The Cross-Section of Stock Returns before CRSP

Guido Baltussen*

Bart P. Van Vliet

Pim Van Vliet

Abstract

This study examines the cross-section of stock returns out-of-sample using a newly created database of U.S. stocks between 1866 and 1926. By augmenting price and dividend data for 1,488 stocks with hand-collected data on market capitalizations, this dataset significantly extends the traditional CRSP sample. Over the 'pre-CRSP' era the relationship between market beta and returns is flat, and most factor premia are sizable and significant. Machine learning methods validate these findings. Importantly, being unaffected by post-publication arbitrage we find that equity factor premia do not materially decay out-of-sample, addressing p-hacking concerns. Additionally, we explore economic explanations of factor premia over the combined pre-CRSP and CRSP samples covering 154 years of data.

This Version: June 2024

Keywords: empirical asset pricing, return anomalies, factor premia, p-hacking, momentum, value, beta, low-risk, size, reversal, machine learning.

JEL classification: G10, G11, N21, N22.

*Baltussen (corresponding author) is from the Erasmus School of Economics, Erasmus University Rotterdam and Northern Trust Asset Management – Quantitative Strategies. Email: baltussen@ese.eur.nl. B. P. Van Vliet is from the Erasmus School of Economics, Erasmus University Rotterdam and Northern Trust Asset Management – Quantitative Strategies. P. Van Vliet is from Robeco Institutional Asset Management. We would like to thank Rob van Bommel, Zhi Da, Amit Goyal, Bill Goetzmann, Campbell Harvey, Edward McQuarrie, and three anonymous referees for valuable contributions and discussions. We thank Amar Soebhag and Pieter Van Der Velde for assistance on collecting the data, and the Erasmus University Rotterdam and Erasmus Research Institute of Management for financial support. Northern Trust Asset Management and Robeco Institutional Asset Management, the firms that employ the authors, offer various multi-factor investment products. The construction of these products may, at times, draw on insights related to this research. The views expressed in this paper are not necessarily shared by Northern Trust Asset Management or Robeco Institutional Asset Management. We welcome comments, including references to related papers we have inadvertently overlooked.

I. Introduction

Seminal studies reveal canonical factors that predict cross-sectional differences in stock returns. Amongst others, Fama and French (1992, 1993) show that size and value are priced stock characteristics. Jegadeesh and Titman (1993, 2001) and Carhart (1997) identify momentum, Ang, Hodrick, Xing, and Zhang (2006), Blitz and Van Vliet (2007), and Frazzini and Pedersen (2014) identify low-risk via (idiosyncratic) volatility or beta, and Jegadeesh (1990), and Lehmann (1990) identify short-term reversal. These factors are by now heavily studied with a wide industry actively allocating to them. At the same time, recent studies by Harvey (2017) and Fama and French (2018), amongst others, have raised concerns regarding the robustness of findings in the cross-section of stock returns, particularly highlighting the issue of p-hacking. Numerous predictive variables found in the cross-section of stock returns that seem important in-sample lose explanatory power, or even fail to hold up, out-of-sample.[1]

To illustrate this point further, Figure 1 summarizes the t-statistics of the CAPM alphas of the canonical equity factors – size, value, momentum, short-term reversal, and low-risk – based on standard 2x3 portfolio sorts across both their respective in-sample periods and subsequent out-of-sample periods. The in-sample period spans the sample period of the original studies (Black, Jensen, and Scholes, 1972, Banz, 1981 and Fama and French, 1992, Basu, 1977 and Fama and French, 1992, Jegadeesh and Titman, 1993, Lehmann, 1990 and Jegadeesh, 1990), while the out-of-sample period spans the period thereafter until the end of 2019. Consistent with the earlier studies, all factors except size exhibit significance over the in-sample period. However, results look materially different over subsequent out-of-sample periods with most factors losing their significance at traditional confidence levels. Only momentum and low-risk factors remain significant, whereas size, value and reversal factors

---

[1] As a case in a point, Harvey, Liu, and Zhu (2016) find 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. Related, Hou, Xue, and Zhang (2020) conduct a large-scale replication study of 447 anomalies and find that 65% are insignificant at the 5% level using conventional critical values and 85% are insignificant using a critical value of three. Chordia, Goyal, and Saretto (2020) show that of about 2.1 million possible trading strategies only a small group survives after correcting for a multiple hypothesis testing bias.

are insignificant out-of-sample. This decay in the performance of equity factors raises concerns about their validity potentially attributable to p-hacking and underscores the need for independent out-of-sample tests.[2]

INSERT FIGURE 1 HERE

In this study we address this challenge and thoroughly examine the canonical equity factors out-of-sample. A complicating factor for independent out-of-sample studies is factor performance decay due to post-publication arbitrage as noted by McLean and Pontiff (2016). They find that out-of-sample performance decay of several anomalies increases due to post-publication arbitrage.[3] To rule out the arbitrage hypothesis effectively, the most powerful way for independent out-of-sample test is to go backwards in time. Interestingly, the period before the start date of the Center for Research in Security Prices (CRSP) dataset (i.e., 1926) provides a unique opportunity. This 'pre-CRSP' period was a period of strong economic growth and rapid industrial developments, with stock markets playing a pivotal role. The available sample period spans a sizeable 61-years (1866-1926), which is comparable to the length used in many existing CRSP-based studies.[4] Consequently, conducting tests in the pre-CRSP era provides a powerful ground for independent out-of-sample tests, while at the same time

---

[2] These worries hold yet while many of the canonical factors have been studied in international markets. For example, although the value factor is significant in European and Asian stock markets, size and momentum yield very mixed results (see Fama and French, 1997, 2012, 2017). Further, international markets are not independent of the U.S., as stock markets and its cross-sectional return patterns are integrated globally to a large degree (e.g. Griffin, Yi and Martin, 2003, Asness, Moskowitz and Pedersen, 2013).

[3] McLean and Pontiff (2016) find a post-publication performance decay of 58% over on average 13 years of post-publication data, which they attribute to arbitrage trading. In contrast, they find a much smaller decay in the post-sample period that is *prior* to publication of the results, although their out-of-sample period is short (on average 5-years), thereby limiting the power of their tests. Jacobs and Muller (2020) find that the United States is the only country with a reliable post-publication decline, while the effect is generally not present in international markets. Further, Jensen, Kelly, and Pederson (2021) argue that the observed post-publication decline in anomaly profits could be a statistical artifact of multiple sequential tests from the same data generating process.

[4] The first empirical tests of the Capital Asset Pricing Model (CAPM) by Black, Jensen, Scholes (1972) and Fama and MacBeth (1973) used a sample period of about 40-years. In the years that followed, the main asset pricing findings were that beta is not significantly related to return, whereas other factors such as dividend yield (Litzenberger and Ramaswamy,1979) and firm size (Banz, 1981) are related to return. These studies relied on samples of 500 to 1,000 NYSE-listed stocks over a 40 to 50 years sample periods. In their seminal paper, Fama and French (1992) use a sample period covering 28.5 years of data (1963-1990), and since most studies to the cross-section of stock returns span the post-1963 CRSP sample period.

negating the arbitrage hypothesis for factor decay, since investors at that time could not have traded on insights from future research.

To this end, we construct a novel database covering the 1,488 stocks traded on the U.S. exchanges between January 1866 and December 1926. The database consists of stock prices, dividend yields, and, importantly, hand-collected market capitalization values. To the best of our knowledge, we are the first to create an extensive dataset for this period that also includes market capitalization values. This is imperative as an historical abundance of small capitalization stocks over the pre-CRSP era could render findings economically less important.[5] We apply a wide range of data quality filters to the pre-CRSP dataset to ensure good data quality, external validity, and sufficient economic importance.

The study focuses on the equity factors most commonly studied that can be constructed over the pre-CRSP sample; market beta, firm size, value, 12-2 months momentum, and short-term 1-month reversal. We test dividend yield as proxy for value, as in the 19th century dividends were widespread, strongly associated with earnings (Braggion and Moore, 2011), and seen as the common valuation technique for stocks (Dowrie and Fuller, 1950, Poitras, 2010). We choose to avoid a wider set of variables to mitigate concerns about p-hacking as much as possible, while the absence of comprehensive and cross-sectional comparable income and balance sheet data records in the pre-CRSP era prevents the inclusion of quality-related factors such as profitability or investments into our analysis.

We start our analysis with Fama-MacBeth (1973) regressions and univariate portfolio sorts, both of which we value-weight to prevent an undue impact of smaller stocks. In line with Black, Jensen and Scholes and Fama-MacBeth we find that market beta is not priced in the cross-section and the CAPM on average fails to explain asset prices: low-beta stocks have positive alpha and high-beta stocks have negative alpha over the 1866-1926 sample. Further, momentum and value carry significant cross-sectional premia. By contrast, size has no

[5] Amongst others, Fama and French (2008), Hou, Xue, and Zhang (2020), and Novy-Marx and Velikov (2022) show that many equity anomalies fail in the post-1963 sample when the smallest stocks are excluded.

significant return spread in portfolio sorts, while short-term reversal is only significant in Fama-MacBeth regressions.

Next, we build 'Fama-French style' market-neutral factor portfolios, double-sorted on size and a factor characteristic.[6] As size is known to interact strongly with other characteristics (e.g., Fama and French, 1992, 2012, and Israel and Moskowitz, 2013) and our historical sample includes sufficient coverage on market capitalization, we are able to control for interaction with size. The main findings are summarized in Figure 2. We find economically substantial and statistically significant premia (as reflected in CAPM alphas) for momentum, value, and low-risk, but insignificance of the size and short-term reversal premia. Additional extensive robustness tests confirm that value, momentum, and low-risk are generally robust and remain significant across testing choices, noting that the size premium becomes significant once the smallest, least liquid, but lowest data quality stocks are included in the sample.

INSERT FIGURE 2 HERE

Comparing the pre-CRSP and CRSP sample results allows for a direct examination of the impact of p-hacking, while controlling for the competing arbitrage hypothesis for factor decay. Data snooping influences factors returns by artificially inflating in-sample returns and reducing their covariances (Linnainmaa and Roberts, 2018, McLean and Pontiff, 2016). We find that out-of-sample decay of stock factor premia is relatively small and statistically insignificant, with premia averaging 4.2% in the pre-CRSP sample and 4.9% during the CRSP sample period. Additionally, we find no evidence for changes in factor correlations. Interestingly, these findings align in size with a 26% out-of-sample decay observed by McLean and Pontiff (2016) over 5-years of post- sample periods yet diverge from Linnainmaa and

[6] Akin to Frazzini and Pedersen (2014) we lever beta-sorted portfolios to be market-neutral.

Roberts (2018)'s findings on accounting-based anomalies for the 1938-1963 period. The absence of significant out-of-sample decay supports the conclusion that the value, momentum, and low-risk factors are not a result of data snooping. Overall, we conclude that value, momentum, and low-risk stand out as persuasive empirical equity factors in out-of-sample analysis.

To further understand the importance of the canonical factors in the cross-section of stock returns, we employ machine learning methods validated in the empirical asset pricing literature, now applied to the new pre-CRSP sample. Gu, Kelly and Xiu (2020) find that machine learning models predict cross-sectional differences in stock returns for the period 1957-2016. These methods are a powerful tool to summarize key predictor variables in a data-driven manner, and hence uncover priced variables beyond the five prominent characteristics without conducting a sizable data-dredging exercise. Over the pre-CRSP sample period we find that machine learning models primarily identify the five canonical variables, albeit through a completely data-driven process with little prior assumptions or guidance. As a result, portfolios sorted based on machine learning model predictions yield no significant added value over and above the canonical equity factors.

Finally, we explore several features of the early sample period to provide insights into economic explanations of stock factor premia. The 1866-1926 period is interesting for several reasons. First, it is characterized by large macroeconomic shocks and market fluctuations, enriching the traditional CRSP sample with additional market and business cycles, allowing for deep-sample insights into macroeconomic risk explanations. Second, delegated asset management was notably absent over this period (Rouwenhorst, 2004), hence providing a natural test on the impact of delegated management or benchmarks as limit to arbitrage. Third, the period allows for an insightful examination of momentum's relation with crash risk, as highlighted by Daniel and Moskowitz (2016), given that crashes are infrequent yet critical events. Fourth, we explore more general downside risk explanations of stock factors, in spirit of Bawa and Lindenberg (1977). Our findings reveal that factor premia generally bear no significant relation to common macroeconomic factors nor the delegated management hypothesis. While momentum is exposed to crash risk, factor premia are hard to align with downside risk explanations.

This study contributes to the body of empirical asset pricing studies utilizing 'pre-samples'. Most notably, two recent studies examine a single factor, momentum, in the cross-section of stock returns before 1926. Geczy and Samonov (2016) study momentum in the U.S. pre-1926 period, identifying a significant momentum premium. However, their dataset lacks both dividends and market capitalization data, which impacts their findings due to the historical prevalence of small-cap stocks.[7] Goetzmann and Huang (2018) find a positive momentum premium in the imperial Russia stock market from 1865 to 1914. Due to the unavailability of data on shares outstanding, they also have had to rely on equal-weighted returns. Unlike these studies, our research (i) uses multiple characteristics, (ii) applies value-weighting, and (iii) includes both the capital appreciation and distribution component of returns.[8] In a broader context, Siegel (1992) gives evidence of equity premia stretching back to 1800, Goetzmann (1993) extends this to 1695, and Golez and Koudijs (2018) go even further back to 1629.

Presumably closest related, Baltussen, Swinkels and Van Vliet (2021) study global factor premia out-of-sample across equity, bond, currency, and commodity markets stretching back to 1800. They find a strong and consistent presence of the majority of global factor premia such as value, momentum, and low-risk in 217 years of data. However, they do not study the cross-section of U.S. stock returns but rather focus on factors across global markets, which

[7] Dividends were historically a major source of return, on average accounting for 51% of the average stock return and 81% of the value-weighted returns.

[8] Other studies examine the out-of-sample performance of accounting-based variables over the early CRSP sample. Linnainmaa and Roberts (2018) show profitability and investment factors are absent over the early CRSP period (i.e., 1926-1963), while Wahal (2019) finds a sizable profitability premium between 1940 and 1963 but an insignificant investment factor premium. Davis, Fama and French (2000) find a positive value premium pre-sample over the 1929-1963 period, while Linnainmaa and Roberts (2018) show size and value factor premia are absent over the early CRSP period.

are largely uncorrelated with ours. Consequently, their findings complement ours, albeit on an independent set of assets.

The remainder of this paper is structured as follows. Section II describes the history of the U.S. equity market and the novel database of U.S. stock prices between 1866 and 1926. Section III analyses the cross-section of stock returns via Fama-MacBeth regressions and portfolio sorts. Section IV analyses the out-of-sample decay of factor premia. Section V discusses historical investability, followed by the machine learning analysis in Section VI. Section VII examines insights into economic mechanisms. Section VII concludes. The Online Appendix provides extensive detail on the sample construction, data quality, historical investability, additional tests, and the machine learning models.

II. The 'pre-CRSP' U.S. stock dataset: 1866 – 1926

The 1866-1926 period was characterized by strong economic growth and industrial developments, laying the foundations of the U.S. becoming the leading economic power in the world. Demand and supply of stock financing grew rapidly resulting in the U.S. stock market experiencing rapid growth. By the early 20th century, the stock market was large relative to the size of the U.S. economy, with a (New York) stock market capitalization to GDP ratio of 174%, about similar levels as observed in 2015 (Neal, 2016). Stocks traded across various exchanges, most notably the New York Stock Exchange (NYSE) and the New York Curb (the predecessor of the AMEX). Most of the trading activity took place on the NYSE, followed by the NY Curb and regional exchanges (mainly Boston and Philadelphia) (Brown et al., 2008, O'Sullivan, 2007). Total annual shares trading volume rose from about 100 million in 1885, to 150 million in 1900, to 250 million in 1915, to 1,151 million shares in 1930. Over two-thirds of trading volume originated from the NYSE, followed by the New York Curb (about 20%), and regional exchanges (about 10% of total). By 1866, 1896, or 1926, 237, 860, or 1,675 number of stocks traded across the major exchanges according to our databases. For more detail on the history of the U.S. equity market we refer to Online Appendix A.

In order to construct a reliable and historically extensive sample we have compiled the pre-CRSP dataset from several sources. The sample spans the period from January 1866 through December 1926 and is at the monthly frequency, containing stock prices, dividend yields, shares outstanding, and market capitalizations for all major stocks traded on the NYSE, NY Curb, and regional exchanges. We use Global Financial Data (GFD) and the Commercial and Financial Chronicle (CFC, the source used to build the CRSP sample as of 1926), which we combine with risk-free rates from Jeremy Siegel's database. The CFC was a well-read financial newspaper containing daily information on stock prices and other characteristics. The GFD stock database has an extensive coverage of historical stocks traded in the U.S. across the NYSE, NY Curb, and regional exchanges. However, GFD did not include reliable information on the number of shares outstanding, which we hand-collected from the CFC. The CFC dates to 1865, implying our start date of 1866 for this study. The sample includes delisted stocks and as such is believed to be free of a survivorship bias. Our dataset construction and verification procedure are described in extensive detail in Online Appendix A. Tables A.3 to A.6 and Figures A.3 to A.6 in the Online Appendix summarize the stocks included in our sample, the return series, market capitalization, dividend, and share issuance characteristics, as well as the industry and exchange compositions.[9] We combine this data with post-1926 data on equity factor returns from CRSP and Kenneth French's website in Section III.

Even though we (and the data vendors) have paid close attention to data quality, the deep historical data tends to be of lesser quality compared to the more recent data, as digital archives and strong requirements on data processes did not exist. Instead, data was

[9] Note we have a limited number (less than 50 or 100) of stocks in our cross-section for about the first six or twelve years of our sample period, making it more difficult to detect the existence of return factors. Even though the average returns need not necessarily be affected, the variation around the average is probably higher due to limited diversification benefits in the factor portfolios.

maintained typically by exchanges, statistical agencies, newspapers, and investor annuals, often in manual writing. Potential data quality issues that could be at work include (manual) misprints and other measurement errors, but also the use of old data, the use of time-averaged prices over a month (often average of the lowest and highest monthly prices), and the timing of dividends sometimes being unknown but assigned at quarter or year ends.

Lesser data quality could influence our tests in a number of ways. On the one hand it could create random measurement errors in our data, thereby, biasing our results towards the null hypothesis that a return factor does not exist. On the other hand, if biases in the data correlate with factor premia, they could create spurious results. For example, Schwert (1990) shows that the use of average of high and low prices over a month generates an artificial AR(1) process in the return series. Further, measurement errors could cause prices to be spuriously inflated, trigger potential reversal (value) profits.

To construct a high-quality dataset, we have taken the following steps. First, we have checked and corrected each data series on potential data errors as outlined in detail in the Online Appendix, Section A. Second, we have verified a random sample of dividends and stock prices from GFD versus CFC data. Third, we construct market indices which we compare against the GFD U.S. stock indices and indices constructed by Schwert (1990) and Goetzmann, Ibbotson, and Peng (2001) (see Table A.3 and Figure A.3 in the Online Appendix). Fourth, we compare the market value distributions of our sample in 1926 versus the CRSP sample (see Figure A.4 in the Online Appendix). Fifth, we apply a number of conservative screens on our data series and remove data points when they do not pass these screens.

These screens include (i) a 'zero return screen' – leaving out data series with more than one zero or missing price return observations in the past 12 months, (ii) a 'return interpolation screen' – leaving out identical returns one month to month, and (iii) a 'stale return screen' – leaving out observation which do not have nine or more differentiating returns over the past 12 months. The first screen filters for data historically available at a non-monthly frequency and reduced liquidity, as assets with lower liquidity or no trades are more likely to have zero-returns. Lesmond, Ogden, and Trzcinka (1999) show that the number of periods with zero-returns is an efficient proxy for liquidity. The second screen filters an unlikely return pattern, exactly identical consecutive monthly returns, which indicate return interpolation. The third screen filters returns which are not updated at the monthly frequency. To this end, we remove an asset at each point in time when over the past 12-months there are less than nine unique monthly returns when rounded to five basis points. We have simulated that such a pattern is unlikely under a normal distribution and the empirical stock return distribution in our universe.[10] Further, we always skip a month between the momentum signal and investing, which removes possible spurious autocorrelation at the monthly frequency.

Please note that these screens mitigate data quality concerns and allow us to select reliable, better tradable data points. We believe this to be of primary importance, but also like to stress that they could potentially bias factor premium estimates downwards if they remove correct data points. Therefore, in the robustness analysis we examine the impact of these screens on factor premium estimates, finding in general limited effect.

Table I, Panel A provides an overview of the sample, while Online Appendix A, Tables A.4 to A.6 show further detail on the sample composition and impact of the data screens. Overall, we have 241,632 unique firm-month observations with market capitalizations, of which 101,949 satisfy our screening criteria. Our cross-section starts with 54 (54) stocks in 1866 and ends with 407 (607) stocks in 1926 after (before) the data quality screens, respectively.[11] Note that the latter number (607) is higher than the sample of CRSP stocks (482 in January 1926) as CRSP only includes NYSE-listed stocks before 1962 (see also McQuarrie, 2009), thereby missing a substantial number of (mostly smaller capitalization)

[10] More specifically, we have randomly drawn 10,000 observations (with replacement) from the normal distribution with mean and volatility equal to the equal-weighted or value-weighted average stock return over our sample period (see Online Appendix, Table A.1), or from the empirical 1926-2019 CRSP market return distribution, and examined the occurrence of this screen. Under these simulations this screen is triggered for less than 0.5% of the observations.

[11] For comparison, in Goetzmann, Ibbotson, and Peng's (2001) old NYSE dataset the number of firms peaked at 114 in May 1883

stocks from the NY Curb and regional exchanges. However, compared to CRSP we include fewer stocks in our final sample due to the use of our data filters and data quality screens, as we choose to focus on stocks with good data quality.[12] In total our sample includes 1,154 (1,488) unique stocks between 1866 and 1926 after (before) the data quality screens, showing that also delisted firms are included in the sample.

INSERT TABLE I HERE

Further, we classify the stocks into five sectors: (i) financials (mostly bank stocks), (ii) infrastructure (mostly railroad stocks), (iii) energy/mining, (iv) utilities, and (v) industrials & miscellaneous stocks. Infrastructure stocks accounted for approximately 80% of the market capitalization between 1866 and about 1890, after which energy/mining, and industrial stocks gained in importance through a series of new issue booms, becoming of similar importance in terms of market capitalization as infrastructure, see Online Appendix A Table A.5. Banks had a large number of listings, but many traded infrequently and had lower market caps. For example, our sample has over 284 stocks pre-filters (54 post-filters) in the banking industry in 1896, but they only contributed to around 10% of the total market capitalization. The Online Appendix contains a further detail on these numbers.

Table I, Panel B presents (annualized) summary statistics on individual stock returns in our sample. The time-series statistics are computed by first value-weighing returns per month for each firm, and then averaging per decade. The value-weighted market index shows an average annual total return of 8.67% and volatility of 11.80% in this period (this compares with an average return of 11.24% and 18.44% volatility of over the period 1927-2019).[13]

[12] We have also checked the impact of data quality screens on the cross-section of stocks in the CRSP universe. Overall, the data quality screens exclude 9.0% of stock observations from CRSP over the period 1927-1930 (note that we need 12-months of observations to apply our screens), comparable to the 12.7% of stocks dropped from our sample in 1926 (see Table I, Panel A). This number drops to 5.2% over the 1926-1962 period, after which the impact become more marginal. Hence, the data quality screens only significantly impact the earliest years of the CRSP sample.

[13] Note that volatility was significantly lower over the pre-CRSP period, as also shown by Goetzmann, Ibbotson, and Peng (2001), Danielsson, Valenzuela, and Zer (2018), and Baltussen, Swinkels and Van Vliet (2021). A

Further, dividends represent 81% of the average stock return (7.05%), similar to the findings of Acheson, Hickson, Turner, and Ye (2009) for United Kingdom and United States stock markets in the 19th century. For comparison, in the CRSP sample, the dividend returns contributed to 32% of the total returns (3.61% of 11.24%).

III. The cross-section of stock returns: 1866 and 1926

Next, we utilize our novel database and examine the cross-section of stock returns over the 1866-1926 period.

Variables

To avoid conducting a large data dredging exercise, we focus on five prominent characteristics that are well-documented in the literature and can be constructed over our sample; beta, size, value, momentum, and short-term reversal (see for example, Fama and French, 1992, 1993, 2015, 2016, 2018, Frazzini and Pedersen, 2014). As accounting data on balance sheet and income statements generally lacks coverage and uniformity in the U.S. before about 1926 (Cohen, Polk, and Vuolteenaho, 2003, Linnainmaa and Roberts, 2018, Wahal, 2019), we cannot reliably test characteristics that require accounting data, such as profitability or investments.[14]

We measure the characteristics by following as closely as possible the common definition in the literature. More specifically, the market factor is constructed by value-weighting all stock returns by month and subtracting the proxy for the risk-free rate. Size we define as the (log) total market capitalization of a firm, and value by the dividend yield over the past year

potential reason is a higher importance of dividends as compared to capital appreciation returns, but we leave an exact examination into the drivers of time-variation in volatility between pre-CRSP and CRSP samples for future research.

[14] U.S. companies listed on te NYSE were required to publish audited accounting statements as of 1932. The standardization of financial statements increased following the establishment of the SEC in 1934, and specific prescriptions regarding the content and format of financial reports was established by the Committee on Accounting Practices in 1939, and Regulation S-X in 1940. We refer to footnote 5 for further background on historical accounting data.

