Springer Series in Reliability Engineering

Hoang Pham *Editor*

Applications in Reliability and Statistical Computing

Springer

Springer Series in Reliability Engineering

Series Editor

Hoang Pham, Department of Industrial and Systems Engineering,
Rutgers University, Piscataway, NJ, USA

Today's modern systems have become increasingly complex to design and build, while the demand for reliability and cost effective development continues. Reliability is one of the most important attributes in all these systems, including aerospace applications, real-time control, medical applications, defense systems, human decision-making, and home-security products. Growing international competition has increased the need for all designers, managers, practitioners, scientists and engineers to ensure a level of reliability of their product before release at the lowest cost. The interest in reliability has been growing in recent years and this trend will continue during the next decade and beyond.

The Springer Series in Reliability Engineering publishes books, monographs and edited volumes in important areas of current theoretical research development in reliability and in areas that attempt to bridge the gap between theory and application in areas of interest to practitioners in industry, laboratories, business, and government.

Now with 100 volumes!

Indexed in Scopus and EI Compendex

Interested authors should contact the series editor, Hoang Pham, Department of Industrial and Systems Engineering, Rutgers University, Piscataway, NJ 08854, USA. Email: hopham@soe.rutgers.edu, or Anthony Doyle, Executive Editor, Springer, London. Email: anthony.doyle@springer.com.

Hoang Pham
Editor

Applications in Reliability and Statistical Computing

Springer

Editor
Hoang Pham
Department of Industrial and Systems
Engineering
Rutgers, The State University of New Jersey
Piscataway, NJ, USA

ISSN 1614-7839 ISSN 2196-999X (electronic)
Springer Series in Reliability Engineering
ISBN 978-3-031-21231-4 ISBN 978-3-031-21232-1 (eBook)
https://doi.org/10.1007/978-3-031-21232-1

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To My Father on His 90th Birthday!

Preface

We're living in an era of fast and unpredictable change. Billions of people are connected to each other through their mobile devices. Data is being collected and processed each day like never before. With 5G and IoT set to generate an estimated 1 billion terabytes of data by 2025, companies continue to search for new techniques and tools that can help them practice data collection effectively in promoting their business. A large portion of this data will come from smart devices, smart communities. The era of big data through reliability and statistical computing with almost all applications in our daily life has experienced a dramatic shift in the past two decades to a truly global industry. The forces that have driven this change are still at play and will continue. Most of the products which affect our daily lives are becoming even more complex than ever.

The book consists of 15 chapters that covers a selection of recent developments and applications on various related topics in reliability and statistical computing. The emphasis of this book is on the practical applications of reliability and statistical methods and techniques in various disciplines using machine learning, risk assessment, modeling and optimization, and other computational methods.

All chapters in the book are written by leading researchers and practitioners in their respective fields with a hope to connect the gap between the theoretical and practical computations in the application areas of reliability and statistical computing.

I acknowledge Springer for this opportunity and professional support. Importantly, I would like to thank all the chapter authors and reviewers for their availability for this work.

Piscataway, NJ, USA
September 2022

Hoang Pham

Contents

Editor and Contributors

About the Editor

Hoang Pham is a Distinguished Professor and former Chairman (2007–2013) of the Department of Industrial and Systems Engineering at Rutgers University. Before joining Rutgers in 1993, he was a Senior Engineering Specialist with the Idaho National Engineering Laboratory, Idaho Falls, Idaho and Boeing Company in Seattle, Washington. His research areas include reliability modeling and prediction, software reliability, and statistical inference. He is editor-in-chief of the International Journal of Reliability, Quality, and Safety Engineering and editor of the Springer Series in Reliability Engineering and has been conference chair and program chair of over 50 international conferences and workshops. Dr. Pham is the author or coauthor of 7 books and has published over 200 journal articles and 100 conference papers, and edited 17 books including Springer Handbook in Engineering Statistics and Handbook in Reliability Engineering. He has delivered over 40 invited keynote and plenary speeches at many international conferences and institutions. His numerous awards include the 2009 IEEE Reliability Society Engineer of the Year Award. He is a Fellow of the IEEE, AAIA, and IISE.

Contributors

Grigore Albeanu "Spiru Haret" University, Bucharest, Romania

Silvia Carpitella Department of Manufacturing Systems Engineering and Management, California State University, Northridge, USA

Antonella Certa Department of Engineering, University of Palermo, Palermo, Italy

Kehan Gao Department of Computer Science, Eastern Connecticut State University, Willimantic, CT, USA

Hitesh Department of Computer Science & Engineering, Amity School of Engineering and Technology, Amity University, Noida, Uttar Pradesh, India

Bui Thanh Hung Data Science Laboratory, Faculty of Information Technology, Industrial University of Ho Chi Minh City, Ho Chi Minh City, Vietnam

Joaquín Izquierdo Institute for Multidisciplinary Mathematics, Universitat Politècnica de València, Valencia, Spain

Umer Khan Amity University, Noida, Uttar Pradesh, India

Taghi Khoshgoftaar Department of Computer and Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL, USA

Gregory Khvatsky Lucy Family Institute for Data & Society, Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN, USA

Junji Koyanagi Graduate School of Tottori University, Tottori-City, Tottori, Japan

Vijay Kumar Department of Mathematics, Amity Institute of Applied Sciences, Amity University, Noida, Uttar Pradesh, India

Valentina Kuskova Lucy Family Institute for Data & Society, University of Notre Dame, Notre Dame, IN, USA

Stephen H.-T. Lihn Quant Research at Atom Investors LP, Austin, TX, USA

Tingnan Lin Department of Industrial and Systems Engineering, Rutgers University, Piscataway, NJ, USA

Henrik Madsen Danish Technical University, Lyngby, Denmark

W. Y. Man Department of Engineering (Avionics), Heliservices (HK) Ltd, Hong Kong SAR, China

Alexandra-Ştefania Moloiu R&D Department, TypingDNA, Bucharest, Romania

Ilyas Mzougui Faculty of Sciences and Technologies, Abdelmalek Essaadi University, Tangier, Morocco

Hoang Pham Department of Industrial and Systems Engineering, Rutgers University, Piscataway, NJ, USA

Florin Popenţiu-Vlădicescu University "Politehnica" of Bucharest & Academy of Romanian Scientists, Bucharest, Romania

Himanshu Sharma Amity University, Noida, Uttar Pradesh, India

Avinash K. Shrivastava International Management Institute, Kolkata, West Bengal, India

Devanshu Kumar Singh Department of Computer Science & Engineering, Amity School of Engineering and Technology, Amity University, Noida, Uttar Pradesh, India

Anna Sokol Lucy Family Institute for Data & Society, Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN, USA

Fengbin Sun Reliability Engineering, Tesla Inc, Palo Alto, CA, USA

Sarah Tasneem Department of Computer Science, Eastern Connecticut State University, Willimantic, CT, USA

Anshul Tickoo Amity University, Noida, Uttar Pradesh, India

Rachel Wesley Department of Industrial and Systems Engineering, Rutgers University, Piscataway, NJ, USA

Eric T. T. Wong Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China

Dmitry Zaytsev University of Notre Dame at Tantur, Jerusalem, Israel

Forecasting The Long-Term Growth of S&P 500 Index

Stephen H.-T. Lihn

Keywords Forecast model · Economic cycle · Mean reversion · CAPE · Stock market · Wavelet

JEL Classification: C38 · C53 · E32 · E37 · E47

1 Introduction

The U.S. stock market has exhibited amazing resilience in the long run. Its long-term growth is a wonderful story of American capitalism. In the past 200 years, it has produced a consistent real return of about 6.6% per year (Fig. 1 and Siegel [20]). However, this wonderful return comes with many ups and downs every decade. In some cases, the market went down more than 50%. In other cases, the market was stagnant for more than a decade. The longest and largest drawdown in history was from 1929 to 1948. More recently, the peak reached in 2000 had not been surpassed until 2012. Making things more intricate, these two large bear markets were preceded by two strongest ten-year bull markets in history. How do we make sense of them? More importantly, are they forecastable?

