P-hacking? Global return factors tested since 1800

Guido Baltussen*

Laurens Swinkels*

Pim Van Vliet*

Abstract

We test six global asset pricing factors (time-series momentum, cross-sectional momentum, value, carry, return seasonality and betting-against-beta) applied to four major asset classes (equities, bonds, commodities and currencies) for a deep historical database going back to 1800. Using a unified testing framework and methods to account for p-hacking, we find the large majority of factors to be highly significant using strict economic, statistical and persistence criteria. T-statistics are generally well above 3, "Bayesianized" p-values that alleviate p-hacking concerns are close to zero, the factors are persistent across subsamples and are generally of similar strength out-of-sample. Further, we find that time-series momentum crowds out cross-sectional momentum. Our findings seem hard to reconcile with explanations based on market risk, downside risk or macroeconomic risks. These results indicate several 'true' global return factors remain that present a challenge to classical asset pricing theories.

First version: April 2017

This Version: February 2018

Keywords: multiple hypotheses testing, p-hacking, return anomalies, predictability, stocks, bonds, currencies, commodities, value, momentum, trend, carry, betting-against-beta, seasonality, style investing.

JEL classification: C11, C12, F31, G11, G12, G15, N210

*Baltussen (corresponding author) and Swinkels are from the Erasmus School of Economics, Erasmus University Rotterdam and Robeco Institutional Asset Management. Email: baltussen@ese.eur.nl Van Vliet and Swinkels are from Robeco Institutional Asset Management. We would like to thank David Blitz, Andrea Frazzini, Winfried Hallerbach, Matthias Hanauer, Matti Keloharju, Ralph Koijen, Martin Martens, Lasse Pedersen and Jeroen van Zundert for valuable discussions. The views expressed in this paper are not necessarily shared by Robeco Institutional Asset Management. We welcome comments, including references to related papers we have inadvertently overlooked.

I. Introduction

In this paper we study the presence of the return anomalies over a long and wide sample spanning the last 217 years across the major markets in equities, bonds, currencies, and commodities. Several recent influential studies document multi-asset anomalies present at the index-level. Most importantly, Moskowitz, Ooi, and Pedersen (2012) show the presence of a 'time-series momentum' anomaly, Asness, Moskowitz, and Pedersen (2013) reveal cross-sectional 'momentum' and 'value' anomalies, Koijen, Moskowitz, Pedersen, and Vrugt (2017) find a 'carry' anomaly, Keloharju, Linnainmaa, and Nyberg (2016) discover a return seasonality anomaly, and Frazzini and Pedersen (2014) document a 'betting-against-beta' (BAB) low-risk anomaly. These studies start around 1965 or later and confirm and extend earlier empirical asset pricing studies which often focus on a single asset class, most often U.S. equities.[1]

On the other hand, a serious concern in finance, and science in general, is 'p-hacking' (see Harvey, 2017). As scientists, we are subject to statistical testing limitations, we are biased in the return anomalies we see published, we have several degrees of freedom (on areas like data manipulation, statistical method, aggregation schemes, results we choose to present) and an incentive to publish. Consequently, several major findings might reflect (a combination of) cherry-picking the most significant results, a publication bias, a multiple hypotheses testing bias, and a type I error in testing (i.e. falsely discovering predictability). 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,*

[1] Several papers show the presence of return factors 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 a 'value' and 'carry' factor.

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."[2]

As a case in a point, Harvey, Liu, and Zhu (2016) find a clear publication bias pattern in the top finance journals, and that of over 300 documented stock-level anomalies many become questionable after analyzing these in a rigorous testing framework that allows for multiple hypotheses testing bias. Hou, Xue, and Zhang (2017) 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 (2017) 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 (2017) show that when tested out-of-sample many equity anomalies are weak and of those that persist typically do at about half of its original size.

The main purpose of this paper is to provide robust and rigorous evidence on the presence or absence of documented anomalies in multi-asset markets from the perspective of 'p-hacking'. Our starting point is six multi-asset return factors which have been published in the Journal of Finance and the Journal of Financial Economics during the period 2012-2017; time-series momentum, cross-sectional momentum, value, carry, return seasonality and betting-against-beta.[3] All studies reported p-values well below 5%

[2] P-hacking is not limited to financial economics. P-hacking is mostly discussed in social sciences and medicine. The Economist discussed the topic in 2013 with the headline title: 'How science goes wrong'. 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 significant psychological studies only 36 could be replicated. In behavioral economics, Camerer et al (2016) find that out of 18 laboratory studies in economics 11 can be replicated.

[3] More specifically, Moskowitz, Ooi and Pedersen (2012, JFE) test 'time-series momentum' and equity futures since 1965, bond futures since 1979, currency forwards since 1971, and commodities futures data since 1965. Asness, Moskowitz and Pederson (2013, JF) test 'cross-sectional momentum' and 'value' using U.S. and international stocks since 1971, and using futures data since 1972 for commodities, 1978 for equity indices, 1979 for currencies, and 1981 for bonds. The global multi-asset 'carry' factor in Koijen, Moskowitz, Pedersen and Vrugt (2017, JFE) starts in 1972. The Keloharju, Linnainmaa, and Nyberg (2016, JF) return 'seasonality' factor uses data since 1963 for individual stocks and data since 1974 for commodities and equity country indices. Frazzini and Pedersen (2014, JFE) document 'BAB' for U.S. stocks since 1926, government bond

for these six factors. We construct a broad and deep historical global multi-asset database compiled from various historical data sources which covers equities, bonds, commodities and currencies as of 1800. This allows us to reexamine the six global return factors over the post-1965 period in a unified testing framework in which we limit the degrees of freedom.[4] In addition, it allows us to examine the pre-sample evidence for the extended period 1800 until 1965. Further, we verify the robustness of our results to common choices made in the construction method. Because the six return factors are difficult to reconcile with classical risk explanations they are often referred to as asset pricing anomalies. Still, the past 50 years were quite exceptional without major wars and only a few large recessions which severely limits the number of 'bad states'. Our extended 200+ years sample includes several wars, and many bear markets (43 years) and recessions (74 years). It therefore allows us to better test if the factors which survive the replication and pre-sample tests, can be explained with market risk, downside risk, or macro risk.

Our findings are as follows. We find consistent and ubiquitous evidence for the large majority of factors. Figure 1 summarizes our main findings by depicting the historical Sharpe ratio over the 'recent sample' (1965-2016) versus the 'pre-sample' (1800-1965). The dashes lines indicate the 1.96 t-value cut-offs traditionally used, or higher at 3.00 as advocated by Harvey, Liu, and Zhu (2016). We replicate five multi-asset factors in the recent sample and find they are also significant in the pre-sample. In terms of statistical significance, most t-values tend to be well above 3 in both samples. In terms of economic significance, most Sharpe ratios are economically substantial, especially at the multi-

---

indices since 1952, equity indices since 1965, corporate bond indices since 1973, currencies since 1975, international stocks since 1984, and commodities since 1989. In this paper we focus on index-level data and uniform asset classes and do not include individual U.S. stocks, international stocks or other company-level securities.

[4] The 1965 cut-off is chosen to align with the earliest data used in the original studies to the multi-asset anomalies we study. Note that our examination over this period differs from the original studies in several aspects. First, our sample period is typically longer (including the most recent data and always beginning in 1965). Second, we have chosen a uniform setup of constructing the factor strategies across the anomalies, and thereby limited the degrees of freedom. Third, our choice of markets is uniform across the anomalies, and typically differs slightly from the original papers. Consequently, we see this period as a robustness check on the original papers.

asset level. Further, in contrast to most out-of-sample studies (see for example Linnainmaa and Roberts, 2018) the economic effects are generally of similar size in the recent sample and out-of-sample periods. As exception, the betting-against-beta effect is present in equity markets in both samples, but evidence is less robust for the cross-section of country bonds[5], commodity and currency markets. In our setup, return seasonality in government bonds and currencies are novel to the literature and verify the existence of a multi-asset global 'seasonality' factor.

INSERT FIGURE 1 HERE

Next, we turn to adjustments on p-values proposed against p-hacking. First, most t-statistics well exceed the 3 threshold advocated by Harvey, Liu, and Zhu (2016).[6] Second, applying a Bayesian perspective advocated by Harvey (2017), using a symmetric and descending minimum Bayes factor with "Perhaps" prior odds ratio 4:1, we find that these adjusted Bayesian, or "Bayesianized", p-values are very close to zero for our extended sample. Third, one downside of the Bayesian approach is the subjectivity required in the formulation of prior odds, subjectivity that typically has substantial impact on the Bayesian p-value. Our next contribution is the formulation of 'break-even' prior odds; those prior odds at which the Bayesian p-value would equal the confidence level chosen. These break-even prior odds remove the need to specify the prior-odds (but do require the confidence level), and allow for an interpretation in terms of prior odds that would be

[5] To stress, Frazzini and Pedersen (2014) find substantially stronger results for BAB applied to bonds across the curve within a single country. Since bond data of various maturities is generally lacking a deep history we leave the BAB factor applied to yield curves within a bond market outside the scope of this study.

[6] Harvey, Liu, and Zhu (2016) conclude their analysis on the multiple testing bias in hypothesis testing in finance by advocating a t-statistic hurdle of 3 for factors to be deemed statistically interesting. Related, Benjamin et al. (2018) 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. They recommend to employ statistical techniques to deal with multiple testing and Bayesian inference. However, this higher t-value is an easy step to implement, and the other techniques may be more technically demanding or there is no general agreement about its use. Harvey, Liu, and Zhu (2016)'s suggestion of increasing the t-statistic to 3 is thus very similar to the proposal of Benjamin et al. (2018).

required to just accept the alternative. The prior odds generally need to be above 99,999 to have less than 5% probability that the null hypothesis is true. Such extremely large odds ratios implies that one needs to be extremely skeptical to disregard the empirical evidence provided in this study.

We advocate that a multi-dimensional approach that also considers subperiods and robustness to degrees of freedom (besides the above statistical and economic criteria) protects against p-hacking and is necessary to reach robust conclusions about financial research.[7] Therefore, we also examine subperiods and robustness checks. We find that the return factors have been positive almost every single decade since 1800, and every single 20 year or 50 year subperiod. Moreover, the factor returns are robust to various degrees of freedom in testing, like the period of rebalancing, accounting for lagged implementation, the exact portfolio construction method, and the trimming of extreme positive returns.

Having established strong, convincing evidence for the global return factors, we next examine the overlap between the return factors and whether they are driven by a common component. We find that most return factors do not share a common component and generally diversify well amongst each other. The exception is time-series momentum encompassing cross-sectional momentum, akin to the findings of Moskowitz, Ooi and Pedersen (2012) over a much shorter sample period. The economic significance of the other return factors is large, with spanning alphas between 7 and 10 percent with double-digit t-values. Further, the time-series momentum, value, carry and seasonality return factors matter substantially in asset allocation, with improvements in the Sharpe ratio of the ex-post mean-variance portfolio going up from 0.34 for a traditional equity-bond portfolio to 0.86-1.53 with inclusion of these return factors. In summary, time-series momentum,

[7] Chordia, Goyal, and Saretto (2017) also make the case for considering economic criteria to protect against p-hacking.

value, carry and return seasonality represent unique, robust and 'true' global factors that drive returns across many markets.