(i.e., dividends over the past 12-months divided by price). The main advantage of dividend yield is that in the 19th century dividends were widespread, strongly associated with earnings (Braggion and Moore, 2011), and seen as the common important valuation tool for stocks (Dowrie and Fuller, 1950, Poitras, 2010).[15] By contrast, earnings and book values are not well available over our sample, making them ill-suited over the pre-CRSP sample.[16] Momentum is measured by the total return of a stock between months $t - 12$ and $t - 2$, as in Jegadeesh and Titman (1993). We define short-term reversal by the past 1-month return, following Jegadeesh (1991). We construct the beta via a regression of a stock's return on the market's excess return over the past 36 months (minimum of 12).[17]

Fama-MacBeth regressions

We start our analysis by estimating monthly Fama and MacBeth (1973) regressions to estimate premia associated with the above stock characteristics without a need to specify portfolio breakpoints or other degrees of freedom. We value-weight each stock-month observation to prevent our results to be skewed to smaller stocks, especially the many small bank stocks present historically. Moreover, value-weighting is shown to be an effective procedure to mitigate the upward biases in regression estimates arising from noise in stock prices (Asparouhova, Bessembinder and Kalcheva, 2013). Table II contains the results, with average slopes multiplied by 100.

INSERT TABLE II HERE

[15] Dividends are a logical and consistent metric to scale a firm's stock price across industries and through time. Fama and French (1998) use dividend yield besides other yield variables when testing the global value premium. Pätäri and Leivo (2017) give an extensive literature review on all value variables used in asset pricing studies including dividend yield.

[16] Cohen, Polk, and Vuolteenaho (2003) study the historical SEC enforcement records and conclude that post-1936 accounting data is of sufficiently high quality to employ in empirical analysis. Wahal (2019) concludes that data related to income statements starts to be of sufficient quality as of 1938, while data related to book value and total assets is of sufficient quality as of 1926.

[17] In Online Appendix C, Table C.1 we also consider volatility and idiosyncratic volatility as alternatives ways to measure 'low-risk' (see Ang, Hodrick, Xing, and Zhang, 2006, Blitz and Van Vliet, 2007). Volatility (idiosyncratic volatility) is measured by the standard deviation of the excess returns (beta-corrected excess returns) of the last 36 months, requiring a minimum of 12 observations. Further, results are qualitatively similar when using betas estimated over 60-month window, or when applying a Dimson (1979) correction by including 1 or 2 months of lagged returns in the beta estimation.

First, we find a flat relationship between market beta and return, with a slope coefficient close to zero (0.05, t-statistic = 0.56). In other words, the CAPM fails in the cross-section of stock returns over the pre-CRSP sample, similar to the findings of amongst others Fama and French (1992) over a more recent sample. We also observe for size no significant relationship between (log) market capitalization and returns (slope = 0.02, t-statistic = 0.50), in line with using value-weights in the regressions. When using equal weights, we observe a negative slope that is marginally significant (-0.08, t-statistic of -1.76), see Online Appendix Table B.1. Further, dividend yield (our proxy for value) carries a positive slope (2.07, t-statistic = 1.84), in direction similar to the results of book-to-market ratio over the CRSP sample period (e.g., Fama and French, 1992). Momentum has a significantly positive slope (0.88, t-statistic = 2.51), while short-term reversal has a significantly negative slope (-2.52, t-statistic = -2.27), again akin to more recent sample results. The last column of Table II shows these results also hold up in a multivariate setting. In Online Appendix Table B.2, we also report results when including share issuance in the regressions, measured as the 1-year change in shares outstanding following Pontiff and Woodgate (2008). As share issuance was relatively rare pre-CRSP, and hence testing power limited, we choose to not include these results in our main tables. We find that share issuance has a significantly negative slope, and above multivariate regression results remain very similar.[18]

Univariate portfolio sorts

Next, we examine the performance of value-weighted univariate portfolios. At the end of every month, we form quintile portfolios that are rebalanced monthly, as our data series are updated at the end of every month (unlike for example post-1926 accounting data, which is

[18] We include a dummy on zero share issuance stocks and a continuous measure on the remaining stocks, as most stocks did not issue or repurchase shares over our sample (on average 71% of firm-month observations have a zero share issuance). This result aligns with the out-of-sample study by Linnainmaa and Roberts (2018), who show share issuance carries a significant premium between 1926 and 1969. We do not consider share issuance in our portfolio sorts, as for most part of the sample we have at most 25 stocks with non-zero issuance, see Online Appendix A, Figure A.5. Consequently, we have to be cautious to interpret the share issuance results as a falsification or verification of results found on more recent data.

typically available at the annual or quarterly frequency). We form quintile portfolios to balance the spread in characteristics across portfolios and the number of stocks within each portfolio. Note that in the first years of our sample we have about 50 stocks in the cross-section, increasing to over 300 in the last years of our sample. In the robustness section we also consider tercile and decile portfolios, although we like to stress that especially the latter have sizable idiosyncratic risks in the earlier years of our sample. For value we group all non-positive dividend stocks in one portfolio and distribute the remaining stocks equally across the other portfolios in case the breakpoint of the first portfolio equals zero.[19] Table III shows the (annualized) excess returns, as well as intercepts and slopes from the CAPM model for each portfolio, as well as for the top minus bottom portfolios.

INSERT TABLE III HERE

The results generally confirm the Fama-MacBeth regression results. The beta-sorted portfolios carry similar average excess returns, with high beta portfolios not significantly outperforming low beta portfolios (t-statistics = 0.59). Consequently, CAPM alphas are significantly positive for low beta portfolios and significantly negative for high beta portfolios, resulting in a -6.81% (annualized) alpha of the high-minus-low beta portfolio (t-statistic = -3.32). Size-sorted portfolios reveal an insignificantly lower return of -2.83% on larger caps over smaller caps (t-statistic = -1.37), a spread that drops to -0.92% (t-statistic = -0.46) when controlling for the higher beta on small caps. High value stocks significantly outperform low value stocks by 5.61% per annum (t-statistic = 2.41). As no-dividend paying firms typically have more volatile stocks with high market betas (see Fama and French, 1993, for similar effect over the more recent period)[20], the CAPM alpha increases to 10.13% (t-statistic = 5.49). Similarly, winner stocks outperform loser stocks by 8.18% per annum (t-statistic = 2.77). As

[19] Note that sometimes at most a handful of stocks have negative dividends, see Online Appendix A.
[20] Zero/low dividend paying stocks have a beta of 1.99 compared to 1.04 for high dividend stocks. For comparison, Fama and French (1993) report a beta of 1.45 for zero-dividend firms and 0.73 for the quintile of firms that pay the highest dividends.

losers typically had a higher beta than winners, the resulting CAPM alpha is 11.53% (t-statistic = 4.16). Post 1-month winners underperform past 1-month losers by 5.31% (t=statistic = -1.93), a spread that becomes insignificant once controlling for market beta (CAPM alpha = -3.26%, t-statistic = 1.21).

2x3 factor portfolios

The above analysis reveals significant differences in the cross-section of stock returns based on characteristics. Next, we follow Fama and French (1993) and construct standard 2x3 portfolios sorted on size and a characteristic. To construct the 2x3 portfolios, every month all stocks in our database are classified as either large or small, using the median cross-sectional market capitalization as breakpoint.[21] Next stocks are sorted on their factor variable within both size groups and split in three portfolios (Low, Medium, High) based on the 30% and 70% percentiles. High always refers to the favorable factor characteristic, being low beta, high dividend yield, high momentum, or low past 1-month return in case of short-term reversal. The exception for this formation is for value, as at most points in time, at least 30% of the smaller capitalization stocks have a 0% dividend yield (and on average less than 1% have a negative dividend yield), see Online Appendix A, Figure A.4. In these cases, stocks with a non-positive dividend yield are assigned to the Low portfolio. The remainder of the stocks are then assigned to the Medium and High portfolios, based on the 50% percentile of the stocks that have a 12-months' dividend yield above 0%. The final factor is created by taking a fifty-fifty long position in large-cap and small-cap High stocks, combined with a fifty-fifty short position in large-cap and small-cap Low stocks. Note that the above procedure differs from Fama and French (1993) by replacing independent sorts by dependent sorts, as the former sometimes produces empty portfolios, especially in the earlier part of our sample. Soebhag, Van Vliet and Verwijmeren (2022) show a limited impact of this choice over the CRSP sample.

[21] Note that we deviate here from the common practice in the asset pricing literature by not NYSE-only based breakpoints, as stocks traded significantly on multiple exchanges, including regional exchanges and the Curb, see Section III and Online Appendix A, Table A.6.

The SMB factor is subsequently constructed by taking the difference, every month, between the simple average of the three small portfolios and the simple average of the three big portfolios across the value sorts (in spirit to Fama and French, 1993, who use book-to-market sorted portfolios). Further, we lever the top-bottom beta portfolio in order to make it market-neutral by levering the long (low beta) leg up and the short (high beta) leg down to a market beta of 1.[22] For simplicity, market betas are estimated full-sample against the market portfolio, but we note that results do not change materially when a 36-months rolling-window estimate is used instead. Estimated betas are floored at 0.5 and capped at 2.0 to limit the effect of estimation noise (we do like to note that this choice does not alter our conclusions but makes the factor more conservative). The 30-day T-bill rate is taken as borrowing and savings rate. This beta-adjustment is in spirit of Frazzini and Pedersen's (2014) Betting-Against-Beta (BAB) factor, while we circumvent the issue raised by Novy-Marx and Velikov (2021) that the size of the BAB stock factor premium is heavily influenced by a large weight to micro-cap stocks.[23] Although the number of stocks is rather limited in the beginning of our sample, adding noise to the portfolios, the above procedure ensures that data is available for every month-portfolio combination.

INSERT TABLE IV HERE

Table IV shows the excess returns, volatilities, t-statistics, CAPM alphas and betas, and t-statistics of the alphas for the 2x3 sorted portfolios. Figure A.7 in the Online Appendix shows the cumulative return of each factor series. The naming convention for the portfolios follows Fama and French, with BETA representing the market-neutral low-beta minus high-beta portfolio. The results generally confirm the results of the Fama-MacBeth regressions and

[22] Note that especially value and momentum also have strong beta differences across the long and short legs, as loser stocks and zero-dividend payers have substantially larger betas. Although these beta spreads are stronger over the 1866-1926 sample compared to the post-1926 sample, we choose to follow common practice and do not lever these factors to hedge out beta exposures.

[23] Novy-Marx and Velikov (2022) argue that the weighting scheme used by Frazzini and Pedersen (2014) biases portfolio weights to equal-weighting, which gives relatively large weights to economically less relevant micro-cap stocks.

univariate portfolio sorts. SMB shows an insignificant premium of 1.17% per annum (t-statistic = 1.15). When correcting for the higher beta of smaller caps the alpha spread becomes an insignificant 1.11% (t-statistic = 1.09). Hence, the size factor is not significantly priced in the cross-section of stock returns pre-1926.[24]

Value (HML) now shows an insignificant premium of 2.76% per annum (t-statistic = -1.40), with the effect being more present in larger caps (3.10%, t-statistic = 1.84). However, as the low dividend stocks have substantially higher betas (the HML beta spread is -0.91, driven especially by small non-dividend payers), the CAPM alpha equals 7.11% per annum, highly significant with a t-statistic of 5.04. In other words, the value factor premium is sizable when controlling for beta exposures. In the subsequent robustness section, we show the robustness of this effect.

Momentum (UMD) shows a sizable and significant average return of 6.13% per annum (t-statistic = -2.76). When controlling for the higher beta on loser stocks (most notably of the small stocks), this spread increases to 9.02% per annum (t-statistic = 4.42). Short-term reversal (ST_REV) displays a significant average return of 4.10% per annum (t-statistic = -1.98). However, when controlling for the higher beta on loser stocks (1-month loser stocks have a 0.33 higher beta than 1-month winners), this spread becomes insignificant (2.54%, t-statistic = 1.25), in line with the univariate portfolio sort results. Finally, BETA shows a sizable and significant premium of 6.63% per annum (t-statistic = 4.16), as low beta stocks offer a similar return as high beta stocks, and hence higher beta-corrected returns. Further, the beta of the BETA portfolio is negative (-0.26), further increasing the CAPM alpha to 7.86%

[24] Hou and van Dijk (2019) show that changes in profitability of small versus large stocks explain the (dis)appearance of the size premium over time. Since we lack data on profitability, we cannot corroborate this conjecture over our sample.

(t-statistic = 5.05), indicating that our procedure for levering the low and high beta portfolio is on the conservative side.[25]

The above results generally show up in both small cap stocks and large stocks. Several studies reveal that average returns on factor portfolios tend to be larger in the small-cap space than in the large-cap space (e.g., Fama and French, 1992, 1993, 2012, 2015, Israel and Moskowitz, 2013). Focusing on CAPM alphas of the long-short factor portfolios we find a higher factor premium amongst small stocks for all four factors, although economic differences with larger cap stocks are limited, being below 1% for all factors.

Robustness to methodological variations and data filters

Next, we examine the robustness of the above portfolio sort findings for common variation in the sorting or portfolio construction procedure. Robustness of portfolio sort results across testing choices is an additional manner to limit the influence of p-hacking. We consider the following variations: univariate sorted tercile and decile portfolios (although idiosyncratic risk in these portfolios tend to be high due to a limited number of stocks per portfolio in especially the early half of our sample), 2x5 size-characteristic sorted portfolios, or 2x3 size-characteristic sorted portfolios that are either equally weighted or sector-neutral by ranking within each sector. Panel A of Table V summarizes the results by means of the top-bottom return spreads and panel B shows CAPM alphas (note that we now also lever the univariate sorted beta long-short portfolios towards market neutrality).

INSERT TABLE V HERE

Overall, we find similar results as in Tables III and IV. The value factor premium is sizable in univariate sorts and when controlling for beta exposures, while momentum and

[25] We can attribute this to the use of capping the estimated betas between 0.5 and 2.0 to prevent overleveraging to extremely estimated betas.

BETA are sizable and significant across all variations. Noteworthy exceptions are a significant size premium in returns spreads and CAPM alphas when equally weighting stocks, and a significant short-term reversal premium in more extreme portfolios (decile or 2x5 sorted portfolios), equally weighted portfolios, and sector-neutral portfolios. However, note that we have many smaller stocks (especially banks) in our sample that get relatively large weight when equally weighting.

Further, to build a high-quality dataset we have applied several data filters. We do acknowledge that these filters might bias factor premium estimates, as less-liquid stocks are excluded from our sample. Next, we assess the robustness of our results to the data filters. The results are summarized in Online Appendix B, Table B.3. First, we test robustness with respect to outliers (for example due to measurement errors) by trimming asset returns at -50% and + 50%. Second, we apply only the zero-return screen and drop the two other data quality screens. Third, we loosen the zero-return screen by leaving out data series with more than three zero return observations in the past 12 months or dropping it altogether. Note that these alternatives will also allow for less liquid stocks to enter the sample. Results are generally similar to the baseline, with short-term reversal and size being significant without the zero-return screen, possibly due to more illiquid stocks entering the sample.

Finally, we include a one-month implementation lag on the characteristics, which removes any impact from the use of the average of high and low prices over a month on momentum, and spuriously inflated prices on value. Note that this lag is on top of the 1-month lag for momentum in the baseline results. Value, momentum, and BETA factor premia remain significant also in this test, size remains insignificant, and short-term reversal (expectedly) drops substantially in returns spread and CAPM alpha. Overall, we conclude that value, momentum, and (low-)BETA equity factor premia are robust to the methodological variations and sample choices.

Spanning tests

Next, we run spanning regressions of each 2x3 long-short factor portfolio on all other factors to examine factor redundancy. Table VI shows the results. SMB has a positive, but insignificant intercept (1.49%, t-statistic = 1.49), akin to the Fama-MacBeth regression results and portfolio sorts. HML has a significantly positive intercept (3.89%, t-statistic = 3.02), with significant negative correlation to the market (as seen above) and also SMB, but positive correlation with UMD and BETA (as high dividend stocks typically also have lower beta). The positive correlation with momentum is due to the high historical relevance of dividends in returns, and hence the momentum measure. In Online Appendix Table C.1, we confirm a significant momentum effect when sorting purely on price momentum (hence ignoring dividend returns in the momentum measure). UMD has a significantly positive intercept (6.39%, t-statistic = 3.19), with significant negative correlation to market, SMB, and ST_REV (akin to results over the CRSP sample). Similarly, BETA has a significantly positive intercept (3.92%, t-statistic = 2.75), despite significant positive correlation to all other factors except short-term reversal. Finally, ST_REV now becomes significantly positive (4.96%, t-statistic = 2.41), as the spanning regression control for the significantly negative loading on UMD and BETA. Overall, we find momentum, value, short-term reversal and (low-)BETA to be non-redundant asset pricing factors.

INSERT TABLE VI HERE

Is there a size premium?

To interpret our results on the size premium, we note a nuanced pattern: a marginally significant coefficient appears in equal-weighted regressions, and a significant size premium emerges in equal-weighted portfolios. However, this premium is not present in our baseline value-weighted portfolio sorts. Similarly, Davis, Fama, and French (2000) and Linnainmaa and Roberts (2018) fail to detect a significant size premium between 1926 and 1963. Interestingly, a significant size premium reappears once we remove the liquidity screen in the value-weighted portfolios, a screen put in place to improve data quality in the pre-CRSP sample (note that these screens have virtually no impact in the CRSP sample). In other words, we find evidence of a significant size premium once the smallest stocks get substantial weight, or when a significant number of illiquid stocks with poorer data quality are included. Yet, as the data quality for these stocks may not be as high, we believe a conservative interpretation of these results is warranted. Consequently, we are cautious about drawing a positive conclusion on the size premium, especially outside the smallest (micro) stocks.

IV. Out-of-sample decay

Several studies reveal evidence of substantial out-of-sample decay of stock factor premia. McLean and Pontiff (2016) show that the performance of trading strategies declines after the publication of research papers that document their discovery. Linnainmaa and Roberts (2018) consider the performance of accounting-based equity anomalies in the period before and after discovery and find a substantial weaker out-of-sample performance for both subsamples. This raises the question how the estimated premia over the post-1926 CRSP-era compared to premia over the 1866-1926 pre-CRSP sample? To study out-of-sample decay, we measure the performance of the 2x3 sorted high-low portfolios over the 1866-1926 'pre-CRSP' and 1927-2019 'CRSP' sample periods and examine returns spreads and CAPM alphas. To this end we reconstruct the 2x3 value-weighted portfolios over the CRSP era (skipping the data quality filters, as this is uncommon for the CRSP data and the CRSP sample is already of good quality).[26] Table VII contains the resulting average top-bottom returns spreads (Panel A) and CAPM alphas (Panel B) of the individual factors and their equally-weighted average, while Figure 1 in the introduction depicts the results.

INSERT TABLE VII HERE

Return spreads and CAPM alphas are generally of similar size over the pre-CRSP and CRSP samples, being not significantly different for most characteristic-sorted portfolios. The exception is ST_REV, having a significantly lower return spread and CAPM alpha over the 1866-1926 period. One explanation for a short-term reversal premium is liquidity provision (e.g., Nagel, 2012), which is seemingly at odds with this finding given the higher illiquidity over the pre-CRSP sample.[27] SMB has an insignificant CAPM alpha over both periods, while HML, UMD and BETA all have significant CAPM alphas over both periods. Return spreads (CAPM alphas) average 4.16% (5.53%) over the pre-CRSP sample and 5.17% (5.92%) over the CRSP sample period, hence differing by an insignificant 1.01% (0.39%). Hence, overall, we find no significant evidence of an out-of-sample decay in stock factor performance. Interestingly, these findings align in size with a 26% out-of-sample decay observed by McLean and Pontiff (2016) over 5-years of post- sample periods[28], but contrast with the results of Linnainmaa and Roberts (2018) on accounting-based anomalies over the 1938-1963 period.

[26] We like to note that factor premia over the CRSP 1927-2019 sample are similar in sign and of about equal size when using the portfolios as published on Kenneth French's data library, with size, dividend and short-term reversal average return spreads differing by less than 40 bps per annum, momentum differing by 85 bps and BETA having a 213 bps lower average return spread in our calculations. Note that the most important differences with Kenneth French are the use of dependent sorts in our calculations (instead of independent sorts) and the inclusion of zero or negative dividend paying stocks in our dividend-sorted portfolios.

[27] An alternative explanation is the presence of indexed investors (Baltussen, Da and Van Bekkum, 2019), investors less prominent during the pre-CRSP sample.

[28] By contrast, they find a post-publication performance decay of 58% over on average 13 years of post-publication data, which they attribute to arbitrage trading. Jacobs and Muller (2020) find that the United States is the only country with a reliable post-publication decline, while the effect is generally not present in international markets. Further, Jensen, Kelly, and Pederson (2021) argue that the observed post-publication decline in anomaly profits could be a statistical artifact of multiple sequential tests from the same data generating process.

Baltussen, Swinkels and Van Vliet (2021) also observe no significant out-of-sample decay for global factor premia over 150 years of out-of-sample data.

To maximize testing power, we also compute the full sample (1866-2019) results, as presented in the last rows of Panel A and Panel B of Table VII. These results confirm the results above, with an insignificant CAPM alpha on SMB of 1.07% (t-statistic = 1.36), and significant CAPM alphas varying between 5.57% (t-statistic = 6.09) for HML and 10.03% (t-statistic = 7.88) for UMD. On average, factor premia are around 5% per annum and highly significant (4.77% return spread, t-statistic = 9.72, 5.67% CAPM alpha, t-statistic = 12.38).

Finally, we examine the correlations of the stock factor premia amongst each other over the pre-CRSP and CRSP sample, as data mining could affect the entire return process, including the correlations amongst anomalies. McLean and Pontiff (2016) and Linnainmaa and Roberts (2018) show that correlations amongst anomalies tend to increase out-of-sample (being either on the 'post-discovery' sample of an anomaly or on the 'pre-discovery' sample). To this end we regress the return spread on each factor series on a constant, the market factor, the average return on all other factor series, and interact these regressors with a dummy for the pre-CRSP sample period. Panel C of Table VII presents the resulting coefficients on the average return on all other factor series and its change across the two samples. Correlations with the other factors do not change significantly for most factors, except for a significant increase for momentum and a significant decrease for short-term reversal. To maximize testing power, we also run a panel regression across all anomalies, following McLean and Pontiff (2016) and Linnainmaa and Roberts (2018), where we cluster standard errors by calendar month and factor to account for correlated errors in the panel. The results show that the coefficients are not significantly different from zero for both the pre-CRSP and CRSP sample, and do not significantly change across both samples. In other words, at odds with a data-mining based explanation we find that stock factor premia do not have different correlations out-of-sample. Overall, we fail to find significant out-of-sample decay in stock factor premia over a 61-years out-of-sample period, leading us to conclude that the value, momentum, low-risk factors are unlikely caused by data snooping.

V. Investment frictions and historical investability of equity factors

Our results show that equity factor premia have robustly existed in 61 years of independent out-of-sample data. At this point, we like to note that the main purpose of this paper is to examine the pricing of several key characteristics in the cross-section of stocks in an economically important out-of-sample period, and thereby provide robust and rigorous long-term evidence on the main factors driving stock returns. A related question is to which extent the documented equity factor premia can be attributed to investment frictions? Most asset pricing models assume frictionless markets, while in reality investors face investments restrictions, leverage constraints, practical or legal boundaries to shorting, and transaction costs. This assumption of frictionless trading has been challenged in the literature, especially for stock-level factor premia which require high amounts of trading in illiquid stocks. For example, Korajczyk and Sadka (2004) examine the impact of frictions on momentum and Avramov, Chordia and Goyal (2006) on short-term reversal. By contrast, Novy-Marx and Velikov (2015) show that simple trade rules are effective cost mitigation techniques, and most anomalies remain significant after transaction costs.

It is commonly assumed that investment frictions were higher in the 19th century than in the 21st century. Although data to assess the exact impact of investment frictions is to the best of our knowledge not available, indications exist that it was not impossible nor extremely expensive to trade in the markets we examine. Several studies, summarized in Online Appendix A, indicate that the U.S. stock market was well-developed, active trading (including shorting) took place in stocks, and trading seemed feasible at limited transaction costs with trading costs in the 19th century not very much different from the 20th century (e.g., Brown et al., 2008). Jones (2002) reports spread estimates for Dow Jones stocks of about 0.5% since

1900, not much different from CRSP-era estimates up to round 1980, and annual share turnover on NYSE stocks being higher between 1900 and 1926 than in 2000.

Although the above suggests that investors could have profited in practice from exposure to the equity factors, our results do not necessarily imply that the factor premia could have been profitably exploited. This study does not examine smarter and possibly better definitions, smart trade rules, nor aspects linked to (limits to) arbitrage and tradability (such as transaction costs, turnover, legal controls, etc.). For example, the use of liquid stocks, introducing smart trade rules, and integrating of multiple factors can all reduce implementation costs significantly (see Novy-Marx and Velikov, 2015). Further, investors do not need to have universal and frictionless access to markets in order to profit from equity factor premia. For example, even a long-only investor with access to a limited number of markets could postpone the buying of a stock, if the particular stock was negative on momentum or overvalued. In other words, investors could have profited from equity factor premia with varying degrees. We leave a thorough assessment of positive factor returns after costs and frictions, or the design of an efficient factor investment strategy for future research.

