ECONOMICS WORKING PAPERS

r-g before and after the Great Wars 1507–2023[*]

Kenneth S. Rogoff[†] Paul Schmelzing[‡]

Economics Working Paper 24113

HOOVER INSTITUTION
434 GALVEZ MALL
STANFORD UNIVERSITY
STANFORD, CA 94305-6010
November 14, 2024

We present new long-run samples of r-g series over centuries for key economies in the international financial system. Across a wide variety of econometric approaches, and including duration-matched constructions, we demonstrate strong evidence of trend stationarity in these series. Although we confirm trend stationarity, we find robust evidence of a major structural break in the first third of the 20th century. A multi-century downward trend in r-g appears to have levelled off in the years around 1930, and since then r-g has shown high volatility coupled with clear upwards pressure: notably, though real interest rates may still appear favorably low, aggregate growth rates are drifting downwards in advanced economies since the interwar period, creating secular pressures on r-g and debt sustainability. Our results stand in contrast to much recent literature, and suggest the need for much more caution in assuming benign trends in global public debt sustainability. At the same time, when adding riskier elements of capital returns, the data lend support for structurally increasing "dynamic efficiency". We then associate the key 1930s inflection to the establishment and growth of welfare states in advanced economies, and the surge in non-defense, non-interest expenditures.

JEL Codes: E4, F3, N20.

---

[*] The views expressed in this paper are solely those of the authors. We thank Dan Smith, Baiyun Jing, and Agustin Kupfer for outstanding research assistance, as well as seminar and conference participants at Northwestern Kellogg, UC Berkeley, and Boston College for very helpful feedback.

[†] Harvard University.

[‡] Boston College, and Hoover Institution Stanford.

1 Introduction

In this paper, we explore the econometric properties of advanced economy interest rate-output growth spreads (r-g) over the very long-run, based on new data. Such an investigation is of major relevance since it is linked to ongoing debates in recent years on the theoretical and practical relevance of environments of negative interest rate-output growth differentials – especially with regards to public debt sustainability. Such debates are also linked to more fundamental debates on "dynamic inefficiency", the idea that returns on capital may structurally fall short of output growth rates, going back to Phelps (1961) and Abel et al. (1989).

The motivation for revisiting the long-run properties in advanced economy r-g differentials is twofold: first, from an applied perspective, the low interest rate period that prevailed over recent decades has spurred an extensive literature investigating the implications for public debt sustainability: specifically, the phenomenon that sovereign real interest rates have fallen below output growth rates has led to influential suggestions that governments may find themselves in a "new regime" of debt sustainability in the 21st century (Blanchard, 2019, 2022). On the contrary, public accountability bodies such as the Congressional Budget Office (CBO) and the International Monetary Fund (IMF) have voiced growing concerns in the post-COVID era over the long-run debt trajectory of advanced economies including the United States, with debt/GDP levels in baseline scenarios expected to reach all-time records surpassing those of World War II by the end of the 2020s. The further rise of political polarization is in principle associated with additional pressures on debt/GDP ratios in the historical context (Funke et al., 2023). As such, long-run r-g trends are of critical importance to evaluate the feasibility of both sides of this debate.

Second, from an empirical side we build on key recent advances in long-run data on interest rates and output growth, which allow for the first time secular reconstructions and analyses of long-run trends and econometric properties of r-g differentials: as we discuss, recent research showed that employing long horizon time series significantly increases the power of statistical tests. In particular, these advances significantly improve the power of standard unit root and structural break tests, which are the focus of much fundamental analyses on macroeconomic and financial time series. On this basis, it is for the first time possible to address robustly the question whether recent years or decades have indeed seen any secular shifts in r-g. Equally important, it allows us to put the recent period into much longer global perspective.

The first part of this paper reconstructs annual-level r-g series for key advanced economies over centuries, including for the U.K., the U.S., Germany, and France, exploiting recent data advancements. We are able to reconstruct such data on a duration-matched basis, for both aggregate g (on which we concentrate) and per capita g bases over centuries – while adding several constructions of "global" r-g trends, mainly for illustrative purposes.

Having a multi-century data set turns out to make an important difference when we proceed to analyze the time series properties of r-g which, as far as we are aware, has not been done for r-g over any long-dated sample. There is a large literature on testing stationarity in the real interest rate itself, mostly on post-war data – albeit a few using spliced data to reach back 100 or 150 years – with none able to reject the unit root, up until using the very-long run horizons that Rogoff et al. (2024a) analyze. Here, having several centuries of r-g data turns out to make a decisive difference, yielding very clear evidence of trend stationarity at the 1% level; moreover, the finding of trend stationarity is economically significant, with half-lives tending to be well under ten years. These results are robust to using a test that allows for a structural break (Zivot and Andrews, 1992); applying the Bai and Perron (1998) test, we do find a small number of mean breaks that align well with historical events and evidence, corresponding particularly to historical inflection points of fiscal policy.