Finally, we explore whether these return factors can be reconciled with downside risk or macroeconomic risks, as theory suggests that expected returns can vary due to downside risk (Lettau, Maggiori, and Weber, 2014) or macroeconomic risk (Fama and French, 1989, Ferson and Harvey, 1991, Campbell and Cochrane, 1999).[8] A concern with many existing studies into these explanations is that they cover up to the past 50 years of data. This period is characterized by peace, growing global prosperity, and only includes a few recessions and periods of social unrest. Consequently, we may wonder how these explanations hold up in our sample, stretching 217 years with many recessions, wars, and bear markets, and we study the sensitivity of the return factors to financial market and macro-economic regimes. We find no positive evidence that downside risk can explain the returns we observe on the multi-asset factors. In the case of carry we find sensitivity to downside risk, but this is at best only a partial explanation, in line with the findings of Koijen, Moskowitz, Pedersen, and Vrugt (2017). Further, we generally cannot reject the null hypothesis that the average return on the multi-asset return factors are the same during recessionary and expansionary periods, crisis and non-crisis periods, bull and bear markets, and volatile and tranquil markets.[9]

This study is not the first to utilize a deep historical samples to study market premiums. For equity and bond premiums, Siegel (1992) gives evidence stretching back to 1800, Goetzmann (1993) to 1695, and Golez and Koudijs (2017) go even further back to 1629. Hurst, Ooi, and Pederson (2017) find a persistent time-series momentum premium going back to 1880, and some other studies provide pre-20th century evidence for

[8] A recent paper by Hodges, Hogan, Peterson, and Ang (2017) argue that equity factor premiums depend on business cycle indicators over the period 1988-2016. Asness, Moskowitz and Pedersen (2013) show that global macroeconomic variables are generally not related to value and momentum returns.
[9] We do acknowledge that the deep historical macro data could contain more noise, which could hurt the testing of macroeconomic explanations. Therefore, we mainly focus on examining multi-asset return factors over a deep historical sample and actively addressing possible p-hacking issues.

individual factor premiums (typically for a particular asset class) such as momentum, and currency carry and value.[10] In general, these historical studies find positive results also for extended sample periods. However, existing studies to deep historical evidence on return factors focusses either on a single factor (e.g. momentum), a single market (often the U.S.), or a single asset class (typically stocks). This means that the results are somewhat fragmented and do not use one common methodology and therefore could be biased to positive results. In this study, we focus on a broad range of factors across a broad range of markets using a simple, robust and uniform testing methodology and several explicit controls and checks against p-hacking.

The remainder of this paper is structured as follows. Section II highlights the investment universes, definitions, portfolio construction method and data. Section III analyses the return anomalies over 217 years and several asset classes, analysis the return anomalies in a Bayesian framework proposed by Harvey (2017), and examines subperiods. Section IV discusses robustness for testing choices. Section V examines common variation amongst the return factors and potential explanations related to downside, macro-economic and financial market risks. Section VI concludes. The Appendix contains more details about the portfolio construction method and the data.

II. Construction and Data

This section describes the investment universes we employ, the definitions of the factors we examine in this study, the portfolio construction method we employ to examine the return factor, and our dataset.

Investment universes. We consider four investment universes, one per major asset class; equity index markets, 10-year government bond markets, developed market

[10] For example, Goetzmann and Huang (2015) find that equity momentum also works in imperial Russia in the period 1865-1914, while carry and value premiums are present in currency markets since 1900 (Doskov and Swinkels, 2015 and Taylor, 2002).

currencies, and commodities. We refer to the Appendix for more details on the markets considered. Within each investment universe, we require at least two markets to be present.

The factors we examine are defines as follows.

Time-series momentum ("Trend"). We use the 12-month-minus-1-month excess return as trend measure for each investment universe. We take a long position if the sign of the past return is positive and short position if it is negative. We skip the last month on momentum investing as this safeguards against potential liquidity issues that might be especially prominent in the earlier parts of our sample. Note that Moskowitz, Ooi, and Pedersen (2012) do not incorporate a one-month skip between the formation and investment periods for their time-series momentum strategies, which they test on liquid futures contracts after 1985. In the robustness section, we also consider a two-month gap to ensure that potential short-term stale prices are not driving our results.

Cross-sectional momentum ("Momentum"). We use the 12-month-minus-1-month excess return as momentum measure for equity indices, bonds, currencies and commodities, following Asness, Moskowitz, and Pedersen (2013). As for time-series trend, we skip the last month to safeguard against liquidity issues, particularly in the earlier parts of our sample. In the robustness section, we also consider a one month delay in implementing the signal..

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 available historically. For government bonds, we use the real yield, which is defined as the 10-year bond yield over 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).[11] Absolute PPP follows Rogoff (1996) and Taylor (2002). Relative PPP is a 5-year reversal of the spot rate corrected for inflation differences defined 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"). For equity we use the excess implied dividend yield priced into the futures versus spot contract as in Koijen, Moskowitz, Pedersen, and Vrugt (2017). 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 52%. For government bonds, we take the slope of the yield curve defined as the 10-year yield minus the short (3-months) yield and omit the quantitatively smaller roll-down on the interest rate curve. For currencies, we use the short-term yield differential (inferred from forwards and before their availability from short-term money market rates), and for commodities the slope of the futures curve. These carry definitions are as in Koijen, Moskowitz, Pedersen, and Vrugt (2017).

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 only those

[11] 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.

equity markets which had the best relative performance in Januaries during the past 20-years and shorts those with lowest relative performance. Keloharju, Linnainmaa, and Nyberg (2016) document a significant return seasonality factor using data since 1963 for individual stocks and data since 1971 for commodities and equity country indices. We also apply their definition to bond and currencies.

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-months of data) relative to the global asset class portfolio return (we refer to the Appendix 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, such that the excess return contains as little of a market effect as possible.

Portfolio construction. We construct factor investment portfolios at the end of every month in spirit of their original papers. However, since details differ across the papers and degrees of freedom are one of the drivers of p-hacking, we choose to apply a uniform method across the factors we study (which might differ slightly from definitions in original paper). Admittedly, this uniform method also has degrees of freedom, and as such we verify the robustness of our results for the most important portfolio parameters in the Robustness section. Further, we have to make some different choices for the sizing of positions since our sample only covers monthly data.

The exact details are summarized in the Appendix. In short, for the time-series momentum factor, which is directional in nature, we go long (short) markets in each asset class when the time-series momentum 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 the 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, Moskowitz, Pedersen, and Vrugt (2017). (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.) Next, from these positions we 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. All portfolios are monthly rebalanced.

Dataset construction. We have compiled our data from several sources in order to have a reliable and historically extensive dataset. Our sample covers 217 years of data, starting on December 31st, 1799 and running till December 31st, 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, the Dimson, Staunton and Marsh database, the Jordà-Schularick-Taylor Macrohistory Database[12], and/or Jeremy Siegel's website.[13] Our dataset construction for each asset and measure is described in more detail below. Table I shows the start and end dates of each factor series. Table A.1 in the Appendix summarizes the markets included in our sample, the start dates of their excess return series (and thereby the momentum, seasonality and BAB measures at most 12-months later), as well as the start dates of the value and carry factor measures per market. All returns we consider are in compounded terms in excess of local financing rates and expressed in U.S. dollars.

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

INSERT TABLE I HERE

INSERT FIGURE 2 HERE

Figure 2 shows that our sample consists of 13 assets across asset classes in 1800: 2 equity markets (U.S., U.K.), 3 bond markets (U.S., U.K., France), 5 commodities (Wheat, Cotton, Cocoa, Copper, Silver), and 3 currencies (GBP/USD, FRF/USD, and USD/USD). The sample increases in 1822 to 18, and then gradually further over time, with several markets entering between 1860 and 1870 to increase the sample size to 36. At the start of World War I in 1914 there are 50 assets. In 1974, the number of assets has increased to 66, and in 1999 the sample decreases from its maximum of 68 to 63, because of the introduction of the euro currency.

Remarks about our dataset. 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 survivorship bias as described in Brown, Goetzmann, and Ross (1995). For example, Argentina is believed to have been an economy of comparable size to large European markets during the beginning of the 1900s, while our data sources do not cover equity or bond markets in Argentina around the 1900s and we do not include this market in our sample. That said, Dimson, Marsh, and Staunton (2008) claim that several Latin American countries combined (including Argentina) were less than 1.5% of the equity market capitalization during the 1900s. Moreover, Goetzmann and Kim (2017) analyze the equity indices in Global Financial Data on survivorship bias in a setting that is arguable more exposed to survivorship bias (i.e.

crashes and rebounds) and show that this is at best a minor concern. Finally, our results are robust across ten year subperiods (including the recent subperiods), which seems hard to reconcile with survivorship bias.

Even though we (and the data vendors) have paid close attention to data quality, it is hard to assess to which extent these worries apply, especially for data further back in history. In general, we expect that poor data quality would be random and therefore increase noise in the data, biasing our results towards the null hypothesis that factor premiums do not exist. In addition to this qualitative argument, we investigate robustness towards trimming extremely positive returns, applying an additional monthly lag between signal measurement and investment strategy implementation, and report results for rolling 10-year periods as well as different sub-samples – splitting the sample in two centuries, four different 50-year periods and the most recent 21st century sample in which data is of best quality. These checks should mitigate concerns about the quality of our data.

III. Return factors since 1800

This section analyses the return factors over 217 years and several asset classes, analysis the return factors in a Bayesian framework, and examines subperiods.

A. Main results

We start by plotting the time-series log-returns of the six multi-asset factors in Figure 3. For comparison the “Trend”, “Momentum”, “Value”, “Carry”, “Seasonality” and “BAB” series are all scaled to a 10% ex post volatility for ease of comparison. We see that, except for BAB, each factor has a positive return. Clearly, this full-sample positive return cannot be attributed to short intervals with positive performance. The relatively smooth cumulative return series indicate that factors have consistently performed well over this

217-year period. For example, it is difficult if not impossible to visually detect performance differences between the pre- and post-1965 periods (which roughly translate to the in-sample and out-of-sample periods). This in contrast to most studies testing anomalies out-of-sample, who generally find a weakening of results out-of-sample. An exception may be the two decades preceding World War I, in which factor returns seem exceptionally high.

INSERT FIGURE 3 HERE

INSERT TABLE II HERE

Next, we are quantifying the significance of the returns presented in Figure 3. In Table II Panel A, we display the full-sample Sharpe ratios of each of the six factors for each of the four asset classes, and their multi-asset combination. The Sharpe ratio is a natural choice, as it relates the average excess returns to the volatility risk associated with earning the return. This is especially important since we are examining excess returns, that can, at least in theory, be leveraged up or down to create any average return. We analyze downside risk, which is not captured by volatility risk, in Section V.B. Table II Panel B contains the Jensen's alphas relative to global equity, government bond, currency, and commodity markets. The Sharpe ratio corrects for volatility risk, but does not take into account the correlation with other asset classes. Jensen's alpha does exactly this. If the factor returns can be explained by correlation to traditional global equity, bond or other markets, they are redundant, even if they have an economically significant Sharpe ratio.

