Stocks for the long run? Evidence from a broad sample of developed markets[⋆]

Aizhan Anarkulova[a], Scott Cederburg[a], Michael S. O'Doherty[b,*]

[a] *Eller College of Management, University of Arizona, McClelland Hall, Room 315R, Tucson, AZ 85721, USA*
[b] *Trulaske College of Business, University of Missouri, 401C Cornell Hall, Columbia, MO 65211, USA*

Abstract

We characterize the distribution of long-term equity returns based on the historical record of stock market performance in a broad cross section of 39 developed countries over the period from 1841 to 2019. Our comprehensive sample mitigates concerns over survivor and easy data biases that plague other work in this area. A bootstrap simulation analysis implies substantial uncertainty about long-horizon stock market outcomes, and we estimate a 12% chance that a diversified investor with a 30-year investment horizon will lose relative to inflation. The results contradict the conventional advice that stocks are safe investments over long holding periods.

Keywords: Long-horizon stock returns, loss probability, survivor bias, easy data bias

JEL classification: C58, G10, G11, G12, G15, N20

[⋆] We thank two referees, Alice Bonaimé, David Brown, Stephen Brown, Luke DeVault, Ryan Flugum, William Grieser, Travis Johnson, Chris Lamoureux, Andrea Rossi, Florent Rouxelin, Richard Sias, Michael Sposi, Mitch Towner, Tiemen Woutersen, Dexin Zhou, and seminar participants at Florida International University, Rutgers University, and the 2020 Paris December Finance Meeting for helpful comments and suggestions. This research was supported by a grant from the Robert J. Trulaske, Sr. College of Business Large Grant Program.

[*] Corresponding author.

E-mail addresses: `aanarkul@email.arizona.edu` (Aizhan Anarkulova), `cederburg@email.arizona.edu` (Scott Cederburg), `odohertym@missouri.edu` (Michael S. O'Doherty).

1. Introduction

Investing in stocks appears to be a very attractive option for long-term investors. The historical equity premium in the United States is quite large (Mehra and Prescott, 1985). Conventional wisdom is that stocks are safe over long holding periods, and long-term loss realizations in the U.S. have been infrequent or non-existent (see, e.g., Siegel, 2014). Fama and French (2018a) estimate a low probability of loss and a high probability of substantial gain for investors with horizons of 20 or 30 years. Provided with this evidence, common investment advice is that young investors with long horizons should invest heavily in stocks. A realized loss over a long investment period, however, is particularly damaging to an investor because it consumes much of her viable saving period without generating wealth. That is, each of us has only one lifetime in which to save for retirement, and poor stock market performance during our working years could lead to a substantial retirement savings shortfall with little time or ability to recover. As such, long-term investors must be wary of the potential for poor performance over their investment horizon.

Although the U.S. historical record is reassuring, there is reason to be cautious when assessing downside risk. The U.S. return history is short. The commonly used sample from the Center for Research in Security Prices spans less than 100 years, which offers limited statistical information about what happens over 30-year horizons. Extending the U.S. sample backward only deepens concerns about survivor bias (Brown et al., 1995). Further, evidence of unexpected increases in equity valuations suggests that the historical performance from the U.S. may be optimistic relative to ex ante expectations (Fama and French, 2002; Avdis and Wachter, 2017). One need not look far for examples of long-term losses in other developed markets. At the close of 1989, Japan's stock market was the largest in the world in terms of aggregate market capitalization.[1] Over the subsequent 30 years from 1990 to 2019, a diversified investment in Japanese stocks produced returns (inclusive of dividends) of -9% in nominal terms and -21% in real terms. Japan's experience is not unique, and several developed countries have realized worse performance or even complete stock market failure (see, e.g., Jorion and Goetzmann, 1999). Given the stark consequences of long-lived poor performance, we seek quantitative evidence on downside risk.

We study the distribution of long-horizon stock market returns. Our analysis aggregates across a broad set of developed economies with the goals of maximizing information about developed-country returns and minimizing potential biases. We use the historical record of stock market performance to estimate the distribution of buy-and-hold returns for a given investment horizon. To implement this analysis, we use a bootstrap simulation approach that resamples from historical data to produce draws of cumulative buy-and-hold returns. We use an expansive sample of 39 developed countries over long historical periods that collectively cover nearly 32,000 months (i.e., nearly 2,700 years) of information about stock returns beginning in 1841 and extending to 2019. These data provide us with a substantially broader view of risk over long horizons compared with a U.S.-centric design.

[1] See https://datacatalog.worldbank.org/dataset/world-development-indicators.

Our sample formation process mitigates two potential biases that affect inferences about expected performance and risk. First, Dimson et al. (2002) describe an easy data bias, wherein continuous stock return data from successful markets are more readily available. Our sample achieves substantially greater coverage of developed country periods compared with previous studies to minimize this bias. Second, Brown et al. (1995) discuss survivor bias, wherein conditioning on eventual market outcomes or return data series with no disruptions produces an upward bias in performance relative to ex ante expectations. We combat survivor bias with our classification of developed countries and our treatment of return data. Our developed country classification is based on information that was available at the time and does not condition on eventual outcomes. Before 1948 countries enter our developed sample when their agricultural labor shares decline below 50%, drawing on evidence about labor patterns from the economics literature (see, e.g., Kuznets, 1973). The U.S., for example, enters on this basis in 1890. After 1948 we use membership in the Organisation for Economic Co-operation and Development (OECD) and its European predecessor, the Organisation for European Economic Co-operation (OEEC). Within the developed country stock markets, we carefully deal with market disruptions and failures. The stock exchanges in several countries temporarily closed during our sample, and the stock market permanently disappeared in Czechoslovakia. Our treatment of return data in these instances reflects investor experiences to minimize survivor bias.

We produce bootstrap distributions of cumulative returns and wealth at horizons ranging from one month to 30 years. Our stationary block bootstrap design allows for features of returns that are important for buy-and-hold investors, including mean reversion and volatility persistence, but also carries the advantage that we remain relatively agnostic about the underlying structure of the return data. We study both nominal and real returns, but we concentrate on real returns because they better reflect the investor experience given a few periods of high inflation in the sample.

Fig. 1 plots bootstrap distributions of cumulative real wealth based on our developed country sample (blue bars) and on the United States sample (gray bars). The wealth levels reflect the outcome of investing $1.00 with a 30-year buy-and-hold strategy. We note that this analysis uses real returns measured in local currencies, such that the developed country distribution is informative about a home-country stock market investment in any developed country. We denote wealth in dollars only for the convenience of specifying a currency when we discuss wealth levels.[2] Both wealth distributions in Fig. 1 feature a prominent right tail, which indicates a high probability of achieving a substantial gain. The expected wealth levels are also similar at $7.38 for developed countries and $8.91 for the U.S. The distributions differ significantly, however, when we compare the left tails. Consistent with the findings in prior literature, the distribution based solely on U.S. data indicates a low 1.2% probability of a loss in buying power over a 30-year horizon. The full sample distribution reflects a much higher probability of loss at 12.1%. Further, the 1st and 5th percentiles for

[2]Throughout the paper, we also consider real returns and wealth levels from investing in global equity markets in U.S. dollars (USD). These results are from the perspective of a global USD investor whose investment performance reflects nominal returns in local currencies, changes in local exchange rates vis-à-vis the dollar, and U.S. inflation.

developed country long-term wealth are only \$0.14 and \$0.47, respectively, indicating that disastrous equity market performance cannot be ruled out for long-term stock market investors.

Evidence from the developed country sample indicates a considerable risk of loss for long-term investors. The distribution suggests that the −21% real return realization in Japan over the past 30 years is not exceedingly rare. In fact, this observation lies in the 9th percentile of the wealth distribution. Although this evidence suggests that Japanese investors were unlucky, their experience appears to be a reflection of the substantial risk exposures of long-term equity investors. An investor who learns about the distribution of 30-year returns using only the U.S. experience, in contrast, would assign a probability of just 0.5% that a return as extreme as the Japanese return realization could occur. The abundance of similar examples suggests that the U.S. distribution is overly optimistic with respect to loss probabilities.

Our findings of elevated loss probabilities for developed country markets relative to the U.S. experience are not driven by early periods, small countries, or small stock markets. In particular, loss probabilities remain elevated as we vary the starting year of the sample between 1841 and 2000, exclude countries with relatively small populations, or exclude countries with lower ratios of market capitalization to gross domestic product (GDP). Our results are also robust to alternative bootstrap assumptions and to measuring real returns in U.S. dollars (USD) rather than in local currencies. In all specifications, our analysis suggests that investors are exposed to substantial risk of loss once we account for information from the broader sample of developed countries.

To explore the economic implications of our findings, we consider a simple portfolio choice application. The investors optimize expected utility by allocating across stocks and a risk-free asset. We find that, relative to investors who use the U.S. sample to form expectations, investors using the developed country sample are considerably less aggressive when investing in stocks. Investors relying on the developed country distribution would also be willing to pay economically large fees of up to 1% per year to maintain their optimal weight rather than adopt the relatively higher weight in stocks implied by the U.S. analysis.

We contribute to a large literature in asset pricing that evaluates empirical phenomena across global asset markets and over long sample periods. The most influential work in this area focuses on economic questions for which traditional U.S. data samples fail to yield definitive answers. In many cases, these studies are motivated by statistical concerns such as survivor and easy data biases, peso problems (Rietz, 1988), and low power of econometric tests to identify significant relations over short time series. For example, Barro (2006), Barro and Ursúa (2008, 2012, 2017), Barro and Jin (2011), and Nakamura et al. (2013) examine the asset pricing implications of rare macroeconomic events using samples of consumption and GDP disasters in a broad cross section of countries. Lundblad (2007) considers the predictability of the market risk premium with lagged expected stock market volatility using a very long time series, and Goetzmann and Jorion (1995) and Golez and Koudijs (2018) adopt a similar approach in their studies on dividend-yield predictability. The literature on the cross section of stock returns contains numerous examples of studies that extend the existing evidence based on U.S. data to global equity markets (e.g., Fama and French, 1998; Asness et al., 2013) and to longer sample periods (e.g., Schwert, 2003; McLean and Pontiff, 2015; Linnainmaa and Roberts, 2018). A large volume of prior work also examines average stock market returns and equity premiums across countries (e.g., Jorion and Goetzmann, 1999; Dimson et al., 2002).

Our topics—the distribution of returns and the probability of loss over long horizons—are well-suited for an expanded-sample analysis. Existing work in this area (e.g., Ibbotson and Sinquefield, 1976; Fama and French, 2018a,b) is based on the historical U.S. experience and provides a useful first step, but the results suffer from the survivor and easy data critiques noted above. Our broader panel of historical returns alleviates these concerns and provides a more reliable characterization of tail outcomes, which are of particular interest for long-term investors.[3] Moreover, the distribution of long-horizon returns in developed markets is a subject in need of quantitative evidence. The findings in our study are relevant for a wide range of academic topics, including optimal portfolio choice and wealth management, optimal consumption and savings behavior in life-cycle models, and the assessment of macroeconomic models of asset prices. Our results also have policy implications for the design of pension systems (see, e.g., Shiller, 2003, 2006; Poterba et al., 2007) and the construction of target-date mutual funds (see, e.g., Viceira, 2008; Pástor and Stambaugh, 2012).

The rest of the paper is organized as follows. Section 2 describes the data on country-level stock market returns and outlines our approach to classifying economies as developed. Section 3 details the bootstrap simulation procedure. Section 4 presents our empirical results on the distribution of returns in developed markets, and Section 5 examines the implications of these results for asset allocation decisions. Section 6 concludes. The Internet Appendix presents additional information on data construction, bootstrap implementation, robustness tests, and portfolio choice design.

2. Data

The primary data for our study are stock market index returns at the country level. We construct these return series using the GFDatabase from Global Financial Data (GFD). Our goal is to produce returns for the broadest available stock market index in each country in terms of both stock inclusion and sample period to best represent the performance of a well-diversified investor. Our initial sample formation approach takes the perspective of an investor who lives in a developed country and invests in her home-country stock market over a predetermined horizon. To this end, we study nominal and real returns on stock market indexes in developed countries. Returns are denominated in local currency, which reflects the experience of a local investor and avoids additional foreign exchange risk. An alternative sample formation approach considers the perspective of a global USD investor. In this case, we measure performance in USD and adjust nominal

[3] A recent literature, including Pástor and Stambaugh (2012), Avramov et al. (2018), and Carvalho et al. (2018), studies the predictive variance of stock returns and finds that stocks may or may not be riskier over longer horizons. These studies emphasize the role of uncertainty about the model parameters and the evolution of expected return while studying long-horizon variance. Our bootstrap method is quite different from the Bayesian approaches in this literature. Although we must make some assumptions to implement the bootstrap, the empirical distribution of returns is not reliant on a particular model or values of model parameters. Our focus on distributions of long-horizon stock returns also provides information about higher-order moments that are important for fully characterizing the risk of stocks.

and real global stock market index returns to account for exchange rate changes over the investment horizon. We note that diversified investments in some countries and periods may have been difficult to achieve for non-domestic investors, but data from these periods are informative about the distribution of developed country returns.

We detail the provenance and attributes of the GFDatabase in the Internet Appendix. Broadly speaking, the GFDatabase provides data for diversified indexes that are created and calculated by stock exchanges (e.g., the Tokyo Stock Price Index from the Tokyo Stock Exchange), by well-known index providers (e.g., the S&P 500 Index), or by GFD directly from original source documents. The country-level index data provided by GFD account for capital changes (e.g., cash dividends, stock dividends and splits, subscription rights, mergers, and exchange offers) in accord with industry standards, such that the data reflect the investor experience. A particularly notable feature of our sample formation approach is that we only consider stock market returns that are inclusive of dividends. Given our focus on cumulative stock market performance over long horizons, including the impact of dividend income is crucial for accurately measuring investment outcomes. Whereas for a single month the effects of dividend income on return mean and variance are small, these effects grow with holding period. The dividend yield substantially affects realized performance over long periods (see, e.g., Dimson et al., 2002), and omitting dividends would lead to a downward-biased performance estimate. Further, uncertainty about dividend income contributes a higher proportion of return variance as investment horizon grows (see, e.g., Avramov et al., 2018), so including dividends allows us to better quantify risk over long horizons. Our cumulative return distributions assume that dividends are reinvested, which is commonly the default option in long-term savings vehicles like 401(k) accounts.

While forming our sample, we are cognizant of two potential biases and take precautions to avoid them to the extent possible. First, an easy data bias described by Dimson et al. (2002) reflects the tendency of researchers to focus on uninterrupted return data that are readily available, such as returns from the U.S. or the U.K. The concern is that these returns are not representative of the broader sample because data availability is an endogenous outcome. We mitigate the easy data bias by creating a sample of developed market returns that is broad in terms of the number of countries and the length of historical periods that are included. Second, a survivor bias described by Brown et al. (1995) arises if we condition on eventual stock market outcomes while constructing our sample. To combat this issue, we identify the periods over which countries are developed based on metrics that would have been available to investors at the time rather than condition on which economies are currently developed. We also carefully recreate, to the extent possible, the historical experiences of investors in markets that had significant disruptions or that ultimately did not survive.

Section 2.1 describes our data sources and return calculations. Section 2.2 discusses the identification of developed countries. Section 2.3 outlines unique data-related issues that require attention in creating our final sample and examines our sample's coverage of the developed period. Section 2.4 presents summary statistics for our return sample.

2.1. Return data

Our primary data source is the GFDatabase from Global Financial Data. The database contains historical time series of stock market and macroeconomic data for a broad set of countries. As we demonstrate in Section 2.3, this data source allows us to create a much broader sample of historical stock returns in developed countries compared with previous studies that consider the distribution of long-horizon stock returns.

To construct returns, we gather five variable types from the database: (i) total return indexes that reflect both price changes and dividends, (ii) price indexes that measure capital gains, (iii) dividend yields at the index level, (iv) consumer price indexes (CPIs), and (v) local currency exchange rates vis-à-vis the U.S. dollar. For periods in which the database contains multiple total return indexes or price indexes for a given country, we select a single index by considering the breadth of market coverage and the length of historical coverage. These index choices are detailed in the Internet Appendix. To calculate a monthly nominal return, we use either a total return index or a combination of a price index and a dividend yield.

The monthly percentage change in a total return index reflects the nominal return,

$$R_{i,t}^{Nominal} = \frac{I_{i,t}^{Total}}{I_{i,t-1}^{Total}}, \tag{1}$$

where $I_{i,t}^{Total}$ is the total return index for country i at the end of month t and $R_{i,t}^{Nominal}$ is the gross nominal return for country i in month t. We use a total return index to calculate returns whenever one is available.

For periods with no total return index data, we combine data from price indexes and dividend yields to produce returns that are inclusive of dividends. Reported dividend yields reflect dividends paid on the index over the prior 12 months,

$$DY_{i,t} = \frac{\sum_{\tau=0}^{11} D_{i,t-\tau}}{I_{i,t}^{Price}}, \tag{2}$$

where $DY_{i,t}$ is the dividend yield for country i reported in month t, $I_{i,t}^{Price}$ is the price index, and $D_{i,t-\tau}$ is the dividend on the index in month $t-\tau$. We use December dividend yields and multiply the dividend yield $DY_{i,t}$ by the price index $I_{i,t}^{Price}$ to calculate the total dividend for the year, $\sum_{\tau=0}^{11} D_{i,t-\tau}$. We assume that this dividend amount is paid equally across months, such that the estimated dividend $\hat{D}_{i,t-\tau}$ for each month $t-\tau$ of the year is $\frac{1}{12}\sum_{\tau=0}^{11} D_{i,t-\tau}$. The nominal return for each month is calculated by combining the capital gain inferred from the price index and the estimated dividend payment. For example, the gross nominal return in December (i.e., $\tau = 0$) is

$$R_{i,t}^{Nominal} = \frac{I_{i,t}^{Price} + \hat{D}_{i,t}}{I_{i,t-1}^{Price}}. \tag{3}$$

We provide additional details on return construction in the Internet Appendix.[4]

[4]In the Internet Appendix, we also conduct tests to assess the validity of our return calculations that use price indexes and dividend yields. During the periods with overlap of a total return index and a combination of a price index and a dividend

Given a nominal return series, we calculate real returns using CPI data. Specifically, let

$$\Pi_{i,t} = \frac{I_{i,t}^{CPI}}{I_{i,t-1}^{CPI}} \tag{4}$$

be the gross realized inflation rate for country i in month t, where $I_{i,t}^{CPI}$ is the CPI level. The gross real return is given by

$$R_{i,t}^{Real} = \frac{R_{i,t}^{Nominal}}{\Pi_{i,t}}. \tag{5}$$

The return in Eq. (5) summarizes the real investment performance of an investor in country i. We also consider real investment outcomes from the perspective of a global USD investor who invests in local markets and makes the required beginning-of-period and end-of period conversions between U.S. dollars and local currency. In this case, the gross real USD return reflects the nominal return expressed in terms of the local currency, local currency appreciation relative to the U.S. dollar, and U.S. inflation over the investment period:

$$R_{i,t}^{Real\ USD} = \frac{R_{i,t}^{Nominal}}{\Pi_{us,t}} \left(\frac{E_t^{USD,i}}{E_{t-1}^{USD,i}} \right), \tag{6}$$

where $E_t^{USD,i}$ is the exchange rate at the end of month t expressed in USD per local currency.

2.2. Developed country classification

Our approach to classifying countries as developed uses information about economic and financial market development that plausibly would have been available to investors in real time. We define economic development based on the degree of industrialization using labor patterns in the pre-1948 period and on economic organization membership in the post-1948 period. We also require a country with a developed economy to have a stock exchange to be included in our developed country sample.

Our sample comprises stock market returns from 40 developed periods in 39 countries. Table 1 lists the developed countries, the years in which they are first classified as developed, the reason for each development classification, and the start and end dates of each country's sample period in our analysis based on development and data availability. The economic development dates and classifications use the following criteria:

1. *Early-period economic development:* The development economics literature identifies country-level patterns in the evolution of labor shares in agriculture, manufacturing, and services through the stages of economic development. As a country industrializes, the labor share in agriculture declines as the labor shares of manufacturing and services increase (see, e.g., Kuznets, 1973). As a simple rule of thumb, we classify a country as developed from the first year before 1948 in which employment data show that less

yield, we calculate nominal returns following both Eq. (1) and Eq. (3). In the full overlapping sample, the resulting return estimates are highly correlated (correlation of 0.98) and have statistically indistinguishable mean (0.97% per month for total return index versus 0.98% for price index and dividend yield) and standard deviation (5.65% versus 5.61%).

than 50% of workers are employed in agriculture.[5] The United Kingdom (1841) is the first to achieve this benchmark, followed in the nineteenth century by the Netherlands (1849), Belgium (1856), France (1866), Norway (1875), Germany (1882), Denmark (1890), Switzerland (1890), the United States (1890), Canada (1891), Argentina (1895), and New Zealand (1896). These dates are broadly consistent with the geographical spread of developments in industrial technology and transportation. Norway's first stock exchange opened in 1881, leading to a slightly later development classification as shown in Table 1, whereas each other country had an established stock exchange when the labor benchmark was achieved. The labor data indicate pre-1948 development dates for 11 additional countries.

2. *Late-period economic development:* We use membership in the OEEC or the OECD to identify countries that become developed after 1948. The OEEC was formed in 1948 to administer the Marshall Plan in the aftermath of World War II, and the OECD replaced the OEEC and expanded membership to non-European countries in 1961. Details on the countries for which organization membership determines the development date follow.

   (a) Four countries–Iceland, Luxembourg, Spain, and Turkey–are classified as developed based on OEEC membership. Luxembourg and Turkey were founding members of the OEEC and are classified as developed starting in 1948. Iceland was also a founding member of the OEEC, but enters our sample of developed countries in 1991 with the establishment of its first stock exchange. Finally, Spain was a late entrant into the OEEC in 1959 on the basis of economic development.[6]

   (b) The founding members of the OECD consisted of the 18 OEEC members along with Canada and the United States. Australia, Japan, and New Zealand, which were already classified as developed based on labor data, joined in the early years of the organization. Of the countries not previously identified as developed, 13 enter our developed sample based on admission to the OECD. Finland was a relatively early entrant in 1969, and the remaining 12 countries have joined the OECD since 1994.

Of the 40 periods in our sample, 37 extend to the end of our sample period in December 2019. The three exceptions are developed periods in Argentina, Chile, and Czechoslovakia. These countries are initially classified as developed based on labor share data, but subsequent outcomes lead us to place ending dates on their developed periods.

[5] Country-level labor data are from Mitchell (1993, 1995, 1998). We select the first year for which labor data are available and the agricultural labor share is less than 50%. Ten countries–Argentina, Australia, Chile, Czechoslovakia, Germany, the Netherlands, New Zealand, Singapore, Switzerland, and the United Kingdom–have less than 50% agricultural labor shares for their first labor datapoint. The sample periods for six of these countries are constrained by return data availability as noted in Table 1. The development dates for the other four–Australia (1901), Germany (1882), New Zealand (1896), and the United Kingdom (1841)–seem reasonable given their development histories and comparable country development dates.

[6] Spain's initial exclusion from the OEEC could be attributable to both political and economic reasons. The government of Spain remained a fascist dictatorship in the immediate aftermath of the World War II period, and the country operated with a largely closed economy. Following political and economic changes, the country began to seek membership in the OEEC in the mid-1950s. Spain was not allowed to join due to its economic policies and development until 1959 (Calvo-Gonzalez, 2006), which is the year that we use for development. Spain's labor force was also much more agriculturally based in the mid-twentieth century compared with its European peers, which is consistent with its later development date.

Despite strong early development, Argentina is currently classified as a developing country by most metrics and Chile only recently gained OECD membership in 2010. At the time Argentina enters as a developed country in 1895, its real per capita GDP exceeded that of the United States.[7] Argentina's relative standing declined, however, through the World Wars, the Great Depression, and the post-World War II period. We remove Argentina as a developed country in 1966 coincident with the Argentine Revolution during which a military coup led to a seven-year period of military dictatorship. By this time, Argentina's real per capita GDP was less than half that of the U.S. Chile enters our developed country list in 1920 with a higher per capita GDP than Austria and Greece, which also enter in that year. Chile experienced particularly slow growth during the Great Depression and the post-World War II period. In 1970, Salvador Allende surprisingly won the presidential election on a platform of nationalizing the Chilean copper industry without compensation, and the stock market dropped by nearly half in response to his election (Girardi and Bowles, 2018). In addition to copper, Allende quickly nationalized large portions of the industrial, infrastructure, service, and financial sectors (*Chile: Allende's Economic Record*, Intelligence Memorandum, Central Intelligence Agency, April 1972). We end Chile's first period of development in 1970, and Chile re-enters our sample in 2010 with its admission to the OECD.

Czechoslovakia is unique in our sample of developed countries as it has the only stock market that ceases to exist. Following the dissolution of the Austro-Hungarian Empire, Czechoslovakia and Austria were the most productive of the emerging countries. Czechoslovakia's per capita GDP was about 25% larger compared with Austria when the country enters our developed sample in 1921. After the German occupation of Czechoslovakia during World War II, a government was formed in 1945 under President Edvard Beneš. The Beneš government nationalized almost all significant companies, and the stock exchange was permanently discontinued in 1946. Czechoslovakia eventually split into the Czech Republic and Slovakia in 1993, and these countries enter our sample with their admissions to the OECD in 1995 and 2000, respectively.

2.3. Special data considerations and data coverage

Forming return series for our developed country sample presents several challenges. Market disruptions caused by economic crises, revolutions, and wars affect data availability, contributing to the potential for the easy data bias of Dimson et al. (2002) and the survivor bias of Brown et al. (1995). In this section, we describe our approaches to the periods in the sample in which clean monthly return data are not readily available. To preview the outcome of this process, Fig. 2 displays our final sample's coverage of the developed periods. The solid blue lines indicate data coverage, and the dotted black lines indicate developed periods with missing data. Fig. 2 illustrates that we have no unintentional gaps in the middle or at the end of any developed country period. We view this feature of our sample as particularly important for minimizing the easy data and survivor biases, given that data availability issues are often concurrent with negative events

[7] Real per capita GDP comparisons are based on the 2018 Maddison Project Database, which is available at https://www.rug.nl/ggdc/historicaldevelopment/maddison/releases/maddison-project-database-2018. See Bolt et al. (2018) for additional details.

such as economic crises or wars. There are, however, some remaining countries for which early historical stock returns inclusive of dividends are not available, and we discuss these cases in more detail below.

For most of our sample, monthly historical return data are readily available. Five situations create exceptions. We briefly introduce these categories and then describe our approach to each situation. We provide full detail on these data items in the Internet Appendix. First, several countries closed their stock exchanges or restricted trading for extended periods during our sample. Second, short gaps in price or dividend yield data occasionally occur in the middle of a sample when a country's stock exchange is open, and these issues tend to occur during important historical events. Third, a period of hyperinflation in Germany produces return observations that are not meaningful from an investor's perspective. Fourth, the stock exchange in Czechoslovakia was permanently discontinued, and the closing prices on the exchange do not reflect the reality of investor outcomes. Fifth, historical data are not available for early periods in some countries.

Table 2 shows periods of stock exchange closures or heavily restricted trading for the countries in our developed sample. For each event, we report the number of months spanned by the period of disruption, the start and end dates of the period, the nominal returns, and the real returns (measured in both local currency and in USD) that were realized when the exchange returned to normal trading. The panels in Table 2 split events into those related to World War I (Panel A), World War II (Panel B), political revolutions (Panel C), financial and banking crises (Panel D), and labor strikes (Panel E).

The U.S. experienced just one disruption that affects monthly data availability. The New York Stock Exchange (NYSE) was closed from July 31 to December 12 in 1914 with the onset of World War I. Within eight trading days of the NYSE closure, the New Street black market emerged to facilitate stock trading (Silber, 2005). The GFDatabase contains return data from the New Street exchange during the NYSE closure (and similar data from black markets in selected other countries). However, Silber (2005) documents that New Street trading volume was considerably lower compared with the NYSE. Further, the majority of investors do not have easy access to black market trading. Most investors who held stocks at day end on July 30, 1914 would have maintained these positions until the market reopened in December, effectively realizing the single five-month nominal return of -2.14% listed in Panel A of Table 2.

The German market during World War II provides another example of market disruption. Stock prices increased during the early war period as Germany notched military victories. With its defeat at Stalingrad, Germany effectively froze stock prices in January 1943. Trading was extremely limited even though the exchange was officially open. The market subsequently closed after the Allied invasion. When prices once again were allowed to float in July 1948, stocks fell. As shown in Panel B of Table 2, the cumulative nominal return for the 67-month period from January 1943 to July 1948 was -87.62%.

