Yale University

EliScholar – A Digital Platform for Scholarly Publishing at Yale

Cowles Foundation Discussion Papers | Cowles Foundation

6-1-2014

Changing Times, Changing Values: A Historical Analysis of Sectors within the US Stock Market 1872-2013

Oliver D. Bunn

Robert J. Shiller

Follow this and additional works at: https://elischolar.library.yale.edu/cowles-discussion-paper-series

 Part of the Economics Commons

Recommended Citation

Bunn, Oliver D. and Shiller, Robert J., "Changing Times, Changing Values: A Historical Analysis of Sectors within the US Stock Market 1872-2013" (2014). *Cowles Foundation Discussion Papers*. 2351.
https://elischolar.library.yale.edu/cowles-discussion-paper-series/2351

This Discussion Paper is brought to you for free and open access by the Cowles Foundation at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Cowles Foundation Discussion Papers by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.

CHANGING TIMES, CHANGING VALUES: A HISTORICAL ANALYSIS OF SECTORS WITHIN THE US STOCK MARKET 1872-2013

By

Oliver D. Bunn and Robert J. Shiller

June 2014

COWLES FOUNDATION DISCUSSION PAPER NO. 1950



COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Changing Times, Changing Values: A Historical Analysis of Sectors within the US Stock Market 1872-2013*

Oliver D. Bunn† and Robert J. Shiller‡

WORKING PAPER
May 2014

Abstract

We construct a price, dividend, and earnings series for the Industrials sector, the Utilities sector, and the Railroads sector from the beginning of the 1870s until the beginning of the year 2013 from primary sources. To infer about mispricings in the sector markets over more than a century, we investigate the forecasting power of the Cyclically Adjusted Price-Earnings (CAPE) ratio[1] for these sectors. With regard to the CAPE ratio, which has originally been devised and employed by Campbell and Shiller (1988, 1998, 2001) as well as Shiller (2005), we define a methodological improvement to this ratio to not only be robust to inflationary changes, but also to changes in corporate payout policy. We then update the original evidence from Campbell and Shiller (1998, 2001) of the return predictability of the CAPE ratio for the overall stock market and furthermore extend this evidence to the three aforementioned sectors individually. Whereas this part of our analysis focuses on each sector of the US economy in isolation, we subsequently construct an indicator from the CAPE ratio that enables us to perform valuation comparisons across sectors. In addition to establishing the prediction of subsequent return differences based on differences in the CAPE-based valuation indicator, we also suggest a hypothetical, historical, and simple value investment strategy that rotates between the three sectors based on the valuation signals derived from the CAPE-based indicator, generating slightly more than 1.09% annualized, inflation-adjusted excess total return over the market benchmark during a period of nearly 110 years.

Keywords:
Cyclically adjusted price-earnings ratio, historic valuation of sectors of US economy, long-term predictability
JEL Codes:
E37, G11, G14, G17, N20

*We thank Barclays Bank PLC for partial financial support of this project. There exists a collaboration with Barclays Bank PLC on ways through which the Cyclically Adjusted Price Earnings (CAPE) ratio, a valuation measure that also plays a prominent role in this paper, may be implemented in an investment. The views herein are those of the authors', and do not necessarily reflect the views of Barclays Bank PLC.

†Barclays Bank PLC. Contact: oliver.d.bunn@gmail.com. The research for this paper was conducted and the paper was written while the author was a graduate student at the Department of Economics at Yale University.

‡Department of Economics, Cowles Foundation for Research in Economics, and International Center for Finance, Yale University. Contact: robert.shiller@yale.edu.

[1]The Cyclically Adjusted Price-Earnings Ratio (CAPE) is also known as the Campbell-Shiller PE(10) Ratio or as the Shiller(10) Ratio.

1 Introduction

1.1 The Cyclically Adjusted Price-Earnings Ratio and Price Fluctuations

There is a substantial literature on the value factor in predicting returns, but this literature has not yet exploited the historically-longest available stock price series, sector indices. Studying the very-long-term value effect with these series allows us to test the value proposition out of its original sample, where out of sample means extending the sample back in time, as well as forward. The long time series enables us to evaluate a very effective measure of mispricing relative to fundamentals, the Cyclically Adjusted Price-Earnings (CAPE) ratio,[2] originally shown to be highly effective in predicting returns in the aggregate US stock market by Campbell and Shiller (1988), who did not, however, look at sector data. Sector data, in contrast to firm-level data, is especially interesting because the sectors represent major, almost macroeconomic, activities, and much analysis, accordingly probably also much mispricing, is sector-related. It is also interesting since we can accurately evaluate a sector's valuation over a longer historical period than has previously been studied in the literature.

Subsequent to Campbell and Shiller (1988), the CAPE ratio has been employed by Campbell and Shiller (1998), Campbell and Shiller (2001), and Shiller (2005). In the analyses since the late 1990s, it has been defined as the ratio of real (inflation-corrected) per-share price divided by a long average of real earnings per share.[3] Our present analysis modifies the definition of the CAPE ratio from Campbell and Shiller (1998), Campbell and Shiller (2001), and Shiller (2005) and considers the ratio real total return per-share price divided by a long average of real scaled earnings per share, where the scaling of earnings takes into account the transition from price return to total return numbers.

In the original analyses, it was concluded that when the average of earnings is taken over the past ten years, the ratio is a significant forecaster of long-term, ten-year stock market returns. When the ratio has been high (i.e., stocks are overpriced) prices tend to come down, not necessarily in the next year or two, but some time over the next ten years, and conversely when the ratio is low. The CAPE ratio differs from the conventional price earnings ratio in that earnings are measured over a much longer interval, and because of the longer-term view they are necessarily inflation-adjusted. The longer interval appears justified, since the earnings in any given year tend to be noisy and influenced by the business cycle. The idea of using price relative to a long average of earnings was not new with Campbell and Shiller (though, it appears that the inflation adjustment was new and important). As long ago as 1934, Benjamin Graham and David Dodd advocated a ratio that is an analogue to CAPE in their classic book Graham and Dodd (1934):

> "A conservative valuation of a stock issue must bear a reasonable relation to average earnings. In addition it must be justified by whatever indications are available as to the future. This approach shifts the original point of departure, or basis of computation, from the current earnings to the average earnings, which should cover a period of not less than five years, and preferably seven to ten years."[4]

The Graham and Dodd book was very widely read and influential, and it is reputed to have inspired a young Warren Buffett into his subsequent stellar investing career. But this ratio, which Graham and Dodd used in the inverse form and called "average earnings on common stock price" was never much picked up. Little mention of it is made in subsequent discussions. It is true that Barron's published a "ratio price to five-year average earnings" for its Barron's 50-Stock Average[5] since the 1950s, but that ratio seems not to have been much noticed, at least in media accounts. The idea of examining a long average of earnings was not original with Graham and Dodd either, as it goes back another couple of decades. In 1911, the Wall Street Journal[6] published an "Index of Railroad Earnings Covering a Ten-Year Period" and suggested that the stock price should be evaluated using this index. Certainly, intuitive analysts looking at earnings data will quickly recognize that any single year's earnings is very noisy, and one needs

[2] The CAPE ratio described here and commonly quoted should not be confused with the "cyclically adjusted price-earnings ratio" of Smithers and Wright (2002), a completely different ratio that seems to have faded from view. Even in that book, the central ratio was the so called "q ratio."

[3] Campbell and Shiller (1988) defined the CAPE ratio as the arithmetic long average of the natural logarithm of real earnings per share minus the natural logarithm of real per-share price.

[4] See Graham and Dodd (1934), page 459.

[5] For example, see Barron's (2012).

[6] See N.N. (1911).

to look at the longer history of earnings. And yet the price earnings ratio that is most commonly quoted is just the price divided by 12-month trailing earnings.

Why has the price-earnings ratio been so durable, despite the fact that such personages as Graham and Dodd dismiss it? The reason may have to do with the fact that the ratio is not always available at the firm level. Many firms are less than ten years old. Because it cannot always be provided, standard printouts of characteristics of individual stocks are likely to omit such a ratio. Moreover, at the firm level, we have to recognize that there may often be special factors that limit the meaning of earnings of as far as ten years ago. The most obvious example is for firms that are still young, even if more than ten years old, and their earnings show a growth trajectory. A CAPE ratio for an individual firm cannot have the same universality, and so it loses the currency of the conventional CAPE. A CAPE ratio for an individual stock seems to need more judgment to interpret, and hence seems to recommend itself only to professional analysts, who often like to keep their important indicators to themselves. In Shiller (2012a), it was argued that investors show an "impulse for conventionality and familiarity." The conventional price-earnings ratio, based on just a year's earnings, is everywhere, and so, by this impulse, may be seen as more trustworthy.

1.2 Historic Sector Rotation Based on the Cyclically Adjusted Price-Earnings Ratio

In the subsequent analysis, we consider the CAPE ratio on the level of sectors, instead of the firm level. Naturally, we face the question about applying the CAPE ratio to sector indices of stock prices (along with other fundamentals), since the stocks that comprise the indices change through time. There is a risk that the earnings of as far as ten years ago might not even be representative of earnings of the companies that are now in the index. However, we take the view of an investor who would like to determine its exposure to certain sectors and for that purpose would like to make use of a long-term valuation ratio like the CAPE ratio, as applied to individual sectors. Therefore, the composition of the underlying sector is only a subordinated concern for this investor, secondary to more global trends of a subset of the overall stock market.

Looking at individual sectors substantially extends the original purpose of the CAPE ratio as initially defined by Campbell and Shiller (1988) as a means to determine the over- and undervaluation of the stock market as a whole. For this purpose, we define an extension of the CAPE ratio for individual sectors, the Relative CAPE indicator, which normalizes the CAPE ratio of a sector relative to its own long-term history and therefore enables comparisons of the over- and undervaluation signals of the CAPE ratio across sectors.

In our long-term analysis of the CAPE ratio when applied to individual sectors, we consider data from the early 1870s until the end of the year 2012 for the sectors "Industrials", "Utilities", and "Railroads" in the United States. In addition, we also replicate our construction for the overall stock market to obtain a market benchmark for reference purposes. We merge data from Cowles (1939) with data from the Security Price Index Record publications by the Standard & Poor's Statistical Service,[7] whereby we supplement the Security Price Index Record's information with data from the Analysts' Handbook, which is also published by the Standard & Poor's Statistical Service.[8] Our goal has been the construction of a US-only sector-specific time series for prices, dividends, and earnings that reaches as far back in time as possible. This has naturally led us to the three sectors Industrials, Utilities, and Railroads because of the availability of data about them in the early phase of our sample. Subsequently, we have aimed to trace exactly these three sectors, in order to be able to investigate the effectiveness of the CAPE ratio (or the Relative CAPE indicator) as a long-term valuation measure over a very long period of time.

Analogous to the findings of Campbell and Shiller (1988, 1998, 2001) for the overall stock market, we initially extend the evidence in favor of the predictive ability of the CAPE ratio for long-term returns to the sector level. Subsequently, we turn our attention to the ability of the Relative CAPE indicator to make predictions for return differences between sectors. Lastly, we translate the predictive evidence about the CAPE ratio and the Relative CAPE indicator into a hypothetical investment strategy. This strategy rotates between Industrials, Utilities, and

[7] See Standard and Poor's Statistical Service (1978-2010). The information contained in the multiple volumes of this publication that we use for the construction of our data has originally been collected by Standard Statistics, which in 1941 merges with Poor's Publishing to form the Standard & Poor's Corporation. For brevity, we will refer to "Standard & Poor's" as publishers of the Security Price Index Record for the entire time period over which we use this data.

[8] See Standard and Poor's Statistical Service (1982-2012). The same disclaimer related to the use of the term "Standard & Poor's" as in footnote 7 applies.

Railroads according to signals obtained from the Relative CAPE indicator, which is complemented with a momentum signal that we design to be of subordinated importance to the value signal provided by CAPE. We find that the real (inflation-adjusted) total return of this deliberately simplistic strategy that only rotates between three different assets beats the real total return on the Standard & Poor's Composite Index in a time span between October 1902 until the beginning of the first quarter of 2013 by an annualized 109 basis points or 1.09%.[9] In other words, whereas the benchmark turns $1 (in inflation-adjusted terms) in October 1902 into nearly $569 (in inflation-adjusted term) in January 2013, our strategy results in approximately $1,756 (again adjusted for inflation).

2 Construction of Price, Dividend, and Earnings Data

2.1 Overview of the Employed Time Series

The process to create the time series of prices, dividends, and earnings for the three sectors as well as for the overall market is generally identical, which is why we will describe it in a generic manner. We aim to construct four different time series, an end-of-quarter price index, a beginning-of-quarter price index, a dividend series, and an earnings series. The following description outlines the key characteristics of our data, whereas we defer a detailed analysis and discussion of the construction to Appendix A.[10]

We use data from Cowles (1939) for the early part of our sample. We then transition to the Security Price Index Record published by Standard & Poor's Statistical Service[11] for prices and earnings after the end of 1925, whereas the dividend data from Cowles (1939) is in use until the end of 1934.[12]. The Security Price Index Record is in use for our price series until its latest current volume, which is the 2010 volume. Price data for the last few years of the sample are from Bloomberg. Because the Security Price Record has ceased the publication of earnings and dividend information for the Railroads sector in 1982 and for the overall stock market as well as the Industrials and the Utilties sector in the third quarter of 2001, we need to switch to an accompanying Standard & Poor's publication, the Analysts's Handbook.[13]

The constructed price indices have quarterly frequency and are designed to reflect the price level of each sector[14] in the beginning and at the end of each quarter. The end-of-quarter series runs from the end of the fourth quarter of 1872 until the end of fourth quarter of 2012, whereas the beginning-of-quarter series starts in the beginning of the first quarter of 1873 and ends in the beginning of the first quarter of 2013.[15] Starting in 1930 for Railroads and in 1932 for the other sectors, the end-of-quarter series refers to the index value at the end of the last trading day of the month. In comparison, the beginning-of-quarter series uses the end of the first trading day of the quarter. Before 1930 and 1932, respectively, we use approximations for the end and the beginning of the quarter that we describe in detail in Appendix A. We normalize all price, dividends, and earnings data such that the nominal price index at the end of the fourth quarter of 1872 equals $1.

In contrast to the two price series, the dividend series measures a payment flow that has accrued over a specific period of time. It also has quarterly frequency and, at the end of each quarter, it expresses the amount of dividends that an investor will have received from a sector during this quarter. The earnings series shares the notion with the dividend series of measuring a flow variable. However, each observation of the earnings series captures a 12-month

[9] The volatility performance of the two strategies is less than 0.5% apart from each other, i.e. the volatility of the two strategies is nearly identical.

[10] Global Financial Data, see Global Financial Data (2012), contains similar data series to the one that we construct from primary sources. In order to have full control over the conventions used in the construction of the data, we have decided to start from the primary sources, particularly because the documentation of the data on Global Financial Data has not provided sufficient detail to be able to comprehend their data construction.

[11] See Standard and Poor's Statistical Service (1978-2010).

[12] For an elaboration on our motives behind the choice of these transition dates, please consult the detailed data description in Appendix A.

[13] See Standard and Poor's Statistical Service (1982-2012).

[14] As already mentioned above, we will construct both price series, the dividend series, and the earnings series for Industrials, Utilities, and Railroads as well as for the overall stock market, which will be represented by the S&P Composite Index. For brevity, we will refer to the three sectors in conjunction with the overall market as "sectors" in the description of the construction of the data.

[15] Technically, our price data starts in the beginning of January 1871. However, the first earnings observation that is uniformly available across all sectors under consideration is for December 1872, which we have therefore chosen as the overall beginning of our sample.

trailing sum of earnings for a sector. Its observations are annual from 1872 until 1934 and switch to quarterly frequency starting with the end of the first quarter of 1935.[16]

An important feature of our data is the lack of interpolated data for dividends and earnings.[17] Dividend data at the end of each quarter has actually occurred during this quarter and earnings information refers to the actual earnings that have accumulated in the past twelve months. Particularly for the kind of predictability analysis that we perform later on, it is crucial not to rely on interpolated data. Moreover, another crucial feature of our data is an emphasis on spot information for the price indices.[18]

Because more precise information is not available otherwise, the monthly price data from Cowles (1939) reports the arithmetic average of a month's high and low price. In contrast, the data from Standard and Poor's Statistical Service (1978-2010) report actual index levels at the end of a trading day. We therefore switch from the averaged data in Cowles (1939) to the actual spot information in Standard and Poor's Statistical Service (1978-2010) at the earliest possible date, which is 1926, to minimize the impact of the data averaging for our later predictability analysis.[19]

2.2 Construction of Total Return Series

The purpose of the dividend series is to turn the price indices (end-of-quarter and beginning-of-quarter) of a sector into total return indices. These indexes represent the value of a portfolio, started at \$1 in the case of the end-of-quarter series,[20] reinvesting all dividends into the same sector. Or, alternatively, the indexes may be thought of as showing the value through time of a share in the sector if it had used share repurchase instead of dividends to make the same cash flow to investors as it in fact did make, so that the investor got exactly the same cash flow, but called share repurchase rather than dividends, and saw the number of shares owned decline through time. Following Miller and Modigliani (1961), we see that the total return index is a valid way of representing the value of a stock, though the index will tend to grow faster than does the size of the business activities. Using total return indices for our analysis eliminates any anomalies due to changes in payout ratios to earnings over the long time interval of our data.

Explicitly, we construct the total return index for the end-of-quarter series as follows: We set the index to \$1 at the end of the fourth quarter of 1872, which is the value of the end-of-quarter price index at this instance. Then, we obtain each value of the total return index at the end of quarter $t+1$ by multiplying the total return index at the end of quarter t by the ratio

$$\frac{P_{t+1}+D_{t+1}}{P_t},$$

where P_t and P_{t+1} denote the level of the original price index at the end of quarters t and $t+1$ respectively and D_{t+1} denotes the dividend series at the end of quarter $t+1$, which stands for the payments that accrue during quarter $t+1$. In other words, we reinvest dividends quarterly and perform the hypothetical reinvestment consideration at the end of the quarter for all dividends that have accrued during the quarter.

The procedure for the beginning-of-quarter series is very similar, although it requires some clarification with regard to the timing of the dividend stream. As we construct the beginning-of-quarter price index to reflect the price of the sector in the very beginning of the quarter, the time period covered between two consecutive beginning-of-quarter observations is, up to minimal differences, identical to the time period covered between the end of the previous quarter and the end of the current quarter. Therefore, we set the beginning-of-quarter total return series index to the level of the price index in the beginning of the first quarter of 1873 and obtain each value of the total return index in the beginning of quarter $t+1$ by multiplying the total return index in the beginning of quarter t by the ratio

[16] Despite the switch in frequency for the earnings series, it is important to keep in mind that the convention of the 12-month trailing sum remains unchanged.

[17] This is in contrast to the online data to be found at Shiller (2012b), which are updated numbers from the data used in Shiller (2005).

[18] This is again in contrast to the online data to be found at Shiller (2012b). That price series contains the average of the month's high and low price for the part of the data that comes from Cowles (1939) and the average of all prices recorded on trading days throughout a month in the later part of the sample.

[19] See Working (1960) for the impact of averaging on the autocorrelation of a time series.

[20] The beginning-of-quarter series starts at a minimally different value reflecting price appreciation or depreciation between the end of the fourth quarter of 1872 and the beginning of the first quarter of 1873.

$$\frac{P'_{t+1} + D_t}{P'_t},$$

where P'_t and P'_{t+1} denote the level of the original price index in the beginning of quarters t and $t+1$ respectively and D_t denotes the dividend series at the end of quarter t, which stands for the payments that accrue during quarter t. This procedure preserves the convention of reinvesting dividends quarterly at the end of the time period, during which they accrue.

2.3 Inflation Adjustment

Because of the length of the time period that we consider, it is necessary to adjust each time series that we use in our subsequent analysis, i.e. the price indices and the total return series as well as the earnings series, for inflation. In order to do so, we use a price series that has already been employed in Shiller (2005)[21] to adjust data spanning an analogous time period for the impact of inflation. Starting in 1913, the price series is the Consumer Price Index for All Urban Consumers (CPI-U), which is published by the US Bureau of Labor Statistics. Before 1913, data comes from a producer price index in Warren and Pearson (1935).[22]

Importantly, at any moment in time, we would like to avoid using data that has not been made available at that moment in time. As the time series that we use for inflation-adjustment will naturally be published with a lag, we have decided to impose a lag in our inflation-adjustment. That is, in order to adjust any value from any of our price indices, total return series and earnings series that is listed in a particular month, irrespective of whether the time series refers to the beginning or the end of the quarter, we lag the inflation-adjustment series by three months. For example, a number listed in December 1955 will be adjusted by the CPI-U from September 1955. Finally, the common price level, to which we refer when we use the term "real" is September 2012.

2.4 Logarithmic Convention

In all our subsequent plots, we employ a logarithmic scale, with e as its base value, a scale that gives equal vertical axis intervals to equal percentage changes. This change is another tribute to the length of our sample, which covers well more than a century. The logarithmic scale represents economic growth more evenly throughout the time period that we consider, as it avoids skewed impressions about the long-term behavior of time series, as these differences may potentially be due to century-long compounding of small differences early in the sample.

3 Historic Sector Classification of the US Economy

3.1 Standard & Poor's Sector Classifications

When constructing our sector dataset, our goal has been to use a sector classification that is not only consistent throughout a long period of time, but also has a long history of publicly reported data on prices, dividends, and earnings. This has naturally led us to the distinction between Industrials, Utilities, and Railroads, a classification employed by both Cowles (1939) and during a substantial part of the period covered by the Security Price Index Record. Overall, this classification persisted from the beginning of the 1870s until 1982. In 1982, Standard & Poor's decided to cease their reporting for Railroads as one of the main sectors of the US economy, as represented by the S&P 500.[23] Instead, S&P commenced to report prices, dividends, and earnings for Transportation, of which Railroads are a part, and Financials.[24] Another change in the sector classification system used by S&P came in August 1999, when

[21] See also the online data section at Shiller (2012b).

[22] See Table 1 on pages 11-14 in Warren and Pearson (1935).

[23] March 1957 is the inception date for the S&P500. Before that, Cowles (1939) and Standard & Poor's report an S&P Composite Index that consists of less than 500 stocks.

[24] See, for example, the information provided in the sector-specific data tables and the introductory history section in the 2004 volume of Standard and Poor's Statistical Service (1978-2010).

S&P, in collaboration with Morgan Stanley Capital International (MSCI), issued the Global Industry Classification Standard (GICS). On the top-level, the so-called sector-level, this classification system consists of ten sectors. Energy, Materials, Consumer Discretionary, Consumer Staples, Health Care, Information Technology, and Telecommunication Services, which were previously part of what was referred to as Industrials, are now considered separate sectors. The remainder of the previous Industrials sector is now paired with the former Transportation sector to form the GICS sector Industrials. Finally, Utilities and Financials have kept their status as separate sectors.[25]

Instead of reflecting the most up-to-date sector classification, our goal when constructing our historic sector dataset has been to trace the performance of the original three sectors Industrials, Utilities, and Railroads through time and provide a consistent and very long-term assessment of these sectors' performance for nearly one and a half centuries. This is why we disregard the introduction of Transportation and Financials as novel top-level distinctions of the US economy in 1982[26] and the inception of the GICS classification in 1999 and preserve the distinction between Industrials, Utilities, and Railroads. Given the novelty of the GICS classification, barely enough time has passed in order to compute a few years worth of CAPE ratios from publicly available data for the sectors according to this classification, which renders a meanignful long-term analysis of the CAPE ratio as a valuation tool infeasible.

Instead of tracing a sector classification through time while relying on publicly available aggregate data, as it is done in this paper, another feasible approach is to start from a recent and up-to-date sector classification, as for example the GICS classification, and extend this classification as far back in time as possible by constructing the appropriate time series from disaggregate information. See Ural et al. (2012) for a construction of GICS-based sector indices within the universe of the S&P500 from individual-firm data, which are then analyzed using the CAPE ratio. Naturally, the limited availability of data leads to a much shorter sample period to analyze, e.g. CAPE ratios in Ural et al. (2012) only cover approximately 30 years compared to the almost 130 years in this paper.

3.2 Universe of Stocks for Sector Indices

An important consideration relates to the universe of stocks, from which the sector-specific information about Industrials, Utilities, and Railroads is derived. Cowles (1939) covers the time period from the beginning of the 1870s until well into the middle of the first half of the twentieth century. Wilson and Jones (1987) in their table A.1[27] provide information about the universe, within which Cowles (1939) reports the price, dividend, and earnings information for the overall stock market as well as for Industrials, Utilities, and Railroads. Whereas the universe only consists of about 50 stocks in the early 1870s, decomposed into about two thirds of Railroad stocks, a tenth of Utilities, and the remainder in Industrials, the stock universe increases to nearly 400 stocks by the middle of the first half of the twentieth century. Both Railroads and Utilities now make up slightly less than a tenth of the universe each, with Industrials representing the remaining portion.[28]

With regard to the transition from the data in Cowles (1939) to the data from the Security Price Index Record, Cowles (1939) explicitly describes in Sections 2 and 4 that the index information is designed to be merged with the data from the Security Price Index Record, however with what he refers to as the "weekly indices". Whereas these weekly indices cover a universe of approximately 400 stocks during the later part of the first half of the 20th century, Standard Statistics also publishes indices based on a so-called "daily series".[29] The universe for this daily index data is only 90 stocks, 50 Industrials, 20 Utilities, and 20 Railroads.

Standard & Poor's has actually later standardized all their indices to reflect the information from the daily series, and therefore has eliminated the weekly indices. This procedural overhaul only affects the early data from the Security Price Index Record, because the universe has been standardized to 500 stocks after the inception of the S&P500 in 1957. This shift is something that has been explicitly mentioned in the methodological section of the Security Price

[25] For a current overview of the GICS classification system, see Standard and Poor's (2012).

[26] Whereas the subsequent analysis of the main part focuses on consistency and and a long-term perspective on sectors of the US economy, Appendix D contains an illustration of the hypothetical investment strategy, which will be described later, if one indeed eliminates the Railroads sector as an independent investment and introduces Transportation and Financials instead.

[27] This table summarizes information contained in Appendix II of Cowles (1939).

[28] For the remainder of this section, it is important to keep in mind that, although the number of stocks can give an indication of the relative importance of sectors, the relevant quantity to compare would be market capitalization. However, neither Cowles (1939) nor Standard and Poor's Statistical Service (1978-2010) contain this information.

[29] See also the discussion in Wilson and Jones (2002) about the very same matter.

Index Record until the volume for 1992, but has been edited in later volumes. In summary, whereas Cowles (1939) considers a universe of approximately 400 stocks with the explicit intention to merge the data with the weekly index data from the Security Price Index Record, our information from Standard & Poor's only reflects 90 stocks until March 1957.

We would like to make two comments about this drop in the size of the universe under consideration. First, whereas the drop from 400 to 90 might sound big, the drop in market capitalization reflected by the indices will for sure have been less dramatic. Along the same lines, the 90 stocks included in the daily series have been chosen with the explicit intention to reflect the US stock market to the closest degree possible, which is why a potential lack of representativeness arising from the decreased stock universe is not too much of a concern. Moreover, it becomes apparent from studying the historic precedents of the Security Price Index Record during parts of the relevant time period[30] that the dividend and earnings information contained in these volumes actually refers to the 90-stock daily series and not to 400-stock weekly indices. Taking into account that we need prices in connection to dividends and earnings, this actually means that we would not have had a choice anyway but to refer to the 90-stock series.

After 1957, all series build on a universe of those 500 stocks that are part of the S&P 500.[31] Initially, the composition was fixed to 425 Industrials, 60 Utilities, and 15 Railroads, which was later adapted to 400 Industrials, 40 Utilities, 20 Transportation (of which Railroads are a part), and 40 Financials. The fixed breakdown of the 500 stocks into sectors was abandoned in April 1988, when S&P adopted a flexible structure of the sectors within the S&P500.

Our focus on sectors within the US economy is the major distinguishing element to the data under consideration in Ibbotson and Sinquefield (1976), Wilson and Jones (1987), Schwert (1990), Goetzmann et al. (2000), and Wilson and Jones (2002), which all exclusively focus on the overall US stock market, and also from the analysis in Jorion and Goetzmann (1999), who look at overall stock markets on an international level.

3.3 Industrials, Utilities, and Railroads through Time

The three sectors at the start of the sample represent three quite distinct economic activities in the United States. At the beginning of our sample, the large conglomerate corporation, in many lines of business, was yet to come, so companies could be neatly divided into separate industries. The clarity of this division however diminishes somewhat through time.

Railroads are the best defined of the three, a single kind of transportation service that until the 1920s dominated land transportation, and dominated the stock market as well. Today it is still a clearly defined consistent industry, though it has lost some of its fundamental elements, notably passenger transportation (taken over in 1971 by the U.S. National Railroad Passenger Corporation, doing business as Amtrak, which is de facto practically a government agency) and mail service, also largely discontinued after 1971.[32] Their relative importance within the overall stock market has also declined over time. As outlined above, whereas two-thirds of the stocks of the S&P Composite Index were Railroads in the beginning of the 1870s, only 15 out of 500 stocks were Railroads when the S&P 500 index started in 1957. Nowadays, it is a level-four GICS classification, a so-called sub-industry. Its index consists of four stocks, Burlington Northern Santa Fe Company, CSX Corporation, Norfolk Southern Corporation, and Union Pacific Corporation.[33]

Utilities define another quite distinct sector. According to the Cowles classification, before 1940, the Utilities sector represented electric, gas, water and heat operating companies, as well as telephone and telegraph, and traction and motor transportation. To a first approximation, this sector represented basic things that one needs to get arrangements for, and to be billed regularly for, when one moves to a new home, and local transportation to and from the home. As such, it has, in recent years, gained the reputation as a boring, stable sector, but it was not always

[30] See Standard Trade and Securities Service (1938-1950).

[31] For the following description, see, for example, the History section of the 2004 volume of Standard and Poor's Statistical Service (1978-2010).

[32] Some U. S. Postal Service use of railways was begun again in the 1980s, and the U.S. Postal Service has participated in container shipments as that industry has developed. In 1993 the industry introduced RoadRailer intermodal equipment that could travel both on highways and railroads. Railroad station post offices were also eliminated by 1977. See United States Postal Service (2012).

[33] See Standard and Poor's (2012) for a detailed description of the GICS codes of the constituents of the S&P500.

so boring. According to GICS, which lists Utilities as one of the ten top-level sectors, Utilities have kept their status as an important part of the US Economy. It currently consists of Electric Utilities, Gas Utilities, Multi-Utilities, Water Utilities, as well as Independent Power Producers and Energy Traders. The sector has therefore preserved its core activities throughout the entire sample period of nearly a century and a half. It has actually become slightly more focused, as its telecommunication and transportation constituents have migrated into separate sectors.

Industrials, in Cowles classification covering the beginning of our sample, were the catchall sector that included advertising, agricultural machinery, airplane transportation, apparel, automobiles and trucks and their parts and accessories, beverages, building equipment, chemicals, coal, containers, copper and brass, cotton and cotton goods, drugs and cosmetics, electrical equipment, fertilizer, food products, gold mining, household products and supplies, lead and zinc, leather, machinery and machine equipment, meat packing, mining and smelting, office and business equipment, oil producing and refining, paper and paper products, radio, phonograph and musical instruments, rayon, retail trade, shipping and shipbuilding, shoes, silk and silk goods, steel and iron, sugar, theatres and motion pictures, tobacco, and wool and woolen goods. It quickly grew from representing less then 20 stocks in the Composite Index in the beginning of the 1870s to representing 425 out of 500 stocks at the inception date of the S&P500, a number that S&P later reduced to 400. Until the creation of the GICS codes in 1999, Industrials were still listed as a single sector. As outlined above, the Industrials sector was split into nearly eight sectors,[34] covering such diverse companies like Black & Decker, Nike Incorporated, and Polo Ralph Lauren in the newly minted Consumer Discretionary sector, eBay Incorporated, Google Incorporated, and Visa Incorporated in the novel Information Technology sector, as well as Boeing, General Electric Company, and Caterpillar Incorporated in the GICS Industrials sector.[35] Our continuation of what was referred to as Industrials is therefore not the GICS Industrials sector, but the so-called "Industrials Composite Index", which is a similar catchall of diverse activities, consisting of all stocks in the S&P500, except for those that are classified as Utilities, Transportation, or Financials.

In Figure 1, we display the price indices[36] for the overall stock market as well as for the three sectors, both in nominal and real terms. As described at the end of Section 2, we employ a logarithmic scale throughout, as a tribute to the length of our sample. However, in order to provide a better sense of the performance of the price indices, we will briefly describe the performance of the different indices over time in non-logarithmic fashion.[37] The nominal index capturing the overall stock market starts at $1 in 1872 and ends up with almost $256, i.e. it increases nearly 256-fold. The real index for the overall stock market ends with the same value by construction.[38]

It starts however at slightly less than $18, which is the inflation-adjusted equivalent of $1 in 1872, reflecting a more than 14-fold increase in real terms. The performance of Industrials is even more impressive than that of the overall stock market. Its nominal price index increases from $1 in 1872 to slightly less than $671 at the end of the fourth quarter in 2012, which translates into an almost 38-fold increase in real terms. In contrast, the performance of Utilities is substantially worse over the entire time period. $1 in 1872 turns into just less than $16 at the end of 2012, which actually turns out to be a decline of nearly 11% in real terms. Railroads, as the final sector that we consider and display in Figure 1 increases about 167-fold in nominal terms and a little bit more than 9-fold in real terms.

In comparison to Figure 1, Figure 2 displays earnings, the second key component of our long-term valuation measure CAPE, both in real and in nominal terms. Earnings growth varies between a more than 65% decrease in inflation-adjusted terms for Utilties and a more than 8-fold, almost 12-fold, and approximately 11-fold inflation-adjusted increase for the Railroads sector, for the Industrials sector, and the overall stock market, respectively. It is striking that the earnings growth numbers are roughly comparable to the increases in the index levels and that no sector ever had negative 12-month trailing earnings. The absence of negative 12-month earnings is particularly striking for the Railroads sector; this must seem strange to those who remember the many troubles and bankruptcies

[34] Again, see Standard and Poor's (2012) for more detailed information about the GICS codes for the constituents of the S&P500.

[35] As outlined above, the GICS sector Industrials consists of some companies that were classified as Industrials before the advent of the GICS codes together with the old Transportation sector.

[36] Note that we use the term "price indices" to refer to indices that do not reinvest dividends. This is in contrast to "total return indices", which are scaled price indices to reflect dividend reinvestment.

[37] The following description centers on the evolution of the price indices, without taking into account reinvested dividends. We will however take a look at the total return indices below.

[38] We use March 2012 as the base month for the inflation adjustment. Because we lag the price level by three months, this is the price level used for the inflation adjustment of the observation at the end of the second quarter of 2012, which is the end of our sample.

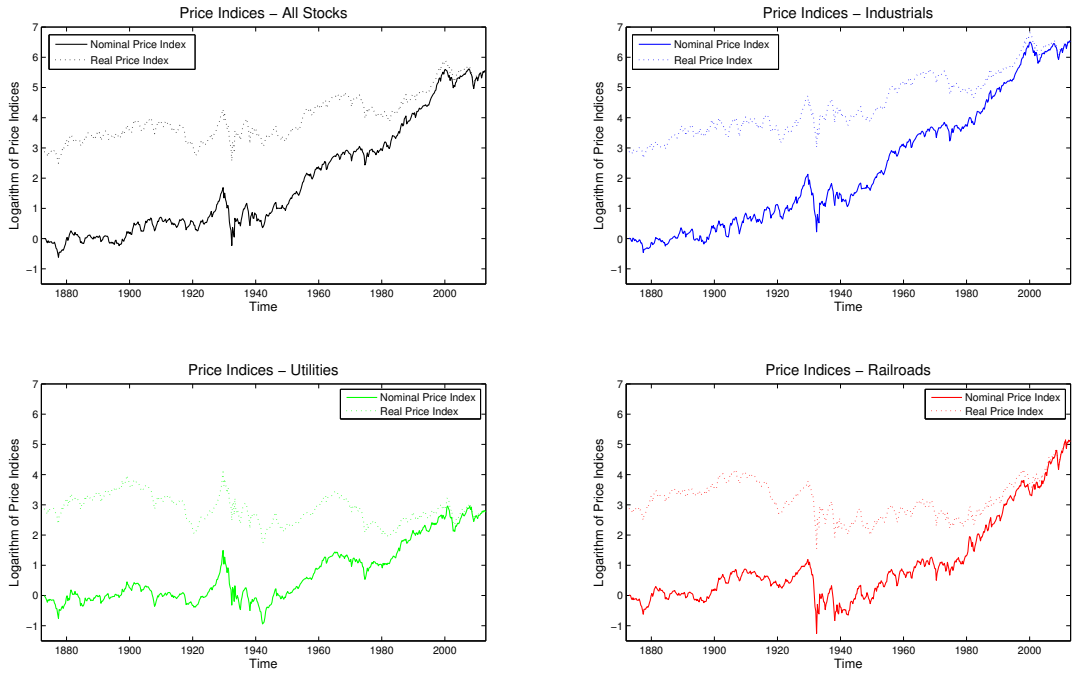


Figure 1: This figure depicts price indices for the overall US stock market (All Stocks), Industrials, Utilities, and Railroads, not reinvesting dividends. The data is quarterly with the first observation occuring at the end of the fourth quarter of 1872 and the last observation at the end of the fourth quarter of 2012. Nominal Data between 1872 and 1925 is taken from series P in the last month of a quarter in Cowles (1939). Between 1926 and 1934, the quarterly price series refer to the level reported by the Security Price Index Record on the last Wednesday of a quarter. Starting in 1935, the price index at the end of a quarter refers to the last trading day of a quarter, again as reported by the Security Price Index Record and complemented by information from the Analysts' Handbook in the case of Railroads after 1982. Inflation-adjustment uses the spliced CPI series, as described in Section 2, with September 2012 as the base month. A nominal number in a given month (end of a quarter) is adjusted using the level of the CPI three months prior, aiming to ensure that the relevant CPI number is known in the month, for which the adjustment occurs. All nominal series are normalized to 1 in the fourth quarter of 1872. Both the nominal and the real series are displayed on a logarithmic scale (natural logarithm).