Disclaimer: The views in this paper are solely the responsibility of the author. They don't reflect the views of the company, nor does this paper contain any propietory data from the company, v1.0, Released on September 30, 2021.

S. H.-T. Lihn (✉)
Quant Research at Atom Investors LP, Austin, TX, USA
e-mail: stevelihn@gmail.com
URL: http://www.linkedin.com/pub/stephen-horng-twu-lihn/0/71a/65

H. Pham (ed.), *Applications in Reliability and Statistical Computing*, Springer Series in Reliability Engineering, https://doi.org/10.1007/978-3-031-21232-1_1

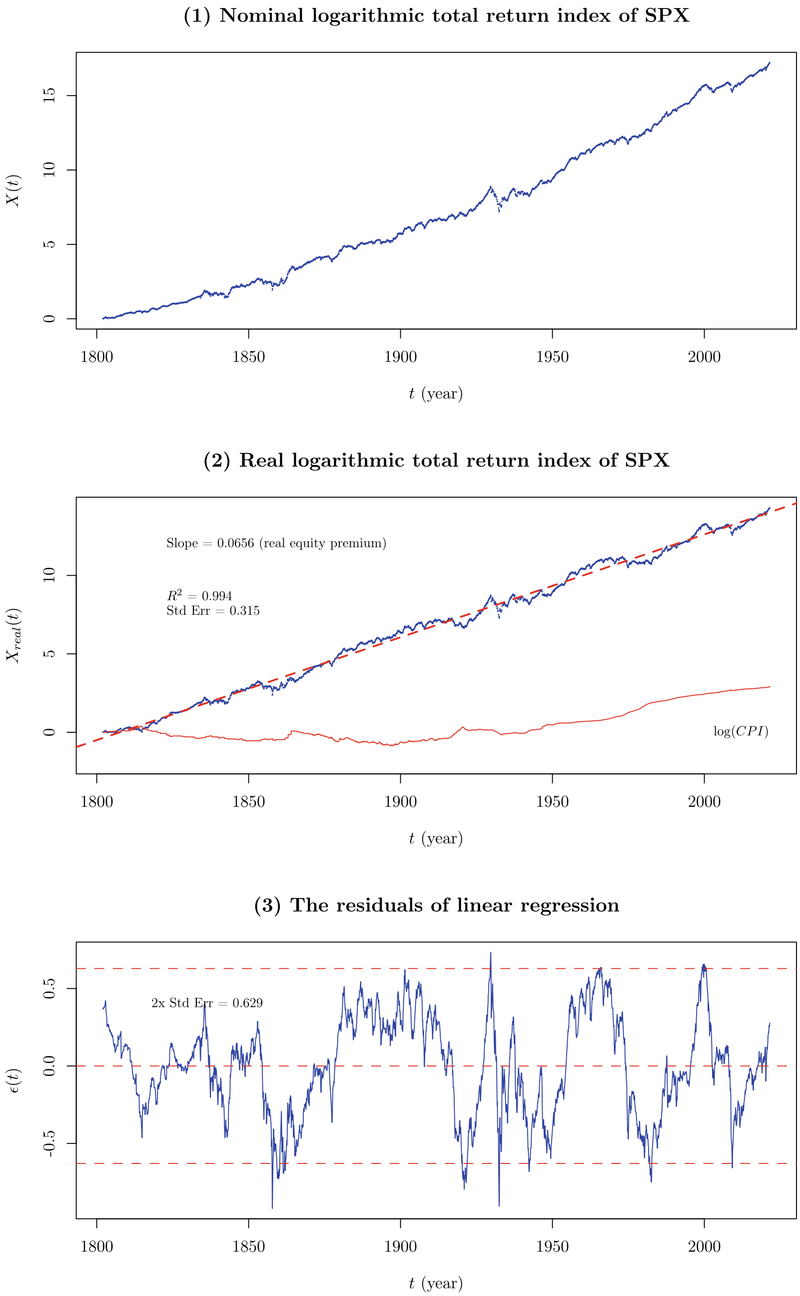

Fig. 1 Panel (1) The nominal total return index for the U.S. stock market $X(t)$ in the logarithmic scale since 1802. Panel (2) shows the real total return index $X_{\text{real}}(t)$. The slope $\beta_{\text{rep}} = 6.55\%$ is the long-term real equity premium over inflation. The linear regression has an impressive $R^2 = 0.994$ with the standard error of 0.32. Panel (3) shows the mean-reverting behavior of the residuals $\epsilon(t)$

For most typical investors, It is believed that the S&P 500 index (SPX) is the best single gauge of the U.S. stock market.[1] This index consists of 500 largest public corporations in the U.S., weighted by their market capitalizations. In a 2017 interview with CNBC,[2] Warren Buffett said, "Consistently buy an S&P 500 low-cost index fund, I think it's the thing that makes the most sense practically all of the time." At the 2021 Berkshire Hathaway annual meeting, he reiterated his conviction, "I just think that the best thing to do is buy 90% in S&P 500 index fund."[3] What is the rationale behind these statements? How much faith should we have in it? What kind of returns can be expected from SPX if we "surrender our freedom", so to speak, of selecting from thousands of stocks, mutual funds and ETFs. This research is intended to answer some of these questions in an econometric setting.

Recent application of trend filtering technique has revealed linear characteristics of market trends [16]. In the short term, the market process is highly lepkurtotic (kurtosis $\gg 3$) and influenced heavily by the underlying volatility process. In the long term, however, the market process is not a random walk process. It is a mean-reverting process with linear growth. More interestingly, when the time horizon is extended to decades, the mean-reverting process is slightly platykurtic (kurtosis < 3), which is strikingly different from the lepkurtotic random walk process observed in the short term.

The mean-reverting process can be confirmed by the model-free wavelet analysis. The Morlet wavelet [13, 14] provides the ability to decipher the market cycles in a financial time series. By applying the wavelet analysis to both the 10-year and 20-year returns, we are able to show that the U.S. stock market exhibited a 36-year cycle after World War II (WWII).

Next, we review the algorithm developed in [10] that separates the mean-reversion component from the linear growth component in the market process. The mean-reversion component is associated with the "cyclically adjusted P/E ratio", aka CAPE [1], in a profound way. The nickname of our model is called "jubilee tectonic model". The "jubilee" name comes from its optimal trend-following window of 45 years and the periodicity of 36 years from the wavelet analysis. The "tectonic" name comes from the hypothesis that there are fault lines in the historical CAPE, which can be calibrated and corrected in this model through statistical learning. Such "model breaks" have been categorically discussed in Chap. 19 of [7]. We apply a more restrictive approach to capture these breaks, and attempt to give them economic interpretation when appropriate.

The forecast of future equity return is an important topic for policy makers and asset allocators. Research from Vanguard [6] found that "many commonly cited signals have had very weak and erratic correlations with actual subsequent returns." CAPE remains one of the most powerful predictors. Even then, it has explained only

[1] https://www.spglobal.com/spdji/en/indices/equity/sp-500/.