In our empirical analysis, we treat each of the events in Table 2 as a single return observation that covers a multi-month period. Each event is characterized by exchange closure or heavily restricted trading. Our treatment reflects that most investors would have been unable to trade, such that they would have had no choice but to wait for the eventual realizations of these longer-term, buy-and-hold returns.

We observe a few historical periods for which return data for a particular country are not readily available while the stock exchange was open. We have price data but no dividend yield data for Austria from 1939 to 1969, Chile from 1967 to 1970, and Czechoslovakia from 1938 to 1945. These periods in Austria and Czechoslovakia notably began during German annexation or occupation, and Chile experienced an economic crisis beginning in 1967 after a period of high economic growth (Rebolledo, 2005). To avoid dropping these periods and creating a systematic bias, we use dividend yield information from these countries in non-missing periods to make informed assumptions about dividends during the missing periods. We are also missing some monthly price index datapoints over short periods in Argentina, France, and Switzerland when the exchanges were open. These issues occur during important periods, as the missing months in France and Switzerland are in the World War I era and those in Argentina surround the military coup that deposed Juan Perón in 1955. Because these missing returns appear to be purely data related, we smooth returns across the missing months in these cases. Making these adjustments allows us to have an uninterrupted return series for each developed country period.

Our sample includes one exception to our practice of computing nominal returns denominated in the primary home currency. Germany gave up the gold standard in August 1914, contemporaneous with the start of World War I, and switched from "gold marks" to "paper marks" as its primary currency. The country experienced bouts of inflation over the 1917 to 1923 period as the value of paper marks fell. Paper marks became nearly worthless with hyperinflation of about 22,000,000,000% in 1923. To varying degrees over this period, gold marks or a close substitute existed as a secondary currency. Rather than using returns measured in paper marks (which produces an annual nominal return in 1923 of about 1,200,000,000,000%), we use an alternative series available in the GFDatabase that is denominated in gold marks over the period from 1917 to 1923. Because CPI data in terms of gold marks are unavailable, we use these returns as both nominal and real returns over this period.

Our sample includes only one country with a stock exchange that permanently closed. Shareholders in Czechoslovakia experienced near total losses during this event. The Prague Stock Exchange closed on May 5, 1945, shortly before the Germans were driven from the country and the Beneš government took control. In addition to shareholder losses from the sudden nationalization of most traded companies, a communist coup in 1948 ended hopes of receiving the compensation for seized stock in these companies that was promised by the Beneš government.[8] Our treatment of this event takes the perspective of an investor in Czechoslovakia

[8]The stock exchange in Prague closed on May 5, 1945, coinciding with a civilian uprising against the Germans. Shortly after, the Soviet Red Army drove out the remaining Germans on May 9, 1945. The new Beneš coalition government began to nationalize assets on May 19, 1945, and most industrial, banking, and insurance companies were nationalized by October 27, 1945 (Winkler, 1994). The stock exchange did not reopen because about 90% of the previously listed stocks had been nationalized, and it was officially discontinued by the Czechoslovak Ministry of Finance in June 1946 (*Foreign Commerce Weekly, Vol. XXIII*, June 8, 1946). Stockholders (except for many individuals of German or Hungarian descent) were to receive compensation for their stock in the form of securities from the Economic Fund of Nationalized Property (Doman, 1950). This fund was designed to receive surplus earnings from nationalized firms and issue securities with interest payments that were guaranteed by the government. In February 1948, the Beneš government was overthrown in a communist coup. The government of the newly formed Czechoslovak Socialist Republic did not recognize any property or compensation rights for

who held stock at the end of March 1943. Stock prices were near all-time highs, suggesting the impending outcome was not anticipated by investors. Severe trading restrictions and price controls were enacted in 1943 and persisted until the exchange closed in 1945, such that selling stock would have been difficult during this period. Lacking an exact figure for the near total losses suffered by investors, we assume a −90% nominal return. This event appears in Table 2 as a 39-month return observation that spans the period that begins with trading restrictions in April 1943 and ends with the permanent discontinuation of the stock exchange in June 1946. This treatment provides a reasonable (and, if anything, an optimistic) representation of the investor experience in Czechoslovakia and is necessary for minimizing survivor bias.

Finally, Table 1 indicates that the sample start date for returns occurs after the development date for 14 of the 40 developed country periods. Several of the 14 countries have short periods with missing returns. A few others, most notably the Netherlands, have relatively longer periods. We classify the Netherlands as developed in 1849 when its agricultural labor share declines below 50%, but monthly price data are only available beginning in 1914. In each country with a missing early period, we lack sufficient data on prices, dividends, or both, such that we are unable to produce monthly returns inclusive of dividends.

Given that the missing returns in the early portions of developed periods represent the only gaps in our sample, it is important to consider their potential effects. A concern discussed by Brown et al. (1995) and Dimson et al. (2002), among others, is that return data are systematically more available for countries with better eventual economic outcomes. As we examine the 14 countries for which we are initially missing data, however, we find that they tend to have slightly higher end-of-sample development, as measured by real per capita GDP, compared with countries with full data samples. A clearer pattern is that data are more likely to be missing for countries that are small, such as Iceland, Luxembourg, and Singapore. In terms of 2019 population, countries with full series are about three times larger on average compared with countries with missing data. Although the precise impact of the missing returns on our findings is unknown, size-related undersampling seems less problematic compared with performance-related undersampling. We also demonstrate in Section 4.2 that our results are robust to excluding small countries from the sample.

Our final sample achieves broad coverage of the developed periods. The 40 periods of development described in Section 2.2 span a total of 36,786 months (over 3,065 years). Our sample contains 31,994 months (over 2,666 years), such that our data cover 87.0% of the full developed period. For comparison, Fama and French (2018a) study the distribution of long-horizon returns using 642 months (over 53 years) of U.S. return data, which cover just 1.7% of the full developed period. Overall, our broad sample contains a wealth of information about long-horizon returns.

2.4. Summary statistics

Table 3 reports summary statistics for our return sample. The statistics in Panel A are based on nominal returns, those in Panel B are based on real returns, and those in Panel C are based on real USD returns.

former stockholders (Glos, 1986).

The number of months and the number of observations can differ for a given country because of the multi-month periods in Table 2. For example, the 1,556 U.S. observations span a 1,560-month period because of the five-month return observation during World War I. We report both the arithmetic mean ($\bar{R}_a$) and the geometric mean ($\bar{R}_g$) for returns as well as the higher-order return moments. The arithmetic and geometric means account for the number of months in the sample rather than the number of observations. The higher moments treat each observation equally. We also show the minimum and maximum return observations during the sample. Finally, each panel has a "Full sample" case that is pooled across all developed periods. We note that (i) cross-country comparisons are somewhat difficult in this setting because of sample period differences and (ii) the recently developed countries have short samples that are likely not representative of long-term expectations.

The nominal returns in Panel A of Table 3 often display extreme behavior. The arithmetic (geometric) mean return for the pooled sample is 0.95% (0.77%) per month, which is slightly higher than the U.S. average of 0.87% (0.75%). The skewness and kurtosis of the full sample, however, indicate that the data include outliers. The maximum return observations exceed 100% for Turkey (127.17%), Germany (128.82%), Austria (300.73%), and Japan (449.38%). Notably, these returns in Austria and Japan are listed in Table 2 as they follow market closures during World War II, and both are accompanied by negative real returns. The worst nominal returns in the sample occur in Iceland with a −71.52% return in October 2008, Portugal with a −80.39% return during the Carnation Revolution that began in 1974, Germany with a −87.62% return in the aftermath of World War II, and Czechoslovakia with our estimated −90.00% return that accompanies the seizure of assets and the official discontinuation of its stock market.

The real returns in Panel B of Table 3 provide a more informative view of the investor experience compared with the nominal returns in Panel A. The arithmetic (geometric) mean of the pooled sample of real returns is 0.55% (0.37%) per month, such that equity returns substantially exceed inflation on average. The U.S. has slightly higher average performance than does the full sample at 0.64% (0.52%), but the U.S. is not an outlier relative to the other countries. This observation is more in line with Dimson et al.'s (2002) findings rather than Jorion and Goetzmann's (1999) conclusion that the high equity premium in the U.S. "appears to be the exception rather than the rule." Relative to nominal returns, real returns have a somewhat lower standard deviation (5.86% per month versus 6.60%) and much lower skewness (1.13 versus 14.16) and kurtosis (43.07 versus 838.83). Moreover, the minimum and maximum real returns in Panel B often differ substantially from the nominal return counterparts in Panel A. As a case in point, Japan's best nominal return observation corresponds to its worst real return, as a large currency devaluation contributed to inflation of about 4,200% in the period following Japan's surrender to end World War II.

The real USD returns in Panel C of Table 3 reflect the real investment performance from investing in local markets in U.S. dollars. From an economic perspective, the results are broadly similar to those in Panel B. Relative to the real local currency returns, the pooled sample of real USD returns exhibits a slightly higher arithmetic mean (0.61% versus 0.55%) and an identical geometric mean (0.37%). The standard deviation of real USD returns is also higher than the standard deviation of real local currency returns for each individual country. This result is expected, however, as the real USD returns incorporate foreign exchange risk.

The summary statistics in Table 3 indicate that accounting for inflation is important for understanding long-term stock returns. Investors ultimately care about the buying power of their savings, and periods of very high inflation can produce nominal returns that are not representative of the investor experience. Given the large effects of inflation within our sample, our empirical analysis in Section 4 is based on real returns and real USD returns. We present results based on nominal returns in the Internet Appendix.

3. Bootstrap simulation design

Our main analysis focuses on distributions of long-term, buy-and-hold stock market returns. This section introduces our bootstrap simulation approach to estimating these distributions. Broadly speaking, the bootstrap method draws with replacement from our sample of 31,465 observations to form a sequence of returns with a horizon of H months. Given a return sequence, we calculate cumulative returns and wealth levels, and we repeat this process to form bootstrap distributions of these variables of interest.

The time-series properties of stock market returns affect buy-and-hold investors, and our approach allows for serial dependencies in returns to account for these effects. Perhaps the most relevant stylized fact for long-term investors is the negative serial correlation in returns, which is commonly referred to in the literature as mean reversion (see, e.g., Poterba and Summers, 1988; Barberis, 2000; Campbell and Viceira, 2002; Siegel, 2014). Mean reversion describes the property of returns that a large positive (negative) return in the current month tends to be followed by relatively low (high) returns in subsequent years. This pattern follows from a valuation framework in which expected stock returns are time varying (see, e.g., Campbell and Shiller, 1988). Although the magnitude of the negative serial correlation is small, the cumulative economic effects can be large. Mean reversion forms much of the basis for arguments that long-horizon stock returns are relatively safe and that younger investors should invest more in stocks compared with older investors.

Whereas the mean reversion property acts over long periods and has the effect of lowering long-horizon variance, additional properties of returns act over shorter periods to increase variance and produce fatter tails. Short-term autocorrelation in nominal returns tends to be positive (Lo and MacKinlay, 1988), which influences cumulative returns over short periods to be relatively riskier. Persistence in measured inflation may also contribute to short-term autocorrelation in real returns. A larger effect on risk in short-term cumulative returns comes from persistence in return variance (Engle, 1982). These patterns may not form a comprehensive list, as other forms of serial dependencies could also affect buy-and-hold investors.

We adopt a block bootstrap design to account for serial dependencies in returns. Relative to alternative approaches to understanding long-horizon risk such as specifying vector autoregressions (see, e.g., Pástor and Stambaugh, 2012; Avramov et al., 2018), the bootstrap approach has the advantage that we do not need to assume a particular form for serial dependencies in expected return and variance. Rather, the underlying structure of the return data is reflected by the return blocks. Several other finance studies use block bootstrap designs to account for time-series dependencies in returns, including Brown and Warner (1980, 1985), Sullivan et al. (1999), Kosowski et al. (2006), Patton and Timmermann (2010), and Martin and Wagner (2019).

Formally, we base our approach on the stationary bootstrap of Politis and Romano (1994). The stationary bootstrap draws blocks of consecutive data from the underlying sample with replacement. The block sizes are randomly determined, but they are controlled by an average block length parameter. To avoid undersampling from any part of the sample, the stationary bootstrap specifies that a block that begins toward the end of a series and that cannot be filled by the remaining data wraps back to the beginning of a series to fill the block. Our longest horizon is 30 years, so we produce return sequences that fill 360 months of returns. Our process for generating the bootstrap return draw in iteration m is as follows:

1. A random block size b is drawn from a geometric distribution with a probability parameter equal to the inverse of the desired mean block length.
2. The return block draw starts with a randomly selected observation from the 31,465 observations in the sample. We denote this observation as $R_{i,t}$, where i indexes countries and t indexes observations within country i. If country i's sample contains return observations $R_{i,t}$ through $R_{i,t+b-1}$ that consist only of single-month returns, the return block draw is $B_b = \{R_{i,t}, R_{i,t+1}, \ldots, R_{i,t+b-1}\}$. Exceptions occur if this sequence of country i's returns includes multi-month observations or if the country's sample ends before observation $t + b - 1$:
   (a) We treat the block size draw b as the number of months rather than the number of observations. If at least one observation in $R_{i,t}$ through $R_{i,t+b-1}$ is a multi-month return observation from Table 2, the number of months spanned by $\{R_{i,t}, R_{i,t+1}, \ldots, R_{i,t+b-1}\}$ is larger than b. In this case, we find the smallest number b^* such that $\sum_{s=0}^{b^*} N_{i,t+s} \geq b$, where $N_{i,t+s}$ is the number of months spanned by $R_{i,t+s}$ (e.g., $N_{i,t+s} = 1$ for most observations and $N_{i,t+s} = 5$ for the five-month U.S. World War I closure). The return block draw is $B_b = \{R_{i,t}, R_{i,t+1}, \ldots, R_{i,t+b^*}\}$. If $R_{i,t+b^*}$ is a multi-month observation, B_b may span more than b months.
   (b) If we denote the last observation of the period in country i as $R_{i,T}$ and $\sum_{s=0}^{T-t} N_{i,t+s} < b$, then $\{R_{i,t}, R_{i,t+1}, \ldots, R_{i,T}\}$ is not sufficient to fill the block B_b. In this case, we draw a random country j from the 40 developed periods discussed in Section 2.2 and use the beginning of its sample for the block. Starting from the beginning of country j's sample ensures that the bootstrap does not undersample these months. Assuming country j's sample has enough data to fill the block (i.e., the data in country j span at least $b - \sum_{s=0}^{T-t} N_{i,t+s}$ months), the block is $B_b = \{R_{i,t}, R_{i,t+1}, \ldots, R_{i,T}, R_{j,1}, R_{j,2}, \ldots, R_{j,b^*}\}$, where b^* is the smallest number such that $\sum_{s=0}^{T-t} N_{i,t+s} + \sum_{s=1}^{b^*} N_{j,s} \geq b$. If the data in country j do not span enough months, we repeat the process by drawing a new random country until the block is filled. We use data from a random country to avoid inadvertently creating serial dependencies across periods where none truly exists. If we were to simply stack country data alphabetically, for example, the simulation would

be under the impression that the 2008 Financial Crisis in Iceland always occurs 12 years before the 1936 to 1938 period of the Anglo-Irish Trade War in Ireland.

3. We add B_b to the bootstrap return vector draw $R^{(m)}$, and we return to step one and repeat the process until $R^{(m)}$ contains at least 360 return observations. The final bootstrap draw in iteration m is $R^{(m)} = \{R_1^{(m)}, R_2^{(m)}, \ldots, R_{360}^{(m)}\}$.

Given the bootstrap return vector draw $R^{(m)}$, for each horizon H we find the smallest number H^* such that $\sum_{s=1}^{H^*} N_s^{(m)} \geq H$, where $N_s^{(m)}$ is the number of months spanned by $R_s^{(m)}$. If the last observation $R_{H^*}^{(m)}$ is a multi-month observation, the returns may span more than H months. This situation mirrors the case of an investor who reaches the end of her horizon during a period in which the exchange is closed, and she must wait for the exchange to reopen before selling her shares. For a \$1.00 buy-and-hold investment, the draw of wealth at the H-month horizon is

$$W_H^{(m)} = \prod_{s=1}^{H^*} R_s^{(m)}. \tag{7}$$

The continuously compounded return is

$$C_H^{(m)} = \sum_{s=1}^{H^*} \log(R_s^{(m)}). \tag{8}$$

We repeat this process for $m = 1, 2, \ldots, 1{,}000{,}000$ to produce the distributions of wealth and cumulative log returns at each horizon.[9]

The block length parameter governs the average size of random block lengths. When the block length parameter is one, all blocks are single return observations and the bootstrap does not capture any aspects of serial dependence in returns. With a short average block length, the bootstrap captures short-run effects like positive autocorrelation and persistent volatility. Short blocks, however, tend to miss out on the long-run mean reversion feature of returns. Using long blocks of return observations from the data allows for longer-term serial correlation to be reflected in our cumulative returns. In the Internet Appendix, we demonstrate these effects using the moments of the cumulative return distributions across different average block sizes. Given the importance of mean reversion for cumulative buy-and-hold returns over long horizons, we choose a base case block length parameter that produces an average block size of 120 months. We also establish robustness to other reasonable choices of the block length parameter and to using an independent bootstrap design in the Internet Appendix.

[9] The bootstrap resamples return data across countries and time. Rather than take the bootstrap samples literally as the experiences of hypothetical investors who shift investments across time and space, we view the bootstrap as a useful statistical tool for making inferences about the population distribution for developed country stock returns given the sample of data.

4. Distributions of stock market returns

In this section, we present results from bootstrap simulations of long-term stock market returns. Section 4.1 focuses on our base case empirical design, which relies on resampling from the full sample of returns in developed stock markets. Section 4.2 presents results for alternative samples. In particular, we assess the quantitative impact of easy data and survivor biases by comparing the base case results with those from samples based on individual countries or from samples conditioned on eventual economic outcomes. We also evaluate the impact of excluding data from early in the sample period, from countries with relatively small populations, or from countries with relatively small stock markets.

4.1. Base case simulation

Table 4 summarizes the distributions of real payoffs across 1,000,000 bootstrap simulations. As described in Section 3, the simulations are based on the full sample of developed country returns over the period from January 1841 to December 2019. We present results for horizons of one month, one year, five years, ten years, 20 years, and 30 years. Each simulation run involves resampling with replacement from the pooled sample of monthly returns using a stationary block bootstrap approach, where the length of each block has a geometric distribution with a mean of 120 months.

For each horizon, Panel A of Table 4 reports the mean and standard deviation of real payoffs (reflecting performance in terms of the local currency), the percentiles of the payoffs, and the proportion of the simulations in which the investor experiences a loss in real terms (i.e., terminal wealth is below $1.00). As expected, the average payoff and the volatility of the payoff both show pronounced horizon effects. The mean real payoff grows from $1.01 at the one-month horizon to $7.38 at 30 years, and the standard deviation of the payoff increases from $0.06 at one month to $13.76 at 30 years. The skewness of payoffs (unreported) also increases with horizon. Whereas the average one-month and one-year payoffs are close to their corresponding median values, average payoffs exceed median payoffs considerably at longer horizons. This effect is apparent in Fig. 3, which plots histograms of real payoffs from the bootstrap simulations. Each of the distributions at the five-year horizon and beyond features a pronounced right tail, reflecting the high probability of an extreme positive cumulative return. These impressive payoffs represent the upside potential of stocks over long holding periods, which financial advisors often tout as the power of compound interest.

The most prominent feature of the results in Panel A of Table 4 and Fig. 3 is the substantial uncertainty over real investment outcomes faced by long-horizon investors. For a ten-year investor, for example, the 1st percentile of real payoff is just $0.13, whereas the 99th percentile is $8.75. The dispersion in the payoff distribution is even more pronounced at the 30-year horizon, as the 1st and 99th percentiles are $0.14 and $53.45, respectively. This variation in investment outcomes has important implications for the consumption, savings, and asset allocation decisions of long-term investors (e.g., investors saving for retirement). The results in Panel A of Table 4 also appear at odds with the conventional advice that stocks are safe for younger investors based on the empirical evidence of strong mean reversion in the aggregate U.S. stock market (see, e.g., Poterba and Summers, 1988; Siegel, 2014). For the 30-year investor, the 1st percentile real payoff of \$0.14 represents a catastrophic investment outcome. Moreover, even the 10th percentile result of \$0.85 represents an economically large loss in real terms.

A complementary way to characterize stock market risk is to examine the probability that investors experience losses over a given horizon. The final column of Table 4 presents the proportion of simulation runs in which the terminal real payoff is less than \$1.00. Fama and French (2018a) present a similar measure in their study based on the distribution of nominal payoffs of U.S. stocks. They find that the likelihood of losing money on a buy-and-hold investment in the aggregate stock market at a horizon beyond 20 years is negligible.

Panel A of Table 4 shows, unsurprisingly, that the likelihood of experiencing a loss is large at short horizons. The proportion of real payoffs less than \$1.00 is 43.2% at one month and 36.8% at one year. This proportion continues to decrease with horizon. Nevertheless, the 20-year and 30-year probabilities of a decline in wealth from stock market investments are still substantial. The loss probability is 15.5% at 20 years and 12.1% at 30 years. These results contrast those in Fama and French (2018a) and also contradict the conventional advice that stocks are safe investments at long horizons. Our findings highlight the importance of guarding against biases related to survival and easy data in characterizing the risks faced by long-term investors. Even at long horizons in the world's most developed markets, investors bear considerable risk of loss.

The payoffs in Panel A of Table 4 do not include the effects of commissions, fees, and expense ratios incurred to gain market exposure. As a practical matter, the cumulative effect of fees is an important consideration for long-term investors. To demonstrate this effect, we take the perspective of a modern-day investor who uses our bootstrap approach to assess the distribution of future real gross (i.e., before fees) returns of financial intermediaries such as mutual funds.[10] The results suggest that mean investment outcomes and loss probabilities can be starkly different for investments in low-fee versus high-fee vehicles. The Vanguard Total Stock Market Index Fund (VTSAX), for example, represents a low-fee vehicle with an annual expense ratio of 0.04%. The cumulative impact of a 0.04% fee on investment results over a 30-year horizon is mild, decreasing the mean payoff from \$7.38 to \$7.29 and increasing the probability of loss from 12.1% to 12.3%. Higher fees have economically significant effects. A mutual fund that charges the asset-weighted (equal-weighted) average expense ratio among active managers of 0.67% (1.11%) has a mean payoff of \$6.04 (\$5.30) and a loss probability of 15.2% (17.5%).[11]

In the Internet Appendix, we also characterize the bootstrap distribution of continuously compounded returns at various investment horizons. These returns are computed following Eq. (8). Fama and French

[10] This analysis assumes that these financial intermediaries generate pre-cost returns that are close to market returns. Lewellen (2011) shows that the portfolio of institutional investor stock holdings closely resembles the market before considering fees. Some institutional investors outperform even after accounting for fees, but the assumption that funds generate market returns before fees is broadly consistent with the experience of the average investor.

[11] The average expense ratios are from the 2019 Morningstar Annual U.S. Fund Fee Study at https://www.morningstar.com/lp/annual-us-fund-fee-study.

(2018a) emphasize that the distribution of continuously compounded returns from the U.S. is close to normal for horizons beyond ten years, such that the wealth distribution is approximately lognormal. We study this issue in the context of our developed country sample and find that the distribution of long-horizon returns is highly non-normal. Even at a 30-year horizon, cumulative log returns exhibit considerable negative skewness and excess kurtosis. These results reflect the pronounced left tail of the return distribution and the associated downside risk for long-term investors.

For comparison, Panel B of Table 4 repeats the analysis in Panel A in terms of real USD payoffs. This approach takes the perspective of a global USD investor whose nominal returns in a given country are adjusted for both exchange rate changes for the local currency vis-à-vis the U.S. dollar and U.S. inflation. All of the conclusions from Panel A based on real local currency returns continue to hold in Panel B. Although the median investment outcomes are similar in Panels A and B, the exchange rate risk reflected in Panel B generates a higher standard deviation of the payoff distribution at each horizon. This increased dispersion also leads to more extreme tail outcomes. For example, the 5th and 95th percentiles of the 30-year real USD payoffs are \$0.34 and \$28.51, respectively, whereas the corresponding figures for real local currency payoffs are \$0.47 and \$23.30. The loss probabilities reported in Panel B are also somewhat higher than the values in Panel A. At the 30-year horizon, the probability of loss for a global USD investor is 14.0% in real terms.

The results for real USD payoffs in Panel B show that our conclusions are, if anything, even stronger for hypothetical global investors who bear exposure to changes in currency exchange rates. To conserve space, the remainder of this section focuses on real local currency returns. We present the corresponding results for real USD returns in the Internet Appendix.

4.2. Alternative samples

The base case examined in Section 4.1 incorporates information from a broad cross section of 39 developed economies over a long sample period from 1841 to 2019. In this section, we compare our base case results with those from research designs that rely on alternative underlying samples. These alternative samples fall into two categories.

First, we construct samples that are subject to survivor and easy data biases to highlight the quantitative importance of our sample formation procedure. As described in Section 2, we take several precautions in forming our base sample to avoid conditioning on eventual market outcomes or the lack of stock market interruptions. These design features include the ex ante identification of developed countries, the calculation of multi-month return observations to account for the effects of market disruptions on investment performance, and the inclusion of a terminal return for Czechoslovakia. Prior studies, in contrast, often focus on investment outcomes using biased samples in terms of survival and easy data availability. There is a large literature, for example, that estimates mean nominal returns, mean real returns, or risk premiums using data from a single country's stock market. As noted by Dimson et al. (2002), many of these studies use data samples concentrated in more recent periods as a result of researchers' preference for data that are not obfuscated by trading halts, wars, hyperinflation, and other extreme events. In Section 4.2.1, we contrast our base case results with those from samples based on the experience of a single country. In Section 4.2.2, we consider broader samples of developed countries that nonetheless condition on end-of-sample survival or the availability of continuous data.

Second, we construct samples to evaluate the robustness of our findings to reasonable sample formation screens. In particular, our base case uses data from a broad cross section of countries over a long sample period. This approach invites potential concerns that the results are driven by data from periods that are less relevant for the prospects of today's investors or by data from countries with smaller stock markets. Sections 4.2.3 and 4.2.4 compare our base case results with those from samples that condition on more recent data, larger populations, and better-capitalized equity markets.

Table 5 provides an overview of the samples considered in the paper. In each case, the table includes a description of the sample formation criteria and summary statistics for the sample size. Panel A details our base case sample of all developed countries. Panels B to F detail the alternative samples, each of which covers a subset of the base case sample. We provide additional descriptions of the sample formation procedures throughout this section and in the Internet Appendix.

4.2.1. Samples based on individual countries

The first two alternative samples are based on a single developed country. We specifically consider samples from the United States over the period from 1890 to 2019 and the United Kingdom over the period from 1841 to 2019. Panel B of Table 5 shows that these samples cover just 4.9% and 6.7%, respectively, of the total months from our developed sample. Brown et al. (1995) note that the U.S. and U.K. markets did not experience the major interruptions that often prevent other countries from being included in studies on long-term returns. As such, benchmarking our base case results to results for these single-country samples provides initial evidence of the effects of conditioning on survival and easy data.

Panel B of Table 6 shows properties of the distribution of real payoffs for the U.S. and U.K. samples. The corresponding full-sample results from Panel A of Table 4 are repeated in Panel A for ease of comparison. The discussion in this section focuses on bootstrap distributions of payoffs at the 30-year horizon. We report results for other horizons in the Internet Appendix.

Table 6 reveals pronounced differences between the distributions based on the full developed sample and those based on the individual countries. The mean of the payoff distribution based on the full developed sample ($7.38) leads the estimate based on the U.K. ($5.28), but trails the estimate from the U.S. ($8.91). The standard deviation of real payoffs from the full sample of $13.76, however, is drastically larger than the estimates based on the U.S. or the U.K. This difference in variance is also reflected in the tail outcomes. The 5th and 95th percentiles from the base case full sample estimation, for example, are $0.47 and $23.30. The corresponding figures are $1.69 and $24.30 for the U.S. sample and $1.24 and $12.97 for the U.K. sample.

Fig. 4 plots the histograms of real payoffs from the base case simulation with those from the alternative simulations with biased samples. The greater dispersion in the distribution based on the full developed sample is evident relative to the single-country distributions for the U.S. (Panel A) and the U.K. (Panel B).