Of course, the Bai-Perron test looks only for mean breaks, but since we are particularly interested in in exploring whether there has been any statistically significant break in the trend – with visual evidence strongly suggestive of such a potential trend break in the long-run reconstructed series in or around the first third of the 20th century (Figures 1 to 3) – we go on to apply the joint trend and intercept break test of Perron and Yabu (2009), an exercise that our long sample greatly facilitates, since it provides far enhanced power.

Overall, our results suggest a most intriguing pattern across all series: namely in the years around the 1930s – across all key economies – r-g differentials appear to change their secular sign. In other words, although we find that a consistent downward trend in r-g that unfolded over more than 300 years ends in the first third of the 20th century, one can make the case that if anything, the trend since has been, on average, upwards (albeit we interpret the statistical evidence cautiously given a high degree of volatility). If this seems surprising to the reader, we note that the relevant "g" in this debt sustainability framework is the aggregate real growth rate of the economy, which has faded over time in advanced economies due to slower population growth – even as the fall in per capita growth is more recent.

Indeed, as we show, much indicates that lower aggregate output growth rates are a key component of the key 20th century break, with a clear post-1930s shift towards lower aggregate output growth rates in advanced economies: the middle of the 20th century exhibited the first clear deceleration of the Industrial Revolution growth path – the only exception in our sample being Germany, which experienced its well-known post-war reconstruction "growth miracle", and therefore saw a delayed inflection towards lower growth rates. The relevant takeaway from the long-run context, crucially, is that existing notions of secularly improving debt sustainability in advanced economies should at the very least be treated much more cautiously, and take into account the possibility of adverse dynamics since the interwar era.

In this sense, this paper challenges the prevailing consensus which takes it as a given that – while volatile – r-g differentials in advanced economies will likely, on average, continue to be historically low, and regularly be negative. Our results suggest that such propositions, which are typically based on post-1980s data, are indeed broadly consistent with data before World War I and the Great Depression, but not necessarily since, and certainly not decisively so. The notion that much suggests that an upwards-sloping r-g path could have emerged since the inflection over the first third of the 20th century is a finding that challenges the idea of structurally improving debt sustainability over recent decades, and suggests that significant reversal pressures on currently low r-g levels remain present. In an intuitive extension, we go on to show that our key findings appear to become, if anything, stronger when incorporating "risky" return components into broad capital return measures as based on long-run private sector data. Such an exercise is doubly important given that, as Acalin and Ball (2023) and Reinhart and Sbrancia (2015) have shown, financial repression has played a significant role in holding down “safe” interest rates over the post World War II period, implying that a low r-g may sometimes mask a significant financial repression tax rather than exceptionally fortuitous debt dynamics.

Part 2 discusses relevant literature on r-g. Part 3 discusses the construction of new very long-run r-g time series, making use of major advances in economic history research over the past decade. Part 4 undertakes econometric exercises and discusses results. Afterwards, in Part 5, we interpret the results in the historical context and place them in context to debates on debt sustainability as well as secular trends in the international financial system: this section emphasizes the analysis of the inflection over the first third of the 20th century, but we also show that we can plausibly link the early modern r-g data to historical public finance dynamics, such as the 1560-1729 r-g era which saw increasingly volatile sovereign default dynamics and a slide into financial repression. Part 5 also includes a closer discussion of the growth component (g) and its role in debt sustainability trends and breaks – here we emphasize evidence of an earlier output growth inflection relative to consensus narratives. This section closes with an extension that incorporates recent data advances on "risky" return elements of the capital stock over time, to more closely approximate the marginal product of capital and its relative evolution over safe return components of the capital stock: here we can address some plausible conclusions to "dynamic efficiency" debates. Part 6 then concludes.

2 Literature

Much recent discussion on interest rate-output differentials is building on fundamental contributions by Blanchard (2019) and Blanchard (2022), who emphasized the feasibility of indefinite public debt rollovers in scenarios where safe interest rates are below growth rates, while the marginal product of capital (MPK) is at least not far above growth rates.[1]

r-g trends have also been recently analyzed in Mauro and Zhou (2021). The authors demonstrate that negative r-g differentials for 55 countries over the past 200 years are not unusual – and that negative r-g differentials have repeatedly preceded sovereign default events. Meanwhile, Lian et al. (2020) assess the relationship between initial debt levels and r-g dynamics, and find that countries with high public debt levels display shorter negative r-g episodes and are generally more prone to experience adverse r-g reversals. In a model setting with overlapping generations, Cao et al. (2024) demonstrate that permanent debt increases can negatively affect growth variables even under r-g negative conditions, mainly due to crowding out effects.

[1] Piketty (2014) centers around "r-g" trends, too, but focuses on capital stock returns and the links to inequality discussions – we do not assess these dimensions in this paper, but Schmelzing (2025) discusses some implications.