VI. Machine learning in the cross-section of stock returns

So far, we have examined the cross-section of stock returns using prominent characteristics and traditional techniques that model returns as a linear function of characteristics. Interestingly, several recent studies show the great promise of machine learning models using dozens of characteristics with non-linear interactions for understanding the cross-section of stock returns. Most notably, Gu, Kelly, and Xiu (2020; henceforth GKX) find that several machine learning models can well-predict cross-sectional differences in U.S. stock returns over the period 1957-2016, with the best performing methods being random forest and neural networks that allow for nonlinear predictor interactions. Leippold, Wang and Zhou (2021) find similar results for the Chinese stock market - world's

2nd stock market in terms of market capitalization - between 2000 and 2020. Machine learning methods are a powerful tool to summarize key predictor variables, including its interactions, in a data-driven manner, and hence uncover priced variables over and above the prominent characteristics without conducting a sizable data-dredging exercise. At the same time, also machine learning models require out-of-sample testing in independent samples, similar to canonical factor models, a challenge we pick up next.

To this end, we apply the most promising machine learning techniques to a wide range of variables we can construct over the pre-CRSP sample. In our tests we largely follow GKX. We utilize the predictors used in their study that can be (reliably) constructed over our sample; dividend yield, 1-month, 6-months, 12-months and 36-months momentum, change in 12-months momentum, (the natural logarithm of) firm size, one-year changes in shares outstanding, beta, beta squared, 36-months total and 12-months idiosyncratic return volatility.[29] Note that this expands on the stock characteristics tested so far, as machine learning methods determine the best (linear or non-linear) combination of characteristics based on validation sample forecast accuracy. In addition, we include the five industry dummies. For comparison, GKX use 94 characteristics (mainly their accounting and daily data related characteristics we cannot include), interaction of each characteristic with market or macroeconomic timeseries variables, and 74 industry sector dummy variables. As machine learning methods generally benefit from a bigger variable set, this likely constrains the opportunity of the machine learning methods in our tests compared to GKX. Nevertheless, machine learning methods should be a powerful tool to identify the key variables that price the cross-section of stock prices.

[29] Following GKX we cross-sectionally rank all stock characteristics period-by-period and map these ranks into the [-1,1] interval, and replace missing characteristics for each stock with the cross-sectional median at each month. Compared to GKX we do not include three variables computable over our dataset; industry momentum and industry-adjusted size, as we include industry dummies, and the percentage of zero trading days, as this variable is not available at daily frequency over the pre-CRSP sample but is employed as data quality filter at the monthly frequency.

To limit the number of tests (and degrees of freedom), we focus on two machine learnings methods: random forests (RF) and neural networks (NN) with 3 hidden layers, as GKX and Leippold, Wang and Zhou (2021) show these tend to be the superior models for predicting stock returns in the cross section. We largely follow GKX in applying RF and NN; conditional expected returns are modelled using the same form over time and across stocks, and do not directly use information from history prior to t or from other stocks than the i^{th}. We use the hyperparameters as reported in Table D.1, a binary cross-entropy prediction evaluation function, early stopping, learning rate shrinkage algorithm, batch normalization, and multiple random seeds in the NN. We split our sample in training, validation and testing samples based on a recursive window. Our training and validation sample is split in a 75-25 ratio, initially starting with a 20-year window. Recursively increasing the training sample, periodically refitting the entire model once per year, and making out-of-sample predictions using the same fitted model over the subsequent year. Each time we refit, we increase the training and validation sample by a year, while maintaining a fixed size rolling sample for validation to tune the parameters. Akin to GKX we choose to not cross-validate in order to maintain the temporal ordering of the data for prediction.[30] Based on the above method we obtain predicted likelihoods of outperformance for month $t+1$ for each stock at the end of month t, which we sort in ascending order at the end of month t and transform into value-weighted quintile portfolios that are held till next month end. We deviate from GKX, who form decile ports, as we have fewer number of stocks in the cross-section. Finally, we construct a zero-net-investment portfolio that buys the stocks with the highest expected return (Q5) and sells the stocks with the lowest expected return (Q1).

[30] More specifically, we divide our 61 years of data into 20 years of initial training sample (1866 - 1885), and 10 years of initial validation sample (1886 - 1895), while using the remaining 31 years (1896 - 1926) for out-of-sample testing. We refit models once per year at year ends to limit computational burden. Hence, each time we refit, we increase the training sample by one year, while rolling the 10-years validation sample forward to include the most recent twelve months.

Table VIII summarizes the results. Shown are the average (annualized) return, Sharpe ratio, and CAPM alpha of the value-weighted quintile and Q5-Q1 portfolios. Akin to GKX we find machine learning models predict cross-sectional differences in U.S. stock returns. The Q5-Q1 return spread for RF is positive but insignificant (3.34%, t-statistic = 1.00), but the CAPM alpha is significantly positive (9.78%, t-statistic = 4.26) as RF tends to select lower-risk stocks. Similarly, NN outperforms RF with an insignificant, but higher return spread (5.05%, t-statistic = 1.58) and a highly significant CAPM alpha of 10.62% (t-statistic = 4.42). These findings align with those of GKX and Leippold, Wang and Zhou (2021) that neural networks tend to be the better machine learning models in the US CRSP sample and the Chinese stock market.[31]

INSERT TABLE VIII HERE

Next, we explore the importance of the characteristics and their interactions selected by the machine learning models. To this end, Figure D.1 shows the variable importance for RF and NN models by tracing the marginal relationships between expected returns and each characteristic. We normalize variable importance within a model to sum to one, giving them the interpretation of relative importance for that particular model. Interestingly, machine learning models are able to select many of the factor measures analyzed in the previous section, but then in a fully data-driven approach with little a priori guidance on the variables to select. Dominant predictive signals include dividend yield, followed by (variations on) momentum variables, beta or other risk variables, and market capitalization. The findings on dividend yield, a characteristic which is mainly important in the RF application, momentum,

[31] GKX also show machine learning models perform better for large stocks relative to small stocks, for annual as opposed to monthly prediction horizons, and NN with shallow learning outperforms deep learning setups. Further, RF and NN also help predicting returns on the market portfolio and (to a lesser extent) various factor portfolios. Leippold, Wang and Zhou (2021) confirm these findings for Chinese stock market. We leave further out-of-sample testing of these findings to future work.

size and risk variables align with GKX.[32] Interestingly, the machine learning models yield largely comparable results out-of-sample over the pre-CRSP period as reported by GKX over the CRSP sample. Finally, we examine the added value of machine learning models over the five prominent characteristics by means of spanning regressions of the Q5-Q1 RF or NN portfolios. The results in Panel B of Table VIII show that portfolios sorted on the machine learning methods load significantly and in the expected direction on the canonical equity factors but yield no significant added value over and above.[33]

VII. Economic explanations

We have documented robust evidence for the pricing of key equity characteristics over the pre-CRSP sample that contains 61 years of out-of-sample data. Next, a natural question is what drives the document returns? Although a full answer to this question is beyond the scope of this paper, the pre-CRSP sample allows for novel insights into economic explanations. To this end, we explore time-series variation in the factor premia since 1866 and consider the role of macroeconomic risks, delegated asset management, crash risk, and downside risk.[34]

A. Macroeconomic risks

The 1866-1926 period is characterized by large macroeconomic shocks and market fluctuations, providing out-of-sample insights into macroeconomic risk explanations of stock factor premia. For example, Asness, Moskowitz and Pedersen (2013) find that value and momentum premia link to macroeconomic risks. On the other hand, Griffin, Ji, and Martin

[32] Leippold, Wang and Zhou (2021) also uncover some differences compared to GKX in variable importance for small versus large stocks, and monthly versus annual return forecasting horizons, which they attribute to larger retail trader base, the large presence of state-owned enterprises, and higher investment frictions in China.

[33] We have also run a separate set of machine learning models that include five macroeconomic variables used in GKX that can be reliably constructed over our sample; the stock market's 36-months variance based on our market return series, the market's dividend-to-price ratio, earnings-to-price ratio, term spread, and inflation, all taken from Baltussen, Swinkels and Van Vliet (2021). Results are comparable when including these macroeconomic predictors and are omitted from the paper for sake of brevity.

[34] Another explanation offered for several of the stock factors is market or funding liquidity risk (see for example Asness, Moskowitz and Pedersen, 2013). Due to the limited availability of deep historical data on the measures used in these studies we choose to not examine such explanations in this paper.

(2003) find no evidence of a relationship between macroeconomic risk and momentum returns. To examine whether macroeconomic risks explain stock-level factor premia we follow Griffin, Ji, and Martin (2003) and examine exposures to, and unconditional pricing of, macroeconomic factors, in the spirit of Chen, Roll and Ross (1986). To this end, we construct the most widely used Chen, Roll and Ross (1986) factors – log changes in industrial production (MP; as in Chen et al. led by 1 month), term spread (UTS), changes in expected inflation (DEI), and unexpected inflation (UI) – for our sample using monthly data.[35] We regress the time series of each stock factor on these macroeconomic variables and obtain coefficients and intercepts. Our sample starts in February 1875 due to the availability of historical U.S. inflation data at the monthly frequency. Table VIII summarizes the results, where we show results for the pre-CRSP period (1875-1926), the CRSP sample period (1927-2019) and the full sample period (1875-2019).

INSERT TABLE IX HERE

If factor premia are driven by macroeconomic risk, then they should exhibit significant sensitivity to the factors proposed by Chen, Roll and Ross (1986). Our findings reveal that the global macroeconomic variables are mostly not significantly related to equity factor returns, or otherwise not consistent over subsamples or subject to the wrong sign.[36] Moreover, the significant stock factors of Section III have positive intercepts that are highly significant and are of similar magnitude to the raw returns over this sample (reported in the column "Actual"). These results suggest that macroeconomic risks have very limited explanatory power for stock factor premia.

[35] We collect our data from the FRED database (https://fred.stlouisfed.org/), and before existence of each series in FRED spline with data from Baltussen, Swinkels and Van Vliet (2021). Akin to Griffin, Ji, and Martin (2003), we omit the default premium, as its historical data availability is limited.

[36] More specifically, value and momentum tend to load positively on MP over the full sample period, but insignificantly so over subsamples. Size and short-term reversal tends to load positively on UTS over the full sample and CRSP sample, and dividend and BETA tend to load negatively on DEI over the same periods.

Next, to examine risk premia attached to each macroeconomic factor and to what extent they can explain factor premia, we apply the Fama and MacBeth (1973) technique on a monthly frequency with stock factors as test assets. We combine the premia with the estimated loadings to decompose the returns on the stock factors into predicted and unexplained components. If the Chen, Roll and Ross factors suffice for explaining stock factor premia, then the difference between the actual and predicted returns (or unexplained) should not be significantly different from zero. The empirical results confirm that none of the stock factors have a significant expected macroeconomic premium, and factor premia are of similar magnitude and significance when controlling for macroeconomic exposures as compared to the raw returns.[37] Overall, this leaves us to conclude that macroeconomic risks do not materially explain stock factor premia.

B. The role of delegated management

Vayanos and Woolley (2013) offer a model of momentum and value premia that originates due to delegated management and cashflows to investment funds. Flows are triggered by changes in fund managers' efficiency, which investors can infer from past performance. Momentum arises because flows exhibit inertia and rational prices underreact to expected future flows. Eventually push prices away from fundamentals causing a value premium. According to this theory, when a delegated management structure is absent, momentum and value premia should be relatively weak. Further, delegated asset management can lead to sub-optimal capital allocation due to misalignment of interest between asset owner and asset manager. Baker, Bradley and Wurgler (2011) show that delegated asset management leads to a flatter risk-return relationship giving rise to a low-risk premium.

[37] An alternative approach to assessing the role of macroeconomic risks is to divide the sample in 'good' and 'bad states' and evaluate factor returns across these states. Online Appendix Table C.2. contains the results for two state indicators: recessions versus expansion, or 12-month equity bear versus equity bull markets. Overall, factor premia vary to a limited extent across economic states but are significantly present in both good and bad states, and typically stronger in the 'good' states of the world.

Interestingly, delegated management was notably absent over the pre-CRSP sample, with only a small number of (typically closed-end) equity mutual funds being available to U.S. investors before 1926. Consequently, our findings of significant value, momentum, and low-risk premia over the pre-CRSP sample presents a clear challenge to any theory solely based on delegated management structures.[38]

C. Crash risk and Momentum

Several studies argue that momentum is exposed to risk on extreme losses (Barroso and Santa-Clara, 2015, Daniel and Moskowitz, 2016). Consequently, momentum might be explained by a peso problem or a compensation for exposure to infrequent crashes. Contradicting evidence to this explanation is provided by Goetzmann and Huang (2018), who show momentum crashes are not present out-of-sample in the imperial Russia stock market. The pre-CRSP sample allows for 61 years of out-of-sample insights from U.S. stocks.

To examine momentum crashes we first examine the distribution of monthly returns for the momentum factor return series. Panel A of Table C.3 in the Online Appendix shows the results for the pre-CRSP, CRSP and combined sample periods. Momentum returns are left skewed and displays excess kurtosis over the 1927-2019 period, but lesser so over the pre-CRSP sample. Nevertheless, the minimum monthly return is sizable with -34.85%. In other words, momentum also displays significant crash risk pre-CRSP. Further, Daniel and Moskowitz (2016) argue that momentum returns resemble a short call option on the market, especially following multi-year market drawdowns. We confirm this payoff pattern over the CRSP and pre-CRSP eras by regressing UMD returns on a past 2-years bear market indicator, the market, its interaction, and its interaction with a contemporaneous up-market indicator, as shown in Panel B of Table C.3. Coefficient estimates over the pre-CRSP and CRSP samples

[38] Note that this finding does not preclude delegated management to drive part of value, momentum, and low-risk factor returns over the CRSP sample.

are fairly similar and show that momentum has well negative betas after bear markets and its returns detract especially in months that markets rebounded subsequently.

Finally, Barroso and Santa-Clara (2015) show that limiting exposure to crash risk substantially improves momentum returns. Panel C of Table C.3 in the Online Appendix report the results of volatility-scaled momentum ('UMD*'; based on volatility over the past 12-months), in spirit of Barroso and Santa-Clara (2015). In line with CRSP-era results, volatility scaled momentum has a materially higher Sharpe ratio and more normal return distribution. Overall, these results confirm that crashes may be an inherent feature of the momentum trade but at the same time an unlikely explanation for its existence.[39]

D. Downside risk

A large and growing literature considers whether stock factor premia compensate investors for downside risk. In this subsection, we consider downside risk explanations via the Downside Risk CAPM (DR CAPM) of Hogan and Warren (1974) and Bawa and Lindenberg (1977). When returns are non-normally distributed and investors display an aversion to lower partial moments such as semi-variance this model will outperform the classical mean-variance CAPM. Within the DR CAPM framework, assets with higher downside betas should have higher expected returns. Another way to interpret downside risk is as a conditional risk factor based on falling markets, also referred to as downstate beta.

A common challenge when estimating downside risk exposures and premia is the general reduction in the number of observations in the left tail. Market crashes do not happen very often. For example, Lettau, Maggiori, and Weber (2014) use a threshold of -1 standard deviation of the equity market return, which results in 55 monthly observations out of 435 in their 1974–2010 sample. To maximize testing power, we combine the pre-CRSP and CRSP samples to have 154 years of data, having many more downside events and hence allowing us

[39] Further, as institutional features such as delegated management was mostly absent pre-CRSP these results suggest they are not causing momentum crashes.

to rigorously examine the hypothesis that downside risk explains factor premia. Our sample includes 185 monthly market states with returns below -1 standard deviation of the equity market return. This relatively large number of observations also enables us to study downside risk even further into the left tail of the distribution. As such, we focus on downside betas with different market return thresholds; a zero-return cutoff ('zero'), a -1 standard deviation cutoff ('1 sigma'), and -2 standard deviations cutoff ('2 sigma').[40]

Table C.4 in the Online Appendix summarizes the results. Shown are the CAPM and DR CAPM betas, differences in beta, alphas, and its t-values per equity factor. The overall picture is that downside risk explains at best a small portion of factor premia. The average downside beta is similar to the regular beta, with differences not larger than 0.20. This leads to DR CAPM alphas that differ mostly 1% from CAPM alphas. Consequently, for all factors except size alphas remain economically and statistically significant. Based on these long-run sample results we conclude that downside market risk does not materially explain stock factor premia.

VIII. Conclusion

We examine the cross-section of stock returns out-of-sample using a newly created database of U.S. stocks between 1866 and 1926. Over this 'pre-CRSP' era, the relationship between market beta and returns is flat, and momentum, value, and low-risk premia are sizable and significant. The size premium is generally insignificant unless the smallest, least liquid, and lowest data quality stocks are included in the sample. We find no evidence that cross-sectional equity factor premia materially decay out-of-sample. Furthermore, recent machine learnings models are successful out-of-sample by selecting the key equity factors. Overall, we conclude that equity factor premia are robust and persuasive empirical

[40] We define downside beta as: $\beta_{down,i} = \frac{E[R_m R_i | R_m \leq threshold]}{E[R_m^2 | R_m \leq threshold]}$.

phenomena, also out-of-sample. Finally, exploring new time-variation in factor premia over 154 years of data, obtained by combining pre-CRSP and CRSP samples, we find these cross-sectional factor premia are hard to align with explanations based on macroeconomic risks, delegated management, crash, or downside risks.

References

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2006). The cross-section of volatility and expected returns. *Journal of Finance*, 61(1), 259-299.

Asness, C. S. (1997). The interaction of value and momentum strategies. *Financial Analysts Journal*, 53(2), 29-36.

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

Asparouhova, E., Bessembinder, H., & Kalcheva, I. (2013). Noisy prices and inference regarding returns. *Journal of Finance*, 68(2), 665-714.

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

Baker, M., Bradley, B., & Wurgler, J. (2011). Benchmarks as limits to arbitrage: Understanding the low-volatility anomaly. *Financial Analysts Journal*, 67(1), 40-54.

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

Baltussen, G., Swinkels, L., & Van Vliet, P. (2021). Global factor premiums. *Journal of Financial Economics*, 142(3), 1128-1154.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, 9(1), 3-18.

Barroso, P., & Santa-Clara, P. (2015). Momentum has its moments. *Journal of Financial Economics*, 116(1), 111-120.

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.

Benston, G., (1969). The Value of the SEC's Accounting Disclosure Requirements. *Accounting Review,* 44(3), 515-532.

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

Braggion, F., & Moore, L. (2011). Dividend policies in an unregulated market: the London Stock Exchange, 1895–1905. *Review of Financial Studies*, 24(9), 2935-2973.

Brown Jr, W. O., Mulherin, J. H., & Weidenmier, M. D. (2008). Competing with the New York stock exchange. *The Quarterly Journal of Economics*, 123(4), 1679-1719.

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance*, 52(1), 57-82.

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

Chen, A. Y., & Zimmermann, T. (2020). Open-source cross-sectional asset pricing. *Working paper*, available at SSRN 3604626.

Chordia, T., Goyal, A., & Saretto, A. (2020). Anomalies and false rejections. *Review of Financial Studies*, *33*(5), 2134-2179.

Cohen, R. B., C. Polk, and T. Vuolteenaho (2003). The value spread. *Journal of Finance,* 58 (2), 609-641.

Cowles, A. (1938). Common Stock Indices, 1871-1937.

Daniel, K., & Moskowitz, T. J. (2016). Momentum crashes. *Journal of Financial Economics*, 122(2), 221-247.

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

Davis, J. L., Fama, E. F., & French, K. R. (2000). Characteristics, covariances, and average returns: 1929 to 1997. *Journal of Finance*, 55(1), 389-406.

Davis, L., & Neal, L. (1998). Micro rules and macro outcomes: the impact of microstructure on the efficiency of security exchanges, London, New York, and Paris, 1800-1914. *American Economic Review*, 88(2), 40-45.

Dowrie, G. W., & Fuller, D. R. (1950) Investments, *John Wiley & Sons*, New York.

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

Fama, Eugene F., and Kenneth R. French. 1992. The cross-section of expected stock returns. *Journal of Finance,* 47(2), 427-465.

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.

Fama, E. F., & French, K. R. (1998). Value versus growth: The international evidence. *Journal of Finance*, 53 (6), 1975-1999.

Fama, Eugene F., and Kenneth R. French. 2015. A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1–22.

Fama, Eugene F., and Kenneth R. French. 2016. Dissecting anomalies with a five-factor model. *Review of Financial Studies*, 29(1), 69-103.

Fama, E. F., & French, K. R. (2018). Choosing factors. *Journal of Financial Economics*, 128(2), 234-252.

Fama, E. F., & French, K. R. (2021). The value premium. *The Review of Asset Pricing Studies*, 11(1), 105-121.

Fohlin, C. (2016). The Venture Capital Divide: Germany and the United States in the Post-War Era. *Working paper*, available at SSRN 2849237.

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

Garvy, G. (1944). Rivals and interlopers in the history of the New York security market. *Journal of Political Economy*, 52(2), 128-143.

Gehrig, T., & Fohlin, C., (2006). Trading costs in early securities markets: The case of the berlin stock exchange 1880–1910. *Review of Finance* 10(4), 585–610.

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

Gibson, T. (1906). The pitfalls of speculation. *Moody Corporation*, New York.

Giesecke, K., Longstaff, F. A., Schaefer, S., & Strebulaev, I. (2011). Corporate bond default risk: A 150-year perspective. *Journal of Financial Economics*, 102(2), 233-250.

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* 130(3), 579-591.

Goetzmann, W., Ibbotson, R. & Peng, L. (2001). A New Historical Database for the NYSE 1815 to 1925: Performance and Predictability. *Journal of Financial Markets,* 4(1), 1–32.

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

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

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, 33(5), 2223-2273.

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

Harvey, C.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.

Hogan, W. W., & Warren, J. M. (1974). Toward the development of an equilibrium capital-market model based on semi variance. *Journal of Financial and Quantitative Analysis*, 9(1), 1-11.

Hou, K., van Dijk, M.A. (2019). Resurrecting the Size Effect: Firm Size, Profitability Shocks, and Expected Stock Returns. *Review of Financial Studies,* 32(7), 2850–2889.

Hou, Kewei, Chen Xue, and Lu Zhang. 2015. "Digesting Anomalies: An Investment Approach." *Review of Financial Studies,* 28(3), 650-705.

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

Israel, R., & Moskowitz, T. J. (2013). The role of shorting, firm size, and time on market anomalies. *Journal of Financial Economics*, 108(2), 275-301.

Jegadeesh, N. (1991). Seasonality in stock price mean reversion: Evidence from the US and the UK. *Journal of Finance*, 46(4), 1427-1444.

Jegadeesh, N., & Titman, S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *Journal of Finance,* 48(1), 65-91.

Jegadeesh, N., & Titman, S. (2001). Profitability of momentum strategies: An evaluation of alternative explanations. *Journal of Finance*, 56(2), 699-720.

Jensen, M. C., Black, F., & Scholes, M. S. (1972). The capital asset pricing model: Some empirical tests.

Jensen, T. I., Kelly, B. T., & Pedersen, L. H. (2021). Is there a replication crisis in finance? *Working paper*, National Bureau of Economic Research.

Jones, C. M. (2002). A century of stock market liquidity and trading costs. *Working paper,* available at SSRN 313681.

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

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

Leippold, M., Wang Q., & Zhou, W. (2021). Machine learning in the Chinese stock market. *Journal of Financial Economics*, forthcoming.

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

Linnainmaa, J. T., & Roberts, M. R. (2018). The History of the Cross Section of Stock Returns. *Review of Financial Studies* 31(7), 2606-2649.

Litzenberger, R. H., & Ramaswamy, K. (1979). The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence. *Journal of Financial Economics*, 7(2), 163-195.

Lowenfeld, H. (1909). All about investment. *Financial Review of Reviews*, London.

Macaulay, F. R. (1938). The social sciences and the unknown future: Is a reprint of the author's introduction to: Some theoretical problems suggested by the movements of interest rates bond yields and stock prices in the United States since 1856. *Working paper,* National Bureau of Economic Research.

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

McQuarrie, E. F. (2009). The Myth of 1926: How Much Do We Know About Long-Term Returns on US Stocks? *The Journal of Investing*, 18(4), 96-106.

Means, G. C. (1930). The diffusion of stock ownership in the United States. *Quarterly Journal of Economics*, 44(4), 561-600.

Michie, R. C., & Michie, R. (2006). *The global securities market: a history*. Oxford University Press.

Nagel, S. (2012). Evaporating liquidity. *Review of Financial Studies*, 25(7), 2005-2039.

Neal, L. (2016). The role of stock exchanges in financial globalization: A historical perspective. *Chapter 6 in Chambers & Dimson (eds) Financial market history: Reflections on the past for investors today*. CFA Research Institute Foundation.

Novy-Marx, R. (2013). The other side of value: The gross profitability premium. *Journal of Financial Economics*, 108(1), 1-28.

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

Novy-Marx, R., & Velikov, M. (2022). Betting against betting against beta. *Journal of Financial Economics*, 143(1), 80-106.

O'Sullivan, M. (2007). The expansion of the US stock market, 1885–1930: historical facts and theoretical fashions. *Enterprise & Society*, 8(3), 489-542.

Pätäri, E., & Leivo, T. (2017). A closer look at value premium: Literature review and synthesis. *Journal of Economic Surveys*, 31(1), 79-168.

Poitras, G. (2010). Valuation of Equity Securities – History, Theory and Application. *World Scientific Publishing Co.*

Poitras, G. (2012). From the renaissance exchanges to cyberspace: A history of stock market globalization. *In Handbook of Research on Stock Market Globalization*. Edward Elgar Publishing.