Comparisons of the payoff percentiles across Panels A and B in Table 6 reveal economically large differences in estimated downside risk faced by stock market investors. The full sample distribution in Panel A, for example, suggests that disastrous outcomes for equity investors are relatively common. The 1st percentile of real payoff is \$0.14, and the 5th percentile is just \$0.47. In contrast, the 1st percentile for the U.S. sample is \$0.96 and the value for the U.K. is \$0.69. The 5th percentile for each single-country distribution is well above \$1.00.

Finally, the long-horizon loss probabilities for the single-country samples are small relative to those for the full sample. Investors lose on equity investments in real terms over a 30-year holding period 1.2% of the time based on the U.S. sample and 3.0% of the time based on the U.K. sample. The estimated distributions based on the U.S. and U.K. samples are much more in accord with the view that stocks are safe investments for long-term, buy-and-hold investors.

Based on these findings, should investors in the U.S. take solace and solely concentrate on the distribution from U.S. data? Our view is that U.S. investors should carefully consider the information from the broad developed country sample while forming ex ante expectations of investment performance. We have four bases for our assertion. First, the U.S. stock market experienced a large, unexpected increase in equity valuations over the past century, such that the historical performance may be optimistic relative to ex ante expectations (Fama and French, 2002; Avdis and Wachter, 2017). The developed country sample includes a variety of trends in valuations that may be informative about possible valuation changes in the future. Second, the short return history of the U.S. provides relatively little information about tail risk in long-term returns. Experiences in other countries can provide insights into the frequency and the distribution of rare events (Jorion and Goetzmann, 1999; Barro, 2006). Third, the use of historical U.S. data is subject to survivor and easy data biases (Brown et al., 1995; Dimson et al., 2002). Long-term stock returns are often shaped by unforeseen political and economic events. Brown et al. (1995) contend, "The fact that most of the continuous markets are in former British colonies such as Australia, Canada, India, South Africa, and the United States is almost certainly an accident of political, and perhaps legal, history. Had the outcome of either world war been different, we might currently be studying the long-term behavior of continental European exchanges." Fourth, faced with alternative probability distributions for long-horizon returns, ambiguity averse individuals may prefer to base their decisions on the least optimistic case (Gilboa and Schmeidler, 1989). If an investor views the estimates from both the developed country sample and the U.S. sample as plausible probability distributions for future U.S. returns, ambiguity aversion would likely cause her to use the developed country distribution because of its relative pessimism about potential losses.

4.2.2. Samples conditioned on survival and continuous data

Panel C of Table 6 presents results for biased samples that, like our base developed country sample, are broad in terms of cross-sectional coverage. In forming these samples, however, we intentionally relax the measures previously put in place to mitigate survivor and easy data problems. The survival sample conditions on 2019 membership in the OECD to consider the effect of survivor bias. Relative to our full sample, the survival sample excludes Argentina, Czechoslovakia, Singapore, and the early development period in Chile. The continuous data sample appeals to the concerns of Brown et al. (1995) and Dimson et al. (2002) about research designs that avoid significant disruptions in return data. In addition to conditioning on 2019 membership in the OECD as in the survival sample, the continuous sample limits each country's return history to the period that avoids major market interruption. In particular, for each country we drop all multi-month observations listed in Table 2 as well as any returns that occur before these observations. The continuous sample only includes single-month observations as a result of this change, and each country's time series represents the continuous monthly return period that extends through 2019. Table 5 shows that the survival sample covers 95.0% of the months in the full sample, whereas coverage for the continuous sample is lower at 71.3%.

The results in Table 6 suggest that conditioning on survival alone has a minimal impact on the estimated distribution of 30-year payoffs. The mean, standard deviation, and percentiles of the payoff distribution are similar for the full sample and the survival sample. Relative to the full sample loss probability of 12.1%, the survival sample loss probability is only marginally lower at 11.1%. These results are perhaps unsurprising given the significant overlap in the underlying data for these two samples shown in Table 5.

Conditioning on both survival and continuous data has a more marked impact. The full sample mean of \$7.38 and standard deviation of \$13.76 are both much lower compared with the continuous sample mean of \$9.78 and standard deviation of \$16.71. Given that the payoff distribution is bounded below by zero, the higher values of mean and standard deviation for the continuous sample are consistent with a bias away from negative and toward positive stock market performance from conditioning on continuous return data. Each percentile of the payoff distribution for the continuous sample exceeds the corresponding value for the full sample. These effects are visibly apparent in Panel D of Fig. 4, as the continuous sample distribution exhibits a dampened left tail and exaggerated probabilities of large gains relative to the full sample distribution. Finally, the continuous distribution suggests substantially muted left tail outcomes, as the loss probability is only 5.2%. In sum, the results for the continuous sample underscore the importance of prudent sample selection techniques. Conditioning on continuous monthly return availability appears to be particularly problematic for assessing the distribution of long-term investment outcomes.

4.2.3. Samples conditioned on data recency

Our base case results in Section 4.1 rely on the full cross section of developed countries, with the earliest return observation dating back to January 1841. This sample provides an expansive view of outcomes from the price formation process in developed countries. One possible concern with this approach, however, is that the experience of investors in the early portion of the sample is less relevant to investment prospects looking forward. Markets at the beginning and the end of our sample period differ in terms of number of listed securities, concentration of firms across industries, trading technology, availability of pricing and financial information on listed firms, trading regulations, investor protections, and many other features. International equity markets in the late nineteenth and early twentieth centuries were also highly concentrated in railroadrelated stocks (Dimson et al., 2002). To address this concern, we consider how our results change if we exclude data from earlier years in the sample.

Panel B of Table 7 reports properties of the distribution of real payoffs for samples that start 40 years apart in 1880, 1920, 1960, and 2000. The base case sample in Panel A starts in 1841, roughly 40 years before the first of these alternative samples. In each case, we use the full cross section of developed countries, but we exclude all observations prior to the indicated start date. The results correspond to a 30-year investment horizon.

The figures reported in Table 7 reveal some discrepancies across samples. In Panel A of Table 7, the 1841 sample has a mean 30-year real payoff of $7.38 and a standard deviation of $13.76. The means and standard deviations for the samples excluding data prior to 1880, 1920, and 1960 are all higher than the corresponding base case values. The means for these three alternative samples range from $7.41 to $8.73, and the standard deviations range from $15.45 to $20.04. The mean ($4.33) and standard deviation ($7.09) for the post-2000 sample, in contrast, are both lower than the corresponding values for the base case. Nonetheless, our conclusions that long-term stock market investments are associated with substantial uncertainty about real payoffs and with high probabilities of economically large losses are robust to the choice of sample start. The bootstrap distributions for the four alternative samples in Panel B have 1st percentile payoffs that range from $0.14 to $0.23 and 5th percentile payoffs that range from $0.39 to $0.66. As with the 1841 base case sample, these outcomes represent catastrophic results for investors.

Fig. 5 presents loss probability as a function of the sample start date for each investment horizon. We consider annual start dates from 1841 to 2000. Each panel of the figure plots two loss probabilities: one for the full sample of developed countries and one for the developed United States sample (starting in 1890). The loss probabilities for shorter investment horizons are relatively insensitive to the sample start. The one-month loss probability for the full developed sample in Panel A, for example, ranges from 41.8% to 44.2%. At the one-year horizon in Panel B, the range is 33.9% to 38.5%.

The sample start has a larger impact on the results for longer investment horizons. The 30-year loss probabilities based on the full developed sample in Panel F of Fig. 5 range from 3.8% (for the sample that starts in 1983) to 18.0% (for the sample that starts in 2000). One prominent feature of the plot in Panel F is that the variation in loss probability becomes much more sensitive to the start date later in the sample period. This result is intuitive, as excluding a good or a bad year of investment performance from a progressively shorter sample has a more pronounced impact on simulated returns. Panel F also shows large drops in loss probabilities as sample start dates progress through World War I and World War II, which is direct evidence of easy data bias for studies that exclude the wartime data. The lowest loss probabilities are concentrated around a 1980 start date, but these loss probabilities of approximately 4.0% have higher bookends of 9.7% with a 1970 start date and 8.1% with a 1990 start date. Most important, the choice of sample start has no impact on the broad conclusion that long-horizon equity investments appear riskier from the perspective of an investor who relies on the historical record of all developed stock markets relative to an investor who relies solely on the historical U.S. experience.

4.2.4. Samples conditioned on population or equity market size

Another potential concern with the evidence from our base case design is that the results are driven by countries that are, by some measure, less economically relevant. Even among developed economies, there are notable discrepancies in population, participation in financial markets, and equity market size relative to economic output. The estimates of long-horizon risk from the base case could be less pertinent for investors in the world's largest equity markets, for example, if country-level dispersion in equity returns were strongly associated with measures of economic relevance.

We address this issue by considering developed samples with additional screens based on population or equity market size. These alternative samples are summarized in Panels E and F of Table 5. We examine two samples based on population and two based on the ratio of equity market capitalization to GDP. Country-level data on population, equity market capitalization, and GDP are from the GFDatabase. To form the samples based on population, we start with our full sample of developed countries, but exclude data for a given country prior to the year in which the country's population reaches a threshold percentage of the total world population. We consider threshold values of 0.2% and 0.5%. As shown in Table 5, these samples cover 73.5% and 40.0%, respectively, of the period for the full developed sample. We use an analogous approach to construct the samples based on market capitalization-to-GDP ratio, with threshold values of 0.5 and 1.0 that produce coverage of 64.5% and 40.5%, respectively.

In the Internet Appendix, we report the initial year of inclusion by country for the samples based on population and equity market size. Seven countries—the Czech Republic, Hungary, Slovakia, Estonia, Slovenia, Latvia, and Lithuania—are entirely excluded from all four alternative samples. The U.S. exceeds the 0.5% population benchmark (as well as the 0.2% population benchmark) at the beginning of its developed period in 1890. The U.S. market capitalization-to-GDP ratio first reaches 0.5 in 1927 and 1.0 in 1998.

Panels C and D of Table 7 report distributional statistics for 30-year real payoffs for the alternative samples based on population and equity market size. The results for these samples, which impose tighter restrictions on economic relevance, are similar to those for the base case design. Each of the four alternative samples reflects considerable risk of investment loss in real terms. The 1st percentile investment outcome ranges from \$0.09 to \$0.24, and the 5th percentile ranges from \$0.33 to \$0.68. Across samples, the loss probability is also in line with the value from the base case and reaches as high as 14.7% for the sample of countries that satisfy the 0.5% population screen.

Overall, the empirical findings in Sections 4.2.3 and 4.2.4 demonstrate the robustness of the results. Conditioning on relatively recent data, on economies with larger populations, or on economies with more developed equity markets has little impact on our conclusions about the risks faced by long-horizon investors.

5. Asset allocation implications

The distributions of cumulative wealth from buy-and-hold stock market investments in Section 4 indicate a significant risk of loss. To provide insight on the economic implications of this finding, we examine a simple portfolio choice problem with a single risky asset and a risk-free asset. We take the perspective of an investor in a developed country and assume that the risky asset is the investor's domestic stock market index. The investor uses information from our broad developed country sample to form expectations about stock market performance. For comparison, we also consider an investor who uses the distribution from the U.S. sample.

We introduce two alternative specifications for the risk-free asset: (i) an inflation-protected risk-free asset and (ii) cash. The inflation-protected risk-free asset is risk free in real terms, whereas cash is risk free only in nominal terms. The inflation-protected risk-free asset maintains its real principal balance, but earns no interest such that a \$1.00 investment retains this value over any holding period. This hypothetical asset corresponds to the baseline of \$1.00 of real wealth that we plot against the cumulative wealth distributions in Section 4. Cash is nominally risk free, but bears inflation risk. Given that inflation is typically positive, a cash position is expected to lose real value over time. For example, the expected real value of \$1.00 in cash held for 30 years is about \$0.35 in our developed country sample. A cash position also carries the risk of realizing a very large real loss if high inflation occurs during the holding period.

The two-asset portfolio choice problem is simple by design to isolate the impact of investors' views about the distribution of domestic stock market performance. In reality, investors also have access to other asset classes. Home bias in stock holdings is well documented in the literature (e.g., French and Poterba, 1991), but investors have the ability to diversify using international equity markets. Government and corporate bonds are also available to investors. Short-term government debt is often used in portfolio choice studies as a risk-free asset, but government bills have occasionally experienced sharp losses that correlate to large crashes in equity markets (see, e.g., Barro, 2006). We opt for the risk-free alternatives discussed above to avoid default risk and data availability issues for government bills in our developed country sample. Including these alternative asset classes alongside the domestic stock market in a multi-asset portfolio choice problem is an interesting topic that is beyond the scope of this paper.

We study asset allocation for an investor with exponential utility over terminal wealth. This investor maximizes expected utility by allocating to the stock market and the risk-free asset,

$$\max_{w} \mathbb{E}\left[\frac{1-e^{-aW_H(w)}}{a}\right], \tag{9}$$

where a is the risk aversion coefficient, w is the weight in stocks, and $W_H(w)$ is real wealth at an H-month horizon given weight w and initial investment of \$1.00. The portfolio is rebalanced monthly between the stock market and the risk-free asset to maintain weights of w and $1-w$, respectively. We numerically solve for the optimal weight by calculating expected utility with respect to the bootstrap distribution of $W_H(w)$.[12] When cash is the risk-free asset, we use realized inflation corresponding to each stock return observation in our bootstrap to determine the real value of cash. The risk aversion parameter a is set to three, and inferences are robust to alternative values. Full details on the calculation of cumulative wealth and results for an alternative investment pattern with an annuity structure are available in the Internet Appendix.

Table 8 shows optimal weights in stocks for investors who use the developed country sample, w_d, and the U.S. sample, w_{us}. To assess the economic importance of differences in optimal weights between the developed country and U.S. cases, we calculate the maximum annualized fee on wealth that each developed country investor would be willing to pay to adopt her own optimal policy rather than that of the U.S. investor. This fee equalizes the developed country investor's utility across using her optimal weight subject to the fee and using the optimal weight from the U.S. data without the fee.

Table 8 shows that the investors using the developed country sample invest less in stocks relative to the U.S. sample investors. At the 30-year horizon, for example, the developed country investor with access to the inflation-protected risk-free asset chooses an optimal weight of 43% in stocks compared with a weight of 75% for the U.S. investor. The developed country investor would be willing to incur a fee up to 0.45% per year to maintain the lower weight. Across all cases, the investors relying on the developed country sample adopt lower optimal weights relative to the U.S. sample investors, and the fees range from 0.10% to 1.02% per year. Consistent with intuition, the increased risk of stocks based on the broad developed country analysis leads to more conservative stock investments, and the differences relative to inferences from the U.S. are economically large.

Table 8 also shows that investors with access to cash invest more in stocks compared with those having access to the inflation-protected risk-free asset. The 30-year developed country investor, for example, chooses a weight of 99% in stocks with cash as the alternative versus just 43% in stocks when the inflation-protected risk-free asset is available. Although cash is nominally risk free, inflation exposes cash to expected real losses and risk. To isolate the effect of the expected real loss, we consider an investor with access to the stock market and an asset for which an investment of \$1.00 produces a guaranteed real payoff of \$0.35 after 30 years, which is the expected real value of a cash investment. This investor has an optimal weight in stocks of 82%, such that much of the difference between the allocations of 43% with the inflation-protected risk-free asset and 99% with cash is attributable to the expected loss from cash. The remaining difference is attributable to the inflation risk of cash and the possibility that stocks provide a hedge against inflation. Overall, the comparisons of the results with real and nominal risk-free assets indicate that stocks are useful for preserving real buying power, such that an investor without access to a truly risk-free asset in real terms

[12] We specify exponential utility rather than power utility for numerical stability. The large left tail in the developed country bootstrap distribution produces occasional draws of extremely low power utility given that this utility function is not bounded below as wealth approaches zero. This feature of power utility causes the optimization to be sensitive to the particular set of bootstrap draws. Using exponential utility avoids the numerical problem because this utility function is bounded between zero and $1/a$. Notwithstanding this issue, inferences from power utility across repeated runs appear to be similar to those from exponential utility.

may still optimally choose a large weight in the stock market despite its considerable risk.

The portfolio choice results suggest that accounting for the broad developed country sample is important for asset allocation. A heightened sense of risk leads to lower stock allocations under the developed country sample relative to the U.S. sample, and the differences are economically important. Our results do not, however, indicate that investors should avoid stocks altogether because of the risk of loss. The availability of alternative assets that provide protection against both inflation and loss of principal is an important factor in determining the attractiveness of stocks.

6. Conclusion

We use a stationary block bootstrap procedure to characterize the distributions of long-horizon returns and payoffs for stock market investors. The bootstrap simulations resample returns from a broad cross section of 39 developed countries over the period from 1841 to 2019. As such, we significantly expand the cross-sectional and time-series dimensions of the samples considered in prior studies (e.g., Fama and French, 2018a,b). This approach allows us to combat concerns over biases attributable to survival (Brown et al., 1995) and easy data availability (Dimson et al., 2002).

Our analysis yields three primary findings. First, the long-term outcomes from diversified equity investments are highly uncertain. Based on the historical record of stock market performance in developed markets, the 5th percentile real payoff (measured in terms of local currency) from a \$1.00 buy-and-hold investment over 30 years is \$0.47, whereas the 95th percentile is \$23.30. This evidence stands in contrast to the conventional view that mean reversion in equity returns makes equity investing relatively safe at long horizons. Second, catastrophic investment outcomes are common even with a 30-year horizon, as the 1st percentile real payoff is \$0.14 and the 10th percentile is just \$0.85. An investor at age 35 saving for retirement, for example, only realizes one draw from the 30-year return distribution, and we estimate a 12.1% chance that this investor will lose relative to inflation. Third, the empirical findings based on the historical record of stock market performance across dozens of developed markets are notably different from those based on the historical U.S. experience. Estimates that rely solely on U.S. data suggest that long-term real investment losses are rare. The contrast in results highlights the importance of guarding against survivor and easy data biases in assessing the distribution of distant payoffs and also has economically significant implications for optimal portfolio choice.

Our study contributes to a large literature that attempts to mitigate survivor concerns with U.S. stock market data by considering the historical record in other markets. Much of this work focuses on estimating average returns and equity premiums (e.g., Dimson et al., 2002; Jordà et al., 2019). These studies yield valuable insights, for example, to managers needing inputs for cost of capital calculations. These insights on expected performance remain important for long-term investors. Given the single-draw nature of long-horizon returns for retirement savers, however, we must also consider higher-order moments and the likelihood and magnitude of extreme tail events. Our study serves as a guide to assessing these potential outcomes, and our findings indicate that investors should carefully consider the significant risk of loss before betting aggressively on equity.

References

Asness, C.S., Moskowitz, T.J., Pedersen, L.H., 2013. Value and momentum everywhere. Journal of Finance 68, 929–985.

Avdis, E., Wachter, J.A., 2017. Maximum likelihood estimation of the equity premium. Journal of Financial Economics 125, 589–609.

Avramov, D., Cederburg, S., Lučivjanská, K., 2018. Are stocks riskier over the long run? Taking cues from economic theory. Review of Financial Studies 31, 556–594.

Barberis, N.C., 2000. Investing for the long run when returns are predictable. Journal of Finance 55, 225–264.

Barro, R.J., 2006. Rare disasters and asset markets in the twentieth century. Quarterly Journal of Economics 121, 823–866.

Barro, R.J., Jin, T., 2011. On the size distribution of macroeconomic disasters. Econometrica 79, 1567–1589.

Barro, R.J., Ursúa, J., 2008. Consumption disasters in the twentieth century. American Economic Review: Papers & Proceedings 98, 58–63.

Barro, R.J., Ursúa, J., 2012. Rare macroeconomic disasters. Annual Review of Economics 4, 83–109.

Barro, R.J., Ursúa, J., 2017. Stock-market crashes and depressions. Research in Economics 71, 384–398.

Bolt, J., Inklaar, R., de Jong, H., van Zanden, J.L., 2018. Rebasing 'Maddison': New income comparisons and the shape of long-run economic development. Unpublished working paper. University of Groningen.

Brown, S.J., Goetzmann, W.N., Ross, S.A., 1995. Survival. Journal of Finance 50, 853–873.

Brown, S.J., Warner, J.B., 1980. Measuring security price performance. Journal of Financial Economics 8, 205–258.

Brown, S.J., Warner, J.B., 1985. Using daily stock returns: The case of event studies. Journal of Financial Economics 14, 3–31.

Calvo-Gonzalez, O., 2006. Neither a carrot nor a stick: American foreign aid and economic policymaking in Spain during the 1950s. Diplomatic History 30, 409–438.

Campbell, J.Y., Shiller, R.J., 1988. The dividend-price ratio and expectations of future dividends and discount factors. Review of Financial Studies 1, 195–228.

Campbell, J.Y., Viceira, L.M., 2002. Strategic Asset Allocation: Portfolio Choice for Long-Term Investors. Oxford University Press, Oxford, UK.

Carvalho, C.M., Lopes, H.F., McCulloch, R.E., 2018. On the long-run volatility of stocks. Journal of the American Statistical Association 113, 1050–1069.

Dimson, E., Marsh, P.R., Staunton, M., 2002. Triumph of the Optimists: 101 Years of Global Investment Returns. Princeton University Press, Princeton, NJ.

Doman, N.R., 1950. Compensation for nationalised property in post-war Europe. The International Law Quarterly 3, 323–342.

Engle, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50, 987–1007.

Fama, E.F., French, K.R., 1998. Value versus growth: The international evidence. Journal of Finance 53, 1975–1999.

Fama, E.F., French, K.R., 2002. The equity premium. Journal of Finance 57, 637–659.

Fama, E.F., French, K.R., 2018a. Long-horizon returns. Review of Asset Pricing Studies 8, 232–252.

Fama, E.F., French, K.R., 2018b. Volatility lessons. Financial Analysts Journal 74, 42–53.

French, K.R., Poterba, J.M., 1991. Investor diversification and international equity markets. American Economic Review: Papers & Proceedings 81, 222–226.

Gilboa, I., Schmeidler, D., 1989. Maxmin expected utility with non-unique prior. Journal of Mathematical Economics 18, 141–153.

Girardi, D., Bowles, S., 2018. Institution shocks and economic outcomes: Allende's election, Pinochet's coup and the Santiago stock market. Journal of Development Economics 134, 16–27.

Glos, G.E., 1986. The Czechoslovak Civil Code of 1964 and its 1982 amendment within the framework of Czechoslovak civil law. New York Law School Journal of International and Comparative Law 6, 215–273.

Goetzmann, W.N., Jorion, P., 1995. A longer look at dividend yields. Journal of Business 68, 483–508.

Golez, B., Koudijs, P., 2018. Four centuries of return predictability. Journal of Financial Economics 127, 248–263.

Ibbotson, R.G., Sinquefield, R.A., 1976. Stocks, bonds, bills, and inflation: Simulations of the future (1976-2000). Journal of Business 49, 313–338.

Jordà, Ò., Knoll, K., Kuvshinov, D., Schularick, M., Taylor, A.M., 2019. The rate of return on everything, 1870–2015. Quarterly Journal of Economics 134, 1225–1298.

Jorion, P., Goetzmann, W.N., 1999. Global stock markets in the twentieth century. Journal of Finance 54, 953–980.

Kosowski, R., Timmermann, A., Wermers, R., White, H., 2006. Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis. Journal of Finance 61, 2551–2595.

Kuznets, S., 1973. Modern economic growth: Findings and reflections. American Economic Review 63, 247–258.

Lewellen, J., 2011. Institutional investors and the limits of arbitrage. Journal of Financial Economics 102, 62–80.

Linnainmaa, J.T., Roberts, M.R., 2018. The history of the cross-section of stock returns. Review of Financial Studies 31, 2606–2649.

Lo, A.W., MacKinlay, A.C., 1988. Stock market prices do not follow random walks: Evidence from a simple specification test. Review of Financial Studies 1, 41–66.

Lundblad, C., 2007. The risk return tradeoff in the long run: 1836-2003. Journal of Financial Economics 85, 123–150.

Martin, I.W.R., Wagner, C., 2019. What is the expected return on a stock? Journal of Finance 74, 1887–1929.

McLean, R.D., Pontiff, J., 2015. Does academic research destroy stock return predictability? Journal of Finance 71, 5–32.

Mehra, R., Prescott, E.C., 1985. The equity premium: A puzzle. Journal of Monetary Economics 15, 145–161.

Mitchell, B.R., 1993. International Historical Statistics: The Americas, 1750-1988. 2nd ed., Macmillan Reference, London.

Mitchell, B.R., 1995. International Historical Statistics: Africa, Asia & Oceania, 1750-1988. 2nd ed., Macmillan Reference, London.

Mitchell, B.R., 1998. International Historical Statistics: Europe, 1750-1993. 4th ed., Macmillan Reference, London.

Nakamura, E., Steinsson, J., Barro, R., Ursúa, J., 2013. Crises and recoveries in an empirical model of consumption disasters. American Economic Journal: Macroeconomics 5, 35–74.

Pástor, Ľ., Stambaugh, R.F., 2012. Are stocks really less volatile in the long run? Journal of Finance 67, 431–477.

Patton, A.J., Timmermann, A., 2010. Monotonicity in asset returns: New tests with applications to the term structure, the CAPM, and portfolio sorts. Journal of Financial Economics 98, 605–625.

Politis, D.N., Romano, J.P., 1994. The stationary bootstrap. Journal of the American Statistical Association 89, 1303–1313.

Poterba, J., Rauh, J., Venti, S., Wise, D., 2007. Defined contribution plans, defined benefit plans, and the accumulation of retirement wealth. Journal of Public Economics 91, 2062–2086.

Poterba, J.M., Summers, L.H., 1988. Mean reversion in stock prices: Evidence and implications. Journal of Financial Economics 22, 27–59.

Rebolledo Leyton, R., 2005. La crisis económica de 1967 en el contexto de la ruptura del sistema democrático. Universum 20, 124–139.

Rietz, T.A., 1988. The equity premium: A solution. Journal of Monetary Economics 22, 117–131.

Schwert, G.W., 2003. Anomalies and market efficiency, in: Constantinides, G.M., Harris, M., Stulz, R.M. (Eds.), Handbook of the Economics of Finance. North-Holland, pp. 939–974.

Shiller, R.J., 2003. Social Security and individual accounts as elements of overall risk-sharing. American Economic Review: Papers & Proceedings 93, 343–347.

Shiller, R.J., 2006. Life-cycle personal accounts proposal for Social Security: An evaluation of President Bush's proposal. Journal of Policy Modeling 28, 427–444.

Siegel, J., 2014. Stocks for the Long Run. McGraw-Hill, New York, NY.

Silber, W.L., 2005. What happened to liquidity when World War I shut the NYSE? Journal of Financial Economics 78, 685–701.

Sullivan, R., Timmermann, A., White, H., 1999. Data-snooping, technical trading rule performance, and the bootstrap. Journal of Finance 54, 1647–1691.

Viceira, L.M., 2008. Life-cycle funds, in: Lusardi, A. (Ed.), Overcoming the Saving Slump: How to Increase the Effectiveness of Financial Education and Saving Programs. Chicago Press, pp. 140–177.

Winkler, P., 1994. The Czechoslovak presidential decrees, 1940-1945. Perspectives 4, 13–23.

Table 1
Developed country sample periods

The table shows developed countries, initial development dates, classification reasons for development, and sample periods. The development year classifications are based on agricultural labor share, organizational membership in the Organisation for European Economic Co-operation (OEEC) or the Organisation for Economic Co-operation and Development (OECD), or the opening of the first stock exchange in an already developed economy. The sample period start date is the later of the development date and the first date with return data.