More generally, public debt sustainability literature has analyzed 20th century dynamics, and posited increasingly favorable sustainability dynamics, with the post-1945 period exhibiting particularly favorable r-g trends: commenting on two centuries of U.S. debt evolution over 1800-1999, Elmendorf and Mankiw (1999) observed that "between these sharp increases [of peacetime U.S. debt/GDP during the Great Depression and the 1980s], the debt-output ratio has generally declined fairly steadily. An important factor behind the dramatic drop between 1945 and 1975 is that the growth rate of GNP exceeded the interest rate on government debt for most of that period." Bohn (1995)'s influential stochastic debt sustainability framework also departed from the observation that "historically, interest rates on "safe" U.S. government bonds have been significantly below the average rate of economic growth". In subsequent extensions, Bohn (1998) stressed the study of long-run time series properties on U.S. debt sustainability data, and found mean-reversion in U.S. primary surpluses depending on initial debt/GDP levels. Later literature studying the post-1945 era generally confirmed this more constructive bent, with e.g. Hall and Sargent (2011) positing a particularly constructive role from the aggregate growth side for debt sustainability dynamics in the U.S. over recent decades – a recent example of a more critical interpretation of U.S. post-war performance, however, is Acalin and Ball (2023), who emphasized the role of financial repression.

Abstracting from sole debt sustainability questions, another strand of the literature places recent trends r-g differentials into the context of dynamic inefficiency theories. Reis (2021) argues that the marginal return on capital (m) remains above the output growth rate as of the early 21st century, so that r<g<m. He argues that in fact the relevant budget constraint depends on the differentials between m-r and m-g. Similarly, Barro (2023) argues that r-g has not been historically positive when r denotes risk-free assets (here real advanced economy Treasury bill returns) – "risky" real equity returns have generally outpaced real output growth rates, which remains consistent with dynamic efficiency conditions.

A unifying theme of virtually all recent literature is the proposition that negative r-g episodes are not an anomaly, that they have not been upended by the global pandemic shock and that, in fact, they may become more frequent over time. Even contributions that question the appropriate methodological basis for debates on dynamic efficiency do not challenge the idea that "safe" sovereign interest rates may remain below output growth rates for extended periods of time as of the early 21st century (Reis, 2021; Barro, 2023).

And a crucial methodological feature of all strands of literature is the fact that it near-exclusively relies only on 20th century data, with even the few exceptions falling measurably short of obtaining sufficient statistical power, and concentrates heavily on the United States (the country which, by construction, exhibits the shortest growth and rates time series), could potentially involve crucial mis-characterizations of the nature of real rate-output differentials.[2]

[2]One recent contribution that demonstrated the key role of sample lengths used new long sample data for advanced economy real interest rates over centuries, via Rogoff et al. (2024a): the authors investigated annual ex ante real interest rates over the period 1318-2022 for eight economies, and found that an extension of the sample length beyond the samples used in traditional literature fundamentally changes the identification of econometric properties associated

3 Long-run empirics – constructing duration-matched r-g series over centuries

3.1 "Safe" r-g reconstructions and trends

Figures 1 and 2 display two key series on r-g over centuries, constructing duration-matched nominal bases series as proposed by Lian et al. (2020). The latter construct r-g series using nominal 10-year maturity government interest rates and nominal aggregate GDP growth rates, both on an annual basis, averaging the nominal growth rate observations over ten years: in other words, g at t reports the average nominal aggregate growth rate over t *to* t+9. For r, the authors use the marginal interest rate rather than the effective interest rate, which is an appropriate approach to reflect changes in market conditions – all our subsequent r-g also report nominal marginal long-maturity rates, consistent not least with Blanchard (2019).[3] Slight variations of r-g measurements exist: while Blanchard (2019) uses marginal nominal interest rates, he adjusts for taxes on government bonds and the maturity structure of debt. Over the long-run, a taxation adjustment is not feasible for our nominal interest rate variable – we focus on nominal yields on a pre-tax basis throughout. As for maturity adjustments, while they are possible for select countries over extended periods of time (e.g. Italy), the overwhelming historical reliance on long-maturity issuance makes our focus on long-maturity marginal rates the consistent one to analyze secular patterns.