[2] https://www.cnbc.com/2017/05/12/warren-buffett-says-index-funds-make-the-best-retirement-sense-practically-all-the-time.html.

[3] https://www.cnbc.com/2021/05/03/investing-lessons-from-warren-buffett-at-berkshire-hathaway-meeting.html.

about 34% of the time variation.[4] Recent forecasts using CAPE have been over-pessimistic. The lofty CAPE issue continues to trouble the academic community, as [24] wrote on Project Syndicate: “It is impossible to pin down the full cause of the high price of the U.S. stock market.” In an attempt to address such issue, Siegel [21] studied six variations: reported earnings, operating earnings, and NIPA profits, in combination with price index portfolio and total return portfolio. The R^2 was increased from 34 to 40% in the best case scenario.

In the jubilee tectonic model, the tectonically adjusted CAPE, plus mean reversion and inflation, form the five-factor econometric model that forecasts long-term equity returns with R^2 above 80%. This model produces different predictions for the future: The original CAPE model predicts below average real returns for the next decade. But the jubilee tectonic model predicts much higher returns and very positive outlook for the next decade.

1.1 Objectives

The key points of this chapter are:

- Setup, global linear regression, and equity risk premium
- Wavelet analysis on periodicity
- Channel deviation framework and CAPE
- The 20-year forecast model.

1.2 Data Sources, Tools, and Abbreviations

This chapter uses the **jubilee** package [11] and the **WaveletComp** package [17] in R to produce the analysis. The S&P 500 data in the **jubilee** package is assembled from several original sources. The main data source is from Shiller’s online data website [23]. The excel file “ie_data.xls” contains monthly averaged prices, dividends, and earnings of SPX since 1871.[5] It also contains consumer price index (CPI) and 10-year Treasury yield (GS10). It derives the real prices, real dividends, and real earnings, and calculates the 10-year CAPE.

The second data source is from Schwert [19], from which we obtain the stock market total return data since January of 1802. The third data source is the annual CPI data since 1800 from Minneapolis FED [12]. The fourth data source is from FRED [9] of St Louis FED, which provides daily and/or monthly online updates for many financial and economic time series.

Frequently used abbreviations are listed below:

---

[4] The 40% R^2 cited in [6] is a result of truncating the CAPE data prior to 1926. Such structural break can be explained by this research.

[5] The word “ie” stands for “Irrational Exuberance”. It is a March 2000 book written by Shiller: https://en.wikipedia.org/wiki/Irrational_Exuberance_(book).

List of Abbreviations

CAPE	Cyclically adjusted P/E ratio
CPI	Consumer price index
ETF	Exchange trade fund
GS10	Ten-year Treasury yield
SPX	The SP 500 index
TRI	Total return index.

2 Total Return Index and Equity Risk Premium

We first define the methodology of calculating stock market's logarithmic total return index (TRI). For the long-term analysis, we work with monthly interval $\Delta t = 1/12$. Assume the market index at time t is $p(t)$ and pays dividend $d(t)$ for the period from t to $t + \Delta t$. The total log-return is $r(t + \Delta t) \equiv \log(p(t + \Delta t) + d(t)) - \log\ p(t)$. And let CPI (t) be the consumer price index (CPI) at time t, we construct the nominal and real TRI in logarithmic scale as

$$
\begin{aligned}
X(t) &= \sum_{t_1 \le \tau \le t} r(\tau), && \text{nominal TRI;} \\
X_{\text{real}}(t) &= X(t) - \log\ \text{CPI}(t), && \text{real TRI.}
\end{aligned}
\tag{1}
$$

where $\{\tau\}$ represents all the months available to our analysis, and t_1 is the inception date of the data, January of 1802.

The above notation of $X(t)$ is the "continuous notation". Empirically, t is discrete. The "discrete notation" states that, at time t_i, the logarithmic index value is X_i. We use both notations depending on the context and the cleanliness of expression. We follow Shiller's convention that each month is identified by the time fraction of $t_i = y(t_i) + (m(t_i) + 1/2)\Delta t$, where $i = 1, 2, 3, \cdots$ is an integer label, $y(t)$ is the calendar year, and $m(t)$ is the month of the year ($m(t) = 0$ for January). X_i is the average price in that month.

Panel (1) of Fig. 1 shows $X(t)$ of the S&P 500 index since January of 1802. The linear trend is obvious, but slightly concave. There are ups and downs. A few of them are quite large. For instance, one in 1860s, one in 1930s, then in 1960–1970s, and more recently in 2000s.

2.1 Equity Risk Premium

The economists often prefer to examine economic quantities in "real" terms, that is, subtracting the effect of inflation. Panel (2) of Fig. 1 shows the more common view in the literatures: the real logarithmic total return index $X_{\text{real}}(t)$ (This reproduces

Figs. 5–4 in [20]). The most notable feature is that $X_{\text{real}}(t)$ can be linearly regressed over the 200-year history, with an impressive $R^2 = 0.994$:

$$X_{\text{real}}(t) \sim \beta_{\text{rep}} t + \alpha_{\text{rep}} + \epsilon(t). \tag{2}$$

The slope β_{rep} is about 6.6% per year (between 1802 and 2021). This is called the **real equity risk premium**. This constant is one of the most celebrated constants in modern financial systems.

However, we must note that no other major equity index exhibits such beautiful linearity over such long history. Geopolitical events, financial bubbles and crashes often caused significant distortion or even disruption to many national indices. Some people may even criticize that the linearity of $X_{\text{real}}(t)$ for SPX carries with it a strong survivorship bias. There is no certainty that it will continue to work, although it has been working quite well for two centuries.

We also note that, on the back of such impressive R^2 is the residuals $\epsilon(t)$ where

$$\epsilon(t) = X(t) - \left(\beta_{\text{rep}} t + \alpha_{\text{rep}}\right) - \log \text{CPI}(t). \tag{3}$$

The residuals $\epsilon(t)$ is illustrated in Panel (3) of Fig. 1. Its standard error is $\sigma = \text{Stdev}(\epsilon(t)) = 0.32$ between 1802 and 2021. Thus its 2σ is ± 0.63, drawn in two red dashed lines. Assume $\epsilon(t)$ is mean-reverting, this implies that $X_{\text{real}}(t)$ will swing around its linear progress $\beta_{\text{rep}} t + \alpha_{\text{rep}}$ between $\pm 2\sigma$ (in 95% confidence) from decade to decade. This large amount of variation is disguised in the semi-log plot of Panel (2).

This work is primarily the study of such "fine structure". A 0.5 downward move in the log scale translates to approximately 50% market drop in a large recession. This can cause massive blowup for funds and companies that have too much leverage. When the lack of growth is stretched over a decade, it puts a lot of pressure on pensions, endowments, and retirement accounts that have significant cash outflow.

Note that the $\pm 2\sigma$ swings in $\epsilon(t)$ typically span several decades. Each cycle is composed of several recessions, which typically occurred every 4–10 years. Recession forecast is a "shorter-term" activity than what is studied here.

We created a more adaptive algorithm than a global linear regression in (2). It is used to build a forecast framework for $X(t)$ a few years into the future.

2.2 *Discussion—A Naive 10-Year Forecast*

In Panel (3), we observe several empirical rules from which we can make a naive 10-year forecast. First, $\epsilon(t)$ oscillates between -2σ and $+2\sigma$. At the dot-com peak of 2000, it touched $+2\sigma$. And at the bottom of 2009 financial crisis, it touched -2σ. Amid the pandemic of 2020, $\epsilon(t)$ was approximately at zero.