Country	Development year	Development classification	Sample period start date	Sample period end date
United Kingdom	1841	Agricultural labor share	1841:01	2019:12
Netherlands	1849	Agricultural labor share	1914:01	2019:12
Belgium	1856	Agricultural labor share	1897:01	2019:12
France	1866	Agricultural labor share	1866:01	2019:12
Norway	1881	Establishment of exchange	1914:02	2019:12
Germany	1882	Agricultural labor share	1882:01	2019:12
Denmark	1890	Agricultural labor share	1890:01	2019:12
Switzerland	1890	Agricultural labor share	1914:01	2019:12
United States	1890	Agricultural labor share	1890:01	2019:12
Canada	1891	Agricultural labor share	1891:01	2019:12
Argentina	1895	Agricultural labor share	1947:02	1966:12
New Zealand	1896	Agricultural labor share	1896:01	2019:12
Australia	1901	Agricultural labor share	1901:01	2019:12
Sweden	1910	Agricultural labor share	1910:01	2019:12
Austria	1920	Agricultural labor share	1925:02	2019:12
Chile period I	1920	Agricultural labor share	1927:01	1970:12
Greece	1920	Agricultural labor share	1977:01	2019:12
Czechoslovakia	1921	Agricultural labor share	1926:01	1946:06
Japan	1930	Agricultural labor share	1930:01	2019:12
Portugal	1930	Agricultural labor share	1934:01	2019:12
Italy	1931	Agricultural labor share	1931:01	2019:12
Ireland	1936	Agricultural labor share	1936:01	2019:12
Singapore	1947	Agricultural labor share	1970:01	2019:12
Luxembourg	1948	OEEC membership	1982:01	2019:12
Turkey	1948	OEEC membership	1986:02	2019:12
Spain	1959	OEEC membership	1959:01	2019:12
Finland	1969	OECD membership	1969:01	2019:12
Iceland	1991	Establishment of exchange	2002:01	2019:12
Mexico	1994	OECD membership	1994:01	2019:12
Czech Republic	1995	OECD membership	1995:01	2019:12
Hungary	1996	OECD membership	1996:01	2019:12
Poland	1996	OECD membership	1996:01	2019:12
South Korea	1996	OECD membership	1996:01	2019:12
Slovakia	2000	OECD membership	2000:01	2019:12
Chile period II	2010	OECD membership	2010:01	2019:12
Estonia	2010	OECD membership	2010:01	2019:12
Israel	2010	OECD membership	2010:01	2019:12
Slovenia	2010	OECD membership	2010:01	2019:12
Latvia	2016	OECD membership	2016:01	2019:12
Lithuania	2018	OECD membership	2018:01	2019:12

Table 2
Multi-month returns

The table details periods of multi-month returns. For each multi-month return observation, the table reports the number of months, the start and end dates of the period, and the nominal and real stock market returns earned over the period. The nominal and real returns are from the perspective of a domestic investor in a representative country, and the real USD returns are from the perspective of a global USD investor. Panels A and B show events corresponding to World War I and World War II, respectively, Panel C shows periods with revolutions, Panel D shows financial and banking crises, and Panel E shows labor strikes.

Country	Months	Start date	End date	Nominal return (%)	Real return (%)	Real USD return (%)
Panel A: World War I						
Australia	6	1914:08	1915:01	−0.45	−0.39	−12.92
Belgium	52	1914:08	1918:11	25.12	−55.91	−36.08
Canada	7	1914:08	1915:02	1.38	−3.59	0.73
Denmark	4	1914:08	1914:11	−2.42	−3.37	−10.67
France	6	1914:08	1915:01	−3.68	−21.68	−16.30
Germany	42	1914:08	1918:01	20.03	−38.87	−41.21
Netherlands	7	1914:08	1915:02	−1.23	−3.50	3.75
Norway	3	1914:08	1914:10	−3.26	−3.81	−14.58
Sweden	4	1914:08	1914:11	−5.91	−8.96	−23.02
Switzerland	24	1914:08	1916:07	−11.00	−27.78	−20.12
United Kingdom	6	1914:08	1915:01	−0.26	−3.30	−12.58
United States	5	1914:08	1914:12	−2.14	−3.11	−3.11
Panel B: World War II						
Austria	2	1938:04	1938:05	6.34	5.62	6.74
Austria	113	1939:07	1948:11	300.73	−19.66	−82.31
Belgium	5	1940:06	1940:10	22.38	12.54	22.51
Belgium	11	1944:08	1945:06	−0.29	−17.08	−31.24
Czechoslovakia	16	1938:10	1940:01	31.95	16.66	54.77
Czechoslovakia	4	1942:01	1942:04	20.59	12.25	16.24
Denmark	2	1940:05	1940:06	−7.64	−10.67	−7.97
France	2	1939:09	1939:10	−2.96	0.53	−10.04
France	10	1940:06	1941:03	94.57	75.61	118.79
Germany	67	1943:01	1948:07	−87.62	−91.10	−99.36
Japan	45	1945:09	1949:05	449.38	−87.15	−95.04
Netherlands	5	1940:05	1940:09	20.63	15.21	21.36
Netherlands	21	1944:09	1946:05	−14.33	−33.15	−42.30
Norway	2	1940:04	1940:05	−16.75	−17.98	−17.05
Switzerland	2	1940:06	1940:07	−3.57	−5.11	−1.71
Panel C: Revolution						
Czechoslovakia	39	1943:04	1946:06	−90.00	−93.90	−95.42
Portugal	35	1974:05	1977:03	−80.39	−89.24	−90.60
Panel D: Financial or banking crisis						
Austria	2	1931:10	1931:11	6.86	6.20	8.35
Germany	2	1931:08	1931:09	−24.58	−23.01	−24.73
Germany	7	1931:10	1932:04	−8.22	1.78	0.87
Greece	2	2015:07	2015:08	−21.53	−20.13	−20.87
Panel E: Labor strike						
France	2	1974:04	1974:05	−6.17	−8.76	−10.10
France	2	1979:03	1979:04	12.79	10.69	7.93

Table 3
Summary statistics

The table reports summary statistics for monthly returns for each developed country and for the pooled sample of all observations. For each country, the table shows the number of sample months, the number of sample observations (i.e., including multi-month observations), the arithmetic average return ($\bar{R}_a$), the geometric average return ($\bar{R}_g$), the standard deviation of return (SD), return skewness (Skew), return kurtosis (Kurt), and the minimum (Min) and the maximum (Max) return. Panel A (Panel B) [Panel C] shows results for nominal returns (real returns) [real USD returns].

	Sample size		Summary statistics for returns						
Country	Months	Observ	$\bar{R}_a$ (%)	$\bar{R}_g$ (%)	SD (%)	Skew	Kurt	Min (%)	Max (%)
Panel A: Nominal returns									
Argentina	239	239	2.11	1.79	8.12	0.44	9.68	−41.47	45.64
Australia	1,428	1,423	0.97	0.89	3.88	−0.88	16.25	−42.13	22.14
Austria	1,139	1,025	0.93	0.66	10.81	20.84	578.05	−32.56	300.73
Belgium	1,476	1,411	0.74	0.62	5.18	0.53	8.75	−31.22	36.17
Canada	1,548	1,542	0.80	0.71	4.24	−0.59	7.45	−28.07	22.87
Chile period I	528	528	1.83	1.63	6.38	0.69	8.36	−31.00	38.27
Chile period II	120	120	0.31	0.22	4.08	0.16	3.08	−10.46	11.28
Czechoslovakia	246	190	0.42	−0.22	7.97	−7.31	90.31	−90.00	31.95
Czech Republic	300	300	1.09	0.81	7.47	−0.19	5.02	−29.44	30.08
Denmark	1,560	1,556	0.67	0.61	3.49	0.06	6.82	−18.47	18.80
Estonia	120	120	1.10	0.96	5.62	3.73	31.98	−11.07	44.82
Finland	612	612	1.34	1.15	6.31	0.20	6.41	−26.88	32.61
France	1,848	1,831	0.90	0.78	5.20	3.35	61.56	−21.82	94.57
Germany	1,656	1,542	0.80	0.47	8.27	3.48	77.49	−87.62	128.82
Greece	516	515	1.38	0.93	9.96	1.74	12.15	−27.83	68.46
Hungary	288	288	1.49	1.19	7.78	−0.07	6.61	−36.06	35.26
Iceland	216	216	0.72	0.29	7.61	−4.52	41.48	−71.52	18.08
Ireland	1,008	1,008	0.98	0.87	4.67	−0.22	7.84	−27.24	28.81
Israel	120	120	0.02	−0.10	4.73	−0.21	3.38	−14.24	12.78
Italy	1,068	1,068	1.18	0.91	7.61	1.69	13.02	−26.44	59.87
Japan	1,080	1,036	1.32	0.92	14.92	26.17	786.37	−20.26	449.38
Latvia	48	48	1.22	1.16	3.56	1.13	5.60	−5.38	13.95
Lithuania	24	24	0.39	0.36	2.62	−0.04	3.01	−5.55	5.03
Luxembourg	456	456	0.93	0.78	5.47	−0.72	6.47	−26.81	18.11
Mexico	312	312	1.23	1.04	6.23	−0.37	4.72	−26.43	19.71
Netherlands	1,272	1,242	0.80	0.67	5.10	0.44	13.52	−23.24	52.45
New Zealand	1,488	1,488	0.86	0.80	3.63	0.02	10.16	−28.29	25.00
Norway	1,271	1,268	0.85	0.72	5.07	−0.20	6.85	−27.42	26.10
Poland	288	288	0.94	0.71	6.82	0.23	7.04	−29.63	37.27
Portugal	1,032	998	0.97	0.65	7.98	2.63	48.05	−80.39	87.83
Singapore	600	600	0.97	0.69	7.40	0.39	9.90	−41.86	47.51
Slovakia	240	240	0.77	0.63	5.29	1.40	10.91	−18.54	33.75
Slovenia	120	120	0.41	0.33	3.97	0.42	5.16	−9.85	17.45
South Korea	288	288	0.96	0.63	8.27	1.12	9.55	−27.47	53.39
Spain	732	732	0.99	0.84	5.44	−0.02	5.08	−25.27	26.95
Sweden	1,320	1,317	0.89	0.78	4.77	−0.16	6.22	−27.11	27.58
Switzerland	1,272	1,248	0.65	0.56	4.29	−0.00	8.73	−24.62	33.78
Turkey	407	407	4.26	3.23	15.74	2.30	15.76	−40.67	127.17
United Kingdom	2,148	2,143	0.66	0.59	3.71	1.07	27.58	−26.51	54.10
United States	1,560	1,556	0.87	0.75	4.95	0.32	13.17	−29.63	42.89
Full sample	31,994	31,465	0.95	0.77	6.60	14.16	838.83	−90.00	449.38

(*continued on next page*)

Table 3 (*continued*)

Country	Sample size: Months	Sample size: Observ	Summary statistics for returns: $\bar{R}_a$ (%)	$\bar{R}_g$ (%)	SD (%)	Skew	Kurt	Min (%)	Max (%)
Panel B: Real returns									
Argentina	239	239	0.19	−0.18	8.53	0.09	7.65	−44.06	41.43
Australia	1,428	1,423	0.66	0.58	3.90	−0.94	16.07	−42.49	23.83
Austria	1,139	1,025	0.40	0.27	5.45	0.28	9.82	−32.63	38.96
Belgium	1,476	1,411	0.36	0.22	5.32	−0.57	15.22	−55.91	31.17
Canada	1,548	1,542	0.57	0.48	4.25	−0.53	7.18	−27.26	23.60
Chile period I	528	528	0.32	0.13	6.15	0.31	6.91	−32.81	30.28
Chile period II	120	120	0.05	−0.03	4.06	0.13	3.07	−10.54	11.05
Czechoslovakia	246	190	0.13	−0.68	7.96	−8.64	104.15	−93.90	16.66
Czech Republic	300	300	0.83	0.55	7.47	−0.19	4.96	−29.25	29.90
Denmark	1,560	1,556	0.39	0.33	3.54	−0.03	6.58	−18.38	18.89
Estonia	120	120	0.91	0.77	5.56	3.71	32.19	−11.43	44.23
Finland	612	612	0.98	0.78	6.31	0.26	6.38	−27.28	32.01
France	1,848	1,831	0.47	0.34	5.10	1.64	30.28	−22.01	75.61
Germany	1,656	1,542	0.65	0.29	8.41	3.27	75.16	−91.10	128.82
Greece	516	515	0.69	0.24	9.95	1.58	10.72	−27.83	65.50
Hungary	288	288	1.05	0.75	7.60	−0.25	6.21	−36.17	30.61
Iceland	216	216	0.37	−0.07	7.66	−4.45	40.76	−72.12	18.18
Ireland	1,008	1,008	0.57	0.46	4.67	−0.29	7.23	−27.26	25.54
Israel	120	120	0.06	−0.06	4.81	−0.17	3.33	−14.50	12.55
Italy	1,068	1,068	0.44	0.17	7.41	1.08	10.39	−34.89	58.61
Japan	1,080	1,036	0.54	0.30	5.96	−2.97	48.51	−87.15	23.22
Latvia	48	48	1.03	0.97	3.54	1.11	5.47	−5.47	13.73
Lithuania	24	24	0.21	0.18	2.61	−0.06	3.29	−6.16	4.86
Luxembourg	456	456	0.74	0.58	5.50	−0.67	6.25	−26.69	18.01
Mexico	312	312	0.58	0.38	6.17	−0.57	5.01	−27.16	17.58
Netherlands	1,272	1,242	0.53	0.40	5.14	0.12	13.24	−33.15	50.24
New Zealand	1,488	1,488	0.56	0.50	3.65	−0.11	9.69	−28.76	23.61
Norway	1,271	1,268	0.52	0.39	5.07	−0.32	6.75	−27.49	25.26
Poland	288	288	0.60	0.38	6.69	0.03	6.23	−29.41	32.71
Portugal	1,032	998	0.50	0.13	8.06	1.99	49.31	−89.24	86.10
Singapore	600	600	0.75	0.47	7.40	0.27	9.40	−41.91	46.80
Slovakia	240	240	0.50	0.37	5.33	1.34	10.58	−18.87	33.34
Slovenia	120	120	0.37	0.29	4.03	0.29	4.61	−10.37	16.19
South Korea	288	288	0.73	0.41	8.20	1.01	8.54	−27.50	49.82
Spain	732	732	0.49	0.34	5.48	−0.01	4.90	−25.71	26.52
Sweden	1,320	1,317	0.59	0.47	4.82	−0.19	6.31	−27.01	28.01
Switzerland	1,272	1,248	0.47	0.38	4.40	−0.22	8.94	−27.78	32.66
Turkey	407	407	1.70	0.70	15.11	2.10	15.80	−42.92	122.90
United Kingdom	2,148	2,143	0.46	0.39	3.74	0.60	21.71	−26.87	50.05
United States	1,560	1,556	0.64	0.52	4.98	0.40	12.94	−29.47	42.52
Full sample	31,994	31,465	0.55	0.37	5.86	1.13	43.07	−93.90	128.82

(*continued on next page*)

Table 3 (*continued*)

Country	Sample size: Months	Sample size: Observ	Summary statistics for returns: $\bar{R}_a$ (%)	$\bar{R}_g$ (%)	SD (%)	Skew	Kurt	Min (%)	Max (%)
Panel C: Real USD returns									
Argentina	239	239	0.36	−0.12	8.68	−2.79	19.14	−57.23	19.90
Australia	1,428	1,423	0.69	0.56	5.03	−0.90	11.26	−45.31	21.54
Austria	1,139	1,025	0.47	0.20	6.76	−2.37	34.30	−82.31	38.82
Belgium	1,476	1,411	0.40	0.24	5.81	−0.22	7.09	−36.08	24.67
Canada	1,548	1,542	0.58	0.46	4.80	−0.55	7.34	−27.92	23.71
Chile period I	528	528	0.40	0.05	8.02	−0.21	14.45	−53.17	45.08
Chile period II	120	120	−0.05	−0.25	6.24	0.09	3.55	−19.83	18.57
Czechoslovakia	246	190	0.58	−0.40	9.03	−5.19	75.38	−95.42	54.77
Czech Republic	300	300	1.16	0.70	9.59	−0.09	4.45	−34.48	37.29
Denmark	1,560	1,556	0.44	0.35	4.29	−0.23	9.02	−28.81	25.97
Estonia	120	120	0.80	0.61	6.36	1.59	14.04	−17.22	39.77
Finland	612	612	1.00	0.78	6.70	0.20	5.68	−28.20	33.68
France	1,848	1,831	0.52	0.32	6.47	2.64	70.08	−55.41	118.79
Germany	1,656	1,542	0.69	0.16	8.70	2.53	67.28	−99.36	127.49
Greece	516	515	0.74	0.23	10.48	1.33	9.73	−34.26	66.57
Hungary	288	288	1.18	0.74	9.17	−0.52	5.25	−38.21	26.46
Iceland	216	216	0.60	0.04	8.86	−3.31	27.20	−74.74	23.34
Ireland	1,008	1,008	0.59	0.45	5.20	−0.31	8.23	−32.57	30.37
Israel	120	120	−0.02	−0.16	5.40	−0.24	3.38	−15.01	13.50
Italy	1,068	1,068	0.66	0.23	8.68	−0.31	18.53	−82.63	59.74
Japan	1,080	1,036	0.66	0.30	6.60	−2.69	45.92	−95.04	27.23
Latvia	48	48	1.15	1.06	4.33	1.03	4.36	−5.44	15.69
Lithuania	24	24	−0.04	−0.10	3.26	0.23	3.98	−8.13	8.09
Luxembourg	456	456	0.77	0.57	6.26	−0.55	6.25	−33.33	23.39
Mexico	312	312	0.62	0.27	8.14	−0.88	5.45	−36.49	18.64
Netherlands	1,272	1,242	0.58	0.43	5.49	−0.73	9.14	−42.30	30.79
New Zealand	1,488	1,488	0.59	0.47	4.85	−0.59	11.24	−37.74	27.50
Norway	1,271	1,268	0.57	0.39	5.90	−0.42	6.45	−30.92	24.49
Poland	288	288	0.77	0.38	8.75	−0.17	4.76	−34.63	32.19
Portugal	1,032	998	0.56	0.16	8.30	1.39	40.49	−90.60	83.52
Singapore	600	600	0.83	0.51	7.96	0.37	8.36	−41.44	49.56
Slovakia	240	240	0.85	0.65	6.43	0.65	8.14	−26.33	35.96
Slovenia	120	120	0.12	−0.02	5.22	−0.00	4.55	−15.76	19.50
South Korea	288	288	0.84	0.31	10.52	1.19	10.26	−31.19	70.73
Spain	732	732	0.60	0.40	6.21	−0.08	4.86	−27.54	29.37
Sweden	1,320	1,317	0.60	0.45	5.45	−0.29	6.11	−28.77	25.70
Switzerland	1,272	1,248	0.55	0.44	4.87	−0.14	7.27	−29.09	33.44
Turkey	407	407	1.91	0.71	16.36	1.59	11.51	−42.00	119.46
United Kingdom	2,148	2,143	0.45	0.35	4.51	0.55	18.77	−29.13	55.81
United States	1,560	1,556	0.64	0.52	4.98	0.40	12.94	−29.47	42.52
Full sample	31,994	31,465	0.61	0.37	6.65	0.53	34.31	−99.36	127.49

Table 4
Bootstrap distributions of payoffs

The table summarizes the distribution of real payoffs from a $1.00 buy-and-hold investment across 1,000,000 bootstrap simulations at various return horizons. The underlying sample is the pooled sample of all developed countries. The real payoffs in Panel A are from the perspective of a domestic investor in a representative country, and the real payoffs in Panel B are from the perspective of a global USD investor. The real payoff for bootstrap iteration m at the H-month horizon is $W_H^{(m)}$. For each horizon, the table reports the mean, standard deviation, and distribution percentiles of real payoffs. The last column in the table shows the proportion of payoff draws that are less than one [$\mathbb{P}(W_H^{(m)} < 1)$]. The bootstrap sampling procedure is based on the stationary bootstrap approach of Politis and Romano (1994) as described in the text. We sample blocks of random length, where the length of each block has a geometric distribution with a mean of 120 months.

	Moments		Percentiles									
Horizon	Mean	SD	1%	5%	10%	25%	50%	75%	90%	95%	99%	$\mathbb{P}(W_H^{(m)} < 1)$
Panel A: Real payoffs												
1 month	1.01	0.06	0.85	0.92	0.95	0.98	1.01	1.03	1.06	1.09	1.16	0.432
1 year	1.08	0.28	0.52	0.72	0.80	0.93	1.06	1.19	1.35	1.49	1.90	0.368
5 years	1.45	0.93	0.19	0.54	0.69	0.95	1.29	1.72	2.34	2.86	4.41	0.283
10 years	2.02	1.76	0.13	0.47	0.69	1.08	1.64	2.44	3.60	4.63	8.75	0.215
20 years	3.88	5.40	0.14	0.43	0.73	1.43	2.63	4.57	7.76	10.89	22.97	0.155
30 years	7.38	13.76	0.14	0.47	0.85	1.94	4.16	8.28	15.58	23.30	53.45	0.121
Panel B: Real USD payoffs												
1 month	1.01	0.07	0.83	0.91	0.94	0.98	1.01	1.03	1.07	1.10	1.18	0.435
1 year	1.09	0.32	0.44	0.69	0.79	0.93	1.06	1.21	1.39	1.57	2.10	0.376
5 years	1.52	1.20	0.16	0.50	0.66	0.93	1.28	1.77	2.51	3.25	5.54	0.297
10 years	2.16	2.34	0.08	0.40	0.63	1.05	1.65	2.56	3.96	5.25	10.71	0.229
20 years	4.34	7.96	0.06	0.34	0.64	1.37	2.65	4.90	8.81	12.94	29.38	0.171
30 years	8.66	22.92	0.06	0.34	0.71	1.82	4.21	9.00	18.22	28.51	71.96	0.140

Table 5
Alternative samples

The table summarizes the alternative samples considered in the paper. For each sample, the table provides the sample name, a description of the sample formation criteria, the number of sample months, and the number of sample months as a percentage of the full developed sample. Panel A shows characteristics for the full sample of developed countries. Panel B details samples corresponding to a single country, and Panel C summarizes samples subject to survivor and easy data biases. The samples described in Panels D, E, and F incorporate screens based on data recency, population, and equity market size, respectively, as described in the table.

		Sample size	
Sample	Description	Months	Coverage (%)
	Panel A: Base case		
Full sample	Full sample of all developed countries	31,994	100.0
	Panel B: Single country		
U.S.	United States from 1890 to 2019	1,560	4.9
U.K.	United Kingdom from 1841 to 2019	2,148	6.7
	Panel C: Survival and easy data		
Survival	Condition sample on 2019 membership in the OECD	30,381	95.0
Continuous	Condition sample on 2019 membership in the OECD and uninterrupted monthly data	22,801	71.3
	Panel D: Sample period		
Post-1880	Exclude data prior to 1880	31,358	98.0
Post-1920	Exclude data prior to 1920	27,747	86.7
Post-1960	Exclude data prior to 1960	18,971	59.3
Post-2000	Exclude data prior to 2000	7,968	24.9
	Panel E: Population		
POP 0.2%	Include countries after population reaches 0.2% of world population	23,523	73.5
POP 0.5%	Include countries after population reaches 0.5% of world population	12,796	40.0
	Panel F: Equity market size		
M/GDP 0.5	Include countries after ratio of market capitalization to GDP reaches 0.5	20,640	64.5
M/GDP 1.0	Include countries after ratio of market capitalization to GDP reaches 1.0	12,960	40.5

Table 6
Bootstrap distributions of 30-year payoffs for biased samples

The table summarizes the distribution of payoffs from a \$1.00 buy-and-hold investment across 1,000,000 bootstrap simulations at the 30-year horizon for alternative samples. The real payoffs are from the perspective of a domestic investor in a representative country. The underlying sample in Panel A is the pooled sample of all developed countries. The underlying samples in Panel B are the United States over the period from 1890 to 2019 (U.S.) and the United Kingdom over the period from 1841 to 2019 (U.K.). The underlying samples in Panel C are the sample conditioned on current membership in the OECD (Survival) and the sample conditioned on current membership in the OECD and continuous data (Continuous). The real payoff for bootstrap iteration m at the H-month horizon is $W_H^{(m)}$. For each horizon, the table reports the mean, standard deviation, and distribution percentiles of real payoffs. The last column in the table shows the proportion of payoff draws that are less than one [$\mathbb{P}(W_H^{(m)} < 1)$]. The bootstrap sampling procedure is based on the stationary bootstrap approach of Politis and Romano (1994) as described in the text. We sample blocks of random length, where the length of each block has a geometric distribution with a mean of 120 months.

	Moments		Percentiles									
Sample	Mean	SD	1%	5%	10%	25%	50%	75%	90%	95%	99%	$\mathbb{P}(W_H^{(m)} < 1)$
Panel A: Base case												
Full sample	7.38	13.76	0.14	0.47	0.85	1.94	4.16	8.28	15.58	23.30	53.45	0.121
Panel B: Single country												
U.S.	8.91	8.49	0.96	1.69	2.29	3.76	6.46	11.03	18.08	24.30	41.90	0.012
U.K.	5.28	4.25	0.69	1.24	1.68	2.68	4.21	6.48	9.95	12.97	21.70	0.030
Panel C: Survival and easy data												
Survival	7.68	14.05	0.17	0.53	0.92	2.05	4.33	8.59	16.24	24.41	55.89	0.111
Continuous	9.78	16.71	0.40	0.97	1.49	2.86	5.60	10.87	20.63	30.93	69.01	0.052

Table 7
Bootstrap distributions of 30-year payoffs with additional sample screens

The table summarizes the distribution of real payoffs from a $1.00 buy-and-hold investment across 1,000,000 bootstrap simulations at the 30-year horizon for alternative samples. The real payoffs are from the perspective of a domestic investor in a representative country. The underlying sample in Panel A is the pooled sample of all developed countries. Panel B (Panel C) [Panel D] presents results for the full developed sample with additional sample screens based on sample start date (population) [ratio of market capitalization to GDP]. The underlying samples in Panels B to D are described in Table 5. The real payoff for bootstrap iteration m at the H-month horizon is $W_H^{(m)}$. For each sample, the table reports the mean, standard deviation, and distribution percentiles of real payoffs. The last column in the table shows the proportion of payoff draws that are less than one [$\mathbb{P}(W_H^{(m)} < 1)$]. The bootstrap sampling procedure is based on the stationary bootstrap approach of Politis and Romano (1994) as described in the text. We sample blocks of random length, where the length of each block has a geometric distribution with a mean of 120 months.

	Moments		Percentiles									
Sample	Mean	SD	1%	5%	10%	25%	50%	75%	90%	95%	99%	$\mathbb{P}(W_H^{(m)} < 1)$
Panel A: Base case												
Full sample	7.38	13.76	0.14	0.47	0.85	1.94	4.16	8.28	15.58	23.30	53.45	0.121
Panel B: Sample period												
Post-1880	7.41	15.45	0.14	0.46	0.83	1.91	4.13	8.30	15.71	23.57	54.41	0.124
Post-1920	8.50	20.04	0.15	0.52	0.95	2.16	4.65	9.43	18.08	27.29	63.15	0.106
Post-1960	8.73	17.73	0.23	0.66	1.09	2.25	4.66	9.49	18.46	28.05	66.19	0.089
Post-2000	4.33	7.09	0.14	0.39	0.65	1.29	2.53	4.88	9.09	13.52	30.11	0.180
Panel C: Population												
POP 0.2%	6.83	14.13	0.11	0.39	0.72	1.73	3.82	7.65	14.40	21.63	50.53	0.143
POP 0.5%	6.74	12.03	0.09	0.33	0.66	1.77	3.95	7.74	14.35	21.28	47.51	0.147
Panel D: Equity market size												
M/GDP 0.5	6.72	8.71	0.22	0.68	1.10	2.21	4.37	8.15	14.28	20.20	39.41	0.087
M/GDP 1.0	5.69	6.50	0.24	0.62	0.99	1.98	3.87	7.09	12.00	16.61	30.93	0.102

Table 8
Asset allocation

The table shows results from asset allocation tests. For each return horizon, the table reports the optimal weight in stocks for an investor who relies on the developed country sample to form expectations about stock market performance (w_d) and the optimal weight in stocks for an investor who relies on the U.S. sample to form expectations about stock market performance (w_{us}). The investors have exponential utility with a risk aversion parameter of three. Each investor allocates across the domestic stock market and a risk-free asset, where the risk-free asset is either the inflation-protected risk-free asset or cash. The table also reports the maximum annualized fee that the investor relying on the developed country sample would be willing to pay to use her optimal weight rather than adopt the optimal weight based on the U.S. sample.