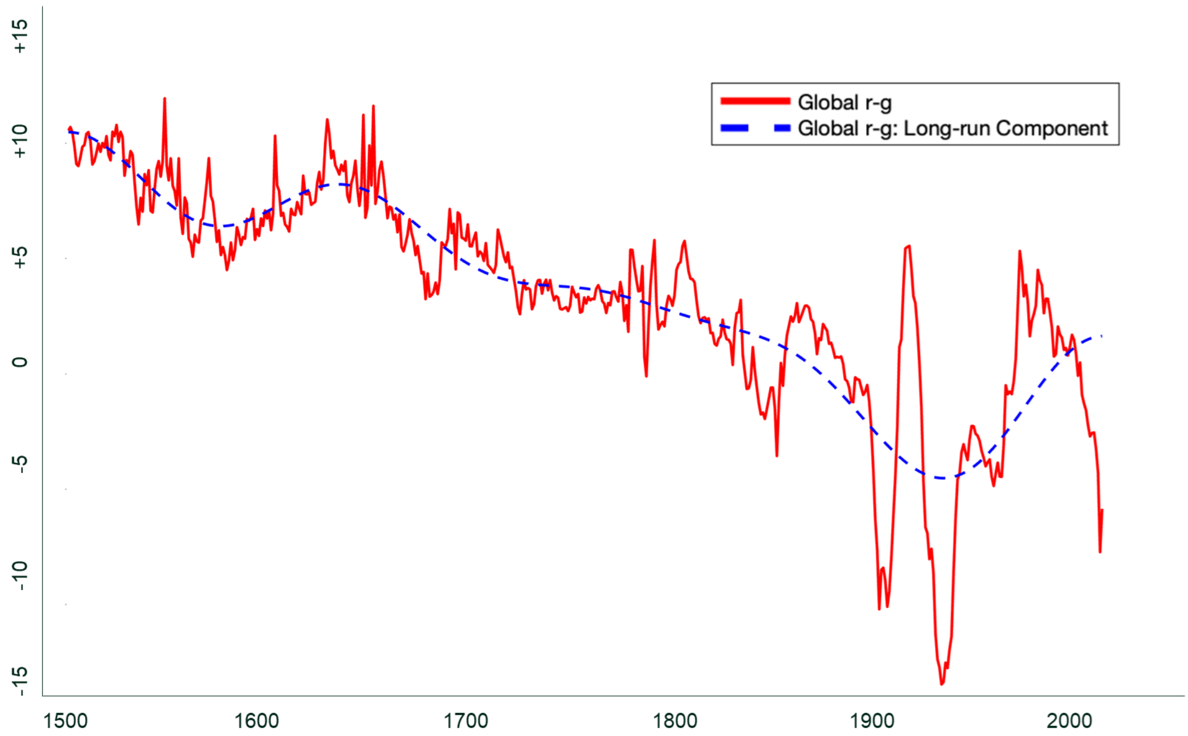
We introduce the "Global AW" bases, as per Figure 1, mainly for illustrative purposes, in the context of existing literature that has analyzed similar constructions of such "global" trends (King and Low, 2014; Jordà et al., 2019; Obstfeld, 2023). These global series are constructed on the basis of arithmetically weighting ("AW") the four key country series (France, U.K., U.S., and Germany), but we note that all key findings are independent of the specific weighting methodology (and robust to, say, weighting the four countries on the basis of GDP shares).

Economic historians have made substantial leaps in developing historical output figures over the past decade. Growth figures for Britain are sourced from Dimsdale and Thomas (2016), who report the figures of Broadberry et al. (2015); growth figures for France starting in 1331 are based on Ridolfi and Nuvolari (2021); growth figures for Germany start in 1501 via Pfister (2022); and U.S. growth – which remains the only "first generation" GDP series and also the shortest – from 1801, is sourced from Sutch (2006). All nominal long-maturity marginal interest rates and associated inflation data are based on the data set of Schmelzing (2025), analyzed further in Rogoff et al. (2024a) and with some earlier discussion Schmelzing (2022). Generally, we start our r-g headline

with long-maturity real rates.

[3]The marginal rate at time t is defined as the interest rate on new borrowing at time t. By contrast, the effective rate at time t is the weighted interest rates being paid on extant debt, excluding external debt. Meanwhile, Mauro and Zhou (2021) in their r-g analysis use effective interest rates, defined as the ratio of the interest bill to the stock of government debt, adjusted by a depreciation component that takes into account external debt – here, the authors use the average public debt stock in the current year and the preceding year as the stock variable for time t. Such a depreciation adjustment is mainly relevant for countries with relevant foreign currency debt issuance – a phenomenon relevant for emerging and peripheral markets, but not the advanced economies we are about to analyze.

Figure 1: Global AW duration-matched r-g, and trend, 1501-2023.