Assume $\epsilon(t)$ will reach $+2\sigma$ in 2030, the annual rate of change of $\epsilon(t)$ is $\sigma/5$ in 10 years. The annual real return of $X(t)$ will be $\sigma/5 + \beta_{\text{rep}}$. If the annual inflation

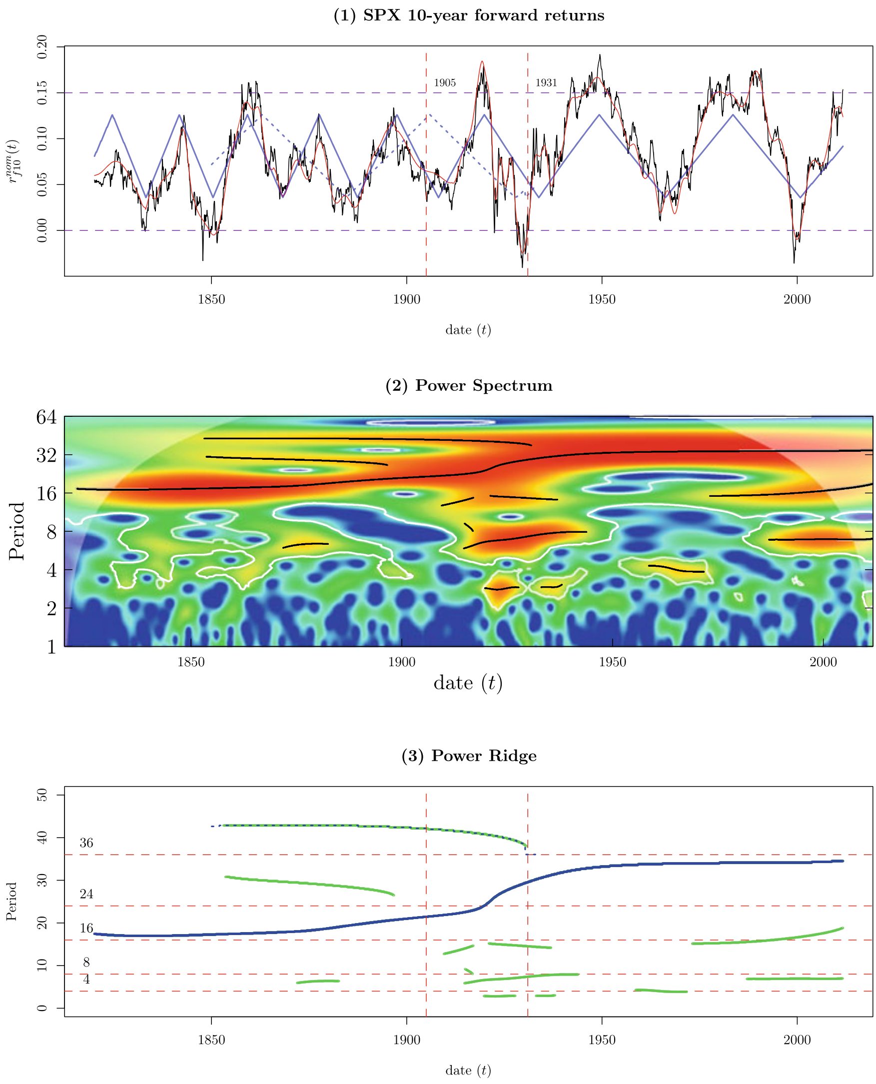

Fig. 2 Wavelet analysis on 10-year forward returns of S&P 500 index, $r_{f10}^{\text{nom}}(t)$. The dominant period is 36 years after WWII. The transition period 1905–1931 is marked by the red vertical dash lines, before which the period is shorter, between 16 and 24 years. The 8-year period was very strong for some intervals, e.g. during the Great Depression years, and between 1980 and 2020

is about 3% for the next 10 years, we arrive at the 10-year forward nominal return $r_{f10}^{\text{nom}}(t)$ of 16%.

This is a pretty naive estimate. Nevertheless, we will show in Panel (1) of Figs. 2 and 3 that 16% is a reasonable average estimate for a long-term bull market. One

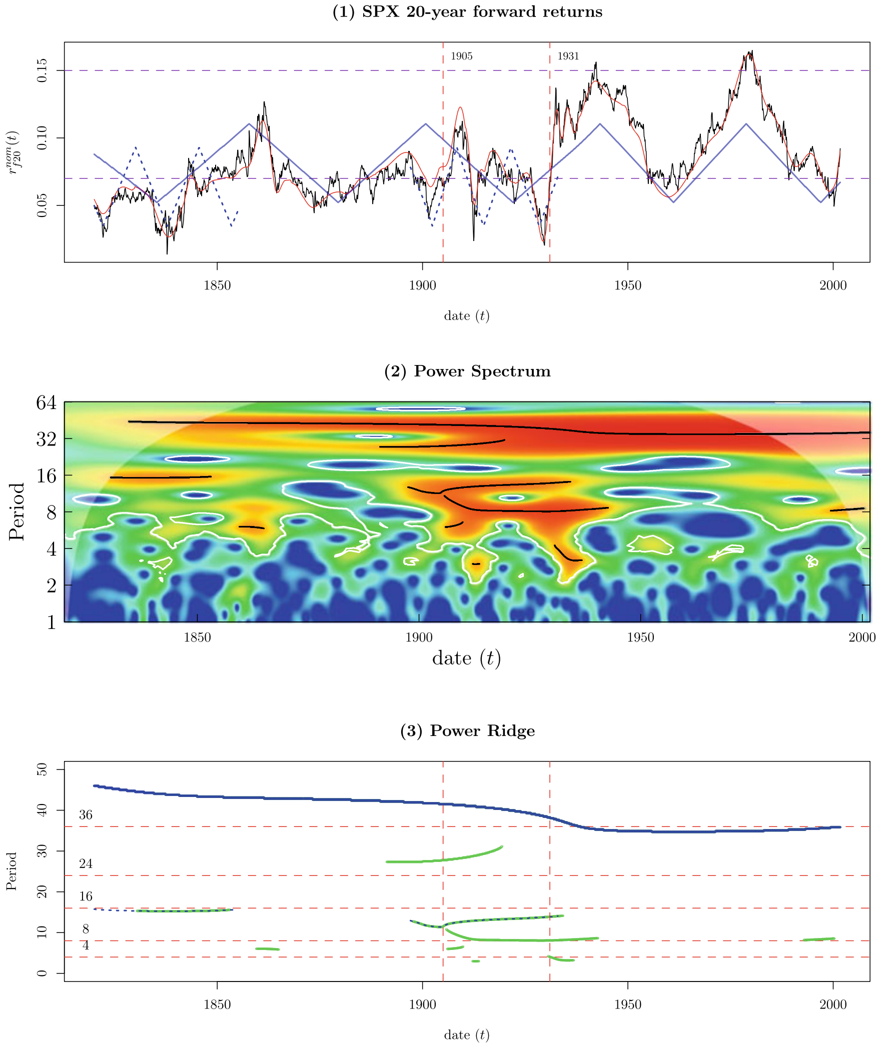

Fig. 3 Wavelet analysis on 20-year forward returns of S&P 500 index, $r^{\mathrm{nom}}_{f20}(t)$. The dominant period is 36 years after the transition year 1931, before which there was no clear dominant period

must remembe that $X(t)$ has the average annual volatility of about 12–13% between 1950 and 2021. In a good year, the return can reach 30%, but in a bad year, the volatility can be as high as 60%.