Pontiff, J., & Woodgate, A. (2008). Share issuance and cross-sectional returns. *Journal of Finance*, 63(2), 921-945.

Rouwenhorst, K. G. (2004). The origins of mutual funds. *Working paper,* available at SSRN 636146.

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

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

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

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

Smith, W. B., & Cole, A. H. (1935). Fluctuations in American business: 1790-1860.

Soebhag, A., Van Vliet, B., & Verwijmeren, P. (2022). Non-standard errors in asset pricing: Mind your sorts. *Working paper*, available at SSRN 4136672.

Vayanos, D., & Woolley, P. (2013). An institutional theory of momentum and reversal. *Review of Financial Studies*, 26(5), 1087-1145.

Wahal, S. 2019. "The Profitability and Investment Premium: Pre-1963 Evidence." *Journal of Financial Economics*, 131(2), 362-377.

Warshow, H. T., 1924, The Distribution of Corporate Ownership in the United States, *Quarterly Journal of Economics*, 39(1), 15-38.

Working, H. (1960). Note on the Correlation of First Differences of Averages in a Random Chain. *Econometrica* 28(4), 916-918.

Figures

Figure 1: Equity factor premia: in-sample and subsequent out-of-sample samples. The figure shows t-statistics of CAPM alphas for the size, value, momentum, short-term reversal, and low-risk factors over original 'in-sample CRSP' periods and subsequent 'out-of-sample CRSP' periods. The dotted line represents the traditional 5% significance level cutoffs. Factors are constructed from top-bottom portfolios from 2x3 size-characteristic-based portfolios. Definitions and in-sample periods generally follow the original studies: Banz, 1981, and Fama and French, 1992; Basu, 1977, and Fama and French, 1992; Jegadeesh and Titman, 1993; Lehmann, 1990, and Jegadeesh, 1990; Black, Jensen, and Scholes, 1972. The in-sample periods are 1926-1990 for Size, 1957-1990 for value, 1965-1989 for momentum, 1934-1987 for short-term reversal, and 1931-1990 for low-risk. The out-of-sample periods spans the period after till the end of 2019. Performance is measured on a monthly frequency.

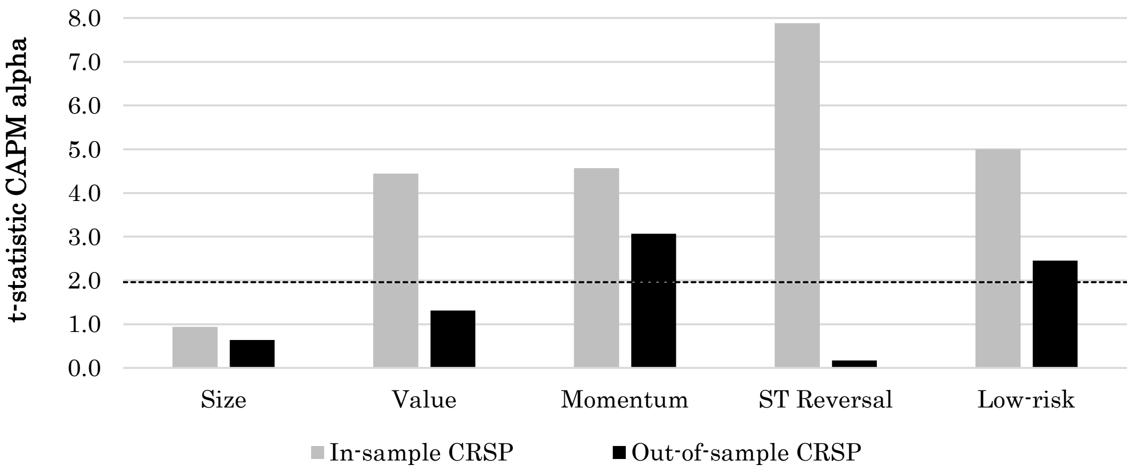
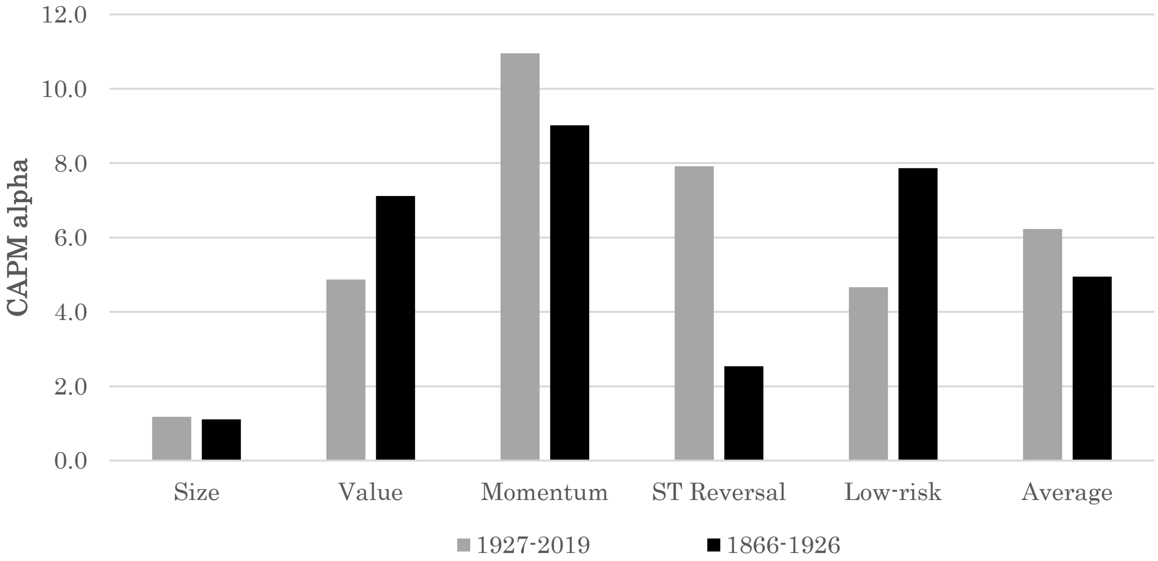


Figure 2: Equity factor premia: pre-CRSP and CRSP samples. The figure shows the average CAPM alphas for the size, value, momentum, short-term reversal ('ST Reversal'), and low-risk factors for the pre-CRSP and full CRSP samples. Factors are constructed from top-bottom portfolios from 2x3 size-characteristic-based portfolios. The pre-CRSP sample starts in January 1866 and ends December 1926. The CRSP sample runs between January 1927, the starting date of UMD, and December 2019. Performance is measured on a monthly frequency.



Tables

Table I: The U.S. stock database – 1866-1926

The table reports descriptive statistics of our sample. Panel A reports the statistics of our sample composition. The first three data columns show the number of stocks included in the cross-section ("No. of Stocks"), the number of stocks that pass our data quality screens ("No. of stocks included"), and the number of stocks that pass our data quality screens as a percentage of the number of stocks ("% of stocks included"). The last three columns show the average market capitalization of the stocks in the cross-section in millions of U.S. Dollars ("MV of stocks ($mln)"), of the stocks that pass our data quality screens ("MV of stocks included ($mln)"), and of the stocks that pass our data quality screens as a percentage of the total market capitalization of stocks ("MV of stocks included (%)"). Panel B reports summary statistics for the return distribution. It presents the sample averages of the value-weighted (annualized) total return, price return and dividend return, as well as the cross-sectional standard deviation ("CS std. deviation"), and 25th, 50th, and 75th percentiles of monthly total returns. The bottom row shows the grand average over our total sample. Statistics are shown per start of every 10-year period in our sample and over our full sample period. The sample runs from January 1866 to December 1926 and is at the monthly frequency.

Panel A: Sample composition

Year	No. of stocks	No. of stocks included	% of stocks included	MV of stocks ($mln)	MV of stocks included ($mln)	MV of stocks included (%)
1866	54	54	100.0%	278	196	70.4%
1876	123	69	56.1%	692	571	82.4%
1886	278	183	65.8%	1,622	1,256	77.4%
1896	455	180	39.6%	2,080	1,463	70.4%
1906	478	206	43.1%	8,412	6,412	76.2%
1916	485	257	53.0%	11,532	9,656	83.7%
1926	607	407	67.1%	18,775	16,406	87.4%
1866-1926	1,488	1,154	77.6%			

Panel B: Return distribution

Year	Total return	Price return	Dividend return	CS std. deviation	25-th percentile	50-th percentile	75-th percentile
1866-1869	6.98	-1.56	8.54	7.55	-2.43	0.55	3.44
1870s	9.88	2.64	7.24	9.72	-2.60	0.38	3.25
1880s	6.53	0.45	6.08	9.59	-3.08	0.47	4.01
1890s	6.97	2.32	4.65	9.51	-3.27	0.35	3.92
1900s	10.85	5.22	5.63	8.97	-2.81	0.42	4.24
1910s	6.93	-0.32	7.25	9.10	-2.77	0.30	3.65
1920-1926	12.74	0.26	12.48	11.09	-3.76	0.34	4.73
1866-1926	8.67	1.62	7.05	9.45	-2.97	0.39	3.89

Table II: Fama-MacBeth regression results

This table presents coefficient estimates from monthly Fama-MacBeth (1973) regressions of excess returns between month *t* and *t+1* against a constant and a series of stock characteristics, as described in Section III. Stock characteristics are measured at the end of month *t* over our sample period from January 1866 to December 1926. We report slope coefficients (multiplied by 100) with standard t-statistics in parentheses, the R^2 of the regressions ("R^2"), and the number of observations ("No. of obs."). Observations are value-weighted. *, ** and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Constant	0.66***	0.43	0.59***	0.61***	0.60***	0.43
t	(9.17)	(0.66)	(3.49)	(5.37)	(5.28)	(0.78)
Beta	0.05					0.09
t	(0.56)					(0.93)
ln(Size)		0.02				0.00
t		(0.50)				(0.01)
Dividend			2.07*			1.85**
t			(1.84)			(2.04)
Momentum				0.88**		1.00***
t				(2.51)		(3.39)
ST Reversal					-2.52**	-3.92***
t					(-2.27)	(-4.06)
R^2	0.12	0.03	0.04	0.06	0.06	0.24
No. of obs.	101,388	101,949	100,604	100,604	101,892	100,604

Table III: Portfolio sorts

The table reports average returns on univariate portfolios sorted by various stock characteristics, as described in Section III. Each month we sort stocks in ascending order into quintile portfolios ("Q1," "Q2," "Q3," "Q4," and "Q5) based on one stock characteristic and compute returns over the subsequent month. Portfolios are value-weighted. The table presents average annualized excess returns (Panel A), CAPM alphas (Panel B), and market betas (Panel C) for each portfolio, as well as the difference between the high portfolio and the low portfolio ("Q5–Q1'). The sample runs from January 1866 to December 1926 and is at the monthly frequency. Numbers in parentheses indicate standard t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level, which we present only for high-low portfolios.

Panel A: Excess return

	Q1	**Q2**	**Q3**	**Q4**	**Q5**	**Q5-Q1**
Beta	7.65	8.81	8.82	8.73	9.67	2.02
t	(8.95)	(8.01)	(5.51)	(3.45)	(2.64)	(0.59)
Size	11.77	8.44	7.92	8.54	8.93	-2.83
t	(4.27)	(3.74)	(4.23)	(4.93)	(6.09)	(-1.37)
Value	6.29	8.79	8.54	8.25	11.90	5.61**
t	(1.88)	(5.61)	(6.55)	(5.94)	(6.60)	(2.41)
Momentum	5.36	6.24	9.43	9.61	13.54	8.18***
t	(1.62)	(3.20)	(6.57)	(6.59)	(6.23)	(2.77)
ST Reversal	11.96	9.29	8.19	8.81	6.64	-5.31*
t	(4.01)	(5.28)	(6.06)	(5.46)	(2.93)	(-1.93)

Panel B: CAPM alpha

	Q1	**Q2**	**Q3**	**Q4**	**Q5**	**Q5-Q1**
Beta	2.32	2.12	0.28	-2.58	-4.49	-6.81***
t	(3.16)	(3.21)	(0.44)	(-2.70)	(-2.65)	(-3.32)
Size	1.40	-1.55	-1.13	-0.41	0.49	-0.92
t	(0.75)	(-1.31)	(-1.22)	(-0.61)	(1.72)	(-0.46)
Value	-7.11	0.85	1.05	0.43	3.02	10.13***
t	(-4.89)	(0.93)	(1.63)	(0.69)	(3.42)	(5.49)
Momentum	-7.17	-3.01	1.44	1.80	4.36	11.53***
t	(-3.83)	(-3.12)	(2.33)	(2.32)	(3.12)	(4.16)
ST Reversal	0.39	0.49	0.44	0.64	-2.87	-3.26
t	(0.22)	(0.59)	(0.75)	(0.73)	(-2.02)	(-1.21)

Panel C: CAPM beta

	Q1	Q2	Q3	Q4	Q5	Q5-Q1
Beta	0.30	0.58	0.97	1.55	2.15	1.85
Size	1.35	1.27	1.08	1.06	0.95	-0.40
Value	1.99	0.85	0.75	0.82	1.04	-0.95
Momentum	1.81	1.12	0.86	0.82	1.10	-0.70
ST Reversal	1.60	1.03	0.81	0.89	1.18	-0.43

Table IV: 2x3 sorted portfolios

The table reports average returns on 2x3 sorted portfolios sorted by size and various stock characteristics, as described in Section III. Every anomaly is constructed as an HML-like factor by sorting stocks first into six portfolios by size and the stock characteristic at the end of every month. The sorts use the 50^{th} percentile breakpoint on market capitalization, and subsequently the 30^{th} and 70^{th} percentile breakpoints on the stock characteristic. The return on the stock factor is the average return on the two high portfolios minus that on the two low portfolios, with "BETA" factor being ex-ante corrected for expected market beta. The high and low labels are chosen based on the 'CRSP-era' studies so that the stocks in the high portfolio earn higher returns than those in the low portfolios. Portfolios are value-weighted. Panel A presents average annualized excess returns ("Return"), standard deviation of returns ("Vol."), and the t-statistic of the average return ("*t*"). Panel B reports CAPM alphas ("Alpha"), beta ("Beta"), and the t-statistic of the CAPM alpha ("*t (alpha)*"). The sample runs from January 1866 to December 1926 and is at the monthly frequency. Numbers in parentheses indicate standard t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level, which we present only for returns spreads and CAPM alphas of the high-low portfolios.

Panel A: Excess return

Size

	Return				Vol.				*t*			
	S	B	SMB		S	B	SMB		S	B	SMB	
Total	9.64	8.47	1.17		14.83	12.96	7.92		(5.08)	(5.11)	(1.15)	

Value

	Return				Vol.				*t*			
	L	M	H	HML	L	M	H	HML	L	M	H	HML
Small	8.62	9.26	11.04	2.42	28.50	10.22	12.38	22.50	(2.36)	(7.08)	(6.96)	(0.84)
Large	6.79	8.73	9.89	3.10*	19.72	9.85	12.60	13.14	(2.69)	(6.92)	(6.13)	(1.84)
Total	7.70	9.00	10.46	2.76	22.86	8.86	11.36	15.35	(2.63)	(7.93)	(7.19)	(1.40)

Momentum

	Return				Vol.				*t*			
	D	M	U	UMD	D	M	U	UMD	D	M	U	UMD
Small	6.50	8.07	13.81	7.31**	28.70	14.49	17.79	25.03	(1.77)	(4.35)	(6.07)	(2.28)
Large	6.23	8.81	11.18	4.95***	17.58	10.41	13.52	14.37	(2.77)	(6.61)	(6.46)	(2.69)
Total	6.36	8.44	12.50	6.13***	21.82	11.66	14.44	17.34	(2.28)	(5.65)	(6.76)	(2.76)

ST Reversal

	Return				Vol.				*t*			
	LR	M	HR	ST_REV	LR	M	HR	ST_REV	LR	M	HR	ST_REV
Small	13.61	8.20	7.65	5.96**	24.66	14.05	20.03	22.72	(4.31)	(4.56)	(2.98)	(2.05)
Large	9.75	8.93	7.51	2.25	17.12	10.49	13.84	14.71	(4.45)	(6.64)	(4.24)	(1.19)
Total	11.68	8.60	7.58	4.10**	19.55	11.52	15.83	16.21	(4.67)	(5.83)	(3.74)	(1.98)

Beta

	Return				Vol.				*t*			
	LB	M	HB	BETA	LB	M	HB	BETA	LB	M	HB	BETA
Small	8.95	8.67	10.45	7.24***	7.78	17.54	31.37	16.17	(8.99)	(3.86)	(2.60)	(3.50)
Large	8.28	8.96	8.83	6.02***	7.30	11.43	22.61	14.03	(8.87)	(6.12)	(3.05)	(3.35)
Total	8.62	8.81	9.64	6.63***	6.46	13.55	25.83	12.46	(10.43)	(5.08)	(2.91)	(4.16)

Panel B: CAPM alpha and beta

Size

	Alpha S	Alpha B	Alpha SMB		Beta S	Beta B	Beta SMB		*t (alpha)* S	*t (alpha)* B	*t (alpha)* SMB	
Total	0.50	-0.61	1.11		1.10	1.09	0.01		(0.54)	(-2.57)	1.09	

Value

	Alpha L	Alpha M	Alpha H	Alpha HML	Beta L	Beta M	Beta H	Beta HML	*t (alpha)* L	*t (alpha)* M	*t (alpha)* H	*t (alpha)* HML
Small	-5.05	3.01	3.55	8.60***	2.05	0.50	0.75	-1.29	(-2.59)	(2.78)	(3.19)	(4.03)
Large	-4.31	1.16	1.32	5.63***	1.51	0.77	0.98	-0.53	(-3.97)	(2.36)	(2.04)	(3.78)
Total	-4.68	2.08	2.43	7.11***	1.78	0.63	0.87	-0.91	(-4.02)	(3.37)	(3.79)	(5.04)

Momentum

	Alpha D	Alpha M	Alpha U	Alpha UMD	Beta D	Beta M	Beta U	Beta UMD	*t (alpha)* D	*t (alpha)* M	*t (alpha)* U	*t (alpha)* UMD
Small	-6.46	-0.50	4.98	11.44***	1.90	0.98	1.03	-0.86	(-2.79)	(-0.44)	(2.98)	(3.88)
Large	-3.97	0.97	2.62	6.59***	1.32	0.83	0.98	-0.34	(-3.78)	(2.08)	(2.88)	(3.71)
Total	-5.22	0.23	3.80	9.02***	1.61	0.90	1.01	-0.60	(-3.76)	(0.39)	(3.58)	(4.42)

ST Reversal

	Alpha LR	Alpha M	Alpha HR	Alpha ST_REV	Beta LR	Beta M	Beta HR	Beta ST_REV	*t (alpha)* LR	*t (alpha)* M	*t (alpha)* HR	*t (alpha)* ST_REV
Small	2.06	-0.41	-2.08	4.13	1.60	0.97	1.22	0.38	(1.01)	(-0.39)	(-1.16)	(1.44)
Large	-0.20	1.05	-1.14	0.94	1.27	0.84	1.00	0.27	(-0.19)	(2.25)	(-1.21)	(0.51)
Total	0.93	0.31	-1.61	2.54	1.44	0.90	1.11	0.33	(0.74)	(0.54)	(-1.40)	(1.25)

Beta

	Alpha LB	Alpha M	Alpha HB	Alpha BETA	Beta LB	Beta M	Beta HB	Beta BETA	*t (alpha)* LB	*t (alpha)* M	*t (alpha)* HB	*t (alpha)* BETA
Small	3.52	-0.94	-4.26	8.92***	0.32	1.20	2.26	-0.35	(4.00)	(-0.70)	(-2.02)	(4.43)
Large	2.40	0.70	-3.56	6.81***	0.42	0.92	1.78	-0.16	(3.43)	(1.45)	(-3.31)	(3.80)
Total	2.96	-0.12	-3.91	7.86***	0.37	1.06	2.02	-0.26	(4.77)	(-0.17)	(-3.08)	(5.05)

Table V: Robustness tests

The table summarizes the robustness test results to methodological variations of equity characteristic portfolio sorts, as described in Section III. We consider the following methodological variations: quintile portfolios ("Quintile"), as in Table III, tercile portfolios ("Tercile"), decile portfolios ("Decile"), 2x3 size-characteristic sorted portfolios ("2x3"), as in Table IV, 2x5 size-characteristic sorted portfolios based on every 20th percentile breakpoint ("2x5"), equally-weighted 2x3 portfolios ("Equal-weighted"), and sector-neutral portfolio that construct 2x3 portfolios by first standardizing each characteristic within industries ("Sector-neutral"). The table presents average annualized excess returns (Panel A), and CAPM alphas (Panel B) of the high-low for each characteristic-sorted portfolio. Portfolios are value-weighted except for the row labelled "Equal-weighted". The sample runs from January 1866 to December 1926 and is at the monthly frequency. Numbers in parentheses indicate standard t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level, respectively.

Panel A: Return spread

	Size	Value	Momentum	ST Reversal	Beta
Quintile	2.83	5.61**	8.18***	5.31*	4.83***
t	(1.37)	(2.41)	(2.77)	(1.93)	(2.47)
Tercile	-0.15	3.34*	5.19**	2.66	5.87***
t	(-0.10)	(1.64)	(2.44)	(1.35)	(3.54)
Decile	5.59**	7.66***	6.08	11.38***	6.17**
t	(2.05)	(2.89)	(1.48)	(2.96)	(2.36)
2X3	1.17	2.76	6.13***	4.10**	6.63***
t	(1.15)	(1.40)	(2.76)	(1.98)	(4.15)
2X5	1.10	4.46**	5.51**	7.00***	5.68***
t	(1.19)	(2.04)	(2.01)	(2.72)	(3.10)
Equal-weighted	2.08**	2.49	6.86***	6.19***	6.61***
t	(2.20)	(1.20)	(2.93)	(2.91)	(4.02)
Sector-neutral	1.05	1.06	5.57***	4.78***	6.34***
t	(1.15)	(0.64)	(3.01)	(2.72)	(4.00)

Panel B: CAPM alpha

	Size	Value	Momentum	ST Reversal	Beta
Quintile	0.92	10.13***	11.53***	3.26	6.73***
t	(0.46)	(5.49)	(4.16)	(1.21)	(3.59)
Tercile	-2.04	7.69***	7.69***	1.18	6.64***
t	(-1.42)	(5.05)	(3.87)	(0.61)	(4.01)
Decile	3.47	12.46***	10.56***	8.27**	8.31***
t	(1.30)	(5.69)	(2.71)	(2.21)	(3.27)
2X3	1.11	7.11***	9.02***	2.54	7.87***
t	(1.09)	(5.04)	(4.42)	(1.25)	(5.05)
2X5	1.70	9.00***	8.75***	5.24**	7.57***
t	(1.86)	(5.41)	(3.42)	(2.07)	(4.33)
Equal-weighted	2.12**	7.18***	10.05***	4.75**	8.00***
t	(2.23)	(4.92)	(4.72)	(2.27)	(5.02)
Sector-neutral	0.87	4.76***	7.95***	3.45**	6.34***
t	(0.94)	(4.01)	(4.67)	(2.00)	(3.97)

Table VI: Spanning regressions

The table summarizes the results of spanning tests for each 2x3 size-characteristic sorted high-low factor return series on all other factor return series. We also include the value-weighted market factor ("Mkt-rf"). The stock characteristics are described in Section III. Portfolios are value-weighted. The sample runs from January 1866 to December 1926 and is at the monthly frequency. Shown are slope coefficients and intercepts (annualized and expressed in percentages) with standard t-statistics in parentheses, the R^2 of the regressions ("R^2"), and residual standard errors from each spanning regression ("s(e)"). Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level, respectively.

	Mkt-rf	**SMB**	**HML**	**UMD**	**ST_REV**	**BETA**
Intercept (ann.)	5.99***	1.49	3.89***	6.39***	4.96**	3.92***
t	(5.62)	(1.49)	(3.02)	(3.19)	(2.41)	(2.75)
Mkt-rf		-1.73	-9.11***	-3.48***	2.70***	2.31***
t		(-4.27)	(-22.32)	(-4.26)	(3.21)	(3.97)
SMB	-2.04***		-4.18***	-1.53*	-0.80	4.47***
t	(-4.27)		(-7.53)	(-1.71)	(-0.87)	(7.26)
HML	-6.43***	-2.49***		2.98***	0.06	4.99***
t	(-22.32)	(-7.53)		(4.35)	(0.37)	(10.93)
UMD	-1.01***	-0.38*	1.23***		-1.62***	1.02***
t	(-4.26)	(-1.71)	(4.35)		(-3.58)	(3.26)
ST_REV	0.75***	-0.19	0.10	-1.55***		-0.95***
t	(3.21)	(-0.87)	(0.37)	(-3.59)		(-3.08)
BETA	1.33***	2.18***	4.07***	2.03***	-1.96***	
t	(3.97)	(7.26)	(10.93)	(3.26)	(-3.08)	
R^2	0.53	0.11	0.60	0.25	0.10	0.26
s(e)	2.35	2.16	2.80	4.36	4.47	3.10

Table VII: Out-of-sample decay

The table reports the results of out-of-sample decay tests for stock factor portfolios. Factors are constructed from top-bottom portfolios from 2x3 size-characteristic-based portfolios. Portfolios are value-weighted. We include the value-weighted market factor ("Mkt-rf"), the stock characteristic-based factors described in Section III, and the equally-weighted average over the stock factor portfolios ("Average"). We estimate average (annualized) returns (Panel A) and CAPM alphas (Panel B) separately over the pre-CRSP ("1866-1926") and CRSP ("1927-2019") samples and examine their difference ("Difference"). The pre-CRSP sample starts in January 1866 and ends December 1926. The CRSP sample runs from January 1927 till December 2019. The rows labelled "1866-2019" present full sample results. Panel C shows the results of regressing the monthly CAPM alphas of each stock factor on all other factors, with the last column ("Panel") containing the results of a combining all stock factors into a panel regression with double (date/factor) cluster-corrected standard errors. Data is at the monthly frequency. Numbers in parentheses indicate standard t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level, respectively.