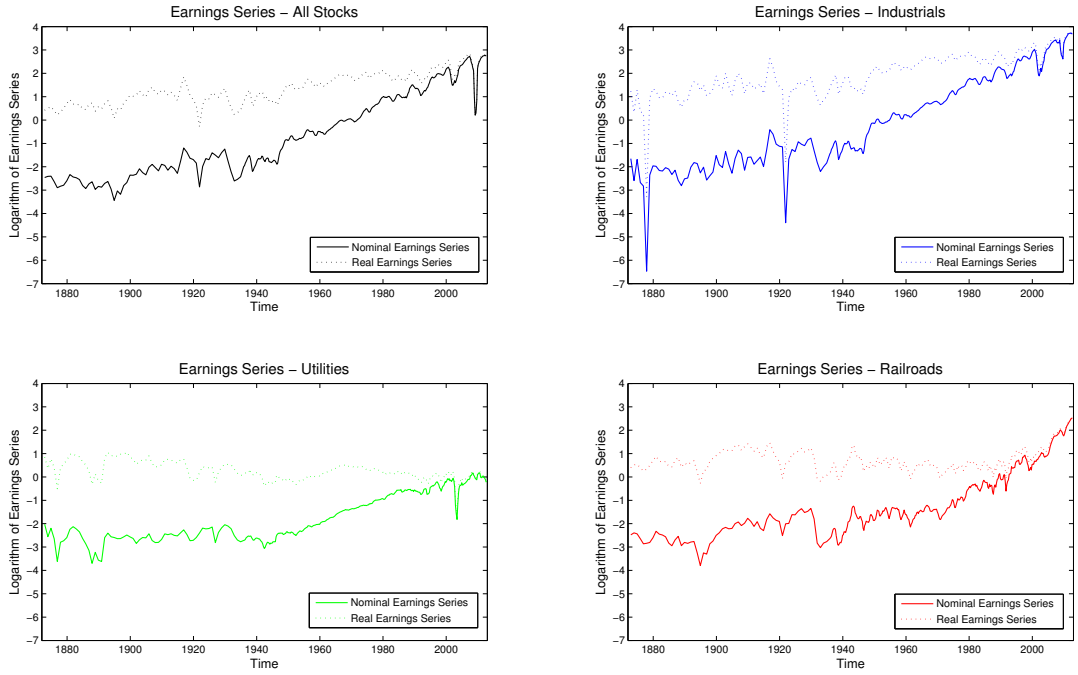


Figure 2: These charts display 12-month trailing earnings information for the overall stock market (All Stocks), Industrials, Utilities, and Railroads. Observations from 1872 until 1934 are annual and constructed from series P and R in Cowles (1939), as outlined in Appendix A. Starting in 1935, observations are diluted quarterly earnings from the Security Price Index Record, which we turn into 12-month trailing sums in quarterly frequency. Information for Railroads after 1982 is from the Analysts' Handbook published by Standard & Poor's Statistical Services, which is the source of all earnings information starting in the fourth quarter of 2001. The procedure for inflation-adjustment is identical to the one applied to the data in Figure 1. The nominal data is normalized to correspond to the normalization of the nominal price series to 1 at the end of 1872. As with the price indices in Figure 1, both the nominal and the real series are displayed on a logarithmic scale (natural logarithm).

of railroads in history, and the complete end to private inter-city passenger service in the U.S. But, we must remember that freight service grew quite steadily and strongly over this whole period, and that passenger service was ultimately driven out by competing modes of transportation, by government regulation, and by subsidies to Amtrak, which could thereby outcompete any private service.

Since our datasets begin at the start of the 1870s, the most of the tumultuous opening decades of the railroad industry are excluded. Stories of wild speculation in railroad shares, and shameless exploitation by railroad promoters of naïve investors, began to appear in the 1830s. Notably, there was a remarkable railroad mania in the United Kingdom which grew through the early 1840s. Railroad prices peaked in late 1845, and then proceeded to lose two thirds of their value by 1850. Before its burst, the bubble attracted investments from many British luminaries of that time, including Charles Babbage (the inventor of the computer), Charles Darwin, William Makepeace Thackery, John Stuart Mill, the Brontë sisters, Benjamin Disraeli, William Gladstone, and Robert Peel. That railroads generated real excitement among investors then is understandable. Some railroads in the 1840s were routinely clocking peak speeds of 75 miles per hour in passenger service, which must surely have impressed travelers of that day. One anonymous writer, a couple decades later, in the 1860s, described the feelings that trains stirred:

> "I can never quite get used to the wonder that is ever passing by my door. Some people, I am told, who have a very fine horse, are, after owning it for years, as proud of its form and its breed as the day on which they bought it. Some love to look on a picture or a statue their life through, others are always lifting a favorite author down from a library shelf to read over and again the same old familiar passages, written in dead or living languages. So I also never can get quite careless and indifferent to the wonder that is passing my door here day and night,–never get rid of my admiration at the precision of all the

> locomotive's movements, - never get rid of my astonishment that it comes to me year in, year out, with almost the regularity of Nature's workmanship. It is due, and it is soon here; it sets down its living load, takes up the living, screams to me, "Coming, coming!" and screams to me, "Good bye!" - a wild creature put into harness by man, the mightiest he has ever controlled, but docile indeed if man uses him wisely."[39]

The emotions this author describes sound familiar, like the emotions we feel today with some of our new twenty-first century gadgets. Railroads were very strong in psychological impact in their time, even to the point of romantification in art, literature and music. Starting in the 1840s, people began to define the era as "the railway age." People grotesquely overrated the importance of the railroad to the economy according to economic historian Robert Fogel.[40]

The emotions regarding railroads and the essential ambiguity of any investment in changing industries must have produced repeated mispricings of their stocks, if not outright bubbles. Railroad "manias" later in the 19th century are described for Argentina, Brazil, China, India and Mexico. But continental Europe is notably absent: throughout Europe, railroads were government projects, apparently because of continental law and the need to exercise rights of eminent domain to start a railroad.

According to Cole and Frickey (1928), US railroad prices more than doubled 1843-4, around the time of the UK bubble, but escaped most of the UK railroad crash in the years after 1845. The big correction in the US railroad market was not until the panic of 1857. That panic was marked by massive railroad stock declines, which led to failures of banks which had lent to them, just as much as the crisis of 2008 was tied to declines in mortgage securities. There was a more than doubling again of railroad prices 1861-4. But at the time our data begins, these most extraordinary stories of railway manias were history. The panic of 1873, at the very beginning of our sample, is not so tied to railroads. This panic is tied by historians to Jim Fisk's and Jay Gould's machinations 1866-8 with Erie Railroad stock, but our data do not show a major collapse of railroad prices in 1873, as can be seen from Figure 1 displaying price indices and Figure 2 displaying the associated earnings series. Railroads by the beginning of our sample had become a mature industry, as a rather dominant part of a larger economy.

Like railroads, utilities had their beginnings in some new and exciting high-tech industries. Water supply was already taken for granted in cities at the beginning of our sample: in fact, supply of running water to homes dates back thousands of years to ancient Rome. But supply of natural gas and then of electricity must have seemed truly revolutionary.

Natural gas was indeed exciting. An observer in 1878 wrote:

> "Gas is rather too unsubstantial a material, it might seem, to create excitement among sober business men ... But, invisible and unsubstantial as this class of chemical agents is, at the present moment nothing in nature or art is more talked about the world over ... As regards the use of gas for fuel, this is a matter which excites the interest of every housekeeper in Christendom. If, the good ladies say, we can only be rid of the nuisance of coal fires, with the attendant ashes, dust, dirt, and labor, we shall regard it as the dawn of the housekeeper's millennium."[41]

As if that excitement weren't enough, it was that same year, 1878, that Thomas Edison founded the Edison Electric Light Company. It was in 1882 that Edison actually began supplying some electricity, and electric light, to homes in New York and London. The process of electrification involved substantial innovation and capital investment, a process that was largely completed, even in rural areas, by the 1920s.

Over time, new inventions would create new flurries of excitement in the Utilities sector. For example, the invention of vaccum tube amplification of telephone signals in the teens brought the telephone to most households by the 1920s: by then the average American made over 200 phone calls a year. These must have seemed like life-changing events, and exciting to investors. It is no surprise that the Utilities index showed the greatest strength in the years leading up to the 1929 stock market crash. Barron's wrote in July 1929, months before the crash:

> "The unbridled speculation in public utilities has now reached the point where it is incapable of interpretation by any reasonable yardstick of appraisal. A tremendous merger movement, virtually a

[39] See Anonymous (1867).
[40] See Fogel (1971).
[41] See Nichols (1878).

> scramble, is taking place. The hysterical psychology now prevalent resembles that which attended the boom for steel stocks at the formation of the United States Steel Corp. in the early days of this century. For the long pull, investors will probably do better to make their commitments when the smoke has cleared way and the big units shake down into their niches."[42]

With the 1929 stock market crash, the Utilities sector suffered a sharp decline. Figure 3 shows inflation-adjusted total return indices for the three sectors as well as for the overall stock market on, again, a logarithmic scale. Between the third quarter of 1929 and the second quarter of 1932, Utilities lost more than 75% of its real total return value, with the decline only being slightly smaller at approximately 55% when looking at a ten-year horizon after the peak in the third quarter of 1929. The Utilities sector had some of the most glamorous and life-changing innovations before 1929, but by the middle third of the twentieth century it too, like Railroads, became a stable, and relatively boring, sector. The overall trend in the real total return performance of both of these sectors is fairly similar during the second half of the twentieth century, with divergences only re-emerging towards the turn of the century, when Railroads clearly outperformed Utilities. Looking at the overall performance of Utilities over the entire sample period, the importance of the total return consideration to reinvest dividends becomes apparent once again. Whereas Utilities actually declined in real terms according to their price level, the more accurate picture provided by the real total return index shows an almost 2,028-fold increase over the full time span.

As impressive as the performance of the real total return series for Utilities is, the performance of the real total return series Industrials sector is even more so, as it increases slightly more than 19,600-fold over the entire period of almost a century and a half. The Industrials sector, as we have noted, is rather more a catchall sector, whose glamour and excitement consists of sparkles here and there, and not something that we can describe well in general terms. In the first half of the twentieth century, there was excitement about radio, which started, by the 1920s, to link all households together in networks of entertainment and news. Around the same time, automobiles became a mass market industry, and this was exciting too, for it freed people from public transportation, so that they could go wherever they wanted, whenever they wanted, and it changed the layout of the nation, made possible the full development of the suburbs as well as the urbanization of the land. But then, the excitement about the automobile in the U.S. was tempered by the advent of much foreign competition. Airlines eventually conspired with the automobile to further displace the railroad. In smaller details, the motion picture became a pastime of millions, countless consumer goods in disposable packages changed the everyday elements of our lives. Computers and mobile phones, and all the other trappings of the information revolution were later comers to the Industrials sector. All of these separate stories came at different times and with different psychological connotations to give a sense of excitement to the Industrials sector.

3.4 Pricing of the Sectors

The value proposition suggests that market prices for the three sectors reflect a combination of genuine information about the future and public misinformation and emotion. The efficient markets hypothesis is by this view best regarded as a half-truth, something always to keep in mind, but never to believe fully. A speculative bubble is a self-fulfilling feedback loop of price increases to further increases, analogous to an epidemic. The initial price increases attract attention to theories that justify the bubble, and bring more people in who wish to buy, causing prices to increase, and these price increases further generate talk and excitement. They are social epidemics, with new era stories spreading as if they were viruses. Eventually, the epidemic must come to an end, for there are no more new people to infect, and the bubble bursts. But, bubbles do not need to be of this extreme variety, just as a disease epidemic does not have to affect all of society, but can be limited to a susceptible subgroup.

The classic stories of the most salient bubbles (the tulipmania, the roaring twenties, etc.) describe them as a preoccupation of all of society, and focal point of attention. But mispricings smaller than bubbles do not necessarily represent a preoccupation and need not be society-wide. Mispricings can pass unnoticed by most people, in contrast to bubbles that make the headlines and may be remembered in popular culture for centuries. Such mispricings can

[42] See The Trader (1929).

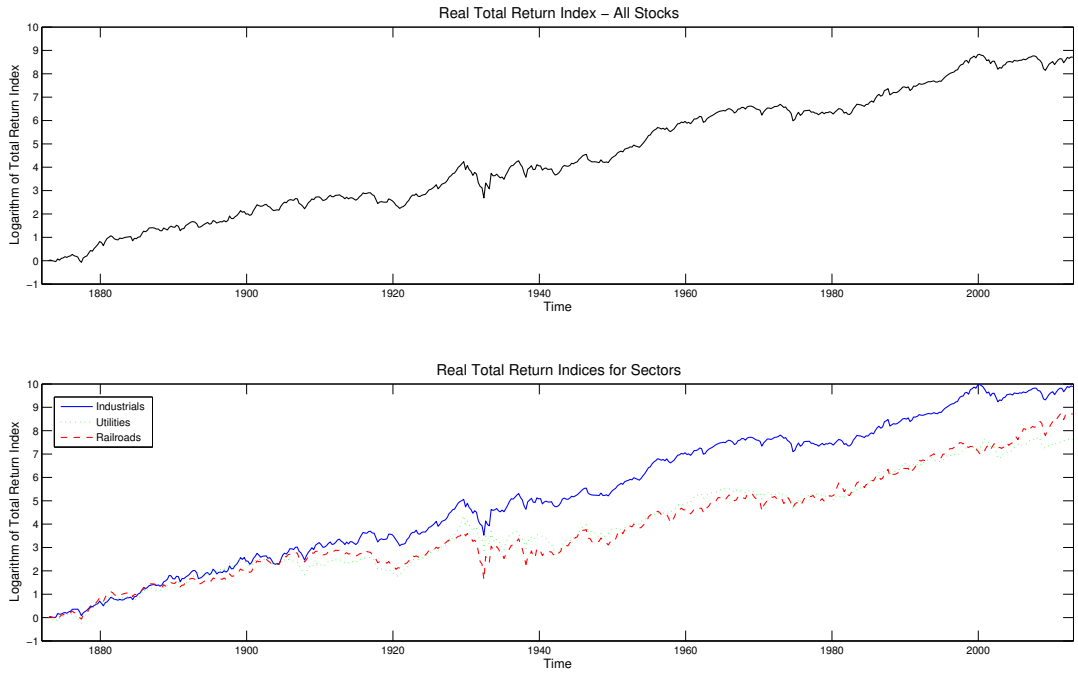


Figure 3: These two charts show a total return index for the overall stock market (top graph) and for the sectors Industrials, Utilities, and Railroads (bottom graph). The total return indices are quarterly starting at the end of the fourth quarter 1872 and ending at the end of the fourth quarter in 2012. The nominal price data is identical to the data displayed in Figure 1. Dividend information from 1873 until 1934 is extracted from series P and series C in Cowles (1939) on a monthly basis, as outlined in Appendix A, and subsequently aggregated into quarterly information. After 1935, dividend information is from the Security Price Index Record, complemented by information from the Analysts' Handbook for Railroads after 1982 and for Industrials, Utilities, and the overall stock market starting in the fourth quarter of 2001. The dividend information after 1935 is initially quoted as 12-month trailing sums, which we turn into quarterly information using quarterly actuals from the 1990s. In consequence, we obtain dividend series of actual quarterly numbers for the overall stock as well as for the three sectors. We construct nominal total return series by assuming that dividends are reinvested at the end of the quarter, in which they occur. Subsequently, we adjust the indices for inflation, applying the same procedure as outlined in the description of Figure 1. Finally, we remain true to the convention to display all time series on a logarithmic scale (natural logarithm).

in principle not even have spectacular price increases, perhaps only outperformance relative to another sector, if that outperformance is seen as justifying and supporting some viral story that suggests the success of that sector.

The mispricings can reflect, to some degree, simple misperceptions, such as the common misperception today that the Railroads sector has something to do with passenger service. Today in the U.S., freight revenues for railroads are roughly thirty times passenger revenues, and all of the intercity passenger revenues go to the government anyway.[43] Yet many people are left with the faulty impression that decline in public interest in passenger rail service is a risk for publicly-traded railroads today. That misperception does not drive everyone, but plausibly enough people to make a difference that will be noticeable in returns.

Figure 3 shows that the Railroads sector has underperformed the stocks reflected in the Industrials sector quite substantially over the full sample period. This has been noted before, e.g. by Keeler (1983), though never studied over such a long time interval as we show. However, it can also be seen, from the relative slopes, that there were decades-long intervals when the Railroads sector outperformed other sectors, including some recent decades. The same has been true for the Utilities vis-a-vis either the Industrials and the Railroads sector, too. Thus, there are opportunities for investors in switching between sectors over time.

Despite the outlined differences between the sectors in Figure 3, it is interesting to observe that the overall growth patterns of the real total return series of the three sectors are overall rather similar. How can this similarity arise, when the industries reflected by the three sectors were so different from each other? Wouldn't one think that random shocks to the three industries would have produced much different returns over a century? In particular, do some industries not grow obsolete and others undergo rebirths? The similarity of long-term trend across the sectors highlight the importance of distinguishing between the industry and the existing firms within an industry, which comprise our index at any point of time, and to distinguish between return and earnings growth. There is a common fallacy of thinking that investing in a sector total return index is much the same as owning a share of all the earnings of a sector. In fact, there is a fundamental difference, partly because the reinvestment of dividends causes the investment to tend to grow faster than the sector, and because of the issuance (or repurchase) of new shares, either by existing companies or by new companies that spring up. Over long intervals of time the difference is substantial.

The fallacy of neglecting what share issuance there will be is as old as the stock market. In fact, this fallacy is fundamentally tied up with mispricings, as the popular intellectual magazine The Spectator wrote in the 1845, just before the peak of the railroad bubble in the U.K:

> "Such an universal hallucination was perhaps never witnessed. There is indeed one fact which accounts for it, though it does not disarm the prospect of its danger; of a large proportion of the schemes it may be said that anyone would be not only justifiable but commendable; the difficulty is not in the inherent badness of all the projects, but in their aggregate enormity; yet as each set of speculators is justified in its own enterprise, it is hard to make it morally responsible for all the rest. Individually, each speculator—we are alluding now to the bona fide and desirable projects—is right; collectively they are all wrong."[44]

4 The Cyclically Adjusted Price-Earnings Ratio and Individual Sectors

4.1 The Construction of the Cyclically Adjusted Price-Earnings Ratio

The fallacy that the quote from The Spectator refers to in the previous section might easily have been spotted by looking at the aggregate value of all railroad stocks, including those that had not yet any earnings, and comparing that with the existing earnings on those stocks. Fortunately, with our price and earnings data, we do not need to collect special new data for this: We can do this systematically at all times for all sectors by looking at the Cyclically Adjusted Price-Earnings (CAPE) ratio. The general rule for this ratio, in line with, for example, the price-earnings ratio, is that low values of the ratio provide a signal of undervaluation, and hence a potentially attractive investment opportunity, whereas high values signal overvaluation, which will supposedly be corrected in the future.

[43] See the description of the of the role of Amtrak for the Railroads sector in the beginning of section 3.
[44] See Littell and Littell (1846) for a reprinted version of the article from The Spectator.

Improving on earlier inspiration from Graham and Dodd (1934),[45] Campbell and Shiller (1988) use the term "cyclically adjusted" to refer to the fact that a ten-year average of earnings is used for the ratio, which is spread over an interval of time much longer than the typical business cycle, and hence is not much affected by such. However, the consideration of data over such an extended period of time necessitates two kinds of corrections. The second correction eliminates the impact of inflation during the ten years by using real numbers for the numerator and denominator of the ratio, a procedure that has already been applied in Campbell and Shiller (1988), when first constructing the original CAPE ratio that forms the basis for the ratio that we will be considering.[46] The first correction acknowledges that CAPE has an innate dependency on corporate payout policy, coming from its long-term view that it takes on earnings. For illustration purposes, let us revert to the individual-company level and consider a very much simplified environment with two identical companies that have not exhausted their growth opportunities. The only difference is that company A does not distribute any earnings to its shareholders, whereas company B pays a substantial fraction, say 50%, of its earnings as dividends. If both companies are exactly identical now (in particular their investment opportunity set coincides), the earnings per share for the two companies will gradually diverge with company B's EPS underperforming those of company A. Additionally, company B's price per share will be lower by the nature of the differing payout policy in future years. However, the crucial aspect is that the future ten-year average of earnings per share will not change as much as the price per share across the two companies, which means that company B's CAPE will be lower than that of company A.

The outlined issue does not only pertain to comparisons across companies (or sectors), but also to comparisons along the time series of a company (or sector). We do not consider it appropriate that a period of increased payout to shareholders should be associated with the CAPE ratio for this company (or sector) providing a signal of a more attractive investment opportunity by virtue of a lower value.

In order to counteract this dependency of CAPE on corporate payout policy, we revert to total return numbers for the computation of CAPE. For the numerator of CAPE, the current price, this amounts to the inclusion of dividends to construct a new time series. The question arises of how to adapt the denominator of CAPE. It is out of question to leave the denominator of CAPE unchanged, as one would divide an increasingly higher number total return price number by a ten-year average of earnings that remains on the same scale, which would lead to the deterioration of the levels of CAPE over time. Therefore, we scale the 12-month trailing earnings per share numbers by the ratio of the total return price to the price return price, i.e.

$$E^{\text{scaled}} = E \cdot \frac{P^{TR}}{P^{PR}} \Leftrightarrow \frac{P^{PR}}{E} = \frac{P^{TR}}{E^{\text{scaled}}}.$$

Another way to think about this change of the 12-month trailing sum of earnings is that it preserves the P/E-ratio of a sector after the price changes from price return numbers to total return numbers.

After having adjusted these scaled earnings for inflation, two more issues need to be addressed until we can compute the denominator of the CAPE ratio. The first issue relates to earnings announcements of companies. In order to access real-time valuation signals from the CAPE ratio, it is crucial to ensure that all information is known at the time that we report a CAPE ratio in a given month. Whereas this complication does not arise for prices, as these are known on the spot, it certainly affects the denominator of the CAPE ratio, which builds on earnings. As companies report earnings for a given period with a lag, we have to make sure that all earnings information used for the computation of the denominator of the CAPE ratio has already been reported at the time, for which we report the CAPE ratio. Generally, we therefore aim to impose a six-month lag on earnings, which is however complicated by the second issue to be addressed, the fact that our earnings observations change frequency from annual observations before 1935 to quarterly observations thereafter. For the details of our solution to this pair of issues, please refer to Appendix B.

We compute the CAPE ratio for the overall stock market as well as for the three sectors Industrials, Utilities, and

[45] See the quote from Benjamin Graham's and David Dodd's textbook Security Analysis, referred to in footnote 3.

[46] The original article, Campbell and Shiller (1988), computes the CAPE ratio as the logarithm of the real price divided by the ten-year average of the logarithm of earnings, i.e. a geometric average of earnings. Our specification does not take this formulation as its starting point, but formulations as they have been employed later in Campbell and Shiller (1998), Campbell and Shiller (2001), and Shiller (2005). The latter formulations have taken hold in the public discourse about CAPE.

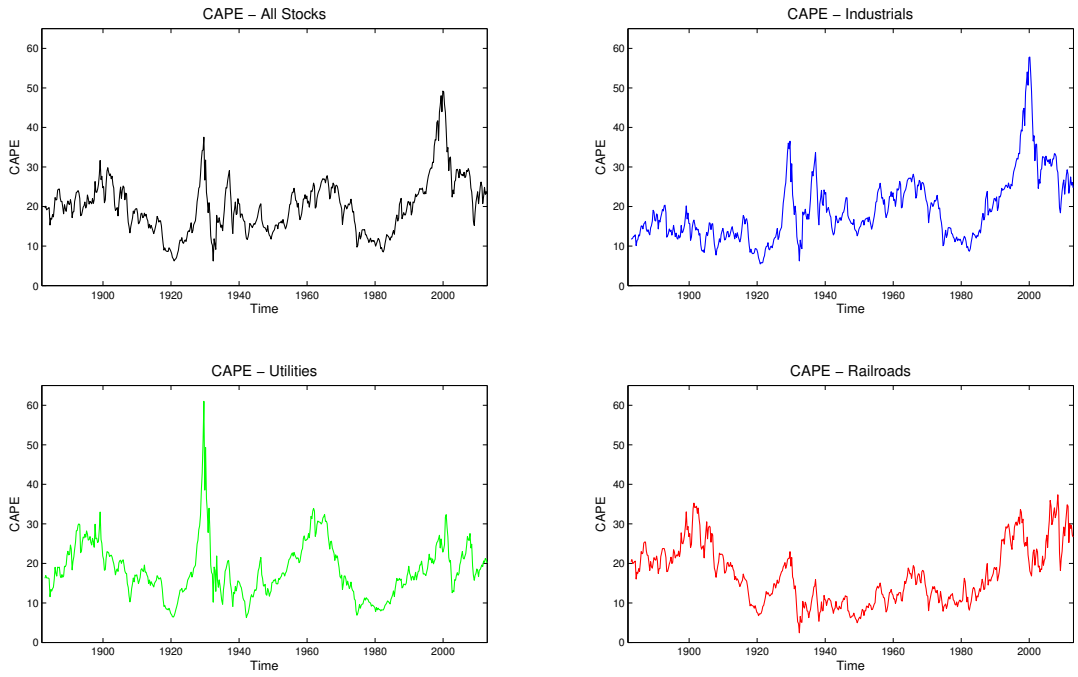


Figure 4: This figure shows the CAPE ratio for the overall stock market (All Stocks), Industrials, Utilities, and Railroads. The quarterly time series start with the fourth quarter of 1882 and end with the second quarter of 2012. As outlined in Section 4, we construct CAPE as the ratio of real total return price (see Figure 3) and a ten-year average of ten observations of scaled, 12-month trailing real earnings numbers (see Figure 2 for the unscaled analogue of these earnings series). The scaling (further explained in Section 4) occurs at the level of nominal numbers, which are subequently inflation-adjusted. When constructing ten-year averages, minor adaptations are necessary to account for the transition from annual earnings to quarterly earnings in 1935, which we describe in detail in Appendix B. In general, we lag all earnings observations after 1936 that enter the long-term averages of earnings by two quarters, in order to guarantee that all earnings that enter a CAPE ratio in the end of a specific quarter are known at that time. Before 1936, where we only have annual earnings observations, we use earnings numbers for the previous year for each quarter of a given year except for the first and earnings numbers from the year before that in the first quarter, effectively establishing a lag ranging between six months (second quarter) and fifteen months (first quarter).

Railroads at the end of each quarter starting with the fourth quarter of 1882[47] and ending with the fourth quarter of 2012. For example, to compute the CAPE ratio at the end of the first quarter of 1995, we use the real total return series[48] at the end of the first quarter of 1995. For the denominator, we use the real earnings series at the end of the third quarter of 1994,[49] the end of the third quarter of 1993, and so on, until the end of the third quarter of 1985. Each of these real earnings numbers will be scaled by the contemporaneous price series to adjust for the usage of the total return series in the numerator. For example, the observation at the end of the third quarter of 1993 will be adjusted using the total return series at the end of the third quarter of 1993 together with the price index at the end of the third quarter of 1993. As the final step in the computation of the CAPE ratio, we form the arithmetic average of the ten scaled, real earnings observations and divide the total return price at the end of the first quarter of 1995 by this average.

We plot the CAPE ratio for the overall market as well as for the three sectors Industrials, Utilities, and Railroads in Figure 4. Note that the CAPE ratio of the three sectors shows a relatively similar pattern across sectors through time, but there are significant differences. Notably, in the 1929 peak, the Utilities sector stood out, because of a sharp

[47] At the starting point for the CAPE ratio, we have to ensure that we have a sufficient length of earnings history. We therefore cannot start the series for the CAPE ratio until ten years into our full sample.

[48] As previously outlined, the inflation adjustment for this observation builds upon the level of the Consumer Price Index (CPI-U) in December 1994.

[49] This number will have been inflation-adjusted by the CPI-U in June 1994. We will not explicitly reference the observations from the CPI-U series for the remaining earnings observations in this example, but these follow the same principle.

increase in the numerator of the ratio, and Utilities' CAPE ratio set the all-time high record in the third quarter of that year with slightly more than 60. In that same quarter, the CAPE ratio for the Industrials sector was high, but much less so, only slightly more than 36. In comparison, the Railroads sector's CAPE ratio at that time was around 20. The other dramatic peak, in the fourth quarter of 1999, was dominated not by the Utilities sector but by Industrials, when the Industrials sector's CAPE ratio reached nearly 58 then. In comparison, the CAPE ratio for Railroads at that time varied between somewhat more than 30 and approximately 15 and the ratio for Utilities barely exceeded 30 at its peak. The all-time record low in our sample was set by Railroads, in the second quarter of 1932, with the CAPE ratio dropping below 2.5. During that time, Industrials's CAPE ratio was also low at close to 6 and Utilities' CAPE ratio came close to 10. We venture that these broad swings in entire sectors of our economy are not entirely due to changes in rational expectations for the future dividends and earnings, and must mean something for subsequent returns.

The CAPE ratio can clearly be categorized as a value measure, in particular as a measure that captures long-term valuation because of its consideration of ten years of earnings data. Its value-focus situates it in the academic tradition originating from Basu (1977), Fama and French (1992), Lakonishok et al. (1994), and Chan et al. (1995), who have investigated value-based investing in the United States, or Chan et al. (1991), who focus on Japan. The most common value measures are however the price-earnings ratio, which neglects the benefits of ten-year averaging of the earnings information alluded to above, and the book-to-market ratio, which interacts balance sheet information and market information and is therefore a different kind of value-variable. Moreover, the literature on value-investing classically looks at value measures on an individual-firm-level, which are then used to sort individual stocks into a value-based portfolio, most commonly in a long-short consideration. We however look at the CAPE ratio as our valuation measure on the aggregate level of a sector already, building on aggregated price- and earnings-information to assess the over-and undervaluation of an entire sector of the US economy, and not of an individual company.

For more recent overviews of the literature on value-investing, see Chan and Lakonishok (2004) as well as Fama (1998) or Campbell (2000), where the latter two surveys focus on theoretical explanations that have been put forward as rationales for the existence of value premia. Most recently, Asness et al. (2012) and Fama and French (2012) have extended the evidence in favor of the effectiveness of value-based investing to a broad international level and have also underscored the existence of value premia in asset classes other than equity.

4.2 The CAPE Ratio, a Railroads Example, and Return Predictability

Looking back at the railroads bubble in the middle of the 19th century, one could speculate about the CAPE ratio for the Railroads sector during that time. While we do not have CAPE data for 1845, the peak of the railroad bubble in the U.K., and while the reliability of the CAPE ratio as an indicator may not be high in the beginning years of a new industry, we can be pretty sure that the ratio of aggregate price to aggregate earnings for all railroads would have been very high then, and CAPE would have served as a warning away from that bubble. Less dramatically, CAPE might have warned us of smaller mispricings in the Railroads sector, or in other sectors, that are generated by investor emotions or prejudices that are a lot weaker than those seen in bubble times, and harder to judge except by reference to some indicator like CAPE.

According to the historian Andrew Odlyzko, arguments like that in The Spectator were made repeatedly by doubters during the U.K. railroad boom of the mid-1840s, but their effect on thinking about the bubble was diminished by another popular argument of that time: that the railroads, initiated in the 1830s in another boom, had appeared to fail at first but were ultimately successful, that, similarly, all the railroads being built in the 1840s should eventually be successful as the economy and demand grew.[50]

That argument must continually have been on the minds of people thinking about railroads well into the twentieth century, until competing modes of transportation began to be seen as ending the seeming inevitability of railroads. The memory of the railroad bubble of the 1840s, embellished in many newspaper stories, must have stuck in many people's minds, too, however less clearly as the decades wore on as memories faded. There was a tug of war between two seemingly plausible stories, the story of railroad inevitability in a growing economy and the story of bubbles,

[50] See Odlyzko (2010).

and no fully objective way to declare a winner. Ultimately, the outcome of such a tug of war is never clear, it results merely in an array of different opinions about whether today's price is good or not. Moreover, there is no simple way to convey this opinion to the multitude, for price changes every day, and so opinions on railroads as investments may change faster than the knowledge of them can disseminate across investors.

The evaluation of the stories we tell about stocks requires considerably more subtle thinking, and cannot be achieved definitively, since the distant future is always diffuse in our minds. Because of such subtlety, we may ultimately find ourselves relying on such an imperfect indicator as the CAPE ratio in judging a bubble, or in judging more modest mispricing. Such an indicator, however imperfect, may loom large in intelligent thinking about the mispricing, because it allows comparisons with other times and other sectors, and there is no more precise way to do quantitative work to inform our judgments.

We will provide evidence for the predictive ability of the CAPE ratio in two ways. At first, we look at each sector in isolation and for the sake of completeness also at the overall stock market (represented by the S&P Composite Index). Our goal in this context is to analyze whether the level of CAPE contains any information about future long-term returns accrued within a specific sector. In the next section, we will extend this analysis to include predictions about differential performance across sectors, outlining how to use the CAPE ratio to assess differences in future returns between different sectors.

However, we would first like to address the question whether the CAPE ratio of a sector contains any information about the very same sector's subsequent return performance. The left panels of Figures 5 through 8 show scatter diagrams for the S&P Composite Index as well as for each of the three sectors.[51] On the x-axis, we display the CAPE ratio of a sector, incorporating the transition to real numbers and to total return numbers. It is measured at the end of each quarter starting in the third quarter of 1902 until the fourth quarter of 2002.[52] The y-axis captures subsequent ten-year returns, as measured from the beginning of the quarter immediately following the end of the quarter, for which the corresponding CAPE ratio of the sector has been computed, to the beginning of the quarter that begins ten years later. That is, the first observation is a ten-year return from the beginning of the fourth quarter of 1902 until the beginning of the fourth quarter of 1912 and the last observation is the ten-year return between the beginning of the first quarter of 2003 until the beginning of the first quarter of 2013. As the time intervals for the data used for the computation of the CAPE ratio and the data for the computation of the corresponding long-term return are truly disjoint, we obtain a genuine and clean assessment of the performance that an investor can expect ten years into the future when knowing the level of CAPE at the end of a given quarter. Finally, we have labeled each point in the scatters by the time, at which CAPE is computed. Each entry contains the year followed by the number of the quarter.

Figure 5 is for the S&P Composite Index. Four caveats are necessary to point out with regard to the figure and subsequent correlation scatters. First, it is important to keep in mind that the y-axis depicts returns over a ten-year horizon. Therefore, any individual long-term return observation is necessarily fairly similar to a subsequent long-term observation, which is measured one quarter later, as successive long-term return observations have an overlap of 39 quarters. This autocorrelation pattern will prevent long-term returns to make sudden shifts. Second and somewhat relatedly, we cannot rule out the existence of a small-sample problem, even though the scatter plots cover more 100 years of data.[53] This is due to the long-term nature of returns depicted in the scatter diagram, which reduces the number of truly independent long-term return observations to approximately ten. Finally, we acknowledge a potential spurious association problem, as we assess the long-term change in prices relative to a ratio, which uses the same price variable in its numerator. This problem should not be mistaken for a predictability problem, which we have addressed by carefully constructing the CAPE ratio from observations that have occurred before the time period over which returns are measured. Nevertheless, spuriousness might lead to some degree of negative association between the CAPE ratio and subsequent returns. Moreover, the randomness of the CAPE ratio introduces a second spuriousness

[51] We will address the information contained in the right panels of Figures 5 through 8 in the subsequent section, but it will turn out to be interesting to compare these panels side-by-side.

[52] Although our sample allows us to compute the first CAPE ratio at the end of the fourth quarter of 1882, we limit the time series for the CAPE ratio to 1902. This is to ensure comparability with the cross-sectional predictability analysis based on the CAPE ratio, which we will outline in the next section.

[53] See, for example, Stambaugh (1999) for the impact of small-sample effects on the analysis of predictability.

channel into the correlation of the CAPE ratio and subsequent long-term returns through the randomness of the CAPE ratio in conjunction with de-facto smallness of the sample.