Table II Panel A shows that Trend, Carry, and Seasonality are the three strongest factors. Their multi-asset Sharpe ratios are above one, and t-statistics above 15. These three factors have Sharpe ratios above 0.40 for each asset class individually, and t-statistics well above 5. The global Value factor has a statistically significant Sharpe ratio for each individual asset class between 0.29 and 0.46, and due to diversification a multi-asset Sharpe ratio of 0.71, with a t-statistic of 10.42. Multi-asset momentum has a similar Sharpe ratio, even though the Commodity asset class does not seem to exhibit momentum. The BAB factor is economically and statistically significantly positive for equities, but is statistically insignificant for the cross-section of country bonds and currencies, and negative and statistically significant for commodities.

That said, our results for the BAB factor are comparable for all asset classes except currencies to Frazzini and Pedersen (2014) over the 'recent' post-1965 sample period (although the exact sample construction and data periods differ to a certain degree). Figure 1 reveals that, over the 1965-2017 period, BAB has an annualized Sharpe ratio of 0.42 for equities, 0.17 for country bonds, 0.12 for commodities and -0.11 for currencies. This compares to 0.51, 0.14, 0.11, and 0.22, respectively, documented in Frazzini and Pedersen (2014). Only the results for equities are significantly positive for the earlier (1800-1965) sample period. This leaves us to conclude that BAB is present in equities, but is weak in country bonds, commodities and currencies. To stress, Frazzini and Pedersen (2014) find substantially stronger results for BAB applied to bonds across the curve within a single country; in a separate U.S. bond sample covering many maturities they find a Sharpe ratio of 0.81. Since bond data of various maturities is generally lacking a deep history we leave the BAB factor inside country bond markets outside the scope of this study.

The Jensen's alphas in Table II Panel B indicate that none of the factor strategies are strongly correlated to global asset class market risk. The alphas are typically about 10 times larger than the Sharpe ratio, which one would expect if there is no correlation with global equity and bond markets, as the factor returns are scaled to a volatility of 10% per annum. The only exception is Commodity Momentum, which had an insignificant Sharpe ratio of 0.07, and is now marginally statistically significant when evaluated against the traditional t-statistic of 1.96. In the following subsections, we address potential concerns on data-mining and multiple testing in more detail.

B. Statistical results in traditional and "Bayesianized" p-values

We report 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 is not answered by frequentist p-values, but can be answered by a Bayesian approach. The challenge with Bayesian methods is that they require a clear sense of the alternative hypothesis, which we often do not have in finance applications. As an alternative to a full-blown Bayesian analysis, Harvey (2017) proposes to use the Minimum Bayes Factor (MBF) as developed by Edwards, Lindman, and Savage (1963). The Bayes Factor connects the prior odds, the odds before having seen the data, with the posterior odds, the odds after having seen the data. The MBF then 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 hypothesis as potential candidates. Bayarri and Berger (1998) come up with the SD-MBF, which follows if we assume that the prior probabilities of alternatives should be symmetric and descending around the null hypothesis. The use of an SD-MBF, which we do throughout, is more conservative than the MBF, that does not have the structure that is common to impose in finance applications. 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 needed, or even an explicit alternative is needed. 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). As such, we also consider an alternative approach; we infer the maximum (i.e. most unfavorable) level of prior odds such that the null hypothesis has a probability of 5% of being true.

INSERT TABLE III HERE

Table III further deals with multiple hypothesis testing as advocated by Harvey (2017). For each asset class factor combination, we report the frequentist p-values, but also the Bayesian p-values with a prior odds ratio of 4:1, prior odds classified by Harvey (2017) as "Perhaps". For all asset class factor combinations the traditional and Bayesian p-values do not cross the 5% threshold.

The difficulty in the Bayesian approach is the assumption on the level of prior odds. This choice is quite subjective, while it typically has 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. In Table III, we display the break-even odds ratio that leads to a posterior probability of 5% for the null hypothesis. For most of our asset class factor combinations, the prior odds ratio is above 99,999-to-1, indicating that one needs to be extremely skeptical in order not to reject the null hypotheses that global factor returns are zero.

C. Subperiod performances

In addition to the full-sample Sharpe ratios and Jensen's alphas, we display here results over subperiods. To reduce the degrees of freedom to pick convenient subperiods, we instead start with displaying the frequency that 10-year rolling window Sharpe ratios are positive in Table IV Panel A. This table shows that for asset class factor combinations with significant full sample Sharpe ratios and Jensen's alpha, the performance is consistent over time with frequencies typically above 80%. With the exception of BAB, all multi-asset factors have at least 94% of the rolling 10-year periods between 1800 and 2017 a positive Sharpe ratio. This consistency of performance over time further strengthens our empirical evidence for the existence of five of the six factor premiums that we examine.

INSERT TABLE IV HERE

Table IV Panel B contains the Sharpe ratios for each half-century and century of our sample. We find that Sharpe ratios are positive for each of these subsamples, and in most cases the t-statistics are above 3.0. This highlights that the Sharpe ratio of multi-asset factors have been strong consistently over our sample period. Further, as also shown in Figure 1, performances of the global return factors are generally comparable over the recent 50 years (or post 1965 period) as for earlier periods. The strong performance across subsamples and results being of similar magnitude in recent periods, should also help to alleviate concerns about data quality.

Note that the limited number of assets per asset class in the first 50 years of our sample period makes it more difficult to detect the existence of 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 long and short portfolios. Hence, even though the lower number of assets makes it less likely that we can reject the null hypothesis of no factor premiums, we still find t-values close to three for the first 50 years of our sample.

IV. Robustness

We have seen a strong statistical and economic presence of return factors over 217 years and almost every 10-year subperiod. Another important aspects to safeguard against p-hacking is verifying the robustness of results to factor definitions and methodological choices with regards to the factor investment strategies, which we examine in this section.

INSERT TABLE V HERE

Table V considers five methodological variations relative to our 'baseline strategies' examined so far. 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 quartile within each asset class (the portfolio construction choice employed by Keloharju, Linnainmaa, and Nyberg, 2016). This simplified portfolio construction only slightly reduces the Sharpe ratios; Carry is affect most with a decrease of just 0.11, while it remains statistically significant. Note that the Trend factor is not affected by definition, as this is a time-series and not a cross-sectional signal. Second, we do not scale the positions using past volatility or beta (in case of BAB), but rather take simple equal notional positions in each asset. This means that more weight is given to the most risky assets within an asset class and the factor strategies become less diversified. As expected, this has a meaningful impact on the Sharpe ratios. For example, for Trend the Sharpe ratio decreases from 1.29 to 0.88, and the corresponding t-statistic from 18.99 to 12.93.

Although the impact is certainly noticeable, the resulting t-values are still highly significant and well above 3.00. Third, we consider robustness to lagging signals by one month before investing. This lowers the Sharpe ratio for each factor except value, but the economic and statistical magnitude of the reduction is limited for each factor except Seasonality. By construction, this factor is linked to a monthly seasonal effect, and delaying the signal by a month indeed reduces the average return to indistinguishable from zero. Fourth, a similar reduction is obtained when rebalancing the portfolios with a quarterly instead of a monthly frequency. Again, this does not hold equally well for Seasonality, which captures only the first-month return of the quarterly rebalancing period. The result is that this factor is still positive and statistically significant, but its economic magnitude is reduced by about a factor three, as could be expected from extending the holding period to three months. Fifth, we test the robustness with respect to outliers by trimming extreme returns, also because extreme returns might be typically hard to harvest. Note that this will also provide more robustness to potential data errors. More specifically, we trim all *strategy* returns before any position sizing in a single market at a maximum of 50%. The Sharpe ratios for each of the factors generally drop by this procedure, but again the resulting t-values are still highly significant and well above 3.00. Unreported results show that the robustness analyses are virtually the same when evaluated by Jensen's alpha instead of Sharpe ratio.

An important remark is in order. The main purpose of this paper is to provide more robust and rigorous long-term evidence on the historical presence of multi-asset factor premiums, utilizing their most simple or basic definitions as put forward in influential papers analyzing recent samples. As such, this study does not examine smarter and possibly better definitions, nor aspects linked to (limits to) arbitrage and tradability (such as transaction costs, turnover, legal controls, capital mobility, etc.). Although there are indications that some of the factor premiums were exploited historically, and their trading was feasible at limited transaction costs, we leave these aspects for future research. Instead, we have documented a consistent, robust existence of factor premiums over 217 years of data, which includes more than 160 years of pre-sample data.[14]

V. Common variation, downside and macro-economic risk explanations

So far we have documented economically and statistically large global factor premiums. This raises the question what is driving these global factor premiums? Even though we do not aim to answer the underlying cause of the global return factors and leave a comprehensive study to the driving forces for further study, we next provide insights into potential explanations related to macro-economic and financial market risks. The advantage of our long-term data series is that it is easier to identify explanations that relate to infrequent observations such as downside or macro-economic risks. We start by examining the common variation across return anomalies and asset classes to study the potential for a common risk-based explanation, whether global return factors can be explained by each other, and how much optimal Sharpe ratio portfolios would allocate to these factors. Subsequently, we analyze explanations based on downside risk, volatility risk and macroeconomic risk.

A. Common variation

We start by showing the average pairwise monthly return correlations for each factor across the four asset classes, and for the six factor series per asset class or their equally weighted multi-asset combination. These can be found in Panel A of Table VI. The average correlation of each factor across asset classes is close to zero, ranging from -0.01 (Value) to 0.06 (Trend). A similar picture emerges for the average correlation for factors

[14] The strong, robust and persistent presence of factor premiums also helps to alleviate concerns about survivorship bias, a bias that is absent in recent data and that is unlikely to have had a material impact during most decades.

within an asset class, with values ranging from 0.02 (Commodities) to 0.19 (Currencies). Further, most individual correlation coefficients between the multi-asset factor series and between each factor-asset class series are also close to zero, as shown in Table A.2. The main exceptions are Trend and Momentum generally correlating positively, and Momentum and Value correlating negatively.

To show the added value of each of the multi-asset return factors, we regress its' time-series on the other multi-asset factor return series. The intercept of this regression is a Jensen's (1968) alpha, and a test on its significance can be interpreted as an expansion of the mean-variance efficient frontier with respect to the other factors; see De Roon and Nijman (2001). All series are scaled to an ex-post volatility of 10% for interpretation purposes. Note that as such, exposures represent partial correlations.