	Inflation-protected risk-free asset			Cash		
	Weight in stocks			Weight in stocks		
Horizon	Developed country sample, w_d	United States sample, w_{us}	Fee (%)	Developed country sample, w_d	United States sample, w_{us}	Fee (%)
1 month	0.55	0.89	0.80	0.95	1.19	0.30
1 year	0.43	0.73	0.66	0.82	1.01	0.23
5 years	0.44	0.76	0.67	0.95	1.07	0.10
10 years	0.45	0.78	0.62	0.99	1.19	0.28
20 years	0.44	0.78	0.56	0.99	1.22	0.69
30 years	0.43	0.75	0.45	0.99	1.21	1.02

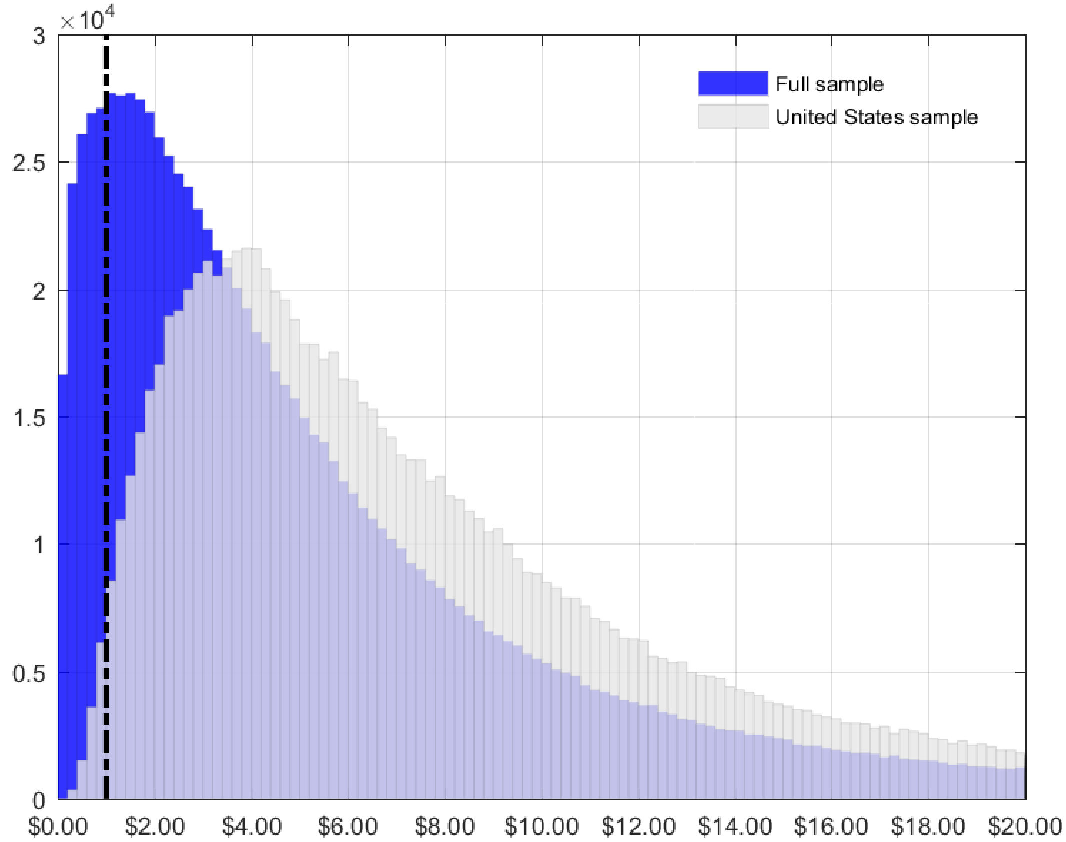

Fig. 1. Cumulative 30-year payoffs. The figure shows histograms of real payoffs across 1,000,000 bootstrap simulations at a return horizon of 30 years. The real payoffs are from the perspective of a domestic investor in a representative country. The underlying sample for the simulated returns is the pooled sample of all developed countries (blue) or the United States sample (gray). The dashed line separates the regions of real loss and gain on a $1.00 initial investment. The bootstrap sampling procedure is based on the stationary bootstrap approach of Politis and Romano (1994) as described in the text. We sample blocks of random length, where the length of each block has a geometric distribution with a mean of 120 months.

Fig. 2. Development periods and data availability by country. The figure details development dates and availability of return data for the developed country sample. The line for each country shows the period over which that country is classified as developed. The development year classifications are based on agricultural labor share, organizational membership in the Organisation for European Economic Co-operation (OEEC) or the Organisation for Economic Co-operation and Development (OECD), or the opening of the first stock exchange in an already developed economy. The dashed portion of each line denotes the period over which return data are missing, and the solid portion denotes the period with valid returns.

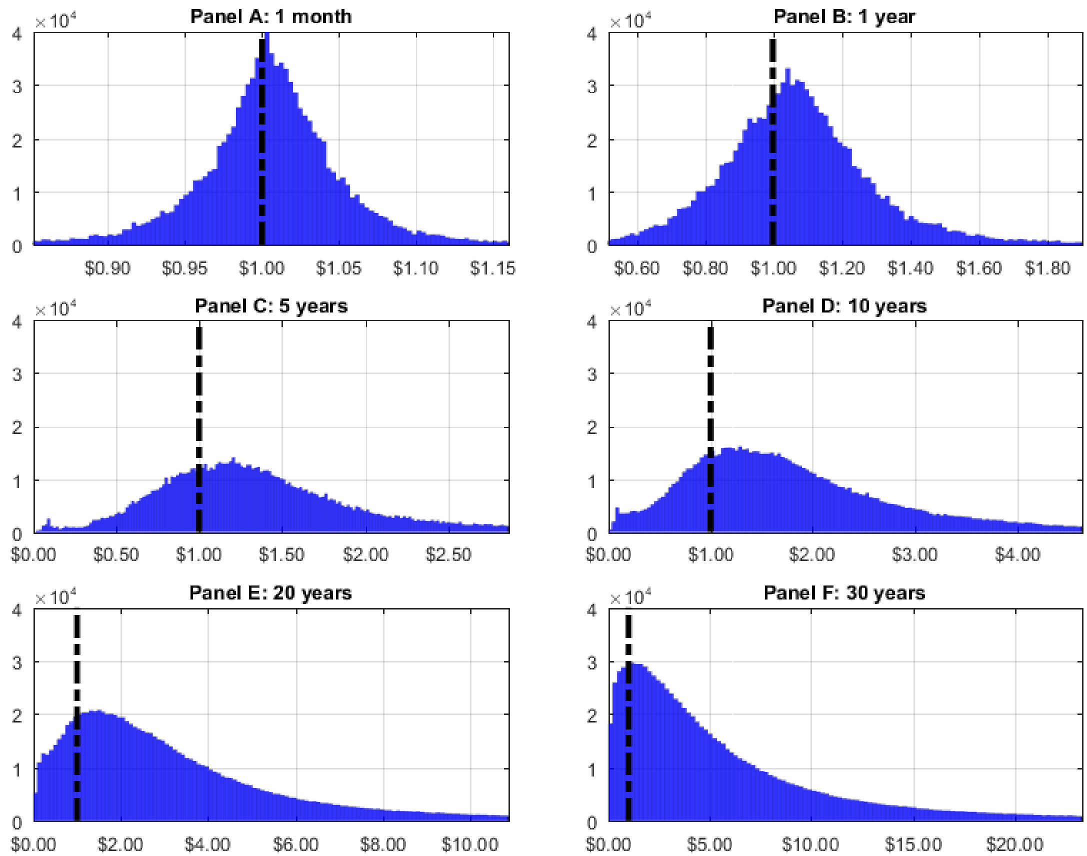

Fig. 3. Cumulative payoffs. The figure shows histograms of real payoffs across 1,000,000 bootstrap simulations at various return horizons. The real payoffs are from the perspective of a domestic investor in a representative country. The underlying sample for the simulated returns is the pooled sample of all developed countries. The dashed line in each plot separates the regions of real loss and gain on a $1.00 initial investment. The bootstrap sampling procedure is based on the stationary bootstrap approach of Politis and Romano (1994) as described in the text. We sample blocks of random length, where the length of each block has a geometric distribution with a mean of 120 months.

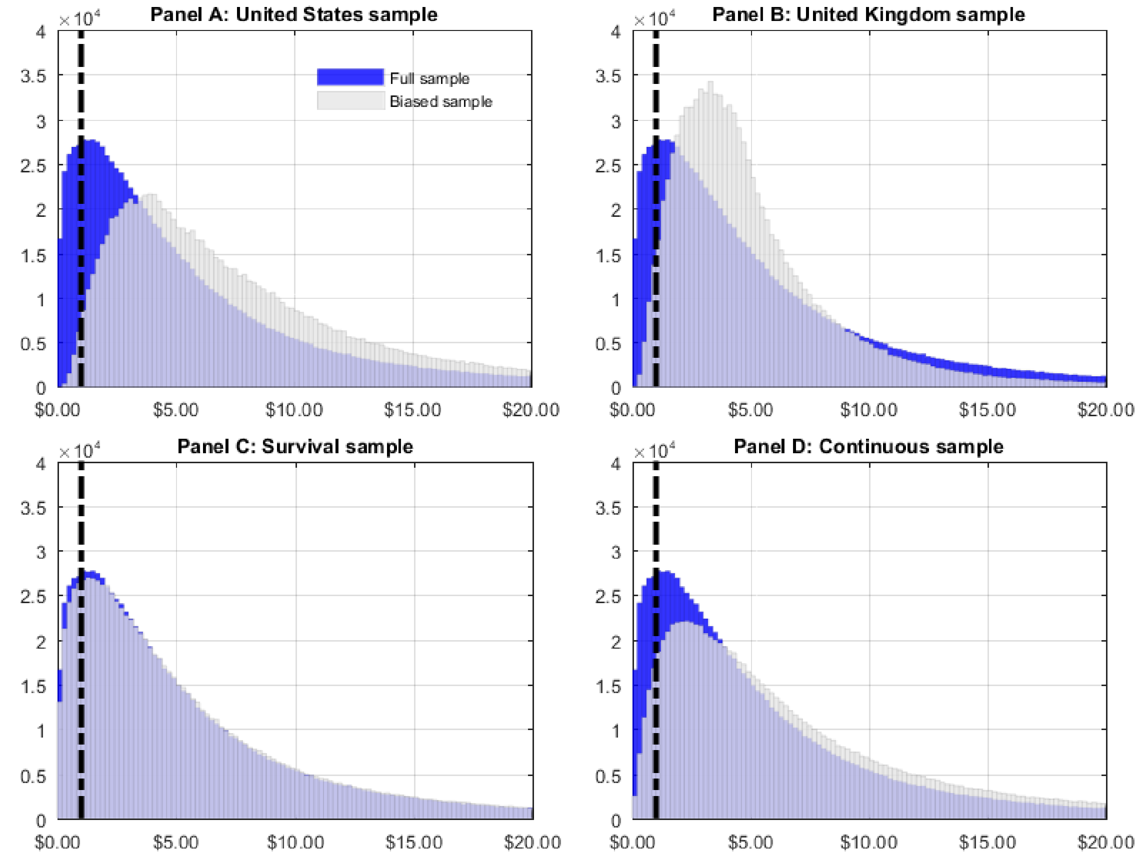

Fig. 4. Cumulative 30-year payoffs for alternative samples. The figure shows histograms of real payoffs across 1,000,000 bootstrap simulations at a return horizon of 30 years. The real payoffs are from the perspective of a domestic investor in a representative country. The blue plot in each panel is the histogram of simulated payoffs based on the pooled sample of all developed countries. The gray plot in Panel A (Panel B) [Panel C] {Panel D} is the histogram of simulated payoffs based on the United States over the period from 1890 to 2019 (the United Kingdom over the period from 1841 to 2019) [the sample conditioned on current membership in the OECD] {the sample conditioned on current membership in the OECD and continuous data}. The dashed line in each plot separates the regions of real loss and gain on a $1.00 initial investment. The bootstrap sampling procedure is based on the stationary bootstrap approach of Politis and Romano (1994) as described in the text. We sample blocks of random length, where the length of each block has a geometric distribution with a mean of 120 months.

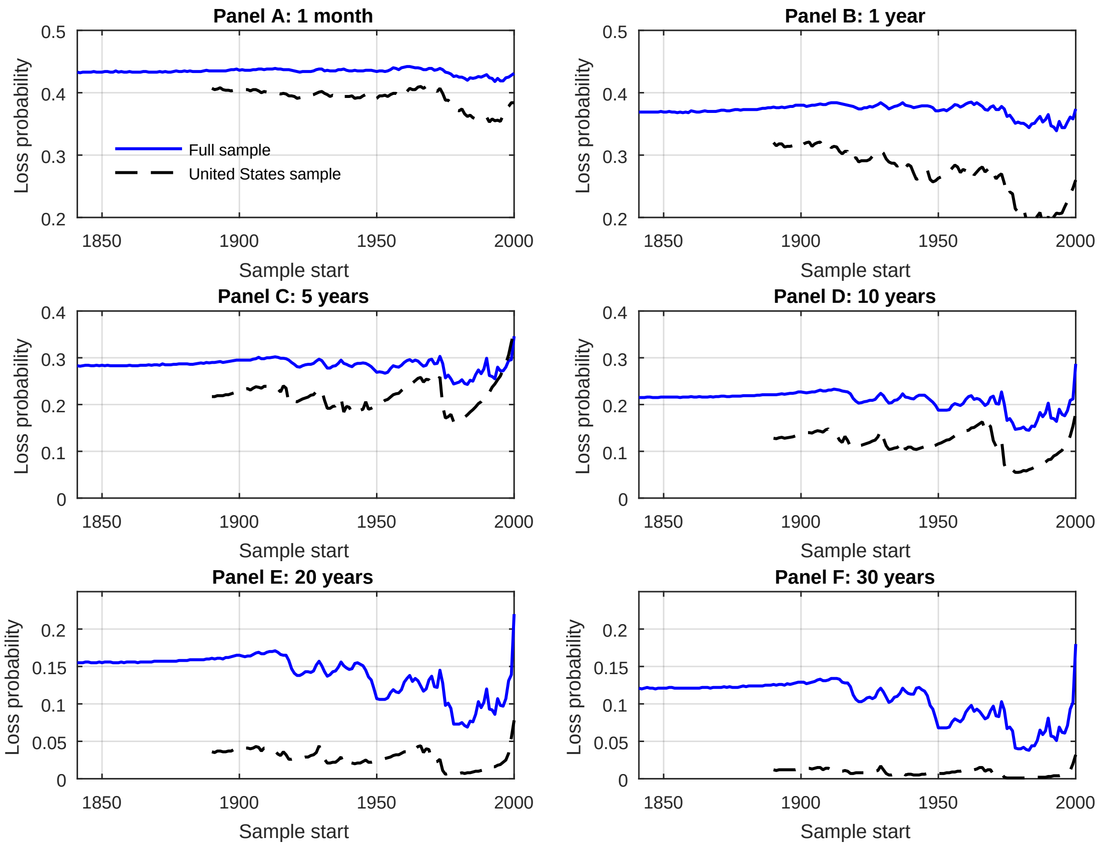

Fig. 5. Loss probabilities for alternative sample start dates. The figure shows the proportion of real payoffs that are less than the initial investment across 1,000,000 bootstrap simulations at various return horizons for alternative sample start dates. The real payoffs are from the perspective of a domestic investor in a representative country. Each panel of the figure corresponds to a specific return horizon. The underlying sample for the simulated returns is the pooled sample of all developed countries (solid line) or the United States sample (dashed line). The bootstrap sampling procedure is based on the stationary bootstrap approach of Politis and Romano (1994) as described in the text. We sample blocks of random length, where the length of each block has a geometric distribution with a mean of 120 months.

Internet Appendix for "Stocks for the long run? Evidence from a broad sample of developed markets"

Aizhan Anarkulova[a], Scott Cederburg[a], Michael S. O'Doherty[b,*]

[a] *Eller College of Management, University of Arizona, McClelland Hall, Room 315R, Tucson, AZ 85721, USA*
[b] *Trulaske College of Business, University of Missouri, 401C Cornell Hall, Columbia, MO 65211, USA*

Abstract

This Internet Appendix provides material that is supplemental to the paper "Stocks for the long run? Evidence from a broad sample of developed markets." Section 1 describes the provenance and attributes of the GFDatabase, provides details on data construction, and presents results from tests to validate our approach to constructing country-level returns. Section 2 examines the robustness of our block bootstrap design parameters. Section 3 presents additional empirical results. Section 4 describes the cumulative wealth calculations used to generate the asset allocation results in the paper and presents additional asset allocation results.

*Corresponding author.
E-mail addresses: aanarkul@email.arizona.edu (Aizhan Anarkulova), cederburg@email.arizona.edu (Scott Cederburg), odohertym@missouri.edu (Michael S. O'Doherty).

Preprint submitted to Journal of Financial Economics *January 18, 2021*

1. Data appendix

Sections 1.1 to 1.4 provide background material on data construction. Section 1.5 discusses empirical tests used to validate our approach to estimating total returns from price indexes and dividend yields for cases in which a total return index is unavailable. Section 1.6 compares our country-level returns data from the GFDatabase with the data used in Jordà et al. (2019).

1.1. GFDatabase from Global Financial Data

The main data source for our study is the GFDatabase from Global Financial Data (GFD). The database provides macroeconomic and financial data with broad coverage of countries and time periods. As shown in Fig. 2 of the paper, our data construction procedure achieves substantial coverage of the periods in which countries in our sample are classified as developed.

The GFDatabase reports stock market data including total return indexes, price indexes, and dividend yields. The computation of these index values by Global Financial Data depends on the country and the time period. For most countries, and especially for more recent time periods, the GFDatabase provides data for indexes that are created and calculated by stock exchanges (e.g., the Tokyo Stock Price Index from the Tokyo Stock Exchange) or well-known index providers (e.g., the S&P 500 Index). In earlier periods for several countries, the GFDatabase includes proprietary total return indexes, price indexes, and dividend yields that are based on data transcribed by Global Financial Data from original historical documents such as newspapers, periodicals, and books. In several cases, GFD has gathered historical data for individual stocks in a country and formed indexes that span early historical periods.

An important issue for ensuring that reported returns reflect stock market performance is the proper handling of capital changes. In private correspondence with Dr. Bryan Taylor and Mike Cerneant from Global Financial Data, we were assured that "GFD handles capital changes in the total return indices that we calculate and that we obtain from other sources. This includes cash dividends, stock dividends and splits, subscription rights, mergers, and exchange offers. We are well aware of how these particular changes impact individual companies and we incorporate these changes into the calculation of returns for individual companies and thus for the index in general."

Return data from the GFDatabase have been used extensively in the finance and economics literature, as the database provides notable coverage of historical periods for a broad set of countries. Recent studies using country-level index data and appearing in premier finance and economics journals include Barro (2006), Berkman et al. (2011), Schularick and Taylor (2012), Colacito and Croce (2013), Nakamura et al. (2013), Rapach et al. (2013), Albuquerque et al. (2015), and Nakamura et al. (2017).

1.2. Data construction

We compute country-level stock market returns inclusive of dividend distributions to ensure that we measure total returns rather than just price returns (i.e., capital gains). For cases in which the GFDatabase contains multiple indexes for a given country, we identify the index with the broadest coverage of firms and with data availability over the longest sample period.

We use total return indexes from GFD for months in which these indexes are available. In these cases, the total nominal return in a given month follows from Eq. (1) in the paper.

If a total return index is not available, we compute total returns from price indexes and dividend yields. Dividend yields in the GFDatabase are recorded either monthly or annually depending on country and period. To estimate country-level dividends, we rely on dividend yields that are reported in December. All annual dividend yield time series from GFD are December observations. For dividend yields available on a monthly basis, we use only the December yields so that cumulative monthly returns accurately reflect the annual dividend yields. The dividend yield data are based on the dividends for the trailing 12 months (i.e., December yields reflect dividends from January to December of the corresponding year). To compute dividends for a given country-year, we multiply the December dividend yield by the price index at the end of December. We then estimate monthly dividends for each of the prior 12 months by dividing the annual dividend amount by 12. This approach relies on the assumption that dividends are issued in 12 equal monthly payments over a given year. The total nominal return in a given month follows from Eq. (3) in the paper.

We use monthly changes in country-level CPIs from GFD to estimate monthly inflation rates. Most of the observations on CPI levels are reported monthly. For cases with missing data (e.g., if the data are reported quarterly or annually over a given period), we estimate the monthly CPI levels through interpolation by assuming a constant monthly inflation rate between reported CPI levels. We compute the gross real return for a given month as the ratio of the gross nominal return and the gross inflation rate following Eq. (5) in the paper.

We use monthly data on local currency exchange rates vis-à-vis the U.S. dollar from GFD to estimate monthly local currency appreciation relative to the dollar. From March 1938 to December 1945, we use German exchange rate data for Austria. The Austrian Schilling was abolished following annexation by Germany, and German Reichsmarks became the local currency during World War II. The GFDatabase reports a constant Austrian Schilling exchange rate vis-à-vis the U.S. dollar over the March 1938 to December 1945 period, so we replace these data to better reflect the currency exchanges that would have been necessary to invest in Austria during this time. We compute the gross real USD return for a given month using the nominal return in local currency, local currency appreciation relative to the U.S. dollar, and U.S. inflation following Eq. (6) in the paper.

Table IA1 reports the total return indexes, the price indexes, and the dividend yields that we use to construct our sample of returns for developed countries. We supplement the GFD data with information from other sources in two cases. First, Luxembourg is missing return data from 2016 to 2019 in the GFDatabase, so we fill in the missing observations using data from the Luxembourg Stock Exchange's official website.[1]

[1] See https://www.bourse.lu/home.

Second, we use annual dividend yields from Jordà et al. (2019) to fill in missing data for Norway from 1914 to 1969 and Portugal from 1934 to 1988. Table IA1 also shows the sample period start and end dates for each country. The data series labels are from GFD, and the subperiods covered by the total return indexes and the subperiods covered by the price indexes and dividend yields combine to form the full sample period for each country (with the exception of Czechoslovakia, which has a multi-month terminal return that extends beyond the GFD sample period as discussed in Section 2.3 of the paper).

1.3. Data adjustments

The following subsections outline additional adjustments required to compute nominal returns and real returns for our developed country sample.

1.3.1. Missing data

For periods in which dividend yield data are missing in the GFDatabase, we estimate yields by examining the reported dividend yields from before and after the data gaps. We apply this correction for missing data in three cases. First, the GFDatabase does not have a dividend yield for Austria from June 1939 to June 1969. The dividend yield for May 1939 is 3.7%, and the yield for July 1969 is 3.5%. We therefore use the average dividend yield of 3.6% to fill in the missing dividend yield data from June 1939 to June 1969. For comparison, the average dividend yield for Austria is 4.3% in the 1930s before the gap and 3.4% in the 1970s after the gap. Second, Chile is missing dividend yield data from 1967 to 1970, which we fill in with a 7.0% yield based on the dividend yield observation in December 1966. Third, Czechoslovakia is missing dividend yield data from April 1938 to March 1943. The dividend yield in Czechoslovakia fluctuates between 1.4% and 2.6% in the three years before the break in the data, so we assume a 2.0% dividend yield for the missing observations.

In a few cases, we use a smoothing procedure to fill gaps in return series. Argentina is missing return data for April, May, November, and December in 1955. We estimate returns for April 1955 and May 1955 using price index data for March 1955 and June 1955 under the assumption of a constant return for April, May, and June of 1955. We make an analogous calculation to fill in the missing data for November 1955 and December 1955. We have either semiannual or annual return data for France from 1915 to 1918 and either quarterly or semiannual data for Switzerland from 1914 to 1920. We treat the intermediate months as missing and smooth returns across the months in each quarterly, semiannual, or annual period.

The price index in Norway is missing for May 1940 and June 1940. We use price index values of 5.779 for May 1940 and 5.930 for June 1940 from Klovland (2004). Finally, as discussed in Section 2.3 of the paper, we estimate a −90% nominal return for the 39-month period from April 1943 to June 1946 in Czechoslovakia.

1.3.2. Hyperinflation in Germany

We use a total return index for Germany from 1917 to 1923 that is denominated in gold marks rather than paper marks. Gold marks were backed by gold, whereas the value of German paper marks effectively went to zero over this period. The CPI data from GFD correspond to the paper marks series. With gold as a base for the index, however, there is presumably negligible inflation. We use the total return series in gold marks for both nominal and real returns. One of the multi-month returns for Germany in Table 2 of the paper spans a 42-month period from August 1914 to January 1918. As such, this multi-month period partially overlaps with the period in which we rely on the gold marks series. We compute inflation for this multi-month return observation using CPI levels from July 1914 and December 1916 and assume a zero inflation rate from January 1917 to January 1918 (i.e., the portion of the multi-month period in which we use returns denominated in gold marks). We also assume zero appreciation of gold marks relative to the U.S. dollar from 1917 to 1923 because both currencies were backed by gold during this period.

The reported data for Germany's total return index indicate a −62.2% return in January 1923 followed by a 333.4% return in February 1923 for a two-month return of 63.7%. We examine alternative sources of information and find that Bittlingmayer (1998) does not show these extreme returns. We treat this sequence of returns as a data error and smooth the two-month return across January 1923 and February 1923.

1.4. Multi-month return periods

Several historical events disrupted stock exchange operations during our sample period. To best reflect the experience of stock market participants during these episodes, we compute multi-month return observations that span the periods of disruption. This approach assumes that an investor could not liquidate her holdings in a given stock market during these times because of market closure or severe restrictions on trading. Table 2 in the paper shows the multi-month return periods in our sample and notes the underlying events.

The events surrounding World War I and World War II impacted stock exchange operations in a number of the countries in our sample. Most of the corresponding multi-month returns listed in Panels A and B of Table 2 cover full periods in which the stock exchange was officially closed. We note four exceptions below. First, although the Swedish exchange was closed from August 1914 to September 1914, the price index values from GFD for July 1914 and October 1914 are identical. Thus, we calculate a four-month return covering August 1914 to November 1914. Second, price controls in Germany started in January 1943 and extended until July 1948, leading us to construct a 67-month return covering this period. Third, Japanese stock market trading stopped in August 1945, and although over-the-counter trading started in May 1946, the stock exchange did not reopen until May 1949. We therefore compute a 45-month return covering September 1945 to May 1949. Fourth, Czechoslovakia's stock exchange was intermittently open during the German occupation. Even when the market was open, it was subject to price controls and limited trading. For this reason, we calculate three multi-month returns associated with the German occupation period in Czechoslovakia: a 16-month return from October 1938 to January 1940, a four-month return from January 1942 to April 1942, and a 39-month return from April 1943 to June 1946.

The other events in our sample that disrupted stock market operations include financial crises, labor strikes, and political revolutions. Both Austria and Germany experienced financial crises in 1931. The Austrian stock exchange closed, and we compute a two-month return from October 1931 to November 1931.

Germany's stock exchange closed in July 1931, reopened for just over two weeks in September 1931, and subsequently closed again until April 1932. We compute a two-month return from August 1931 to September 1931 and a seven-month return covering October 1931 to April 1932. Greece's stock market was closed for five weeks starting from June 29, 2015 because of the country's financial crisis, so we compute a two-month return from July 2015 to August 2015. The exchange in France was closed because of labor strikes in April 1974 and March 1979. In Portugal, trading stopped on April 25, 1974 with the start of a military coup, and stock trading resumed on March 7, 1977. We therefore compute a 35-month return for Portugal that covers May 1974 to March 1977.

1.5. Internal validation of total returns data

If total return indexes are missing in the GFDatabase, we estimate total returns using price indexes and dividend yields. To validate this approach, we compute total returns based on total return indexes and compare them with total returns estimated from price indexes and dividend yields. We make these comparisons over the periods for which total return indexes, price indexes, and dividend yields are all available. Table IA2 lists the total return indexes, price indexes, and dividend yields that we use for our validation tests as well as the testing period for each country.[2]

Table IA3 reports our test results. The table shows correlations of returns calculated using total return indexes and returns computed from price indexes and dividend yields for each country and for the pooled sample. The correlation for the pooled sample is 0.98, and Israel's correlation of 0.86 represents the minimum value across individual countries. In addition, the table compares average monthly return and standard deviation across the two approaches for computing returns. None of the differences in mean return is statistically significant at the 10% level. For the comparisons of standard deviations, only the differences in return volatilities for the Czech Republic and Israel are statistically significant at the 10% level. The difference in mean return for the pooled sample is only one basis point per month, and the difference in standard deviation of returns is only four basis points. For both comparisons in the pooled sample (i.e., means and volatilities), we fail to reject the null hypothesis of equal performance.

1.6. External validation of total returns data

We perform an additional analysis that compares the return data in our sample with another data source with broad coverage of countries and periods. Specifically, we examine returns from the overlapping periods in our sample and the sample of Jordà et al. (2019).[3] Jordà et al. (2019) use a variety of sources for historical return data across countries. Many of these data sources are different from those used by Global Financial Data in the construction of their database. For example, post-war data for Japan in the GFDatabase are

[2]Because there are no comparable indexes for Germany from 1917 to 1923, we exclude the German returns over this period from the validation tests.

[3]The data from Jordà et al. (2019) are available at http://www.macrohistory.net/data/. We thank the authors for making these data available.

available for the Nikkei 225 Index and the Tokyo Stock Price Index, whereas Jordà et al. (2019) use annual data from the Statistical Yearbooks published by the Statistics Bureau of Japan. As such, the comparison in this section provides external validation of the historical return data in both databases.