In order to address these caveats, we perform a simulation study for each of the correlation scatters involving the CAPE ratio and long-term returns, whose procedure closely matches the kind of simulation study that Campbell and Shiller (1998, 2001) perform. We hereby create simulated data for the CAPE-ratio by estimating an AR(1)-process for the natural logarithm of the underlying data for the CAPE-ratio.[54]We hereby correct the autoregressive coefficient of the logarithmic CAPE-ratio using the formula in Orcutt and Winokur Jr. (1969), which corrects for a bias as analyzed in Kendall (1954) and Marriott and Pope (1954).[55] Moreover, we simulate quarterly returns as random walks in the natural logarithm of the underlying inflation-adjusted total-return index.[56] Imposing the assumption of a random walk for the price process makes returns inherently unpredictable, which is a core assumption of our simulation exercise. Ultimately, we are interested in rejecting the notion that the association between our valuation ratio and subsequent long-term returns can arise from purely unpredictable returns.

The simulated normally distributed error-terms in both the AR(1)-equation for the CAPE-ratio as well as in the random walk evolution for the real, total-return index match the mean and the standard deviation of their empirical counterparts. Additionally, we set up their joint distribution's covariance to match that of the empirical counterpart. When calculating the empirical quantities, we rely on samples for the CAPE ratio as well as for the total-return index starting in the third quarter of 1902, which is the same timeframe displayed from Figures 5 onwards.[57] Finally, for both the evolution of the simulated CAPE ratio as well as the simulated total return indices, we use the actual starting value of each of the underlying series.

Running $100,000$ iterations of the simulation exercise, we obtain a distribution of the correlation between the CAPE ratio and subsequent long-term returns. This will then allow us to determine a p-value corresponding to the actual correlation value as the probability mass that falls below the actual correlation value. This is what we will henceforth refer to as the simulation-based p-value. It captures the probability that the observed correlation is compatible with the notion that we aim to reject of unpredictable returns for the overall market or for the respective sector of the overall market, respectively.[58]

We perform the simulation exercise independently for the overall stock market as well as for the three sectors Industrials, Utilities, and Railroads, as long as we consider the predictive ability of the CAPE ratio for the overall market as well as for the three sectors independently from each other. This will be different, once the CAPE ratio will play a role in allocations across sectors in the later analyis, which we will then elaborate on in greater detail.

Figure 5 shows a clearly negative association between both CAPE and subsequent long-term returns. That is, historically, low values of CAPE, indicating undervaluation, are followed by higher long-term returns and, vice versa, high values of CAPE and Relative CAPE indicating overvaluation correlate with lower subsequent ten-year returns.[59]

[54]Taking the natural logarithm of the underlying CAPE-series allows us to better control for the skewness in the empirical distribution of the ratio, making the assumption of normally distributed error terms that is then incorporated in the simulation more reasonable.

[55]See MacKinnon and Smith Jr. (1998) for an exposition of this kind of bias correction for autoregressive estimation.

[56]The argument in favor of the natural logarithm rests again on the skewness of the empirical distibution of the percentage changes in the underlying real, total-return index.

[57]When we later investigate correlation structures for shorter timeframes, we preserve the original horizon for the estimation of the empirical quantities and simply shorten the time series of the simulated processes to reflect the shorter time horizon.

[58]As pointed out, the simulation procedure outlined above closely resembles the approach pursued in Campbell and Shiller (1998, 2001), which is related to the procedure in Goetzmann and Jorion (1993); Nelson and Kim (1993) or in Campbell and Shiller (1989). The simulations in Campbell and Shiller (1998, 2001) aim at understanding the predictive ability of the dividend-price ratio, and not of the CAPE ratio, which is - on the overall market level - also analyzed in Campbell and Shiller (1998, 2001), but not the subject of the simulation analysis of the statistical validity of the predictive evidence provided by the scatter diagrams of the valuation ratio and subsequent long-term returns.Nelson and Kim (1993) further the robustness of the predictive ability of the dividend-price ratio in the presence of small-sample concerns, mostly relying on an AR(1)-structure. Relatedly, Goetzmann and Jorion (1993) concern themselves with the predictive ability of the dividend yield, but rely on simulations of the return process to infer the dividend yield from the actual, historic dividend-price ratio. Campbell and Shiller (1989) explicitly include a close variant of the CAPE ratio (with a thirty-year horizon for earnings) in their simulation exercise, but only to develop a better understanding of their VAR parameter restrictions implied by linearized present value models connecting dividends and returns.

[59]Displaying the scatter diagrams of long-term returns and a valuation ratio such as the CAPE ratio naturally touches on the more general academic discussion of the validity of predictive regressions, see for example Welch and Goyal (2008) and Campbell and Thompson (2008). Welch and Goyal (2008), investigating a multitude of predictive variables, raise doubts about the out-of-sample performance of this kind of regression in general. In contrast, Campbell and Thompson (2008) introduce sensible and fairly non-restrictive restrictions on predictive regressions (related to the sign of the prediction variables and/or to the sign of the implied equity premium) and strongly argue in favor of the validity of predictive regressions. Interestingly, the version of the CAPE ratio that they use proves to be one of the

The correlation between CAPE and long-term return is almost −57% with a simulation-based p-value of 0.0213. This evidence is consistent with the results presented in Campbell and Shiller (1998, 2001) and extends their results by roundabout another decade, even in the presence of the modifications that we have undertaken to the version of the CAPE ratio in these analyses. The statistical simulation furthermore underscores that the degree to which this result arises from statistical spuriousness is very small at best.

Figure 6 displays the information contained in Figure 5 in a summarized and more pronounced manner.[60] It starts from the same pairs of observations for the CAPE ratio and the subsequent, annualized real total ten-year return that Figure 5 shows. It however then treats the complete time series (from the third quarter of 1902 until the fourth quarter of 2002) of CAPE ratios as an empirical distribution and splits the CAPE observations in tertiles. The bars represent averages of the corresponding ten-year returns for all the observations of the CAPE ratio that fall into a specific tertile. On a technical note, by "average" we refer to the geometric mean of the non-annualized real total ten-year returns, which is subsequently annualized. The notion that the CAPE ratio negatively predicts future returns should be reflected in the tertile bars in the form of decreasing bars from the first to the third tertile. This is clearly the case in the left panel of Figure 6, which displays a clear monotonicity from the first to the third tertile. The smallest observations of the CAPE ratio are associated with an average annualized total real return of approximately 10%, which exceeds the average for the second tertile by more than 4% annually and the average of the thrd tertile by nearly 7% annually.

The evidence in the scatter diagrams for the Industrials and the Utilities sector, shown in the left panel of Figures 7 and 9 respectively, is nearly identical to that for the overall stock market. One observes a strongly negative correlation of approximately −54% for Industrials with an associated simulation-based p-value of 0.0378 and an even slightly stronger effect for Utilities with a correlation of −63%, which translates into a simulation-based p-value of 0.0072.

With regard to the averaged representation of long-term real total returns for tertiles of the distribution of the CAPE ratio, the evidence for Industrials and Utilities is also coherent when compared to the graph for the overall stock market. Both charts show a monotonous decline of the average long-term returns with the first decile's return exceeding the third decile's by approximately 5.5% for Industrials and 8% for Utilities. The only qualitative difference is the behavior of the average return for the middle tertile, which resembles the first tertile's much more for the Industrials sector than it does for the Utilities sector.

Finally, the evidence for the Railroads sector in Figure 11 is consistent with the evidence from the other two sectors and also from the overall stock market. This sector provides the comparably weakest overall evidence in favor of the predictive ability of CAPE for subsequent long-term returns, with a correlation of nearly −35% that is associated with quite a high simulation-based p-value of 0.2283. As can be seen from the time labels of the scatter plots, the main impact for the weaker evidence of the scatter plot comes from the 1990s, which was a period with a comparatively high level of CAPE. At the same time, the long-term returns reaching into the beginning of the twentieth century have been quite substantial, in contrast to the overvaluation signal given by CAPE.

The tertile-based evidence for the Railroads sector underlines the predictive ability of the CAPE ratio for subsequent returns, showing a solid degree of monotonicity in average returns. The degree, to which the average for the first tertile exceeds that of the third tertile, is even the strongest among all three sectors with more than 8%.

In conjunction, these correlations and the accompanying bar chart representations are new confirmation of our theory, never before tested on sector data and never investigated for such a long time period. However, the CAPE ratio is not a spotless predictor of returns, far from it: Its annual return predictions (instead of the ten-year predictions shown in Figures 5 through 12) exhibit weaker correlation. Correlations for this shorter time period are slightly higher −24% for the overall stock market, slightly stronger than −24% for Industrials, close to −20% for Utilities, and approximately −23% for Railroads. We will refrain from specifying the corresponding simulation-based significance results at this point and refer instead to section 6, in which we analyze the sensitivity of our correlation analysis more generally. There, we will not only investigate variations in the horizon over which we calculate subsequent returns, but also in the time horizon used to smooth earnings in the denominator of the CAPE ratio. We will thereby vary

strongest predictive variables (see tables 1 and 2 in Campbell and Thompson (2008)).

[60] We will again focus on the left panel initially and revert back to the information contained in the right panel, once we have formally introduced the concept of the Relative CAPE indicator.

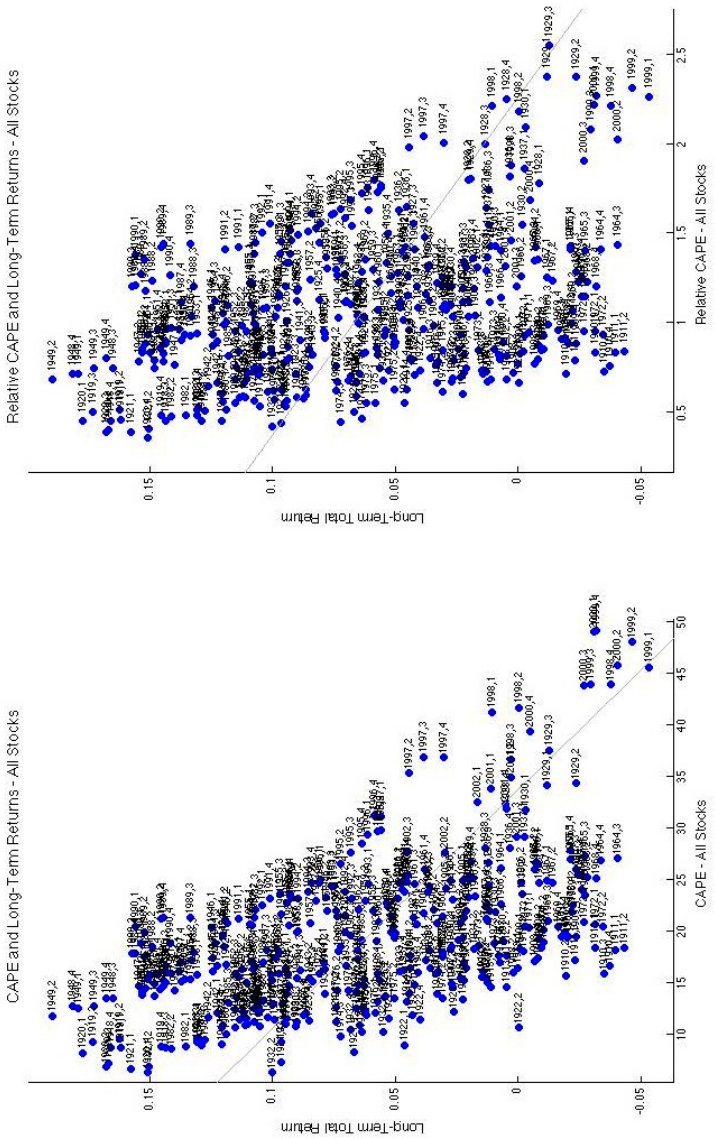


Figure 5: The left panel depicts a scatter diagram of the CAPE ratio (discussed in Section 4) of the overall stock market and the right panel depicts the analogous scatter using Relative CAPE (discussed in Section 5). On the x-axis of the left panel, we display the CAPE ratio of the overall market, incorporating the transition to real numbers and to total return numbers. It is measured at the end of each quarter starting in the third quarter of 1902 until the fourth quarter of 2002. The x-axis of the right panel captures the Relative CAPE of the overall market, using the same timing convention as for CAPE ratio. The y-axis, which is common to both panels, captures subsequent ten-year total returns, as measured from the beginning of the quarter immediately following the end of the quarter, for which the corresponding CAPE ratio or Relative CAPE indicator of the sector has been computed, to the beginning of the quarter that begins ten years later. The first observation is a ten-year return from the beginning of the fourth quarter of 1902 until the beginning of the fourth quarter of 1912 and the last observation is the ten-year return between the beginning of the first quarter of 2003 until the beginning of the first quarter of 2013. We have labeled each point in the scatters by the time, at which the CAPE ratio or the Relative CAPE indicator, respectively, is computed. Each entry contains the year followed by the number of the quarter.

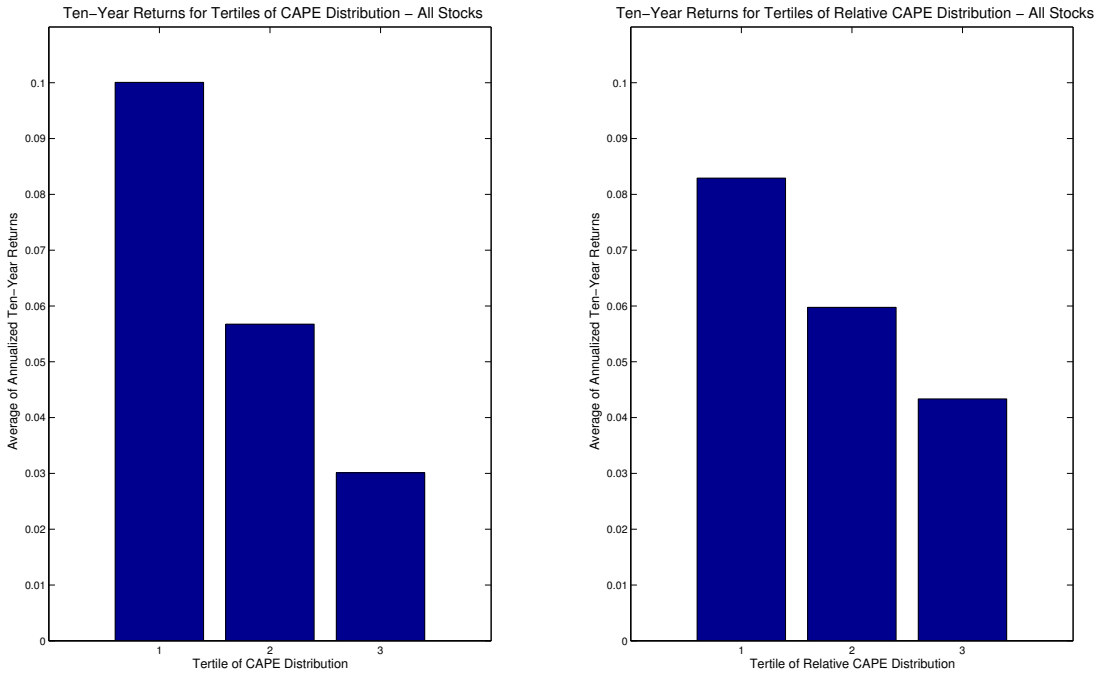


Figure 6: This figure displays the same information about the overall US stock market as Figure 5, but in a more aggregate manner. In the left panel, we split the distribution of CAPE ratios for the overall stock market from the third quarter of 1902 until the fourth quarter of 2002 into three tertiles. For each tertile, we geometrically average the corresponding, subsequent, non-annualized ten-year return numbers and annualize the average. The right panel is conceptually identical to the left panel, but uses the Relative CAPE indicator instead of the CAPE ratio.

the smoothing horizon to range from the horizon used for the classical P/E valuation ratio (1 year) until we have reached the horizon underlying the CAPE ratio (10 years).

Reconsidering the evidence from Figure 11 for Railroads in a little bit more detail, it becomes apparent that the Railroads sector's scatter seems to have been the strongest among the three sectors in terms of tightness of fit of the scatter in the earlier years, but then to have shifted substantially after the mid-1990s. Before the 1990s, the entire scatter was shifted to the left relative to the scatters of the other two sectors. After the 1990s, the entire scatter moved up to the right to rejoin the region of Industrials, as displayed in Figure 7. Fortunately, the CAPE ratio rose even dramatically higher for Industrials around that time, and so a comparison across sectors based on the two sectors CAPE ratio around 2000 would have given exactly the right advice: to move out of Industrials into Railroads then. However, we do not wish to rely on that lucky example, and wish somehow to correct CAPE for its tendency to go through long swings, as we saw with Railroads. A few further aspects are noteworthy from Figure 4, the plot of the sectors' CAPE ratio over time. Railroads had the lowest CAPE ratio for a half century, from the 1920s to the 1970s. Utilities had the lowest CAPE ratio for most of the time since the late 1970s. But, equally apparent, investing in the lowest CAPE stocks would not appear to have been advantageous, since, as we have seen in Figure 3, Industrials outperformed overall, while their CAPE was usually higher than the others.

5 CAPE-Based Comparisons Across Sectors

5.1 Varying Growth Prospects Across Sectors

The shift in the scatter of points that we see for Railroads is troublesome. It means that using CAPE to predict returns, or to distinguish between projected returns, for sectors may be unreliable. The problem is that a sector, as illustrated by Railroads, might show low CAPE for a long period of time because its CAPE is, at that time in history, naturally lower. Let us consider why, in the example, average CAPE, and the entire scatter of points, might

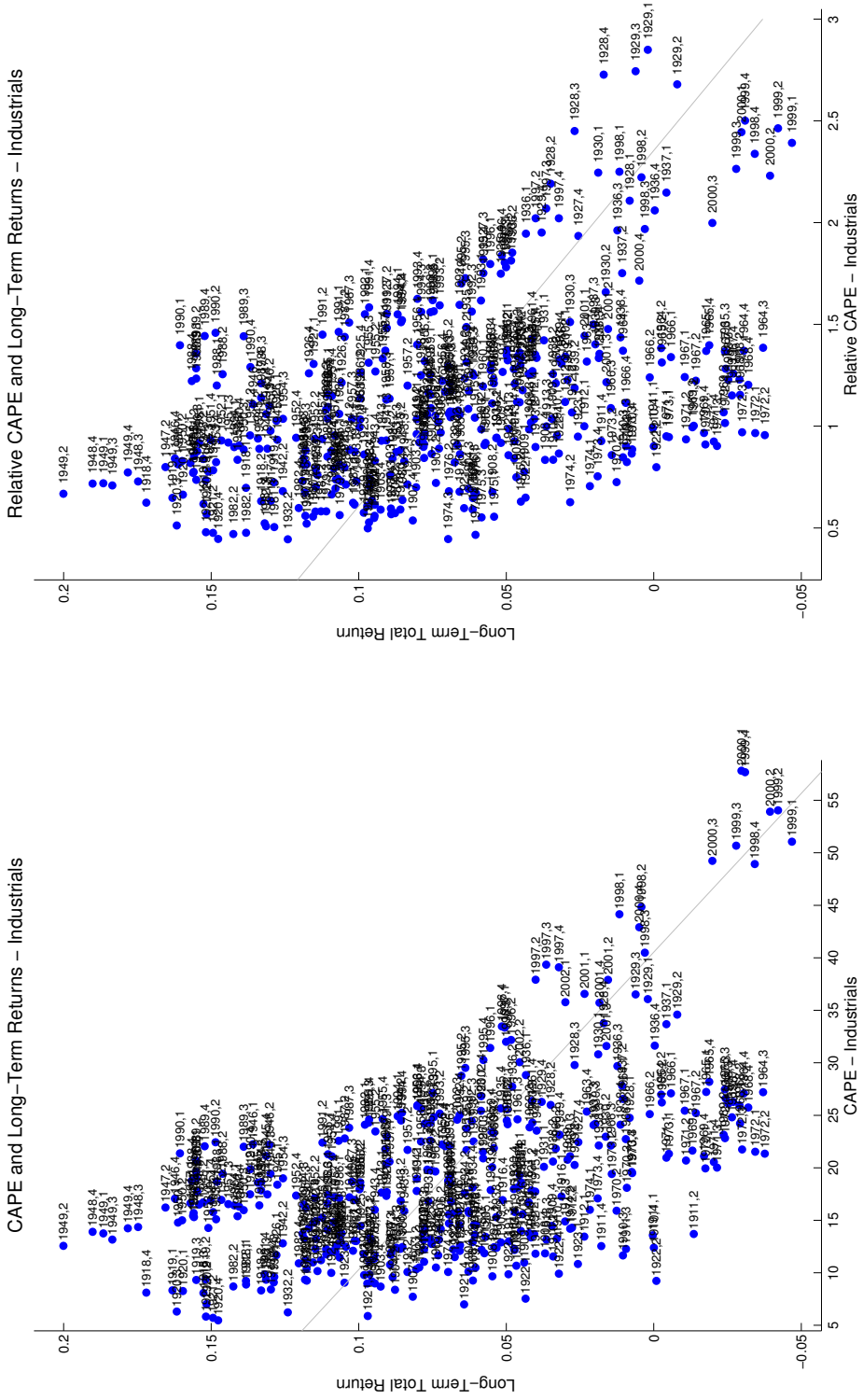


Figure 7: This figure is the analogue of Figure 5 for the Industrials sector instead of the overall stock market.

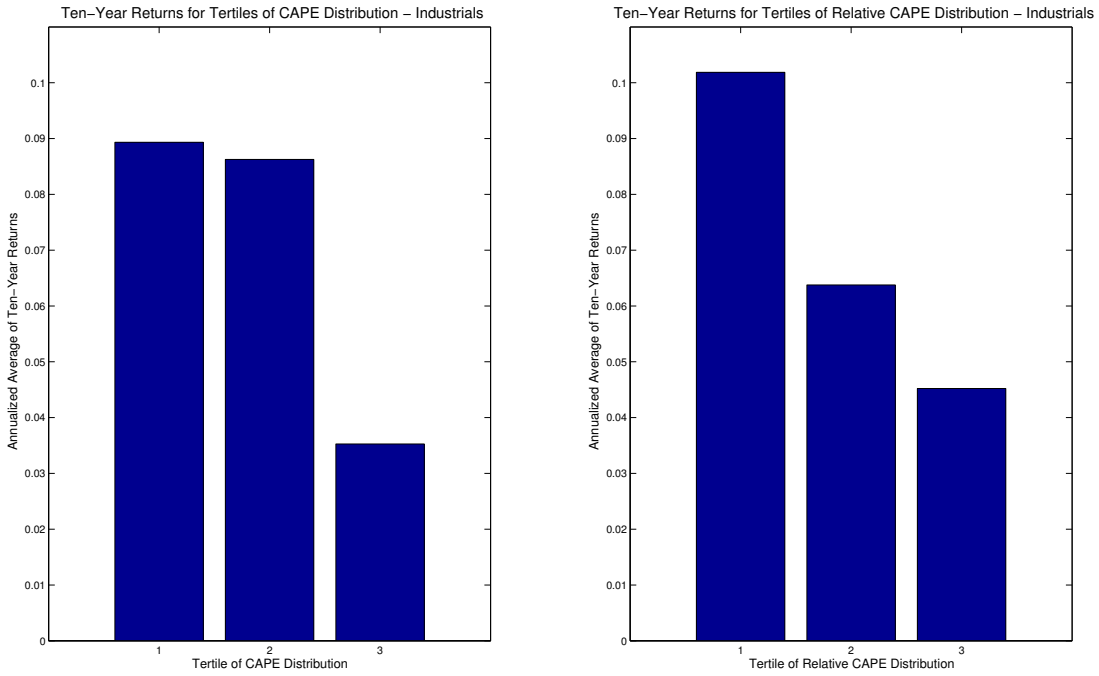


Figure 8: This figure is the conceptual analogue of Figure 6, summarizing the information from Figure 7 for the Industrials sector.

have been lower for the period before the 1990s.

First we must ask, what, after all, are railroads, in terms of their engineering and economic significance? Railroads are distinguished, of course, from other modes of transportation by their use of steel rails and steel wheels. That represents a significant invention in terms of transport efficiency. Wrought iron rails were invented in 1820 and steel rails in 1857, as part of a lively period of invention of better surfaces for transportation vehicles in the early nineteenth century. There were other inventions to reduce friction of that time, including plank roads and canals. Ultimately, steel rails won the competition for durable low friction, and remain in heavy use today. Railroad transport, with unimpeded right of way, avoids energy loss from frequent starts and stops, and the close-spacing of railway cars, closer than is safe to maintain with cars and trucks, lowers wind friction. Hence it is possible to economize on motive energy, by a factor of as much 20 to 1 between railroads and trucks. Today, basic railroad freight rates are typically less than three cents per ton mile.

Also, the development of techniques to create a smooth railroad, and signaling systems to prevent accidents allowed greater speeds. But another invention came in soon after, that of the pneumatic tire, patented by Robert William Johnson in 1846, but not produced commercially until 1889. Tired vehicles, while far less energy efficient, have the advantage that they can be driven smoothly on ordinary roads, which are subsidized by governments, and are not confined to a predefined rail path. Yet other inventions, those of the automobile, and the expressway, appeared by some in the twentieth century eventually to make the railroad obsolete.

The basic technology of steel rails and steel wheels remains today very significant. And yet other inventions have encroached on its aspect of uniqueness. This fact has apparently been a factor in the lower CAPE ratio that the Railroads sector has exhibited in much of the twentieth century. The relatively less inspirational trend for the Railroads sector became readily apparent relatively early in the twentieth century, though the trend was never entirely clear. A 1931 Barron's article noted that:

> "There has come an increasingly prevalent belief that the railroad star is setting; that they are in process of obsolescence; that they may go the way of the ox-cart. People have viewed with alarm the increase in highway competition from automobiles, buses, and trucks; the growing threat of gasoline and natural-gas pipelines, a newer development, and the possibility of inroads on traffic by airways and

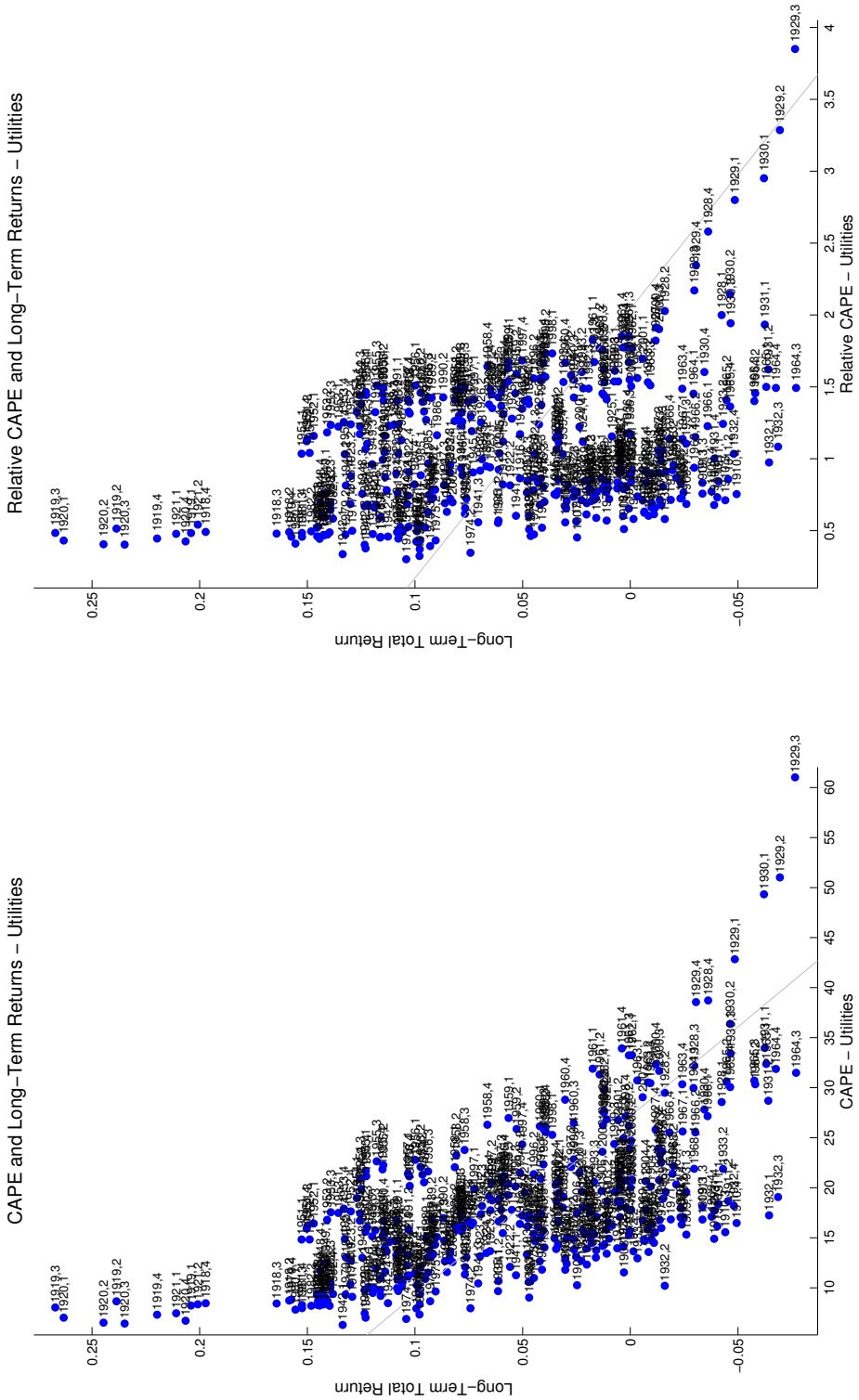


Figure 9: This figure is the analogue of Figure 5 for the Utilities sector instead of the overall stock market.

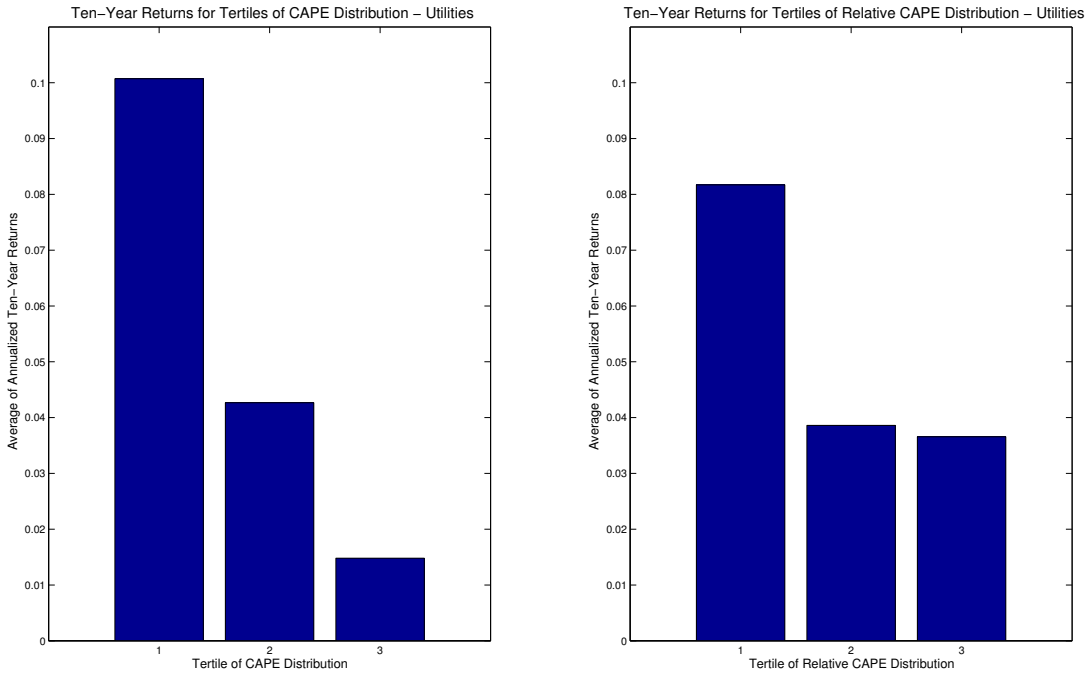


Figure 10: This figure is the conceptual analogue of Figure 6, summarizing the information from Figure 9 for the Utilities sector.

waterways."[61]

The article argues that the end of railroads was by no means imminent, but that the effects of the newfound competition had been that "the rate of increase of railroad traffic in the decade ended in 1929 has been much retarded in comparison with previous decades. The upward curve flattened out."[62]

The article interpreted the lower price-earnings ratio for railroad stocks, when compared with other stocks at the time, as reflecting this expected weakened future trajectory, or, at least as reflecting the risk of a lower trajectory. Indeed, Figure 3 shows that the Railroads sector has been on a lower trajectory for nearly the next half-century after this Barron's article was written, so some degree of lower expectations for railroads would have been in an approximate sense ex-post rational. This rational expectation might then account for lower CAPE, and for a scatter of points like the one in the left panel of Figure 11 that is shifted down.

As can be seen from Figure 4, the CAPE ratio for the Railroads sector was pronouncedly lower than that for the Utilities and the Industrials sector from the mid-1920s to the mid-1970s. Given the arguments presented in the above Barron's article, would it make sense to preferentially invest in the Railroads over the Utilities and the Industrials sector on the basis of CAPE over that whole period? Perhaps we can credit the markets with enough efficiency to take proper account of such obvious long-term trends or risks like the effects of the growth of automobile.

5.2 The Construction of the Relative CAPE Indicator

When the CAPE ratio makes somewhat shorter-term swings, it would appear to be more logical to attribute many of these movements to public overreaction to news, and more likely to expect to see these movements later reversed. We do not take the view that there is a single scale of values for the level of the CAPE ratio, on which all sectors can be compared meaningfully. In contrast, we attribute to each sector its own scale of values of CAPE, from which over- and undervaluation of this particular sector *relative to its own long-term history* is derived. When we make comparisons across sectors, we therefore construct a standardization of the CAPE ratio, referred to as the Relative

[61] See Hovey (1931).
[62] See Hovey (1931).

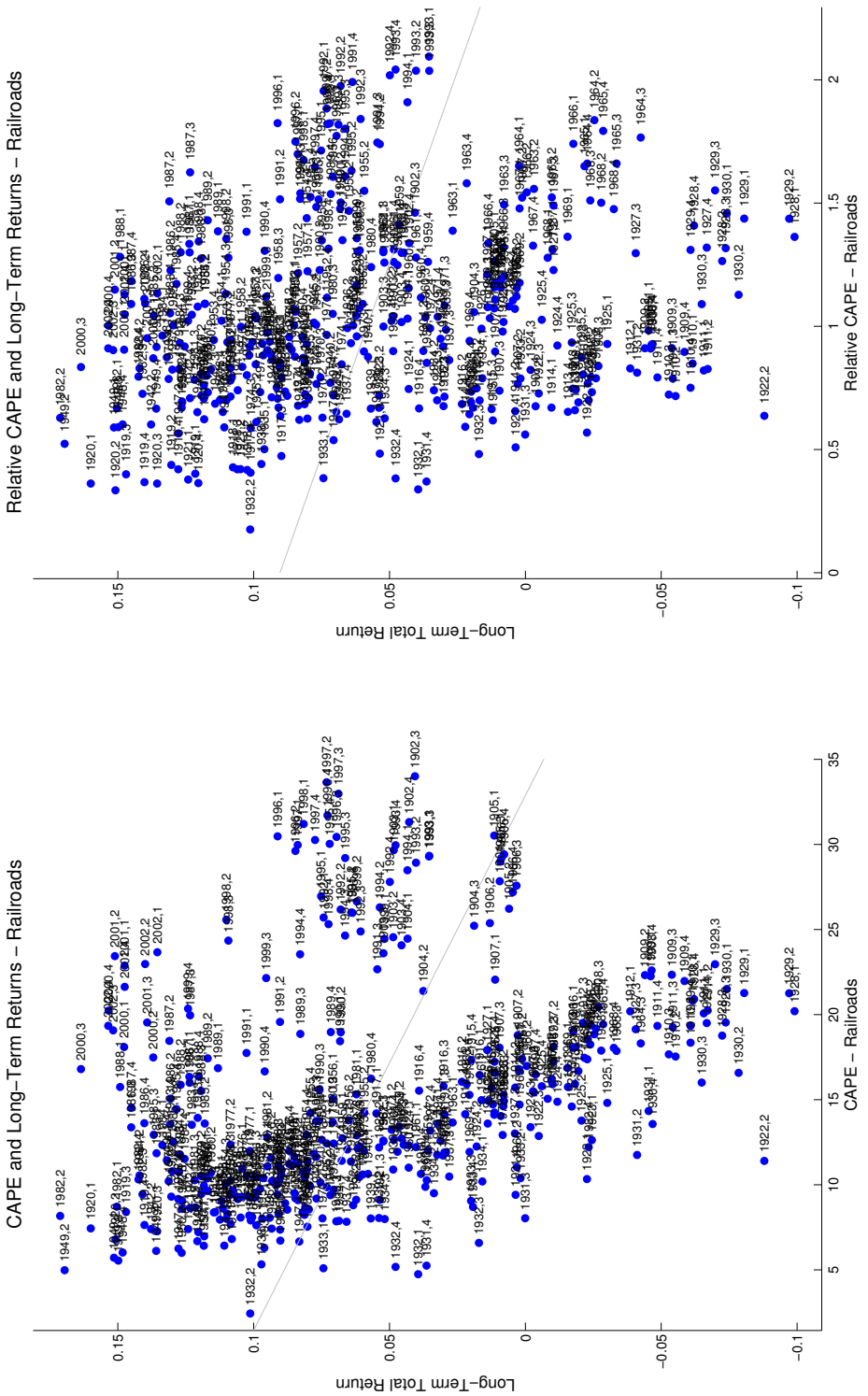


Figure 11: This figure is the analogue of Figure 5 for the Railroads sector instead of the overall stock market.