Table VI, Panel B shows that the trend and momentum factor are highly correlated with statistically and economically significant betas to one another just above 0.5. The trend factor subsumes the momentum factor, in the sense that the annual alpha of 8.38 of trend is economically and statistically significant, but the annual alpha of momentum is below zero with -1.79. These results on trend and momentum are remarkably similar to the empirical results presented by Moskowitz, Ooi, and Pedersen (2012) over the period 1985 to 2009. They find a beta of 0.66 when regressing trend on momentum, and an alpha of trend that remains economically and statistically significant, while an alpha of momentum that is economically small and negative but no longer statistically significant. This leaves us to conclude that trend encompasses momentum in multi-asset space.

INSERT TABLE VI HERE

The other factors have only economically small exposures to one another, even though these are often statistically significant due to our long sample period. The explanatory power is low, with R^2-s around or below 10%. The alphas of each of these factors remain economically and statistically significantly positive, indicating long positions in optimal asset allocations.

Finally, we consider an optimal asset allocation problem in which we maximize the full-sample Sharpe ratio, restricting the portfolio weights to be non-negative and sum to unity. The first row of panel C of Table VI show the mean-variance optimal portfolio over global equity and bonds. Since the Sharpe ratio of government bonds is close to zero, the optimal Sharpe ratio portfolio is fully invested in the global equity market, which yields an annual Sharpe ratio of 0.34. The next rows add the multi-asset return factors one by one. We see that portfolio allocations range from 67% to 85%, with the exception of BAB with optimal allocation of zero. For the five factors with positive allocation, the Sharpe ratios increase to 0.84 or higher. The last row adds the six multi-asset return factors together. We see that the result is an almost equally-weighted allocation to Trend, Value, Carry, and Seasonal. As we have seen in the spanning tests in Panel B, Momentum is crowded out by Trend, and BAB's standalone multi-asset performance is too weak to be included in the optimal portfolio. Further, we have examined the impact of replacing the multi-asset BAB series with the equities BAB series, as BAB is confirmed for equities. The results (unreported) reveal an optimal portfolio weight of 56% for equities BAB when added to a portfolio of global equities and bonds, and a 1% weight for equities BAB if we also include the other factors.

It is evident that investors who could have added these multi-asset factor premiums, could have benefited the past 217 years. That said, we do not aim to make explicit claims on the tradability of the factor strategies, as that is outside the scope of this paper. Instead, our focus is on testing the robustness of factor premiums in a deep, long sample covering various economics and market regimes. However, various evidence exists that asset markets in the 17th and 18th century are in many important dimensions similar to today's, like the existence of active and liquid derivatives markets with the ability to go short, liquidity, a well-developed system of trading on margin, and a centralized exchange with clearing mechanism (see Harrison 1998, Koudijs, 2016, Lean, 1985). In fact, a large part of trading activity was related to futures and options. Active markets in commodity futures (and also options) existed since the 16th century in Amsterdam (Pentram, 2014), and active markets in equity futures (and option) emerged in the 18th century, with active call and put markets dating back to the late 1600s in Amsterdam and London (Harrison, 2004). Further, the 18th and 20th century stock markets shared many important behavioral and institutional characteristics (Harrison, 1998, Koudijs, 2016). For example, futures settled in (sometimes fractions of) difference (i.e. money), there was an active shorting market in which shorting via futures played a major role, trades could be done on margin, and clients could often arrange to borrow from the broker if faced with a price move against their position. Further clearinghouse mechanism were present, and calculations by Koudijs (2016) reveal very limited transaction costs on equities and futures, sometimes close to zero and not higher than on the NYSE around the turn of the millennium. Lean (1985) provides evidence that capital markets of the eighteenth century were closely integrated even by early twentieth-century standards, and integrated even during financial crises that affected one market more severely than the other, as evident from arbitrage on prices of identical assets (equities and bonds) that are traded simultaneously in multiple markets. Indications exists that trading costs prior to the 20th century were not as high as often assumed (Koudijs, 2016), and cross-border capital flows were significant (Neal, 1991). Jobst (2009) documents that the Austro-Hungarian Central Bank was extensively using currency forward markets during the late 19th and early 20th century.

B. Downside risk

The large and growing literature on the return anomalies considers whether they compensate investors for crash risk or downside risk. In this subsection we consider downside risk controls via alphas relative to a downside risk model. The CAPM either assumes normal returns or mean-variance preferences of investors. These joint assumptions are quite strict and can be challenged. The Downside Risk (DR) CAPM was developed by Bawa and Lindenberg (1977) and is more flexible to account for non-normal returns (Fama, 1965) and non-linear marginal utility of investors (Tsiang, 1972). In this downside risk model, the mean-variance beta is replaced by downside beta in the following equation:

$$Ri\text{-}Rf = alpha + downside\_beta*(Rm\text{-}Rf)$$

When the DR CAPM holds, assets with higher downside betas should have higher expected returns similar to the CAPM. Downside beta can be specified against different thresholds such as zero, or lower thresholds deeper into the left tail of the return distribution. Recently, Lettau, Maggiori and Weber (2014) use the DR CAPM model and find that the carry strategy has higher downside beta compared to the regular beta. A common challenge with modelling downside risk, is a general reduction in the number of observations. Especially market crashes do not happen very often. Therefore this extended multi-century sample is well suited to verify the hypothesis if downside risk can explain factor premiums.

INSERT TABLE VII HERE

Table VII shows the betas and downside betas of the 6 factor premiums. We find that the carry strategy is indeed more sensitive to falling equity markets and even more so when markets drop significantly by 1 standard deviation or more. The regular beta of carry is 0.09, while the downside 'crash' beta of carry is 0.23. This finding is in line with the results of Lettau, Maggiori and Weber (2014). For seasonal we observe a small increase in downside beta, while for all other factors we do not observe a pattern of higher downside risk. The CAPM alpha of carry is 12.28 percent and statistically significant. This alpha goes down to 11.75 percent for the DR CAPM, a decrease of 0.53 percent in alpha. Still, the alpha remains statistically positive and only a small portion of the carry premium can be explained. Thus downside risk can only partially explain a part of the carry premium, while it cannot explain other premiums. The explanatory power of the DR model is limited which makes economic sense. If the DR model holds, the price for a unit of downside beta equals the equity market premium which is 3.05 percent in this sample. To rationalize factor premiums of 10 percent, they should have downside betas of at least 3 to bring these large alphas down to zero, beta differences well beyond what we observe in the data. As robustness, we estimated downside betas over a rolling sub-sample analysis of 10 years, finding 99% of all rolling downside betas distributed between -0.8 and +0.7. Therefore an explanation based on downside risk seems at best a partial explanation.

C. Macroeconomic risk and bad-states

This subsection describes the performance of factor premiums over economic and market regimes. 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.[15] A major strength of our sample (due to over 200 years of data) is the presence of many observations within and across various economic and market regimes, like recessions, crises episodes, bull and

[15] 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 a main candidate 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 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 premia.

bear markets, and turbulent and tranquil periods. Consequently, we can study the long-run behavior and robustness of factor premiums across these regimes.

To this end, we construct indicators that capture different market states 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 6 months of a calendar year are in a recession, and as expansionary otherwise.
- Global crises versus non-crises 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.
- Equity bull versus bear market periods, where we mark calendar years as a bull (bear) market when calendar year global equity return series were positive (negative).
- High volatile versus tranquil periods, where we mark calendar years as 'Turbulent' when the calendar year sum of squared monthly global equity returns is in its full sample top quartile, and as 'Tranquil' otherwise.

INSERT TABLE VIII HERE

The results of the regime-dependent return analyses are summarized in Table VIII. The 74 recession and 143 expansion years have statistically indistinguishable returns for Trend, Momentum, and Carry, with positive return in all states. Value and Seasonality have 3.91% (t-value = 2.73) and 4.99% (t-value = 3.49) higher returns in recessions. Nevertheless, their returns during expansion are still large economically and statistically highly significant. For BAB, we find that the sign of return switches across regimes. The average return is negative (-4.17%, t-value: -3.60) during recessionary years and positive (2.34%, t-value 2.80) during expansionary periods, leading to a statistically significant difference of 6.50% per annum.

When we compare crisis periods with non-crisis periods, we find that only for Momentum and Carry the returns are statistically different from zero. Both factors perform worse during crisis periods with 3.75% and 6.46% per annum. Separating bull and bear markets leads to a statistically significant return difference over these two states for the Trend factor, with a 4.83% higher return during bear markets. Further, Carry and Seasonality have lower returns during turbulent periods than during tranquil periods. But again returns on these global factors are economically large and highly significant across all states. Finally, BAB experiences positive returns during turbulent periods (5.05%, t-value = 3.71) and negative returns during tranquil periods (-1.51%, t-value = -1.93), with a return difference of 6.56% (t = 4.18).

In summary, except for multi-asset BAB, each of the global return factors has an economically large and statistically highly significant return across macroeconomic 'bad' (recessions, crisis) and 'good' (expansions, non-crisis) states, and financial market 'bad' (bear or turbulent markets) and 'good' (bull or tranquil markets) states. Furthermore, the returns differences across these regimes are indistinguishable from zero for most global return factors.

VI. Conclusion

We provide comprehensive evidence on six global asset pricing factors (time-series momentum, cross-sectional momentum, value, carry, return seasonality and betting-against-beta) across four major asset classes (equities, bonds, commodities and currencies) utilizing a deep historical international database going back to 1800. We find that most factors offer economically high Sharpe ratios and alpha's that are generally very persistent over time, and are generally of similar size over the recent sample and more distant out-of-sample periods. As main exception we find that the betting-against-beta factor works in equity markets, but is weak in the cross-section of country bonds, commodity and currency markets. From a statistical perspective, we find that t-statistics generally are well above 3. Further, utilizing new methods to account for p-hacking we find that "Bayesianized" p-values (that alleviate p-hacking concerns) close to zero, and 'break-even' prior odds to generally be above 99,999. These factors are not spanned by each other, except for time-series momentum crowding out cross-sectional momentum.

The resulting multi-asset return factors are at best partially driven by downside risk explanations. Further, they are consistently present across various macroeconomic and financial market states. Consequently, our results seem hard to reconcile with arguments based on risk, although we are cautious on such an interpretation because risk exposures and especially risk premiums are not directly observable. Instead, we interpret our findings as providing no positive evidence on a relationship between return factors and risk. We leave further study to explanations and economic drivers of the documented return factors to future research.

Our results have strong implications for research in asset pricing. The literature on asset pricing theory and return predictability have often evolved separately by asset class or factor, where most studies focus on a single asset class, market or factor at a time, ignoring the universality of return factors across asset classes and the existence of at least four multi-asset return factors. Our findings reveal that a theory that aims to model variation in expected returns should consider multiple asset classes and factors simultaneously. Further, the documented multi-asset return factors are important controls for empirical studies to new asset pricing factors. All in all, our study shows a strong, robust and persistent presence of economically important global return factors within the major asset classes.

References

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

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

Baltussen, G., Z. Da, & S. Van Bekkum. (2017). Indexing and Stock Market Serial Dependence around the World. *Journal of Financial Economics*, forthcoming.

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. J., 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.J., & Berger, J.O. (1998). Quantifying surprise in the data and model verification, in J.M. Bernardo, J.O. Berger, A.P. David, and A.F.M. Smith, eds: *Bayesian Statistics* 6, Oxford University Press: Oxford, U.K.

