, M., 2016. Two centuries of price-return momentum. Financial Analysts Journal 72(5), 32-56.

Geczy, C., Samonov, M., 2017. Two centuries of multi-asset momentum (equities, bonds, currencies, commodities, sectors and stocks). Unpublished working paper.

Goetzmann, W., 1993. Patterns in three centuries of stock market prices. Journal of Business 66(2), 249-270.

Goetzmann, W., Huang, S., 2018. Momentum in Imperial Russia. Journal of Financial Economics 130(3), 579-591.

Goetzmann, W., Kim, D., 2018. Negative bubbles: what happens after a crash. European Financial Management 24(2), 171-191.

Goetzmann, W., Li, L., Rouwenhorst, K., 2005. Long-term global market correlations. Journal of Business 78(1), 1-38.

Golez, B., Koudijs, P., 2018. Four centuries of return predictability. Journal of Financial Economics 127(2), 248-263.

Electronic copy available at: https://ssrn.com/abstract=3325720

Goyal, A., and Jegadeesh, N., 2018. Cross-sectional and time-series tests of return predictability: what is the difference? Review of Financial Studies 31(5), 1784-1824.

Griffin, J., Ji, X., Martin, J., 2003. Momentum investing and business cycle risk: evidence from pole to pole. Journal of Finance 58(6), 2515-2547.

Harrison, P., 1998. Similarities in the distribution of stock market price changes between the eighteenth and twentieth centuries. Journal of Business 71, 55-79.

Harvey, C., 2017. Presidential address: the scientific outlook in financial economics. Journal of Finance 72(4), 1399-1440.

Harvey, C., Liu, Y., Zhu, H., 2016. … and the cross-section of expected returns. Review of Financial Studies 29(1), 5-68.

Hjalmarsson, E., 2010. Predicting global stock returns. Journal of Quantitative and Financial Analysis 45(1), 49-80.

Hou, K., Xue, C., and Zhang, L., 2020. Replicating anomalies. Review of Financial Studies 33(5), 2019-2133.

Hurst, B., Ooi, Y., Pedersen, L., 2017. A century of evidence on trend-following investing. Journal of Portfolio Management 44(1), 15-29.

Ilmanen, A., 1995. Time-varying expected returns in international bond markets. Journal of Finance 50(2), 481-506.

Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. Journal of Finance 48(1), 65-91.

Jensen, M., 1968. The performance of mutual funds in the period 1945-1964. Journal of Finance 23(2), 389-416.

Jobst, C., 2009. Market leader: the Austro-Hungarian Bank and the making of foreign exchange intervention, 1896-1913. European Review of Economic History 13(3), 287-318.

Jordà O., Schularick, M., Taylor, A., 2017. Macrofinancial history and the new business cycle facts. In: Eichenbaum, M., Parker, J., (Eds) NBER Macroeconomics Annual 2016. Chicago: University of Chicago Press, Volume 31.

Keloharju, M., Linnainmaa, J., Nyberg, P., 2016. Return seasonalities. Journal of Finance 71(4), 1557-1590.

Keloharju, M., Linnainmaa, J., Nyberg, P., 2018. Seasonal reversals in expected stock returns. Unpublished working Paper.

Koijen, R., Moskowitz, T., Pedersen, L., Vrugt, E., 2018. Carry. Journal of Financial Economics 127(2), 197-225.

Karolyi, G., Wu, Y., 2018. A new partial-segmentation approach to modeling international stock returns. Journal of Financial and Quantitative Analysis 53(2), 507-546.

Korajczyk, R., Sadka, R., 2004. Are momentum profits robust to trading costs? Journal of Finance 59 (3), 1039-1082.

Electronic copy available at: https://ssrn.com/abstract=3325720

Koudijs, P., 2016. The boats that did not sail: asset price volatility in a natural experiment. Journal of Finance 71(3), 1185-1226.

Lesmond, D., Ogden, J., Trzcinka, C., 1999. A new estimate of transaction costs. Review of Financial Studies 12, 1113-1141.

Lettau, M., Maggiori, M., Weber, M., 2014. Conditional risk premiums in currency markets and other asset classes. Journal of Financial Economics 114(2), 197-225.