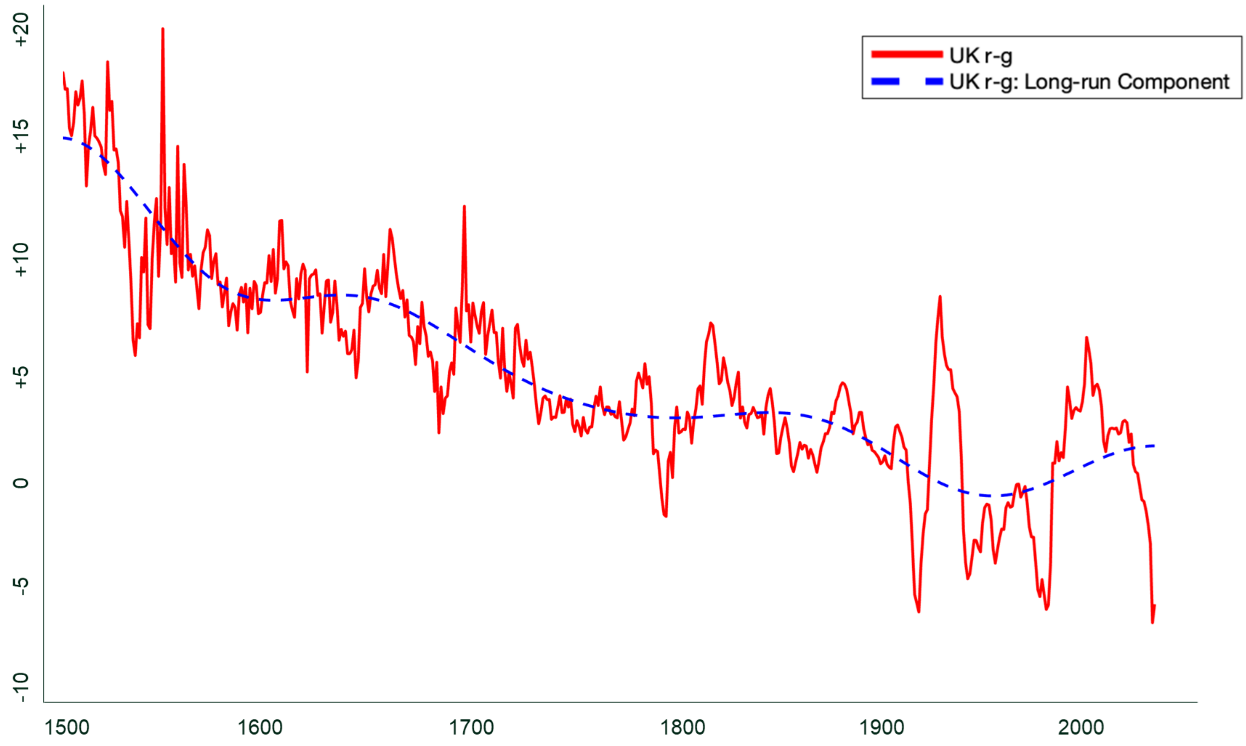
Notes: Annual data based on sources per Appendix section 1. Growth data g refers to aggregate output growth. Blue line is Mueller-Watson smoothed trend, via Müller and Watson (2018). To obtain the "Global AW" series, country series for U.K., France, Germany, and U.S. are arithmetically weighted (AW).

series in the early 1500s to obtain the greatest country coverage and comparability, though earlier data is available for all countries but the U.S.

Figure 2 for the United Kingdom and Figure 3 for France are generally similar to Figure 1: both begin measurements in the early modern period at absolute levels between 10-15%, and both generally decline by similar averages, of 2.8 (France) and 2.6 (U.K.) basis points per annum (taking the respective full observation periods). For Germany, the equivalent figure is -2.0 basis points p.a.

Equally, we can already observe in all three charts that negative r-g observations are a decidedly "modern" phenomenon: for instance, the first negative r-g observation on the annual level for the U.K. dates to the year 1786; for Germany, it occurs in 1792; for France, it dates to the year 1838. However, the visual inspection equally raises an intriguing hypothesis for the post-1914 era: specifically, it appears that at or around this point the centuries-long downward tendency of r-g across all series appears to have sharply moderated – or even, on some level, have gone into reverse. Certainly Figure 1 appears most dramatic in this respect, showing a steep trend reversal on the Müller and Watson (2018) trend basis (blue dashed lines). While substantial volatility around the trend is an obvious feature, it is prima facie grounds for entertaining the hypothesis of a *rising* secular r-g trend more seriously going forward, and to use tests that appropriately allow

Figure 2: Duration-matched U.K. r-g, 1525-2023.



Notes: Annual growth data to 2016 based on Broadberry et al. (2015) and Dimsdale and Thomas (2016), from 2017 via ONS (2024). Blue line is Mueller-Watson smoothed series, via Müller and Watson (2018).

for both slope and intercept breaks.