The return data from Jordà et al. (2019) are annual, so we annualize the returns in our sample to facilitate comparison. We compound returns within each year, and we calculate real returns by compounding returns and CPI separately and then adjusting the annual returns for inflation. If a year ends during a multi-month period (see Section 1.4), we form a multi-year observation in the annual return dataset. During these periods, we also compound the annual returns from Jordà et al. (2019) to calculate a comparable multi-year observation. We use annualized CPI inflation to compute real returns based on the nominal returns from Jordà et al. (2019).

Our sample overlaps with the sample from Jordà et al. (2019) for 16 countries. Jordà et al.'s (2019) data span 1870 to 2015, so the sample start date for each country is the later of 1870 and the start of our developed sample period. As previously noted, we use a return index in gold marks in Germany for the 1917 to 1923 period. Jordà et al. (2019) report returns in paper marks, so we omit the 1917 to 1923 period in Germany from this analysis.

Table IA4 shows summary statistics for the two historical return samples for each country and for the pooled sample. The table reports the number of years and the number of observations (i.e., including multi-year observations) of overlap across the samples. For each sample, we report the arithmetic and geometric means, the standard deviation, and the minimum and maximum return. We also compute the correlation between the GFDatabase returns and the Jordà et al. (2019) returns. Panel A reports statistics for nominal returns, and Panel B provides statistics for real returns.

The summary statistics in Table IA4 show that the return data from the two databases have very similar characteristics. Return correlation exceeds 0.90 for nearly all countries and for the pooled sample. The pooled sample means and standard deviations closely match across the databases, and the two databases agree cross-sectionally about which countries have had relatively high or low average returns and standard deviations during the shared sample periods. The two databases also report similar extreme returns. In particular, the minimum returns reported for each country are very close in magnitude, which is important given our focus on characterizing the left tails of the cumulative return distributions. Based on this analysis, we conclude that the quality of return data from both the GFDatabase and the Jordà et al. (2019) database is supported by the external validation exercise.

2. Bootstrap appendix

Section 2.1 discusses the choice of the mean bootstrap block length parameter for the base case. Section 2.2 evaluates the robustness of our main results to changes in the block length parameter.

2.1. *Bootstrap design*

The purpose of the block resampling scheme is to reflect the time-series properties of weakly dependent country-level return observations. Serial dependencies in returns that are included in the same return block will be reflected by the block bootstrap draws. Based on prior literature, we anticipate that the data are characterized by multiple types of serial dependencies that could be important to long-term investors. In Section 3 of the paper, we specifically note the potential effects of short-term positive autocorrelation, persistence in volatility, and long-term mean reversion on buy-and-hold returns. Additional forms of dependence may exist in the data, and we do not need to take a stance on these patterns as long as they are often included in the same return blocks in the bootstrap.

The optimal block length parameter is not ex ante obvious. The return blocks should be long enough to allow for longer-term effects like mean reversion to be reflected in the cumulative return distribution but short enough to allow the bootstrap to generate simulated return sequences that are not directly observed in the historical data. We ultimately select 120 months as our base case block length parameter. As described further below, this block length allows for the effects of mean reversion and produces return distributions that are similar to those from other specifications with relatively long block lengths. We also show the robustness of our findings to alternative block lengths in Section 2.2.

To initially examine the effect of the block length parameter, we consider the moments of the distribution of cumulative log returns. Anticipating that the standard deviation of cumulative log returns could decrease with longer block lengths as long blocks of return data can reflect more mean reversion, we examine whether and how the distribution moments stabilize across longer block lengths. Fig. IA1 plots the standard deviations of cumulative log return distributions. The six panels correspond to the six return horizons considered in the paper, and we study standard deviation as a function of the bootstrap block length parameter. The solid blue line corresponds to the developed country sample, and the dashed black line corresponds to the U.S. sample. We consider block length parameters varying from one month (which is equivalent to i.i.d. resampling) to 240 months.

Fig. IA1 reveals that the results exhibit some sensitivity to the block length parameter. The standard deviation of each log return distribution initially increases as the block length parameter increases from one to around ten to 20 months. The increase in risk likely reflects positive short-term autocorrelation and persistence in volatility. Especially with the longer horizons of 20 or 30 years, the standard deviation decreases as the block length parameter further increases. This pattern likely reflects the role of mean reversion, which is a longer-term effect that can be picked up by larger return blocks. The U.S. sample displays a greater reduction in risk as the block length increases compared with the developed sample. For example, at the 30-year horizon for the developed sample, the standard deviation with a block size parameter of 240 months is 1.12, which is comparable in magnitude with the one-month block value of 1.21. In the U.S. sample, the 240-month block length parameter produces a standard deviation of only 0.71 relative to the one-month block standard deviation of 0.94. In an unreported analysis, we find that variance ratios calculated following Poterba and Summers (1988) tend to be low in the U.S. relative to many of the other developed countries, which is consistent with the longer block lengths having a larger effect in the U.S. sample owing to stronger mean reversion in this sample. For our base case of developed countries, the standard deviation at each horizon stabilizes with longer block lengths. Similar patterns exist for skewness and kurtosis of the cumulative log return distributions. As such, we choose a base case block length parameter of 120 months, which allows the bootstrap to reflect mean reversion while still generating sufficient resampling.

2.2. Impact of mean block length

To construct bootstrap distributions of payoffs and continuously compounded returns, we resample with replacement from the sample of returns in developed markets using a stationary block bootstrap approach. Our base case design draws blocks of consecutive returns, where the length of each block has a geometric distribution with a mean of 120 months. A block resampling procedure is appropriate for stationary weakly dependent time series (Politis and Romano, 1994b). There are several reasons to believe that the country-level time series of real returns in our sample are weakly dependent. Notably, the empirical evidence of time-varying volatility in stock market returns is overwhelming (Bollerslev et al., 1994), and several studies document evidence of mean reversion in long-horizon returns (e.g., Poterba and Summers, 1988; Siegel, 2014). The mean block length for the base case design of 120 months accounts for these features of the data, but we also consider the sensitivity of the results to other choices for the block length.

Table IA5 summarizes the distribution of real payoffs at various investment horizons for alternative choices of mean block length in the bootstrap procedure. The panels of the table are organized by horizon, and each panel considers i.i.d. resampling and mean block lengths of 12, 120 (base case), and 240 months. The i.i.d. design in the top row of each panel corresponds to the approach used in Fama and French (2018a,b). The block length parameter does not have a major impact on the results, particularly at short investment horizons. There are some noticeable differences at longer horizons. For the 30-year distributions in Panel F, the mean and standard deviation of real payoff are highest for the bootstrap design with a 12-month mean block length. This design is also associated with the most extreme tail outcomes and the largest loss probability. These features of the bootstrap distribution likely reflect that a 12-month block length incorporates effects from time-varying volatility and short-term autocorrelation but not from long-term mean reversion. In Panel F, the payoff moments, payoff percentiles, and loss probability for our 120-month block base case design are similar to the corresponding values for the i.i.d. resampling design. The decrease in payoff volatility for the 120-month block length relative to the 12-month block length is consistent with longer blocks reflecting mean reversion in the data.

Fig. IA2 shows loss probability as a function of mean block length for each investment horizon. For horizons of five years and beyond, the loss probability tends to spike around a block length of 12 months. The general conclusion, however, is that the block length parameter has a minor impact on the results. The 30-year loss probability for the developed sample ranges from 11.1% to 15.3% across block lengths compared with only 0.6% to 3.8% for the U.S. sample.

3. Additional results appendix

This appendix presents supplementary empirical results. Section 3.1 considers the bootstrap distribution of nominal payoffs. Section 3.2 characterizes the bootstrap distribution of continuously compounded returns at various investment horizons. Section 3.3 presents results for the alternative samples considered in Section 4.2 of the paper. Section 3.4 reproduces the analyses in Section 4.2 of the paper using real USD returns rather than real local currency returns.

3.1. Nominal payoffs

As discussed in Section 2.4 of the paper, our primary analysis focuses on real returns and payoffs because periods of hyperinflation can lead to nominal performance that poorly reflects the true economic experience of investors. For completeness, Table IA6 reports statistics for the bootstrap distributions of nominal payoffs.

3.2. Continuously compounded returns

This section evaluates properties of the bootstrap distributions of continuously compounded returns for the full sample of developed countries and for the U.S. sample. This analysis is motivated by Fama and French's (2018a) finding that the distribution of continuously compounded returns from the U.S. is close to normal for horizons beyond ten years, such that the wealth distribution is approximately lognormal. We are specifically interested in whether or not this result extends to our developed country sample.

Our base case bootstrap simulation procedure described in Section 3 of the paper uses a stationary block bootstrap approach to draw continuously compounded returns according to Eq. (8) in the paper. If the bootstrap procedure were to draw individual single-month observations with replacement, then the continuously compounded returns would be sums of independent and identically distributed (i.i.d.) log returns. In this setting, the central limit theorem states that the distribution of continuously compounded returns converges toward the normal distribution as the return horizon increases.

Our stationary bootstrap approach differs from monthly i.i.d. resampling in two dimensions. First, we resample return observations in blocks of random length rather than individual observations. Second, our bootstrap design incorporates the multi-month observations in Table 2 of the paper. These features of our procedure take us outside the scope of the classical central limit theorem. Politis and Romano (1994a), however, show that a stationary resampling scheme applied to weakly dependent random variables leads to distributions that converge to normal distributions asymptotically. Thus, it seems reasonable to expect our bootstrap distributions of continuously compounded returns from developed markets to converge to normal with horizon.

Table IA7 allows us to assess the rate of this convergence. Panel A of this table reports moments of continuously compounded real returns based on the full sample across the 1,000,000 bootstrap simulations at each horizon. At the one-month horizon, returns are highly non-normal, as indicated by the skewness and kurtosis statistics of -8.14 and 315.99, respectively. Skewness declines in magnitude as the investment horizon increases, but each distribution exhibits considerable negative skewness. For example, skewness is −1.90 at one year, −0.74 at ten years, and −0.44 at 30 years. Similarly, kurtosis decreases with horizon, but even at 30 years the kurtosis statistic of 4.02 well exceeds the value of 3.00 for normally distributed log returns.

Fig. IA3 plots the kernel smoothed density of simulated continuously compounded returns at each horizon. Each panel of the figure also shows the normal density with mean and variance equal to those of the simulated returns. The evidence in Panel A of Table IA7 and Fig. IA3 suggests that the return distribution converges to normal as the horizon increases, but the convergence remains far from complete. The distribution of continuously compounded real returns displays negative skewness and excess kurtosis at the horizons relevant to the majority of investors. Further, the pronounced left tail reflects downside risk that is particularly harmful for long-term investors.

For comparison, we also examine the rate at which the distribution of continuously compounded returns from the U.S. bootstrap simulations converges to normal. In Panel B of Table IA7, one-month log returns are negatively skewed and exhibit excess kurtosis. The magnitude of skewness for the U.S. sample declines toward zero as the holding period increases. The 30-year continuously compounded returns remain negatively skewed, but the skewness estimate of −0.02 is small in magnitude. The kurtosis of 30-year returns for the U.S. sample is 3.04. This estimate is smaller than the kurtosis value of 4.02 for the full developed sample and is also close to the value of 3.00 for normally distributed log returns.

Fig. IA4 presents kernel smoothed densities for the simulated log returns based on the U.S. sample. At 20-year and 30-year horizons, the distributions of investment outcomes appear mostly indistinguishable from their corresponding normal distributions. These results are consistent with those from Fama and French (2018a), who use a post-1963 U.S. sample to demonstrate that bootstrap distributions of continuously compounded nominal returns converge to normal at long horizons.

3.3. Alternative samples

In Section 4.2 of the paper, we compare the bootstrap distribution of real payoffs based on the full sample of developed countries with the corresponding bootstrap distributions from alternative samples. This section provides supplementary information and results for the analysis in the paper.

Table IA8 reports the sample start date for each developed country in the base sample and in the alternative samples with screens for population size and the ratio of aggregate market capitalization to GDP.

Tables 6 and 7 in the paper summarize the distributions of real payoffs at a 30-year investment horizon for alternative underlying samples. For completeness, Table IA9 presents results for the alternative samples corresponding to the six return horizons considered in the paper.

3.4. Real USD payoffs

Fig. 1 in the paper shows histograms of 30-year real payoffs based on the pooled sample of developed countries and on the U.S. sample. These payoffs are from the perspective of a domestic investor in a representative country and are measured in local currency. Fig. IA5 summarizes the distributions of 30-year real USD payoffs from the perspective of a global USD investor.

In Section 4.2 of the paper, we characterize the distributions of real payoffs at a 30-year investment horizon for alternative underlying samples. The empirical results are presented in Tables 6 and 7. Tables IA10 and IA11 report corresponding results based on real USD payoffs.

4. Portfolio choice appendix

In Section 5 of the paper, we examine asset allocation with a two-asset design. The investors use the bootstrap distribution of stock market payoffs to form expectations about stock market performance and allocate their portfolios to a single risky asset (i.e., a country-level equity index) and a risk-free asset. Section 5 of the paper focuses on investors making single lump sum contributions, and this appendix extends these results to investors making contributions as monthly annuities. Section 4.1 provides details on the calculations of cumulative wealth for investors who follow the lump sum and annuity contribution patterns. Section 4.2 presents the empirical results for the annuity investors.

4.1. Cumulative wealth calculations

We begin with the lump sum investors. In the cases with an inflation-protected risk-free asset, an investment in the risk-free asset maintains its value in each period. Given that the portfolio is rebalanced each month to maintain a weight in stocks of w, the cumulative wealth in the portfolio at an H-month horizon in bootstrap draw m is given by

$$W_H^{(m)}(w) = \prod_{s=1}^{H^*} (wR_s^{(m)} + (1-w)), \tag{IA1}$$

where H^* is the smallest number such that $\sum_{s=1}^{H^*} N_s^{(m)} \geq H$. To calculate cumulative wealth in the cases with cash, we need to account for the effects of inflation on the real value of cash. In the bootstrap method described in Section 3 of the paper, we store gross realized inflation $\Pi_{i,t}$ alongside each gross real return observation $R_{i,t}$ while forming draws. Each iteration m of the bootstrap thus produces a bootstrap return vector draw $R^{(m)} = \{R_1^{(m)}, R_2^{(m)}, \ldots, R_{360}^{(m)}\}$ and a bootstrap inflation vector draw $\Pi^{(m)} = \{\Pi_1^{(m)}, \Pi_2^{(m)}, \ldots, \Pi_{360}^{(m)}\}$. If $R_s^{(m)}$ is one of the multi-month return observations listed in Table 2 of the paper, then $\Pi_s^{(m)}$ is a multi-month cumulative inflation observation. Cumulative wealth with the cash alternative is

$$W_H^{(m)}(w) = \prod_{s=1}^{H^*} \left(wR_s^{(m)} + \frac{1-w}{\Pi_s^{(m)}} \right) \tag{IA2}$$

for all draws in which $\Pi_{H^*}^{(m)}$ (i.e., the last realized inflation observation to be included in the bootstrap draw) is a single-month inflation observation. If $\Pi_{H^*}^{(m)}$ is a multi-month observation and $\sum_{s=1}^{H^*} N_s^{(m)} > H$, then Eq. (IA2) would cause the cash balance to continue to be impacted by inflation even after the horizon H has been reached. We require the investor to remain invested in stocks during these periods to mirror an investment in a closed stock market, but we allow the investor to withdraw the cash balance upon reaching the horizon H without incurring any additional inflation. We therefore recalculate the last inflation observation $\Pi_{H^*}^{(m)}$ in these cases such that the new quantity only reflects inflation from the beginning of the multi-month period through the horizon date H. We then calculate cumulative wealth according to Eq. (IA2).

We proceed to the annuity investors. These investors contribute \$1/H at the beginning of each month of the H-month holding period, and they rebalance the portfolio each period to maintain a weight in stocks of w. Cumulative wealth with access to an inflation-protected risk-free asset is recursively calculated for $s = 1, 2, \ldots, H^*$,

$$W_s^{(m)}(w) = \left(W_{s-1}^{(m)} + \frac{1}{H}\right)\left(wR_s^{(m)} + (1-w)\right) + \frac{N_s^{(m)} - 1}{H}, \tag{IA3}$$

where $W_0^{(m)} = 0$ and $W_H^{(m)} = W_{H^*}^{(m)}$. The first term reflects that wealth from the end of last period and this month's annuity contribution are invested with weights w in the stock market and $1 - w$ in the inflation-protected risk-free asset. If $N_s^{(m)} = 1$ such that $R_s^{(m)}$ is a single-month observation, the second term in Eq. (IA3) is equal to zero and the first term fully reflects the evolution of wealth over the month. When $N_s^{(m)} > 1$, the stock market is closed in the intermediate months of the multi-month period. The investor continues to save \$1/H per month in this case, but this entire new investment is placed in the inflation-protected risk-free asset because the stock market is closed. A total of $N_s^{(m)} - 1$ additional contributions take place after the multi-month period has started, which accounts for the second term in Eq. (IA3). After the multi-month period ends, the wealth balance is rebalanced to invest in the stock market with weight w and the inflation-protected risk-free asset with weight $1 - w$. An exception occurs if $\sum_{s=1}^{H^*} N_s^{(m)} > H$ because $N_s^{(m)} - 1$ additional contributions would cause the investor to contribute more than \$1.00 in aggregate during the holding period. In these cases, we replace $N_s^{(m)} - 1$ with $H - \sum_{s=1}^{H^*-1} N_s^{(m)}$ in Eq. (IA3) such that the investor stops new contributions upon reaching the horizon H. For annuity investors with a cash alternative, the wealth calculation is similar but accounts for the effect of inflation on the real value of cash. The wealth calculation is recursive for $s = 1, 2, \ldots, H^*$,

$$W_s^{(m)}(w) = \left(W_{s-1}^{(m)} + \frac{1}{H}\right)\left(wR_s^{(m)} + \frac{1-w}{\Pi_s^{(m)}}\right) + K_s^{(m)}, \tag{IA4}$$

where $K_s^{(m)}$ is an additional cash term that we calculate if $N_s^{(m)} > 1$ (and $K_s^{(m)} = 0$ if $N_s^{(m)} = 1$). The calculation of $K_s^{(m)}$ takes into account that contributions are made each month and uses the monthly inflation observations within the multi-month period to calculate the cumulative real value of any cash that is contributed during the multi-month period. Exceptions occur if $\sum_{s=1}^{H^*} N_s^{(m)} > H$, and we modify both

$\Pi_s^{(m)}$ and $K_s^{(m)}$ in these cases to ensure that contributions stop at the horizon H and that the cash balance is not subject to additional inflation after the horizon H.

In Section 5 of the paper and in Section 4.2 below, we report the maximum annualized fee that the developed country investor would be willing to pay to maintain her optimal weight in stocks rather than adopt the optimal U.S. investor weight. These fees are applied to the entire wealth balance in each period, such that each period in the wealth calculations from Eqs. (IA1) to (IA4) is further multiplied by $(1 - f/12)^{N_s^{(m)}}$. To find the value of f, we first calculate utility using the optimal U.S. weight and no fee for the investor who evaluates expected utility using the developed country sample. We then calculate utility for the developed country investor under the optimal developed country weight and a range of values of f. We report the value of f that equalizes utility across these two cases.

4.2. Annuity investors

The asset allocation results in Table 8 of the paper correspond to investors who contribute \$1.00 as a lump sum at the beginning of the holding period. We find that investors using the developed country sample to form expectations about stock market outcomes invest less in stocks relative to investors relying on the U.S. sample. We also find that the differences in optimal weights are economically significant. Table 8 specifically shows that the maximum annualized fees on wealth that developed country investors would be willing to pay to adopt their own optimal policies rather than those of the U.S. investors are economically large.

The results in Table IA12 indicate that our findings in Table 8 of the paper are robust to an annuity investment pattern rather than a lump sum approach. An annuity investor contributes a total of \$1.00 throughout the holding period as a monthly annuity. That is, an investor with an H-month horizon contributes \$1/H each month. As with the lump sum case, the annuity investor rebalances monthly to ensure that the weight in stocks remains at w. Outcomes from annuity investments in the stock market are generally less volatile compared with lump sum investments as much of the invested capital has less time to grow and investors can more easily recover from crashes that occur early in the holding period. The optimal weights in stocks are higher for each investor type under the annuity plan compared with the lump sum plan, but developed country investors remain less aggressive relative to U.S. investors and the associated fees remain high.

References

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

Barro, R.J., 2006. Rare disasters and asset markets in the twentieth century. Quarterly Journal of Economics 121, 823–866.

Berkman, H., Jacobsen, B., Lee, J.B., 2011. Time-varying rare disaster risk and stock returns. Journal of Financial Economics 101, 313–332.

Bittlingmayer, G., 1998. Output, stock volatility, and political uncertainty in a natural experiment: Germany, 1880-1940. Journal of Finance 53, 2243–2257.

Bollerslev, T., Engle, R.F., Nelson, D.B., 1994. ARCH models, in: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics. North-Holland. volume 4, pp. 2959–3038.

Colacito, R., Croce, M.M., 2013. International asset pricing with recursive preferences. Journal of Finance 68, 2651–2686.

Fama, E.F., French, K.R., 2018a. Long-horizon returns. Review of Asset Pricing Studies 8, 232–252.

Fama, E.F., French, K.R., 2018b. Volatility lessons. Financial Analysts Journal 74, 42–53.

Jordà, Ò., Knoll, K., Kuvshinov, D., Schularick, M., Taylor, A.M., 2019. The rate of return on everything, 1870–2015. Quarterly Journal of Economics 134, 1225–1298.

Klovland, J., 2004. Historical stock price indices in Norway 1914–2003, in: Eitrheim, Ø., Klovland, J.T., Qvigstad, J.F. (Eds.), Historical Monetary Statistics for Norway 1819–2003. Norges Bank Occasional Papers, Oslo, pp. 329–347.

Nakamura, E., Sergeyev, D., Steinsson, J., 2017. Growth-rate and uncertainty shocks in consumption: Cross-country evidence. American Economic Journal: Macroeconomics 9, 1–39.

Nakamura, E., Steinsson, J., Barro, R., Ursúa, J., 2013. Crises and recoveries in an empirical model of consumption disasters. American Economic Journal: Macroeconomics 5, 35–74.

Politis, D.N., Romano, J.P., 1994a. Limit theorems for weakly dependent Hilbert space valued random variables with application to the stationary bootstrap. Statistica Sinica 4, 461–476.

Politis, D.N., Romano, J.P., 1994b. The stationary bootstrap. Journal of the American Statistical Association 89, 1303–1313.

Poterba, J.M., Summers, L.H., 1988. Mean reversion in stock prices: Evidence and implications. Journal of Financial Economics 22, 27–59.

Rapach, D.E., Strauss, J.K., Zhou, G., 2013. International stock return predictability: What is the role of the United States? Journal of Finance 68, 1633–1662.

Schularick, M., Taylor, A.M., 2012. Credit booms gone bust: Monetary policy, leverage cycles, and financial crises, 1870–2008. American Economic Review 102, 1029–1061.

Siegel, J., 2014. Stocks for the Long Run. McGraw-Hill, New York, NY.

Table IA1
Data series

The table summarizes the data series used to compute returns for each developed country in the sample. The table reports the sample period start and end dates and lists the total return index, price index, and dividend yield series used to construct returns (along with the corresponding start and end dates). For a given country, we use the total return index over the periods for which it is available, and we otherwise rely on price index and dividend yield data to estimate returns. The data symbols correspond to those in the GFDatabase from Global Financial Data. The total return index for Luxembourg for the period from 2016:12 to 2019:12 is from the Luxembourg Stock Exchange (https://www.bourse.lu/home). The dividend yields for Norway from 1914:02 to 1969:12 and Portugal from 1934:01 to 1988:01 are from Jordà et al. (2019).

			Total return index			Price index and dividend yield			
Country	Sample start	Sample end	Total return index	Start	End	Price index	Dividend yield	Start	End
Argentina	1947:02	1966:12				_IBGD	SYARGYM	1947:02	1966:12
Australia	1901:01	2019:12	_AORDAD	1901:01	2019:12				
Austria	1925:02	2019:12	_ATXTRD	1970:01	2019:12	_WBKID	SYAUTYM	1925:02	1969:12
Belgium	1897:01	2019:12	_BCSHD	1897:01	2019:12				
Canada	1891:01	2019:12	_TRGSPTSE	1891:01	2019:12				
Chile period I	1927:01	1970:12				_IGPAD	SYCHLYM	1927:01	1970:12
Chile period II	2010:01	2019:12	_IPSAD	2010:01	2019:05	_IGPAD	SYCHLYM	2019:06	2019:12
Czechoslovakia	1926:01	1946:06				CZINDXM	SYCZEYM	1926:01	1937:11
						CZINDEXM	SYCZEYM	1937:12	1943:03
Czech Republic	1995:01	2019:12	_PXTRD	1995:01	2019:12				
Denmark	1890:01	2019:12	_OMXCGID	1890:01	2019:12				
Estonia	2010:01	2019:12	_OMXTGID	2010:01	2019:12				
Finland	1969:01	2019:12	_OMXHGID	1969:01	2019:12				
France	1866:01	2019:12	TRSBF250D	1866:01	2019:12				
Germany	1882:01	2019:12	_CDAXD	1882:01	2019:12				
Greece	1977:01	2019:12	_RETMD	1977:01	2019:12				
Hungary	1996:01	2019:12	_BUXD	1996:01	2019:12				
Iceland	2002:01	2019:12	_OMXIGID	2002:07	2019:12	_OMXIPID	SYISLYM	2002:01	2002:06
Ireland	1936:01	2019:12	_IVRTD	1936:01	2019:12				
Israel	2010:01	2019:12	TRISRSTM	2010:01	2019:11	ILTLVGD	SYISRYM	2019:12	2019:12
Italy	1931:01	2019:12	_BCIPRD	1931:01	2019:12				
Japan	1930:01	2019:12	_TOPXDVD	1930:01	2019:12				
Latvia	2016:01	2019:12	_OMXRGID	2016:01	2019:12				
Lithuania	2018:01	2019:12	_OMXVGID	2018:01	2019:12				
Luxembourg	1982:01	2019:12	_LUXXRD	1985:01	2016:11	_LUXXD	SYLUXYM	1982:01	1984:12
			See table caption	2016:12	2019:12				
Mexico	1994:01	2019:12	_IRTD	1994:01	2019:12				

(*continued on next page*)

Table IA1 (*continued*)

			Total return index			Price index and dividend yield			
Country	Sample start	Sample end	Total return index	Start	End	Price index	Dividend yield	Start	End
Netherlands	1914:01	2019:12	_AAXGRD	1914:01	2019:12				
New Zealand	1896:01	2019:12	_NZGID	1896:01	2019:12				
Norway	1914:02	2019:12	_OSEAXD	1970:01	2019:12	_OBXPD	See table caption	1914:02	1969:12
Poland	1996:01	2019:12	_WIGD	1996:01	2019:12				
Portugal	1934:01	2019:12	_BVLGD	1988:02	2019:12	_IBTAD	See table caption	1934:01	1988:01
Singapore	1970:01	2019:12	_TFTFSTD	1970:01	2019:12				
Slovakia	2000:01	2019:12	_SAXD	2000:01	2019:12				
Slovenia	2010:01	2019:12				_SBITOPD	SYSVNYM	2010:01	2019:12
South Korea	1996:01	2019:12	TRKORSTM	1996:01	2019:12				
Spain	1959:01	2019:12	_BCNPR30	1959:01	2019:12				
Sweden	1910:01	2019:12	_OMXSBGI	1910:01	2019:12				
Switzerland	1914:01	2019:12	_SSHID	1914:01	2019:12				
Turkey	1986:02	2019:12	TRRBILED	1986:02	2019:12				
United Kingdom	1841:01	2019:12	_TFTASD	1841:01	2019:12				
United States	1890:01	2019:12	_SPXTRD	1890:01	2019:12				

Table IA2
Internal validation test data series

The table reports the total return index, price index, and dividend yield data used to test the validity of our return construction approach. For each country, we compare returns estimated from two approaches: (i) using the total return index and (ii) using the price index and dividend yield. The final testing sample includes all periods between the sample start date and the sample end date for which the total return index, the price index, and the dividend yield are available. The data symbols correspond to those in the GFDatabase from Global Financial Data.