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

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.

Blitz, D. C., & van Vliet, P. (2007). The Volatility Effect. *The 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 Studies in the Quantity Theory of Money, ed. Milton Friedman, 25-117. Chicago: University of Chicago Press.

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

Campbell, J. Y., & Cochrane, J. H. (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior, *Journal of Political Economy* 107, 205-251.

Chordia, T., Goyal, A. & Saretto, A. (2017). p-Hacking: Evidence from Two Million Trading Strategies. SSRN working paper 3017677.

De Roon, F. A., & Nijman, T. E. (2001). Testing for mean-variance spanning: A survey. *Journal of Empirical Finance*, 8, 111-155.

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

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

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

Fama, E. F. (1965). The behavior of stock-market prices. *The Journal of Business*, *38*(1), 34-105.

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

Fama, E. F., & French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics* 25, 23-49.

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

Ferson, W. E. & Harvey, C. R., (1991). The Variation of Economic Risk Premiums. *Journal of Political Economy* 99, 385-415.

Frehen, R. G., Goetzmann, W. N., & Rouwenhorst, K. G. (2013). New evidence on the first financial bubble. *Journal of Financial Economics*, *108*(3), 585-607.

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

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

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

Goetzmann, W.N., & Huang, S. (2018). Momentum in Imperial Russia, *Journal of Financial Economics*, forthcoming

Goetzmann, W.N., & D. Kim. (2017). Negative Bubbles: What Happens After a Crash. *European Financial Management*, forthcoming.

Golez, B., & Koudijs, P. (2017). Four centuries of return predictability. *Journal of Financial Economics*. forthcoming.

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

Harrison, P. (2004). What Can We Learn for Today from 300-Year-Old Writings about Stock Markets?. *History of Political Economy* 36(4), 667–688.

Harvey, C.R., (2017). Presidential Address: The Scientific Outlook in Financial Economics. *Journal of Finance*, 72 (4), pp. 1399–1440.

Harvey, C.R., Liu, Y. & Zhu, H. (2016). … and the Cross-Section of Expected Returns, *Review of Financial Studies*, 29(1), 5-68.

Hou, K., Xue, C., and Zhang, L. (2017). Replicating anomalies. *NBER Working Paper* 23394.

Hurst, B., Ooi, Y. H., & Pedersen, L. H. (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. *The 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.C. (1968). The Performance of Mutual Funds in the Period 1945-1964, *Journal of Finance* 23, 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., M. Schularick, & A. M. Taylor. 2017. Macrofinancial History and the New Business Cycle Facts. *NBER Macroeconomics Annual* 2016, volume 31, edited by Martin Eichenbaum and Jonathan A. Parker. Chicago: University of Chicago Press.

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

Koijen, R. S., Moskowitz, T. J., Pedersen, L. H., & Vrugt, E. B. (2017). Carry, *Journal of Financial Economics*, forthcoming.

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

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

Linnainmaa, J. T., & Roberts, M. R. (2017). The History of the Cross Section of Stock Returns. Working paper.

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. C., (1973). An intertemporal capital asset pricing model. *Econometrica* 41, 867-887.

Moskowitz, T.J., Ooi Y.H., & Pedersen, L.H. (2012). Time series momentum, *Journal of Financial Economics*, 228-250.

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

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

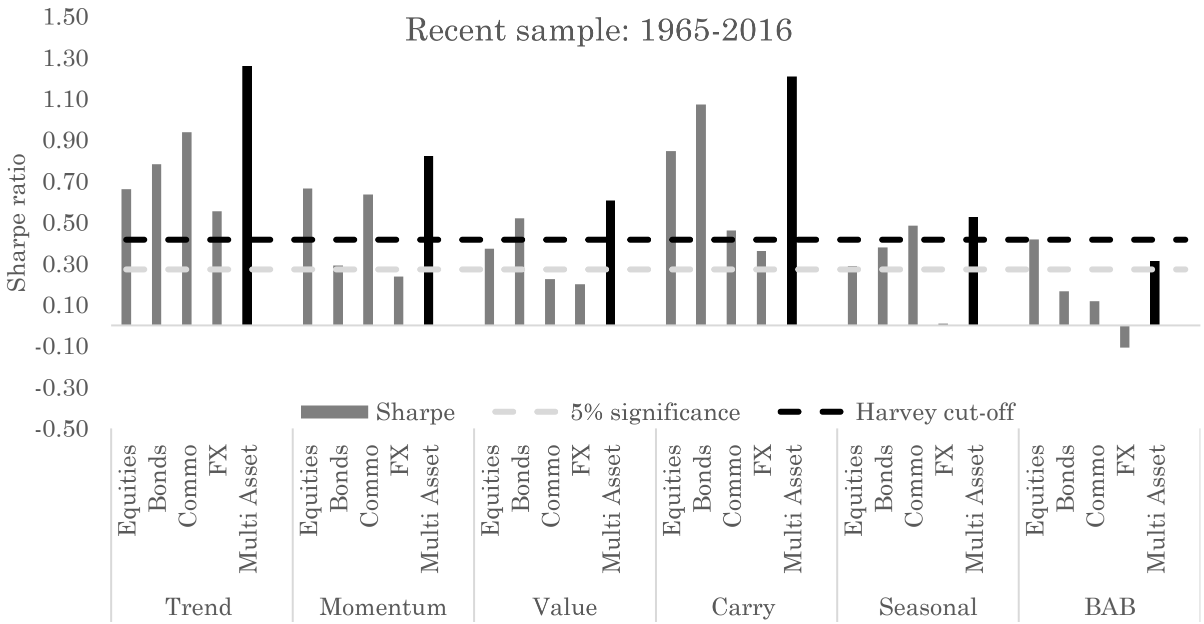
Taylor, A. M. (2002). A century of purchasing-power parity. *Review of Economics and Statistics*, 84(1), 139-150.

Tsiang, S. C. (1972). The rationale of the mean-standard deviation analysis, skewness preference, and the demand for money. *The American Economic Review*, *62*(3), 354-371.

Figures

Figure 1: Multi-asset factor returns since 1800. The figure shows the annualized Sharpe ratios for the six return factors for each of the individual four asset classes (grey color) and for the multi-asset combination (black color). 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. Panel A shows the results for the 'recent sample' period, spanning 1965 to 2016, and Panel B shows the results for the 'pre-sample' period (except for Commodities carry; black-white dashed bar), spanning 1800 till 1965. The individual factor portfolios are long-short portfolios applied on equity indices ("Equities"), 10-year maturity government bond indices ("Bonds"), commodities ("Commodities"), and currencies ("FX"). The multi-asset ("Multi Asset") factor portfolio consists of an equally-weighted combination of the four individual factor portfolios.

Panel A:



Panel B:

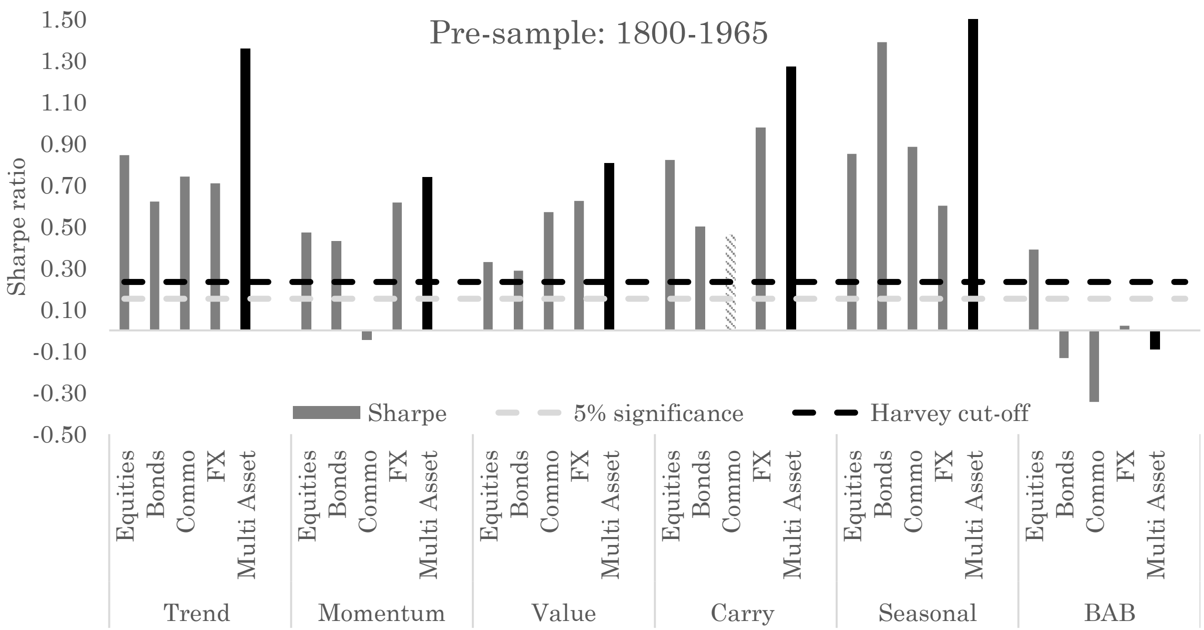


Figure 2: Global financial markets since 1800. The figure shows the number of markets included in the investment universe of the sample at each point in time. The sample starts in December 1799 and ends December 2016 and is at the monthly frequency. Covered are equity indices ("Equities"), 10-year maturity government bond indices ("Bonds"), commodities ("Commodities"), and currencies ("FX").