Linnainmaa, J., Roberts, M., 2018. The history of the cross section of stock returns. Review of Financial Studies 31(7), 2606-2649.

Little, R., 1988. Missing-data adjustments in large surveys. Journal of Business and Economic Statistics 6(3), 287–296.

McLean, R., and Pontiff, J., 2016. Does academic research destroy stock return predictability? Journal of Finance 71(1), 5-32.

Menkhoff, L., Sarno, L., Schmeling, M., Schrimpf, A., 2012. Currency momentum strategies. Journal of Financial Economics 106(3), 660-684.

Menkhoff, L., Sarno, L., Schmeling, M., Schrimpf, A., 2017. Currency value. Review of Financial Studies 30(2), 416-441.

Merton, R., 1973. An intertemporal capital asset pricing model. Econometrica 41, 867-887.

Moskowitz, T., Ooi, Y., Pedersen, L., 2012. Time series momentum. Journal of Financial Economics 104(2), 228-250.

Muir, T., 2017. Financial crises and risk premia. Quarterly Journal of Economics 132(2), 765-809.

Neal, L., 1985. Integration of international capital markets: quantitative evidence from the eighteenth to twentieth centuries. Journal of Economic History 45(2), 219-226.

Neal, L., 1987. The integration and efficiency of the London and Amsterdam stock markets in the eighteenth century. Journal of Economic History 47(1), 97-115.

Novy-Marx, R., Velikov, M., 2015. A taxonomy of anomalies and their trading costs. Review of Financial Studies 29(1), 104-147.

Novy-Marx, R., Velikov, M., 2018. Betting against betting against beta. Unpublished working paper.

Open Science Collaboration, 2015. Estimating the reproducibility of psychological science. Science 349(6251), aac4716.

Poitras, G., 2009. From Antwerp to Chicago: the history of exchange traded derivative security contracts. Revue d'Histoire des Sciences Humaines 20(1), 11-50.

Post, T., Van Vliet, P., 2006. Downside risk and asset pricing. Journal of Banking and Finance 30(3), 823-849.

Rogoff, K., 1996. The purchasing power parity puzzle. Journal of Economic Literature 34(2), 647-668.

Electronic copy available at: https://ssrn.com/abstract=3325720

Rubin, D., 1986. Statistical matching using file concatenation with adjusted weights and multiple imputations". Journal of Business and Economic Statistics 4(1), 87–94.

Schwert, G., 1990. Indexes of United States stock prices from 1802 to 1987. Journal of Business 63(3), 399-426.

Schwert, G., 2011. Stock volatility during the recent financial crisis. European Financial Management 17(5), 789-805.

Shumway, T., 1997. The delisting bias in CRSP data. Journal of Finance 52(1), 327-340.

Shumway, T., Warther, V., 1999. The delisting bias in CRSP's Nasdaq data and its implications for the size effect. Journal of Finance 54(6), 2361-2379.

Siegel, J., 1992. The equity premium: stock and bond returns since 1802. Financial Analysts Journal 48(1), 28-38.

Stambaugh, R., Yu, J., Yuan,Y., 2012. The short of it: investor sentiment and anomalies. Journal of Financial Economics 104, 288–320.

Stulz, R., 1981. On the effects of barriers to international investment. Journal of Finance 36, 923–934.

Taylor, A., 2002. A century of purchasing-power parity. Review of Economics and Statistics 84(1), 139-150.

Welch, I., 2019. Reproducing, extending, updating, replicating, reexamining, and reconciling. Critical Finance Review 8(1-2), 301-304.

Working, H., 1960. Note on the correlation of first differences of averages in a random chain. Econometrica 28(4), 916-918.

Zhang, C., Jacobsen, B., 2013. Are monthly seasonals real? A three-century perspective. Review of Finance 17(5), 1743-1785.

Zinkina, J., Andreev, A. Mosakova, E., 2017. The nineteenth century as the cradle of capitalism. In: Grinin, L., Ilyin, I., Herrmann, P., and Korotayev, A. (Eds), Globalistic and globalization studies: Global evolution, historical globalistics and globalization studies. (eds) 'Uchitel' Publishing House, Volgograd, 173-180.

Electronic copy available at: https://ssrn.com/abstract=3325720