Country	Testing period start date	Testing period end date	Total return index	Price index	Dividend yield
Australia	1901:01	2019:12	_AORDAD	_AORDD	SYAUSYM
Austria	1970:01	2019:12	_ATXTRD	_WBKID	SYAUTYM
Belgium	1897:01	2019:12	_BCSHD	_BSPTD	SYBELYM
Canada	1891:01	2019:12	_TRGSPTSE	_GSPTSED	SYCANYTM
Chile	2010:01	2019:05	_IPSAD	CLIPSAM	SYCHLYM
Czech Republic	1995:01	2019:12	_PXTRD	_CTXUSDD	SYCZEYM
Denmark	1921:01	2001:06	_OMXCGID	_CSEID	SYDNKYM
	2001:07	2019:12	_OMXCGID	_OMXCPID	SYDNKYM
Finland	1969:01	2019:12	_OMXHGID	_OMXHPID	SYFINYM
France	1866:01	2019:12	TRSBF250D	_CACTD	SYFRAYM
Germany	1882:01	2019:12	_CDAXD	_CXKXD	SYDEUYM
Greece	1977:01	2019:12	_RETMD	_ATGD	SYGRCYM
Iceland	2002:07	2006:12	_OMXIGID	_OMXIPID	SYISLYM
Ireland	1936:01	2019:12	_IVRTD	_ISEQD	SYIRLYM
Israel	2010:01	2019:11	TRISRSTM	ILTLVGD	SYISRYM
Italy	1931:01	2019:12	_BCIPRD	_BCIID	SYITAYM
Japan	1930:01	2019:12	_TOPXDVD	_TOPXD	SYJPNYM
Luxembourg	1985:01	1994:12	_LUXXRD	_LUXXD	SYLUXYM
Mexico	1994:01	2019:12	_IRTD	_MXXD	SYMEXYM
Netherlands	1914:01	2019:12	_AAXGRD	_AAXD	SYNLDYAM
New Zealand	1896:01	2019:12	_NZGID	_NZCID	SYNZLYM
Poland	1996:01	2019:12	_WIGD	_WIG20D	SYPOLYM
Portugal	1988:02	2006:05	_BVLGD	_IBTAD	SYPRTYM
Singapore	1972:01	2019:12	_TFTFSTD	_FTSTID	SYSGPYM
South Korea	1996:01	2019:12	TRKORSTM	_KS11D	SYKORYM
Spain	1959:01	2019:12	_BCNPR30	_SMSID	SYESPYM
Sweden	1910:01	2019:12	_OMXSBGI	_OMXSPID	SYSWEYM
Switzerland	1918:01	2019:12	_SSHID	_SPIXD	SYCHEYM
Turkey	1986:02	2019:12	TRRBILED	_XU100D	SYTURYM
United Kingdom	1933:07	2019:12	_TFTASD	_FTASD	_DFTASD
United States	1890:01	2019:12	_SPXTRD	_SPXD	SYUSAYM

Table IA3
Internal validation test results

The table reports results of tests to assess the validity our approach of computing returns from price index and dividend yield data. For each country, the table shows the number of monthly observations for which the total return index, the price index, and the dividend yield are available and the correlation between nominal returns computed using the total return index (R) and nominal returns computed using the price index and the dividend yield (R^*). The table also compares the arithmetic mean and the standard deviation of the total return index series with the corresponding statistics for the return series based on price index and dividend yield. The p-value for the comparison of means (volatilities) corresponds to a t-test (F-test) for difference in average return (variance).

			Comparison of means			Comparison of volatilities		
Country	Observ	Corr	$\bar{R}_a$ (%)	$\bar{R}_a^*$ (%)	p-value	$\sigma(R)$ (%)	$\sigma(R^*)$ (%)	p-value
Australia	1,424	0.997	0.98	0.99	0.932	3.88	3.86	0.861
Austria	600	0.971	0.74	0.76	0.948	5.54	5.22	0.144
Belgium	1,401	0.974	0.75	0.76	0.989	5.12	5.07	0.711
Canada	1,541	0.995	0.80	0.81	0.961	4.24	4.22	0.902
Chile	113	0.994	0.37	0.68	0.558	4.08	3.80	0.464
Czech Republic	300	0.942	1.09	0.87	0.732	7.47	8.39	0.045
Denmark	816	0.931	0.90	0.87	0.887	4.35	4.26	0.542
Finland	612	0.983	1.34	1.32	0.948	6.31	6.21	0.706
France	1,771	0.976	0.85	0.86	0.940	4.76	4.63	0.235
Germany	1,450	0.945	0.83	0.83	0.966	4.59	4.71	0.311
Greece	514	0.977	1.43	1.31	0.837	9.92	9.24	0.109
Iceland	54	0.943	3.21	2.97	0.818	5.69	5.18	0.501
Ireland	1,008	0.999	0.98	0.98	0.999	4.67	4.68	0.940
Israel	119	0.863	0.02	0.49	0.387	4.76	3.56	0.002
Italy	1,056	0.997	1.24	1.25	0.963	7.55	7.55	0.985
Japan	1,032	0.999	1.05	1.06	0.965	5.81	5.81	0.981
Luxembourg	120	0.993	1.41	1.68	0.640	4.35	4.68	0.435
Mexico	300	0.995	1.28	1.31	0.948	6.32	6.54	0.553
Netherlands	1,242	0.997	0.82	0.81	0.975	5.05	5.06	0.951
New Zealand	1,483	0.985	0.86	0.84	0.843	3.63	3.65	0.825
Poland	288	0.974	0.94	0.83	0.856	6.82	7.41	0.160
Portugal	220	0.932	0.63	0.70	0.886	5.59	5.53	0.870
Singapore	576	0.944	0.93	0.99	0.890	7.52	7.13	0.209
South Korea	288	0.978	0.96	0.76	0.772	8.27	7.96	0.517
Spain	732	0.955	0.99	1.07	0.776	5.44	5.39	0.807
Sweden	1,320	0.992	0.89	0.91	0.890	4.76	4.71	0.708
Switzerland	1,224	0.993	0.68	0.68	0.994	4.38	4.33	0.696
Turkey	407	0.984	4.26	4.24	0.982	15.75	15.72	0.979
United Kingdom	1,038	0.998	0.99	1.00	0.972	4.87	4.86	0.948
United States	1,560	0.999	0.87	0.87	0.977	4.95	4.95	0.979
Full sample	24,609	0.980	0.97	0.98	0.956	5.65	5.61	0.232

Table IA4
External validation test results

The table reports summary statistics for annual returns for each developed country with a return sample that overlaps with the sample from Jordà et al. (2019). For each country, the table shows the number of sample years and the number of sample observations (i.e., including multi-year observations). The table also shows the following summary statistics for our sample and for Jordà et al.'s (2019) sample: the arithmetic average return ($\bar{R}_a$), the geometric average return ($\bar{R}_g$), the standard deviation of return (SD), the minimum (Min) and the maximum (Max) return, and the correlation between the return samples (Corr). Statistics for the pooled sample of all observations are also reported. Panel A (Panel B) shows results for nominal returns (real returns).

	Sample size		Summary statistics for returns										
			Global Financial Data					Jordà et al. (2019)					
Country	Years	Observ	$\bar{R}_a$ (%)	$\bar{R}_g$ (%)	SD (%)	Min (%)	Max (%)	$\bar{R}_a$ (%)	$\bar{R}_g$ (%)	SD (%)	Min (%)	Max (%)	Corr
Panel A: Nominal returns													
Australia	115	114	12.70	11.30	17.60	−40.38	66.80	11.93	10.57	17.21	−40.38	63.70	0.99
Belgium	118	113	10.03	7.75	23.41	−47.56	117.31	10.67	8.07	25.04	−56.06	125.80	0.93
Denmark	126	126	9.46	7.63	21.14	−47.96	105.86	11.10	9.62	18.65	−57.96	81.64	0.72
Finland	47	47	20.51	15.21	37.64	−51.31	167.04	19.47	14.17	37.46	−51.31	167.04	0.99
France	146	146	12.30	9.76	27.16	−40.89	211.62	8.38	6.26	23.13	−40.89	125.38	0.90
Germany	124	118	10.83	7.20	27.94	−86.20	137.27	11.24	7.16	29.83	−89.43	137.76	0.97
Italy	85	85	15.94	11.71	33.32	−46.47	132.22	15.88	11.32	35.35	−46.84	160.20	0.97
Japan	86	82	16.61	11.99	40.31	−40.62	257.57	13.74	10.61	32.31	−27.82	226.33	0.81
Netherlands	100	98	10.75	8.44	22.47	−49.46	73.19	10.90	8.39	24.60	−49.84	130.07	0.82
Norway	100	100	11.88	8.49	30.02	−52.59	179.47	10.17	7.74	23.27	−54.06	92.80	0.88
Portugal	81	78	15.54	8.02	44.47	−89.21	225.14	11.82	6.43	34.28	−86.70	136.88	0.85
Spain	57	57	14.14	11.42	25.32	−39.38	91.51	14.15	11.55	25.36	−36.95	113.26	0.91
Sweden	106	106	11.99	9.68	22.55	−38.60	71.23	12.11	9.80	22.55	−39.27	69.76	1.00
Switzerland	99	99	9.17	7.13	21.05	−34.05	72.82	9.21	7.40	19.91	−34.05	61.36	0.97
United Kingdom	145	144	9.32	7.82	19.55	−51.75	152.12	10.14	8.64	19.52	−50.20	149.61	0.96
United States	126	126	11.00	9.19	19.32	−43.86	52.89	10.91	9.26	18.47	−40.30	52.64	0.99
Full sample	1,661	1,637	12.02	9.17	26.81	−89.21	257.57	11.44	8.84	24.95	−89.43	226.33	0.90

(*continued on next page*)

Table IA4 (*continued*)

	Sample size		Summary statistics for returns										
			Global Financial Data					Jordà et al. (2019)					
Country	Years	Observ	$\bar{R}_a$ (%)	$\bar{R}_g$ (%)	SD (%)	Min (%)	Max (%)	$\bar{R}_a$ (%)	$\bar{R}_g$ (%)	SD (%)	Min (%)	Max (%)	Corr
Panel B: Real returns													
Australia	115	114	8.58	7.10	17.46	−42.50	53.53	7.82	6.39	17.10	−42.50	50.67	0.99
Belgium	118	113	5.20	2.68	22.80	−63.88	81.80	5.63	2.98	23.50	−60.28	88.90	0.94
Denmark	126	126	5.76	3.91	20.47	−49.17	92.27	7.20	5.83	17.04	−49.17	71.30	0.73
Finland	47	47	15.35	10.00	37.17	−52.92	161.38	14.37	9.02	37.06	−52.92	161.38	0.99
France	146	146	6.31	3.97	23.10	−42.54	126.59	3.09	0.65	22.76	−50.55	111.39	0.91
Germany	124	118	9.21	5.33	28.08	−90.61	157.17	9.58	5.29	29.63	−92.80	156.69	0.97
Italy	85	85	5.73	1.89	27.86	−70.76	89.81	5.52	1.54	28.12	−72.78	96.89	0.97
Japan	86	82	8.94	3.57	31.34	−93.15	143.11	6.39	2.28	23.84	−93.75	81.82	0.71
Netherlands	100	98	7.36	5.08	21.78	−50.43	66.07	7.41	5.03	23.02	−50.43	99.72	0.82
Norway	100	100	7.72	4.42	28.65	−53.58	166.92	6.10	3.70	22.55	−55.02	80.07	0.89
Portugal	81	78	8.52	1.09	38.31	−95.18	168.71	5.41	−0.40	31.03	−94.07	102.75	0.85
Spain	57	57	7.38	4.49	25.12	−40.24	76.91	7.41	4.60	25.19	−43.64	97.00	0.92
Sweden	106	106	8.09	5.72	22.15	−40.01	68.99	8.20	5.83	22.17	−40.00	67.53	1.00
Switzerland	99	99	7.08	4.93	21.36	−36.39	69.24	7.11	5.20	20.14	−37.83	56.30	0.98
United Kingdom	145	144	6.24	4.73	18.07	−59.54	101.69	7.02	5.52	17.82	−58.23	99.69	0.95
United States	126	126	8.08	6.22	19.51	−38.09	53.43	7.99	6.29	18.64	−38.81	50.72	0.99
Full sample	1,661	1,637	7.51	4.60	24.65	−95.18	168.71	6.99	4.28	23.22	−94.07	161.38	0.91

Table IA5
Bootstrap distributions of payoffs for alternative block sampling lengths

The table summarizes the distribution of real payoffs from a $1.00 buy-and-hold investment across 1,000,000 bootstrap simulations at various return horizons for alternative mean block sampling lengths. The real payoffs are from the perspective of a domestic investor in a representative country. Each panel of the table corresponds to a specific return horizon. Each panel presents results for i.i.d. sampling and sampling with mean block lengths of 12 months, 120 months (base case), and 240 months. The underlying sample is the pooled sample of all developed countries. The real payoff for bootstrap iteration m at the H-month horizon is $W_H^{(m)}$. For each horizon and block sampling length, the table reports the mean, standard deviation, and distribution percentiles of real payoffs. The last column in the table shows the proportion of payoff draws that are less than one [$\mathbb{P}(W_H^{(m)} < 1)$]. The bootstrap sampling procedure is based on the stationary bootstrap approach of Politis and Romano (1994b) as described in the text. For the designs other than i.i.d. sampling, we sample blocks of random length, where the length of each block has a geometric distribution with the indicated mean block length.

	Moments		Percentiles									
Block length	Mean	SD	1%	5%	10%	25%	50%	75%	90%	95%	99%	$\mathbb{P}(W_H^{(m)} < 1)$
Panel A: 1 month												
1 month (i.i.d.)	1.01	0.06	0.85	0.92	0.95	0.98	1.01	1.03	1.06	1.09	1.16	0.432
12 months	1.01	0.06	0.85	0.92	0.95	0.98	1.01	1.03	1.06	1.09	1.16	0.432
120 months	1.01	0.06	0.85	0.92	0.95	0.98	1.01	1.03	1.06	1.09	1.16	0.432
240 months	1.01	0.06	0.85	0.92	0.95	0.98	1.01	1.03	1.06	1.09	1.16	0.432
Panel B: 1 year												
1 month (i.i.d.)	1.07	0.22	0.62	0.76	0.83	0.93	1.05	1.19	1.33	1.43	1.70	0.385
12 months	1.08	0.27	0.54	0.73	0.81	0.93	1.06	1.19	1.34	1.48	1.87	0.372
120 months	1.08	0.28	0.52	0.72	0.80	0.93	1.06	1.19	1.35	1.49	1.90	0.368
240 months	1.08	0.28	0.52	0.72	0.80	0.93	1.06	1.19	1.35	1.49	1.90	0.368
Panel C: 5 years												
1 month (i.i.d.)	1.40	0.66	0.33	0.59	0.72	0.95	1.28	1.70	2.21	2.60	3.61	0.288
12 months	1.46	0.94	0.20	0.52	0.66	0.93	1.29	1.77	2.39	2.92	4.54	0.300
120 months	1.45	0.93	0.19	0.54	0.69	0.95	1.29	1.72	2.34	2.86	4.41	0.283
240 months	1.45	0.94	0.19	0.54	0.70	0.96	1.29	1.71	2.33	2.86	4.41	0.280
Panel D: 10 years												
1 month (i.i.d.)	1.95	1.38	0.16	0.52	0.70	1.05	1.61	2.45	3.56	4.47	6.96	0.226
12 months	2.13	2.11	0.12	0.42	0.61	1.00	1.64	2.62	4.04	5.34	9.47	0.249
120 months	2.02	1.76	0.13	0.47	0.69	1.08	1.64	2.44	3.60	4.63	8.75	0.215
240 months	1.99	1.65	0.14	0.48	0.71	1.10	1.65	2.41	3.50	4.44	8.62	0.209
Panel E: 20 years												
1 month (i.i.d.)	3.79	4.27	0.14	0.47	0.74	1.38	2.56	4.68	8.00	11.04	20.36	0.159
12 months	4.54	7.44	0.10	0.35	0.61	1.27	2.62	5.22	9.77	14.36	30.94	0.189
120 months	3.88	5.40	0.14	0.43	0.73	1.43	2.63	4.57	7.76	10.89	22.97	0.155
240 months	3.70	4.59	0.16	0.45	0.77	1.47	2.64	4.39	7.25	10.05	20.97	0.147
Panel F: 30 years												
1 month (i.i.d.)	7.35	11.32	0.13	0.48	0.85	1.86	4.05	8.53	16.51	24.52	51.78	0.123
12 months	9.69	27.53	0.09	0.35	0.67	1.68	4.16	9.88	21.43	34.34	85.44	0.152
120 months	7.38	13.76	0.14	0.47	0.85	1.94	4.16	8.28	15.58	23.30	53.45	0.121
240 months	6.73	11.24	0.17	0.53	0.91	2.04	4.20	7.86	13.95	20.29	43.79	0.111

Table IA6
Bootstrap distributions of nominal payoffs

The table summarizes the distribution of nominal payoffs from a $1.00 buy-and-hold investment across 1,000,000 bootstrap simulations at various return horizons. The underlying sample is the pooled sample of all developed countries. The nominal payoffs are from the perspective of a domestic investor in a representative country. The nominal payoff for bootstrap iteration m at the H-month horizon is $W_H^{(m)}$. For each horizon, the table reports the mean, standard deviation, and distribution percentiles of nominal payoffs. The last column in the table shows the proportion of payoff draws that are less than one $[\mathbb{P}(W_H^{(m)} < 1)]$. The bootstrap sampling procedure is based on the stationary bootstrap approach of Politis and Romano (1994b) as described in the text. We sample blocks of random length, where the length of each block has a geometric distribution with a mean of 120 months.

	Moments		Percentiles									
Horizon	Mean	SD	1%	5%	10%	25%	50%	75%	90%	95%	99%	$\mathbb{P}(W_H^{(m)} < 1)$
1 month	1.01	0.07	0.86	0.93	0.95	0.98	1.01	1.03	1.06	1.09	1.17	0.396
1 year	1.13	0.35	0.58	0.77	0.85	0.98	1.09	1.23	1.41	1.58	2.18	0.291
5 years	1.98	2.87	0.40	0.71	0.89	1.17	1.54	2.10	3.05	4.01	8.32	0.147
10 years	5.00	31.78	0.33	0.82	1.09	1.60	2.37	3.73	6.22	9.40	33.36	0.080
20 years	77.92	3,001.12	0.45	1.21	1.81	3.17	5.74	11.18	23.64	42.95	269.57	0.036
30 years	1,027.13	340,838.32	0.72	2.00	3.23	6.58	14.16	32.76	83.23	175.81	1,637.38	0.017

Table IA7
Bootstrap distributions of continuously compounded returns

The table summarizes the distribution of continuously compounded returns across 1,000,000 bootstrap simulations at various return horizons. The underlying sample in Panel A (Panel B) is the pooled sample of all developed countries (the United States over the period from 1890 to 2019). The real returns are from the perspective of a domestic investor in a representative country. For each horizon and sample, the table reports the mean, standard deviation, skewness, and kurtosis of continuously compounded real returns. The bootstrap sampling procedure is based on the stationary bootstrap approach of Politis and Romano (1994b) as described in the text. We sample blocks of random length, where the length of each block has a geometric distribution with a mean of 120 months.

Horizon	Mean	SD	Skew	Kurt
Panel A: Full sample				
1 month	0.0038	0.0640	−8.141	315.986
1 year	0.0453	0.2583	−1.899	23.464
5 years	0.2293	0.5573	−0.911	8.021
10 years	0.4522	0.7396	−0.743	5.875
20 years	0.8982	0.9946	−0.525	4.497
30 years	1.3448	1.1912	−0.445	4.022
Panel B: United States sample				
1 month	0.0052	0.0496	−0.380	10.996
1 year	0.0615	0.1905	−0.584	5.115
5 years	0.3105	0.3931	−0.114	3.041
10 years	0.6207	0.5197	−0.069	2.842
20 years	1.2407	0.6902	−0.035	2.887
30 years	1.8619	0.8079	−0.025	3.043

Table IA8
Alternative sample periods

The table shows the sample start dates for developed countries in the alternative samples with screens based on population or equity market size. The alternative samples are described in Table 5 of the paper.

Country	Base sample start date	POP 0.2% start date	POP 0.5% start date	M/GDP 0.5 start date	M/GDP 1.0 start date
United Kingdom	1841:01	1841:01	1841:01	1846:01	1880:01
Netherlands	1914:01	1914:01	—	1914:01	1914:01
Belgium	1897:01	1897:01	—	1897:01	1908:01
France	1866:01	1866:01	1866:01	1881:01	1999:01
Norway	1914:02	—	—	2005:01	—
Germany	1882:01	1882:01	1882:01	1999:01	—
Denmark	1890:01	—	—	1997:01	2015:01
Switzerland	1914:01	1914:01	—	1927:01	1928:01
United States	1890:01	1890:01	1890:01	1927:01	1998:01
Canada	1891:01	1891:01	1943:01	1901:01	1928:01
Argentina	1947:02	1947:02	1947:02	—	—
New Zealand	1896:01	—	—	1899:01	1931:01
Australia	1901:01	1901:01	—	1931:01	1950:01
Sweden	1910:01	1910:01	—	1910:01	1997:01
Austria	1925:02	1925:02	—	2006:01	—
Chile period I	1927:01	1927:01	—	1927:01	1928:01
Greece	1977:01	1977:01	—	1998:01	1999:01
Czechoslovakia	1926:01	1926:01	1926:01	—	—
Japan	1930:01	1930:01	1930:01	1937:01	1988:01
Portugal	1934:01	1934:01	—	1998:01	—
Italy	1931:01	1931:01	1931:01	1999:01	—
Ireland	1936:01	—	—	1944:01	1964:01
Singapore	1970:01	—	—	1970:01	1970:01
Luxembourg	1982:01	—	—	1984:01	1988:01
Turkey	1986:02	1986:02	1986:02	1999:01	—
Spain	1959:01	1959:01	1959:01	1997:01	2007:01
Finland	1969:01	—	—	1997:01	1998:01
Iceland	2002:01	—	—	2002:01	2004:01
Mexico	1994:01	1994:01	1994:01	—	—
Czech Republic	1995:01	—	—	—	—
Hungary	1996:01	—	—	—	—
Poland	1996:01	1996:01	1996:01	—	—
South Korea	1996:01	1996:01	1996:01	1999:01	2007:01
Slovakia	2000:01	—	—	—	—
Chile period II	2010:01	2010:01	—	2010:01	2010:01
Estonia	2010:01	—	—	—	—
Israel	2010:01	—	—	2010:01	—
Slovenia	2010:01	—	—	—	—
Latvia	2016:01	—	—	—	—
Lithuania	2018:01	—	—	—	—

Table IA9
Bootstrap distributions of payoffs for alternative samples

The table summarizes the distribution of real payoffs from a $1.00 buy-and-hold investment across 1,000,000 bootstrap simulations at various return horizons for alternative underlying samples. The real payoffs are from the perspective of a domestic investor in a representative country. Each panel of the table corresponds to a specific sample, and the samples are summarized in Table 5 of the paper. The real payoff for bootstrap iteration m at the H-month horizon is $W_H^{(m)}$. For each horizon and sample, the table reports the mean, standard deviation, and distribution percentiles of real payoffs. The last column in the table shows the proportion of payoff draws that are less than one [$\mathbb{P}(W_H^{(m)} < 1)$]. The bootstrap sampling procedure is based on the stationary bootstrap approach of Politis and Romano (1994b) as described in the text. We sample blocks of random length, where the length of each block has a geometric distribution with a mean of 120 months.

	Moments		Percentiles									
Horizon	Mean	SD	1%	5%	10%	25%	50%	75%	90%	95%	99%	$\mathbb{P}(W_H^{(m)} < 1)$
Panel A: U.S. sample												
1 month	1.01	0.05	0.87	0.93	0.95	0.98	1.01	1.03	1.06	1.08	1.12	0.407
1 year	1.08	0.20	0.62	0.76	0.84	0.96	1.08	1.20	1.32	1.40	1.56	0.320
5 years	1.47	0.59	0.54	0.71	0.82	1.04	1.40	1.77	2.24	2.57	3.32	0.217
10 years	2.12	1.15	0.59	0.77	0.92	1.31	1.86	2.68	3.68	4.26	5.85	0.127
20 years	4.38	3.32	0.70	1.11	1.41	2.16	3.47	5.55	8.54	10.65	16.37	0.036
30 years	8.91	8.49	0.96	1.69	2.29	3.76	6.46	11.03	18.08	24.30	41.90	0.012
Panel B: U.K. sample												
1 month	1.00	0.04	0.89	0.94	0.97	0.99	1.00	1.02	1.04	1.06	1.09	0.398
1 year	1.06	0.15	0.71	0.83	0.88	0.99	1.05	1.13	1.23	1.32	1.51	0.293
5 years	1.34	0.43	0.50	0.71	0.86	1.05	1.29	1.55	1.90	2.10	2.71	0.197
10 years	1.76	0.79	0.52	0.76	0.95	1.24	1.61	2.08	2.85	3.29	4.34	0.129
20 years	3.06	1.98	0.56	0.95	1.23	1.81	2.60	3.72	5.41	6.94	10.38	0.058
30 years	5.28	4.25	0.69	1.24	1.68	2.68	4.21	6.48	9.95	12.97	21.70	0.030
Panel C: Survival sample												
1 month	1.01	0.06	0.85	0.92	0.95	0.98	1.01	1.03	1.06	1.08	1.16	0.430
1 year	1.08	0.28	0.53	0.72	0.81	0.94	1.06	1.19	1.35	1.48	1.90	0.365
5 years	1.47	0.96	0.22	0.54	0.69	0.96	1.30	1.73	2.36	2.90	4.52	0.278
10 years	2.05	1.81	0.16	0.48	0.70	1.09	1.66	2.48	3.66	4.72	8.94	0.211
20 years	3.98	5.37	0.16	0.46	0.76	1.47	2.70	4.68	7.99	11.23	23.62	0.147
30 years	7.68	14.05	0.17	0.53	0.92	2.05	4.33	8.59	16.24	24.41	55.89	0.111
Panel D: Continuous sample												
1 month	1.01	0.06	0.85	0.92	0.95	0.98	1.01	1.03	1.06	1.09	1.16	0.427
1 year	1.09	0.27	0.55	0.73	0.80	0.93	1.07	1.21	1.37	1.50	1.90	0.362
5 years	1.51	0.87	0.36	0.59	0.73	0.98	1.34	1.81	2.49	3.04	4.46	0.262
10 years	2.21	1.79	0.30	0.62	0.80	1.18	1.78	2.67	3.94	5.08	9.16	0.173
20 years	4.66	5.77	0.33	0.73	1.05	1.82	3.16	5.47	9.35	13.16	26.67	0.091
30 years	9.78	16.71	0.40	0.97	1.49	2.86	5.60	10.87	20.63	30.93	69.01	0.052
Panel E: Post-1880 sample												
1 month	1.01	0.06	0.85	0.92	0.95	0.98	1.01	1.03	1.06	1.09	1.16	0.434
1 year	1.08	0.28	0.52	0.72	0.80	0.93	1.06	1.19	1.35	1.49	1.91	0.373
5 years	1.46	0.94	0.19	0.53	0.69	0.95	1.29	1.73	2.35	2.88	4.47	0.286
10 years	2.02	1.78	0.13	0.46	0.68	1.07	1.64	2.45	3.62	4.67	8.79	0.219
20 years	3.89	5.55	0.13	0.42	0.72	1.41	2.62	4.58	7.81	10.96	22.94	0.158
30 years	7.41	15.45	0.14	0.46	0.83	1.91	4.13	8.30	15.71	23.57	54.41	0.124

(*continued on next page*)

Table IA9 (*continued*)