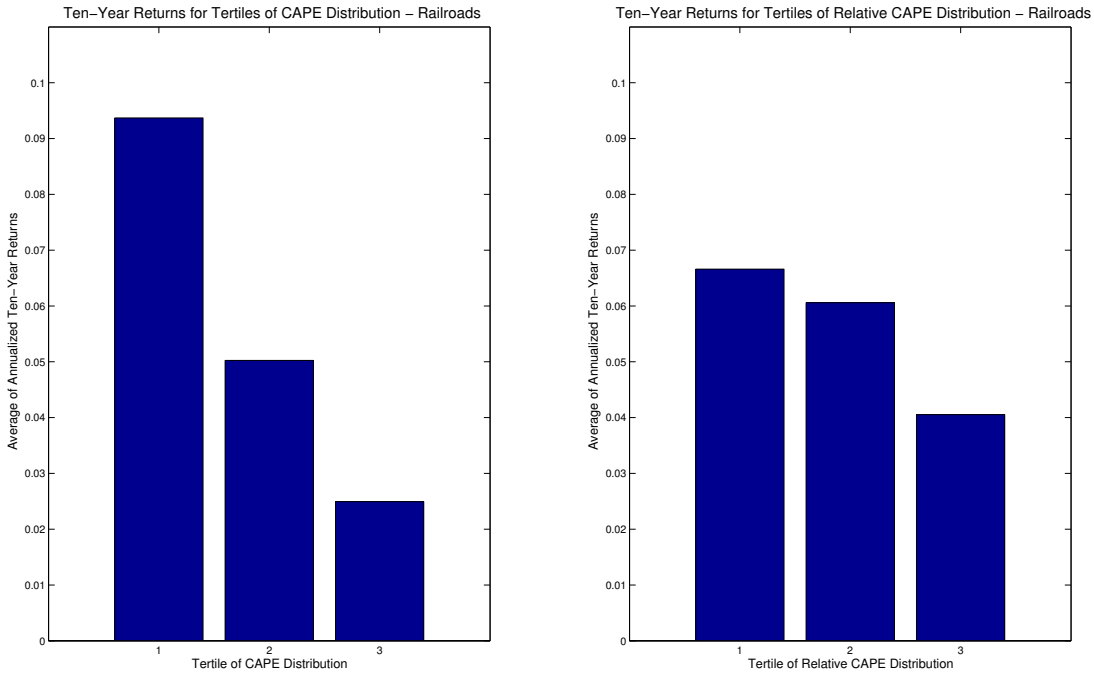


Figure 12: This figure is the conceptual analogue of Figure 6, summarizing the information from Figure 11 for the Railroads sector.

CAPE indicator, by dividing the current period's CAPE by its own twenty-year average,[63] which has previously been winsorized at the five-percent level.[64]

The division by the twenty-year average of CAPE removes both long-term trends and intermediate-term cycles, leaving only the relatively higher frequency movements in CAPE. In terms of its valuation signal, it controls for the issue of sectors having different growth trajectories, like the Railroads sector between the 1920s and the 1970s. At the same time, it also erases other kinds of idiosyncracies of sectors that might affect the level of the non-standardized CAPE ratio, like, for example, different accounting standards across sectors, which might affect a sector's earnings reports. Figure 13 shows the Relative CAPE indicator for the overall market as well as for the three sectors.

The question arises how closely the Relative CAPE indicator matches the behavior of the original CAPE ratio and how much of the return predictability evidence outlined in Section 4 persists after transitioning from the CAPE ratio to the Relative CAPE indicator. We do not expect the CAPE ratio and the Relative CAPE indicator to be perfectly correlated, as the Relative CAPE indicator corrects for some deficiencies of the CAPE ratio with regard to across-sector comparisons. On the other hand, it would be comforting not to see the Relative CAPE indicator overturn all the information by the CAPE ratio. It turns out that the correlation between the two is about 85% for the overall stock market, approximately 79% for the Industrials sector, almost 90% for the Utilities sector, and about two thirds for the Railroads sector, the sector that we have found to require the biggest correction in terms of the information provided by the plain CAPE ratio.

The right hand side panels of Figures 5 through 12 replace the CAPE ratio with the Relative CAPE indicator in the predictive analysis for long-term returns for the overall market as well as for the three sectors. Overall, the patterns across the two panels for each of the four figures is similar. The correlations for the overall market equals approximately -39%, which is approximately a 17% drop when compared to the evidence for the CAPE ratio. We apply the same simulation ($100,000$ repititions) as in section 4 in order to assess the significance of the correlation result against a simulated distribution derived under the assumption of no return predictability, whereby we perform the outlined calculation of the Relative CAPE indicator for each simulation of the time series of the CAPE ratio. The

[63] The twenty-year average contains the current quarter's CAPE and the previous 79 quarterly observations of a sector's CAPE ratio.

[64] The winsorization then sets the four highest CAPE observations of the 80 CAPE observations used for the computation of the long-term average to the level of the fifth-highest observation and the four lowest observations to the fifth-lowest observation.

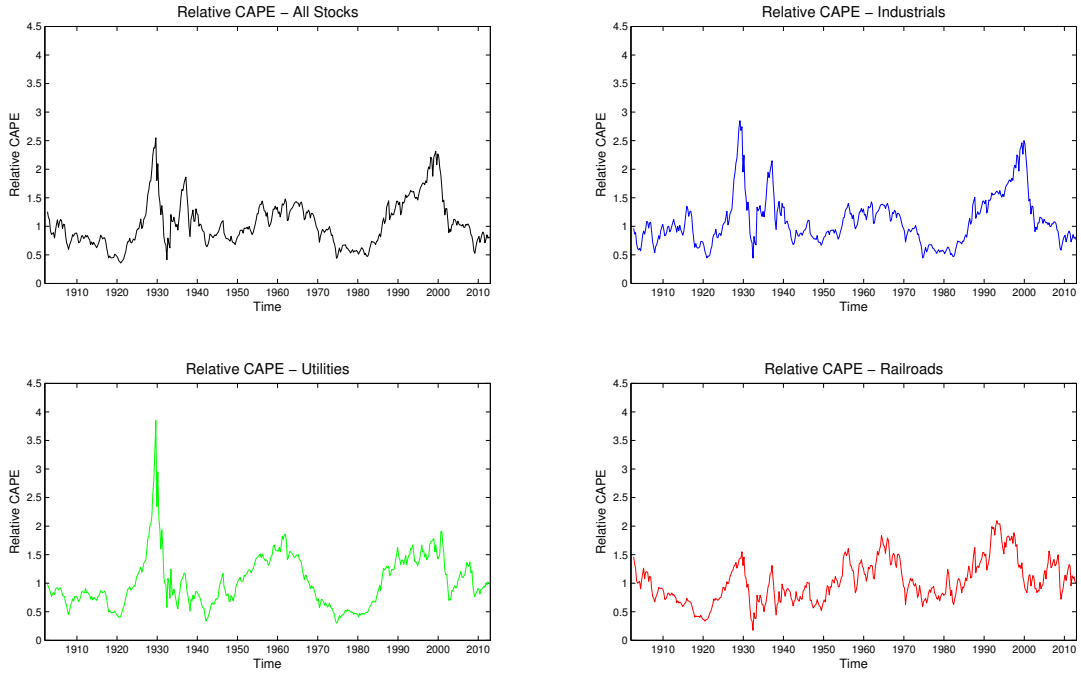


Figure 13: Whereas Figure 4 depicts the CAPE ratio for the overall stock market (All Stocks), Industrials, Utilities, and Railroads, this figure shows the Relative CAPE indicator. The frequency is again quarterly and observations cover the time period from the end of the third quarter in 1902 until the end of the fourth quarter of 2012. Relative CAPE in a given quarter for each of the four time series is the ratio of this quarter's CAPE and a modified, twenty-year average of that time series' history of CAPE ratios, up to and including the current CAPE ratio. We modify a time series' 80 CAPE observations (twenty years of quarterly data) that enter the long-term average of CAPE ratios in the sense that we winsorize the top and the bottom at the 5% level.

resulting simulation-based p-value for the overall stock market is 0.0689, which still provides considerable evidence of predictability, even in the presence of the outlined normalization of the CAPE ratio for the overall stock market. Reconsidering the market-based evidence of the predictive ability of the Relative CAPE from the perspective of the tertile-bars paints a much more calming picture about the effectiveness of the Relative CAPE indicator, even for the overall stock market. Although the differences in average returns are less pronounced compared to the bars for the CAPE ratio, there is still a clear monotonicity and the first tertile's average exceeds that of the third tercile by approximately 4%.

Industrials display a negative correlation between the Relative CAPE indicator and subsequent long-term returns that exceeds -48%, which is only an drop (in the strength of the association) of about 5% compared to the evidence for the CAPE ratio. The associated simulation-based p-value is 0.0281, emphasizing the incompatibility of the observed correlation with the assumption of unpredictable returns. Figure 8 complements the evidence, again displaying a clear monotonicity pattern across the tertiles as well as a difference between the average return for the first and the third tertile of approximately 6%. The evidence in favor of the predictive ability of the Relative CAPE indicator is similarly strong for the Utilities sector. Even though the correlation number in Figure 9 moves more than 20% closer to zero correlation to a level slighty less than -39%, the simulation-based p-value low at 0.0671. Moreover, the difference between the average return of the first and third tertile in Figure 10 is approximately 4.5%.

Finally, the correlation of the Railroads sector drops from approximately -34% to approximately -20%, a correlation that translates into a simulation-based p-value of 0.2656, which means that the Railroads sector continues to provide the relatively weakest evidence in favor of a sector's CAPE ratio to predict subsequent returns. Figure 12 complements the notion of return predictability for this particular sector by showing a clear, albeit somewhat attenuated in comparison to the left panel, monotonicity across the tertiles with a maximum difference in the average returns close to 3%.

Taking all of this evidence into account, the normalization involved in the creation of the Relative CAPE indicator appears to leave the overall predictive ability of the CAPE ratio unchanged, while at the same time addressing important concerns on the path to using the CAPE ratio to assess the performance of sectors relative to each other. Observe in Figure 13 that the Relative CAPE indicator is not consistently lower for Railroads compared to Utilities or Railroads between the mid 1940s and the 1970s. Moreover, after the mid-1990s, the shift upwards in the CAPE ratio for Railroads, as evidenced by in Figure 4, is offset by an increase in the denominator, the long-term average of the Railroads sector's CAPE ratio, making for a more consistent fit of the scatter plot in Figure 11 when using Relative CAPE.

Table 1 concludes the discussion of the interaction between the CAPE ratio or the Relative CAPE indicator of a specific sector and that same sector's long-term return. It employs the same methodology as has been used for Figures 5 through 12, but investigates the impact of different starting dates for the sample, for which the correlations are computed. The first column, included as a reference, contains the correlations depicted in Figures 5, 7, 9, as well as 11, whereas the second column computes the correlations when the sample is restricted to the period between the fourth quarter of 1935 and the fourth quarter of 2012. During this period, we solely rely on information from the Security Price Index Record and the Analysts' Handbook publications from Standard & Poor's. This does however not mean that we completely disregard data from Cowles (1939), as we still rely on it for the computation of the CAPE ratios and the history of the CAPE ratios needed for the calculation of the Relative CAPE indicator. The third column of Table 1 limits the data to the last 50 years of the sample to give a sense of the most recent history. Simulation-based p-values are included in parentheses below the associated correlation numbers.

The reported results in Table 1 exhibit robustness with regard to the choice of the time horizon. All correlations are negative and rather similar in size across the different sample periods, although the pattern for the simulation-based p-values outlines the impact of the sample length on the significance of the correlation coefficients against the assumed unpredictability for returns that the simulation exercise is based on. The Railroads sector actually displays the most interesting pattern, in that the correlations involving the CAPE ratio shrink towards zero, the shorter the sample becomes. This has already become apparent from Figure 11 that clearly shows a difference in the left panel brought about by the Railroads sector's behavior at the end of the 20th century. At the same time, the right panel of Figure 11 outlines the benefits of the construction of the Relative CAPE indicator, as it shows a much more

Correlation with Subsequent 10-Year Return	Start in Q3 of 1902		Start in Q4 of 1935		Start in Q4 of 1962	
	CAPE	Relative CAPE	CAPE	Relative CAPE	CAPE	Relative CAPE
All Stocks	-56.56% (0.0213)	-39.01% (0.0689)	-64.64% (0.0262)	-45.68% (0.0915)	-61.43% (0.1581)	-36.06% (0.3194)
Industrials	-53.95% (0.0378)	-48.10% (0.0281)	-61.31% (0.0514)	-47.56% (0.0841)	-58.06% (0.2237)	-36.13% (0.3315)
Utilities	-62.91% (0.0072)	-39.44% (0.0671)	-64.92% (0.0281)	-35.01% (0.1827)	-81.32% (0.0084)	-46.30% (0.2212)
Railroads	-34.46% (0.2283)	-20.43% (0.2656)	-21.92% (0.5543)	-49.03% (0.0740)	-18.16% (0.7453)	-50.96% (0.1797)

Table 1: This table shows the results for an analysis of the correlation between the CAPE ratio and the Relative CAPE indicator, respectively, on the one hand and the subsequent annualized ten-year total return on the other hand. We perform this analysis for the overall stock market, the Industrials sector, the Utilities sector, as well as the Railroads sector. The first column reports the correlations depicted in Figures 5, 7, 9, as well as 11, whereas the second and third column change the start date of the analysis, leaving the end date fixed at the end of the second quarter of 2002, which is the last quarter, for which we can compute a ten-year return. The fourth quarter of 1935 refers to the first quarter, at which we use data from S&P resources only, and the fourth quarter of 1962 is the quarter that allows us to single out the final fifty years of the dataset. Each cell lists a correlation number accompanied by a simulation-based p-value in parentheses.

consistent scatter diagram. This is confirmed by the results in Table 1 for Railroads in connection with the Relative CAPE indicator, as the correlations become substantially stronger associated with actually smaller simulation-based p-values, as the sample decreases in size.

5.3 Long-Term Return Differences between Sectors and the Relative CAPE Indicator

Whereas the previous analysis has focused on the interaction between the CAPE ratio or the Relative CAPE indicator of a specific sector and that same sector's long-term return, we will now focus attention on pairwise comparisons across sectors. It is the standardization involved in the construction of the Relative CAPE indicator that enables us to sensibly compare this indicator across sectors. We will therefore address the question whether a sector that appears more undervalued than another sector according to a comparison of the respective Relative CAPE indicators, has higher subsequent long-term returns, and vice versa.

For this purpose, we construct two variables that measure differences across sectors. The first variable is the percentage difference of the Relative CAPE indicators of two sectors, defined as the division of the Relative CAPE indicator of the one sector and the Relative CAPE indicator of the other sector, with 1 being subtracted from the resulting ratio. If, for example, the percentage difference between the Relative CAPE indicator of the Industrials sector and the indicator of the Utilities sector is less than zero, the Industrials sector is relatively more undervalued than the Utilties sector, an assessment that is reversed in case that the percentage difference is bigger than zero. The second variable is the difference of the logarithm of the non-annualized long-term total returns for the two sectors under consideration, where we again choose a time horizon of ten years and consider inflation-adjusted returns. As previously, the time horizon, over which the return is computed, starts in the beginning of the quarter subsequent to the quarter, for which the corresponding Relative CAPE indicators are assessed, providing a genuine separation between valuation and subsequent returns.

Figure 14 focuses on the Industrials and the Utilities sector and displays the association between the percentage difference of the Relative CAPE indicators and the long-term return differences in two different ways. Whereas the top panel shows the time series of both differences, the bottom panel shows a scatter diagram. Both panels start

with the difference in the Relative CAPE indicators at the end of the third quarter of 1902 and end at the end of the fourth quarter of 2002. Correspondingly, the first long-term return is computed between the beginning of the fourth quarter of 1902 and the beginning of the fourth quarter of 1912, whereas the final long-term return captures the period between the beginning of the first quarter of 2003 and the beginning of the first quarter of 2013.

The top panel shows that the difference in the Relative CAPE indicator for the two sectors moves in the opposite direction from the differences in subsequent long-term returns and the scatter diagram confirms the negative correlation. The correlation coefficient for the two difference series for Industrials and Utilities turns out to be almost -40%. Percentage differences of the Relative CAPE indicator clearly account for the biggest differences in long-term returns, with Utilities outperforming as measured by the ten-year returns around 1920 and around 2000 and Industrials outperforming around 1930 and in the late 1980s.

Similar to the simulations addressing the statistical significance of the correlation between the CAPE ratio as well as the Relative CAPE indicator and subsequent long-term returns for the overall market as well as for the sectors *individually*, we have performed simulations for the present situation of cross-sectional predictions. For this purpose, we simulate both sectors' quarterly real, total-return indices simultaneously, matching their respective empirical distributions. Moreover, analogously to the simulation for an individual sector, we impose an AR(1) structure for the natural logarithm of each sectors' CAPE ratio. In addition to matching the mean and the standard deviation of the error terms of the underlying valuation processes, we match the four-by-four covariance matrix arising from the two processes of the sectors under consideration.

The simulated realizations of the innovations provide us with $100,000$ paths for the returns and the valuation ratios. We subsequently build the Relative CAPE indicator as well as the long-term total return index for each of the two sectors to then compute the differences in the way that they are being displayed in the subsequent correlation plots. With regard to the correlation of approximately -40% as displayed in Figure 14, it turns out that this simulation procedure yields an associated p-value of 0.0698, providing fairly definite evidence against the notion that the apparent negative association is compatible with unpredictable sector returns.

Figure 15 is concerned with the pairwise comparison of the Industrials and the Railroads sector, but is otherwise analogous to Figure 14. The correlation between the two time series is also negative, as predicted, at slightly lower than -10%. As in the case of the comparison between the Industrials and the Utilities sector in Figure 14, the bigger oscillations between the Industrials sector and the Railroads sector are well accounted for by the percentage differences of the Relative CAPE indicator. In much of the years of the Great Depression, in the 1930s, the Industrials sector had a higher Relative CAPE indicator than did the Railroads sector and, as our theory predicts, subsequent ten-year returns were lower. Around 1980, the Railroads sector had a relatively higher Relative CAPE indicator. This was the time of the 1980 Staggers Act, which deregulated railroads, and of other actions weakening railroad labor unions. Moreover, in 1995, the railroads' regulator, the Interstate Commerce Department, a 1887 creation of Congress, was eliminated, replaced only with a much weaker Surface Transportation Board. These events manifestly deregulated railroads and got them onto a more profit-oriented trajectory after the mid-1990s. But it appears that the stock market overreacted at the time of the 1980s Staggers Act, so that the Relative CAPE indicator of the Railroads sector was higher than that for the Industrials sector, and indeed, the Industrials sector fared much better in terms of subsequent ten-year returns. After the late 1990s, the increased earnings produced by deregulation finally had their impact by raising the denominator of the CAPE ratio of the Railroads sector, at the same time that widespread enthusiasm for technology stocks and other components of the Industrials sector shifted the CAPE ratio, and the Relative CAPE indicator accordingly, of the Industrials sector way above that of the Railroads sector. The suggested shift in investment from the Industrials sector to the Railroads sector was, as we now know with hindsight, a very suitable step to take at that time.

The simulation-based p-value for this cross-sectional correlation result is quite high at 0.4185, raising concerns that the displayed assocation might very well be compatible with a framework in which returns are completely unforecastable. It will however turn out that this evidence is, as in the subsequent cross-sectional correlation plot involving the Utilities and the Railroads sector, dominated by the earliest part of the sample. Indeed, the notion of unpredictable returns will be convincingly rejected when restricting attention to forecasting cross-sectional returns starting in the 1930s or even starting in the 1960s, which is evidenced by Table 2 below.

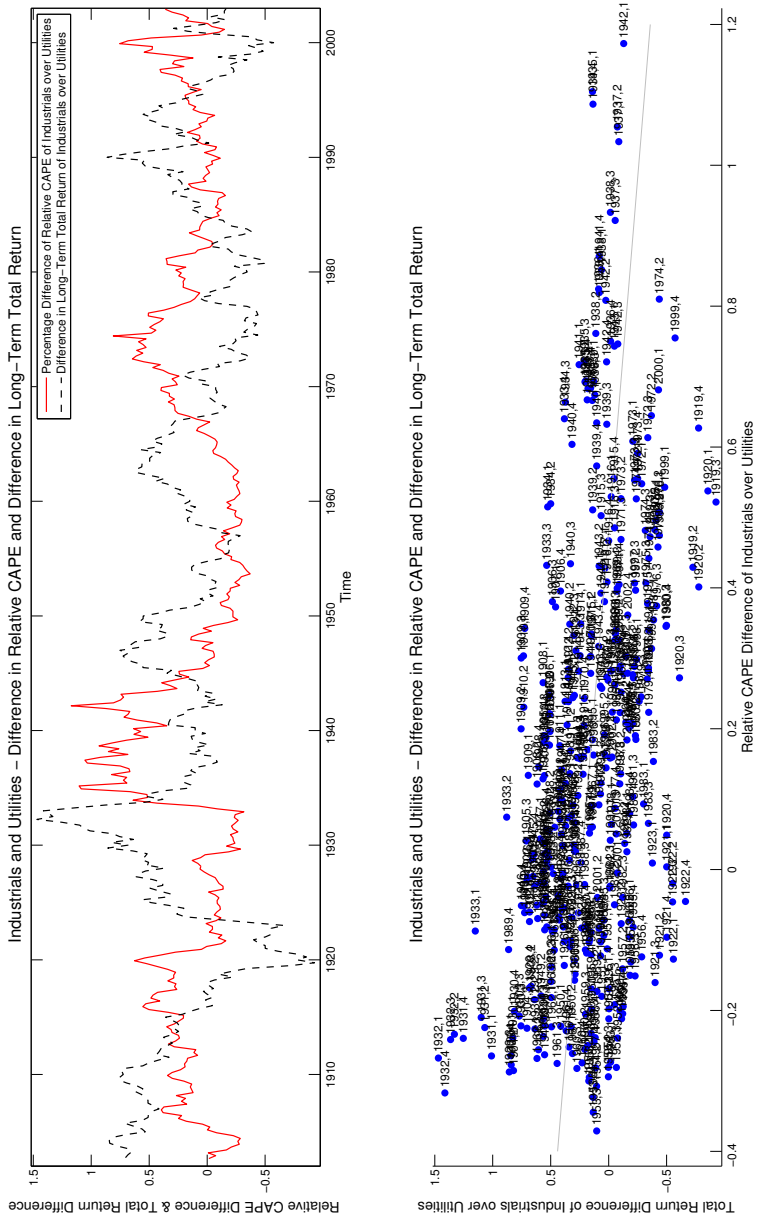


Figure 14: This figure analyzes percentage differences in the Relative CAPE indicator between the Industrials and the Utilities sector and displays their relation with differences in subsequent ten-year returns for the two sectors. The Relative CAPE indicators for the two sectors are quarterly in frequency and their time series starts at the end of the third quarter of 1902. Because of the consideration of subsequent ten-year returns, the last observation of this indicator that is incorporated in this figure is from the end of the second quarter of 2002. We compute the percentage difference between the two indicators as the Relative CAPE indicator for Industrials divided by the indicator for Utilities minus one. In order to bring both time series to a very similar scale, we compute the ten-year return difference as the log-difference of the real total gross return (non-annualized) in the Industrials sector minus the real total gross return (non-annualized) in the Utilities sector, leading to a close approximation to the return difference. The left panel shows the time series of both difference series (dashed black line: return, solid red line: Relative CAPE), whereas the right panel restates the very same information in a scatter diagram.

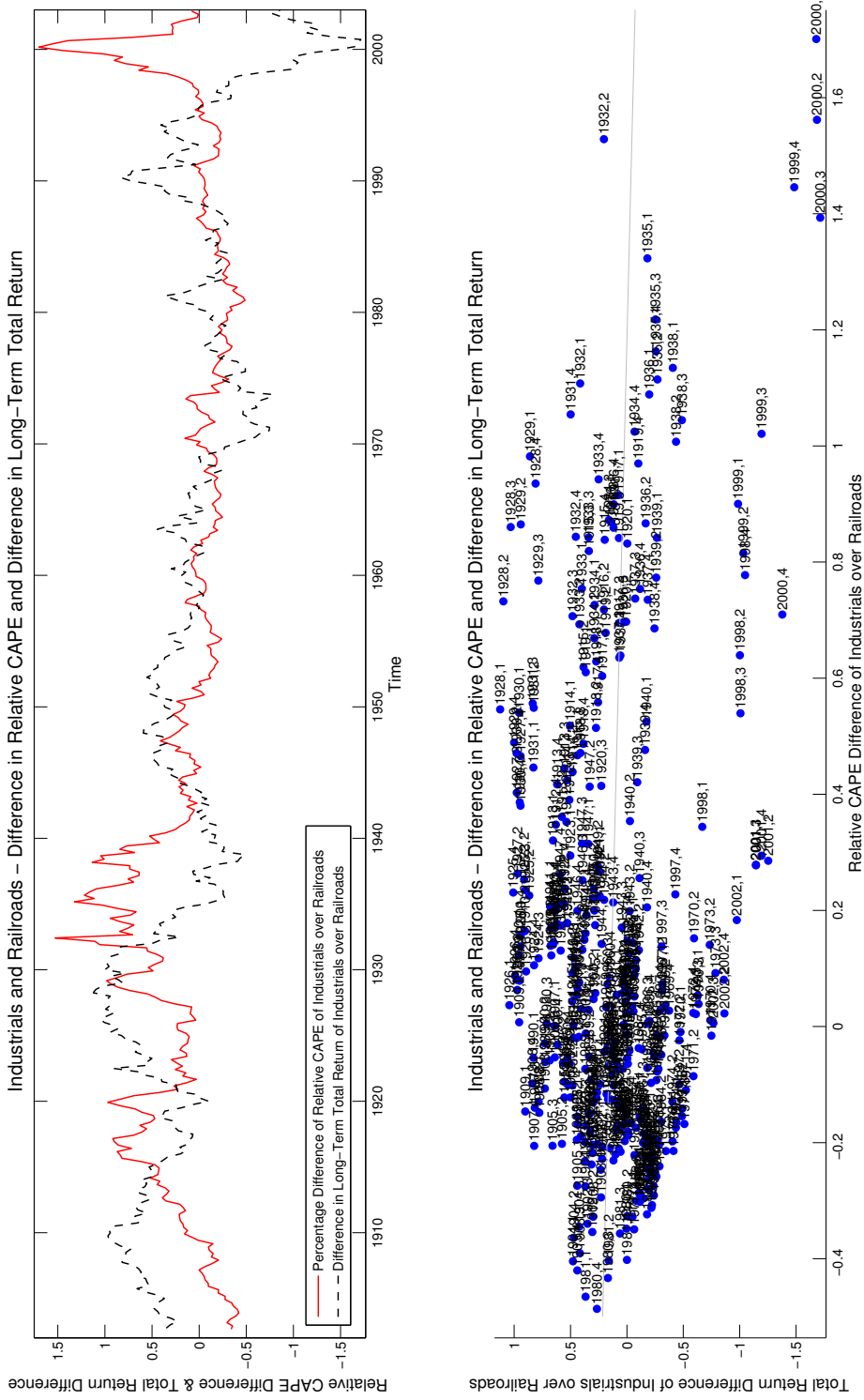


Figure 15: This figure is the analogue of Figure 10, looking at differences between Industrials and Railroads instead of Industrials and Utilities.

Correlation between Differences in Relative CAPE Indicator and Differences in Long-Term Return	Start in Q3 of 1902	Start in Q4 of 1935	Start in Q4 of 1962
Industrials and Utilities	-39.56%	-43.02%	-75.12%
	(0.0698)	(0.1163)	(0.0188)
Industrials and Railroads	-10.30%	-45.66%	-73.16%
	(0.4185)	(0.0989)	(0.0263)
Railroads and Utilities	-3.91%	-31.20%	-70.11%
	(0.5174)	(0.2279)	(0.0392)

Table 2: This table shows the results for an analysis of the correlation between the percentage difference in the Relative CAPE indicator for a pair of sectors and the difference in the logarithm of the non-annualized ten-year total gross return. We perform this analysis for three pairs, Industrials and Utilities, Industrials and Railroads, as well as Railroads and Utilities. The first column reports the correlations depicted in Figures 10 through 12, whereas the second and third column change the start date of the analysis, leaving the end date fixed at the end of the second quarter of 2002, which is the last quarter, for which we can compute a ten-year return. The fourth quarter of 1935 refers to the first quarter, at which we use data from S&P resources only, and the second quarter of 1962 is the quarter that allows us to single out the final fifty years of the dataset.

Figure 16 contains the final pairwise comparison, investigating differences in the Relative CAPE indicator and subsequent long-term returns for the Railroads sector paired with the Utilities sector. The correlation evidence between these sectors is the weakest, only amounting to an overall total of approximately -4% with an associated simulation-based p-value of 0.5174. It however becomes apparent from the top panel of Figure 16 that much of the lack of strength of the overall correlation can be attributed to the first forty years of the sample, in which the Relative CAPE indicator does a poor job of capturing subsequent long-term return differences. The evidence in favor of a negative association between differences in the Relative CAPE indicator and subsequent long-term returns changes in the second half of the 20th century. For example, percentage differences in the indicator capture the long-term return profile in the 1970s and shortly after the beginning of the 1980s very well. Moreover, differences in the indicator convincingly deliver the correct prediction in the 2000s when they identify the Railroads sector as undervalued compared to the Utilities sector.

Given the evidence of Figure 16, it appears that the correlation between the differences in the Relative CAPE indicator for the Railroads and the Utilities sector, when compared to differences in subsequent long-term returns ought to decrease, i.e. become stronger, when looking only at the more recent part of the sample. Table 2 provides this information for the same time periods that have been displayed in Table 1. Indeed, the correlation becomes substantially stronger, decreasing to less than -70% for the comparison between the Railroads and the Utilities sector. This change is however not confined to the comparison between these two sectors. The correlations drop markedly for all three pairwise comparisons, surpassing the -70% mark for each pair, when attention is limited to only the last fifty years of the sample. Marked drops in the simulation-based p-values accompany the decreasing trend in correlations, as the sample period restricts attention more and more on the most recent time period. The marked drops in p-values show that the increase in the strength of the correlation between differences in the Relative CAPE indicator and subsequent long-term return differences is much too strong to be solely attributed to a decrease in the length of the underlying sample. Focusing on the last fifty years of the sample, there is therefore very convincing evidence that the degree to which differences in the Relative CAPE indicator are capable of predicting differences in long-term returns fundamentally rejects the notion that returns - and by extension differences in returns - are unpredictable.

6 Predicting the Long-Term Growth of Ten-Year Average Earnings

Starting from the evidence presented in Figure 4 that the CAPE ratio for the overall stock market as well as for the three sectors Industrials, Utilities, and Railroads is mean-reverting, one might formulate alternative hypotheses about the predictive ability of the CAPE ratio. Instead of the hypothesis that it is the numerator of the CAPE ratio that

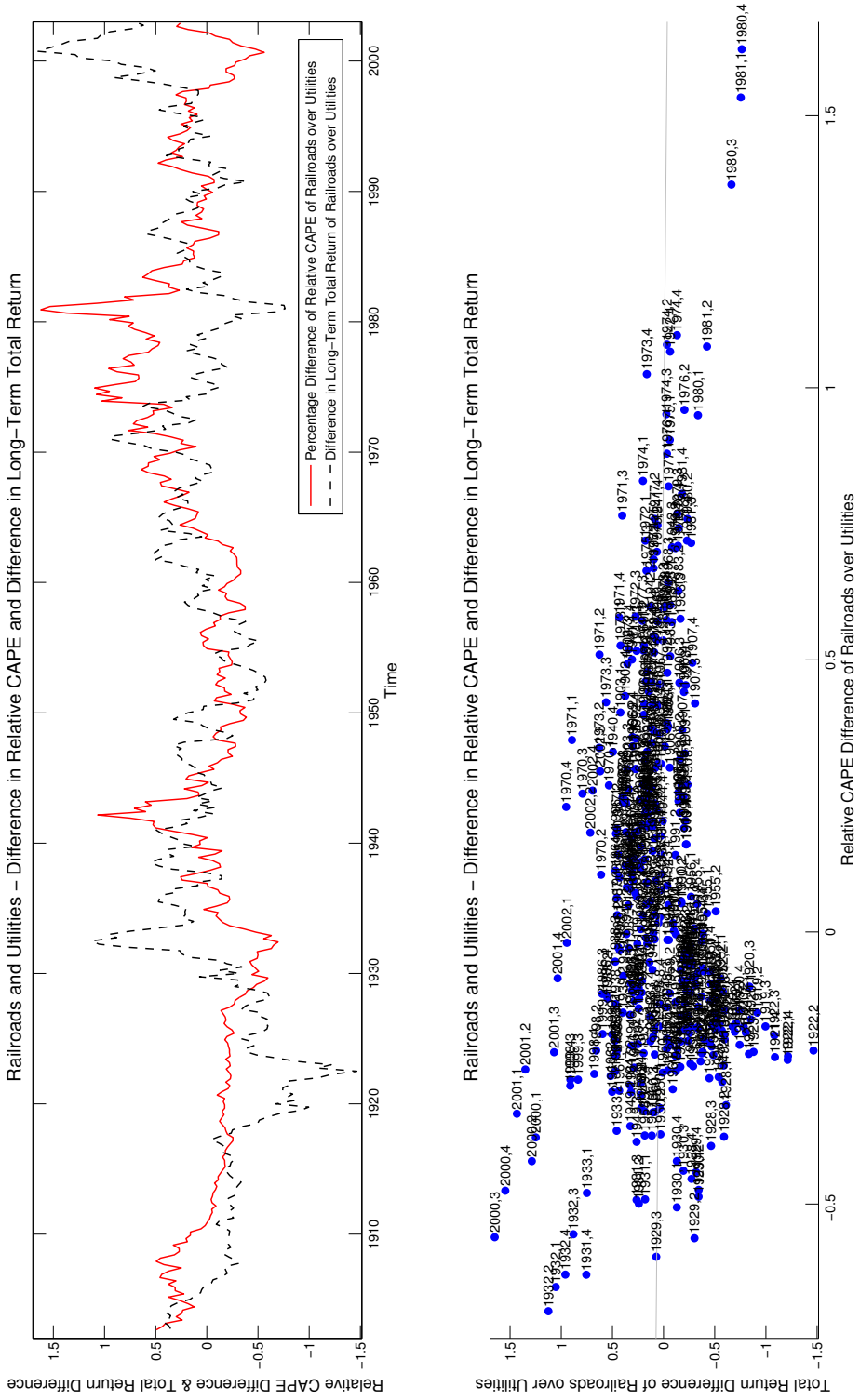


Figure 16: This figure is the analogue of Figure 10, looking at differences between Railroads and Utilities instead of Industrials and Utilities.

brings the CAPE ratio back to its mean, which translates into the predictive ability of the CAPE ratio for subsequent (long-term) returns, one might postulate that it could also be the numerator of the ratio - the ten-year average of inflation- as well as total-return-adjusted earnings - that ensures reversion to the mean. Correspondingly, this would translate into predictability of the CAPE ratio for the (long-term) change in this long-term earnings average. As the lack of return predictability rests at the heart of the Efficient Markets Hypothesis, these kinds of predictability patterns in subsequent earnings would be the natural explanation for any variation in the CAPE ratio.

On the level of the overall stock market, this argument has also been presented in Campbell and Shiller (1998, 2001), who start out from the implications of the premise of the mean-reversion-property of valuation ratios. For both the dividend-price ratio as well as their version of the CAPE ratio, they present evidence against the notion of earnings predictability and in favor of return predictability. It is the purpose of this section to extend the analysis in Campbell and Shiller (1998, 2001) and complement the convincing evidence from Section 4 for the ability of the CAPE ratio to forecast long-term real total returns by disproving the notion that the CAPE ratio has a strong impact on the subsequent growth in the long-term average of earnings. Starting from the mean-reversion-property of the CAPE ratio, the predicted sign of the association is positive, i.e. the hypothesis is that low values of the CAPE ratio, for example, predict negative growth in the long-term average of earnings, and vice versa.