	SMB	HML	UMD	ST_REV	BETA	Average
Panel A: Return spread						
1866-1926	1.17	2.76	6.13***	4.10**	6.63***	4.22***
t	(0.91)	(1.45)	(2.85)	(2.29)	(4.94)	(5.45)
1927-2019	2.70***	0.83	8.69***	8.97***	4.66***	5.07***
t	(2.59)	(0.54)	(4.98)	(6.19)	(4.29)	(8.08)
Difference	-1.53	1.93	-2.55	-4.87**	1.96	-0.85
t	(-0.92)	(0.79)	(-0.92)	(-2.12)	(1.14)	(-0.85)
1866-2019	2.09**	1.59	7.67***	7.04***	5.44***	4.73***
t	(2.58)	(1.33)	(5.67)	(6.25)	(6.44)	(9.71)
Panel B: CAPM alpha						
1866-1926	1.11	7.11***	9.02***	2.54	7.87***	5.46***
t	(0.90)	(5.04)	(4.50)	(1.45)	(5.96)	(7.66)
1927-2019	1.18	4.87***	10.96***	7.92***	4.67***	5.85***
t	(1.18)	(4.26)	(6.75)	(5.61)	(4.37)	(10.14)
Difference	-0.07	2.25	-1.94	-5.39**	3.20*	-0.39
t	(-0.04)	(1.24)	(-0.75)	(-2.40)	(1.89)	(-0.43)
1866-2019	1.07	5.57***	10.03***	5.87***	5.82***	5.62***
t	(1.36)	(6.09)	(7.88)	(5.32)	(6.91)	(12.30)
Panel C CAPM alpha correlations of anomaly with other anomalies						
1866-1926	-0.08	0.33***	0.17*	-0.41***	0.44***	0.03
t	(-1.57)	(5.35)	(1.82)	(-5.85)	(7.48)	(0.19)
1927-2019	-0.23***	0.19***	-0.40***	-0.08	0.49***	-0.04
t	(-5.18)	(3.27)	(-5.11)	(-1.15)	(9.13)	(-0.23)
Difference	0.16**	0.15*	0.57***	-0.33***	-0.05	0.07
t	(2.40)	(1.73)	(4.68)	(-3.40)	(-0.60)	(0.41)

Table VIII: Machine learning and asset pricing

In this table, we report the performance of prediction-sorted portfolios based on two machine learning models: a Random Forest (RF) model, and a Neural Network with 3 layers (NN). Inputs are all characteristics used by Kelly and Xiu (2020) that can be computed based on our sample, with which next month returns are predicted. All stocks are sorted into quintiles based on their predicted returns for the next month. Results are computed over the 40-year out-of-sample period. Shown in Panel A are per quintile ("Q1",..,"Q5") and top-bottom portfolio ("Q5-Q1") the average realized (annualized) monthly returns (Avg. return"), their standard deviations ("Vol."), their Sharpe ratios ("SR"), and their CAPM alphas ("CAPM alpha"). Panel B summarizes the results of spanning tests of the top-bottom RF or NN portfolios on the factor return series. All portfolios are value weighted. Numbers in parentheses indicate standard t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level, respectively.

Panel A: Machine learning portfolio sorts

	Random Forest				Neural Network (3 layers)			
	Avg. return	SR	CAPM alpha	*t*	Avg. return	SR	CAPM alpha	*t*
Q1	7.02	0.13	-6.46***	(-3.47)	6.39	0.11	-6.37***	(-3.40)
Q2	8.96	0.27	-2.01	(-1.43)	7.22	0.20	-3.07**	(-2.37)
Q3	8.79	0.39	0.13	(0.14)	7.81	0.32	-0.81	(-1.07)
Q4	8.11	0.42	0.50	(0.68)	7.96	0.39	0.20	(0.25)
Q5	10.36	0.68	3.31***	(3.70)	11.43	0.72	4.25***	(4.07)
Q5-Q1	3.34	0.16	9.78***	(4.26)	5.05	0.25	10.62***	(4.42)
t	(1.00)				(1.58)			

Panel B: Spanning regressions

	RF	NN
Intercept (ann.)	0.10	0.17
t	(0.72)	(1.14)
Mkt-rf	-0.56***	-0.45***
t	(-9.78)	(-7.53)
SMB	-0.10	-0.27***
t	(-1.45)	(-3.71)
HML	0.85***	0.70***
t	(15.33)	(11.94)
UMD	0.18***	0.27***
t	(5.16)	(7.51)
ST_REV	0.29***	0.37***
t	(8.63)	(10.45)
BETA	0.19***	0.29***
t	(4.00)	(5.73)
R^2	0.77	0.72
s(e)	2.96	3.12

Table IX: Macroeconomic risk and factor returns

The table summarizes the explanatory power of macroeconomic risk for stock factor returns using methods outlined in Griffin, Ji, and Martin (2003). We regress the benchmark-adjusted returns of each stock factor on the following macroeconomic variables of Chen, Roll and Ross (1986): industrial production growth (MP), term premium (UTS), change in expected inflation (DEI), and unexpected inflation (UI). The coefficients and annualized intercept ("Interc. (ann.)") of the regression are shown in the table. We combine the resulting loadings against macroeconomic risks with estimates of risk premia of these risks (estimated using Fama and MacBeth on the 2x3 sorted individual and factor portfolios) to get the predicted return originating from an unconditional macroeconomic risk model ("Pred"). The table further contains the historical average annual return ("Act.") and the differences with predicted returns (i.e., the unexplained return; "Diff."). We estimate results separately over the pre-CRSP and CRSP samples. The pre-CRSP sample starts in February 1875 and ends December 1926. The CRSP sample runs from January 1927 till December 2019. The combined sample runs from February 1875 till December 2019. Both samples are at the monthly frequency. Numbers in bold are significant at the 5% level, while parentheses indicate standard t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level, respectively.

Factor	Period	MP	UTS	DEI	UI	Interc. (ann.)	*t*	Act.	Pred	Diff.	*t*
Size	1875-1926	**-0.03**	-0.16	-0.52	1.32	**3.07*** **	(2.68)	1.66	0.24	1.42	(0.60)
	1927-2019	0.01	**0.27**	1.36	-0.54	**-2.96***	(-1.76)	1.18	2.38	-1.20	(-0.72)
	1875-2019	0.00	**0.17**	1.18	0.16	-0.03	(-0.03)	1.29	1.61	-0.32	(-0.40)
Value	1875-1926	**0.03**	0.20	6.47	-1.55	**5.47*** **	(3.39)	**7.07**	0.13	6.94	(1.83)
	1927-2019	0.01	0.00	**-5.80**	0.77	**5.08*** **	(2.90)	**4.87**	0.48	**4.39** **	(2.31)
	1875-2019	**0.02**	0.04	**-4.69**	0.29	**4.44*** **	(3.78)	**5.52**	0.36	**5.17*** **	(5.58)
Momentum	1875-1926	-0.01	-0.01	1.99	-1.06	**8.99*** **	(3.67)	**8.34**	0.38	**7.96*** **	(3.46)
	1927-2019	**0.03**	-0.15	1.91	-0.87	**11.45*** **	(4.67)	**10.96**	1.62	**9.34*** **	(4.40)
	1875-2019	**0.02**	-0.02	2.29	**-1.65**	**8.59*** **	(5.13)	**9.91**	1.18	**8.73*** **	(6.63)
ST Reversal	1875-1926	**0.04**	0.01	1.31	-2.54	0.14	(0.06)	2.22	0.16	2.06	(0.55)
	1927-2019	**-0.02**	**0.21**	-0.46	0.22	**5.94*** **	(3.13)	**7.92**	-0.10	**8.02*** **	(6.13)
	1875-2019	-0.01	**0.22**	-0.71	-0.81	**3.98*** **	(2.82)	**5.93**	-0.01	**5.94*** **	(5.36)
BETA	1875-1926	0.00	0.18	-6.61	1.18	**8.70*** **	(4.79)	**8.92**	-0.95	**9.87*** **	(4.52)
	1927-2019	0.01	0.04	**-3.41**	0.17	**3.96*** **	(2.80)	**4.67**	1.10	**3.57** **	(2.18)
	1875-2019	0.01	0.01	**-4.18**	0.56	**5.75*** **	(5.32)	**6.11**	0.36	**5.74*** **	(6.76)

Online Appendix

In the main text we have analyzed factor premia in cross-section of U.S. stocks over a unique, novel sample between 1866 and 1926. In this Online Appendix we describe our dataset and the history of the U.S. stock market in more detail in Section A, including the cross-sectional dataset construction procedure and additional summary statistics, present results on the robustness to data filters or data quality screens in Section B, show additional results in Section C, and provide more detail on our machine learning tests in Section D.

Online Appendix A: The Pre-CRSP (1866-1926) sample

A brief history of the U.S. equity market

One of the first form of organized trading in U.S. stocks date to 1792, when the origins were laid for the New York Stock Exchange (NYSE) by 21 brokers and 3 firms agreed to maintain exclusive dealings and minimum commissions. This "Buttonwood Agreement" eventually evolved 25 years later in the NYSE. Soon, the first railroad stock was listed in New York (1830) and within two decades the exchange became predominantly a market for railroad securities (Garvy, 1944), where also banks stocks were well-represented. By the end of 1838, over 300 stocks were trading in the United States. The NYSE grew rapidly and had an annual business in excess of three billion dollars in 1867. In 1869, the NYSE merged with the Open and Gold Boards, and became the dominant exchange for trading stocks in New York and one the three leading exchanges in the world (Davis and Neal, 1998). Memberships now became tradable, and aspiring members could purchase seats from retiring members. Besides on the NYSE, stocks traded on the New York Curb, which later became named the American Stock Exchange (AMEX), and several regional exchanges.[41]

As the U.S. economy developed, demand for and supply of stock financing grew rapidly, with the U.S. stock market experiencing rapid growth between the early 1880s and late 1920s. Neal (2016) shows that in the early 20th century the New York stock market was large relative to the size of the U.S. economy, with a stock market capitalization to GDP ratio of 174%, about similar levels as observed in 2015. Most of the trading activity took place on the NYSE, followed by the NY Curb (the predecessor of the AMEX) and regional exchanges (mainly Boston and Philadelphia) (Brown et al., 2008, O'Sullivan, 2007). Total annual shares trading volume rose from about 100 million in 1885, to 150 million in 1900, to 250 million in 1915, to 1,151 million shares in 1930. Over two-thirds of trading volume originated from the

[41] The Curb market represents the market outside of general market operations. Trading took place outside the exchanges, on the street curb.

NYSE, followed by the New York Curb (about 20%), and regional exchanges (about 10% of total). In dollars, trading volume on the U.S. exchanges was $26.5 billion in 1920 and $49.5 billion in 1926 (O'Sullivan, 2007). On these exchanges 237, 860, and 1,675 number of stocks traded in 1866, 1896, 1926 according to our databases, respectively.

The 19th and 20th century markets shared many important behavioral and institutional characteristics (Harrison, 1998, Koudijs, 2016). Traded equities could quite readily be bought or sold across exchanges via stock dealer firms, traded via derivatives and options, could be bought on margin, and an active market existed for shorting stocks with well-known short speculators (see for example Gibson, 1906, Brown et al., 2008, Poitras, 2012). Major technological innovations such as the telegraph in 1844, the transatlantic cable (1866), the introduction of the ticker tape (1867), the availability of local telephone lines (1878), and direct phone links via cables around 1890 facilitated the growth in the depth and breadth of NYSE trading activity (Poitras, 2012, Fohlin, 2016). These innovations gave rise to a liquid and active secondary market for stocks and other securities, like corporate bonds (Giesecke et al., 2011). With the introduction of the transatlantic cable and ticker tape, price quotations were quite instantly known from coast to coast and on the other side of the Atlantic (Garvy, 1944, Hoag, 2006). Hoag (2006) notes that historical markets priced securities so well that transatlantic steamship crossing times can be recovered from stock prices. In the second half of the 19th century, the increased communication networks were utilized by several firms for arbitrage as prices on different exchanges were rapidly known, and increased brokerage and market making activities due to enhanced market liquidity. Investors had access to a wide range of reputable sources of information such as the Commercial and Financial Chronicle, newspapers, and monthly bulletin of all recorded prices on major exchanges and quarterly or semi-annual supplements which listed all the major companies and gave detailed information on securities issued by them (Giesecke et al., 2011). A sizable industry of financial analysts provided assessments of assets and financial markets, while also investment advice developed quickly and was not dissimilar to what we observe today (e.g., Lowenfeld, 1909).

Further, trading costs in the 19th century seem not very much different from 20th. Brown et al. (2008) shows trading costs were limited for many stocks. The median bid-ask spread for NYSE stocks remained fairly constant between 1885 and 1926 at 2.0% for most of the period, but the higher-volume stocks and NYSE stocks that also traded at other exchanges had about a quarter of these costs, or even often traded at the minimum tick of 1/8th. Jones (2002) reports spread estimates for Dow Jones stocks of about 0.5% since 1900, not much different from CRSP-era estimates up to round 1980, and annual share turnover on NYSE stocks being higher between 1900 and 1926 than in 2000. Fohlin (2016) reports that in the decade prior to World War I, quoted spreads at the NYSE averaged about 2%, but the median spread was 86 basis points, and a quarter of trades took place with spreads less than 36 basis points.[42]

Stock ownership was spread over many investors with stock data being well available. Market participants in early U.S. stock market mostly were wealthy individuals, but also banks and insurance companies[43], retail investors, investment trusts, and arbitrage players. In the 19th century, stock ownership was largely dominated by the rich. However, stock ownership expanded rapidly as of around 1900 from the rich to the less rich, making the middle class an important factor. Warshow (1924) and Means (1930) estimated that the number of stockholders grew from 4.4 million in 1900, to 8.6 million by 1917, to 18 million by 1928, driven by amongst other entrepreneurs and large trusts unloading their stock upon the public, financial education campaigns teaching the less wealthy to save and invest, and larger incomes of the wage-earning classes. Broad market indices were introduced around 1885

[42] Related, Gehrig and Fohlin (2006) find low trading costs in the German stock market, a major stock market in the 19th and 20th century, among a nearly comprehensive set of stocks trading in Berlin for four benchmark years (1880, 1890, 1900, and 1910). Koudijs (2016) finds that trading costs in the 1770s and 1780s Dutch stock market were substantially lower than over recent decades.

[43] For example, O'Sullivan (2007) reports $781 million of bank security holdings in utility and industrial companies in 1920.

when Charles Dow began publishing a daily index of actively traded, large capitalization stocks, with the Commercial and Financial Chronicle (and later Wall Street Journal) being a well-read financial newspaper containing daily information on stock prices, volumes, and other characteristics.

Dataset construction

We have compiled our data from several sources in order to obtain a reliable and historically extensive dataset. Our sample covers 61 years of data on monthly stock prices, dividend yields and market capitalizations for all major stocks traded on the NYSE, NY Curb and regional exchanges. The sample spans the period from January 1866 through December 1926 and is at the monthly frequency. We build our dataset from the Commercial and Financial Chronicle (CFC, the source also used to build the CRSP sample as of 1926) and Global Financial Data (GFD), which we combine with risk-free rates from Jeremy Siegel's website. Note that we choose to overlap our sample partly with CRSP over 1926, as characteristics like momentum and beta require at least one year of data, and as such are not tested in CRSP over 1926. Below we further outline the data sources and the construction of each series we use in detail.

Data sources and items: We collect dates, company identifiers, company names, monthly stock prices, dividends, price returns and monthly total returns from GFD, all adjusted for stock splits. GFD handles splits, dividends and stock dividends by adjusting the total and price return series with the relevant multipliers, which we have verified by also calculating returns ourselves for a random subset of 100 stocks. We calculate dividend yields by subtracting the monthly price return from the monthly total return, also to capture negative dividends. One important note to make is that several companies engaged in 'assessments'. These are basically reverse dividends, in which companies called for capital upon its shareholders to pay for the difference between par and nominal amounts. The GFD stock database has an extensive coverage of historical stocks traded in the U.S. across the NYSE and regional exchanges, as well as stocks traded in the Over-The-Counter (OTC) markets and includes delisted stocks. As such it is relatively free of a survivorship bias or an exchange bias (i.e., focusing on a specific exchange, while historically many exchanges varied in importance). GFD has covered United States stock prices from 1791 till date. As a downside, this database did not include number of shares outstanding.

We manually collect shares outstanding from the CFC, the first national business newspaper in the United States. The CFC was a weekly newspaper founded in 1865 representing the industrial and commercial interest of the United States. The Fraser library has published a digital archive of this newspaper online, with articles from July 1st, 1865, to August 23, 1962, implying our start date of 1866 for this study. These articles contain company names, prices, dividends, par value outstanding, and size of par value, both for stocks and bonds. We retrieve par value outstanding and size of par value from the CFC. Figures A.1 and A.2 show pages of the CFC in 1865 and 1925. Note that the first few years, December 1925 - January 1928, of the monthly stock data from the Center for Research in Securities Prices (CRSP) database were also gathered from the Commercial and Financial Chronicles' Bank and Quotation Section and Public Utility Compendium. The following 33 years (February 1928 - December 1960) were assembled from an expansion of this section, the Bank and Quotation Record. From CFC we collect par value outstanding and par value of a share via the following procedure.

To keep data collection efficient, we apply the following procedure. We start by collecting the CFC data in five-year periods of e.g., 1865, 1870, ..., 1925. If data items differ in value between five-year periods, we continue by also collecting the data items for every year in between. The main assumption behind this methodology is as follows: if the number of shares outstanding in year 1 is equal to the number of shares outstanding in year 6, every value in between is likely to have the value found at year 1 and the same and interpolated accordingly.

We have performed 100 random checks to verify this methodology, with a 100% success rate, confirming the efficacy of the above approach. Most of the interpolation is done for stocks in the banking industry, as most banks did not have frequent changes in their number of shares outstanding. The data on shares outstanding per year has been compared with past and future values at the time of entry. The companies' shares outstanding are calculated as the amount of par value outstanding divided by the par value of a share. Most shares were issued at a par value of 100 dollars before 1926, with however several stocks breaking up of their par-values into 50, 25, 10, 5 and even 1 dollar shares post World War I. Table A.1 shows an example of the data-collection procedure.

Table A.1: Data collection example

This table displays how entries have been added to create the data set of number of shares outstanding for the period 1866-1926. "Found as" contains the name of the company as found in the CFC, "Industry" refers to the industry the company belongs to, and columns 1869 through 1874 show the number of shares outstanding. The "Company" and "Found as" names are abbreviated, for example Continental (NY) was Continental national bank of New York (NY). The wide format allows for direct comparison when filling the table with entries

Company	Found as	Industry	1869	1870	1871	1872	1873	1874
Continental (NY)	Continental	Bank	20,000	20,000	20,000	20,000		
NY NH Railroad	NY and NH	RR	90,000	90,000	90,000			
Penn Coal Co.	Pennsylvania Coal	Misc		64,000	80,000	80,000	80,000	80,000
Morris & Essex RR Co.	Morris and Essex	RR	157,602	157,602	273,344	273,965	280,162	283,309
Sw Rr Georgia	SW (Georgia)	RR		39,399	38,773	38,773	38,773	

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

- Misprints and other measurement errors. This could cause prices to be spuriously inflated, trigger potential value profits.

- Reported prices in our databases are not necessarily transaction prices, but bid prices, ask prices, average prices of the day or month, or average of daily or monthly high and low prices. The use of bid or ask prices creates artificial short-term reversal effects, while the use of average prices over a month creates an artificial AR(1) process (see Working, 1960, Schwert, 1990). Working (1960) shows that such time averaging does not induce autocorrelation beyond a one-month horizon, and therefore does not preclude testing for momentum effects, provided that one skips a month between the end of the formation period and the beginning of the holding period.
- Missing data, which have sometimes been solved by interpolating, or padding, prices or returns known at a lower frequency to the monthly frequency.
- The timing of equity dividends was not always known historically. As a solution, to construct return series, they have sometimes been distributed to fixed points over the year, often year ends. For equities this can result in high returns during 'assigned dividend' months, while returns may be artificially low on the actual ex-dividend months (as prices may drop to reflect the dividend payment). This could generate spurious seasonality in stock returns.

Data selection: We applied several data filters on the database to focus on common stocks that are economically comparable and interesting to be used in factor research. First, we filter all securities in the GFD database by excluding every instrument that does not have "United States" as home country. This filter excludes many international stocks. Second, we exclude all instruments that do not have "United States Dollar" as currency. Third, we remove every instrument that is not a *common* stock (noteworthy is that the word stock was historically also used for debt claims, common stocks refer to equity claims), and remove bonds, real estate investment trusts (REITs), American depositary receipts (ADRs), certificates, preferred stocks and other financial instruments. Fourth, we remove all stocks listed on OTC exchanges

(PK, OTC, BB, QBB, QX) from our analyses, in line with CRSP-era studies.[44] Fifth, we require each stock to have at least 12 monthly return observations and remove observations after the stock price dropped below one dollar, or after receiving a return of -100% in one month. These choices imply the more illiquid, less traded stocks may drop out of our sample. As these are often microcap stocks, this possibly influences SMB return estimates. That said, we believe these choices to be important to optimize the quality of the pre-CRSP dataset. Sixth, a comparison between the GFD and CRSP database between 1926 and 2018 revealed that GFD includes many stocks that went bankrupt, and which traded as penny stocks in the years following bankruptcy. To remove these stocks from our sample we eliminate those stock observations that have had a previous two months' return of at least -70%, as this filter largely eliminates the difference between GFD and CRSP. As data collection is a very labor-intensive process, we applied data filters one to six before collecting the number of shares outstanding. Seventh, we require stock to have shares outstanding (and hence market capitalization) available over the previous year-end. Table A.2 shows the exclusion criteria in greater detail. Finally, we drop NYSE stock observations for the period July 1914 – December 1914 from our sample and for our data quality screens, as the NYSE was closed for over this period due to World War I.

Using the filters described above, we collect 22,493 yearly number of shares outstanding observations from the CFC.[45] In total, we have collected data for 1,488 U.S. common stocks from CFC. As a result, the combined shares outstanding and GFD data between 1866 and 1926 contains 1,488 unique common stocks with market capitalization values.

[44] The GFD database includes stocks traded on the Philadelphia, Boston, or Chicago exchanges, but many of these stocks are not covered in the CFC and hence we lack their market capitalization data. Moreover, many stocks from these exchanges have gaps in their monthly returns or traded as penny stocks. More specifically, of the 196, 523, or 121 unique stocks featured in GFD from Philadelphia, Boston, or Chicago, respectively, 69, 130, or 17 have sufficient coverage and market capitalization data available.

[45] Further, we have collected about 34,000 observations spread over 2,777 U.S. OTC stocks, as they also represented a tradable market. For example, O'Sullivan (2007) reports that the OTC market in stocks accounted for about $2,5 (3,5) billion in trading volume in 1920 (1926), or 6% (7%) of the value of exchange sales in that year.

Table A.2: Sample exclusion criteria

This table outlines the filters we have applied to the Global Financial Data stock data set.

Exclusion Criteria	Description
1. Domestic stocks only	If one financial instrument is not from the United States, it is excluded from the sample.
2. Domestic currency only	If one financial instrument is denominated in a currency other than United States dollar, it is excluded from the sample.
3. Common stocks only	If one financial instrument is not a common stock, it is excluded from the sample. Instruments excluded are: American depositary receipts (ADRs), corporate bonds, exchange traded funds (ETFs), government bonds, municipal bonds, preferred stocks, preferred convertibles, preferred trusts, real estate investment trusts (REITs), rights, scrips, state bonds, units, and warrants.
4. Stocks from non-OTC exchanges only	If one stock is listed on an over-the-counter exchange, it is excluded from the sample. OTC-exchanges include: BB, OTC, PK, QBB and QX.
5. Qualified stocks only	If one stock has less than 13 monthly return observations, it is excluded from the sample. Additionally, observations are removed after the stock price has dropped below one dollar, or after receiving a return of -100% in one month.
6. Remove bankruptcy listings	If one stock those had a *previous* two months' return equal to or lower than -70% it is excluded afterwards
7. Stocks with market capitalization only	If one stock does not have market capitalization, it is excluded from the sample.

Further, we applied a number of conservative screens on our data series and remove data points when they do not pass these screens, as outlined in Section III of the paper. These screens reduce the impact of data issues such as missing monthly data, reduced liquidity or non-tradability ('zero return screen'), the possibility that prices or returns known at a lower frequency have been interpolated to the monthly frequency ('return interpolation screen'), and the possibility that returns are stale or update infrequently ('stale return screen'). These screens eliminate 23.4% of the equity observations, of which the large bulk is due to the zero-return screen and eliminating missing returns.