Horizon	Moments		Percentiles									$\mathbb{P}(W_H^{(m)} < 1)$
	Mean	SD	1%	5%	10%	25%	50%	75%	90%	95%	99%	
Panel F: Post-1920 sample												
1 month	1.01	0.06	0.85	0.92	0.95	0.98	1.01	1.03	1.06	1.09	1.17	0.434
1 year	1.09	0.29	0.52	0.71	0.80	0.93	1.07	1.21	1.37	1.52	1.95	0.376
5 years	1.50	0.99	0.18	0.54	0.69	0.95	1.31	1.79	2.45	3.01	4.70	0.286
10 years	2.13	1.95	0.13	0.48	0.71	1.10	1.70	2.59	3.84	4.99	9.47	0.208
20 years	4.27	6.10	0.14	0.45	0.78	1.52	2.81	5.03	8.67	12.21	25.65	0.141
30 years	8.50	20.04	0.15	0.52	0.95	2.16	4.65	9.43	18.08	27.29	63.15	0.106
Panel G: Post-1960 sample												
1 month	1.01	0.06	0.84	0.92	0.94	0.98	1.01	1.04	1.07	1.09	1.17	0.440
1 year	1.09	0.30	0.52	0.70	0.79	0.92	1.07	1.21	1.38	1.53	1.99	0.382
5 years	1.50	1.03	0.28	0.56	0.69	0.94	1.29	1.77	2.49	3.08	4.98	0.289
10 years	2.14	2.02	0.21	0.54	0.73	1.08	1.68	2.58	3.84	4.97	9.75	0.211
20 years	4.34	6.48	0.20	0.56	0.86	1.54	2.79	5.01	8.79	12.50	26.96	0.129
30 years	8.73	17.73	0.23	0.66	1.09	2.25	4.66	9.49	18.46	28.05	66.19	0.089
Panel H: Post-2000 sample												
1 month	1.00	0.06	0.85	0.91	0.94	0.98	1.01	1.04	1.06	1.08	1.14	0.430
1 year	1.06	0.23	0.48	0.67	0.76	0.92	1.07	1.20	1.33	1.43	1.66	0.373
5 years	1.33	0.77	0.23	0.48	0.62	0.87	1.20	1.61	2.10	2.56	4.27	0.347
10 years	1.67	1.33	0.18	0.45	0.62	0.93	1.38	2.01	2.91	3.74	7.07	0.288
20 years	2.69	3.23	0.15	0.40	0.62	1.09	1.86	3.17	5.30	7.50	15.11	0.220
30 years	4.33	7.09	0.14	0.39	0.65	1.29	2.53	4.88	9.09	13.52	30.11	0.180
Panel I: POP 0.2% sample												
1 month	1.01	0.06	0.85	0.92	0.95	0.98	1.01	1.03	1.06	1.08	1.16	0.435
1 year	1.08	0.29	0.52	0.71	0.80	0.93	1.06	1.19	1.34	1.48	1.90	0.369
5 years	1.44	0.96	0.17	0.52	0.67	0.94	1.28	1.70	2.31	2.83	4.29	0.294
10 years	1.98	1.83	0.11	0.42	0.65	1.04	1.62	2.38	3.52	4.54	8.86	0.231
20 years	3.69	5.40	0.11	0.36	0.64	1.32	2.50	4.35	7.37	10.40	22.41	0.177
30 years	6.83	14.13	0.11	0.39	0.72	1.73	3.82	7.65	14.40	21.63	50.53	0.143
Panel J: POP 0.5% sample												
1 month	1.01	0.07	0.84	0.92	0.94	0.98	1.01	1.03	1.06	1.09	1.17	0.430
1 year	1.08	0.31	0.50	0.71	0.80	0.94	1.06	1.19	1.35	1.48	1.93	0.359
5 years	1.44	0.90	0.10	0.52	0.68	0.97	1.29	1.70	2.28	2.83	4.16	0.271
10 years	1.98	1.69	0.09	0.35	0.65	1.08	1.65	2.39	3.52	4.51	8.17	0.220
20 years	3.66	4.68	0.10	0.29	0.60	1.37	2.58	4.37	7.40	10.32	20.81	0.174
30 years	6.74	12.03	0.09	0.33	0.66	1.77	3.95	7.74	14.35	21.28	47.51	0.147
Panel K: M/GDP 0.5 sample												
1 month	1.01	0.05	0.86	0.92	0.95	0.98	1.01	1.03	1.06	1.08	1.14	0.424
1 year	1.07	0.23	0.54	0.73	0.81	0.94	1.07	1.19	1.33	1.44	1.75	0.354
5 years	1.41	0.68	0.33	0.59	0.72	0.97	1.30	1.71	2.23	2.67	3.66	0.271
10 years	1.94	1.29	0.20	0.57	0.75	1.12	1.66	2.41	3.44	4.20	6.41	0.196
20 years	3.62	3.52	0.20	0.58	0.88	1.55	2.70	4.49	7.21	9.64	17.10	0.124
30 years	6.72	8.71	0.22	0.68	1.10	2.21	4.37	8.15	14.28	20.20	39.41	0.087

(*continued on next page*)

Table IA9 (*continued*)

	Moments		Percentiles									
Horizon	Mean	SD	1%	5%	10%	25%	50%	75%	90%	95%	99%	$\mathbb{P}(W_H^{(m)} < 1)$
						Panel L: M/GDP 1.0 sample						
1 month	1.01	0.05	0.86	0.92	0.95	0.98	1.01	1.03	1.06	1.08	1.14	0.427
1 year	1.07	0.22	0.53	0.72	0.80	0.94	1.06	1.19	1.33	1.44	1.71	0.363
5 years	1.37	0.61	0.36	0.57	0.71	0.95	1.29	1.68	2.12	2.47	3.35	0.289
10 years	1.82	1.10	0.23	0.56	0.73	1.07	1.59	2.30	3.19	3.85	5.53	0.213
20 years	3.23	2.88	0.23	0.55	0.82	1.45	2.48	4.08	6.34	8.36	14.15	0.139
30 years	5.69	6.50	0.24	0.62	0.99	1.98	3.87	7.09	12.00	16.61	30.93	0.102

Table IA10

Bootstrap distributions of 30-year USD payoffs for biased samples

The table summarizes the distribution of payoffs from a \$1.00 buy-and-hold investment across 1,000,000 bootstrap simulations at the 30-year horizon for alternative samples. The real payoffs are from the perspective of a global USD investor. The underlying sample in Panel A is the pooled sample of all developed countries. The underlying samples in Panel B are the United States over the period from 1890 to 2019 (U.S.) and the United Kingdom over the period from 1841 to 2019 (U.K.). The underlying samples in Panel C are the sample conditioned on current membership in the OECD (Survival) and the sample conditioned on current membership in the OECD and continuous data (Continuous). The real payoff for bootstrap iteration m at the H-month horizon is $W_H^{(m)}$. For each horizon, the table reports the mean, standard deviation, and distribution percentiles of real payoffs. The last column in the table shows the proportion of payoff draws that are less than one [$\mathbb{P}(W_H^{(m)} < 1)$]. The bootstrap sampling procedure is based on the stationary bootstrap approach of Politis and Romano (1994b) as described in the text. We sample blocks of random length, where the length of each block has a geometric distribution with a mean of 120 months.

	Moments		Percentiles									
Sample	Mean	SD	1%	5%	10%	25%	50%	75%	90%	95%	99%	$\mathbb{P}(W_H^{(m)} < 1)$
						Panel A: Base case						
Full sample	8.66	22.92	0.06	0.34	0.71	1.82	4.21	9.00	18.22	28.51	71.96	0.140
						Panel B: Single country						
U.S.	8.91	8.49	0.96	1.69	2.29	3.76	6.46	11.03	18.08	24.30	41.90	0.012
U.K.	4.90	4.89	0.50	0.94	1.29	2.13	3.50	5.89	9.86	13.49	24.54	0.058
						Panel C: Survival and easy data						
Survival	8.94	21.54	0.06	0.39	0.77	1.92	4.35	9.22	18.83	29.69	75.52	0.130
Continuous	11.02	24.09	0.32	0.84	1.33	2.68	5.56	11.59	23.46	36.74	88.19	0.065

Table IA11
Bootstrap distributions of 30-year USD payoffs with additional sample screens

The table summarizes the distribution of real payoffs from a \$1.00 buy-and-hold investment across 1,000,000 bootstrap simulations at the 30-year horizon for alternative samples. The real payoffs are from the perspective of a global USD investor. The underlying sample in Panel A is the pooled sample of all developed countries. Panel B (Panel C) [Panel D] presents results for the full developed sample with additional sample screens based on sample start date (population) [ratio of market capitalization to GDP]. The underlying samples in Panels B to D are described in Table 5 of the paper. The real payoff for bootstrap iteration m at the H-month horizon is $W_H^{(m)}$. For each sample, the table reports the mean, standard deviation, and distribution percentiles of real payoffs. The last column in the table shows the proportion of payoff draws that are less than one [$\mathbb{P}(W_H^{(m)} < 1)$]. The bootstrap sampling procedure is based on the stationary bootstrap approach of Politis and Romano (1994b) as described in the text. We sample blocks of random length, where the length of each block has a geometric distribution with a mean of 120 months.

	Moments		Percentiles									
Sample	Mean	SD	1%	5%	10%	25%	50%	75%	90%	95%	99%	$\mathbb{P}(W_H^{(m)} < 1)$
Panel A: Base case												
Full sample	8.66	22.92	0.06	0.34	0.71	1.82	4.21	9.00	18.22	28.51	71.96	0.140
Panel B: Sample period												
Post-1880	8.76	32.19	0.06	0.34	0.70	1.80	4.20	9.07	18.50	29.10	74.25	0.143
Post-1920	10.35	27.08	0.06	0.38	0.81	2.11	4.92	10.67	21.92	34.52	88.32	0.122
Post-1960	10.68	27.37	0.22	0.66	1.11	2.37	5.06	10.78	22.12	35.13	91.21	0.088
Post-2000	5.72	17.87	0.11	0.33	0.56	1.18	2.51	5.40	11.48	19.04	53.19	0.207
Panel C: Population												
POP 0.2%	7.87	20.74	0.04	0.26	0.58	1.61	3.85	8.21	16.50	25.93	66.36	0.164
POP 0.5%	7.86	18.13	0.02	0.18	0.50	1.60	3.88	8.34	17.05	26.82	65.51	0.170
Panel D: Equity market size												
M/GDP 0.5	7.11	10.80	0.15	0.58	0.98	2.05	4.23	8.31	15.35	22.41	47.09	0.103
M/GDP 1.0	5.85	7.34	0.18	0.52	0.87	1.84	3.76	7.17	12.66	17.82	34.62	0.120

Table IA12

Asset allocation for annuity investment

The table shows results from asset allocation tests. For each return horizon, the table reports the optimal weight in stocks for an investor who relies on the developed country sample to form expectations about stock market performance (w_d) and the optimal weight in stocks for an investor who relies on the U.S. sample to form expectations about stock market performance (w_{us}). The investors have exponential utility with a risk aversion parameter of three and make contributions as a monthly annuity. Each investor allocates across the domestic stock market and a risk-free asset, where the risk-free asset is either the inflation-protected risk-free asset or cash. The table also reports the maximum annualized fee that the investor relying on the developed country sample would be willing to pay to use her optimal weight rather than adopt the optimal weight based on the U.S. sample.

	Inflation-protected risk-free asset			Cash		
	Weight in stocks			Weight in stocks		
Horizon	Developed country sample, w_d	United States sample, w_{us}	Fee (%)	Developed country sample, w_d	United States sample, w_{us}	Fee (%)
1 month	0.62	0.89	0.41	1.03	1.19	0.15
1 year	0.67	1.09	0.88	1.27	1.50	0.20
5 years	0.68	1.11	0.82	1.27	1.53	0.16
10 years	0.68	1.10	0.80	1.36	1.51	0.07
20 years	0.65	1.05	0.76	1.08	1.49	0.39
30 years	0.62	1.03	0.79	1.04	1.41	0.91

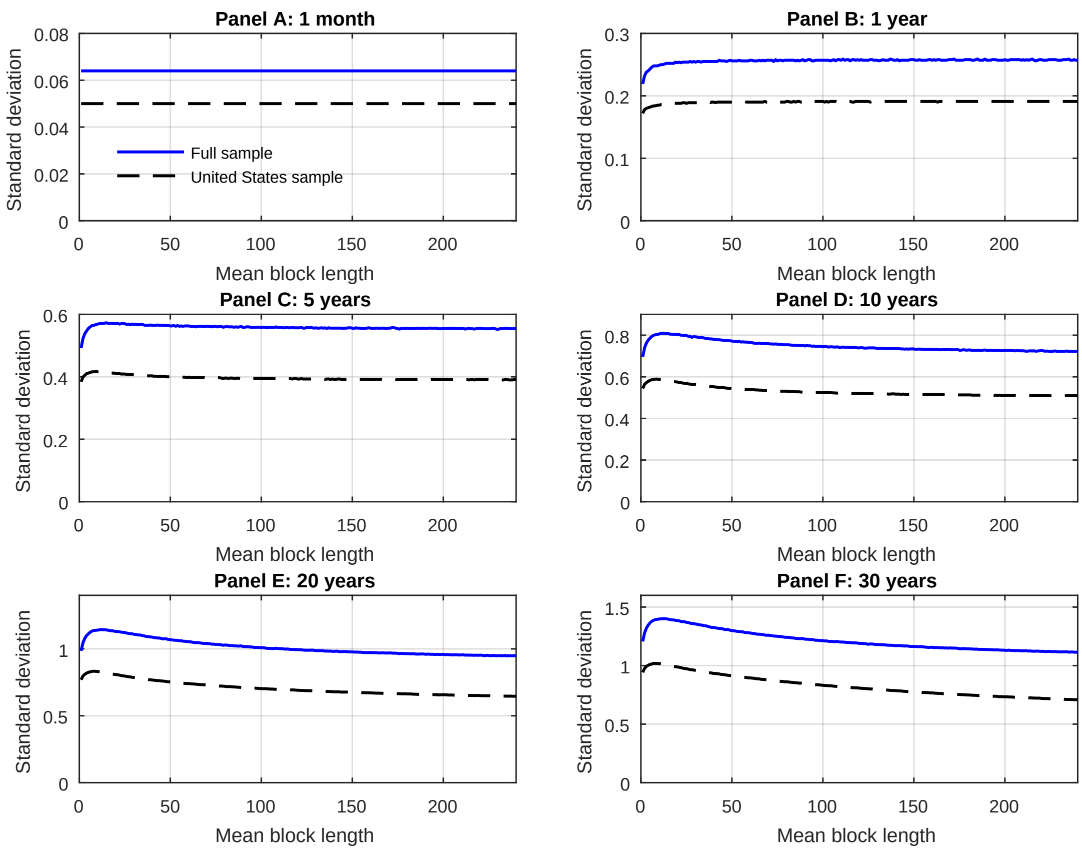

Fig. IA1. Standard deviation of continuously compounded returns for alternative block sampling lengths. The figure shows the standard deviation of continuously compounded returns across 1,000,000 bootstrap simulations at various return horizons for alternative mean block sampling lengths. The real returns are from the perspective of a domestic investor in a representative country. Each panel of the figure corresponds to a specific return horizon. The underlying sample for the simulated returns is the pooled sample of all developed countries (solid line) or the United States over the period from 1890 to 2019 (dashed line). The bootstrap sampling procedure is based on the stationary bootstrap approach of Politis and Romano (1994b) as described in the text.

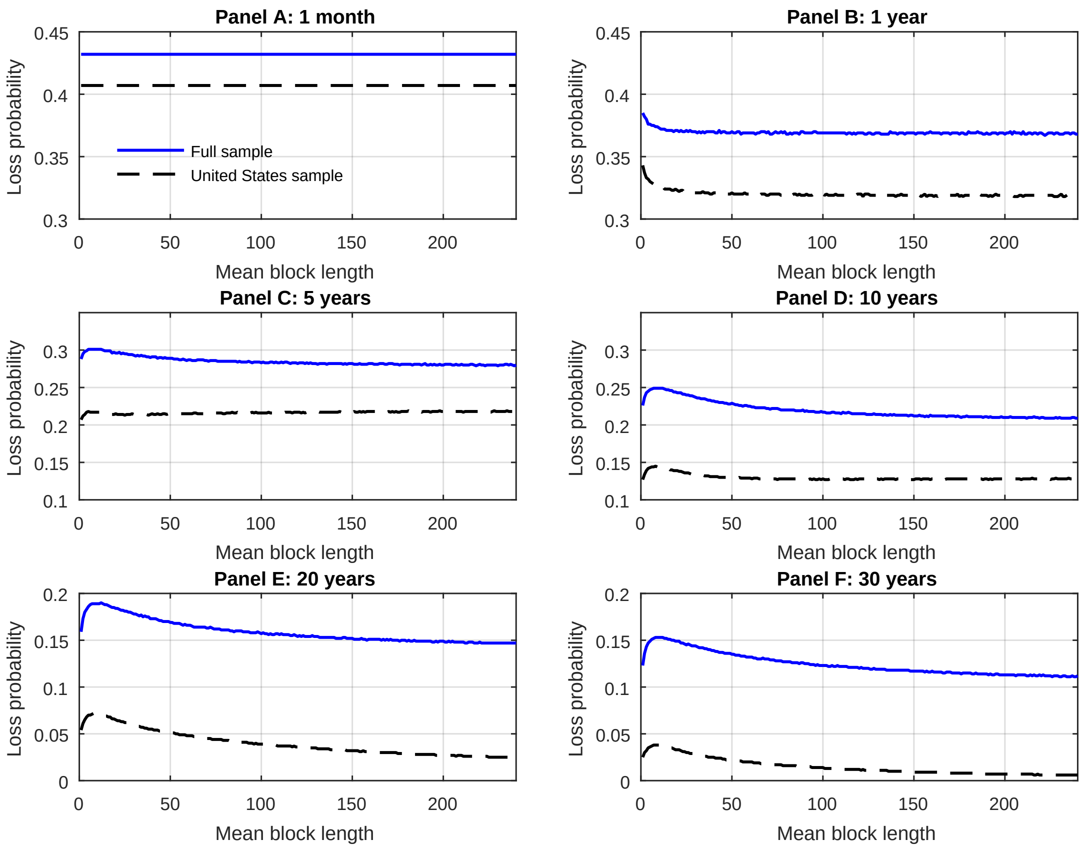

Fig. IA2. Loss probabilities for alternative block sampling lengths. The figure shows the proportion of real payoffs that are less than the initial investment across 1,000,000 bootstrap simulations at various return horizons for alternative mean block sampling lengths. The real payoffs are from the perspective of a domestic investor in a representative country. Each panel of the figure corresponds to a specific return horizon. The underlying sample for the simulated returns is the pooled sample of all developed countries (solid line) or the United States over the period from 1890 to 2019 (dashed line). The bootstrap sampling procedure is based on the stationary bootstrap approach of Politis and Romano (1994b) as described in the text.

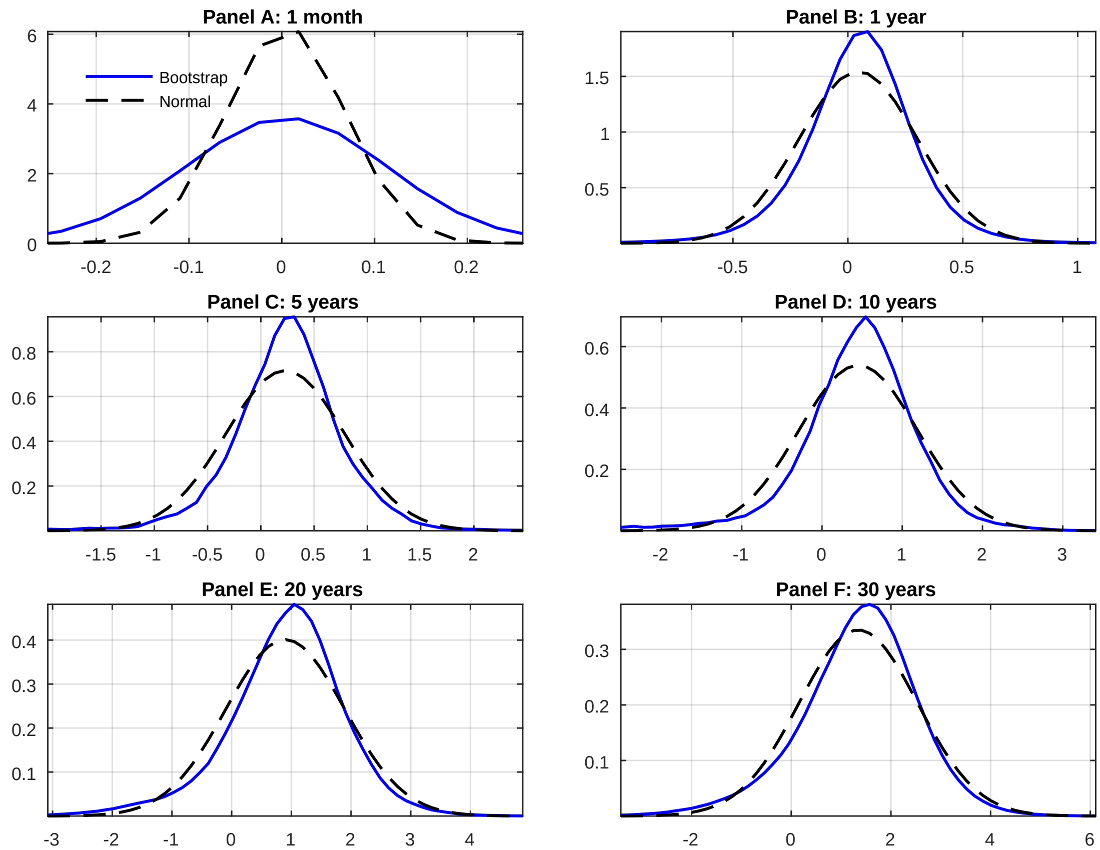

Fig. IA3. Continuously compounded returns. The figure shows distributions of continuously compounded real returns across 1,000,000 bootstrap simulations at various return horizons. The underlying sample for the simulated returns is the pooled sample of all developed countries. The real returns are from the perspective of a domestic investor in a representative country. In each panel, the solid line is the kernel smoothed density of simulated returns, and the dashed line is a normal density with mean and variance equal to those of the simulated returns. The bootstrap sampling procedure is based on the stationary bootstrap approach of Politis and Romano (1994b) as described in the text. We sample blocks of random length, where the length of each block has a geometric distribution with a mean of 120 months.

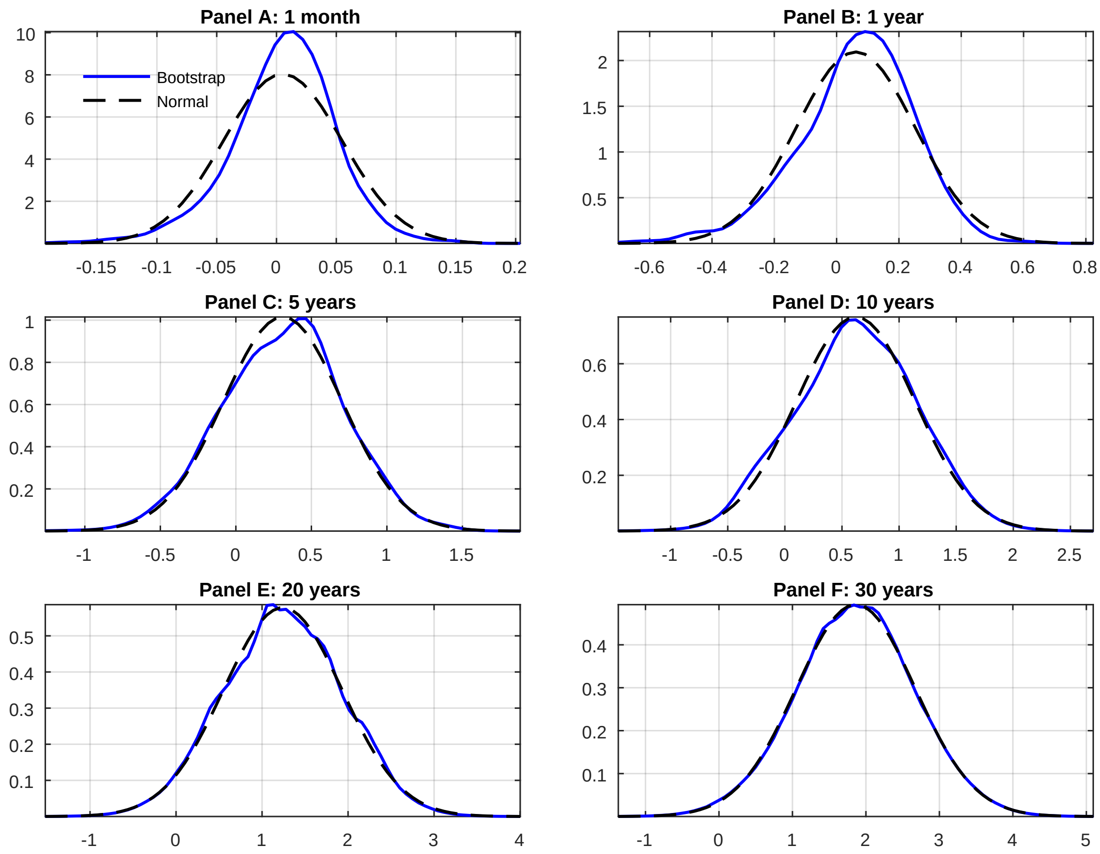

Fig. IA4. Continuously compounded returns for the United States sample. The figure shows distributions of continuously compounded real returns across 1,000,000 bootstrap simulations at various return horizons. The underlying sample for the simulated returns is the United States over the period from 1890 to 2019. In each panel, the solid line is the kernel smoothed density of simulated returns, and the dashed line is a normal density with mean and variance equal to those of the simulated returns. The bootstrap sampling procedure is based on the stationary bootstrap approach of Politis and Romano (1994b) as described in the text. We sample blocks of random length, where the length of each block has a geometric distribution with a mean of 120 months.

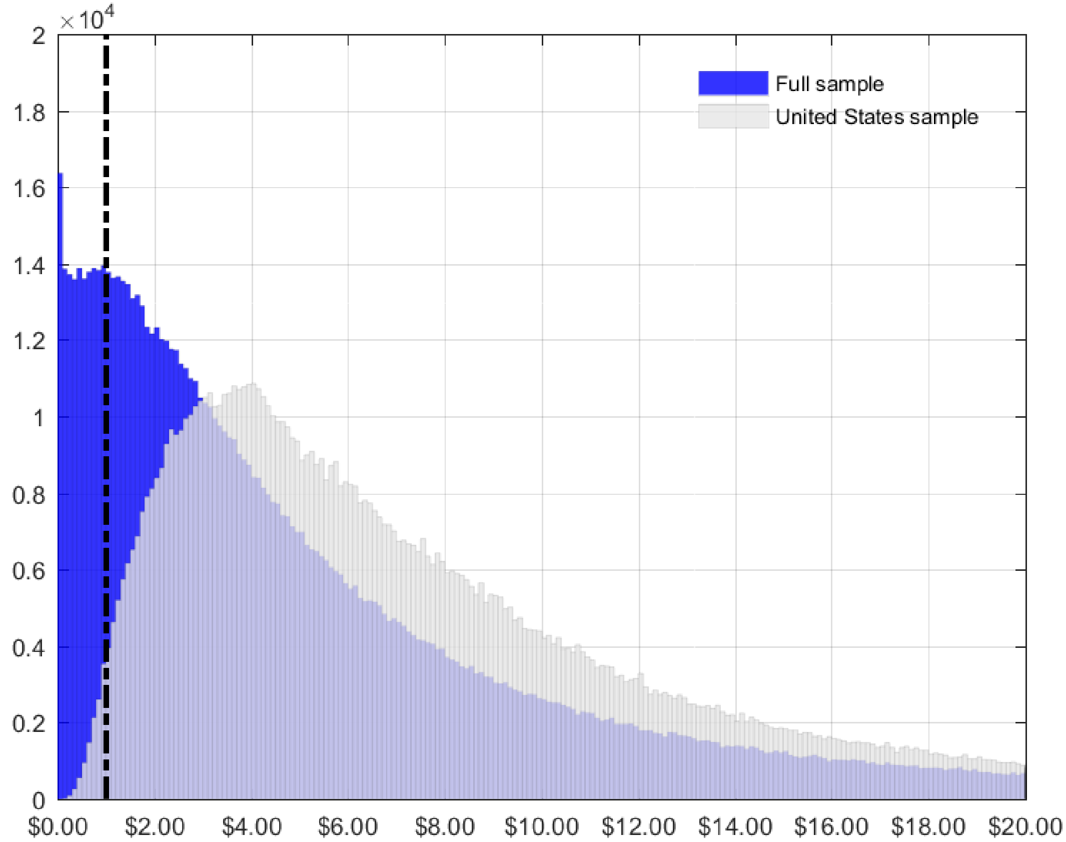

Fig. IA5. Cumulative 30-year payoffs for global USD investors. The figure shows histograms of real payoffs across 1,000,000 bootstrap simulations at a return horizon of 30 years. The real payoffs are from the perspective of a global USD investor. The underlying sample for the simulated returns is the pooled sample of all developed countries (blue) or the United States sample (gray). The dashed line separates the regions of real loss and gain on a $1.00 initial investment. The bootstrap sampling procedure is based on the stationary bootstrap approach of Politis and Romano (1994b) as described in the text. We sample blocks of random length, where the length of each block has a geometric distribution with a mean of 120 months.