Practically, we use the time series of earnings averages used in the construction of the CAPE ratio, as outlined in Section 4. As earnings are only available annually until the middle of the 1930s, we restrict the time series under consideration to start with the ten-year average ending in the first quarter of 1936 in order to ensure sufficient quarterly variation in the long-term average of earnings. The particular observation at the end of the first quarter of 1936 is the average used for the CAPE ratio at the end of the third quarter of 1936 because of the earnings lag to take into account corporate earnings reporting. We then caculate the annualized ten-year change in this ten-year average.[65] Each pair of observations to assess the predictive ability of the CAPE ratio for subsequent earnings growth uses a given observation of the CAPE ratio together with the subsequent ten-year growth rate of the numerator of this very observation of the CAPE ratio. All scatters thereby start with the CAPE ratio at the end of the third quarter of 1936 and end with the CAPE ratio at the end of the fourth quarter of 2002.

The simulation from Section 4 that allowed us to test the degree to which the scatter evidence is compatible with the notion of unpredictable returns will also enable us to calculate simulation-based p-values for the association between the CAPE ratio and subsequent (long-term) growth rates in 10-year average earnings. For each pair of the simulated logarithm of the simulated real total-return index and the simulated logarithm of the CAPE ratio, we obtain the corresponding simulation of 10-year average (real and total-return adjusted) earnings by taking the difference. From that, we can calculate the ten-year change and calculate a distribution for the correlation of the CAPE ratio and long-term changes in the 10-year average earnings, appropriately truncating the time series for the CAPE ratio and the earnings growth rate to reflect the shortened underlying time series. Importantly, the simulated p-value now refers to the probability mass of the distribution that exceeds the underlying correlation, as our conjectured association between the CAPE ratio and earnings growth is positive. Finally, as in Section 4, we can extend the correlation analysis to the Relative CAPE indicator, by simply replacing the CAPE ratio by this indicator, as outlined above.

Figure 17 displays the connection between the CAPE ratio and subsequent long-term earnings growth for the overall stock market. The correlation coefficient is -14.55% in the left panel, which focuses on the CAPE ratio. This correlation actually has the wrong sign to be compatible with the notion that it is the denominator of the CAPE ratio that ensures the mean-reversion property of the valuation ratio. Moreover, the simulation-based p-value derived under the assumption that returns are unpredictable is 0.9993, i.e. there are barely any iterations of the simulation that yield a correlation as small as the underlying data. In conjunction with the evidence from the return scatters from Section 4, the evidence convincingly points to the numerator of the CAPE ratio, and not the denominator, as the predominant force behind the mean-reversion of the CAPE ratio for the overall stock market.

The right panel of Figure 17, using the Relative CAPE indicator instead of the CAPE ratio, shows a very similar pattern to the left panel. The correlation is -4.58% and again has the wrong sign. The associated simulation-based

[65] For the sake of completeness, we would like to point out again that, as explained in Section 4 and Appendix B, the time series of averaged earnings used in the denominator of the CAPE ratio at the transition from annual to quarterly earnings actually uses slightly varying time horizons for the earnings average, ranging from 10 years to ten 10.75 years. We will however ignore this complication subsequently and refer to the denominator of the CAPE ratio as a ten-year average.

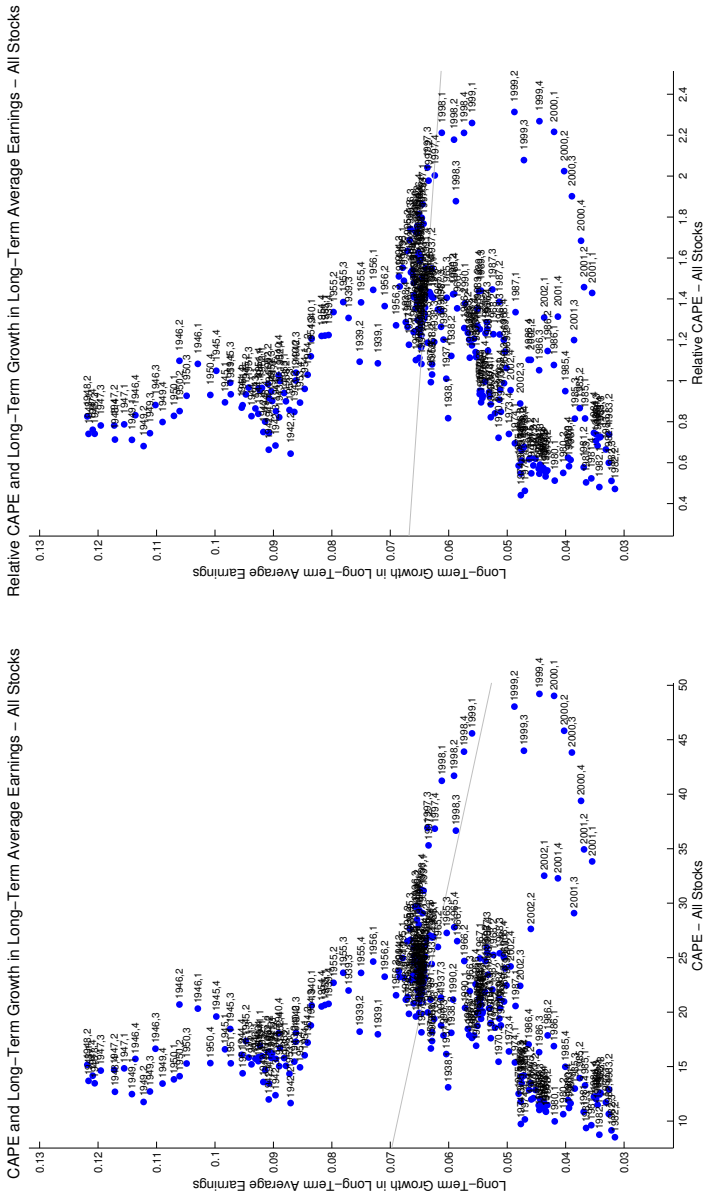


Figure 17: The left panel depicts a scatter diagram of the CAPE ratio (discussed in Section 4) of the overall stock market and the right panel depicts the analogous scatter using Relative CAPE (discussed in Section 5). On the x-axis of the left panel, we display the CAPE ratio of the overall market, incorporating the transition to real numbers and to total return numbers. It is measured at the end of each quarter starting in the third quarter of 1902 until the fourth quarter of 2002. The x-axis of the right panel captures the Relative CAPE of the overall market, using the same timing convention as for CAPE ratio. The y-axis, which is common to both panels, captures the annualized ten-year growth rate of the denominator of the CAPE ratio, which in principle is a ten-year average of the 12-month trailing sum of earnings, as measured from the current observation of the CAPE ratio to the corresponding CAPE ratio ten years later. For details with regard to the actual calculation of the denominator of the CAPE ratio, please refer to Section 4 as well as to Appendix B. We have labeled each point in the scatters by the time, at which the initial CAPE ratio or the Relative CAPE indicator, respectively, at the beginning of the ten-year period is computed is computed. Each entry contains the year followed by the number of the quarter.

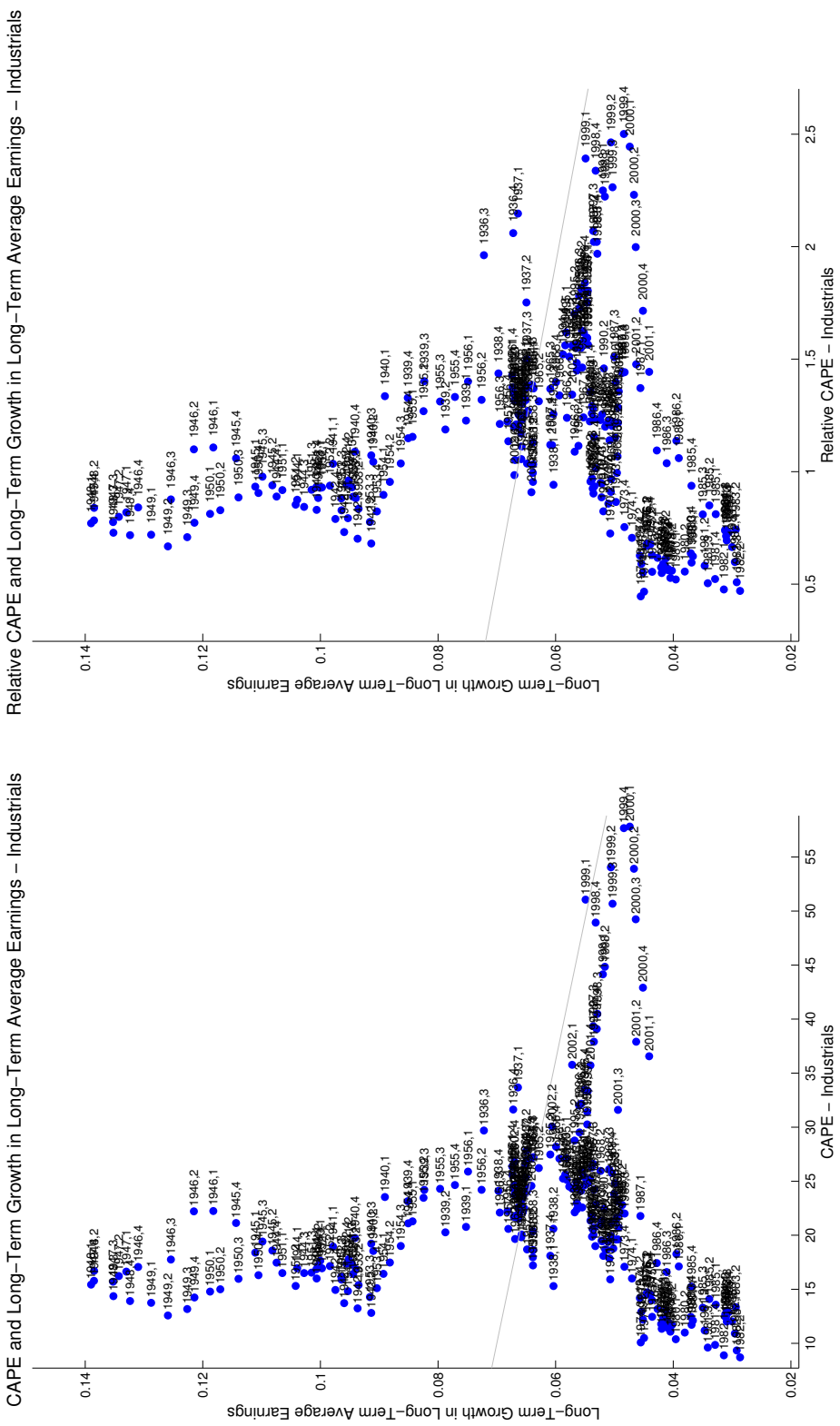


Figure 18: This figure is the analogue of Figure 18 for the Industrials sector instead of the overall stock market.

p-value is 0.9929.

Figure 18 displays scatters of valuation ratios and subsequent long-term growth in the ten-year average of earnings for the Industrials sector. It is again the case both the correlation for the CAPE ratio in the left panel as well as the Relative CAPE indicator in the right panel are negative at -12.82% and -11.69%, respectively, contrary to the hypothesized sign. Moreover, there is again considerable evidence against the notion of earnings predictability from the simulation-based p-values, which amount to 0.9992 and 0.9958, respectively.

When looking at the Utilities sector, it becomes evident that the correlation for both the scatter involving the CAPE ratio as well as the Relative CAPE indicator is in the hypothesized direction. The correlations equal 36.85% in the panel for the CAPE ratio and 26.22% on the right hand side that uses the Relative CAPE indicator. However, even though their sign is consistent with the predictability and their absolute value is sizeable, it turn out that the associated simulation-based p-values of the correlations are 0.8328 for the left panel and 0.8952 for the right panel, preserving the consistency of the evidence against the notion that the mean-reversion of the CAPE ratio arises from the earnings average in the denominator.

For the Railroads sector as the final sector, the correlation pattern is curious, as the sign flips from positive for the CAPE ratio to negative when applying the transformation to the Relative CAPE indicator. The correlation values are 21.14% for the CAPE ratio and -12.60% for the Relative CAPE indicator, with associated simulation-based p-values of 0.9512 and 0.9962, again rejecting the notion of the predictive ability of the CAPE ratio for subsequent earnings growth rates.

Given the degree to which the correlation results change in Sections 4 and 5 when one considers a shorter time horizon, it is worthwhile reconsidering the association between the CAPE ratio and the Relative CAPE indicator with the long-term growth rate of ten-year average earnings when limiting attention to the last fifty years of the time period under consideration.[66]It turns out that this limitation causes the correlation pattern for the overall stock market, for the Industrials sector, as well as for the Railroads sector to become much more consistent with the conjecture that the CAPE ratio as well as the Relative CAPE indicator are positively associated with subsequent growth in 10-year average earnings. Explicitly, the correlations are 40.10% for the CAPE ratio and 59.67% for the Relative CAPE indicator in the case of the overall stock market, 50.69% and 54.83%, respectively, for the Industrials sector, and 52.73% and 14.33%, respectively, in the case of the Railroads sector. However, the simulation-based p-values continue to speak against a strong and considerable notion of earnings growth predictability, as they amount to 0.6426 for the CAPE ratio and 0.4024 for the Relative CAPE indicator for the overall stock market, 0.5402 and 0.4970, respectively, for the Industrials sector, as well as 0.4495 and 0.8861 for the Railroads sector. Curiously, the correlation for the Utilities sector, which has had the predicted sign when looking at the longer sample, is now negative at -0.0442 for the CAPE ratio and -0.3950 for the Relative CAPE indicator, which translates into p-value of 0.9693 and 0.9977, respectively.

Overall, it becomes very evident that the question whether it is the numerator (prices) or the denominator (10-year average earnings) that ensures the mean-reversion of the CAPE ratio can clearly be answered un favor of the numerator, therefore establishing the notion of return predictability for the CAPE-ratio not only for the overall stock market, as already analyzed for data up to the 1990s by Campbell and Shiller (1998, 2001), but also for the three sectors Industrials, Utilities, and Railroads.

7 Sensitivity Study for the CAPE-Based Correlation Analyses

When analyzing the predictive ability of the CAPE ratio as well as the Relative CAPE indicator for subsequent long-term returns, we have fixed the return-horizon, over which we consider subsequent returns, to ten years. While this is certainly motivated by the notion that the CAPE ratio by its very definition is supposed to capture the long-term over- or undervaluation of the overall stock market or a constituent sector, the consideration of ten-year returns might appear ad-hoc.

[66]This limitation corresponds to the third scenario analyzed above in Sections 4 and 5. Observe that the overall horizon analyzed for the correlation analysis involving earnings vaguely corresponds to the second scenario, as we cannot extend these correlation analyses farther back because of the annual reporting frequency of earnings.

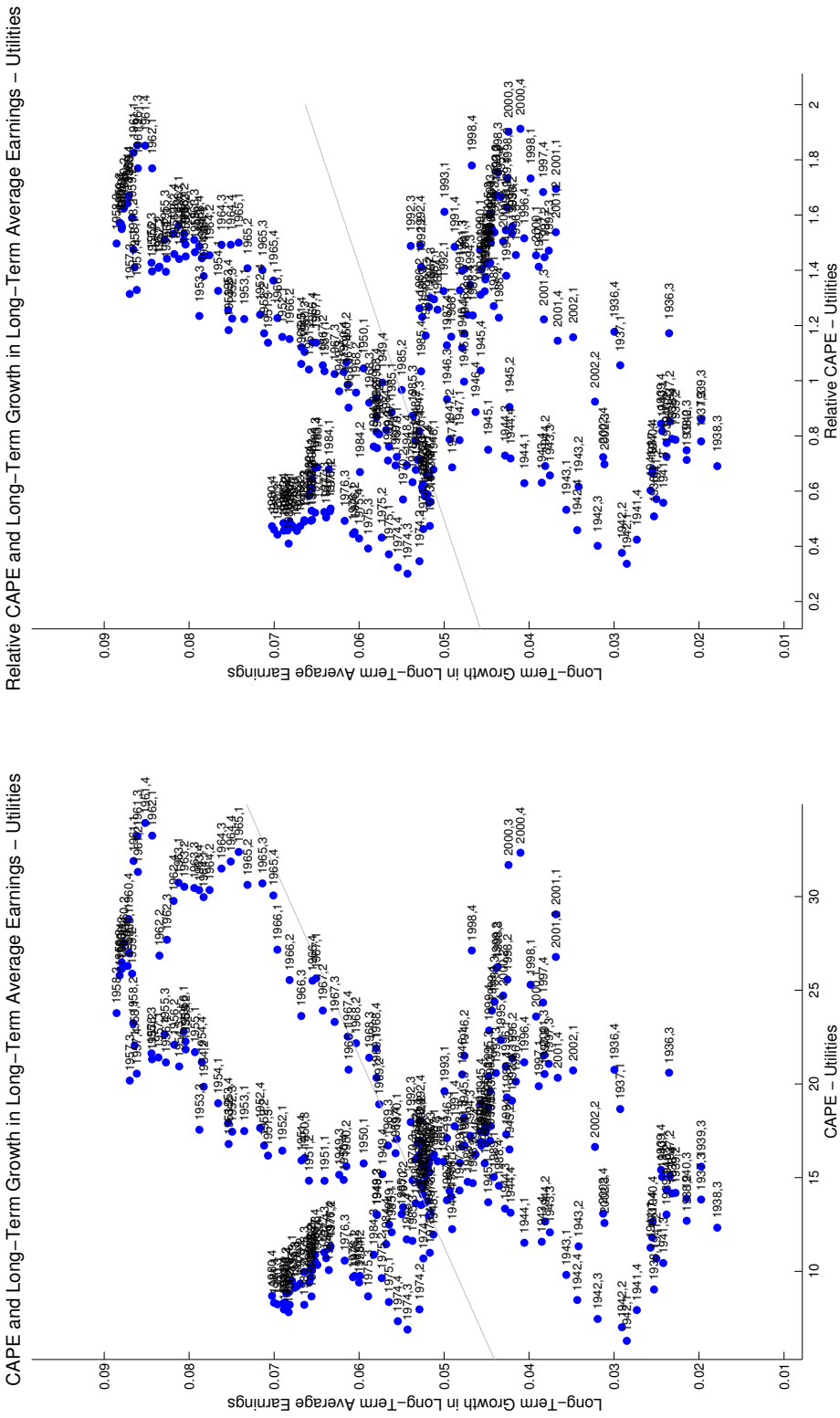


Figure 19: This figure is the analogue of Figure 5 for the Utilities sector instead of the overall stock market.

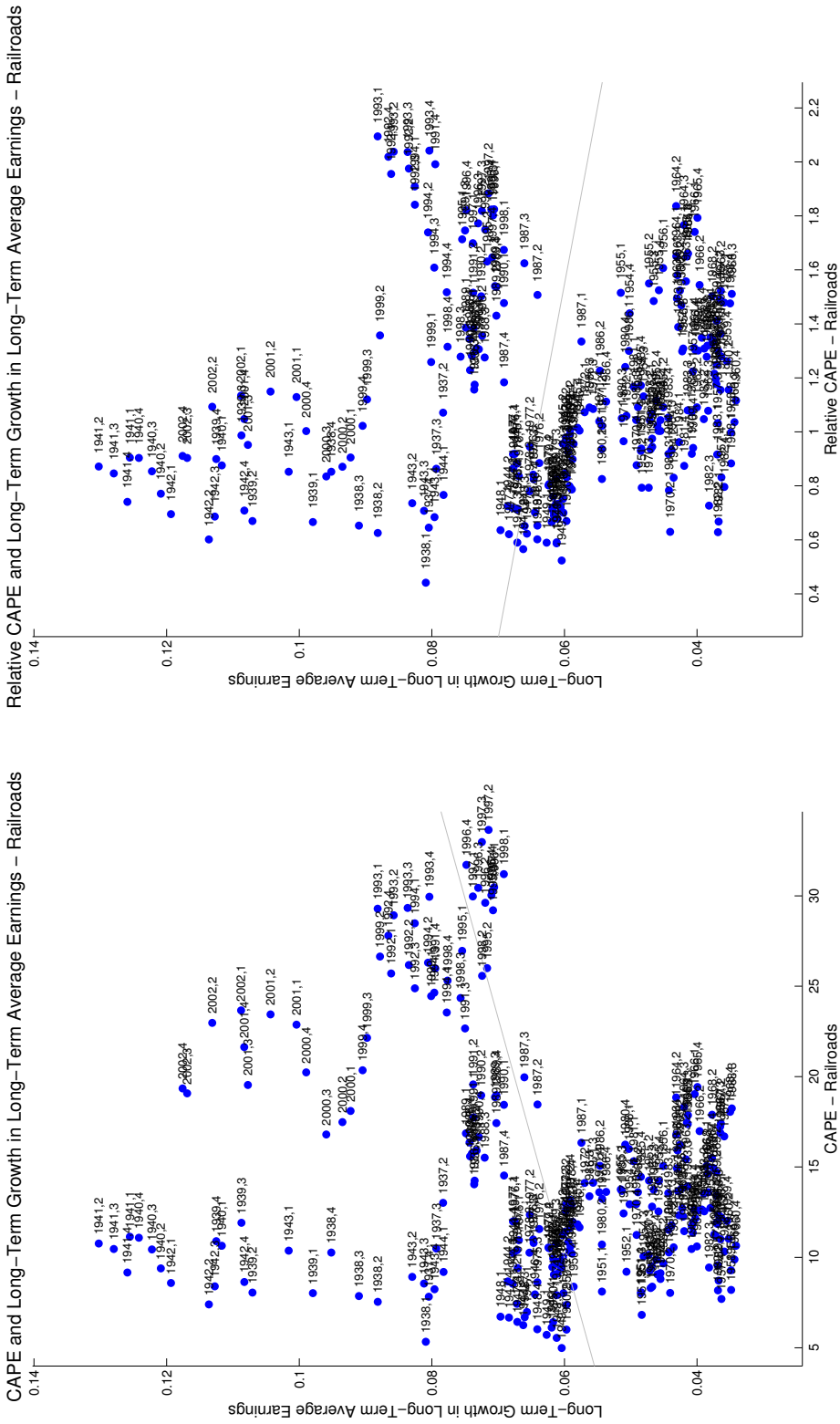


Figure 20: This figure is the analogue of Figure 5 for the Railroads sector instead of the overall stock market.

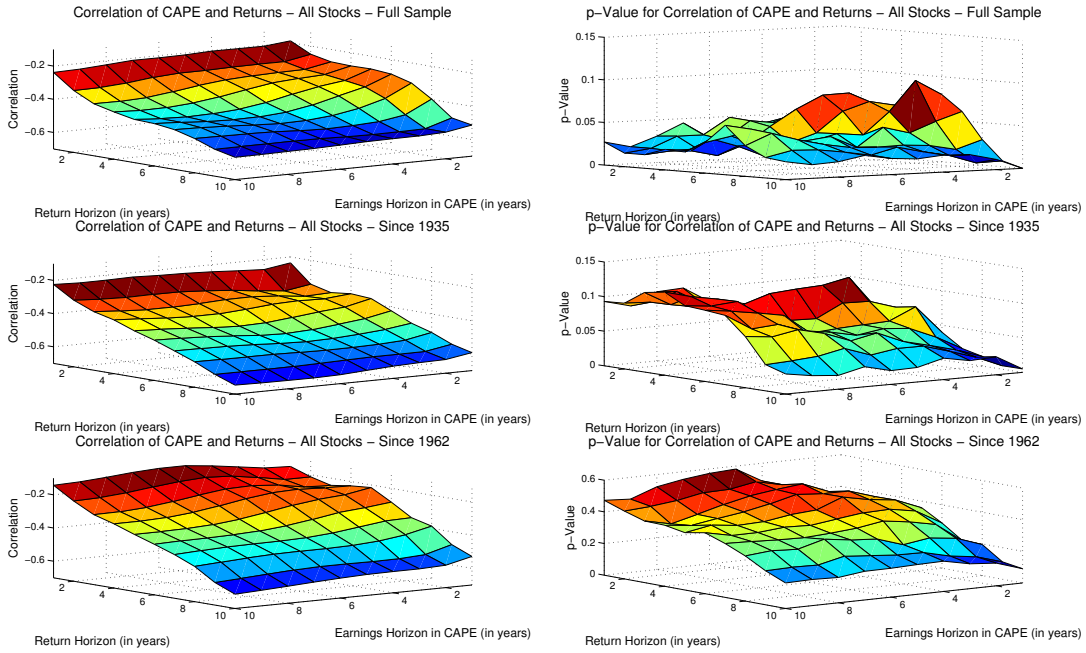


Figure 21: This figure displays a sensitivity study of the correlation analysis as displayed in the left panel of Figure 5. The horizon for subsequent returns varies between 1 and 10 years in annual increments (left axis), which is the same variation imposed on the horizon for the average of earnings in the denominator of the CAPE ratio (right axis). 10 years for the earnings horizon correspond to the CAPE ratio as defined above, whereas 1 year leads to the more classic price-earnings ratio. Except for the differences in horizons, the procedure to calculate the long-term returns as well as the variant of the CAPE ratio is as described in Sections 3 and 4. The full sample time series for the CAPE ratio in each scenario homogenously starts in the third quarter of 1902 and ends in the fourth quarter of 2002 and subsequent returns are considered correspondingly. The resulting correlation as well as the associated p-values are displayed in the top two panels. The calculation for the simulation-based p-values follows the description in Section 4 with minor adaptations to reflect the changes in horizons. The middle and the bottom pair are analogous to the top pair, but restrict the sample under consideration to the time starting with the fourth quarter of 1935 and the fourth quarter of 1962, which are the same starting points as in Section 4.

Moreover, the very definition of the CAPE ratio also involves a choice on a time horizon, which is the horizon for the calculation of the earnings-average in the denominator of the valuation ratio. While the current definition is certainly in line with the way that the ratio is used in Campbell and Shiller (1988, 1989, 1998, 2001) and lies at the upper end of the spectrum for longer-term earnings to be used in conjunction with prices in valuation exercises as advcated by Graham and Dodd (1934),[67] the use of ten-yar averages for earnings is nevertheless an ad-hoc assumption.

In order to analyze the impact of these two assumptions on (a) the correlation between the CAPE-ratio as well as the Relative CAPE indicator and subsequent returns for the overall stock market and the three individual sectors and (b) the correlation between differences in the Relative CAPE indicator and differences in subsequent returns when one takes a cross-sectional perspective, we perform a sensitivity study for these correlation analyses. For this purpose, we vary both the subsequent return horizon as well as the horizon, over which we compute the average of earnings in the denomiator of the CAPE ratio, between 1 and 10 years in annual increments. Except for these differences in the earnings horizon, we excatly follow the procedures for the calculation of the CAPE ratio and of the Relative CAPE indicator as outlined in Sections 4 and 5, respetively. Moreover, we preserve the notion of inflation-adjusted returns that have been made invariant to corporate payout policy through the transition to total returns as detailed in Section 2. Finally, we guarantee the comparability of the correlation results across different horizon pairs by truncating all samples used for the calculation of the correlations to have a first CAPE[68] observation (for the ratio

[67] See the respective quote in the introductory paragraphs.

[68] Even though the CAPE ratio is formally defined in terms of ten-year average earnings, we will for ease of reference refer to the ratio originating from all variations in the underlying earnings horizon as the CAPE ratio.should

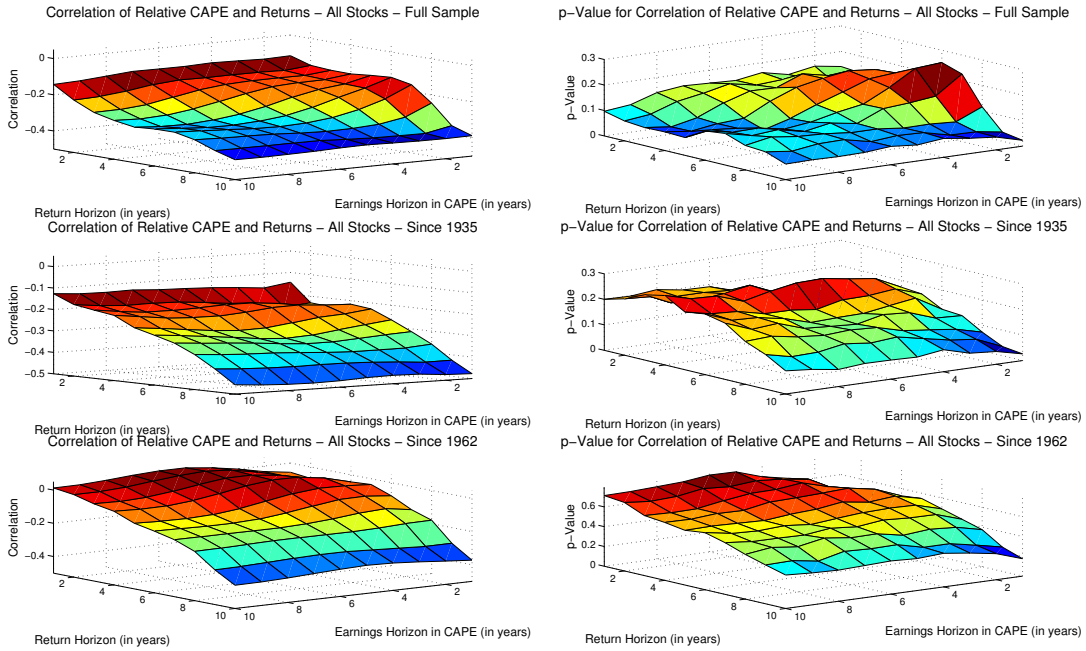


Figure 22: This figure is analogous to Figure 21, but it employs the Relative CAPE indicator instead of the CAPE ratio for the sensitivity study of the correlation involving the overall stock market.

as well as for the Relative CAPE indicator) at the end of the third quarter of 1902 and the last one in December 2002. Correspondingly, subsequent returns begin in the beginning of the fourth quarter of 1903 and the last return observation originates from the beginning of the first quarter of 2003.

Observe that the original analysis from Sections 4 and 5 corresponds to the pair (10,10) for the return horizon and the earnings horizon. Moreover, the consideration of only a single year for the average of earnings provides a version of the more classic price-earnings and allows us to directly compare the price-earnings ratio against the CAPE ratio using a ten-year average of earnings.[69] Given the additional variation in the horizon for returns, we cannot only perform this comparison for a single long-term horizon, but for a whole range of horizons ranging from rather short-term (1 year) to long-term (10 years). In the figures below detailing the results of the sensitivity analysis, we initially focus on the relation between the variants of the CAPE ratio and subsequent returns for the overall stock market as well as for the Industrials, Utilities, and Railroads sector and display the correlation numbers as well as their associated simulation-based p-values. We obtain these p-values by repeating the simulation methodology introduced in Section 4 in connection with the correlation analysis involving the CAPE ratio and in Section 5 when analyzing correlations involving the Relative CAPE indicator, with only minor modifications reflecting the changes in the horizons for returns and the averages or earnings. Finally, we display all results for three choices of the time horizon, the full sample since 1902, the sample since the fourth quarter of 1935, and the final fifty years of the sample. These are exactly the same scenarios as already considered in Tables 1 and 2.

Figure 21 contains the sensitivity study for the correlation between variants of the CAPE ratio and subsequent returns for the overall stock market. It is indeed the point (10,10), which corresponds to the original correlation analysis as captured by Figure 5, that displays the most negative correlation. Both increases in the return horizon for subsequent returns as well as shorter-term averages for the earnings entering the denominator of the CAPE ratio lead to increases in correlation, i.e. a less pronounced association of the respective variant of the CAPE ratio with subsequent returns. Focusing on the comparison between the price-earnings ratio (earnings-horizon of 1 year) and the formal CAPE ratio (earnings-horizon of 10 years) for the full sample, the evidence for the overall stock market

[69] As we apply the inflation- as well as the total-return adjustment for earnings at the already aggregated level of 12-month trailing earnings, it will be the case that both adjustments nearly cancel out when the horizon for the average of earnings is only a single year, as analogous adjustment operations apply to both the numerator as well as the denominator of the CAPE ratio. The fact that the operations do not cancel ou texactly is due to the six-month lag that we impose on earnings.

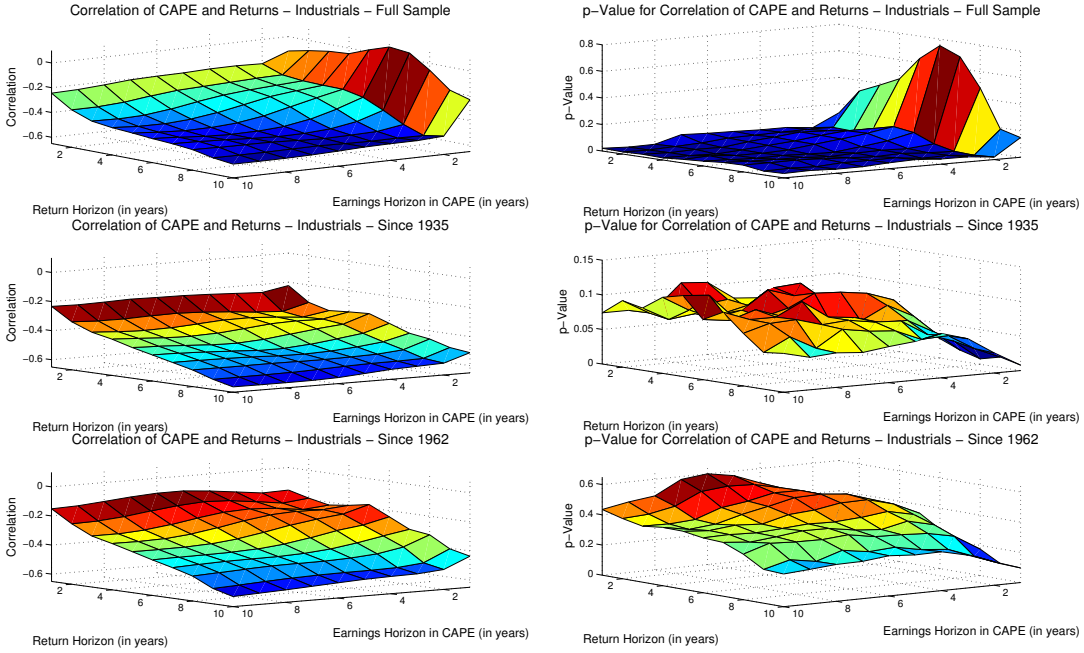


Figure 23: This figure is the analogue to Figure 21 for the Industrials sector instead of the overall stock market.

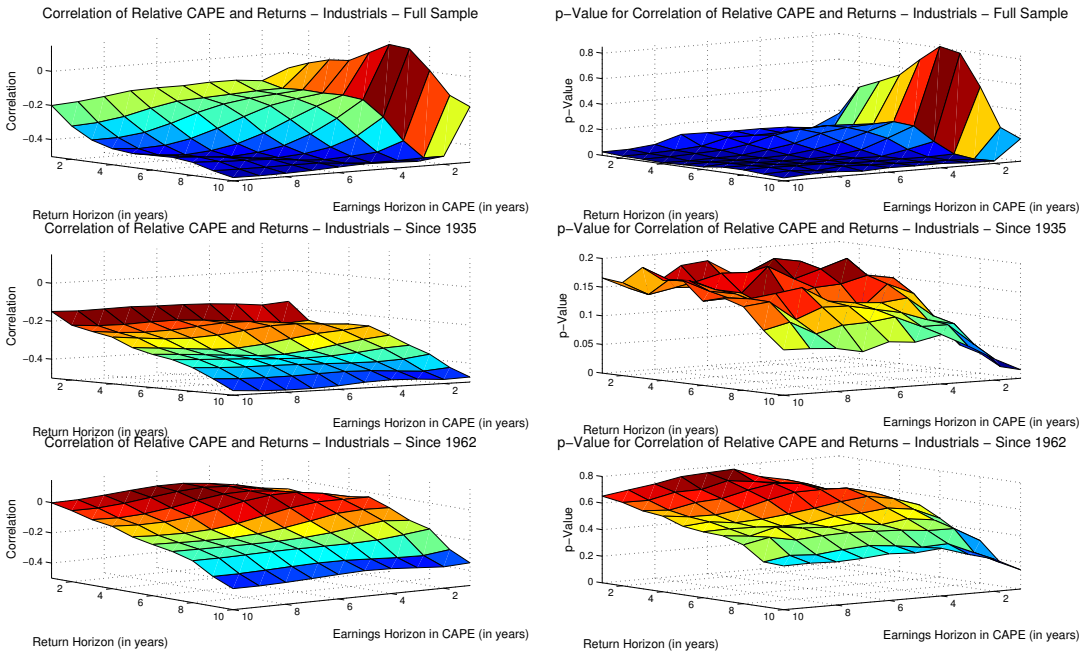


Figure 24: This figure is analogous to Figure 22, but it employs the Relative CAPE indicator instead of the CAPE ratio for the sensitivity study of the correlation involving the Industrials sector.