3 Wavelet Analysis on the 36-Year Long Term Cycle

In this section, we use the Morlet wavelet [13, 14] to show the 36-year long-term cycle in the U.S. stock market after WWII. This pattern can be observed in both the 10-year and 20-year returns with very little model assumptions. Recognition of such long-term cycle can greatly demystify the behavior of the stock market, e.g. the bull markets in the 1950s, and 1980–90, and the bear markets during 1970s and 2000s.

The **WaveletComp** package in R is used to perform the wavelet analysis. The advantage of this package is its simple user interfaces and beautiful graphical outputs. We briefly explain the main features of the wavelet theory, according to [18].

3.1 Introduction to the Wavelet Transform

The "mother" Morlet wavelet is defined as

$$\psi\left(t\right) \equiv \pi^{-1/4} e^{iwt} e^{-t^2/2}, \tag{4}$$

where the "angular frequency" w is set to 6. This is the preferred value in the literatures since it is approximately 2π. This wavelet can be thought of as the composite of a Fourier component e^{iwt} and a Gaussian component $e^{-t^2/2}$. The Fourier component captures the phase of a wave.

The wavelet transform of a time series x_t is defined as its convolution with a set of "wavelet daughters" $\psi\left(\frac{t-\tau}{s}\right)$. The daughters are generated from the mother wavelet by translation in time by τ and scaling by s. Each convoluted wave is

$$\text{Wave}\left(\tau, s\right) \equiv \sum_t x_t \frac{1}{\sqrt{s}} \psi^* \left(\frac{t-\tau}{s}\right), \tag{5}$$

where $*$ denotes the complex conjugate. Since x_t in our case is monthly data, τ is shifted in the unit of dt =1/12 (year).

For scaling, the choice of the set of s determines the coverage in the frequency domain, called "periods" $\{s_j\}$. It is a fractional power of 2, a "voice" in an "octave" with $1/dj$ determining the number of voices per octave:

$$s_j = s_{\min} 2^{j \cdot dj},\ j = 0 \ldots J, \tag{6}$$

where $s_{\min}$ is set to 1 (year), and dj is set to 1/128. The maximum of s_j is set to 64 (year), which determines $J = 768$. These settings allow us to analyze periods from 1 year to 64 years, that covers our target period of interest: 36 years.

The power spectrum is defined as [4]

$$\text{Power}(\tau, s) \equiv \frac{1}{s} |\text{Wave}(\tau, s)|^2 . \tag{7}$$

The power ridges are the s locations of local maximums in Power (τ, s) at a given τ [3]. The **WaveletComp** package has a built-in utility to identify statistically significant power ridges in the entire spectrum. For our purpose, the most interesting power ridge is the ridge of global maximum: $\left\{s_{\max}(\tau) = \text{argmax}_s \text{ Power}(\tau, s)\right\}$.

The instantaneous or local wavelet phase characterizes the periodic phenomena:

$$\text{Phase}(\tau, s) \equiv \text{Arg}(\text{Wave}(\tau, s)) , \tag{8}$$

We can follow the phase of global maximum power ridge $s_{\max}(\tau)$ over τ(assume it meets certain continuity condition) to understand the long-term periodicity of the market:

$$\text{Phase}_{\max}(\tau) \equiv \text{Arg}(\text{Wave}(\tau, s_{\max}(\tau))) , \tag{9}$$

By transforming the phase via the triangle wave function $f(\theta) = 1 - \frac{2}{\pi}$ arccos $(\cos(\theta))$, where $\theta = \text{Phase}_{\max}(\tau)$, the periodicity of interest can be clearly illustrated.

The time series can be smoothed and reconstructed by summing over a set of waves:

$$(x_t) = \frac{dj \cdot \sqrt{dt}}{0.776 \cdot \psi(0)} \sum_s \frac{1}{\sqrt{s}} \text{Re}(\text{Wave}(\tau, s)) . \tag{10}$$

The reconstruction factor 0.776 is adopted from [25] as an empirically suggested constant for the full reconstruction.

Financial time series is known to have high noise-to-signal ratio. Proper shrinkage during reconstruction (smoothing and/or denoising) can enhance the signal of interest. The wavelet shrinkage is performed by either filtering out s smaller than a certain threshold, or dropping weaker waves according to the strength of the power spectrum.

3.2 *Wavelet Regression of the 10-Year Returns*

The 10-year forward returns $r^{\text{nom}}_{f10}(t)$ is analyzed in this section. We emphasize that the input data is model-free. The only parametrization is the choice of the return window: 10 years. The wavelet analysis is shown in Fig. 2. From the "Power Ridge" chart in Panel (3), we observe that the dominant period was 36 years after WWII.

In both Figs. 2 and 3, the charting conventions are as follows:

Panel (1) shows the time series x_t ($r^{\text{nom}}_{f10}(t)$ and $r^{\text{nom}}_{f20}(t)$) in the black line, and the reconstructed (x_t) in the red line. The triangle phase $f(\theta)$ of the strongest power ridge is drawn in the solid blue line, and the secondary in the dashed blue line. Two vertical red dashed lines are drawn at 1905 and 1931—two fault line locations from the 20-year forecast model in Sect. 6.2.

Both the 10-year and 20-year returns could not exceed 15–17% for too long. This level marks the rampant bull market. On the other hand, the 10-year returns rarely went below 0%. The 20-year returns also appear to have a floor at 5–7%.

Panel (2) shows the power spectrum Power (τ, s). The y-axis is the period τ. The color spectrum illustrates the power level where red is high and blue is low. The power ridges are drawn in black lines.

Panel (3) show the power ridges with the guided red dashed lines at the ladders of 4, 8, 16, 24, 36 years. The strongest power ridge is drawn in the solid blue line, and the secondary in the dashed blue line. The remaining ridges in the green lines.

There was a fundamental change in the periodicity before WWI and after WWII. We conjecture this might be related to the transition of the world power from Europe to Washington. Prior to WWI, the period is about 16–24 years, much shorter than 36 years.

3.3 Wavelet Regression of the 20-Year Returns

As we see above, the 36-year period is the natural frequency of the long-term mean-reversion cycles. The regression on the 20-year returns requires the least tectonic adjustments. This gives us the strong incentive to explore the 20-year returns here, even though most financial analysis stops at the 10-year returns.

The wavelet analysis on the 20-year forward returns, $r_{f20}^{\mathrm{nom}}(t)$, is shown in Figure 3. We can clearly observe the 36-year period after the transition year 1931 from the "Power Ridge" chart in Panel (3).

In Panel (1), before 1931, the 20-year returns were pretty flat, around 7%. Most of the smaller fluctuations were smoothed out. In Panel (3), during the Great Depression years, the 8-year period was very strong. But before 1905 and after 1931, there was almost no power distributed in any of the secondary periods. This is consistent with our observation that $r_{f20}^{\mathrm{nom}}(t)$ removed most of the short-term fluctuations and preserved the most important long-term signals.

The 36-year period began to emerge after the 1929 crash. It went through two cycles after WWII. As of this writing, the market is at the bottom of this cycle, and is about to revert from a bear market to a bull market.

4 Channel Deviation Framework

In this section, we lay out the channel deviation framework, in which $X(t)$ is decomposed into the smooth channel moving average $\alpha(T)$, the channel return $R(T)$, and the mean-reverting channel deviation $Y(T)$. We show how the optimal look-back duration $\Delta T_b = 45$ is chosen for the S&P 500 index.