Data verification procedure: We have taken the following steps to check the quality of each data series and clean for obvious measurement errors. First, we have randomly checked 100 observations in GFD against prices and dividends reported in the CFC. Similarly, we have verified the GFD data against 126 matched listings in the International Center for Finance at Yale database.[46] These checks all verified the GFD data. Second, we have manually verified extreme returns (>100%, <-50%), dividends, and changes from year to year in number of shares outstanding, and when due to a data error corrected. We have checked changes from year to year in number of shares outstanding. For example, if the value of shares outstanding in 1870 divided by the value in 1869 is equal to 0.1, or 10, there was a high chance a zero to many or a zero to few was added to the value in 1870. These values were checked again in the data sources and when erroneous replaced with the correct value. Third, we have compared the number of stocks, overall, per exchange and per sector with other sources, like O'Sullivan (2007) and Michie (2006) and found them to be roughly in line. Fourth, we have compared GFD against CRSP over the post-1926 sample in terms of number of firms and average returns, causing us to apply data quality filter six described above. Fifth, we checked each series on gaps, the level and dynamics in the first- and second-order autocorrelations. Finally, we built an industry classification starting from GFD subgroups: Financials (Finance & real estate), Energy/Mining (Materials & Energy), Industrials & Other (all other), Infrastructure (Transports), and Utilities. Subsequently, we have manually checked company names against classifications in CFC, and when available descriptions of company practices. This verification led us to reclassify several companies compared to GFD.[47]

Survivorship and delisting biases: The sample includes delisted stocks and as such is believed to be free of a survivorship bias. A related issue is the possibility of a delisting bias within the database. If large negative returns are not well documented, for example in case of bankruptcy or a default, this tends to overstate the returns of risky assets and understate

[46] International Center of Finance at Yale University (http://icf.som.yale.edu/old-new-york-stock-exchange-1815-1925).

[47] More specifically, 12 stocks were reclassified from Infrastructure to Industrial & Other, two stocks were reclassified from Financials to Industrials & Other, and one stock was reclassified from Industrials & Other to Infrastructure. We follow the classification of GFD in case of conglomerates and firms changing industries over time.

the returns of less risky assets. For example, the CRSP database contained a delisting bias for many years before it was detected and cleaned by Shumway (1997). This bias was most severe among small risky stocks, thus leading to an overestimation of the size premium (Shumway and Warther, 1999). A possible delisting bias in general overstates the returns of risky assets thus leading to a potential underestimation of the BETA premium in particular. We believe this to be of limited concern. First, we have stocks entering and exiting the sample over time but have stocks that experience bankruptcy being maintained in the sample for several years. We apply data filter six to manage these observations. Second, our approach of using value-weights limits the potential impact of a delisting bias.

Other studies to U.S. stock prices pre-1926: We are not the first to use historic data of the United States stock exchanges before 1926, but to our knowledge we are the first to use market capitalizations throughout our sample period in constructing factor portfolios and use the Commercial and Financial Chronicle (CFC) as a data source. Other studies have mainly used different sources. Goetzmann, Ibbotson, and Peng (2001) use The New York Shipping List, The New York Herald, and The New York Times and collect end of month equity prices and combine these with semi-annual dividend announcements of The New York Commercial, The Banker's Magazine, The New York Times, and The New York Herald. Their collected data have a few gaps, 1822, part of 1848, 1849, and 1866, all of 1867, January 1868 and July 1914 to December 1914. Golez and Koudijs (2018) use the data of Cowles III et al. (1938) for the period 1871-1925.[48] Unfortunately, the original Cowles data were lost and only the monthly indices remain. Schwert (1990) spliced the monthly stock returns of Smith and Cole (1935), Macaulay (1938), and Cowles III et al. (1938) and created a monthly stock return index from 1802 to 1925.[49] Geczy and Samonov (2016) use a combination of GFD, the International

[48] The monthly Cowles indices are available at: https://som.yale.edu/faculty-research/our-centers-initiatives/international-center-finance/data/historical-cowles.
[49] The monthly index of William Schwert is available at: http://schwert.ssb.rochester.edu/mstock.htm.

Center for Finance at Yale (ICF), and Inter-University Consortium for Political and Social Research (ICPSR) databases to study price momentum in U.S. stock markets between 1800 and 1925. However, their sample lacks dividend and market capitalization data, implying they have to rely on equal-weighted price returns and are consequently plagued by the abundance of small caps and banks historically. None of the previously discussed studies can consistently use market capitalization values to weight their market indices or construct factor portfolios.

Summary statistics: First, we present summary statistics of the value-weighted and equal-weighted market returns between 1866 and 1926, as shown in Table A.3. Shown are the average (annualized) return and volatility by decade and over the full 1866-1926 period. We compare these with other available U.S. equity return series from Schwert (1990), which was value-weighted between 1863 and 1885 and price-weighted thereafter, and Goetzmann, Ibbotson and Peng (2001), which was price-weighted over the entire sample but excludes dividends. As both series end in 1925, we append the series over 1926 with the constructed market returns from our database. Figure A.3 depicts the resulting series. We find that U.S. stock returns are generally of comparable magnitudes across our and the Schwert data-series, while the Goetzmann et al. series lag by about the average dividend yield in our sample. The average yearly value-weighted total (excess) return of the early sample was 8.67% (4.78%) and dividends contributed to 81% of this return, the average yearly value-weighted dividend return was 7.05%. For comparison, when equally-weighted the yearly total (excess) return is 9.42% (5.53%) and dividends contributed 51% to this return. As the equal-weighted index shows, the influence of smaller market capitalization stocks is positive on the total return and negative on the dividend return, due to larger companies having higher dividend yields. Furthermore, we compare our market returns with the CRSP sample between 1927 and 2019, again finding similar statistics. The difference in the value-weighted total or excess market return between the early (1866-1926) sample and the CRSP sample is 2.57% (8.67% versus 11.24%) or 3.18% (4.78% versus 7.96%), both of which do not do not significantly differ from each other. In the CRSP sample, dividends contributed 32% to the total returns, as the yearly value-weighted total returns were 11.24% and the yearly value-weighted dividend returns were 3.61%. This shift from 81% to 32% shows that the structure of total returns of investments changed over the 19th and 20th century.

Second, we summarize the distribution of the key variables in our final dataset. Figure A.4 compares the U.S. stock market capitalization distribution (by plotting the timeseries average of the monthly cross-sectional distribution statistics) of the stocks in our sample with the CRSP sample, finding overall similarly distributed market capitalizations. Figure A.5 depicts the number of dividend payers versus zero-dividend versus negative dividend stock, split per small (market capitalization below median) and large (market capitalization above median) stocks. Figure A.6 repeats the same exercise for share issuance. Figure A.7 shows the 20th, 50th, and 80th percentiles of key characteristics at each point in time.

Third, we report details on our dataset composition. Table A.4 shows the number of stocks in our sample before and after our filters. Tables A.5 and A.6 shows the number of stocks and the cross-sectional composition of market capitalizations per sector and exchange. Key sectors were infrastructure stocks (especially railroads), industrials, mining, and utilities (for example, telephone and telegraph stocks). Railroads where the most important stocks in terms of market capitalization for the first 30 years of our sample (see also Garvy, 1944). This changed around 1890 when the industrial stocks and mining stocks started dominating the stock exchanges (see also Garvy,1944). In the early 1860s, mining securities made their appearances on the stock markets, these included oil, copper, and gold mining stocks (Garvy, 1944). Banks became very prominent in the lists of U.S. stocks traded in the early part of the 20th century (see also Goetzman, Ibbotson and Peng, 2001). In 1896, the number of banks reported in the CFC increased considerably, with above 50% of all the stocks being bank stocks from 1896 to 1910. However, for the most part of our sample (up to the 1920s) bank stocks were not widely traded and represent relatively low market capitalizations, which has been attributed to their double liability characteristic (i.e. stockholders of a failing bank could lose not only the amount they had spent in purchasing the shares but could also be assessed an amount up to the par value of the shares they owned) and their relatively low dividend payments (O'Sullivan, 2007).[50] For example, our sample has over 250 stocks in the banking industry after 1896, but they only contributed to around 10% of the total market capitalization. Note that when creating an equal-weighted index, the index return will largely be driven by the (historically less important) banking industry.

Regional exchanges gained in importance mainly as of the 1900s and presented a sizable fraction of market capitalization (increasing from 7% in 1866 to 30% in 1906, while dropping to 22% in 1926). The New York Curb market (the predecessor of the AMEX) gained importance as of the mid-1920s (close to the start of CRSP) presented a small fraction of the market capitalizations (1% in 1926). The NYSE had conservative listing requirements, precluding it from admitting issuers other than the largest and most well-established companies and, at that time, such companies in the United States tended to be railroads (O'Sullivan, 2007). Outside of the NYSE many small, and typically more thinly traded securities were listed on the New York Curb and regional exchanges, mostly banks (financials) and textile companies (industrials). By the 1880s, the NYSE was largely an exchange for railroad stocks, with the most actively traded stocks on the NYSE being generally railroads or Western Union (Brown et al., 2008, see also Goetzmann Ibbotson and Peng, 2001). Ten years later, railroads continued to dominate the ranks of NYSE stocks but energy/mining stocks, industrial and utility stocks had grown considerably in importance. Most of the utility companies that were added to the Exchange in the period 1886–1895 were

[50] See for example, Michie (2006, p. 104): "*bank and insurance stocks …. These did not generate sufficient turnover to justify space on the trading floor and the attention of members, and so were also traded outside on the street or curb market.*"

traction companies, telephone, telegraph and cable companies, and electric and gas companies.[51]

[51] GFD and CFC also contain data on OTC stocks, which we have mostly collected but excluded from the sample employed in this paper. The OTC market was a sizable market in terms of number of listings. Our sample includes 2,777 unique OTC stocks (compared to 1,488 non-OTC stocks), but they are typically small and thinly traded, and have might have opaque stock structures and governance. O'Sullivan (2007) reports that the OTC market in stocks accounted for about $2,5 (3,5) billion in trading volume in 1920 (1926), or 6% (7%) of the value of exchange sales in that year. Further, the amount of OTC stocks in our sample increased a lot in 1896, driven by the banking and quotation record of the CFC starting to report prices and shares outstanding data of bank stocks across almost every state. There were a huge number of bank stocks in the late nineteenth and early twentieth centuries in the United States, but as with listed bank stocks these were typically small and little traded.

Figure A.1: Example Commercial and Financial Chronicle 1865

860 THE CHRONICLE. [December 30, 1865.

RAILROAD, CANAL, AND MISCELLANEOUS STOCK LIST.

COMPANIES.	Stock outstanding.	Dividend. Periods.	Dividend. Last p'd.	Market. Bid.	Market. Askd
Railroad.					
Albany and Susquehanna......100	1,347,192				
Alleghany Valley.............. 50	1,947,600				
Alton and St. Louis............100	800,000	Quarterly.	Jan ..1¾	92	
Atlantic & Great Western, N. Y.100	919,153				
do do Pa...100	2,500,000				
do do Ohio.100	5,000,000				
Baltimore and Ohio............100	13,188,902	April and Oct	Oct...4	111½	115
Washington Branch....100	1,650,000	April and Oct	Oct...5	127	
Bellefontaine Line..... 100	4,434,250	Feb. and Aug	Aug..3		
Belvidere, Delaware............100	997,112				
Berkshire......................100	600,000	Quarterly.	Oct...1¾		
Blossburg and Corning......... 50	250,000	June & Dec.	Dec. .2½		
Boston, Hartford and Erie......100	8,500,000			11½	13
Boston and Lowell.............500	1,830,000	June & Dec.	Dec..3½	94	95
Boston and Maine..............100	4,076,974	Jan. and July	Jan ..4	118½	120
Boston and Providence.........100	3,160,000	Jan. and July	Jan ..5	125	126
Boston and Worcester..........100	4,500,000	Jan. and July	Jan ..4½	130	132
Brooklyn Central...............100	492,150				
Brooklyn City.. 10	1,000,000	Feb. and Aug	Aug..3½	180	
Brooklyn City and Newtown...100	366,000				
Buffalo, New York, and Erie...100	850,000	Jan. and July	July..3½		
Buffalo and State Line.........100	2,200,000	Feb. & Aug.	Aug..5		190
Burlington and Missouri River.100	1,000,000				
Camden and Amboy............100	6,472,400	Jan. and July	Jan...5	125½	126
Camden and Atlantic........... 50	378,455				
do do preferred.. 50	682,600				
Cape od..................... 60	681,665	Jan. and July	July..3½		
Catawissa.................... 50	1,150,000			43	43½
o preferred 50	2,200,000	Feb. & Aug.	Aug..3½	72	75
Central of New Jersey..........100	5,600,000	Quarterly.	Oct...2½	118	120
Central Ohio....................					
Cheshire (preferred). 100	2,085,925			48	49
Chester Valley................ 50	871,900			102	104¾
Chicago and Alton.............100	1,783,100	Feb. & Aug.	Aug..3½	104½	105
do preferred....100	2,425,200	Feb and Aug.	Aug..3½		109
Chicago Burlington and Quincy.100	8,376,510	May & Nov.	N.5c&20s	113½	114
Chicago and Great Eastern.....100					
Chicago, Iowa and Nebraska....100	1,000,000				
Chicago and Milwaukee........100	2,250,000			53	
Chicago and Northwestern 100	13,160,927			35½	35½
do do pref..100	12,994,719	June & Dec.	June..3½	61⅞	61½
Chicago and Rock Island.......100	6,000,000	April and Oct	Oct...5	107¼	107¾
Cincinnati and Chicago Air Line100	1,106,125				
Cincinnati, Hamilton & Dayton.100	3,000,000	May and Nov.	Nov..5	98	100
Cincinnati and Zanesville.......100	2,000,000			12	
Cleveland, Columbus, & Cincin.100	6,000,000	Feb. and Aug	Aug..5		125
Cleveland, Painesville & Ashta.100	4,000,000	Jan. and July	Jan		
Cleveland and Pittsburg 50	5,253,625	Jan. and July	Jan.'66 4	82½	82⅝
Cleveland and Toledo.......... 50	4,654,800	April and Oct	Oct...5	113	113¼
Columbus & Indianapolis Cent.100					125
Columbus and Xenia...........100	1,490,800	Jan. and July	July..5		
Concord............ 50	1,500,000	Jan. and July	July..3½	120	
Concord and Portsmouth.......100	250,000	Jan. and July	July..3½	126	
Coney Island and Brooklyn.....100	500,000				
Connecticut and Passumpsic.. 100	892,900				
do do pref.100	1,255,200	Jan. and July	July..3	74	
Connecticut River 100	1,591,100	Jan. and July	July..4	103½	
Covington and Lexington......100	1,582,169				
Dayton and Michigan 100	2,316,705			20	
Delaware. 50	406,132	Jan. and July	July..3		
Delaware, Lacka., & Western .. 50	6,832,950	Jan. and July	Jan...3	165	170
Des Moines Valley.............100	1,550,000				
Detroit and Milwaukee.........100	952,350				
do do pref.....100	1,500,000				
Dubuque and Sioux City.......100	1,751,577				
do do pref.....100	1,982,180				
Eastern, (Mass) 100	3,155,000	Jan. and July	Jan...3	99¾	100
Eighth Avenue, N. Y...........100	1,000,000	Quarterly.	Oct		
Elmira, Jefferson, & Canandagua100	500,000	Feb. and Aug	Aug..2½		
Elmira and Williamsport 50	500,000	Jan. and July	Jan...2½	52	53
do do pref... 50	500,000	Jan. and July	Jan...3½	82	86
Erie 100	16,400,100	Feb. & Aug.	Aug..4	95⅞	96
do preferred 100	8,535,700	Feb. & Aug.	Aug..3½	85½	87
Erie and Northeast... 50	400,000	Feb. & Aug.	Aug..5		
Fitchburg.......................100	3,540,000	Jan. and July	July..3	106	
Forty-sec'd St. & Grand St. F'y.100	750,000	April and Oct	Oct 5..		
Hannibal and St. Joseph.......100	1,900,000			30	
do do pref...100	5,253,836				
Hartford and New Haven.......100	2,350,000	Quarterly.	Oct ...3		
Housatonic 100	820,000				
do preferred..........100	1,180,000	Jan. and July	July..4		
Hudson River..................100	6,218,042	April and Oct	Oct...4	108⅞	108⅝
Huntingdon and Broad Top 50	617,500				
do do pref. 50	190,750	Jan. and July	July..3½		
Illinois Central.................100	22,888,900	Feb. and Aug	Au5&10s	121½	132½
Indianapolis and Cincinnati 50	1,689,900	April and Oct	Oct...4	99	
New York and Boston Air Line.100	788,047				
New York Central..............100	24,386,000	Feb. and Aug	Aug..3	96⅞	96½
New York and Harlem 50	5,085,050				
do preferred............ 50	1,500,000	Jan. and July	Jan,..4		
Niagara Bridge & Canandaigua.100	1,000,000	Jan. and July	Jan...3		
New York and New Haven.....100	2,980,839	Quarterly.	Jan...4	114½	115
New York Providence & Boston100	1,508,000	Quarterly.	Jan...3	122½	125
Ninth Avenue...................100	795,360				
Northern of New Hampshire...100	3,068,400	June and Dec	Dec 4	88½	90
Northern Central............... 50	3,344,800	Quarterly.	Nov 2	88	90
North Pennsylvania............ 50	3,150,150			61	62
Norwich and Worcester........100	2,338,600	Jan. and July	July..4		100
Ogdensburg & L. Champlain...100	3,077,000			41	42
Ohio and Mississippi...........100	21,250,000			28⅝	28½
do preferred..100	2,979,000	January.	Jan ..7	70	
Old Colony and Newport.......100	3,609,600	Jan. and July	Jan...4	100	100
Oswego and Syracuse 50	482,400	Feb. and Aug	Aug..4		
Panama (and Steamship).......100	7,000,000	Quarterly.	Jan ..6	235	40
Peninsula......................100					
Pennsylvania................... 50	20,000,000	May and Nov	May. .5	113½	113¾
Philadelphia and BaltimoreCent100	218,100				
Philadelphia and Erie 50	5,013,054			60¾	61
Philadelphia and Reading 50	20,072,323		De.'65-10	106⅝	106⅞
Phila., Germant'n, & Norrist'n. 50	1,358,100	Apr. and Oct	Oct. .4	106½	106¾
Phila., Wilmington & Baltimore 50	8,657,300	Apr. and Oct	Oct ..5	119	120
Pittsburg and Connellsville..... 50	1,770,414				
Pittsburg, Ft. Wayne & Chicago100	8,181,126	Quarterly.	Jan...2½	105⅞	106
Portland, Saco, and Portsmouth100	1,500,000	Jan. and July	July..4	94	100
Providence and Worcester......100	1,700,000	Jan. and July	July..4½		
Racine and Mississippi.........100					
Raritan and Delaware Bay......100	2,360,700				
Reading and Columbia.......... 50	501,890				
Rensselaer and Saratoga 50	800,000	Jan. and July	July..4		
Rome, Watertown & Ogdensb'g100	1,774,175	Jan. and July	Jan...5		
Rutland and Burlington 100	2,233,376				
St. Louis, Alton, & Terre Haute100	2,300,000			39	43
do do pref.100	1,700,000	Annually.	May..7	70	
Sandusky, Dayton, and Cincin..100	2,989,090				
do do pref.100	354,866	Feb. and Aug	Aug..3		
Sandusky, Mansfield & Newark100	862,571				
Schuylkill Valley............... 50	576,000	Jan. and July	July..5		
Second Avenue (N. Y.).........100	650,000	Apr. and Oct		66	
Shamokin Valley & Pottsville.. 50	869,450	Feb. and Aug	Aug..3		70
Sixth Avenue (N. Y.)...........100	750,000	Quarterly.		130	136
Syracuse, Binghamton & N. Y.100	1,200,130				
Terre Haute and Richmond..... 50	1,900,150	Jan. and July	Jan...6		
Third Avenue (N. Y.)...........100	1,170,000	Quarterly.	Oct......		
Toledo, Peoria, and Warsaw...100	1,700,000				
do do 1st pref.100	1,700,000				
do do 2d pref.100	1,000,000				
Toledo, Wabash and Western.. 50	2,442,350	June and Dec	June.3	42	45
do do preferred. 50	984,700	June and Dec	Dec. .3½		65
Tioga..........................100	125,000	Jan. and July	July..3½		
Troy and Boston................100	607,111				
Troy and Greenbush............100	274,400	June and Dec	Dec ..3		
Utica and Black River..........100	811,560	Jan. and July	Jan ..4		
Vermont and Canada...........100	2,860,000	June and Dec	Dec ..4	94	95
Vermont and Massachusetts....100	2,214,225		Jan...2	43	44
Warren 50	1,408,300	Jan. and July	Jan...3	93½	95
Westchester and Philadelphia.. 50	684,036				
Western (Mass)................100	5,665,000	Jan. and July	Jan...4	137½	140
Worcester and Nashua.........83⅓	1,141,000	Jan. and July	July..3	100	
Wrightsville, York & Gettysb'g 50	317,050	Jan. and July	July..1		
Canal.					
Chesapeake and Delaware...... 25	1,343,563				
Chesapeake and Ohio 25	8,228,595				
Delaware Division.............. 50	1,633,350	Feb. and Aug	Aug..3	62	63
Delaware and Hudson 100	10,000,000	Feb. and Aug	Aug.10	144	146
Delaware Junction (Pa.).......100	398,910				
Delaware and Raritan..........100		Jan. and July	Jan...5		
Lancaster and Susquehanna.... 50	200,000				
Lehigh Navigation.............. 50	4,282,950	May and Nov	Nov..5	109	110½
Monongahela Navigation....... 50	726,800				
Morris (consolidated) 100	1,025,000	Feb. and Aug	Feb..6	82	83
do preferred 100	1,175,000	Feb. and Aug	Feb..5	120	121
North Branch 50	138,086				121
Schuylkill Navigation (consol.). 50	1,908,207			53	54
do preferred. 50	2,888,805	Feb. and Aug	Aug..3½	66	67
Susquehanna and Tide-Water.. 50	2,050,070			18	20
Union.......................... 50	2,750,000			2	5
do preferred 50				40	42
West Branch and Susquehanna.100	1,000,000	Jan. and July	July..5		
Wyoming Valley............... 50	700,000	May & No	Nov. 4	112	116
Miscellaneous.					
American Coal................. 25	1,500,000	Feb. and Aug	Aug..4		72
American Telegraph 100			10		

Figure A.2: Example Commercial and Financial Chronicle 1925

60 BANKS AND TRUST COMPANIES [VOL. 121.

NEW JERSEY—(Concluded)

	Capital.	Surplus & Profits.	Gross Deposits.	Par.	Bid.	Ask.
Long Branch—	$	$	$		Per	share
Citizens' Nat Bank	100,000	240,000	2,450,000	100	325	-----
Long Branch Bkg Co	150,000	187,557	2,429,316	50	200	220
Morristown—					Per	share.
First National Bank	200,000	373,678	5,013,494	100	t240	-----
National Iron Bank	200,000	214,482	5,289,541	50	t100	-----
American Trust Co	150,000	122,157	1,750,549	100	t140	-----
Morristown Trust Co	600,000	1,049,371	8,733,592	100	t275	-----
Mt. Holly—					Per	share.
Mt Holly Nat Bank	100,000	88,822	773,171	25	38	40
Union Nat Bank	100,000	216,386	1,387,085	50	135	150
Farmers' Trust Co	200,000	133,290	1,053,445	100	120	125
Mt Holly S D & Tr	100,000	179,324	709,007	100	135	150
Newark—						
American Nat Bank	500,000	1,095,436	15,598,126	100	425	-----
Broad & Market N B	200,000	550,196	8,235,724	100	425	-----
Cit N Bk & Tr Co	200,000	118,763	1,232,336	100	160	-----
Lincoln Nat Bank	600,000	159,331	1,622,704	100	220	-----
Mer & Mfrs N Bk	1,350,000	2,061,854	14,572,182	100	320	-----
Mutual Bk of Rosev	200,000	223,517	2,331,492	100	190	-----
Nat Newark & Essex Banking Co	2,500,000	1,630,707	32,634,935	100	280	-----
National State Bank	500,000	1,057,126	6,228,152	100	375	-----
North Ward Nat Bk	400,000	976,368	12,162,245	109	450	-----
City Trust Co	300,000	408,656	4,778,671	100	330	-----
Clinton Trust Co	400,000	465,188	7,692,076	100	315	-----
Federal Trust Co	2,500,000	1,851,032	23,339,292	100	410	-----
Fidelity Union Tr Co	5,250,000	5,209,189	76,248,383	100	525	-----
Ironbound Trust Co	500,000	779,117	13,832,704	100	500	-----
Liberty Trust Co	200,000	122,369	2,298,251	100	160	-----
Newark Trust Co	200,000	107,800	2,140,510	100	300	-----
So Side N B & T Co	200,000	g56,000	415,000	100	150	-----
Springfield Av Tr Co	200,000	427,755	7,593,437	100	400	-----
Vailsburg Trust Co	200,000	90,797	1,765,874	100	210	-----
Washington Trust Co	300,000	421,851	3,714,821	100	320	-----
Weequahic Trust Co	200,000	178,109	2,238,323	100	300	-----
West Side Trust Co	600,000	785,673	8,343,189	100	475	-----
New Brunswick					Per	share
Cits Nat Bk of N Br	250,000	g50,000	1,246,714	100	125	-----
Nat Bank of N J	500,000	1,003,631	12,808,754	100	325	350
Peoples Nat Bank	200,000	297,387	3,795,505	100	290	300
Middlesex TG&T Co	100,000	121,000	2,200,000	100	175	195
New Brunsw Tr Co	300,000	396,738	5,187,486	100	270	-----
North & West Hudson—						
First Nat Bk, Town of Union	150,000	g65,998	3,728,014	100	160	-----
First N Bk, West NY	100,000	g149,721	3,967,428	100	160	-----
Commonw'lth Tr Co	600,000	g629,160	7,621,753	100	300	-----
Guttenberg B & T Co	100,000	g135,409	3,992,625	100	350	-----
Weehawken Tr Co	600,000	g200,000	8,639,737	100	175	-----
Highland Trust Co	300,000	g191,962	4,832,712	100	240	-----
Passaic—					Per	share
Merchants Bank	100,000	125,907	1,897,899	100	-----	240
Passaic N Bk&Tr Co	1,500,000	2,091,996	22,994,385	100	-----	225
City Trust Co	200,000	257,832	4,213,134	100	-----	-----
Hobart Trust Co	300,000	343,226	3,854,754	100	450	-----
People's Bk & Tr Co	400,000	972,476	9,110,413	100	375	400
Service Trust Co	400,000	260,000	1,500,000	100	-----	180
Paterson—					Per	share
First National Bank	600,000	813,394	8,301,216	100	415	-----
Paterson Nat Bank	1,200,000	1,118,580	14,835,058	100	280	290
Second Nat Bank	750,000	999,992	12,166,076	50	225	-----
Nat Bank of Amer	500,000	332,129	3,709,408	100	200	205
Paterson Sav Inst	1,000,000	1,621,300	23,854,500	25	145	160
Citizens' Trust Co	500,000	714,699	9,569,369	100	360	-----
Franklin Trust Co	150,000	410,035	3,226,369	100	360	-----
Hamilton Trust Co	600,000	556,135	10,871 482	100	300	-----
U S Trust Co	350,000	1,390,651	18,496,652	100	580	-----
Plainfield—						
City National Bank	150,000	347,043	6,699,691	100	-----	t200
First National Bank	200,000	339,265	5,881,270	100	-----	t230
Plainfield Trust Co	609,300	857,954	12,986,123	100	-----	t225
State Trust Co	100,000	159,368	3,301,619	100	190	-----
Trenton—						
Broad St Nat Bank	250,000	953,085	10,330,515	100	400	-----
Capital City Tr Co	150,000	g140,676	1,228,624	100	200	-----
First National Bank	500,000	1,233,438	9,248,718	100	350	-----
Hanover Trust Co	200,000	167,454	3,354,419	100	200	-----
Mechanics' Nat Bk	1,000,000	1,655,157	17,593,422	50	260	-----
Trenton Banking Co	750,000	1,288,408	13,784,727	50	175	-----
Colonial Trust Co	100,000	188,382	3,587,887	100	225	-----
Mercer Trust Co	200,000	500,731	5,595,772	100	275	-----
Trenton Trust Co	750,000	1,132,009	10,902,440	100	225	-----
Wilbur Trust Co	100,000	50,495	1,641,538	100	165	-----

NEW YORK—(Continued).