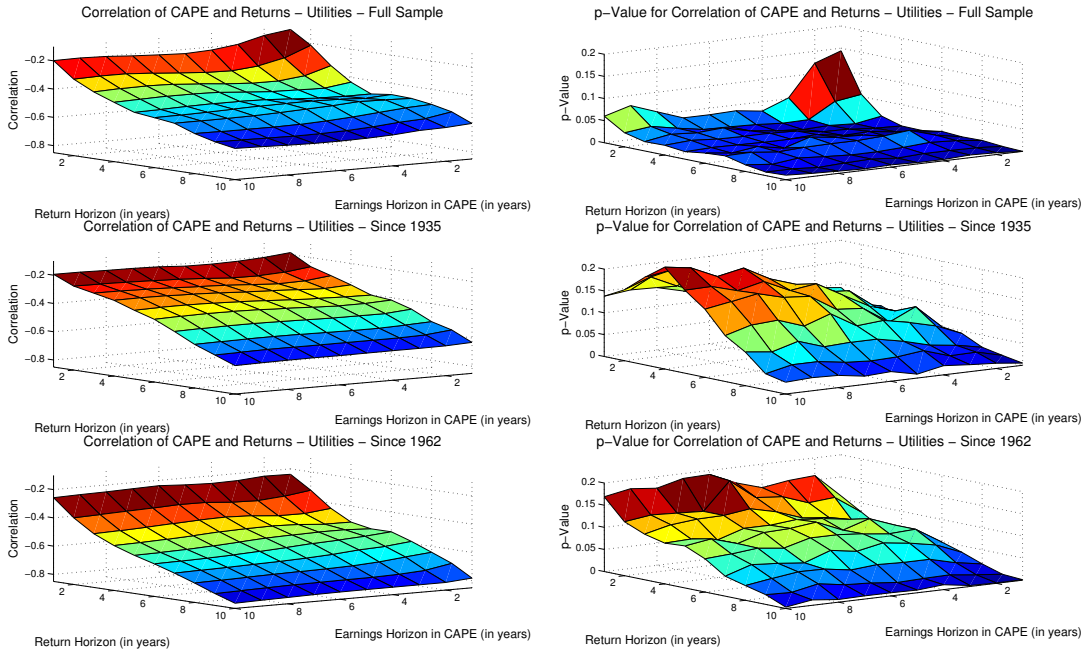


Figure 25: This figure is the analogue to Figure 21 for the Utilities sector instead of the overall stock market.

clearly underlines the appropriateness of the long-term averaging of earnings, as the correlations for the CAPE ratio are consistently lower than those for the price-earnings ratio for all return horizons under consideration. The associated simulation-based p-values for the full sample reject the notion that the increased strength in the negative association between the CAPE ratio and subsequent returns for longer return horizons is solely an artefact of statistical spuriousness, as the p-values remain nearly constant for varying return horizons, particularly for the original CAPE ratio (earnings horizon of 10 years). Moreover, the correlations for the original CAPE ratio are not only stronger, but they also exhibit lower p-values than the price-earnings ratio for the full sample. This does not necessarily hold for the shortened samples, but - as evidenced by the two bottom right hand side panels - the statistical significance of the correlation results for the overall market suffers rather substantially in general, which causes us to discount the seemingly lower p-values of the price-earnings ratio for the shorter samples.

If instead of the CAPE ratio, one considers the Relative CAPE indicator's correlation with subsequent returns, as displayed in Figure 22, the evidence in terms of the correlation for the full sample is very similar to the evidence presented in Figure 21 for the CAPE ratio. Moreover, the improvement in the p-values of the original CAPE ratio when compared to the price-earnings ratio is quite substantial when one looks across all return horizons jointly. For the shortened samples in the bottom panels, the evidence for the strength of the negative association between the valuation measure and subsequent returns persists, whereas the difference across horizons for the earnings average becomes less pronounced. The latter effect is however less problematic given the substantially increased p-values for the shortened sample, as also evidenced by Figure 21.

Figures 23 and 24 present the results of the sensitivity study for the Industrials sector. In the context of this sector, it is particularly noteworthy that the correlation evidence for the price-earnings ratio (earnings horizon of 1 year) for both the CAPE ratio as well as the Relative CAPE indicator is dramatically worse than for earnings horizons bigger than a year. However, even beyond the one-year horizon, there is steady decline in the correlation as one uses longer averages for the earnings in the denominator of the CAPE ratio, particularly in the surface plot involving the Relative CAPE indicator in the full-sample-scenario.

In the case of the Utilities sector, presented in Figures 25 and 26, there is substantial evidence on how the association between variants of the CAPE ratio as well as of the Relative CAPE indicator strengthens, as the horizon for subsequent returns increases. In the sensitivity analyses for the CAPE ratio, these higher degrees of association are even so strong, as to overcome the shortening of the underlying sample and still to produce very low p-values.

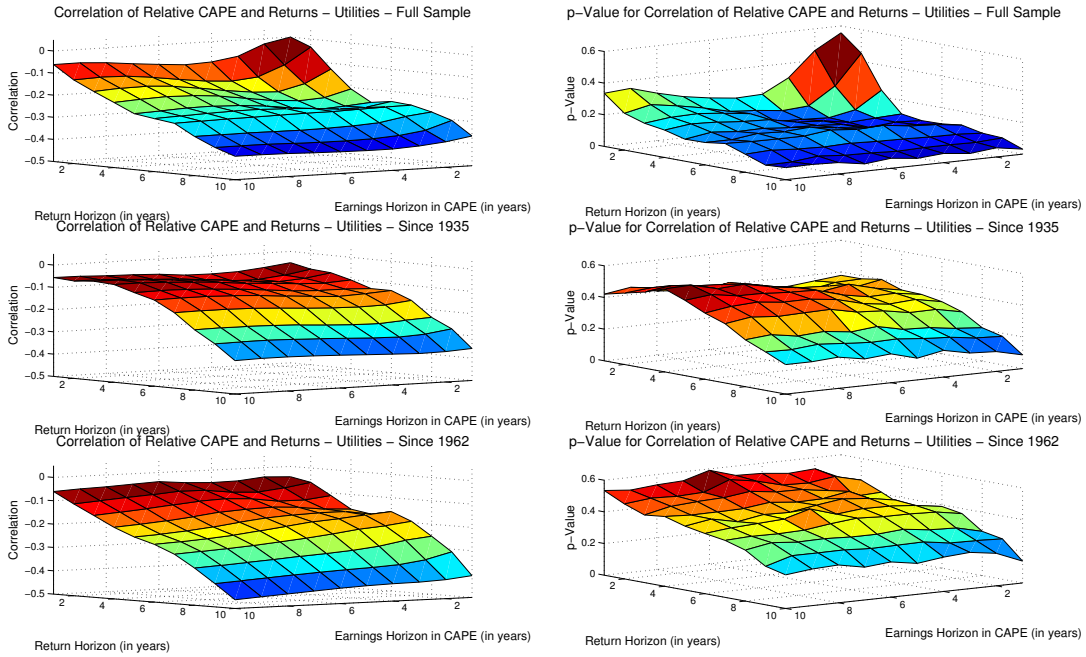


Figure 26: This figure is analogous to Figure 25, but it employs the Relative CAPE indicator instead of the CAPE ratio for the sensitivity study of the correlation involving the Utilities sector.

With regard to changes in the earnings horizon for the denominator of the CAPE ratio, there is evidence for the superiority of the CAPE ratio when compared with the price-earnings ratio, particularly for the full sample and the scenario of only the most recent fifty years of the sample (since 1962), although less pronounced than for the overall stock market and the Industrials sector.

In contrast, the drop from shorter earnings averages to longer averages and ultimately a ten-year average is quite pronounced for the Railroads sector for both the CAPE ratio as well as the Relative CAPE indicator, as displayed by Figures 27 and 28, respectively. This is especially true for longer return horizons, although some caution is necessary to assess te results for longer return horizons, as it is generally not the case that increase in the strength of the negative association between either the CAPE ratio or the Relative CAPE indicator and subsequent returns outweighs the more pronounced statistical issues, which is why the Railroads sector generally shows increases in the associated simulation-based p-values, as the return horizon increases toward ten years.

Subsequent to the sensitivity study on the level of the overall stock market and on the level of individual sectors, we now turn attention to the kind of cross-sectional analysis that we have performed in Section 5, that is the correlation between percentage differences in the Relative CAPE indicator compared to logarithmic differences in subsequent returns for pairs of sectors. Except for minor adaptations to reflect the changes in the horizons for returns and the average of earnings, which will enter the CAPE ratio and subsequently the Relative CAPE indicator for a given sector, we follow the methodology outlined in Section 5 for this part of the sensitivity study. The same also applies to the calculation of the simulation-based p-values. Moreover, we look at a uniform sample across all instances of the sensitivity study by focusing on samples for the Relative CAPE indicator from the end of the third quarter of 1902 until the end of the fourth quarter of 2002. We then consider returns from the beginning of the subsequent quarter until the specified return horizon correspondingly and calculate the logarithmic difference between the non-annualized returns of two sectors. As in the previous sensitivity studies, we display three scenarios, the full sample since 1902, the sample since 1935, and finally the final fifty years worth of data starting in 1962.

Considering the comparison between the Industrials sector and the Utilities sector, it first becomes apparent that the correlation value at $(10, 10)$, corresponding to the original setting from Section 5, remains the lowest, even across the different time horizons under consideration. An interesting pattern arises in the comparison of these three time horizons. Whereas the full sample already shows that longer return horizons and also that longer-term averages for

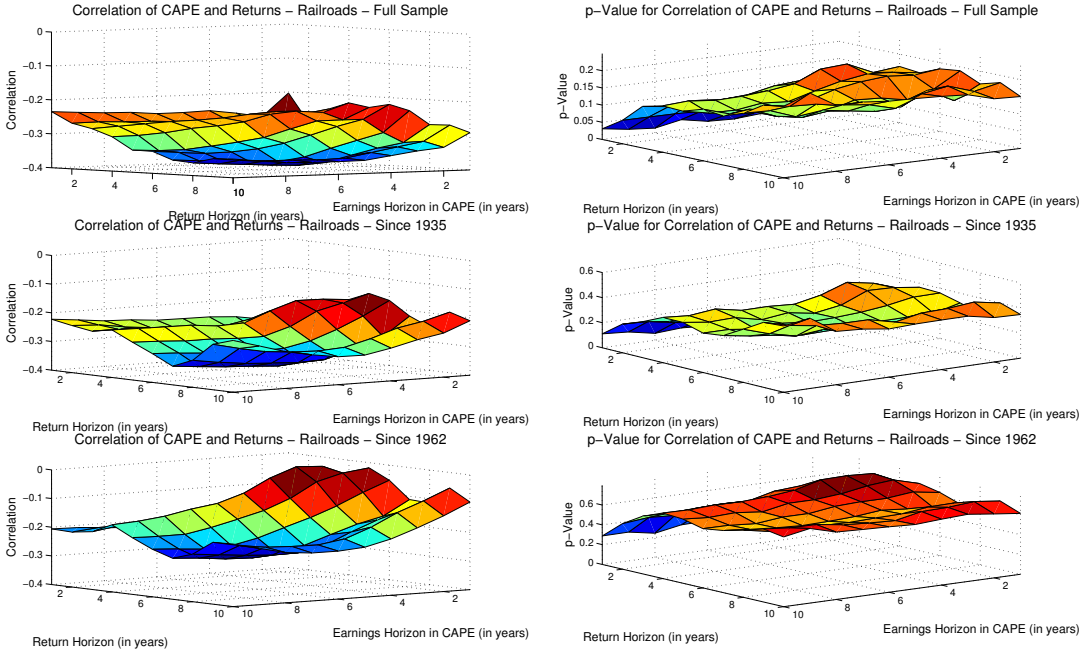


Figure 27: This figure is the analogue to Figure 21 for the Railroads sector instead of the overall stock market.

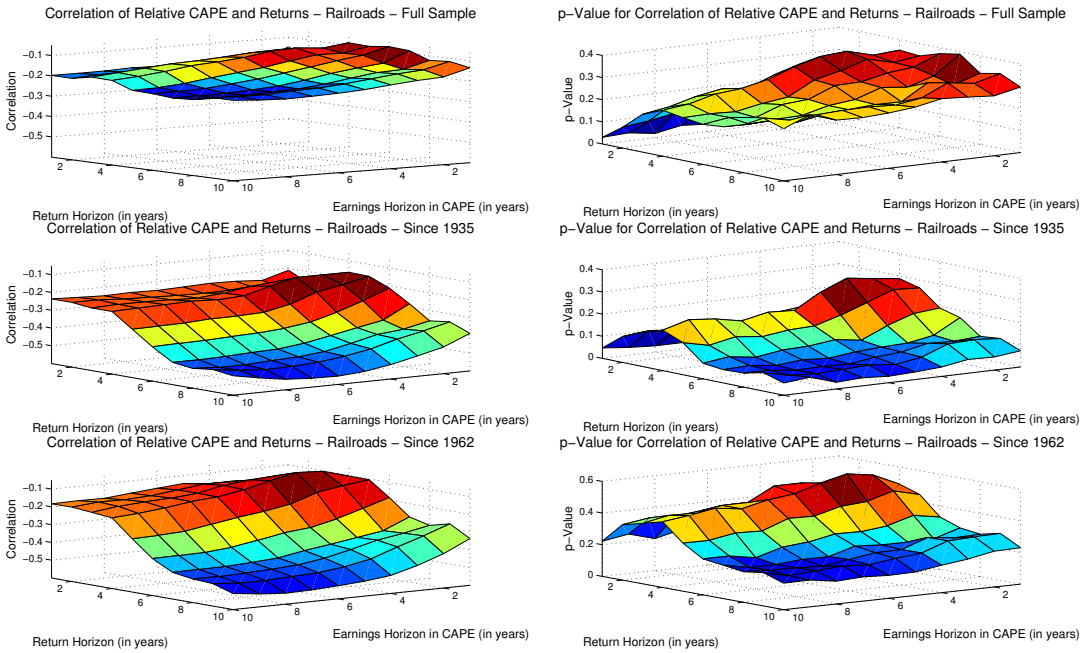


Figure 28: This figure is analogous to Figure 21, but it employs the Relative CAPE indicator instead of the CAPE ratio for the sensitivity study of the correlation involving the Railroads sector.

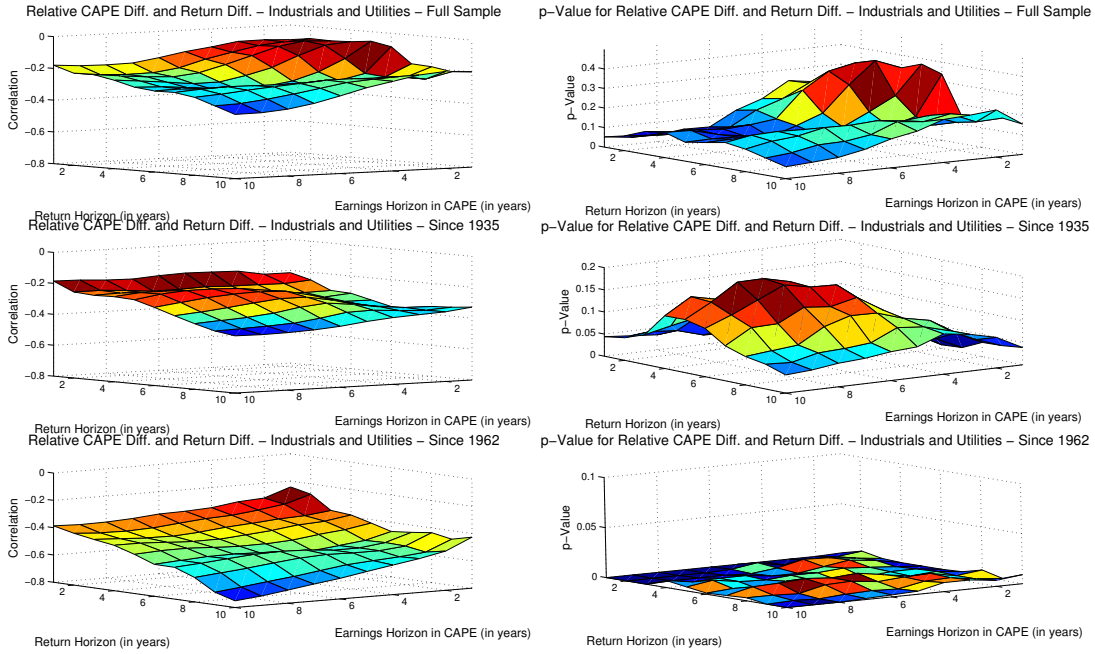


Figure 29: This figure displays a sensitivity study of the correlation analysis as displayed in the cross-sectional comparison between the Industrials and the Utilities sector of Figure 14. The horizon, over which we calculate subsequent returns, varies between 1 and 10 years in annual increments (left axis). We then compare the differences in these returns to differences in variants of the Relative CAPE indicator. For the calculation of the Relative CAPE indicator, we impose the same variation between 1 and 10 years in annual increments for the horizon of the average of earnings in the denominator of the CAPE ratio (right axis). 10 years for the earnings horizon correspond to the Relative CAPE indicator as defined above, whereas 1 year leads to a version of the Relative CAPE indicator derived from the more classic price-earnings ratio. Except for the differences in horizons, the procedure to calculate the long-term returns as well as the variants of the Relative CAPE indicator is as described in Sections 3 through 5. The full sample time series for the Relative CAPE indicator of each of the two sectors in each scenario homogenously starts in the third quarter of 1902 and ends in the fourth quarter of 2002 and subsequent returns are considered correspondingly. The resulting correlation as well as the associated p-values are displayed in the top two panels. The procedure for the simulation-based p-values follows the decsription in Section 5.3 with only minor adaptations for the changes in the horizons. The middle and the bottom pair are analogous to the top pair, but restrict the sample under consideration to the time starting with the fourth quarter of 1935 and the fourth quarter of 1962, which are the same starting points as in Section 4.

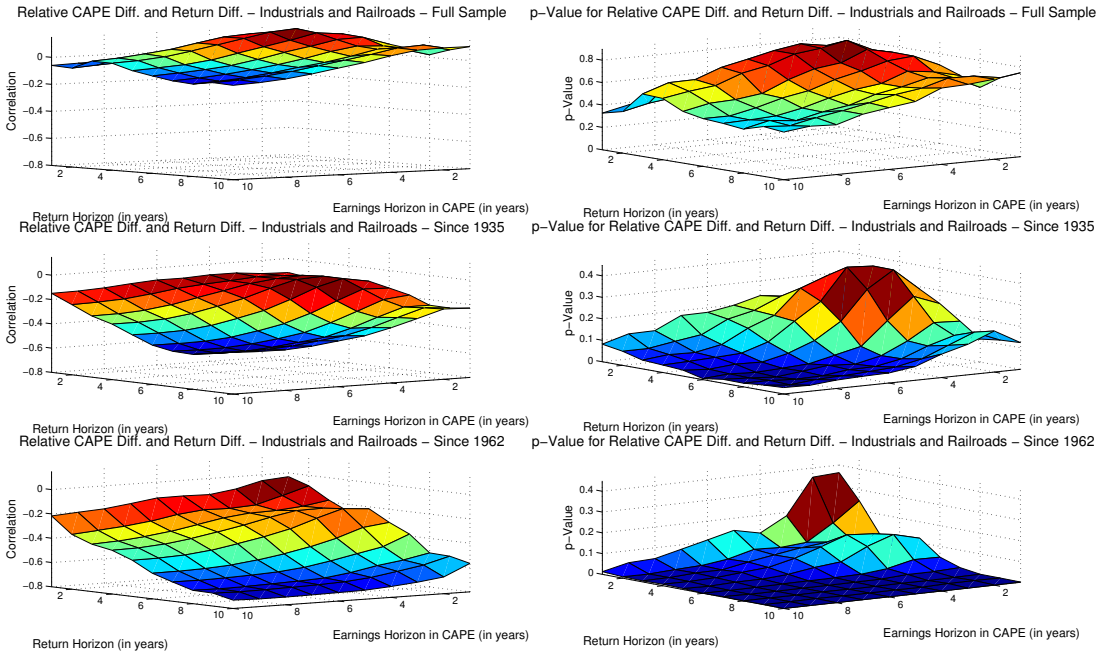


Figure 30: This figure is the analogue to Figure 29 displaying the sensitivity study for differences in the Relative CAPE indicator as well as subsequent returns the Industrials sector and the Railroads sector instead of the differences between the Industrials and the Utilities sector.

the earnings in the denominator of the CAPE ratio both lead to a stronger negative association between differences in the Relative CAPE indicator and differences in subsequent returns, these effects definitely become substantially stronger for the sample that focuses on the last fifty years of data. Moreover, while the p-values in the full sample tend to support the notion that the correlation results are not driven statistical spuriousness and are in fact not compatible with the assumption of unpredictable returns, the negative association for the final fifty years is supported much more strongly by the simulation results.

Focusing again on the correlation evidence, we would like to point out that - when focusing on the ten-year return horizon - there is a difference between the price-earnings ratio (earnings horizon of 1 year) and the CAPE ratio (earnings horizon of 10 years) that amounts to nearly 30%. This is strong evidence in favor of the notion of long-term earnings averages for value investing criteria, in particular given the context of cross-sectional comparisons of this sensitivity study.

When focusing on the comparison between the Industrials and the Railroads sector (Figure 30) as well as on the comparison between the Railroads and the Utilities sector (Figure 31), the same patterns as for the comparison between the Industrials and the Utilities sector arise. The superiority of the notion of 10-year average earnings in the denominator of the CAPE ratio as well as of the notion of ten-year subsequent returns to assess the long-term investment performance becomes the most pronounced for the final fifty years of the overall sample. The evidence in favor of the Relative CAPE indicator being able to predict cross-sectional return differences is then convincingly supported by the simulation-based p-values, which strongly indicate that this association is incompatible with the assumption of unpredictable returns, whereas this incompatibility is not nearly as convincingly established for the upper two panels. Finally, particularly in the shorter-sample scenario, there continues to be a sizeable difference in the degree of the negative association between differences in the Relative CAPE indicator and subsequent return differences when comparing the more conventional price-earnings ratio with the CAPE ratio, reiterating the usefulness of the notion of earnings averages in the computation of a long-term value measure, as it has been alluded to by Graham and Dodd (1934).

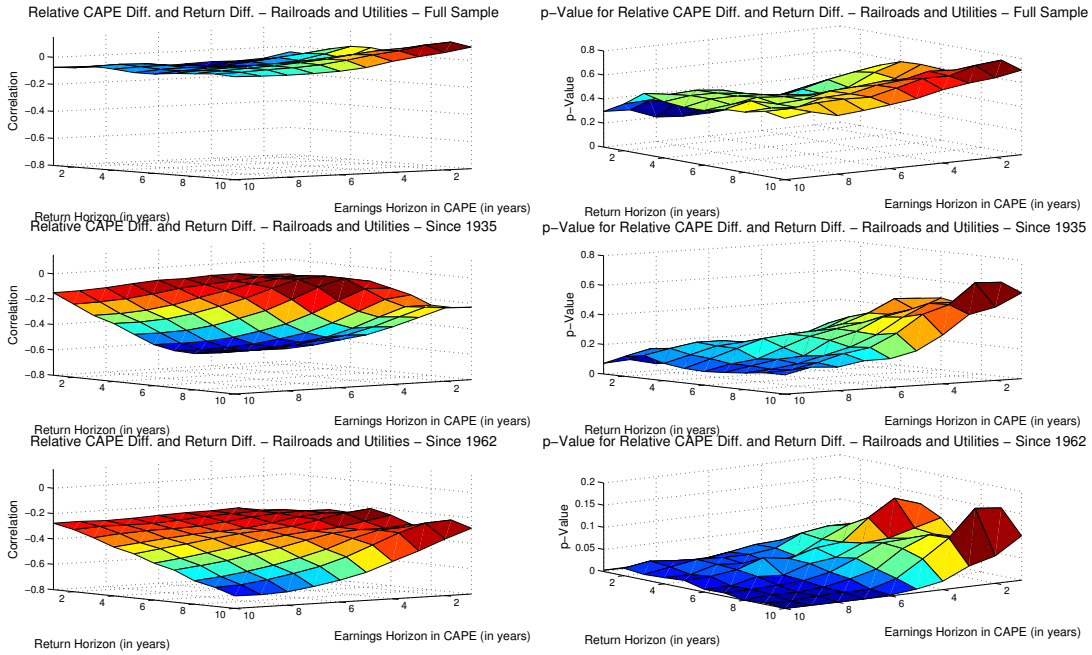


Figure 31: This figure is the analogue to Figure 29 displaying the sensitivity study for differences in the Relative CAPE indicator as well as subsequent returns the Railroads sector and the Utilities sector instead of the differences between the Industrials and the Utilities sector.

8 A Hypothetical Strategy Based on the Relative CAPE Indicator

8.1 Description of the Strategy

In this section, we explore a way to exploit the qualitative evidence in favor of CAPE as a signal of over- and undervaluation in terms of a simple, hypothetical investment strategy, which rotates between the three sectors Industrials, Utilities, and Railroads. Given that our strategy starts from only three assets, an important consideration for our strategy is diversification. Because of this, it appears reasonable to construct a strategy that does not just build on a single sector at each rebalancing date, as it would unnecessarily expose a hypothetical investor to the idiosyncratic risk of that particular sector. The second characteristic of our strategy is quarterly rebalancing, i.e. we re-evaluate the hypothetical investment position as frequently as our dataset permits. Doing so, we always incorporate the latest available information in the allocation decision. Third, we consider a long-only strategy, as we consider short-selling, particularly short-selling the sectors, as highly impractical until the very late part of the sample under consideration. Finally, we are committed to the principle of simplicity for our hypothetical strategy.

We determine the allocation between the three sectors at each rebalancing date by a two-step process. In the first step, we apply the core tenet of our strategy, the CAPE ratio or more explicitly the concept of the Relative CAPE indicator, which allows us to discriminate the sectors by their degree of over- and undervaluation. In particular, we use the Relative CAPE indicator of a sector to determine whether we consider it as overvalued. The second step on the way to the allocation for a specific quarter involves a momentum consideration. This consideration arises because of the so-called value trap. With quarterly rebalancing, CAPE - as a measure of over- and undervaluation - has by its very nature as a long-term gauge of value a diminished ability to distinguish between a sector that is undervalued and will rise again subsequently and a sector that appears undervalued and will continue to fall in value in the short-term. This kind of distinction is exactly what the momentum consideration provides. However, we are keen on ensuring that this consideration plays a complementary role in the investment strategy, with CAPE clearly being the dominant part.

The best way to understand our strategy is to think in terms of "investment scores" for each of the three sectors under consideration. We start from an equally weighted allocation between the three sectors and allocate each of

the three sectors a score of 2.[70] For the sector with the highest Relative CAPE indicator, we subtract 1 from the respective sector's score. In contrast, we add 1 to the score of the sector with the lowest Relative CAPE indicator. This leaves us with an ordering of scores from 1 through 3 that is determined by the Relative CAPE indicator and that overweights the sector that is identified by the indicator as undervalued and underweights the sector that the indicator points to as the relatively most overvalued.

Now, we look at the momentum of each of the three sectors, as the percentage change in the real total return price of a sector over the past four quarters. If a sector's momentum is the lowest among the three for at least two consecutive quarters, this sector's score will be set to zero, concluding the second and final step of the allocation process. To determine the weights of the sectors from the allocation scores, we simply divide the score of each of the three sectors by the sum of the scores of the sectors for this rebalancing date.

We would like to point out that there are two ways, by which we disadvantage the momentum step compared to the CAPE step. First, the momentum reduction, if any, eliminates one sector, but leaves the weight difference that the comparison of the Relative CAPE indicator introduces between the remaining two sectors unchanged. In that sense, the predominant impact on portfolio weights comes from the Relative CAPE indicator, leaving a subsidiary role to the momentum consideration. Second, it is not even the case that there is a momentum reduction in each and every step. Only under certain circumstances, we implement a momentum reduction at all. One way to think about our requirement that a sector is the worst momentum-performer for two consecutive quarters in a row before a score reduction is applied is as a "probationary phase". If a sector has the worst momentum, it is put on "momentum probation" and will lose 1 score if its momentum is still comparatively the worst. If however the order of the sectors in terms of their momentum changes from one quarter to the next, its "probation" is lifted and another sector enters into its "probationary phase".

Observe the specific way in which we make use of momentum. Normally and classically, the momentum variable serves as a sorting criterion for individual stocks that has direct portfolio implications as determining a long-portfolio, which may be complemented by a short-portfolio of the worst momentum-performers, an idea that has been put forward by Jegadeesh and Titman (1993). Daniel (2011) provides evidence for the long-term success of momentum-based investing, but also finds evidence for the skewness of the distribution of momentum-based investment returns, as there are few instances, in which the strategy exhibits substantial crashes. Asness et al. (2012) as well as Fama and French (2012) further substantiate the generality of the momentum idea, with the latter two also considering international evidence as well as other asset classes.[71] Asness et al. (2012) and Fama and French (2012) actually do not just look at momentum in isolation, but consider it in conjunction with a value-based investment approach, an idea that has also been employed by Asness (1997). What is however common to these analyses, particularly in Asness et al. (2012) and Fama and French (2012), is the approach to construct portfolios based on each of the two effects in isolation, which are then subsequently aggregated linearly into a common portfolio. This is different from our approach that explicitly incorporates a value-based metric, which is then corrected by a momentum-characteristic, leading to a non-linear, two-step aggregation of the two effects. We explicitly use momentum as a secondary variable, acting on assets, sectors in this case, that have already undergone a value-based manipulation. As a one-sided variable, with low momentum leading to elimination, we use it more for its corrective ability than for its portfolio selection ability.

Moreover, we deviate from the classic construction of momentum portfolios, as pursued in all of the above references, that operates on the level of individual assets, which are then combined into portfolios. In contrast, we apply our momentum-consideration at the level of sectors directly. This approach bears similarity to the analysis in Bhojraj and Swaminathan (2006), who also consider momentum on an aggregate level, in their case on the level of country indices and not sectors of the US sectors economy as presented here. The sector-based approach also bears resemblance to the argument in Moskowitz and Grinblatt (1999) and Scowcroft and Sefton (2005) that characterize

[70] The choice of an equally weighted allocation as the starting point is necessitated by our lack of information about the sectors' market capitalization.

[71] Novy-Marx (2012) also considers a wide range of countries as well as asset classes in his analysis of momentum-like features of asset returns. Importantly however, he breaks with the convention to look at returns that span a time period that lasts right until the time of the portfolio sort (or a single month prior to the sorting time), but looks at returns as measured until six months prior to the portfolio sort.

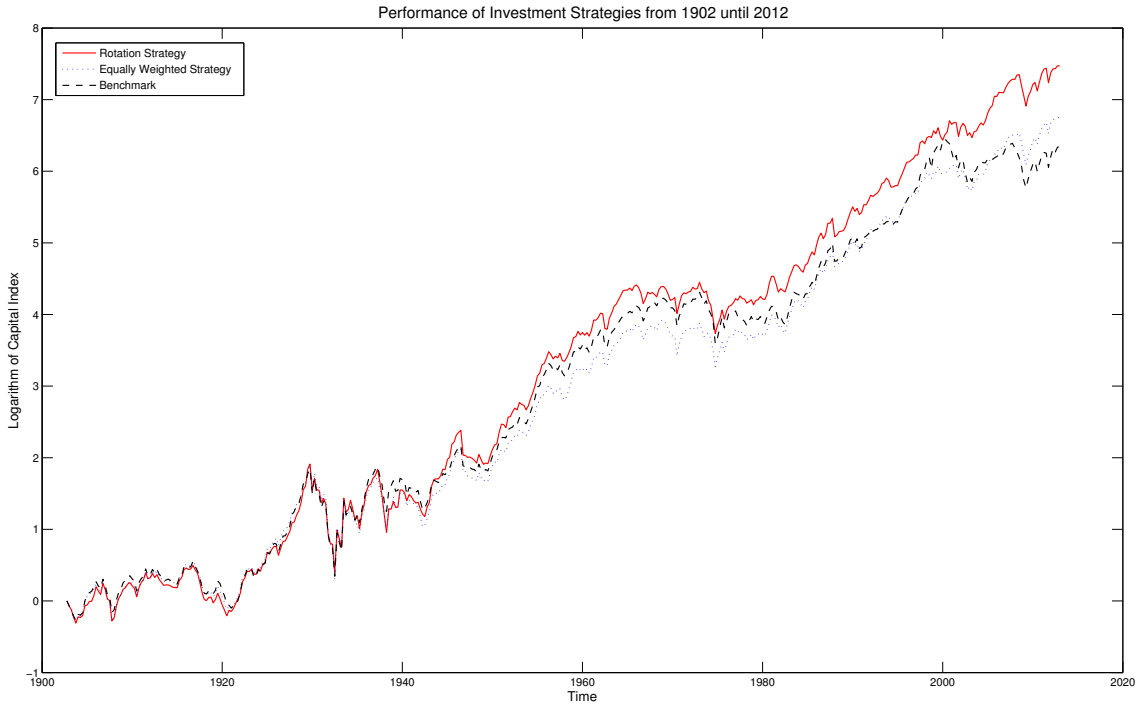


Figure 32: This figure shows the performance of three different portfolio strategies on a log-scale. Investment commences in the beginning of the first quarter of 1903. Rebalancing quarterly, all three strategies end in the beginning of the first quarter of 2013. The strategy represented by a dashed black line invests into the S&P Composite Index at each rebalancing date, which we consider as our main benchmark. The second benchmark, marked as a dotted blue line, is an equally weighted strategy that puts a weight of one third into each of the three sectors at each rebalancing date. Finally, the solid red line represents the CAPE-based sector strategy as outlined in Section 6.

momentum as an industry-effect by isolating the characteristics of the individual assets from industry-wide effects. In contrast, by applying the momentum-consideration on the sector-level directly, we omit this attribution and focus on the industry/sector effect directly. By doing so, our application of the momentum effect is actually very close in spirit to Moskowitz et al. (2012), who do not pursue the classic momentum-based portfolio sort but look at the momentum of an individual asset to determine its impact on this specific asset's future returns.

8.2 Performance of the Hypothetical Investment Strategy

We will begin the investigation of our historical, hypothetical investment strategy in the first quarter of 1903. The allocation of this quarter will be determined by valuation measures complemented by the aforementioned momentum consideration at the end of the fourth quarter of 1902. Rebalancing quarterly, the last quarter, during which we take a look at the performance of our hypothetical strategy is the fourth quarter of 2012 according to valuation and momentum rankings as of the end of the third quarter of 2012. As benchmarks, we report both the performance of the overall stock market, also in cumulative inflation-adjusted total return levels as well as the performance of an equally weighted strategy that puts a weight of one third on each of the three sectors at each rebalancing date.

As it becomes evident from Figure 32, the hypothetical CAPE-based rotation between Industrials, Utilities, and Railroads leads to consistently higher levels of cumulative, inflation-adjusted capital almost throughout the entire time period under investigation. It is the period until the 1940s, in which our strategy tracks either of the two benchmarks fairly closely, only to substantially outperform these thereafter.

Not looking at the performance in log-scale but in terms of money obtained, an investment into our simplistic sector-rotation strategy based only on the three sectors turns \$1 in the beginning of the fourth quarter of 1902 into approximately $\$1,756$ in the beginning of the first quarter of 2013. In comparison, an investment into the S&P Composite turns \$1 into approximately \$569 over the same time horizon. If, in contrast to a benchmark like the S&P

Performance Statistics for Portofolio Strategies	Start in Q3 of 1902	Start in Q4 of 1935	Start in Q4 of 1962
CAPE-Based Sector Rotation	7.01%	8.07%	7.29%
(Return / Volatility / Information Ratio)	20.52%	17.20%	15.57%
	34.17%	46.92%	46.84%
Equally Weighted Strategy	6.32%	7.12%	6.85%
(Return / Volatility / Information Ratio)	20.42%	16.38%	15.12%
	30.93%	43.49%	45.31%
Market Benchmark	5.92%	6.39%	5.47%
(Return / Volatility / Information Ratio)	20.15%	16.88%	16.43%
	29.39%	37.84%	33.31%

Table 3: This table shows the performance statistics when one limits the time horizon under consideration. The first column reports the statistics derived from the performance of the strategies as displayed in Figure 32. The second and third column change the start date of the analysis, leaving the end date fixed at the beginning of the third quarter of 2012, which is the end of the overall sample. The fourth quarter of 1935 refers to the first quarter, at which we use data from S&P resources only, and the second quarter of 1962 is the quarter that allows us to single out the final fifty years of the dataset. The rows refer to different strategies, with our hypothetical strategy located in the top row and the two benchmarks located below. Each cell contains the annualized, inflation-adjusted total return, the annualized volatility as derived from the inflation-adjusted total return series, and the ratio of return and volatility, which is referred to as Information Ratio.

Composite that is weighted according to market value, one considers an equally weighted benchmark, which simply allocates evenly across the three sectors under consideration at each rebalancing date, \$1 grows to about \$856, less than half of the amount accrued by the CAPE-based strategy.

Figure 3 has displayed the inflation-adjusted total return performance of the three sectors. As it has become evident, the Industrials sector has been the sector with the best performance among the three. Analyzing an investment into this one sector alone in connection with the performance of our sector-rotation strategy, the Industrials sector would turn \$1 in the beginning of the fourth quarter of 1902 into approximately $\$1,463$ (as usual, inflation-adjusted) in the beginning of the first quarter of 2013, which would still fall short by almost \$300 compared to the performance of our rotation-strategy. Moreover, by relying solely on one sector, an investor would have exposed herself to substantial non-diversification risk, which is by definition much less strong in the allocation that the sector-rotation induces.