4.1 Mean-Reversion Decomposition

For a given time series that is predominantly in a linear trend, such as the total return index $X(t)$ in (1), we assume it is composed of a linear process and a mean-reverting process. The goal of this framework is to decompose $X(t)$ into these two processes while maintaining causality.

Let ΔT_b be the duration of the look-back channel. At time T, we apply linear regression

$$X(t) \sim \alpha(T) + R(T)(t - T), \text{ where } t \in [T - \Delta T_b, T], \tag{11}$$

to obtain $\alpha(T)$, which is called **channel moving average** (CMA), and $R(T)$, which is called **channel return**. Then we derive the **channel deviation** at time T as $Y(T) = X(T) - \alpha(T)$. One can view $Y(T)$ and $\alpha(T)$ as the decomposition of $X(T)$, where $\alpha(T)$ is linear and non-stochastic, and $Y(T)$ is mean-reverting. $R(T)$ is the instantaneous rate of change of $\alpha(T)$.

$Y(T)$ is of paramount importance in this framework. We will show that log-CAPE mean-reverts in similar pattern and scale to $Y(T)$ in Sect. 5. Since $\alpha(T)$, $R(T)$, and thus $Y(T)$ are causal, they can be used for forecasting after time T, as shown in Sect. 6.

4.2 Closed Form Solution

There are closed form solutions for $\alpha(T)$, $R(T)$, and $Y(T)$ in the discrete notation. (11) is the ordinary least squares (OLS) optimization. Let $\langle t_i - T\rangle$ be the mean of $t_i - T$ for $t_i \in [T - \Delta T_b, T]$, and N is the sample size of t_i, we have $\langle t_i - T\rangle = \frac{N+1}{2}\Delta t \underset{N\gg 1}{\approx} -\frac{1}{2}\Delta T_b$, and $\text{var}(t_i) = \frac{1}{12}\left(N^2 + N\right)\Delta t^2 \underset{N\gg 1}{\approx} \frac{1}{12}\Delta T_b^2$. Then

$$\begin{aligned}
R(T) &= \frac{\text{cov}(X_i, t_i)}{\text{var}(t_i)} = \text{cor}(X_i, t_i)\frac{\text{stdev}(X_i)}{\text{stdev}(t_i)} \underset{N\to\infty}{\approx} \frac{\sqrt{12}}{\Delta T_b}\text{cor}(X_i, t_i)\,\text{stdev}(X_i),\\
\alpha(T) &= \langle X_i\rangle - R(T)\langle t_i - T\rangle = \langle X_i\rangle + \sqrt{3\left(\frac{N+1}{N}\right)}\text{cor}(X_i, t_i)\,\text{stdev}(X_i)\\
&\underset{N\gg 1}{\approx} \langle X_i\rangle + \frac{1}{2}R(T)\,\Delta T_b.
\end{aligned} \tag{12}$$

The main feature in $R(T)$ and $\alpha(T)$ is the covariance between X_i and t_i in the channel. Given the same Stdev (X_i), $R(T)$ is maximized by the best Cor (X_i, t_i), which is 1 when X_i is perfectly linear to t_i.

$\alpha(T)$ is the result of the optimal linear predictor. The first term in $\alpha(T)$ is the moving average $\langle X_i\rangle$. The second term introduces the "correction" for the trend, which is non-zero as long as Cor $(X_i, t_i) \neq 0$. The sign of the "correction" is given by the sign of Cor (X_i, t_i).

Equation (12) leads to the closed form of the channel deviation,

$$Y(T) = X(T) - \langle X_i \rangle - \sqrt{3\left(\frac{N+1}{N}\right)} \operatorname{cor}(X_i, t_i) \operatorname{stdev}(X_i) \tag{13}$$

This equation is out-of-sample, thus is causal. Also note that this framework is scale independent. The outputs don't vary much with regard to different data sampling frequency.

4.3 Optimal Choice of Look-back Channel at 45 Years

The look-back channel ΔT_b is the only hyperparameter in this framework. It should be chosen such that the outputs are least biased. The wavelet analysis shows that the channel must be longer than 36 years. Based on our empirical experimentation, we know it is between 30 and 50 years. We provide one version of optimization that we use to determine $\Delta T_b = 45$.

For a given $\Delta T_b < 60$, we calculate $Y(T)$ for all T's between 01/1862 and 12/2017. We then calculate the skewness and kurtosis of $Y(T)$ for such ΔT_b. We seek the optimal ΔT_b that produces the lowest kurtosis and zero skewness with a tolerance of randomness. The kurtosis and skewness are shown in Fig. 4.

This turns out to be a relatively simple optimization problem to solve. When ΔT_b is small, the kurtosis is very high and the skewness is negative. As ΔT_b increases, the

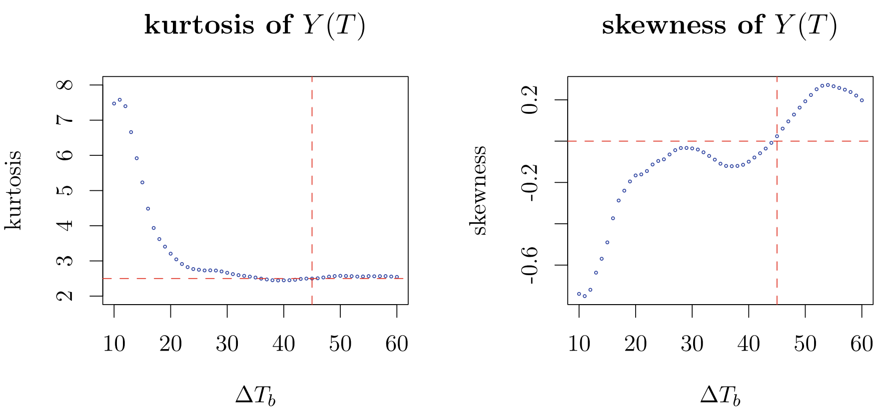

Fig. 4 Optimization of the look-back channel ΔT_b. The left panel shows the kurtosis of $Y(T)$ forms a plateau around 2.5 when $\Delta T_b > 35$. The right panel shows the skewness of $Y(T)$ crosses zero at $\Delta T_b = 45$, which we choose to be the optimal look-back period

kurtosis decreases towards 3 and the skewness increases towards zero. When $\Delta T_b > 21$, the kurtosis decreases below 3, that is, the system transitions from lepkurtotic to platykurtic. When $\Delta T_b > 35$, the kurtosis forms a plateau around 2.5. The kurtosis reaches its minimum of 2.445 at $\Delta T_b = 39$, but the skewness doesn't cross zero until $\Delta T_b = 45$ at which point the kurtosis is at 2.498, slightly higher than the absolute minimum. We determine that $\Delta T_b = 45$ is the optimal choice.

4.4 Discussion on the Outputs

Figure 5 shows the result of $\alpha(T)$, $Y(T)$, and $R(T)$ at $\Delta T_b = 45$. We first note that $Y(T)$ oscillates between ± 0.5 with a periodicity of approximately 40 years. The periodicity is particular clear by observing the legs of $Y(T)$. The market swings violently during two periods: From 1929 to 1933, the oscillation almost reaches ± 1.0. From 2000 to 2009, the oscillation is as large as ± 0.75. We will elaborate more on the periodicity and amplitude of $Y(T)$ in Sect. 5.1.