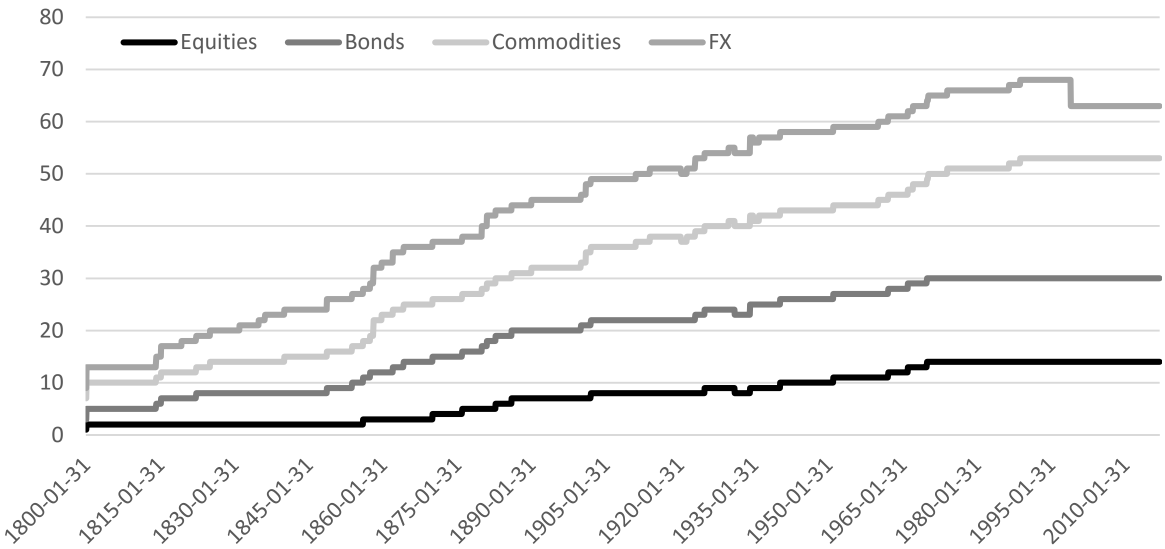
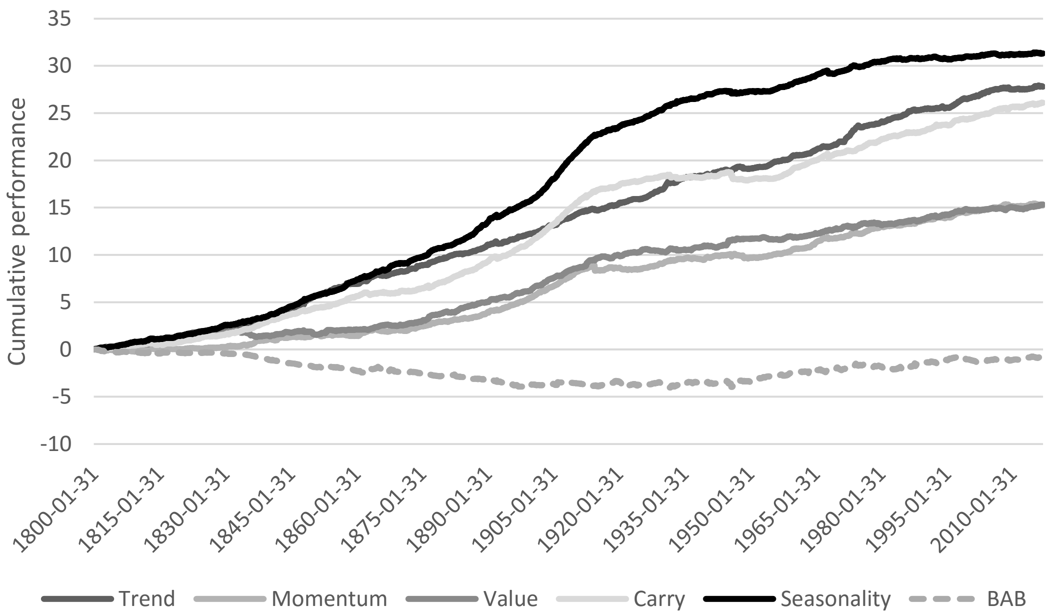


Figure 3: 217 years of multi-asset factor returns. The figure shows the cumulative performance of the six multi-asset factor premiums ("Trend", "Momentum", "Value", "Carry", "Seasonality" and "BAB"). The cumulative performance is shown in log-terms. The sample starts in December 1799 and ends December 2016, performance is measured on a monthly frequency. For comparison the "Trend", "Momentum", "Value", "Carry", "Seasonality" and "BAB" series are all scaled to a 10% ex post volatility for ease of comparison.



Tables

Table I: A 217 year sample

The table shows series starting date for each of the individual factor-asset class strategy combinations. Covered are equity indices ("Equities"), 10-year maturity government bond indices ("Bonds"), commodities ("Commodities"), currencies ("FX"), and the multi-asset factor series ("Multi Asset"), for the trend, momentum, value, carry, seasonality and BAB return factor series. The sample starts in December 1799 and ends December 2016 and is at the monthly frequency.

Starting date	Trend	Momentum	Value	Carry	Seasonality	BAB
Equities	1799-12-31	1801-03-31	1856-01-31	1805-02-28	1801-02-28	1801-04-30
Bonds	1799-12-31	1800-03-31	1872-03-31	1799-12-31	1800-02-28	1800-04-30
Commodities	1800-03-31	1800-03-31	1799-12-31	1968-01-31	1800-02-28	1800-04-30
FX	1800-03-31	1800-04-30	1871-03-31	1799-12-31	1800-03-31	1800-05-31
Multi Asset	1799-12-31	1800-03-31	1799-12-31	1799-12-31	1800-02-28	1800-04-30

Table II: Performance of multi-asset factor portfolio's within and across asset classes

The table summarizes the historical performance of the return factors. Shown are per factor per asset class the historical annualized Sharpe ratio (Panel A) and Jensen's alpha relative to relative to global excess equity, bond, currency and commodity market returns (Panel B). The alpha's are scaled to a 10% ex-post volatility for ease of comparison. The sample starts in December 1799 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 parentheses indicate t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level.

Panel A:

Sharpe ratio	Trend	Momentum	Value	Carry	Seasonality	BAB
Equities	0.75***	0.47***	0.29***	0.77***	0.67***	0.33***
	(11.09)	(6.92)	(3.68)	(11.27)	(9.80)	(4.81)
Bonds	0.61***	0.34***	0.30***	0.57***	1.12***	-0.12*
	(8.92)	(5.00)	(3.66)	(8.38)	(16.50)	(-1.74)
Commodities	0.74***	0.07	0.43***	0.43***	0.74***	-0.27***
	(10.96)	(1.07)	(6.34)	(2.98)	(10.85)	(-3.99)
FX	0.63***	0.46***	0.46***	0.77***	0.40***	-0.07
	(9.23)	(6.79)	(5.53)	(11.29)	(5.91)	(-1.07)
Multi Asset	1.29***	0.71***	0.71***	1.21***	1.45***	-0.04
	(18.99)	(10.43)	(10.42)	(17.81)	(21.39)	(-0.57)

Panel B:

Jensen's alpha	Trend	Momentum	Value	Carry	Seasonality	BAB
Equities	7.71***	5.29***	2.61***	7.77***	6.90***	3.42***
	(11.30)	(7.76)	(4.43)	(11.53)	(10.14)	(5.03)
Bonds	6.71***	4.07***	2.41***	6.03***	11.79***	-0.57
	(9.82)	(5.96)	(4.31)	(8.85)	(17.25)	(-0.83)
Commodities	8.29***	1.40**	4.83***	1.07***	7.83***	-2.07***
	(12.19)	(2.06)	(7.06)	(3.26)	(11.45)	(-3.09)
FX	6.63***	5.21***	3.51***	7.91***	4.40***	-0.35
	(9.71)	(7.63)	(6.32)	(11.64)	(6.45)	(-0.51)
Multi Asset	13.38***	7.70***	7.77***	12.15***	14.86***	0.01
	(19.55)	(11.29)	(11.41)	(17.88)	(21.76)	(0.01)

Table III: Statistical perspectives on multi-asset factor portfolio's within and across asset classes

The table summarizes various statistical perspective on the historical performance of the factors. Shown are per factor per asset class 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 starts in December 1799 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").

		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	0.02	0.00	0.00	0.00
	BE-odds	> 99,999	> 99,999	8.96	> 99,999	> 99,999	864
Bonds	p-value	0.00	0.00	0.00	0.00	0.00	0.08
	Bayesian-p	0.00	0.00	0.02	0.00	0.00	0.69
	BE-odds	> 99,999	2,121	8.36	> 99,999	> 99,999	0.09
Commodities	p-value	0.00	0.28	0.00	0.00	0.00	0.00
	Bayesian-p	0.00	0.80	0.00	0.15	0.00	0.01
	BE-odds	> 99,999	0.05	> 99,999	1.04	> 99,999	27.48
FX	p-value	0.00	0.00	0.00	0.00	0.00	0.28
	Bayesian-p	0.00	0.00	0.00	0.00	0.00	0.80
	BE-odds	> 99,999	> 99,999	31,621	> 99,999	> 99,999	0.05
Multi Asset	p-value	0.00	0.00	0.00	0.00	0.00	0.57
	Bayesian-p	0.00	0.00	0.00	0.00	0.00	0.78
	BE-odds	> 99,999	> 99,999	> 99,999	> 99,999	> 99,999	0.05

Table IV: Sub-period performances of multi-asset factor portfolio's within and across asset classes

The table summarizes the historical sub-period performances of the of the return factors. Shown are per factor per asset class the percentage of rolling 10-year period with positive Sharpe ratios ("P(SR_{10y}>0)"; Panel A), as well as performance per specific 50-year period, the 1800s, 1900s, and 2000s, and the full sample (Panel B). The sample starts in December 1799 and ends December 2016 and is at the monthly frequency. Covered in Panel A 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 B focuses on the Multi Asset factor return series. Numbers in parentheses indicate t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level, where in Panel A p-values are based on Newey-West to account for overlapping observations.

Panel A:

P(SR_{10y}>0)	**Trend**	**Momentum**	**Value**	**Carry**	**Seasonality**	**BAB**
Equity	98%***	87%***	71%***	94%***	88%***	82%***
Bond	99%***	80%***	84%***	89%***	92%***	34%*
Commo	95%***	61%	87%***	96%***	92%***	24%***
FX	91%***	86%***	72%**	83%***	67%**	43%
Multi Asset	100%***	94%***	95%***	94%***	99%***	44%

Panel B:

Sharpe ratio	Trend	Momentum	Value	Carry	Seasonality	BAB
1800-1849	1.31***	0.39***	0.41***	1.13***	1.54***	-0.59***
	(9.29)	(2.78)	(2.87)	(7.98)	(10.91)	(-4.18)
1850-1899	1.41***	0.88***	1.09***	1.59***	2.19***	-0.38***
	(9.96)	(6.24)	(7.74)	(11.27)	(15.50)	(-2.71)
1900-1949	1.37***	0.69***	0.89***	1.11***	2.20***	0.08
	(9.69)	(4.86)	(6.27)	(7.86)	(15.53)	(0.57)
1950-1999	1.67***	1.08***	0.60***	1.41***	0.80***	0.37***
	(11.78)	(7.64)	(4.27)	(10.01)	(5.68)	(2.60)
1800-1899	1.36***	0.66***	0.78***	1.37***	1.86***	-0.45***
	(13.62)	(6.55)	(7.79)	(13.66)	(18.59)	(-4.53)
1900-1999	1.51***	0.86***	0.75***	1.24***	1.50***	0.22**
	(15.09)	(8.56)	(7.50)	(12.40)	(14.95)	(2.18)
2000-2016	0.67***	0.46*	0.35***	1.05***	0.24	0.17
	(2.77)	(1.90)	(1.46)	(4.34)	(1.00)	(0.70)
Full sample	1.29***	0.71***	0.71***	1.21***	1.45***	-0.04
	(18.99)	(10.43)	(10.42)	(17.81)	(21.39)	(-0.57)

Table V: Robustness of performances of multi-asset factor portfolio's within and across asset classes

The table summarizes the robustness of the historical performance of the return factors to strategy construction choices. Factors are tested with the following different implementation methodologies: Quartile (Q1-Q4) long-short portfolios ("Quartile portfolios"), no volatility weighting of individual markets ("No volatility scaling"), a 1-month implementation lag ("One month lagged signals"), a quarterly rebalancing period ("Quarterly rebalance period"), and by capping returns of factor returns in a particular market at 50% ("Capping extreme returns"). Shown are the historical annualized Sharpe ratio for each factor for the equally weighted multi-asset combination across the four asset classes. The sample starts in December 1799 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.