The outlined investment performance of our strategy translates into an inflation-adjusted, annualized, compounded total return of about 7.01%, compared to approximately 5.92% for the S&P Composite and 6.32% of the equally weighted benchmark. With annual volatility of about 20.52%, it follows that our rotation strategy's information ratio is approximately 34.17%, which exceeds the information ratio of the S&P Composite by nearly five percentage points (20.15% volatility and 29.39% information ratio for this benchmark) and the information ratio of the equally weighted benchmark by slightly more than three percentage points (20.42% volatility for the equally weighted strategy leading to an information ratio of 30.93%).

When limiting, analogous to Tables 1 and 2, the time horizon under consideration to start farther into the sample, while leaving the end of the performance analysis unchanged (which is the beginning of the third quarter of 2012 in this case), a consistent picture of the outperformance of our simplistic strategy arises. It actually turns out that the outperformance compared to the market benchmark becomes even stronger as we limit the time horizon under consideration, leading up to more than 180 basis points of outperformance, which translates into an almost 14% higher information ratio.

Focusing again on the longest time interval for the analysis of our hypothetical investment strategy, Figure 33 displays the portfolio weights that our rotation strategy allocates to the three underlying sectors at each rebalancing date. The weight on the Industrials sector appears at the top in red, the weight on Utilities is depicted in blue in the middle, and the weight on Railroads is green at the bottom of the figure. In the beginning of the analyzed time period, our strategy tends to overweight Industrials, only to underweight this sector in the 1910s. During the 1920s, and in particular before the 1929 stock market crash, the Relative CAPE indicator initially causes the strategy

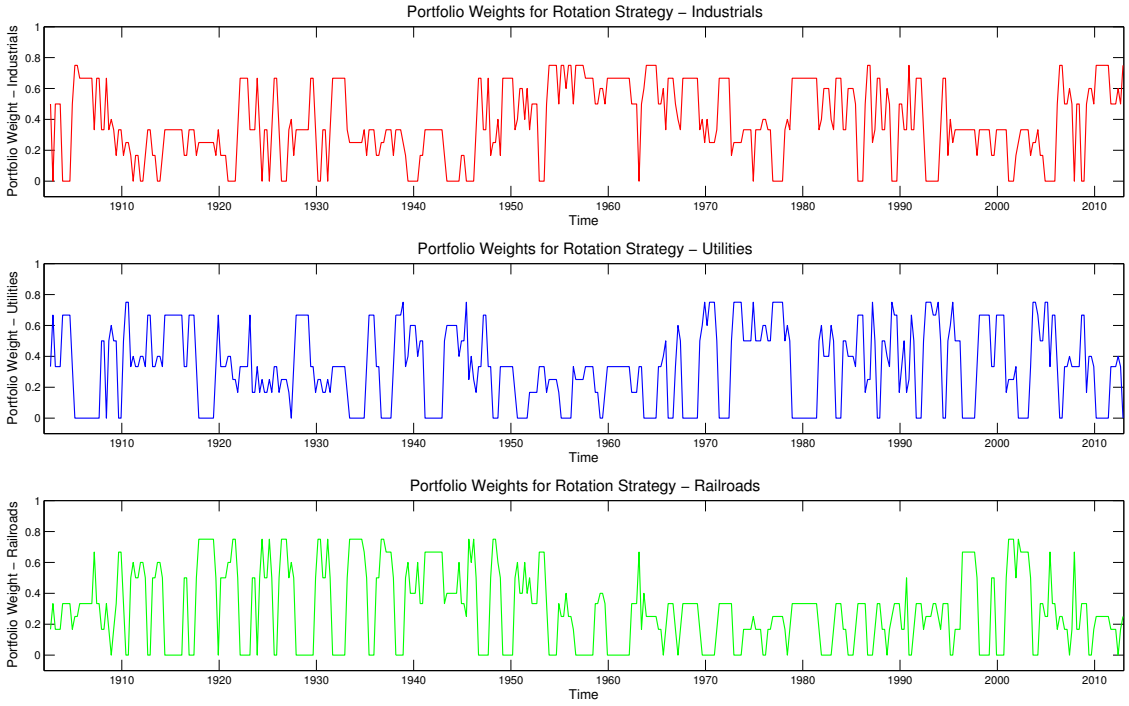


Figure 33: This figure depicts the portfolio weights for the investment strategy that rotates between Industrials, Utilities, and Railroads according to the sectors' Relative CAPE indicator, supplemented by a momentum consideration, as outlined in Section 8. Each time series starts at the end of the fourth quarter of 1902 and runs until the end of the third quarter of 2012. Their frequency is quarterly. The top graph corresponds to the Industrials sector, the middle one to the Utilities sector, and the bottom one to the Railroads sector.

to underweight the Utilities sector, whereas Railroads are eliminated by the momentum consideration. Despite the dramatic peak in the CAPE ratio for the Utilities sector in the late 1920s, as shown in Figure 4, the Industrials sector turns out to be more overvalued in the years immediately preceding 1929, as its CAPE ratio increases faster in a shorter period of time compared to the Utilities sector. However, at the onset of the 1929 stock market crash, Utilities are the most overvalued, resulting in it being underweighted until the middle of the 1930s. Another period of underweighting the Utilities sector starts - subsequent to a period of an underweight Industrials sector between the mid-1930s and the mid-1940s - in the very late 1940s until the mid-1960s. What follows is an almost thirty-year period of the Railroads sector not being considered very favorably by the Relative CAPE indicator and also by the additional momentum-consideration. In the mid-1990s, the Relative CAPE indicator starts picking up the overvaluation from the technology boom, which impacts the Industrials sector. It then takes until the middle of the 2000s for the Industrials sector to appear more favorably. During the final quarters of the sample, the Industrials sector is the most favored sector by the Relative CAPE indicator, followed by the Utilities sector, and lastly by the Railroads sector.

Figure 34 assesses the excess performance of the sector-based sector rotation strategy compared to the equally weighted as well as the market benchmark. It does so by splitting the sample period into non-overlapping five-year periods, positioning the intervals such that the end of the sample period (beginning of the first quarter of 2013) coincides with the end of a five-year interval. In consequence, the last interval captures the five-year period between the beginning of the first quarter of 2008 and the first quarter of 2013, whereas the previous interval goes back to the beginning of the first quarter of 2003.[72] All displayed returns are non-annualized. The left hand side of the figure focuses on the comparison of the rotation strategy to the equally weighted benchmark, whereas the right hand side focuses on the comparison of the rotation strategy to the benchmark. The two top panels display the actual five-year total returns, whereas the bottom panels show differences.

[72]This procedure leads to a single quarter in the very beginning of the sample not being included in any five-year interval.

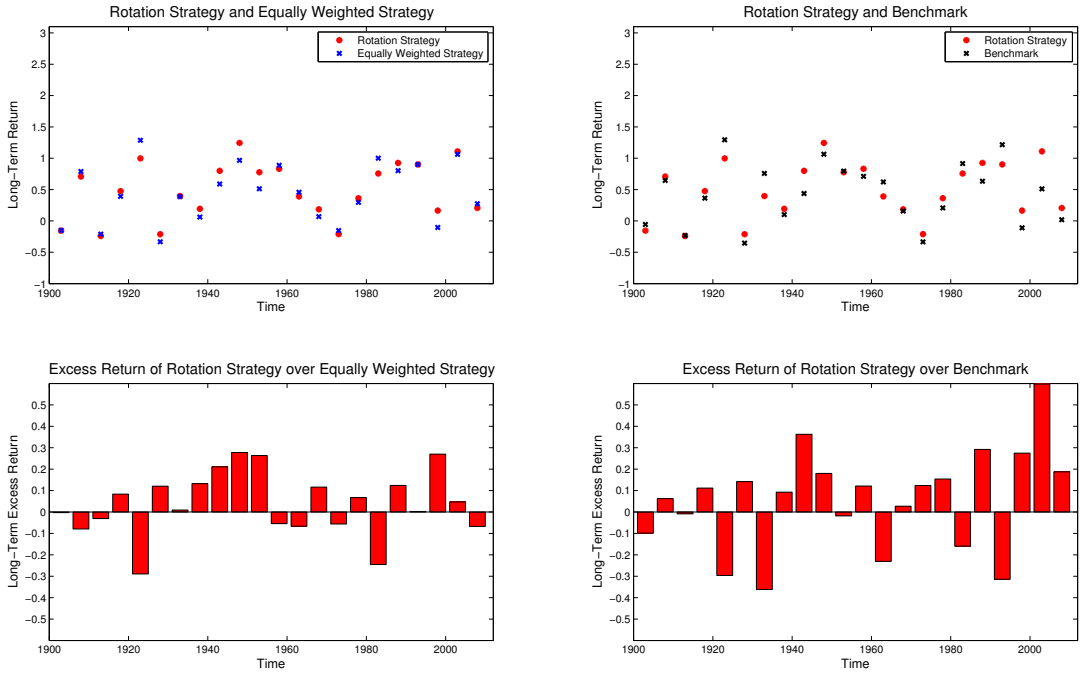


Figure 34: This figure displays the long-term performance of the hypothetical rotation strategy as compared to the equally weighted benchmark (left two panels) and to the benchmark (right two panels). The time period between the beginning of the fourth quarter of 1902 and the beginning of the first quarter of 2013 is split into non-overlapping five-year intervals, with the very first quarter not being included in any five-year interval. The return numbers are inflation-adjusted, net total return quantities, which are not annualized. They are derived from the time series displayed in Figure 32. The top panels display the returns for both the strategy as well as the respective benchmark, whereas the bottom panels display the difference.

From the difference bar charts, it becomes apparent that the outperformance of our hypothetical rotation strategy is not concentrated in a particular time period, but is actually fairly consistent throughout the entire time period under investigation. Expectedly, there are five-year periods, in which the performance of our rotation strategy falls short of one or even both of the benchmarks, but other time periods convincingly compensate for this lower performance and lead to the overall performance pattern that Figure 32 conveys. It is moreover worth emphasizing how the most dramatic underperformance periods come about, e.g. the difference to the equally weighted strategy during the periods starting in 1922 and 1982 as well as the difference to the benchmark during the period starting in 1932 and also in 1982. It is not the case that our rotation strategy performs negatively, whereas the comparison strategies strive. In contrast, it is actually the case that both strategies under comparison perform very well in each of the mentioned time periods, as evidenced by the top panels, but that the rotation strategy performs slightly less impressively.

By the nature of the long-term focus that our valuation measure, the Relative CAPE indicator, has, it is completely natural to expect a more cautious portfolio approach that will of course impede performance when markets are soaring. However, it will also guarantee more stable performance during times of market decline, as evidenced for example by the outperformance after the burst of the technology bubble around the turn of the twentieth century.

As a final piece of information to assess the performance of our sector rotation strategy compared to the S&P Composite as well as compared to the equally weighted benchmark, we consider what we term an "outperformance ratio". For various time horizons, ranging from 1 quarter to 40 quarters, we look at rolling, and therefore necessarily overlapping, time periods. For each of the different time horizons, we count the number of times that our strategy outperforms the benchmark for the rolling time periods of fixed length. This results in an outperformance ratio associated with a specific time horizon. As our strategy is supposed to capture long-term value, we expect the outperformance ratio to be an increasing function of the length of the time period over which performance is considered.

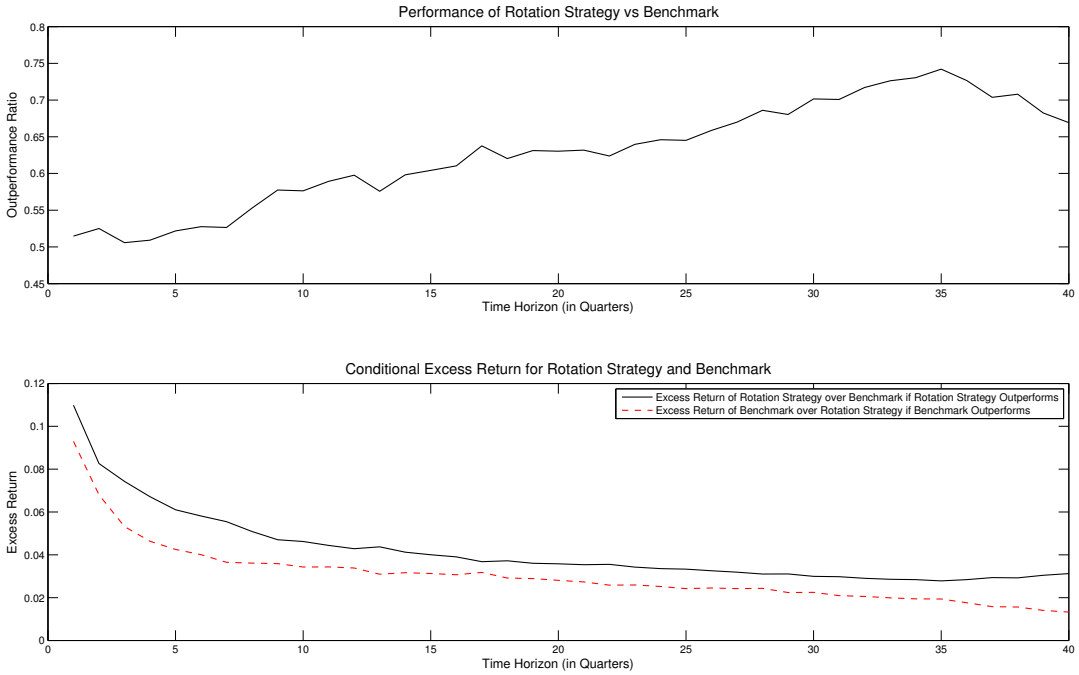


Figure 35: This figure provides additional information about the performance of our hypothetical rotation strategy when compared to the market benchmark. The top panel computes a ratio that we refer to as "outperformance ratio". For a fixed time horizon, it measures the number of times that our rotation strategy outperforms the benchmark over that time horizon. The first count occurs at the beginning of the sample, which is the beginning of the fourth quarter of 1902, and the last count ends with the beginning of the first quarter of 2013. We assess the outperformance for the fixed time horizon, where we look at overlapping time-periods for different outperformance assessments. One fixed time interval results in a single ratio and the plot in the top panel displays the ratio against the time horizon used for the ratio calculation. The bottom panel measures the associated conditional outperformance. For each fixed time horizon, the performance difference (difference in non-annualized, real total return) between the rotation strategy and the benchmark is attributed to the black solid line when the rotation strategy outperforms and to the red dashed line when the rotation strategy underperforms. (The red dashed line measures the excess of the benchmark over the rotation strategy) After having considered all overlapping time intervals for a fixed time horizon, we average the excess returns in the two groups geometrically and annualize the result to enable comparisons across time horizons. Both panels display the resulting quantities for varying time horizons, ranging from 1 quarter to 40 quarters.

This is indeed what the black line in the top graph of Figure 35 displays. Whereas for very short time horizons, our strategy outperforms the benchmark only slightly more than 50% of the time, this ratio increases fairly consistently to approximately two thirds for the longest time period considered, which looks at rolling 10-year windows.

The bottom panel aims to address the potential concern that, although its outperformance ratio is increasing, our rotation strategy still does not outperform the benchmark around 30% of the time, even over a time period of 10 years. The two lines in the bottom graph therefore measure average conditional excess returns. The black line captures the conditional excess return of the sector rotation over the benchmark within those rolling time-windows, in which it actually outperforms the benchmark. This series of excess returns associated with a fixed time horizon is then geometrically averaged and the result is annualized to allow for comparisons across different time horizons. The resulting excess return quantity is then plotted against the length of the time horizon under consideration. Analogously, the red line captures the conditional outperformance of the benchmark over the rotation strategy, i.e. the shortfall in performance of our strategy against the length of the rolling time windows. Two properties become apparent with regard to these conditional excess returns. First, both curves decrease as the length of the time horizon increases. This means that the performance profile becomes more stable, as the outperformance in each direction gets less extreme, with increases in the time horizon, over which performance is measured.[73] This effect is to be expected,

[73]It is important to keep in mind that we have annualized the return differences (after averaging) to make excess returns comparable

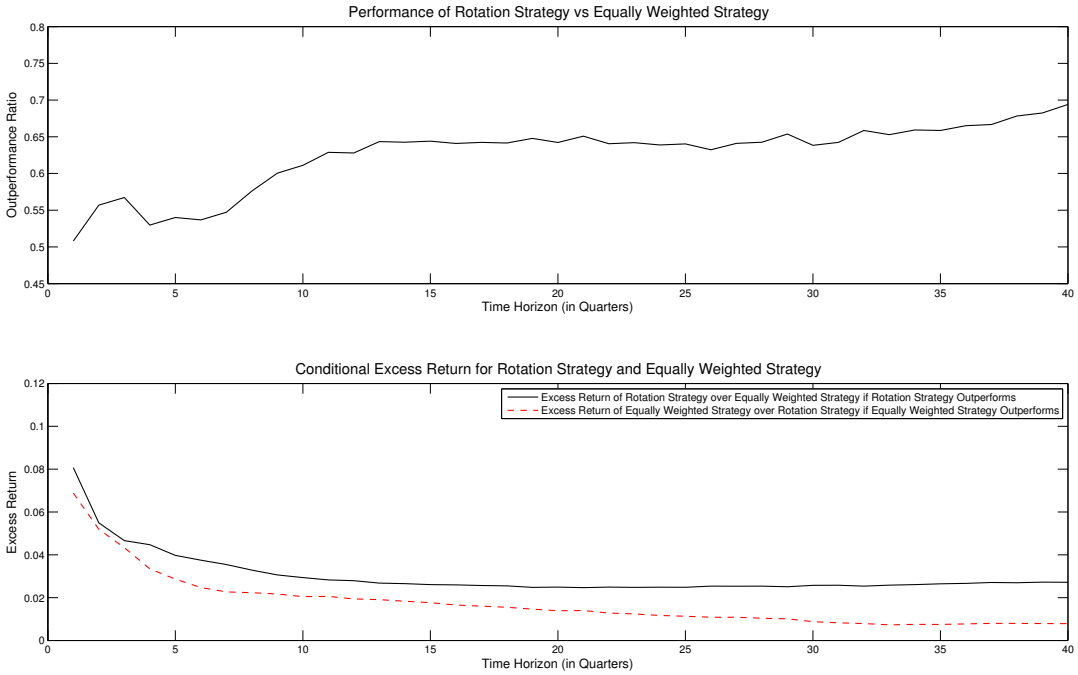


Figure 36: This figure is analogous to Figure 35. It does however not compare the performance of the CAPE-based sector rotation to the market benchmark, but to the equally weighted benchmark.

as a longer-term focus ought to result in a more stable return profile. Second, it is crucial to note that the black line, referring to the conditional excess return of the sector allocation strategy consistently exceeds the red line, referring to those instances, in which the benchmark outperforms. That is, it is not only true that our strategy outperforms the benchmark more often, as evidenced by the top graph, but also that the outperformance of the sector allocation strategy is higher than the underperformance during times, in which the benchmark has higher returns. Moreover, during the longest time horizons, where the outperformance ratio actually falls slightly from above 70% to below 70%, one can clearly see that the associated excess return markedly picks up, i.e. the conditional discrepancy compensates for the drop in the outperformance ratio.

Figure 36 shows an analogous plot to Figure 35, performing a comparison between our rotation strategy and the equally weighted benchmark. One observes a very similar pattern, in that the outperformance ratio increases from slightly more than 50% to nearly 70%, accompanied by the conditional excess return of the CAPE-based sector allocation strategy consistently exceeding the conditional excess return of the benchmark.

9 Conclusion

The contribution of this paper is threefold. First, we have constructed a record of price, dividend, and earnings indices from primary sources - Cowles (1939) as well as various editions of the Standard & Poor's Security Price Index Record and the Standard & Poor's Analysts' Handbook - starting in the early 1870s and ending in the end of the year 2012.

This data has enabled us to investigate the valuation of the US stock market for almost 130 years[74] on the more granular level of sectors and has additionally given us the opportunity to consider the relative valuation of these components, which is the second contribution of this paper. The CAPE ratio has not only proven to be effective in the prediction of the performance of the sectors individually, but has also the potential to assess the relative return

across different time horizons.

[74]The full sample captures about 140 years of data, which results in about 130 years of data for the CAPE ratio, as this ratio requires ten years of earnings. As the Relative CAPE indicator involves twenty years of history of the CAPE ratio, our investment strategy only covers about 110 years.

performance across sectors, based on differences in the Relative CAPE indicator. Building on comparisons of the Relative CAPE indicator across sectors, we have devised a hypothetical, historical investment strategy overweighting the sector with the lowest Relative CAPE indicator and underweighting the one with the highest indicator. The only additional ingredient, which we construct to be of subsidiary importance compared to the CAPE-based weighting, is the elimination of a sector based on a momentum consideration, which is necessary because of the so-called value trap.

Finally, on methodological grounds, we extend the definition of the Cycically Adjusted Price-Earnings (CAPE) ratio to not only account for inflationary effects that occur because of the consideration of a long-term earnings history, but also to correct for effects from corporate payout policy. Moreover, our construction of the Relative CAPE indicator is necessary to use a long-term valuation measure based on the CAPE ratio to perform comparisons across assets.

Our results using over a hundred years of data are consistent with the notion that major sectors of the stock market show frequent mispricings that can be exploited in an investment strategy that generally leads to better results than holding the market portfolio. Entire sectors do appear to go through short periods of mispricing as suggested by behavioral theories. The durations of these periods of mispricing are often measured in just a few years - requiring continued vigilance for sector allocations to achieve better returns.

Bibliography

Anonymous, March 1867. A Girl at a Railway Junction's Reply. Every Sunday: A Journal of Choice Reading 3 (61), 281.

Asness, C. S., 1997. The Interaction of Value and Momentum Strategies. Financial Analysts Journal 53 (2), 29–36.

Asness, C. S., Moskowitz, T. J., Pedersen, L. H., 2012. Value and Momentum Everywhere. New York University - Working Paper.

Barron's, July 2012. Barron's 50-stock average. `http://online.barrons.com/public/page/9_0210-barr50stock.html`.

Basu, S., 1977. Investment Performance of Common Stocks in Relation to their Price-Earnings Ratios: A Test of the Efficient Markets Hypothesis. The Journal of Finance 32 (3), 663–682.

Bhojraj, S., Swaminathan, B., January 2006. Macromomentum: Returns Predictability in International Equity Indices. The Journal of Business 79 (1), 429–451.

Campbell, J. Y., 2000. Asset Pricing at the Millenium. The Journal of Finance 55 (4), 1515–1567.

Campbell, J. Y., Shiller, R. J., 1988. Stock Prices, Earnings, and Expected Dividends. The Journal of Finance 43 (3), 661–676.

Campbell, J. Y., Shiller, R. J., 1989. The Dividend-Ratio Model and Small Sample Bias: A Monte Carlo Study. Economics Letters 29, 325–331.

Campbell, J. Y., Shiller, R. J., Winter 1998. Valuation Ratios and the Long-Run Stock Market Outlook. Journal of Portfolio Management, 11–26.

Campbell, J. Y., Shiller, R. J., April 2001. Valuation Ratios and the Long-Run Stock Market Outlook: An Update. National Bureau of Economic Research Working Paper 8221.

Campbell, J. Y., Thompson, S. B., 2008. Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? The Review of Financial Studies 21 (4), 1509–1531.

Chan, L. K., Hamao, Y., Lakonishok, J., 1991. Fundamentals and Stock Returns in Japan. The Journal of Finance 46 (5), 1739–1764.

Chan, L. K., Jegadeesh, N., Lakonishok, J., 1995. Evaluating the Performance of Value versus Glamour Stocks: The Impact of Selection Bias. Journal of Financial Economics 38 (3), 269–296.

Chan, L. K., Lakonishok, J., 2004. Value and Growth Investing: Review and Update. Financial Analysts Journal 60 (1), 71–86.

Cole, A. H., Frickey, E., 1928. The Course of Stock Prices 1825-66. Review of Economic Statistics 10, 117–139.

Cowles, A., 1939. Common-Stock Indexes, 2nd Edition. Principia Press, Inc., Bloomington, IN.

Daniel, K., 2011. Momentum Crashes. Columbia University - Working Paper.

Fama, E. F., 1998. Market Efficiency, Long-Term Returns, and Behavioral Finance. Journal of Financial Economics 49 (3), 283–306.

Fama, E. F., French, K. R., 1992. The Cross-Section of Expected Stock Returns. The Journal of Finance 47 (2), 427–465.

Fama, E. F., French, K. R., 2012. Size, Value, and Momentum in International Stock Returns. Journal of Financial Economics 105 (3), 457–472.

Fogel, R., 1971. Railroads and American Economic Growth. Cambridge University Press, Cambridge, UK.

Global Financial Data, August 2012. Global Financial Data Database. `https://www.globalfinancialdata.com/index.html`.

Goetzmann, W. N., Ibbotson, R. G., Peng, L., July 2000. A New Historical Database for the NYSE 1815 to 1925: Performance and Predictability. Yale International Center for Finance Working Paper.

Goetzmann, W. N., Jorion, P., June 1993. Testing the Predictive Power of Dividend Yields. The Journal of Finance 48, 663–679.

Graham, B., Dodd, D. L., 1934. Security Analysis. McGraw-Hill Companies, Inc., New York, NY.

Hovey, D., May 1931. What of the Railroad Future?: Not a Question of Obsolescence but Evolution. Barron's 11 (21), 3–4.

Ibbotson, R. G., Sinquefield, R. A., January 1976. Stocks, Bonds, Bills, and Inflation: Year-by-Year Historical Returns (1926-1974). The Journal of Business 49 (1), 11–47.

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

Jorion, P., Goetzmann, W. N., 1999. Global Stock Markets in the Twentieth Century. The Journal of Finance 54 (3), 953–80.

Keeler, T. E., 1983. Railroads, Freight, and Public Policy. Brookings Institution, Washington, DC.

Kendall, M. G., 1954. Note on the Bias in the Estimation of Autocorrelation. Biometrika 41, 403–404.

Lakonishok, J., Shleifer, A., Vishny, R. W., 1994. Contrarian Investment, Extrapolation, and Risk. The Journal of Finance 49 (5), 1541–1578.

Littell, E., Littell, R. S., 1846. The Railway Madness and its Treacherous Keepers. The Living Age 8, 61.

MacKinnon, J. G., Smith Jr., A. A., 1998. Approximate Bias Correction in Econometrics. Journal of Econometrics 85, 205–230.

Marriott, F. H. C., Pope, J. A., 1954. Bias in the Estimation of Autocorrelations. Biometrika 41, 393–402.

Miller, M., Modigliani, F., 1961. Dividend Policy, Growth, and the Valuation of Shares. The Journal of Business 34, 411–433.

Moskowitz, T. J., Grinblatt, M., 1999. Do Industries Explain Momentum? The Journal of Finance 54 (4), 1249–1290.

Moskowitz, T. J., Ooi, Y. H., Pedersen, L. H., 2012. Time Series Momentum. Journal of Financial Economics 104, 228–250.

Nelson, C. R., Kim, M. J., June 1993. Predictable Stock Returns: The Role of Small Sample Bias. The Journal of Finance 48, 641–661.

Nichols, J. R., December 1878. The Gas Excitement. Boston Journal of Chemistry 12 (18), 211.

N.N., January 12, 1911. Index of Railroad Earnings Covering a Ten-Year Period. The Wall Street Journal, 6.

Novy-Marx, R., 2012. Is Momentum Really Momentum? Journal of Financial Economics 103, 429–453.

Odlyzko, A., 2010. Collective Hallucinations and Inefficient Markets: The British Railway Mania of the 1840s. University of Minnesota - Working Paper.

Orcutt, G. H., Winokur Jr., H. S., 1969. First Order Autoregression: Inference, Estimation, and Prediction. Econometrica 37, 1–14.

Schwert, G. W., 1990. Indexes of U.S. Stock Prices from 1802 to 1987. The Journal of Business 63 (3), 399–426.

Scowcroft, A., Sefton, J., 2005. Understanding Momentum. Financial Analysts Journal 61 (2), 64–82.

Shiller, R. J., 1992. Market Volatility. The MIT Press, Cambridge, MA.

Shiller, R. J., 2005. Irrational Exuberance, 2nd Edition. Princeton University Press, Princeton, NJ.

Shiller, R. J., 2012a. Finance and the Good Society. Princeton University Press, Princeton, NJ.

Shiller, R. J., July 2012b. Online data. `http://www.econ.yale.edu/~shiller/data.htm`.

Smithers, A., Wright, S., 2002. Valuing Wall Street: Protecting Wealth in Turbulent Markets. McGraw-Hill Companies, Inc., New York, NY.

Stambaugh, R. F., 1999. Predictive Regressions. Journal of Financial Economics 54, 375–421.

Standard and Poor's, July 2012. GICS Classification. `http://www.standardandpoors.com/indices/gics/en/us`.

Standard and Poor's Statistical Service, 1978-2010. Security Price Index Record. Various Years.

Standard and Poor's Statistical Service, 1982-2012. Analysts' Handbook. Various Years.

Standard Trade and Securities Service, 1938-1950. Standard Statistical Bulletin. Various Years.

The Trader, July 1929. What of the Market?: The Trader Gives His Views. Barron's 29, 2.

United States Postal Service, July 2012. Mail by rail. `http://about.usps.com/who-we-are/postal-history/mail-by-rail.pdf`.

Ural, C., Lazanas, A., Zhuang, J., Staal, A., September 2012. Sector Selection Based on the Cyclically Adjusted Price-Earnings (CAPE) Ratio. Barclays Research Paper.

Warren, G. F., Pearson, F. A., 1935. Gold and Prices. John Wiley and Sons, New York, NY.

Welch, I., Goyal, A., 2008. A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. The Review of Financial Studies 21 (4), 1455–1508.

Wilson, J. W., Jones, C. P., April 1987. A Comparison of Annual Common Stock Returns: 1871-1925 with 1926-85. The Journal of Business 60 (2), 239–258.

Wilson, J. W., Jones, C. P., July 2002. An Analysis of the S&P500 Index and Cowles's Extensions: Price Indexes and Stock Returns, 1870-1999. The Journal of Business 75 (3), 505–533.

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

A Construction Details for Price, Dividend, and Earnings Data

A.1 Overview of Data Sources

Tables 4 and 5 provide a detailed record of the data sources used to construct the nominal price, dividend, and earnings series that, as emphasized in Section 2, form the basis for our later analysis of the overall stock market, the Industrials sector, the Utilities sector, as well as of the Railroads sector. They allow for the construction of a quarterly price series in the beginning and the end of each quarter, a quarterly dividend series, and an earnings series, which is annual until the end of 1935 and quarterly thereafter. Details of the construction follow in subsequent sections.

In principle, three data sources are involved in the construction of our data. For the first part of our sample until the mid-1920s for prices and earnings and until the 1930s for dividends, we rely on Cowles (1939). Thereafter, we switch to the Standard & Poor's Security Price Index Record and rely on its information as long as it is provided. For prices, this enables us to use data until the latest publication of this record, which is the 2010 publication. For dividends and earnings, we need to switch to the Standard & Poor's Analysts' Handbook for dividends and earnings information of Railroads starting in 1982 and for the overall stock market as well as the remaining two sectors in the fourth quarter of 2001.

A.2 Prices

Cowles (1939) provides a monthly time series, referred to as series P, between December 1872[75] and December 1938 for the overall stock market as well as for the three sectors under consideration.[76] Importantly, it needs to be kept in mind that the price data from Cowles (1939) is computed as the simple average between the low and the high price for the month.[77] For the beginning-of-month series, we use the monthly price for the beginning of the quarter and for the end-of-month series the monthly price reported in the last month of the quarter. Given the reporting convention of averaging the high and low price of the month, we are aware that both of these numbers have to be taken as approximations.[78]

[75] As already mentioned in Section 2, the price data actually starts in January 1871 already. Because the first uniform earnings observation across the overall stock market and the three sectors under consideration does not occur until the end of the year 1872, we disregard the first 23 monthly price observations.

[76] Because of World War I, Cowles (1939) does not report any price data between August and November 1914, as the New York Stock Exchange was closed during that time. This also affects the total return indices that we, as decsribed in the next section, use for the construction of the dividend series. In consequence, there is no observation at the end of the third quarter of 1914 as well as in the beginning of the fourth quarter of 1914. With regard to the dividend series, we observe an unusually high dividend for the overall stock market, the Industrials, the Utilities, as well as the Railroads sector at the end of the fourth quarter of 1914, which leads us to assume that this dividend captures both the third and the fourth quarter of 1914, with reinvestment occuring at the end of the fourth quarter of 1914. With regard to the later investment strategy, we simply omit the beginning of the fourth quarter of 1914 as a rebalancing date and preserve the allocation from the beginning of the third quarter of 1914. In contrast, the actual evolution of the capital for any strategy is averaged geometrically to fill in the missing value in the third quarter of 1914. Finally, as it will become apparent in the subsequent section on earnings, the earnings information from Cowles (1939) is annual, whose reporting is therefore not affected by the closure of the New York Stock Exchange between August and November.

[77] See page 17 of Cowles (1939).

[78] Wilson and Jones (2002), focusing on the overall stock market data, suggest a correction of the averaged monthly series to reflect end-of-month information more accurately. For this purpose, they rely on accompanying information from the Dow index. We have decided not to follow their procedure, as it relies on the assumption that the different samples underlying the Cowles (1939) and later on Standard Statistics data follow nearly identical price paths, which is not verifiable. Therefore, the magnitude of the approximation error by using the averaged monthly Cowles (1939) can, in our view, not be sensibly compared to the approximation error that their analysis introduces. Moreover, their analysis only applies to the overall stock market and does not provide any guidance on how to proceed for the sector data that we consider.

Information for All Stocks, Industrials, and Utilities	Time Period and Source
Prices	December 1872 - December 1938: Cowles (1939)
	January 1926 - December 1931: Security Price Index Record 1980 [weekly data]
	January 1932 - December 2001: Security Price Index Record 2002 [only until September 2001 for Utilities]
	January 2000 - December 2001: Security Price Index Record 2006 [only for Utilities]
	January 2002 - December 2007: Security Price Index Record 2008
	January 2008 - December 2009: Security Price Index Record 2010
	January 2010 - January 2013: Bloomberg
Dividends	Q1 1873 - Q4 1934: Cowles (1939)
	Q1 1935 - Q3 2001: Security Price Index Record 2002
	Q1 1996 - Q3 1999: Analysts' Handbook, Monthly Supplement for December 1999
	Q4 1999 - Q3 2001: Analysts' Handbook, Monthly Supplement for December 2001
	Q4 2001 - Q1 2005: Analysts' Handbook, Monthly Supplement for June 2005
	Q2 2005 - Q3 2008: Analysts' Handbook, Monthly Supplement for December 2008
	Q4 2008 - Q1 2012: Analysts' Handbook, Monthly Supplement for June 2012
	Q2 2012 - Q4 2012: Analysts' Handbook, Monthly Supplement for January 2013
Earnings	1872 - 1925: Cowles (1939)
	1926 - 1934: Security Price Index Record 2002
	Q1 1935 - Q3 2001: Security Price Index Record 2002
	Q4 2001 - Q1 2005: Analysts' Handbook, Monthly Supplement for June 2005
	Q2 2005 - Q3 2008: Analysts' Handbook, Monthly Supplement for December 2008
	Q4 2008 - Q2 2012: Analysts' Handbook, Monthly Supplement for September 2012

Table 4: This table lists the resources used for the construction of the nominal time series for prices (both for the beginning-of-the-quarter and the end-of-the-quarter series), dividends, and earnings, as referred to in Section 2.1 for the overall stock market, for the Industrials sector, and for the Utilities sector. "Security Price Index Record" refers to the appropriate volume of Standard and Poor's Statistical Service (1978-2010), whereas "Analysts' Handbook" is a reference to Standard and Poor's Statistical Service (1982-2012). The information for the price series is quoted monthly for ease of reference. Ultimately, we construct two quarterly price series, a beginning-of-quarter series and an end-of-quarter series. The quarterly dividend information from Cowles (1939) is not readily available but needs to be inferred from price return and total return information therein. There is an overlap between January 1926 and December 1938 in the price data that is used to re-base the data from Cowles (1939) to match it with the S&P data. Another overlap in the price series occurs for the Utilities sector between January 2000 and September 2001, which is needed to re-base the S&P information after September 2001 to match it with the prior information.