	Capital.	Surplus & Profits.	Gross Deposits.	Par.	Bid.	Ask.
Buffalo—	$	$	$		Per	share
Liberty Bank	2,500,000	4,172,662	53,230,170	100	335	-----
Manuf'rs & Trad Nat	2,000,000	2,711,852	53,624,419	100	400	-----
People's Bank	1,000,000	1,167,735	25 598,859	100	270	-----
Com-So Side N Bk	550,000	431,334	11,609,081	100	240	-----
Buffalo Trust Co	2,500,000	3,482,175	55,854,318	100	440	450
Fidelity Trust Co	1,000,000	1,943,231	30,291,213	100	400	-----
Marine Trust Co	11,250,000	13,355,072	141,680,375	100	335	370
Elmira—					Per	share
Merchants' Nat Bk	250,000	253,582	3,150,691	100	225	-----
Second Nat Bank	400,000	914,874	8,185,544	100	305	-----
Chemung Can T Co	600,000	958,674	8,871,829	100	275	-----
New York City— *are of date Nov. for National and of banks and trust found in our "Rai*	*Deposits N 28 1925. Nov 14 19 companies lway and I*	*ew York Ci Surplus an '25 for Stat in New Yo ndustrial"*	*ty banks are d profits are e banks. A rk City and Section, pag*	*repo of da divi Broo e244*	*rted ne te Sept. dend r klyn m .*	*t and 28 '25 ecord ay be*
	$	$	$		Per	share
Amalg Bank of N Y	200,000	151,278	d5,536,715	----	-----	-----
Amer Ex-Pac Nat Bk	7,500,000	12,625,380	143,608,000	100	475	485
Amer Union Bank	1,200,000	429,509	d8,595,046	100	205	220
Bank of America	6,500,000	5,223,884	93,623,000	100	335	-----
Bank of Europe	450,000	488,199	d10,322,203	100	-----	-----
Bank of Manhat Co	10,000,000	14,732,900	154,488,000	50	225	230
Bank of U S	4,000,000	2,683,238	d68,214,064	100	285	295
Bank of Wash Hgts	200,000	604,380	9,067,000	100	725	-----
Berardini State Bank	150,000	783,788	d1,022,806	----	-----	-----
Bowery Nat Bank	250,000	928,200	5,229,000	100	850	950
Broadway Cent Bank	300,000	205,234	d6,235,910	100	245	-----
Bronx Borough Bank	150,000	727,172	7,547,266	100	675	-----
Bronx National Bank	300,000	252,000	a7,098,600	100	350	450
Bryant Park Bank	200,000	213,389	d2,462,989	100	210	230
Capitol Nat Bank	2,000,000	890,000	a20,075,600	100	215	225
Chase National Bank	20,000,000	29,008,050	392,103,000	100	563	568
Cent Mercantile Bk	1,500,000	1,105,009	d21,161,726	100	330	-----
Chatham Phenix Nat Bk & Tr Co	13,500,000	13,231,500	215,828,000	100	358	363
Chelsea Exch Bank	1,500,000	722,764	d16,094,510	100	215	225
Chemical Nat Bank	4,500,000	17,597,100	120,279,000	100	710	725
Coal & Iron Nat Bk	1,500,000	1,755,003	18,208,000	100	345	355
Colonial Bank	1,200,000	2,787,800	32,770,000	100	550	-----
Commonwealth Bank	600,000	1,089,659	13,677,000	100	355	370
Continental Bank	1,000,000	1,161,388	7,185,000	100	250	-----
Corn Exchange Bank	10,000,000	14,558,000	204,424,000	100	570	580
Cosmopolitan Bank	400,000	226,844	d8,887,671	100	190	-----
Eastern Exch Bank	100,000	28,450	d1,341,582	----	-----	-----
East River Nat Bank	2,500,000	2,375,200	43,992,000	100	355	370
Federation Bk of N Y	750,000	842,724	d11,484,828	----	-----	-----
Fifth Avenue Bank	500,000	2,905,364	25,288,000	100	2350	-----
First National Bank	10,000,000	71,199,679	222,732,000	100	2950	3000
Franklin Nat Bank	800,000	483,371	a4,877,214	----	160	170
Garfield Nat Bank	1,000,000	1,766,100	19,129,000	100	380	-----
Gimbel Bros Bank	100,000	106,600	e1,114,600	----	-----	-----
Grace Nat Bk of N Y	1,000,000	1,798,200	10,462,000	100	270	-----
Greenwich Bank	1,000,000	2,594,680	24,689,000	100	425	-----
Hamilton Nat Bank	1,000,000	350,000	a9,378,000	100	218	225
Hanover Nat Bank	5,000,000	25,443,000	110,082,000	100	1100	1120
Harriman Nat Bank	1,000,000	1,326,500	a39,540,500	100	480	490
Internat Union Bank	250,000	208,391	d3,547 414	----	-----	-----
Lebanon Nat Bank	250,000	80,400	a1,185,600	----	-----	-----
Liberty Nat Bank	1,500,000	662,900	a7,901,400	----	-----	-----
Longacre Bank	200,000	102,088	d4,853,153	----	-----	-----
Madison State Bank	200,000	47,177	d2,303,606	----	-----	-----
Mech & Met Nat Bk	10,000,000	17,380,479	183,668,000	100	418	425
Mutual Bank	500,000	764,600	e15,706,600	100	465	-----
Nat American Bank	1,000,000	603,700	a8,025,600	100	175	185
Nat Butch & Drov	1,000,000	863,400	a9,204,100	25	170	180
Nat Bk of Commerce	25,000,000	40,021,600	327,280,000	100	365	370
National City Bank	50,000,000	63,149,175	726,665,000	100	585	595
National Park Bank	10,000,000	24,375,400	137,807,000	100	505	515
New Netherland Bk	600,000	372,538	d13,319,761	100	270	280
Penn Exchange Bank	200,000	55,061	d2,327,328	100	124	134
Peoples Comm'l Bk	100,000	57,393	d2,437,114	100	-----	-----
Port Morris Bank	100,000	100,422	d2,514,698	----	200	-----
Prisco State Bank	150,000	69,200	e1,543,500	----	-----	-----
Public Nat Bank	4,000,000	6,702,700	a104636 000	100	625	650
Seaboard Nat Bank	5,000,000	8,558,441	125,096,000	100	690	710
Seventh Nat Bk	k1,000,000	88,100	a4,664,200	100	170	180
Standard Bank	200,000	243,987	d5,423 897	100	480	500
State Bank	3,500,000	5,867,562	105,076,000	100	775	825
Trade Bank of N Y	500,000	241,085	d2,909,827	----	145	155
United Nat Bk in NY	1,000,000	571,500	a11,209,300	----	210	225
World Exch Bank	100,000	42,400	e1,227,500	----	-----	-----
Trust Co. returns da	*te Nov. 14*	1925				
American Trust Co	4,000,000	3,009,770	45,120,502	100	-----	-----
Anglo South Am Tr	1,000,000	584,061	9,264,075	----	-----	-----
Banca Com It Tr Co	1,000,000	584,332	8,999,515	----	-----	-----
Banco di Sicilia Tr Co	500,000	15,133	1,681,798	----	-----	-----
Bank of N Y & Tr Co	4,000,000	12,8C7,853	71,844,790	100	625	635
Bankers Trust Co	20,000,000	30 391,589	372,277,384	100	573	578
Bronx County Tr Co	825,000	420,741	11,471,796	100	230	255
Brotherhood Locom Eng Co-Op Tr Co	700,000	259,958	6,062,628	----	-----	-----

Figure A.3: U.S. stock market returns: 1866-1926. The figure shows the cumulative value of a one-dollar investment from 1866 through 1926 in the U.S. stock market. Shown are the value-weighted ('VW') or equal-weighted ('EW') cumulative U.S. market stock return as constructed in this paper ("Market"), the index of Schwert (1990; "Schwert") and the index of Goetzmann et al. (2001; "Goetzmann et al."). The y-axis is on a logarithmic scale.

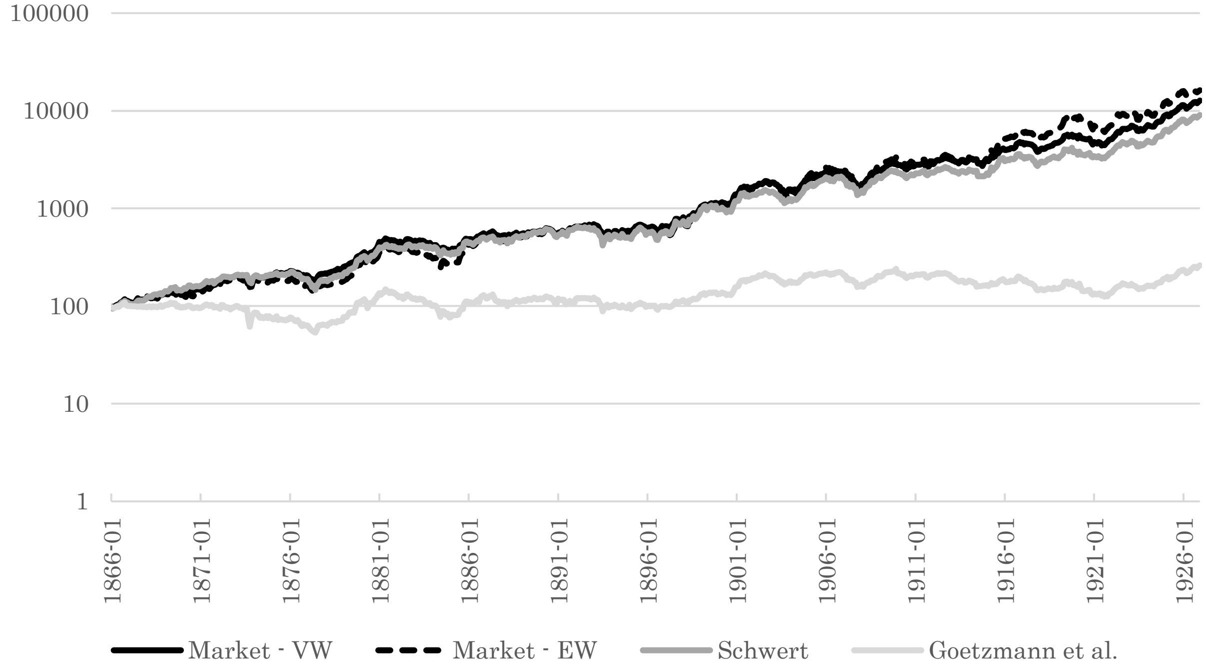


Figure A.4: Distribution of market capitalization of U.S. stocks: 1866-1926 versus 1927-2019. The figure shows the time-series average of the monthly cross-sectional distribution of (the natural logarithm of) stocks' market capitalizations for our sample (1866-1926; 'Pre-CRSP') and the CRSP sample (1927-2019; 'CRSP').

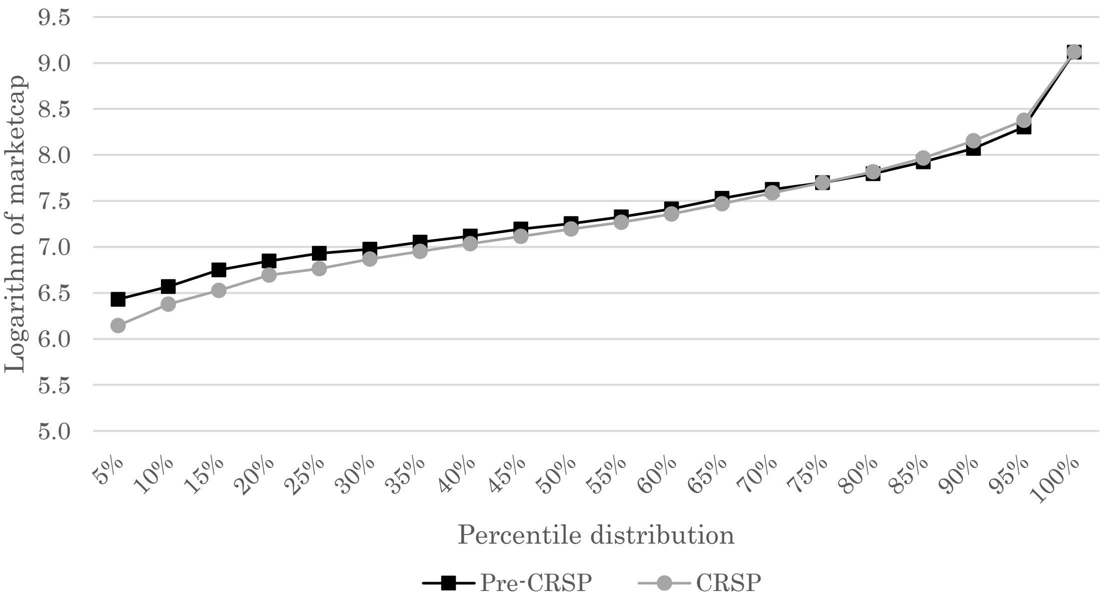


Figure A.5: Distribution of dividend paying stocks: 1866-1926. The figure shows per month in our sample the number of dividend payers versus zero-dividend versus negative dividend stock, split per small (market capitalization below median) and large (market capitalization above median) stocks. The sample runs from 1866 till 1926.

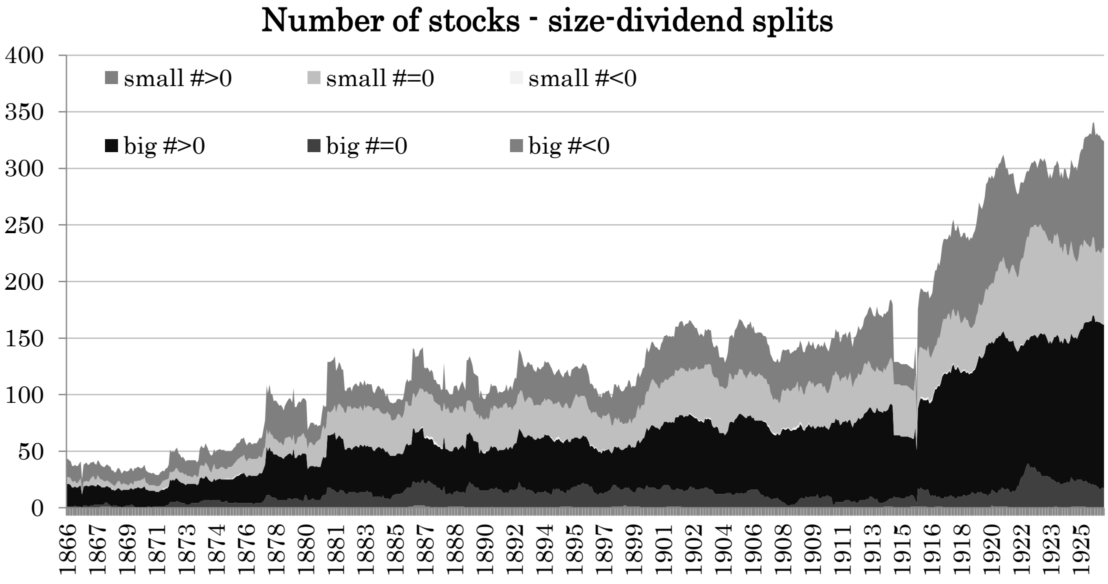


Figure A.6: Distribution of share issuance: 1866-1926. The figure shows per month in our sample the number of stocks with positive, zero or negative share issuance, split per small (market capitalization below median) and large (market capitalization above median) stocks. The sample runs from 1866 till 1926.

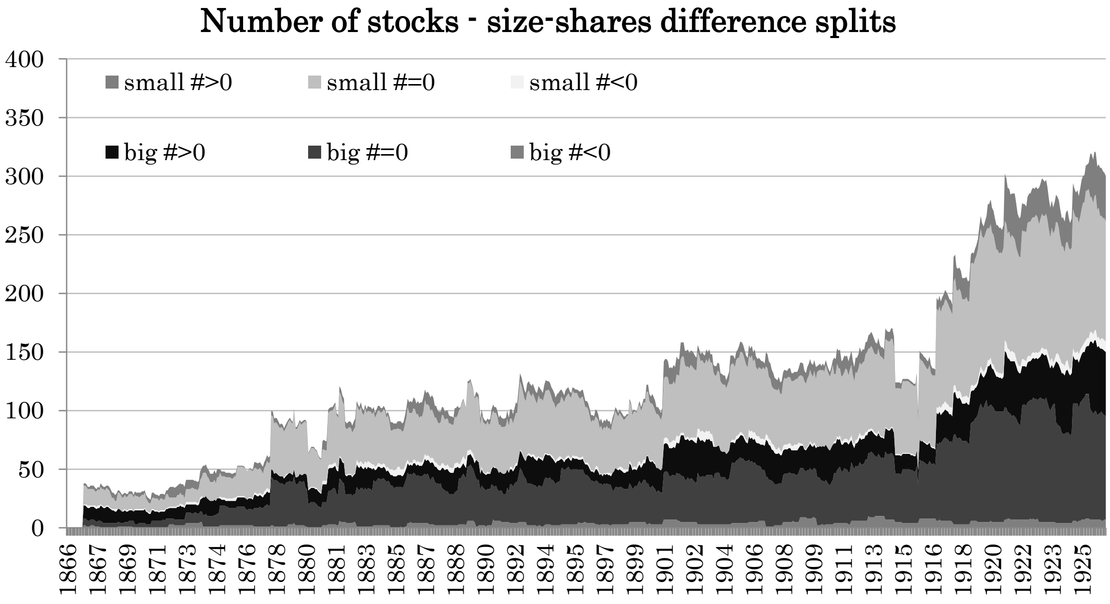


Figure A.7: Cross-sectional distribution of characteristic variables: 1866-2019. The figure shows per month in our sample the 20th (bottom black line), 50th (grey line), and 80th (top black line) percentiles for several key characteristics. The sample runs from 1866 till 2019, with the dotted vertical line indicating the pre-CRSP versus CRSP sample cutoff date.

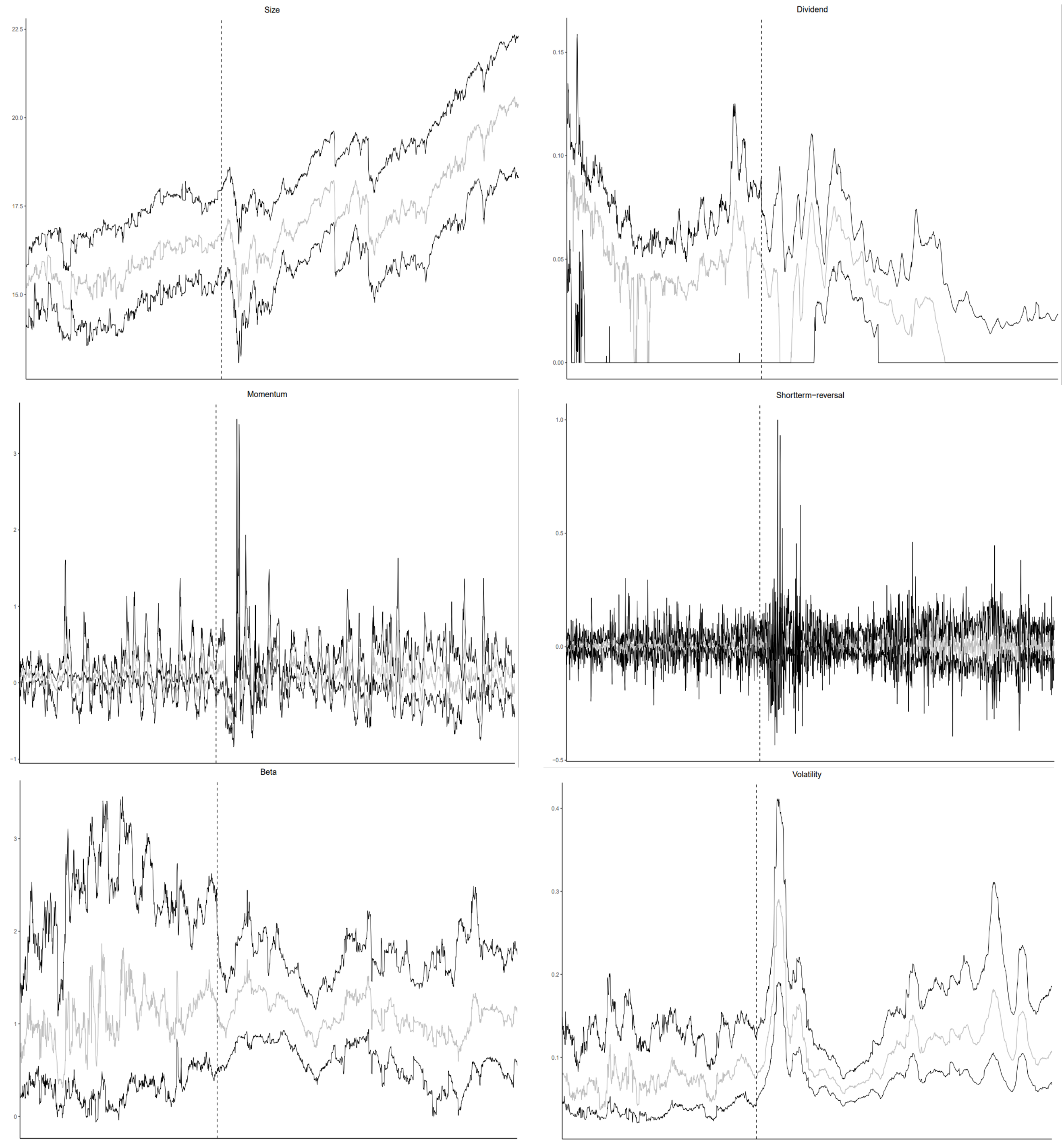


Figure A.8: U.S. equity factor returns: 1866-1926. The figure shows the cumulative value of a one-dollar investment from 1866 through 1926 for the size ('SMB'), value ('HML'), momentum ('UMD'), short-term reversal ('ST_REV'), and low-risk ('BETA') factors. Factors are constructed from top-bottom portfolios from 2x3 size-characteristic-based portfolios. The y-axis is on a logarithmic scale.