Sharpe ratio	**Trend**	**Momentum**	**Value**	**Carry**	**Seasonality**	**BAB**
Baseline	1.29***	0.71***	0.71***	1.21***	1.45***	-0.04
	(18.99)	(10.43)	(10.42)	(17.81)	(21.39)	(-0.57)
Quartile portfolios	1.29***	0.68***	0.63***	1.10***	1.37***	0.01
	(18.99)	(9.94)	(9.26)	(16.22)	(20.20)	(0.17)
No volatility scaling	0.88***	0.38***	0.61***	0.88***	1.12***	-0.18***
	(12.93)	(5.60)	(8.94)	(12.92)	(16.46)	(-2.64)
One month lagged signals	1.09***	0.61***	0.73***	0.80***	-0.13*	0.01
	(15.99)	(8.94)	(10.81)	(11.76)	(-1.91)	(0.18)
Quarterly rebalance period	1.10***	0.58***	0.72***	1.06***	0.51***	-0.04
	(16.21)	(8.53)	(10.65)	(15.58)	(7.50)	(-0.59)
Capping extreme returns	0.99***	0.51***	0.66***	1.05***	1.35***	-0.04
	(14.54)	(7.55)	(9.65)	(15.49)	(19.92)	(-0.59)

Table VI: Common variation and portfolio value added of multi-asset factor portfolio's

The table summarizes the common variation of the global return factors and the added value from a portfolio perspective. Shown in panel A are the average pairwise monthly return correlations for each factor across the four asset classes ("Factor"), and for the six factor series per asset class and the equally weighted multi-asset combination across the four asset classes ("Asset class"). Panel B shows the results of spanning tests for each factor return series on all other factor return series for the equally weighted multi-asset combination across the four asset classes. Panel C contains the results of a mean-variance portfolio optimization. Shown are the portfolio weights, average annual return ("Avg. return"), and Sharpe ratio of the maximum Sharpe ratio portfolio with the restrictions that the exposure to each factor cannot be negative and all exposures together should sum to unity. The return series are scaled to a 10% ex-post volatility for ease of interpretability. The sample starts in December 1799 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: Correlations

	Trend	**Momentum**	**Value**	**Carry**	**Seasonality**	**BAB**
Factor	0.06***	0.04**	-0.01	0.02	0.03	0.01
	Equities	**Bonds**	**Commodities**	**FX**	**Multi Asset**	
Asset class	0.09***	0.07***	0.02	0.19***	0.11***	

Panel B: Spanning test

	Trend	**Momentum**	**Value**	**Carry**	**Seasonality**	**BAB**
Intercept	8.38***	-1.79***	8.96***	7.57***	9.98***	-1.95**
(annualized)	(13.20)	(-2.76)	(11.52)	(10.07)	(13.52)	(-2.44)
Trend		0.53***	-0.07***	0.02	0.09***	0.09***
		(32.56)	(-3.05)	(0.93)	(4.04)	(3.97)
Momentum	0.55***		-0.09***	0.16***	0.12***	0.01
	(32.56)		(-3.77)	(6.81)	(5.18)	(0.41)
Value	-0.05***	-0.06***		0.01	0.00	0.09***
	(-3.05)	(-3.77)		(0.78)	(0.23)	(4.48)
Carry	0.02	0.11***	0.02		0.23***	0.02
	(0.93)	(6.81)	(0.78)		(12.01)	(1.10)
Seasonality	0.07***	0.09***	0.00	0.23***		-0.01
	(4.04)	(5.18)	(0.23)	(12.01)		(-0.73)
BAB	0.06***	0.01	0.09***	0.02	-0.01	
	(3.97)	(0.41)	(4.48)	(1.10)	(-0.73)	
R2	33.98	35.63	2.58	10.01	10.55	1.66

Panel C: Portfolio weights and ex-post Sharpe ratio

Equities	Bonds	Trend	Momentum	Value	Carry	Seasonality	BAB	Avg. return	Sharpe ratio
100%	0%							3.64%	0.34
19%	0%	81%						11.50%	1.38
31%	0%		69%					6.37%	0.84
33%	0%			67%				6.30%	0.86
15%	0%				85%			11.28%	1.28
15%	0%					85%		13.33%	1.53
100%	0%						0%	3.64%	0.34
6%	0%	26%	0%	22%	19%	26%	0%	11.73%	2.22

Table VII: Downside risk perspectives on multi-asset factor portfolio's

The table reports the beta's, downside beta's, annualized CAPM alpha's and downside-CAPM alpha's for each factor for the equally weighted multi-asset combination across the four asset classes. The (downside) betas are calculated versus the excess return of the equity market portfolio and uses the LPM(*2,threshold*) definition. We depict downside betas for thresholds of 0, -1 and -2 standard deviations away from zero. The sample starts in December 1799 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.

Beta's	Trend	Momentum	Value	Carry	Seasonal	BAB
Beta	0.01	-0.01	-0.07	0.09	0.04	0.06
DBeta_0	-0.01	0.02	-0.08	0.16	0.09	0.04
DBeta_1std	0.00	0.04	-0.09	0.20	0.10	0.04
DBeta_2std	-0.04	0.01	-0.08	0.23	0.10	0.05

Alpha's	Trend	Momentum	Value	Carry	Seasonal	BAB
Alpha_CAPM	13.34***	7.61***	7.84***	12.28***	14.87***	-0.12
	(19.65)	(11.21)	(11.58)	(18.17)	(21.93)	(-0.17)
Alpha_DB0	13.39***	7.48***	7.86***	12.04***	14.72***	-0.04
	(19.73)	(11.02)	(11.62)	(17.76)	(21.68)	(-0.06)
Alpha_DB1	13.38***	7.41***	7.91***	11.86***	14.65***	-0.03
	(19.71)	(10.90)	(11.68)	(17.42)	(21.56)	(-0.05)
Alpha_DB2	13.52***	7.52***	7.85***	11.75***	14.67***	-0.08
	(19.89)	(11.07)	(11.60)	(17.19)	(21.60)	(-0.12)

Table VIII: Macro-economic and market states performance of multi-asset factor portfolio's

The table summarizes the historical performance of global return factors across various macro-economic and market sub-periods. Sub-periods are examined at an annual frequency and include recession versus non-recession, global crisis versus non-crisis, bear and bull equity markets, and Q2-Q4 low-volatility 'Tranquil' periods versus Q1 high-volatility 'Turbulent' periods. Shown are per specific sub-period the historical annualized (annualized) return for the equally weighted multi-asset combination across the four asset classes. The sample starts in December 1799 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.

	Trend	**Momentum**	**Value**	**Carry**	**Seasonality**	**BAB**
Recession	13.47***	6.59***	10.15***	14.15***	18.32***	-4.17***
(74 years)	(11.59)	(5.67)	(8.74)	(12.18)	(15.79)	(-3.60)
Expansion	13.31***	8.08***	6.24***	11.81***	13.33***	2.34***
(143 years)	(15.91)	(9.65)	(7.48)	(14.12)	(15.97)	(2.80)
Difference	0.16	-1.49	3.91***	2.34	4.99***	-6.50***
	(0.11)	(-1.04)	(2.73)	(1.64)	(3.49)	(-4.56)
Crisis	13.09***	4.72***	9.28***	7.70***	13.63***	-0.48
(52 years)	(9.44)	(3.41)	(6.69)	(5.57)	(9.83)	(-0.34)
Non-crisis	13.45***	8.47***	7.04***	14.16***	15.48***	0.30
(165 years)	(17.28)	(10.88)	(9.05)	(18.24)	(19.88)	(0.39)
Difference	-0.36	-3.75**	2.23	-6.46***	-1.85	-0.78
	(-0.23)	(-2.36)	(1.41)	(-4.07)	(-1.16)	(-0.49)
Bear market	17.24***	8.75***	9.70***	13.26***	15.02***	-2.09
(43 years)	(11.32)	(5.74)	(6.36)	(8.69)	(9.85)	(-1.37)
Bull market	12.41***	7.27***	7.05***	12.45***	15.04***	0.66
(174 years)	(16.39)	(9.59)	(9.30)	(16.42)	(19.83)	(0.87)
Difference	-4.83***	-1.48	-2.65	-0.81	0.02	2.76
	(-2.84)	(-0.87)	(-1.55)	(-0.48)	(0.01)	(1.62)
Tranquil periods	12.65***	7.79***	8.16***	13.79***	16.66***	-1.51*
(163 years)	(16.15)	(9.95)	(10.42)	(17.63)	(21.34)	(-1.93)
Turbulent periods	15.54***	6.88***	5.81***	9.04***	10.10***	5.05***
(54 years)	(11.42)	(5.05)	(4.27)	(6.65)	(7.44)	(3.71)
Difference	2.89*	-0.91	-2.35	-4.75***	-6.56***	6.56***
	(1.84)	(-0.58)	(-1.49)	(-3.03)	(-4.19)	(4.18)

Appendix

Portfolio construction procedure

After having 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 time-series momentum factor, which is directional in nature, we go long (short) markets in each asset class when the time-series momentum measure is positive (negative), following Moskowitz, Ooi, and Pedersen (2012). For the other factors, who are all cross-sectional in nature, we rank the markets in each investment universe based on the factor measure and take the 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 (2017). (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 time-series momentum add up to zero at each point in time:

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

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 betting-against-beta only), in the same spirit as Asness, Moskowitz, and Pedersen (2013), Frazzini and Pedersen (2014), and Moskowitz, Ooi, and Pedersen (2012) but fit 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 estimate at the maximum of the 10% quantile of volatility estimates per asset class or 2.5%, 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. Further, we 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). Again, we floor this estimate to prevent extreme leverage at twice its long-run average. This approach takes an ex ante view of portfolio construction, as available in real-time. However, our results are not materially different if we simple scale by the in-sample ex-post volatility, as applied in Koijen, Moskowitz, Pedersen and Vrugt (2017).

We rebalance the portfolios each month based on the signals and various 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 such '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.[16]

We have one additional portfolio construction rule, motivated by practical considerations. Due to problems with exceptionally high levels of data uncertainty, we exclude hyperinflation episodes from our sample by excluding assets from countries with ex-ante levels of monthly inflation over 50 percent. This definition follows Cagan (1956)

[16] 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.

and is widely accepted. 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 during 1920-1921, and France, Italy, and Japan during the post-World War II years (1946-1949/1950 and 1943-1948 for Italy), and South Korea during 1950-1956.

Dataset construction

Financing rates: Our main measure for the financing rates are the 3-month LIBOR rates (sourced from Bloomberg and Datastream), spliced with (in order of usage) Eurodollar rates from Datastream, 3-month Treasury-bill rates and commercial paper yields from Global Financial Data, short rates from Dimson, Staunton and Marsh, Macrohistory and for the U.S. with data from Jeremy Siegel. When all are unavailable we splice with lagged Treasury-bill returns.