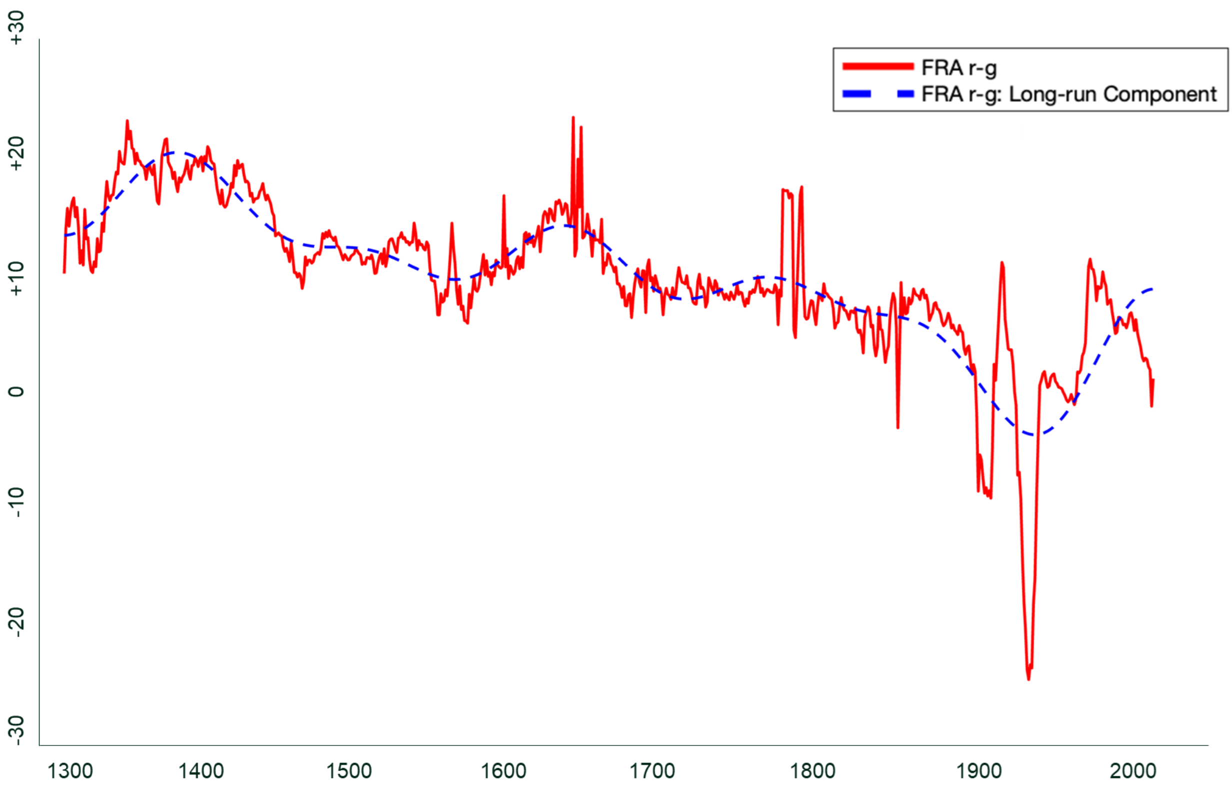
The dashed line in both series is the long-run smoothed non-linear trend found by applying Müller and Watson (2018); although the standard error bands using this approach are characteristically large, we note already that the mean estimate points upwards over the last century, foreshadowing the result we will get later in a formal trend break test, as we shall see in section 5.1.

Meanwhile, our reconstructions reveal many historical examples of fairly sharp (visual) reversals in duration-matched r-g, for instance during the 1550s or during the 1690s (the Glorious Revolution and Louis XIV's European Wars), the Napoleonic Wars, or the 1860s (the Crimean and Prussian Wars). In all these cases, fiscal outlays surged rapidly and r-g increases of more than 4% were observed within a decade of a new multi-decade r-g low – whether some of these instances rise to the level of structural breaks is something we will also test more formally therefore, to potentially find common drivers of such occurrences.

3.2 Further refinements

By focusing on long-maturity marginal government interest rates in our base case r-g constructions, we have thus far left out short-maturity components. Evidence suggests that over the 20th century, with the introduction of regular consolidated short-maturity public debt issuance, the share of

Figure 3: Duration-matched French r-g, 1331-2023.



Notes: Annual aggregate output growth data based on Ridolfi and Nuvolari (2021), duration-matched, with r series based on Schmelzing (2025).

short-maturity debt in total public debt stocks rose meaningfully and may have peaked in multiple advanced economies around the 1980s and 1990s (Francese and Pace, 2008).[4] In any case, while by necessity being limited to much shorter samples over time (consolidated short-maturity public debt is a relatively recent phenomenon), it could still be of interest to construct r-g measures that took into account the dynamic evolution of the average maturity of public debt stocks in advanced economies, and construct maturity-weighted r-g series: however, we note that the evidence points towards real short-maturity rates exhibiting identical econometric features with regards to trend stationarity, though that is not by necessity to be expected given term spread variation over time (Rogoff et al., 2024b). While adding much nuance, we would not therefore expect meaningful changes to underlying properties and conclusions presented here from such an exercise, therefore – especially not with regards to our results on trend stationarity.

[4]Similar developments are evident from time series that reconstruct U.S. and U.K. average maturity of public debt (Hall and Sargent, 2011; Ellison and Scott, 2020).