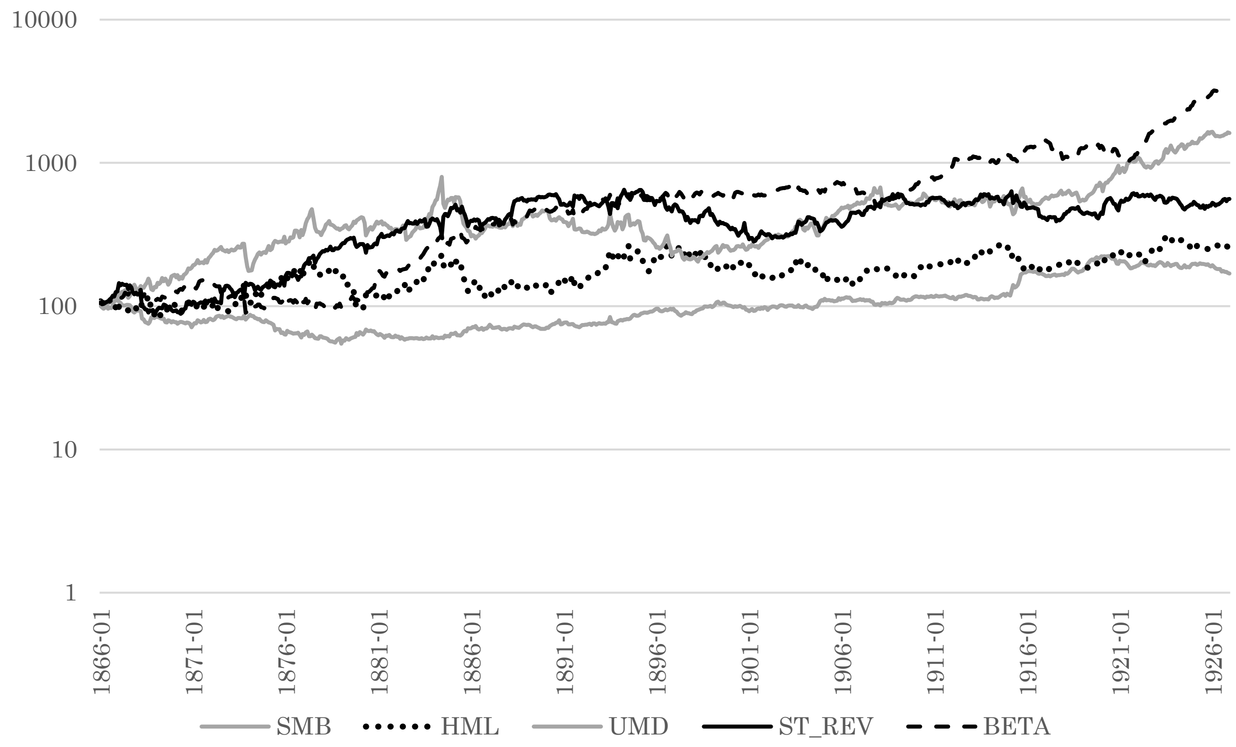


Table A.3: Sample summary statistics

The table summarizes the return series we use in our sample. Shown are the average annualized total return and volatility ('Std. deviation') of the value-weighted ('VW') or equal-weighted ('EW') market index as constructed in this paper, the equally weighted index of Schwert (1990), and the price-weighted index of the price appreciation (i.e., excluding dividends) on NYSE stocks of Goetzmann et al. (2001). Results are shown per calendar decade and over our full sample period (1866-1926). The last row shows the results over the CRSP sample period (1927-2019) based on the value-weighted market index from CRSP.

	Our sample - VW		Our sample - EW		Schwert (1990)		Goetzmann, Ibbotson and Peng (2001)	
Year	Average return	Std. deviation	Average return	Std. deviation	Average return	Std. deviation	Average return	Std. deviation
1866-1869	6.98	10.59	5.21	11.40	9.54	9.41	-0.59	7.77
1870s	9.88	11.03	8.75	13.24	8.03	11.91	2.49	16.74
1880s	6.53	11.93	8.44	15.08	7.38	13.22	2.14	16.20
1890s	6.97	11.99	7.44	15.59	6.91	17.87	2.03	13.13
1900s	10.85	13.30	13.03	15.29	10.39	15.25	6.63	10.69
1910s	6.93	10.97	10.31	13.44	5.49	13.40	-3.28	10.14
1920-1926	12.74	12.19	10.57	16.26	12.59	12.77	6.86	11.88
1866-1926	8.67	11.80	9.42	14.54	8.33	13.97	2.39	13.14
1927-2019	11.24	18.44						

Table A.4: Impact sample filters

The table shows the number of unique stock observations included in our sample before the various data quality filters (see Table A.2) and before and after the data quality screens, the percentage of stocks included in our final sample relative to the sample before data quality filters (i.e., 'Stocks with MV'), and the percentage of market capitalization ('MV') included. Results are per December of the start year of every 10-year period and over our full sample period (1866-1926).

Year	All	Domestic	Common stock	Non-OTC	Qualified stock	Stocks with MV	Data quality screens	% of stocks included	MV of stocks included (%)
1866	396	393	253	237	83	54	54	100.0%	70.4%
1876	1,223	1,218	880	632	275	123	69	56.1%	82.4%
1886	1,138	1,124	984	733	392	278	183	65.8%	77.4%
1896	2,514	2,496	2,312	860	540	455	180	39.6%	70.4%
1906	3,152	3,099	2,691	1,011	566	478	206	43.1%	76.2%
1916	3,570	3,444	2,775	1,432	613	485	257	53.0%	83.7%
1926	4,401	4,159	3,207	1,675	651	607	407	67.1%	87.4%
1866 - 1926	12,369	11,904	8,765	4,819	1,872	1,488	1,154	77.6%	

Table A.5: Sample distribution: sectors

The table shows the number of unique stock observations included in our sample split per sector and over all sectors combined ('Total'). Results are shown before (Panel A) and after (Panel B) the data quality screens. Panel C (Panel D) shows the (relative percentage) market capitalization composition (in millions of U.S. Dollars) of the stocks included in Panel B. Results are per December of the start year of every 10-year period and over the average over all months in our full sample period (1866-1926).

Panel A: Number of stocks - pre-data quality screens

Year	Energy/Mining	Financials	Industrial & Other	Infrastructure	Utilities	Total
1866	7	13	-	33	1	54
1876	5	42	1	65	10	123
1886	6	126	2	137	7	278
1896	14	284	20	112	25	455
1906	40	265	55	90	28	478
1916	63	226	87	77	32	485
1926	111	249	159	68	20	607
Average	35	158	50	91	18	344

Panel B: Number of stocks - final sample

Year	Energy/Mining	Financials	Industrial & Other	Infrastructure	Utilities	Total
1866	7	13	-	33	1	54
1876	3	8	1	52	5	69
1886	5	57	2	113	6	183
1896	10	54	15	83	18	180
1906	33	42	47	64	20	206
1916	58	39	76	56	28	257
1926	104	82	149	54	18	407
Average	30	37	43	67	14	185

Panel C: Market capitalization composition - final sample

Year	Energy/Mining	Financials	Industrial & Other	Infrastructure	Utilities	Total
1866	8	32	-	155	1	195
1876	7	43	5	505	10	571
1886	4	129	22	1,064	37	1,256
1896	38	144	185	931	164	1,463
1906	1,487	262	990	3,091	582	6,412
1916	3,265	414	1,982	2,997	998	9,656
1926	4,145	1,470	4,460	3,925	2,407	16,406
Average	985	240	915	1,616	434	4,040

Panel D: Relative market capitalization composition - final sample

Year	Energy/Mining	Financials	Industrial & Other	Infrastructure	Utilities	Total
1866	4%	16%	-	79%	1%	100%
1876	1%	8%	1%	89%	2%	100%
1886	0%	10%	2%	85%	3%	100%
1896	3%	10%	13%	64%	11%	100%
1906	23%	4%	15%	48%	9%	100%
1916	34%	4%	21%	31%	10%	100%
1926	25%	9%	27%	24%	15%	100%
Average	13%	9%	13%	60%	7%	100%

Table A.6: Sample distribution: exchanges

The table shows the number of unique stock observations included in our sample split per exchange (NYSE, Curb, or regional exchanges) and over all exchanges combined ('Total'). Results are shown before (Panel A) and after (Panel B) the data quality screens. Panel C (Panel D) shows the (relative percentage) market capitalization composition (in millions of U.S. Dollars) of the stocks included in Panel B. Results are per December of the start year of every 10-year period and over the average over all months in our full sample period (1866-1926).

Panel A: Number of stocks - pre-data quality screens

Year	NYSE	Curb	Regional	Total
1866	42	-	12	54
1876	96	1	26	123
1886	183	2	93	278
1896	222	8	225	455
1906	232	15	231	478
1916	247	19	219	485
1926	348	25	234	607
Average	198	9	137	344

Panel B: Number of stocks - final sample

Year	NYSE	Curb	Regional	Total
1866	42	-	12	54
1876	47	-	22	69
1886	121	-	62	183
1896	113	-	67	180
1906	132	3	71	206
1916	171	12	74	257
1926	292	14	101	407
Average	128	11	53	185

Panel C: Market capitalization composition - final sample

Year	NYSE	Curb	Regional	Total
1866	182	-	13	195
1876	491	-	79	571
1886	1,083	-	173	1,256
1896	1,165	-	298	1,463
1906	4,515	1	1,897	6,412
1916	6,847	105	2,704	9,656
1926	12,604	239	3,563	16,406
Average	2,985	163	988	4,040

Panel D: Relative market capitalization composition - final sample

Year	NYSE	Curb	Regional	Total
1866	93%	-	7%	100%
1876	86%	-	14%	100%
1886	86%	-	14%	100%
1896	80%	-	20%	100%
1906	70%	0%	30%	100%
1916	71%	1%	28%	100%
1926	77%	1%	22%	100%
Average	80%	1%	19%	100%

Online Appendix B: Data quality analyses

Table B.1: Fama-MacBeth regression results (equal-weighted)

This table presents coefficient estimates from monthly Fama-MacBeth (1973) regressions of excess returns between month *t* and *t+1* against a constant and a series of stock characteristics, as described in Section III. Stock characteristics are measured at the end of month *t* over our sample period from January 1866 to December 1926. We report slope coefficients (multiplied by 100) with standard t-statistics in parentheses, the R^2 of the regressions ("R^2"), and the number of observations ("No. of obs."). Observations are equal-weighted. *, ** and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Constant	0.69***	2.05**	0.73***	0.65***	0.69***	2.09***
t	(8.51)	(2.56)	(3.51)	(4.65)	(4.87)	(3.45)
Beta	0.08					0.08
t	(0.88)					(0.81)
ln(Size)		-0.08*				-0.10***
t		(-1.76)				(-2.72)
Dividend			1.09			2.78***
t			(0.75)			(2.90)
Momentum				0.55		0.66**
t				(1.56)		(2.17)
ST Reversal					-4.41***	-6.56***
t					(-4.38)	(-6.96)
R^2	0.09	0.02	0.03	0.05	0.06	0.20
No. of obs.	101,388	101,949	100,604	100,604	101,892	100,604

Table B.2: Fama-MacBeth regression results including share issuance

This table presents coefficient estimates from monthly Fama-MacBeth (1973) regressions of excess returns between month *t* and *t+1* against a constant and a series of stock characteristics, as described in Section III. Stock characteristics are measured at the end of month *t* over our sample period from January 1866 to December 1926. We report slope coefficients (multiplied by 100) with standard t-statistics in parentheses, the R^2 of the regressions ("R^2"), and the number of observations ("No. of obs."). Observations are value-weighted. *, ** and *** indicate significance at the 10%, 5%, and 1% level, respectively.

	(1)	*(2)*
Constant	0.81***	1.04*
t	(5.92)	(1.82)
Beta		0.10
t		(1.07)
ln(Size)		-0.03
t		(-0.98)
Dividend		2.01**
t		(2.13)
Momentum		0.85***
t		(2.89)
ST Reversal		-3.85***
t		(-4.05)
D(Issuance=0)	-0.11	-0.10
t	(-1.34)	(-1.33)
Issuance	-0.92**	-0.74**
t	(-2.22)	(-2.02)
R^2(%)	0.05	0.28
No. of obs.	92,857	92,857

Table B.3: Robustness of equity factors: data quality filters

The table summarizes the robustness test results to screens and controls on data quality of equity characteristic portfolio sorts. We consider the following variations: the combination of the zero return, the return interpolation and stale return screens ("Baseline"), the addition of trimming individual stock returns at -50% and +50% on the Baseline ("Trimming extreme returns"), applying only the zero return liquidity screen ("Zero return screen"), applying a loser version of the zero-return screen allowing for a maximum of 3 out of 12 zero monthly returns ("Zero return screen (3/12) screen"), no liquidity screens (hence including all stocks in the portfolio sorts; "No liquidity screen"), and the application of a one-month additional lag between signal and portfolio formation ("1-month lag"). The table presents average annualized excess returns (Panel A), and CAPM alphas (Panel B) of the high-low for each characteristic-sorted portfolio. Portfolios are value-weighted. The sample runs from January 1866 to December 1926 and is at the monthly frequency. Numbers in parentheses indicate standard t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level, respectively.

Panel A: Return spread

	Size	Value	Momentum	ST Reversal	BETA
Baseline	1.17	2.76	6.13***	4.10**	6.63***
t	(1.15)	(1.40)	(2.76)	(1.98)	(4.15)
Trimming extreme return	0.41	3.75**	6.88***	3.56*	6.90***
t	(0.42)	(2.00)	(3.26)	(1.81)	(4.43)
Only zero return screen	1.17	2.78	6.14***	4.11**	6.62***
t	(1.16)	(1.41)	(2.77)	(1.98)	(4.15)
Zero return screen (3/12) screen	1.27	2.06	6.40***	5.34***	6.25***
t	(1.26)	(1.05)	(3.00)	(2.86)	(3.99)
No liquidity screen	1.42	2.15	6.16***	5.30***	4.41***
t	(1.39)	(1.25)	(3.58)	(3.41)	(2.67)
1-month lag	0.93	2.80	4.31**	0.06	6.27***
t	(0.90)	(1.45)	(2.01)	(0.03)	(3.76)

Panel B: CAPM alpha

	Size	Value	Momentum	ST Reversal	BETA
Baseline	1.11	7.11***	9.02***	2.54	7.87***
t	(1.09)	(5.04)	(4.42)	(1.25)	(5.05)
Trimming extreme return	0.45	7.95***	9.64***	2.11	8.17***
t	(0.46)	(5.98)	(4.98)	(1.09)	(5.39)
Only zero return screen	1.11	7.13***	9.03***	2.54	7.87***
t	(1.09)	(5.05)	(4.43)	(1.25)	(5.05)
Zero return screen (3/12) screen	1.56	6.84***	9.33***	3.85**	7.99***
t	(1.55)	(5.05)	(4.76)	(2.10)	(5.35)
No liquidity screen	3.07***	5.98***	8.27***	4.22***	7.63***
t	(3.46)	(4.86)	(5.18)	(2.76)	(5.84)
1-month lag	0.84	7.11***	6.85***	-1.98	7.46***
t	(0.81)	(5.09)	(3.41)	(-1.06)	(4.55)

Online Appendix C: Additional results

Table C.1: Momentum and low-risk equity factors

The table summarizes the results of portfolio sorts based on various measures of momentum (Panel A; total return momentum, and price momentum) or low-risk (Panel B; volatility, idiosyncratic volatility, and beta). Volatility (idiosyncratic volatility) is measured by the standard deviation of the excess returns (beta-corrected excess returns) of the last 36 months, requiring a minimum of 12 observations. Beta is estimated over a 36-month window, requiring a minimum of 12 observations. We show results from the following sorting procedures: quintile portfolios ("Quintile"), as in Table III, tercile portfolios ("Tercile"), decile portfolios ("Decile"), 2x3 size-characteristic sorted portfolios ("2x3"), as in Table IV, and 2x5 size-characteristic sorted portfolios based on every 20th percentile breakpoint ("2x5"). The table presents average annualized excess returns spreads ("Return spread") and CAPM alphas ("CAPM alpha") of the high-low for each characteristic-sorted portfolio, each leg we lever based for the market beta following the procedure of the BETA factor construction. Portfolios are value-weighted. The sample runs from January 1866 to December 1926 and is at the monthly frequency. Numbers in parentheses indicate standard t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level, respectively.

Panel A: Momentum

	Momentum		Price Momentum	
	Return spread	CAPM alpha	Return spread	CAPM alpha
Quintile	8.18***	11.53***	5.09*	7.80***
t	(2.77)	(4.16)	(1.85)	(2.95)
Tercile	5.19**	7.69***	4.27**	6.33***
t	(2.44)	(3.87)	(2.07)	(3.21)
Decile	6.08	10.56**	4.83	8.09**
t	(1.48)	(2.71)	(1.22)	(2.10)
2X3	6.13***	9.02***	5.00**	7.35***
t	(2.76)	(4.42)	(2.31)	(3.59)
2X5	5.51**	8.75***	4.91*	7.72***
t	(2.01)	(3.42)	(1.84)	(3.05)

Panel B: Low-risk

	Volatility		Idiosyncratic volatility		Beta	
	Return spread	**CAPM alpha**	**Return spread**	**CAPM alpha**	**Return spread**	**CAPM alpha**
Quintile	3.89**	4.52***	3.99**	3.99**	4.83***	6.73***
t	(2.52)	(2.93)	(2.54)	(2.52)	(2.47)	(3.59)
Tercile	4.40***	4.40***	3.01**	3.01**	5.87***	6.64***
t	(3.39)	(3.37)	(2.49)	(2.48)	(3.54)	(4.01)
Decile	3.91**	6.03***	6.58***	6.58***	6.17**	8.31***
t	(2.01)	(3.28)	(3.45)	(3.42)	(2.36)	(3.27)
2X3	5.22***	6.02***	4.73***	4.80***	6.63***	7.87***
t	(3.84)	(4.47)	(3.74)	(3.77)	(4.15)	(5.05)
2X5	4.66***	6.34***	4.88***	5.52***	5.68***	7.57***
t	(3.04)	(4.39)	(3.38)	(3.84)	(3.10)	(4.33)

Table C.2: Stock factor returns in 'good' and 'bad' states

The table summarizes the historical performance of stock factors across 'good' and 'bad' states based on macroeconomic and market sub-periods. Sub-periods examined are at the annual frequency and include recession versus non-recession, and bear and bull equity markets. Shown are historical (annualized) market-adjusted returns per macroeconomic state for each stock factor. The column "Dif." contains the differential factor returns between bad and good states. We estimate results separately over the pre-CRSP, CRSP and combined samples. The pre-CRSP sample starts in January 1866 and ends December 1926. The CRSP sample runs from January 1927 till December 2019. The combined sample runs from January 1866 till December 2019. Both samples are at the monthly frequency. Numbers in parentheses indicate t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level, respectively.

		Recession/expansion				Bear/bull market			
		Rec.	**Exp.**	**Diff.**	***t***	**Bear**	**Bull**	**Diff.**	***t***
Size	1866-1926	-1.47	3.78	-5.26**	(-2.60)	-1.06	2.17	-0.60	(0.47)
	1927-2019	3.41	0.72	2.70	(0.92)	-0.20	1.77	-0.10	(0.16)
	1866-2019	0.16	1.47	-1.31	(-0.77)	-0.23	1.65	-0.16	(0.17)
Value	1866-1926	8.58	5.60	2.98	(1.06)	4.73	8.28	1.93***	(3.59)
	1927-2019	0.43	5.79	-5.36	(-1.76)	2.22	6.01	1.06*	(1.78)
	1866-2019	5.73	5.49	0.24	(0.12)	4.00	6.28	2.44**	(2.04)
Momentum	1866-1926	9.87	8.13	1.74	(0.43)	2.97	11.96	0.84***	(3.22)
	1927-2019	2.20	12.78	-10.57**	(-2.48)	6.79	12.76	2.31***	(4.21)
	1866-2019	7.21	11.28	-4.07	(-1.47)	5.79	11.96	2.54***	(3.51)
ST Reversal	1866-1926	6.98	-2.05	9.03**	(2.25)	12.30	-2.23	3.52*	(1.78)
	1927-2019	15.95	6.25	9.69***	(2.94)	9.63	7.18	4.23***	(3.56)
	1866-2019	10.07	4.03	6.03**	(2.52)	10.40	3.82	5.27*	(1.92)
BETA	1866-1926	6.64	9.14	-2.49	(-0.81)	-1.60	12.49	-0.60***	(5.84)
	1927-2019	-0.28	5.69	-5.97**	(-2.43)	2.30	5.68	1.36**	(2.46)
	1866-2019	4.24	6.51	-2.27	(-1.24)	1.14	7.93	0.76***	(3.26)

Table C.3: Momentum crashes

The table summarizes the results of momentum crash analyses. Panel A reports the return distribution of Momentum (UMD). Panel B reports the results of regression specification (3) of Daniel and Moskowitz (2016). Panel C reports the return distribution of Momentum volatility-scaled momentum (UMD*) of Barroso and Santa-Clara (2015). Returns ('Ret.') and Sharpe ratios ('SR') are annualized, other numbers (skewness; 'Skew.', kurtosis; 'Kurt.', minimum return; 'Min.') are monthly. 'Min.*' represent the minimum monthly return of UMD* scaled to the same volatility of UMD. We estimate results separately over the pre-CRSP, CRSP and combined samples. The pre-CRSP sample starts in January 1866 and ends December 1926. The CRSP sample runs from January 1927 till December 2019. The combined sample runs from January 1866 till December 2019. Both samples are at the monthly frequency. Numbers in parentheses indicate t-values. Asterisks are used to indicate significance at a 10% (*), 5% (**) or 1% (***) level, respectively.

Panel A: Return distribution Momentum

		Ret. (%)	SR	Skew.	Kurt.	Min. (%)	Min*. (%)
UMD	1866-1926	6.12	0.35	-0.98	5.15	-34.85	-34.85
	1927-2019	8.64	0.53	-2.74	24.53	-49.23	-49.23
	1866-2019	7.68	0.46	-1.98	15.83	-49.23	-49.23

Panel B: Market timing regression results Momentum

		a_0	α_B	β_0	β_{B}	$\beta_{\mathrm{B},U}$	R^2
UMD	1866-1926	0.87***	1.24***	-0.06	-1.21***	-0.53***	0.37
		(4.73)	(2.78)	(-0.95)	(-13.71	(-5.31)	
	1927-2019	0.88***	1.19***	0.07**	-0.68***	-0.31***	0.30
		(6.53)	(2.95)	(2.17)	(-14.93)	(-6.55)	
	1866-2019	0.88***	0.61**	0.04	-0.78***	-0.28***	0.29
		(7.88)	(2.13)	(1.44)	(-18.92)	(-6.73)	

Panel C: Return distribution volatility-scaled Momentum

		Ret. (%)	SR	Skew.	Kurt.	Min. (%)	Min*. (%)
UMD*	1866-1926	5.88	0.46	-0.38	6.67	-21.22	-28.91
	1927-2019	10.80	0.86	-0.71	5.57	-27.24	-35.74
	1866-2019	8.88	0.70	-0.58	5.95	-27.24	-36.20

Table C.4: Downside risk and stock factor returns

The table shows the beta and downside betas of the six stock factors returns. The downside beta is calculated versus the excess return of the equity market portfolio and uses three downside risk thresholds: zero, -1 standard deviation (1 sigma) and -2 standard deviation (2 sigma). The difference between regular CAPM beta (β) and downside beta (β^-) indicates the amount of additional downside risk. The annualized CAPM alpha (α) and downside risk CAPM alphas (α^-), both in percent with t-stats added. We estimate results over the combined pre-CRSP and CRSP samples at the monthly frequency. The combined sample runs from December 1866 till December 2019.

Factor	DR threshold	β^-	β	$\beta^- - \beta$	α	t-stat	α^-	t-stat
Size	zero	0.14	0.16	-0.014	1.00	1.27	1.09	1.39
	1 sigma	0.15	0.16	-0.009	-	-	1.05	1.34
	2 sigma	0.17	0.16	-0.015	-	-	0.89	1.13
Value	zero	-0.52	-0.59	0.064	5.34	5.84	4.91	5.34
	1 sigma	-0.47	-0.59	0.115	-	-	4.56	4.92
	2 sigma	-0.40	-0.59	0.190	-	-	4.05	4.27
Momentum	zero	-0.22	-0.35	0.137	10.04	7.91	9.11	7.11
	1 sigma	-0.22	-0.35	0.136	-	-	9.11	7.11
	2 sigma	-0.24	-0.35	0.112	-	-	9.28	7.26
ST Reversal	zero	0.18	0.18	-0.002	5.75	5.21	5.76	5.23
	1 sigma	0.18	0.18	0.004	-	-	5.72	5.19
	2 sigma	0.14	0.18	-0.034	-	-	5.98	5.42
BETA	zero	0.00	-0.05	0.053	5.72	6.83	5.36	6.37
	1 sigma	0.03	-0.05	0.089	-	-	5.11	6.04
	2 sigma	0.08	-0.05	0.131	-	-	4.83	5.64

Online Appendix D: Machine learning tests

Table D.1: Machine learnings hyperparameters

The table summarizes the hyperparameters used in the Random Forest (RF) and Neural Network with three hidden layers models (NN).

	RF	NN3
Prediction evaluation	Binary Cross-Entropy	Binary Cross-Entropy
Hyper parameters	Depth = 3	Dynamic learning rate
	#Trees = 100	starting at 0.005
	#Features in each split = 9	decreasing after 10 epochs
		Batch size = 128
		Epochs = 100
		Patience = 5
		Adam Para. = default
		Ensemble = 10

Figure D.1: Variable importance by machine learning model: 1866-1926. The figure shows the most influential variables in each machine learning model: Random Forest (RF) and Neural Network with three hidden layers (NN3). Variable importance is the average over all training samples and within each model normalized to sum to one. The sample runs from 1866 to 1926.