Equities: We source equity futures and indices price and return data 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 (2017). Next, we backfill these data with equity index level data downloaded from Global Financial Data. From the same sources we obtain dividend yields. 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.

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. We apply a two-months lag to inflation numbers to mimic its real-time availability. The markets we consider are the major developed bond markets around the globe.

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, Dimson, Staunton and Marsh, and for the U.S. with data from Jeremy Siegel. Purchasing Power Parity data is obtained from the OECD website, and before 1971 with data from Macrohistory. 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.

Commodities: We source commodity futures price and return data from Bloomberg, spot prices from Bloomberg and Datastream, both spliced with futures data and spot data from Global Financial Data. For carry, we use the front futures and second futures prices. We use the following contracts based on their general usage and liquidity. Futures data available from GFD extents before 1964 for the following 6 markets; Wheat, Corn, Cotton, Cocoa, Coffee and Copper. Further, we spline the futures data of the front contract with spot data before their inception dates. Due to restrictions on tradability we ignore 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).

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 USD, and for commodities we also use an equal-weighted portfolio of all commodities 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.[17]

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

Table A1: A 217 year sample

The table shows the start dates of the individual return series, value and carry measures of the markets included in the investment universe. Covered in Panel A are equity indices ("Equities"), 10-year maturity government bond indices ("Bonds"), commodities ("Commodities"), and currencies ("FX").

Starting date	Country	Index	Returns	Value	Carry
Equities	U.S.	S&P500	1800-02-28	1801-01-31	1805-01-31
	U.K.	FTSE100	1799-12-31	1923-12-31	1799-12-31
	Germany	DAX	1870-01-31	1869-12-31	1874-12-31
	France	CAC	1856-01-31	1855-12-31	1860-12-31
	Japan	TOPIX	1886-01-31	1886-12-31	1890-12-31
	Australia	ASX200	1882-10-31	1882-10-31	1887-09-30
	Canada	TSE60	1934-01-31	1934-01-31	1938-12-31
	Switzerland	SMI	1966-02-28	1966-01-31	1971-01-31
	Sweden	OMX	1902-01-31	1871-12-31	1906-12-31
	Spain	IBEX	1876-01-31	1875-12-31	1880-12-31
	Italy	MIB	1925-01-31	1925-01-31	1929-12-31
	the Netherlands	AEX	1951-01-31	1969-07-31	1955-12-31
	Hong Kong	HANGSENG	1969-12-31	1972-12-31	1974-12-31
	South Korea	KOSPI200	1962-02-28	1963-01-31	1967-01-31
Bonds	U.S.	10Y bond	1799-12-31	1872-02-29	1799-12-31
	U.K.	10Y bond	1799-12-31	1872-02-29	1799-12-31
	Germany	10Y bond	1815-03-31	1869-03-31	1815-02-28
	France	10Y bond	1800-02-28	1872-02-29	1800-01-31
	Japan	10Y bond	1880-01-31	1872-02-29	1879-12-31
	Australia	10Y bond	1857-06-30	1872-02-29	1857-06-30
	Canada	10Y bond	1934-03-31	1872-02-29	1934-02-28
	Switzerland	10Y bond	1900-01-31	1893-12-31	1893-12-31
	Sweden	10Y bond	1853-10-31	1872-02-29	1854-01-31
	Spain	10Y bond	1881-01-31	1872-02-29	1880-12-31
	Italy	10Y bond	1862-01-31	1872-02-29	1861-12-31
	the Netherlands	10Y bond	1814-03-31	1872-02-29	1814-02-28
	New Zealand	10Y bond	1923-02-28	1915-10-31	1923-01-31
	Norway	10Y bond	1822-04-30	1872-02-29	1822-03-31
	Belgium	10Y bond	1848-09-30	1872-02-29	1848-08-31
	Denmark	10Y bond	1864-03-31	1872-02-29	1864-02-29

Starting date	Country	Index	Returns	Value	Carry
Commodities	-	Soy oil	1911-02-28	1915-07-31	1968-04-30
	-	Corn	1858-02-28	1862-07-31	1968-06-30
	-	Cocoa	1799-12-31	1799-12-31	1967-12-31
	-	Oil (WTI)	1859-10-31	1864-03-31	1983-05-31
	-	Oil (Brent)	1988-10-31	1974-08-31	1989-02-28
	-	Cotton	1799-12-31	1799-12-31	1968-02-29
	-	Gold	1974-01-31	1974-01-31	1975-01-31
	-	Copper	1800-02-28	1804-07-31	1994-10-31
	-	Heating Oil	1967-02-28	1971-07-31	1986-06-30
	-	Gasoline	1986-07-31	1987-07-31	1986-12-31
	-	Coffee	1825-02-28	1829-07-31	1973-02-28
	-	Kansas Wheat	1970-03-31	1983-06-30	1970-02-28
	-	Aluminum	1901-01-31	1905-06-30	1997-08-31
	-	Cattle	1858-02-28	1862-07-31	1965-04-30
	-	Live Hog	1858-02-28	1862-07-31	1987-01-31
	-	Nickel	1901-01-31	1905-06-30	1997-08-31
	-	Zinc	1840-02-29	1844-07-31	1997-08-31
	-	Natural Gas	1890-01-31	1894-06-30	1990-05-31
	-	Soybeans	1913-12-31	1918-05-31	1968-04-30
	-	Sugar	1960-02-29	1964-07-31	1968-02-29
	-	Silver	1799-12-31	1799-12-31	1975-01-31
	-	Soymeal	1929-11-30	1934-04-30	1968-04-30
	-	Wheat	1799-12-31	1799-12-31	1968-06-30
FX	-	AUD/USD	1834-12-31	1871-02-28	1834-11-30
	-	CAD/USD	1934-03-31	1871-02-28	1934-02-28
	-	CHF/USD	1831-01-31	1871-02-28	1830-12-31
	-	DEM/USD	1815-03-31	1871-02-28	1815-02-28
	-	EUR/USD	1999-01-31	1871-02-28	1815-02-28
	-	GBP/USD	1799-12-31	1871-02-28	1799-12-31
	-	JPY/USD	1880-01-31	1871-02-28	1879-12-31
	-	NOK/USD	1819-05-31	1871-02-28	1819-04-30
	-	NZD/USD	1923-02-28	1920-09-30	1923-01-31
	-	SEK/USD	1836-03-31	1871-02-28	1836-02-29
	-	FRF/USD	1800-02-28	1871-02-28	1800-01-31
	-	ITL/USD	1862-01-31	1871-02-28	1861-12-31
	-	ESP/USD	1881-01-31	1871-02-28	1880-12-31
	-	NLG/USD	1814-03-31	1871-02-28	1814-02-28
	-	BEF/USD	1848-09-30	1871-02-28	1848-08-31

Correlations of factors

Table A.2: Common variation per global return factor

The table shows the individual correlations coefficients of the global return factors. Shown are the monthly return correlations between the equally weighted multi-asset combination of each factor series across the four asset classes (Panel A), and between each factor-asset class series (Panel B). The sample starts in December 1799 and ends December 2016 and is at the monthly frequency.

Panel A:

	Trend	Mom	Value	Carry	Seasonal	BAB
Trend						
Momentum	0.56					
Value	-0.16	-0.35				
Carry	0.16	0.17	0.04			
Seasonal	0.08	0.11	-0.11	0.09		
BAB	0.12	0.05	0.18	0.10	0.04	

Panel B:

		Trend				Momentum				Value				Carry				Seasonal				BAB			
		EQ	BND	COM	FX	EQ	BND	COM	FX	EQ	BND	COM	FX	EQ	BND	COM	FX	EQ	BND	COM	FX	EQ	BND	COM	FX
Trend	Equity																								
	Bond	0.24																							
	Commo	0.09	0.07																						
	FX	0.08	0.11	0.15																					
Momentum	Equity	0.29	0.11	0.08	0.17																				
	Bond	0.12	0.28	0.10	0.11	0.13																			
	Commo	0.05	-0.02	0.64	0.08	0.02	0.06																		
	FX	0.14	0.08	0.13	0.59	0.13	0.17	0.05																	
Value	Equity	0.06	0.07	-0.05	-0.07	-0.22	-0.02	0.02	-0.01																
	Bond	0.00	-0.03	-0.08	-0.02	-0.02	-0.14	-0.07	-0.07	0.03															
	Commo	0.01	0.02	-0.30	-0.11	-0.05	-0.05	-0.49	-0.04	0.02	0.03														
	FX	0.08	-0.05	-0.14	-0.22	0.02	-0.09	-0.13	-0.31	0.04	0.09	0.06													
Carry	Equity	0.03	0.12	0.00	0.03	0.04	0.02	-0.02	0.02	0.06	0.02	-0.08	-0.01												
	Bond	0.03	0.16	0.00	0.07	0.04	0.20	-0.02	0.06	0.01	0.27	-0.03	-0.04	0.14											
	Commo	0.08	0.02	0.20	-0.02	0.02	0.03	0.28	0.03	0.01	0.08	-0.21	-0.06	-0.06	0.01										
	FX	0.09	0.04	-0.04	0.02	0.06	0.01	-0.04	0.05	-0.03	0.02	0.00	0.21	0.08	0.19	-0.01									
Seasonal	Equity	0.09	0.00	-0.01	0.06	0.07	-0.01	0.03	0.02	-0.05	-0.06	-0.10	0.06	0.06	0.01	0.07	-0.02								
	Bond	-0.05	-0.01	0.02	0.01	0.04	0.11	0.02	0.00	-0.02	0.00	-0.02	-0.03	0.09	0.12	-0.04	0.06	0.03							
	Commo	-0.03	0.03	0.04	0.00	-0.04	0.02	0.11	-0.01	-0.03	-0.02	-0.07	-0.03	0.04	0.02	0.00	-0.04	0.04	0.00						
	FX	0.02	0.03	0.03	0.15	0.04	0.04	-0.02	0.09	-0.03	0.01	-0.03	-0.06	0.02	0.05	-0.06	0.01	0.06	0.07	0.00					
BAB	Equity	0.40	0.07	0.04	0.02	0.06	0.06	0.06	0.03	0.16	0.01	-0.04	0.09	0.02	-0.03	0.08	0.12	0.15	0.06	-0.03	0.03				
	Bond	0.02	0.05	-0.02	0.02	-0.07	-0.07	0.02	0.01	0.10	0.05	0.05	0.06	0.02	0.07	0.01	0.05	-0.05	-0.06	0.01	0.00	0.06			
	Commo	-0.03	0.04	0.03	-0.03	0.02	0.12	-0.04	0.05	-0.04	0.10	-0.05	0.01	0.00	0.03	0.08	0.06	-0.01	0.05	0.04	0.02	-0.01	-0.02		
	FX	0.13	-0.01	-0.04	-0.08	-0.03	-0.02	0.00	-0.02	0.07	0.00	0.06	0.33	-0.05	-0.16	0.08	0.03	0.07	-0.05	-0.06	-0.11	0.13	0.04	0.10	