Table 1: ADF-GLS / Elliot, Rothenberg and Stock Unit Root Test, aggregate r-g basis					
Series	Time-period	test statistic	critical values		
		(t-value for α_1)	1%	5%	10%
U.K. r-g	1573-2023	-5.1505	-3.48	-2.89	-2.57
GER r-g	1501-2023	-4.4189	-3.48	-2.89	-2.57
FRA r-g	1331-2023	-3.8128	-3.48	-2.89	-2.57
U.S. r-g	1801-2023	-3.652	-3.48	-2.89	-2.57
Global AW r-g	1501-2023	-4.6897	-3.48	-2.89	-2.57

Notes: The table reports the results of the ADF-GLS test as per Elliott et al. (1996). All r-g series refer to aggregate measures of g. For a lag of 2 time periods with a trend and intercept. H_0 = series is non-stationary. H_A = Series is stationary. For t-value < critical value, series is stationary. Model: $y_t = \alpha_1 y_{t-1} + \alpha_2 \Delta y_{t-1} + \alpha_3 \Delta y_{t-2} + \epsilon_t$

4 Long-run statistical properties of r-g, 1500 - 2023

4.1 ADF-GLS stationarity

First, we test our duration-matched r-g series using the classic ADF-GLS methodology (Elliott et al., 1996) to establish whether r-g displays trend stationarity. Table 1 displays the results for aggregate measures of g and confirms trend stationarity for all country-series, as well as for the global aggregated series: while Table 1 assumes a time trend, there is strong evidence of stationarity even when dropping the assumption of a time trend.[5] Even for arguably the most important individual country series, the U.K., the significance is at the 5% level – and in fact all other countries exhibit significance at the 1% level. In Appendix section 6, we apply an alternative test that is robust to allowing for a structural break (Zivot and Andrews, 1992), and find very similar results.

4.2 Bai-Perron test

Next, we test our duration-matched r-g series using the conventional structural break test of Bai-Perron (Bai and Perron, 1998).

Per Table 2, which is implemented with a deterministic trend and allows for up to five breaks, we observe that all aggregate r-g series feature between two and four breaks. All breaks except for the U.S. 1979 break occur in 1931 and earlier – we recall, however, that the U.S. data is decidedly shorter than all other series, and is the only one not based on "new generation" national account series: when restricting structural break analyses to only one single break, the U.S. data indeed tilts more heavily towards a later, late 1970s break (see in particular Table 3). However, generally, a first clear cluster of r-g breaks can be detected in the 1564-1729 period, one that we will analyze in its fiscal historical context. A second – broader – cluster is apparent in the 1907-1931 period, where all series consistently show a break. These clusters are robust to alternative test specifications: we

[5] See Appendix section 2, table A.1, which confirms trend stationarity without the assumption of a time trend at the 1% level for all series except for the U.K.

Table 2:		
	Bai-Perron Test Results (For Deterministic Trend)	
Series	Breakpoint(s) (obs no.)	Breakpoint(s) (years)
Global AW r-g	124, 433	1623, 1931
France r-g	132, 235, 342, 579	1461, 1564, 1667, 1909
Germany r-g	330, 408	1629, 1907
U.K. r-g	157, 359	1729, 1931
U.S. r-g	63, 96, 131, 179	1863, 1896, 1931, 1979

Notes: The table reports the results of the Bai-Perron test as per Bai and Perron (1998). The r-g series here refer to aggregate measures of g, and assume a deterministic trend. Model: $y_t = \beta_{j0} + \beta_{j1} t + \varepsilon_t$, Where j=1,. . . ,5 is the number of breaks. Up to 5 dates could be selected by the model.

will analyze these two historical context and plausibility of these suggested break date clusters further in a subsequent section.[6]

4.3 Perron and Yabu 2009

Third, we test the econometric properties of the long-run r-g series with recourse to Perron and Yabu (2009), who present a test for structural changes in a trend function of a time series without any prior knowledge of whether the noise component is stationary or integrated.[7] This method tests for a single structural change in both intercept and slope – a feature that is particularly well suited for our r-g series given that Figures 1-3 visually invited such a proposition, per the Müller and Watson (2018) trends above. The difference between this method and previous Bai-Perron calculations is most importantly that Bai and Perron (1998) only tests for the intercept, whereas Perron and Yabu (2009) is ideal for our purposes as we are interested in *both* the intercept and trend r-g breaks – in other words, whereas the question of whether the direction of global debt sustainability has changed fundamentally. Like the Bai-Perron test, the Perron and Yabu (2009) also does not impose any priors on when a break might have occurred.

A limitation of the Perron and Yabu (2009) test is that, unlike Bai and Perron (1998), it only allows for a single break: the code finds the most likely point of a break and provides a test statistic to determine if that break is significant. If the test statistic is insignificant, then there are no breaks.

Table 3 reports distinct joint breaks in intercept and trend for all respective series. Importantly, we observe that all these breaks are clustered between 1907-1933 and that all of them are significant at the 1% level: the single exception is observed for our shortest series, the United States, where the

[6] In the Appendix, we also report results assuming no deterministic trend (Section A.2).

[7] We use the underlying Matlab code for Perron and Yabu (2009) as per the 2021 release via Pierre Perron's website, see: https://blogs.bu.edu/perron/codes/.