Secondly, we observe that $R(T)$ has three plateaus in history. The first plateau is at 5% before 1860. The second plateau is at 7.13% from 1880 to 1950. The third plateau is at 10.52% from 1970 to now. The values of plateau are determined by zeroth order `genlasso::trendfilter` utility in R.[6] At 600 degrees of freedom, we round the output of beta to 3 digits, and select the largest clusters of beta that have repeated more than 50 months. The average of beta from each cluster is the mean of the plateau. The 10.52% return of the third plateau is often quoted in the literatures and media as the long-term expected return of SPX. Here we provide a proper context in terms of $R(T)$.

5 Relation Between Channel Deviation and CAPE

In this section, we show that $Y(t)$, $R(t)$, and CPI have large explanatory power on CAPE, even though their data generating processes (See Chap. 1 of [7]) don't seem to be related at all. The log-CAPE can be decomposed by a four-factor model with a high R^2.

5.1 Regression of Log-CAPE

Let $\mathrm{CAPE}_{\Delta T}(t)$ denote the ΔT-year CAPE where $\Delta T = 10, 20$. In Panel (1) of Fig. 6, it is shown that $\log(\mathrm{CAPE}_{10}(t))$ and $\log(\mathrm{CAPE}_{20}(t))$ are very similar (the

[6] See also https://cran.r-project.org/web/packages/genlasso/vignettes/article.pdf.

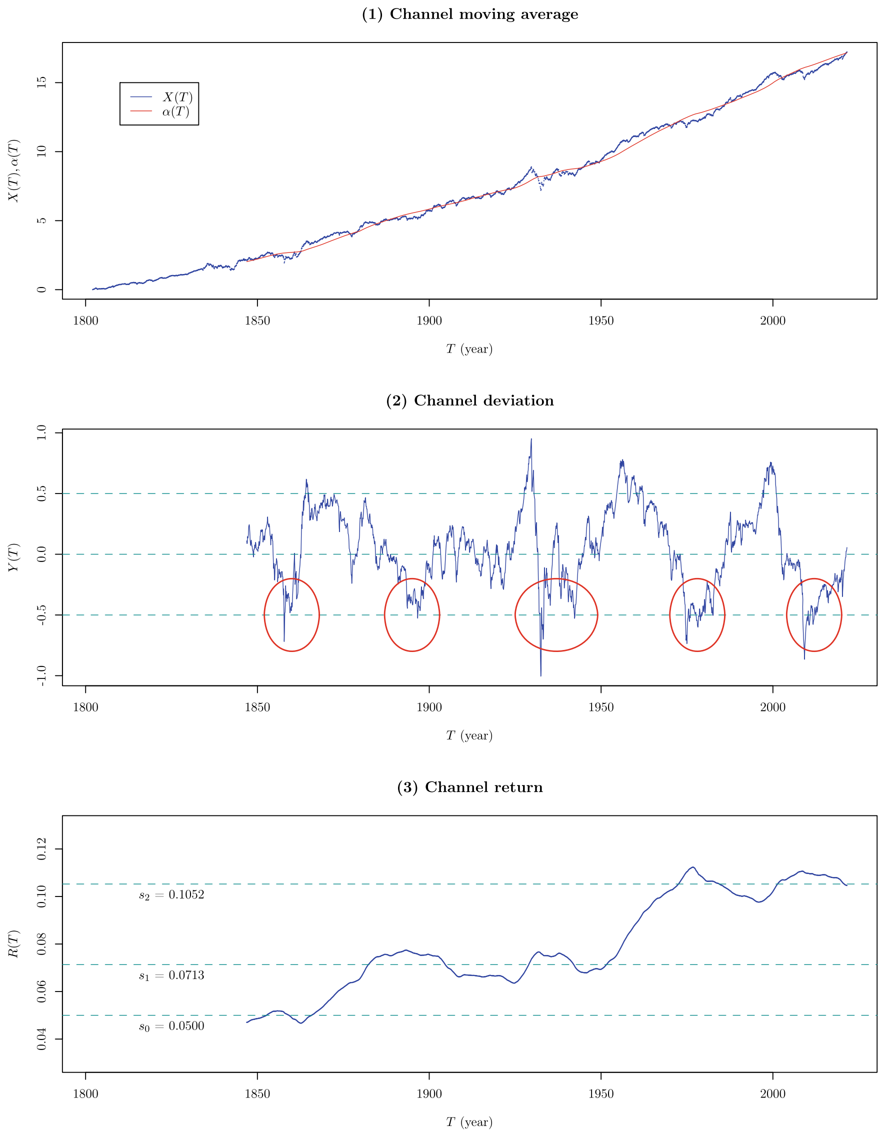

Fig. 5 Optimally decomposed $\alpha(T)$, $Y(T)$, and $R(T)$ at $\Delta T_b = 45$. The legs of $Y(T)$ are drawn in red circles in Panel (2) to illustrate the periodicity. The levels of plateau in $R(T)$, $s_0 = 0.05$, $s_1 = 0.0713$, $s_2 = 0.1052$, are calculated from zeroth order `genlasso::trendfilter` utility. The 10.52% of s_2 is often quoted as the long-term expected return of SPX

blue and cyan lines). Also note that $Y(t)$ is in the same scale of $\log(\mathrm{CAPE}_{\Delta T}(t))$. Hence, we focus on the 20-year model.

And let $\mathrm{CPI}_{10}(t)$ and $\mathrm{CPI}_{20}(t)$ denote the 10 and 20-year log-returns of CPI. That is,

$$\mathrm{CPI}_{\Delta T}(t) = \frac{\log \mathrm{CPI}(t) - \log \mathrm{CPI}(t - \Delta T)}{\Delta T}. \tag{14}$$

In Panel (2), $\mathrm{CPI}_{10}(t)$, $\mathrm{CPI}_{20}(t)$ and $R(t)$ are shown. $\mathrm{CPI}_{10}(t)$ is more volatile than $\mathrm{CPI}_{20}(t)$. $R(t)$ is the long-term moving average of nominal equity returns. It is shifted down by the equity risk premium β_{rep} (6.6%), and we observe it is approximately the long-term (40 years) inflation rate. These three factors constitute the inflation inputs for the regression model.

We perform the following linear regression for t between 1/1881 and 12/2020:

$$\log(\mathrm{CAPE}_{20}(t)) \sim \beta_0 + \beta_1 Y(t) + \beta_2 R(t) + \beta_3 \mathrm{CPI}_{10}(t) + \beta_4 \mathrm{CPI}_{20}(t) + \varepsilon, \tag{15}$$

which results in a high R^2 of 0.82. The summary of linear model from R is shown below:

```
a ← lm(log.cape20 ~ eqty.lm.y + eqty.lm.r + cpi.logr.10 +
cpi.logr.20, data=df) summary(a)
```

```

Call:
lm(formula = log.cape20 ~ eqty.lm.y+eqty.lm.r+cpi.logr.10+
    cpi.logr.20, data = df)

Residuals:
     Min       1Q   Median       3Q      Max
-0.34672 -0.16100 -0.02993  0.18624  0.45322

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.00789    0.02938   34.30   <2e-16 ***
eqty.lm.y     0.93990    0.01713   54.87   <2e-16 ***
eqty.lm.r    25.16626    0.36828   68.33   <2e-16 ***
cpi.logr.10  -3.84196    0.28691  -13.39   <2e-16 ***
cpi.logr.20 -11.73114    0.42353  -27.70   <2e-16 ***
---
Signif. codes: 0 '***'0.001 '**'0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1923 on 1555 degrees of freedom
  (1067 observations deleted due to missingness)
Multiple R^2:   0.82,    Adjusted R^2:   0.8196
F-statistic:  1771 on 4 and 1555 DF,  p-value: < 2.2e-16
```

All four factors are highly significant. More than three quarters of information in log-CAPE is contained in the linear combination of our mean reversion analytics and past inflations.