Information for Railroads	Time Period and Source
Prices	December 1872 - December 1938: Cowles (1939)
	January 1926 - December 1931: Security Price Index Record 1980 [weekly data until end of 1929]
	January 1932 - December 1975: Security Price Index Record 2002
	January 1976 - December 1985: Security Price Index Record 1986 [middle of first week for beginning of month, middle of last week for end of month, proper month end since 1983]
	January 1986 - December 1989: Security Price Index Record 1992 [middle of first week for beginning of month, proper month end for end of month]
	January 1990 - December 1995: Security Price Index Record 1996
	January 1996 - December 2001: Security Price Index Record 2001
	January 2000 - December 2003: Security Price Index Record 2004
	January 2004 - December 2007: Security Price Index Record 2008
	January 2008 - December 2009: Security Price Index Record 2010
	January 2010 - January 2013: Bloomberg
Dividends	Q1 1873 - Q4 1934: Cowles (1939)
	Q1 1935 - Q4 1981: Security Price Index Record 2002
	Q1 1982 - Q3 1984: Analysts' Handbook, Monthly Supplement for December 1984
	Q4 1984 - Q1 1988: Analysts' Handbook, Monthly Supplement for June 1988
	Q2 1988 - Q3 1991: Analysts' Handbook, Monthly Supplement for December 1991
	Q4 1991 - Q2 1995: Analysts' Handbook, Monthly Supplement for September 1995
	Q3 1995 - Q1 1999: Analysts' Handbook, Monthly Supplement for June 1999
	Q2 1999 - Q3 2001: Analysts' Handbook, Monthly Supplement for December 2001
	Q4 2001 - Q4 2002: Analysts' Handbook, Monthly Supplement for March 2003
	Q1 2003 - Q3 2006: Analysts' Handbook, Monthly Supplement for December 2006
	Q4 2006 - Q2 2010: Analysts' Handbook, Monthly Supplement for September 1995
	Q3 2010 - Q1 2012: Analysts' Handbook, Monthly Supplement for June 2012
	Q2 2012 - Q4 2012: Analysts' Handbook, Monthly Supplement for January 2013
Earnings	1872 - 1925: Cowles (1939)
	1926 - 1934: Security Price Index Record 2002
	Q1 1935 - Q4 1981: Security Price Index Record 2002
	Q1 1982 - Q3 1984: Analysts' Handbook, Monthly Supplement for December 1984
	Q4 1984 - Q1 1988: Analysts' Handbook, Monthly Supplement for June 1988
	Q2 1988 - Q3 1991: Analysts' Handbook, Monthly Supplement for December 1991
	Q4 1991 - Q2 1995: Analysts' Handbook, Monthly Supplement for September 1995
	Q3 1995 - Q1 1999: Analysts' Handbook, Monthly Supplement for June 1999
	Q2 1999 - Q3 2001: Analysts' Handbook, Monthly Supplement for December 2001
	Q4 2001 - Q4 2002: Analysts' Handbook, Monthly Supplement for March 2003
	Q1 2003 - Q3 2006: Analysts' Handbook, Monthly Supplement for December 2006
	Q4 2006 - Q2 2010: Analysts' Handbook, Monthly Supplement for September 1995
	Q3 2010 - Q2 2012: Analysts' Handbook, Monthly Supplement for September 2012

Table 5: This table is the analogue of Table 4, describing the data sources used for the construction of the price, dividends, and earnings data of the Railroads sector. Conventions are identical to those used for Table 4. The same price overlaps exist as for the Utilities sector in Table 4.

Starting in 1926, we obtain price data from Standard and Poor's Statistical Service (1978-2010). When constructing the beginning-of-month series as well as the end-of-month series, we use spot prices taken from the end of business days, moving away from the averaging convention in Cowles (1939). For the overall stock market as well as for the Industrials and the Utilities sector, this amounts to using price data from the first business day of the first month of a quarter for the beginning-of-quarter series and data from the last business day of the last month of a quarter for the end-of-quarter series. Prior to 1932, we only have access to data from Wednesday of each week. In this case, we use the first Wednesday of a quarter for the beginning-of-quarter series and the last Wednesday of a quarter for the end-of-quarter series. The situation for the Railroads sector bears two differences. First, we only have to rely on weekly data between 1926 and 1929, as the daily price series for this sector starts already in 1930. However, for a period between 1979 and 1989, we have only been able to uncover weekly data together with end-of-month information between 1983 and 1989. We were therefore forced to rely on data from Wednesdays for the beginning-of-quarter series between 1979 and 1989 and for the end-of-quarter series between 1979 and 1983.

The price series undergoes one re-basing for the overal stock market and the Industrials sector and a total of two re-basing operations for Utilities and Railroads. A re-basing that is common to all occurs between the data from Cowles (1939) and the information provided in the Security Price Index Record starting in 1926. In each of the four cases, for the overall stock market as well as the three sectors, we take monthly data from Cowles (1939) between January 1926 and December 1938 together with data from the end of the month from the Security Price Index Record for the same time period and compute the mean of the ratio. This mean is then used to scale the data from Cowles (1939). The merged series contains data from Cowles (1939) until the end of 1925 and data from the Security Price Index Record thereafter. The second re-basing only affects the Utilities and the Railroads sector, for which Standard & Poor's rebases their indices around 2000. Between January 2000 and September 2001, we take data from both the beginning of each month and the end of each month and again compute the mean of the ratio between the two series. After scaling the latter part of the sample, we use the earlier data until September 2001 when we switch to the re-scaler information.

In a final step, we re-scale all merged series (beginning-of-quarter as well as end-of-quarter), so that the value of the nominal end-of-quarter index in the beginning of the sample (December 1872) is equal to \$1.

A.3 Dividends

The purpose of the dividend series for the overall stock markets as well as for the Industrials, the Utilities, and the Railroads sector is to construct total return indices from the price indices, whose construction is described above.[79] Because of this, it is essential that the dividend series matches the frequency of the price indices (both the beginning- and the end-of-quarter series), which is quarterly. Unfortunately, the frequency of the dividend data that Cowles (1939) provides explicitly is annual, which is why we have to infer the dividend information from the interplay between the monthly price index, referred to as series P in Cowles (1939), and the monthly total return index, referred to as series C in Cowles (1939).

Explicitly, consider a generic sector and denote the price index in month t by P_t, the dividend accrued between month t and month $t+1$ by D_{t+1}, and the total return index in month t by C_t. It follows that

$$\frac{P_{t+1}+D_{t+1}}{P_t} = \frac{C_{t+1}}{C_t} \Leftrightarrow D_{t+1} = \frac{C_{t+1}}{C_t}P_t - P_{t+1},$$

which enables us to infer the dividend for month $t+1$. We obtain monthly dividend data from January 1873 (prices from Cowles, at least as far as we consider them, start in December 1872) until December 1934. In contrast to the price and earnings data, we do not switch to the Security Price Index Record in 1926, as this source only provides annual dividend information between 1926 and 1934 that, as mentioned above, is inappropriate for our purposes. As the underlying stock universe of Cowles (1939) is designed to match the universe of stocks in the Security Price Index Record closely in the 1930s[80], we are confident that the more frequent dividend observations from Cowles (1939) will nearly accurately reflect the actual occurrences referenced in the Security Price Index Record.

[79] For the construction procedure of the total return indices, see Section 2.

[80] See our discussion on Section 2.

We aggregate the monthly dividend information into quarterly numbers and then apply the re-basing factor of the Cowles data that has been constructed from overlapping price data, as described in the previous section.

Starting in 1935, we use dividend information from the Security Price Index Record that is however reported as 12-month trailing sums. In order to properly disaggregate these 12-month sums into quarterly information, we rely on accompanying information from the monthly supplements of the Analysts' Handbook, which provides actual quarterly dividends starting in 1996. Using these quarterly numbers, we are able to infer the actual quarterly dividend observations from the 12-month sums all the way back from 1996 until 1935. Importantly, this kind of procedure does not introduce any kind of forward-looking bias, as it does not introduce any new information, but merely serves as a disaggregating device for existing information.

As a second-to-last manipulation, it is necessary to point out that the re-basing for Utilities and Railroads around 2000 also needs to be imposed on the dividends, using the factor derived from price data. Finally, the dividend series for the entire sample period needs to be scaled to reflect the normalization of the price at the end of fourth quarter 1972 to \$1.

A.4 Earnings

From 1871 until 1925, earnings come from Cowles (1939). Observations are annual and refer to 12-month trailing sums. In reporting the earnings data, Cowles (1939) differentiates between the overall set of companies that are incorporated in the price data (series P for index data as well as series C for total return data) and a smaller set of companies, for whom earnings information is available to construct sector earnings indices. For ease of reference, Cowles (1939) provides annual price averages (matching the frequency of the earnings data) data for both the overall set of companies as well as for the set of companies incorporated in the earnings series. This enables us to construct an approximation to the earnings of the overall set of companies by multiplying an annual earnings observation by the ratio of the price average of the overall set divided by the average of the smaller set of companies with earnings data.[81]

The earnings data from Cowles (1939) is then scaled to re-base it to the level of the data from the Security Index Record, using the re-basing factor derived from the price data as described above.

The earnings data after 1926 comes from the Security Price Index Record and is later complemented by information from the Analysts' Handbook, according to the information provided in Tables 4 and 5. It comes from the earnings per share series that is adjusted to the price index.[82] Whereas the observations 1926 until 1934 are annual, the frequency switches to quarterly in 1935. We employ the actual quarterly series that we subsequently aggregate into 12-month trailing sums. The Security Price Index Record provides both actual quarterly earnings numbers and 12-month trailing sums. However, Standard & Poor's emphasizes in the Security Price Index Record[83] that for the period 1935-41, the 12-month trailing sums are based on annual earnings of the companies in the index at the end of each quarter. In order to avoid this unnecessary imprecision, we have decided to aggregate quarterly earnings numbers and to consistently apply this convention thorughout the entire time period, for which we rely on S&P data.

As with the dividend data, two final manipulations need to be addressed. First, the re-basing for Utilities and Railroads around 2000 also needs to be imposed on the earnings for these two sectors, using the factor derived from the relevant price data. Finally, the earnings series for the entire sample period again needs to be scaled to reflect the normalization of the price of the respective sector (as well as for the price of the overall stock market in the case of the overall market data) at the end of fourth quarter 1972 to \$1.

[81] Shiller (1992) applies the same procedure for the overall market data from Cowles (1939). The ratio of annual earnings and the annual average of price data for the companies with earnings is also referred to as series R in Cowles (1939). Shiller (1992) then multiplies the R series by the annual average of the P series, which is exactly the same operation as described in the text.

[82] Comparing the earnings numbers in the Security Price Index Record to the reported numbers in the Analysts' Handbook, it turns out that the earnings numbers are diluted earnings per share number, an information that the Security Price Index Record does not provide explicitly.

[83] See, for example, page 462 in the 2004 volume of Standard and Poor's Statistical Service (1978-2010).

B Construction Details for the Denominator of the CAPE Ratio

This section relates to Section 4.1 and deals with an unaddressed detail concerning the selection of the earnings observations that enter the CAPE ratio. The problem is that our earnings observations change frequency in the beginning of 1935. 12-month earnings numbers are annual up to and including 1935 and quarterly afterwards. This issue is further complicated by the fact that we impose a six-month lag on the earnings observations, in order to address the look-ahead problem.

To address this issue, we therefore distinguish between three cases in the construction of the denominator of the CAPE ratio. At the end of each quarter up to and including the second quarter of 1936, we solely rely on the annual earnings observations from Cowles (1939) to compute a CAPE ratio, applying the same procedure as outlined for the pre-1935 period in Section 4.2. For this part of the sample, this procedure implies that the earnings lag that we impose varies between six months (for a CAPE computation at the end of the second quarter of a year) and 15 months (for a CAPE computation at the end of the first quarter of a year). Clearly, the limited availability of the earnings data necessitates what we have just outlined. However, the impact of the varying lag is limited by the fact that the maximum lag (15 months) is close to only a tenth of the total time period, over which we consider earnings (10 years).

For any CAPE number computed from the third quarter of 1936 until the second quarter of 1945, we have to deal with the transition from annual to quarterly numbers in the year 1935. As an example, consider the earnings numbers for the computation in the fourth quarter of 1936. Because of the three-month lag, the first earnings observation to consider would be from the end of the second quarter of 1936. As this is a 12-month trailing sum, it reaches all the way back to the beginning of the third quarter of 1935. Now, we cannot use the earnings observation at the end of the second quarter of 1935, as it reaches well into the year 1934, for which we only have an annual earnings observation, which would result in double-counting some earnings observation. For this purpose, we have singled out the quarterly earnings from 1935, have adjusted them for inflation and have individually (at the quarterly level) scaled them to reflect total returns. Within our example, we add two quarters of these quarterly earnings into the earnings sum used for the denominator of the CAPE ratio, which gets us to the beginning of the year 1935. Now, we continue summing 12-month trailing earnings, until we have collected ten 12-month trailing observations. Of course, this results in ten and a half years of earnings for the denominator of CAPE, which is why we divide the sum of these earnings by 10.5 instead of 10. In general, it will depend on the quarter of consideration whether we average exactly ten years of earnings.[84]

If, instead of a fourth quarter during the period between the third quarter of 1936 and the second quarter of 1945, we select earnings observations to be used in the computation of the CAPE ratio for a first quarter, we apply an analogous procedure to the one described above and we end up with 10.75 years of earnings (using observations from the first through third quarter of 1935). Finally, for a third quarter, we obtain 10.25 years, whereas no adjustment is necessary at all for any earnings numbers involved in the computation of a second-quarter CAPE ratio between the third quarter of 1936 and the second quarter of 1945.

C Investment Strategy for Changing Sector Universe

In the main text, we have adhered to the principle of consistency with regard to the sector classification that we use. We have aimed to provide very long-term evidence for the predictive ability of the CAPE ratio and the derived Relative CAPE indicator and have translated this evidence into a hypothetical rotation strategy between the Industrials, Utilities, and Railroads sector from 1902 until 2013. In this section, we would like to emphasize the robustness of our hypothetical strategy when we actually attempt to incorporate changes in the reporting of top-level sectors of the US economy.

[84] Another approach to deal with the change of the frequency of the earnings observation could have been the application of the procedure pre-1935, as described in the previous paragraph, for another ten years until the full earnings history of the CAPE ratio only refers to quarterly numbers. This procedure however completely disregards much of the detail of the quarterly earnings observations between the mid-1930s and the mid-1940s, which we find undesirable.

For this purpose, we consider the switch of S&P from the reporting of quantities for the Industrials, Utilities, and Railroads sector to the reporting of numbers for the Industrials, Utilities, Transportation, and Financials sector.[85] The Railroads sector hereby becomes a part of the Transportation sector. For the Transportation sector, we use data from Standard and Poor's Statistical Service (1978-2010) between 1975 and 2001 and from Bloomberg subsequently to obtain a time series of the CAPE ratio that starts in the fourth quarter of 1985. Similarly, data from Standard and Poor's Statistical Service (1978-2010) and Standard and Poor's Statistical Service (1982-2012) in conjunction with price index information from Bloomberg for the most recent years result in the analogous time series for the Financials sector. Our construction methodology for this valuation ratio are thereby identical to the procedure outlined in Sector 4.

Figure 37 plots the CAPE ratio for the Transportation and the Financials sector together with the appropriately shortened time series for the original three sectors as well as for the overall stock market, which we repeat from Figure 4 for ease of comparison. There is quite a similarity between the CAPE ratio of the Transportation sector and the CAPE ratio of its component, the Railroads sector. Some noteworthy differences arise in the late 2000s, when Railroads are more overvalued as expressed by their CAPE ratio before their CAPE ratio drops more dramatically. During the increase in valuation after the turn of the decade, it is then the Transportation sector that rises more extremely, while the increase in valuation for the Railroads sector is more subdued. The Financials sector's valuation increases before the turn of the century, a tendency that is in line with the pattern for the overall stock market and the Industrials sector, which at that time makes up a substantial portion of the overall market. Its decrease in valuation is however much less dramatic than for the overall stock market and the Industrials sector. Finally, it is noteworthy that the Financials sector valuation in the years prior to the 2007-2009 subprime crisis is nt experiencing any major upwards movement. The crisis itself leads to further drops in valuation, from which the Financials sector has not recovered until the end of the sample at the end of 2012.

With regard to the calculation of the Relative CAPE indicator for the Transportation and the Financials sector, we make a minor adjustment to the methodology outlined in Section 5 above. If we required twenty years worth of observations of the CAPE ratio for the two sectors, we would not be able to make any reasonable inference from the time series of their Relative CAPE indicators, as we would have less than five years worth of data. We therefore perform the first calculation for the indicator at the end of the third quarter of 1990 already, with only five years worth of observations for the respective CAPE ratios.[86] The drawback however is the fact that the notion of a long-term average of the CAPE ratio for the two sectors becomes substantially less meaningful in the presence of a smaller sample to take the mean over. For any subsequent quarter, we consider a longer and longer sample of the CAPE ratio to calculate the Relative CAPE indicator over, until we reach the end of the second quarter of 2010, which is the first quarter with a full twenty years of history for the CAPE ratio.

In terms of the hypothetical investment strategy, we only have to make one minor modification to accommodate the fact that our rotation considers four sectors instead of three. We start with the same "investment score" of 2 as in Section 8 at each rebalancing date that involves four sectors. As in Section 8, the score of the one sector with the lowest Relative CAPE indicator is then raised by 1, whereas we drop the score for the one sector with the highest Relative CAPE indicator by one. This leaves us with a very similar ordering to that described in Section 8 except for the fact that there are two "middle" sectors with an unchanged score, in contrast to one in the initial setting. Subsequently, we apply the same momentum-methodology as in Section 8 with the identical "probationary phase" to control the impact of the momentum consideration.

The switch from the three-sector rotation to the four-sector rotation occurs with the availability of the Relative CAPE indicator for the Transportation and the Financials sector, as outlined above. That is, our rotation strategy rotates between the Industrials, Utilities, and Railroads sector starting with the initial portfolio allocation in the beginning of the fourth quarter of 1902 up to and including the portfolio rebalancing in the beginning of the third

[85] Another change in the top-level reporting for sectors of the US economy is the introduction of the GICS classification system in 1999. As pointed out above, we will neglect this change, as the available data after this change is barely sufficient to calculate a single CAPE ratio, which consequently renders the consideration of the Relative CAPE indicator and a longer-term history of an investment strategy infeasible.

[86] For the sake of completeness, we would like to point out we preserve the winsorization at the 5% level at both the bottom and the top of the respective samples for the CAPE ratio.

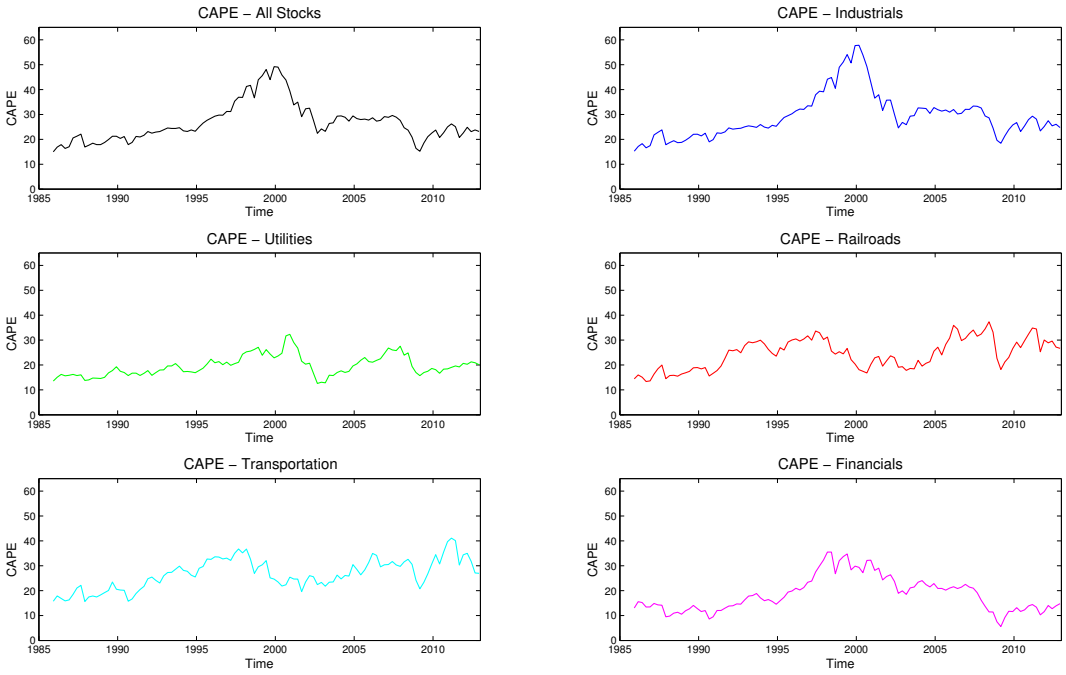


Figure 37: This figure shows the CAPE ratio for the overall stock market (All Stocks), Industrials, Utilities, Railroads, Transportation, and Financials. The construction of the CAPE ratio follows section 4. The quarterly time series start with the fourth quarter of 1985 and end with the second quarter of 2012. For the time period under consideration, the information in the panels for the overall stock market, the Industrials, the Utilities, and the Railroads sector is identical to the information from Figure 4, but is included here to simplify the comparison to the Transportation and the Financial sector.

quarter of 1990. From the rebalancing in the beginning of the fourth quarter of 1990 onwards, the strategy then takes into account the changed sector universe that then includes the Industrials and Financials sector.[87]

Figure 39 displays the performance of the rotation strategy, which is very similar to the performance of the original CAPE-based sector rotation strategy, establishing the robustness of the outlined CAPE-based sector allocation to the inclusion of changes in the sector classification methodology. The annualized geometric average of the inflation-adjusted total return of the CAPE-based strategy amounts to 6.59%, which is approximately 40 basis points lower than the return of the origina strategy, but still nearly 70 basis points annually above the return of the market benchmark, as measured over a time period of approximately 110 years. The volatility of the CAPE-based strategy at 20.61% is roughly comparable to that of the market benchmark at 20.15%. Correspondingly, the information ratio of the rotation strategy exceeds that of the market benchmark by nearly 2.60% (31.95% versus 29.39%). The outperformance in terms of the information ratio of the CAPE-based sector compared to the equally weighted benchmark remains about the same as in the original performance comparison at approximately 3%. The annual geometrically averaged return of this benchmark amounts to 5.99% with an accompanying volatility of 20.66%.

Figure 40, which is analogous to Figure 34, displays the excess returns of the CAPE-based sector rotation strategy with the sector switch over the benchmark strategies. Focusing on the rightmost comparisons that involve the modified set of sectors for the rotation, it becomes apparent that the only five-year period, in which the CAPE-based sector rotation falls somewhat behind either of the benchmarks, is a period, in which both benchmarks perform very well and the sector rotation just does not quite perform as well, and not a period with opposite return patterns. This evidence points again to the more careful allocation implied by the long-term focus of the CAPE ratio and the corresponding Relative CAPE indicator and is very consistent picture with the evidence presented for the original sector rotation.

Finally, Figures 41 and 42 are the analogues of Figures 35 and 36, displaying the outpermance ratio, as defined in

[87] We impose the same kind of change in the underlying constituents of the strategy for the equally weighted benchmark to preserve the comparability with the CAPE-based rotation strategy.

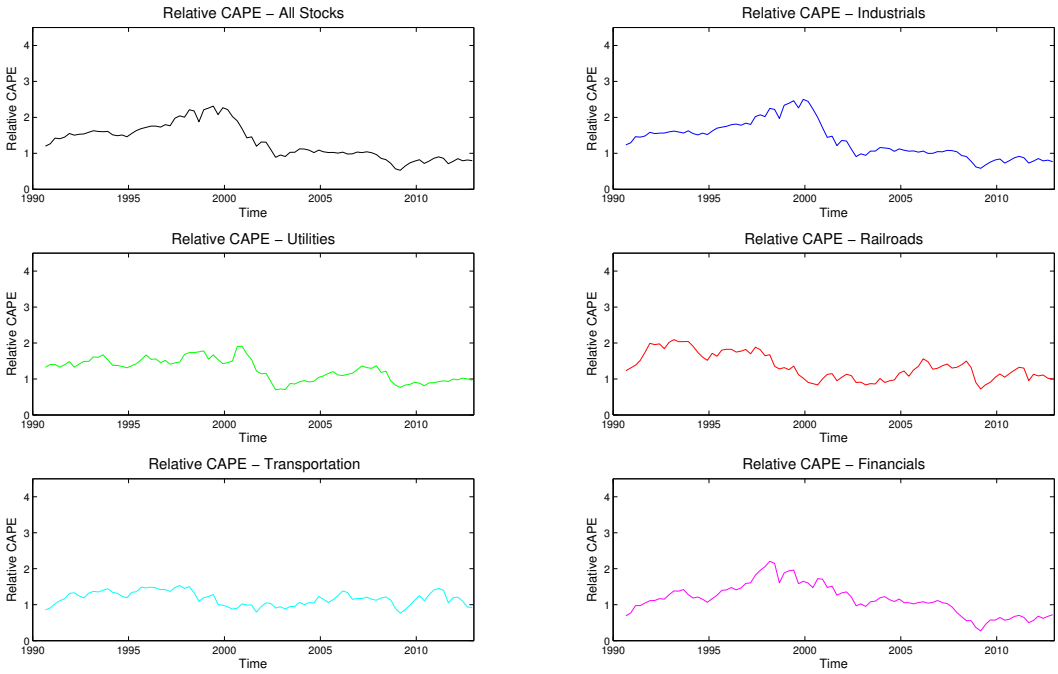


Figure 38: Analogous to the transition from Figure 4 to Figure 13, this figure contains the Relative CAPE indicators corresponding to the CAPE ratios displayed in Figure 37. The frequency of the observations is again quarterly and observations cover the time period from the end of the third quarter in 1990 until the end of the fourth quarter of 2012. Relative CAPE in a given quarter for each of the four time series is the ratio of this quarter's CAPE and a winsorized, twenty-year average of that time series' history of CAPE ratios, up to and including the current CAPE ratio, where the winsorization occurs at the 5% level. The Transportation and the Financials sector do not have twenty years of history for their respective CAPE ratios until the third quarter of 2004. Before this quarter, we use whatever time series is available for the normalization of the CAPE ratio (which remains winsorized at the 5% level).

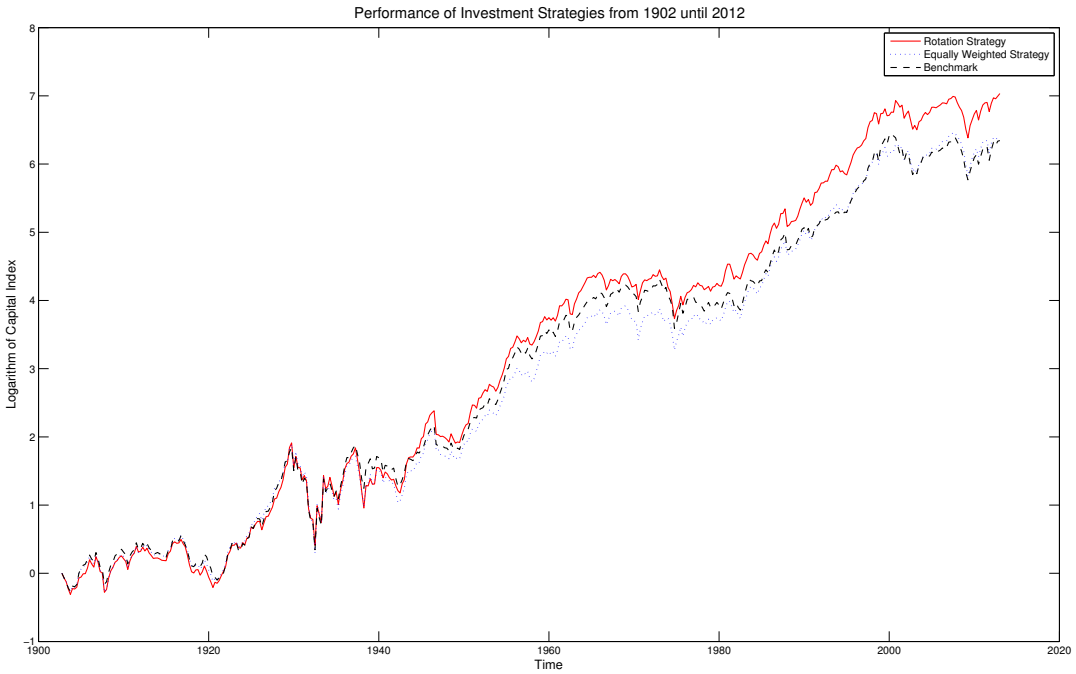


Figure 39: This figure shows the performance of three different portfolio strategies on a log-scale. Investment commences in the beginning of the first quarter of 1903. Rebalancing quarterly, all three strategies end in the beginning of the first quarter of 2013. The strategy represented by a dashed black line invests into the S&P Composite Index at each rebalancing date, which we consider as our main benchmark. For the equally weighted benchmark, marked as a dotted blue line, as well as for the CAPE-based sector rotation strategy, displayed as a solid red line, we utilize the sectors Industrials, Utilities, and Railroads for the rebalancing from the first quarter of 1903 up to and including the rebalancing in the beginning of the third quarter of 1990 and the sectors Industrials, Utilities, Transportation, and Financials thereafter until the last rebalancing in the beginning of the fourth quarter of 2012. The equally weighted benchmark puts equal on each utilized sector during each rebalancing and the CAPE-based strategy assigns weight according to the procedure outlined in Section C.

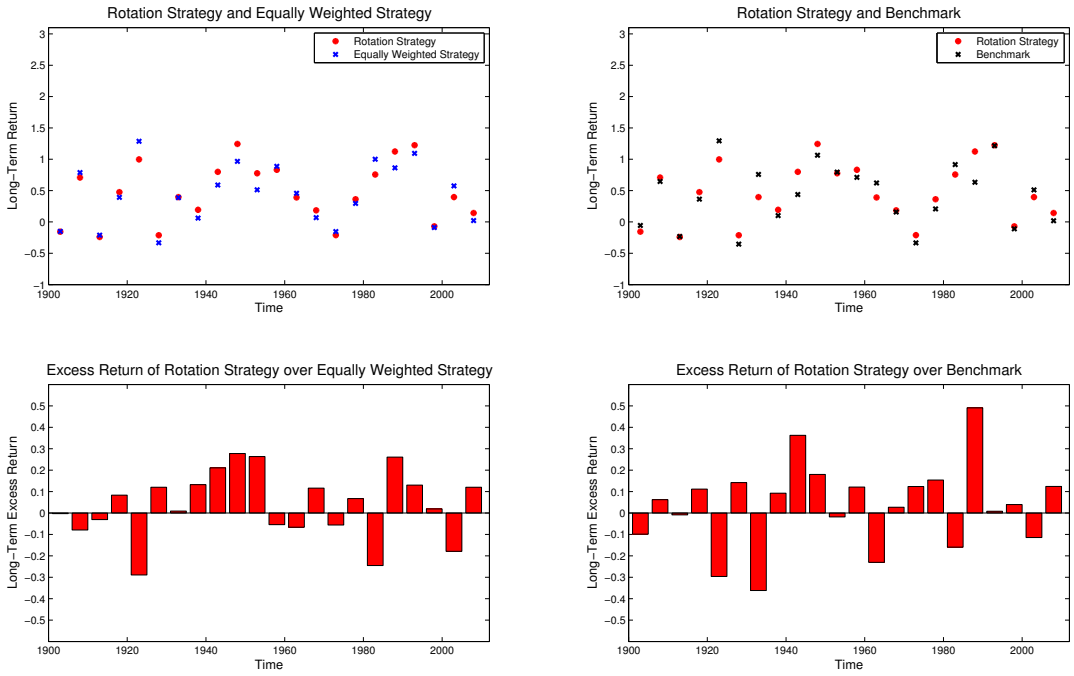


Figure 40: This figure displays the long-term performance of the hypothetical rotation strategy that initially rotates between the Industrials, Utilities, and Railroads sectors (up to and including the rebalancing in the beginning of the third quarter of 1990) and subsequently between the Industrials, Utilities, Transportation and Financials sectors when compared to the equally weighted benchmark (appropriately reflecting the available sectors at each rebalancing date) in the left two panels and to the benchmark in the right two panels. The time period between the beginning of the fourth quarter of 1902 and the beginning of the first quarter of 2013 is split into non-overlapping five-year intervals, with the very first quarter not being included in any five-year interval. The return numbers are inflation-adjusted, net total return quantities, which are not annualized. They are derived from the time series displayed in Figure 39. The top panels display the returns for both the strategy as well as the respective benchmark, whereas the bottom panels display the difference.

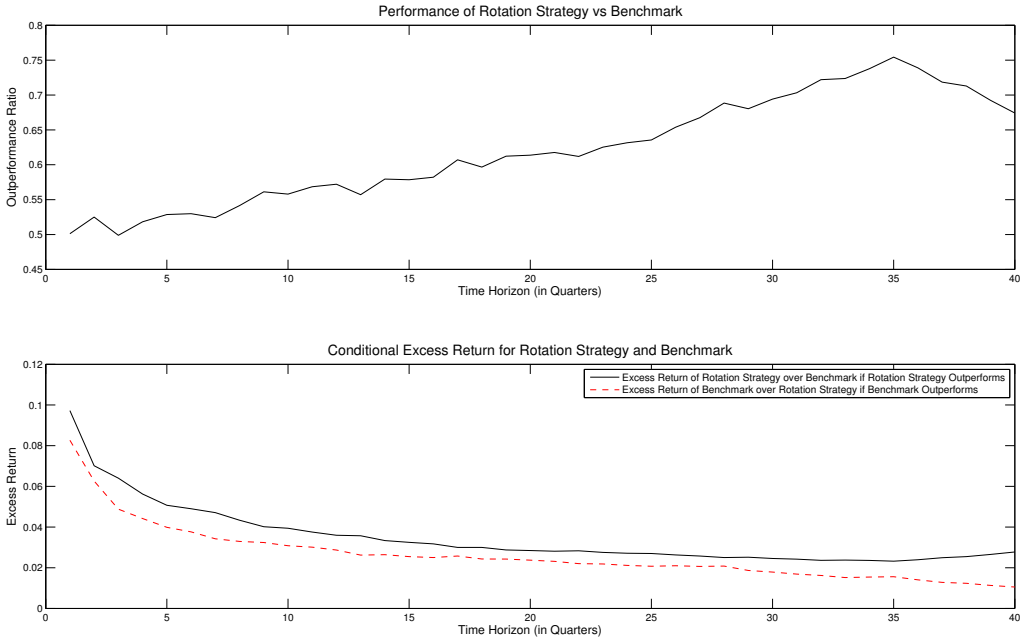


Figure 41: This figure is the analogue of Figure 35 for the CAPE-based Sector Rotation Strategy that initially rotates between the Industrials, Utilities, and Railroads sector up to and including the rebalancing in the beginning of the third quarter of 1990 and between the Industrials, Utilities, Transportation, and Financials sector thereafter.

Section 8, as well as the patterns in the annualized conditional outperformance of the CAPE-based sector rotation strategy when compared to both the market benchmark as well as the equally weighted benchmark. All results are very similar - to the degree of almost being indistiguishable - to the results presented for the original CAPE-based sector allocation strategy building on the Industrials, Utilities, and Railroads sector,[88] which again provides evidence for the robustness of the CAPE-based sector allocation methodology with regard to adaptations in the sector classification.

[88]Some degree of similarity is of course to be expected, as the two strategies share the same performance from 1902 until 1990.

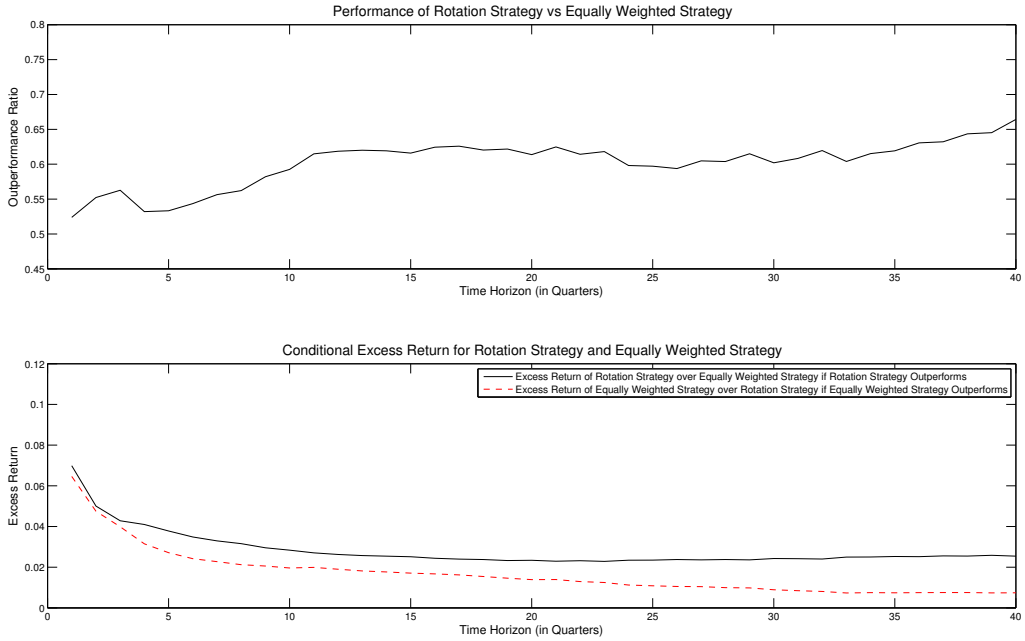


Figure 42: This figure is the analogue of Figure 36 for the CAPE-based Sector Rotation Strategy that initially rotates between the Industrials, Utilities, and Railroads sector up to and including the rebalancing in the beginning of the third quarter of 1990 and between the Industrials, Utilities, Transportation, and Financials sector thereafter. The same switch occurs for the equally weighted benchmark.