The result is shown in Panel (3) of Fig. 6.

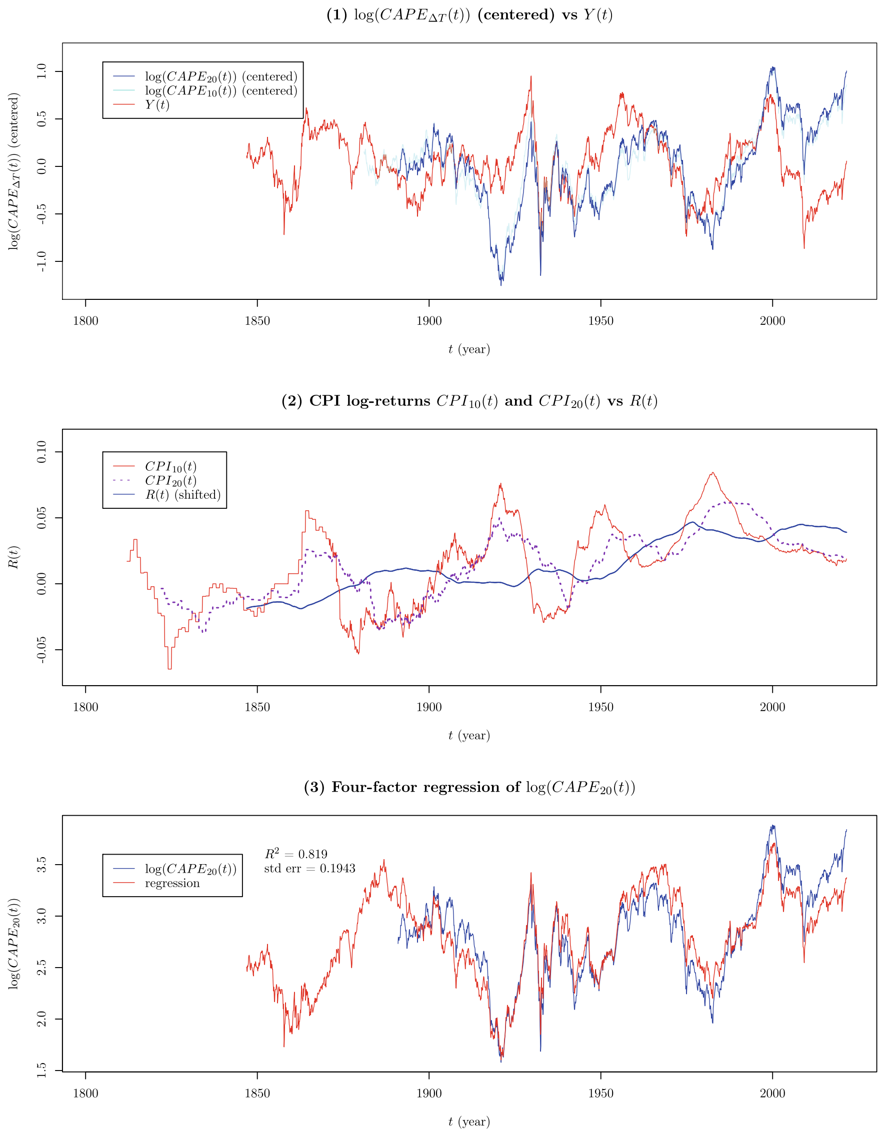

Fig. 6 Linear regression of the 20-year log-CAPE by the four factors: $Y(t)$, $R(t)$, $\text{CPI}_{10}(t)$ and $\text{CPI}_{20}(t)$. Panel (1): Comparison of centered log-CAPE and $Y(t)$, showing their similarity and in the same scale. Panel (2): $R(t)$, $\text{CPI}_{10}(t)$ and $\text{CPI}_{20}(t)$ as the inflation inputs to supplement the differences between log-CAPE and $Y(t)$. Here $R(t)$ is shifted down by the real equity premium. Panel (3): The result of regression on $\log(\text{CAPE}_{20}(t))$ in the four-factor model

5.2 Discussion

We illustrated the inner workings of the four-factor regression by Panel (1) and Panel (2) of Fig. 6. Panel (1) shows that log $(\mathrm{CAPE}_{20}(t))$ is almost in the same scale as $Y(t)$, and this is confirmed by the coefficient $\beta_1 = 0.94$ in Eq. (15).

There are times that log-CAPE moves below $Y(t)$ (e.g. in 1920s, 1950s, and early 1980s) and other times that log-CAPE moves above $Y(t)$ (e.g. in 1900s and 2000s). Their differences are made up by $\mathrm{CPI}_{10}(t)$ and $\mathrm{CPI}_{20}(t)$. This is confirmed by the negative correlation ($\beta_3 = -3.8$ and $\beta_4 = -11$) in the summary statistics above. This is shown graphically in Panel (2). We observe that, whenever $\mathrm{CPI}_{10}(t)$ and $\mathrm{CPI}_{20}(t)$ are above $R(t)$, $Y(t)$ tends to be above log $(\mathrm{CAPE}_{20}(t))$, and vis versa.

This anti-correlation between log-CAPE and inflation is one of the two main reasons why CAPE is perceived at a lofty level since 2000. The high CAPE reading is a reflection of ultra-low inflation in the past two decades.

The second reason is that log-CAPE is positively correlated to $R(t)$ with $\beta_2 \approx 25$. Since $R(t)$ is currently at the third plateau, it also contributes to the high level of CAPE. From 1950 to 1970, the market was transitioning from the second plateau to the third, the difference in $R(t)$ is $s_3 - s_2 \approx 3.4\%$. Multiplying it by $\beta_2 \approx 25$, its impact on log-CAPE is 0.85, which is translated to 130% higher CAPE. In 1970s and 1980s, this effect was muted because of the high inflation. Going into 1990s, the high tide of inflation receded and CAPE began to move much higher.

However, in order to justify such high level of equity returns and valuation, it seems to imply that the future inflation will have to be much higher.

5.3 Tectonic CAPE

We introduce the concept of tectonic CAPE, in which we hypothesize wars and national policy changes in the past might have resulted in significant dislocations in the data generating processes of CAPE and CPI (Chap. 19 of [7]). We use nonlinear optimization technique to uncover these dislocations. However, we do this only sparingly so that we don't overfit the data.

At a specific time t_i^{adj}, the amount $\Delta_i \log \mathrm{CAPE}_{20}$ should be added to $\log(\mathrm{CAPE}_{20}(t))$. These adjustments are called the "fault lines", and the adjusted CAPE is called "tectonic CAPE".

Formally, the tectonically adjusted log-CAPE is

$$\log\left(\mathrm{CAPE}_{\Delta T}^{\mathrm{adj}}(t)\right) = \log(\mathrm{CAPE}_{\Delta T}(t)) + \sum_{i=1\cdots N} \begin{cases} 0, & t < t_i^{\mathrm{adj}}; \\ \Delta_i \log \mathrm{CAPE}_{\Delta T}, & t \geq t_i^{\mathrm{adj}}. \end{cases} \tag{16}$$

Lihn [10] showed that the 20-year model requires smaller amount of "fault line adjustments" than the 10-year model. The interpretation is that many economic shocks tend to average out much better in 20 years than 10 years.