Table 3:						
Perron and Yabu (2009) Test Results (Structural Change in Intercept and Slope)						
Series	Breakpoint (obs no.)	Breakpoint (year)	Test Statistic (W-RQF)	Critical Values		
				10%	5%	1%
France Aggregate r-g	579	1909	18.5657	2.48	3.12	4.47
Germany Aggregate r-g	407	1907	8.5631	2.48	3.12	4.47
U.K. Aggregate r-g	406	1931	13.2972	2.48	3.12	4.47
U.S. Aggregate r-g	178	1978	17.6323	2.48	3.12	4.47
Global AW Aggregate r-g (1501-)	433	1932	4.5729	2.48	3.12	4.47

Notes: Model: H_0: Series does not contain structural change. H_A: Series contains structural change. For test statistic > critical value, the series has a structural change. Model 3: $Y_t = a_0 + a_1 \cdot DU + b_0 \cdot t + b_1 \cdot DT + \varepsilon_t$, where $DU = 1(t > TB)$ and $DT = 1(t > TB) \cdot (t - TB)$.
Obtained by minimizing the sum of squared residuals from a regression of the relevant series on a constant, a time trend, a level shift, and a slope shift dummies.

year 1978 is identified.[8]

Another commonly used methodology to assess structural breaks is represented by the Chow test (Chow, 1960), which of course requires one to impose priors on where breaks might have occurred, and as such has somewhat fallen out of fashion. However, for completeness, in the Appendix (section 3), we report results on the Chow test methodology for both global r-g and the U.K., testing 1981 jointly with other key historical potential break dates; that Appendix also reports half-lives of all key series, which we have already referred to.[9]

Taken together, these econometric results appear to suggest strongly that, when one looks at a long enough time series, r-g differentials should generally be modelled as stationary variables that exhibit mean reverting behavior. However, a cluster of relevant structural breaks appears to be confirmed for the interwar period, with a trend break especially prominent in the period 1907-1932, affecting all countries under investigation. We will discuss the rationale for why such a break should have occurred then in the following section.

[8] In the Appendix (Section A.4), we display separately the tables for breaks in the trend only, and breaks just in the intercept.

[9] Half-life results following the methodology of Steinsson (2008) suggest general ranges of 5-12 years for r-g variables, details in Appendix section 4.

4.4 Monte-Carlo exercise

The Appendix, section 8, features detailed results of a Monte-Carlo exercise. This exercise is mainly designed to explore to what extent the use of long samples really matters for the identification of "true" r-g econometric properties. We observe there that the identification of trend stationarity clearly improves the true positive identification of r-g econometric properties, across a wide range of previous forecasting horizons in the literature – underpinning the relevance of using long sample r-g series as we introduced here. We show that in the large samples we introduce in this paper, the correctly specified model can reject the null with frequency 0.97, close to the nominal size of the test. Once sample size drops to n=100, rejection frequency decreases substantially. On the other hand, the mis-specified model without trend cannot reject the null even in large samples.

Generally, these observations raise the possibility that shorter samples – with which existing r-g literature has near-exclusively operated thus far – mis-specify not just stationarity features but also break dates. In practical terms, the Monte-Carlo exercise suggests that even the 75 or at best 150 year samples for debt sustainability variables used in seminal works such as Bohn (1998) or Blanchard (2019) apparently fall short of the robustness our multi-century samples offer.[10] For 150-year series, say, the false positive of identifying a unit root still ranks in the order of 60-70%, per the results. The upside of using very long-run debt sustainability measures is therefore far from anecdotal, but it appears that they indeed decisively improve standard econometric exercises and the identification of time series properties.

5 Secular r-g trends, the 16th century inflection, and the 1930s breaks

We have briefly mentioned evidence of several sharp r-g reversals that are for the first time measurable in our long-run data – some of them are even visually apparent e.g. on the Mueller-Watson basis, others are econometrically suggested. In addition to the historical context on trends, we look more granularly at these major reversal episodes, assessing the public finance and political context, including fiscal sources. Afterwards, we turn to an analysis of the key formal break episode confirmed across both country and global series, the 1930s.

5.1 Fiscal histories – secular context and the 1564-1729 cluster

It is obvious that the public debate half a millennium ago did not engage with the concept of "r-g" at all, and thus primary fiscal sources will not reveal any firmly testable statements – but at the same time, our new data should align with the general contours and event study evidence that generations of fiscal historians studied for all our advanced economies. The plausibility of general trends and sharp movements as compared to these sources is the focus of this section – not least to provide a more stringent "reality check" of key inflection points and the political-fiscal context of

[10] Bohn (1998) emphasizes the need to use long sample measures for his example of U.S. data, but even his time series exercises "only" span the period 1916-1995, 79 annual observations in other words.

econometric dates suggested. In the Bai-Perron exercise via Table 2, we identified two particularly notable "clusters" of r-g breaks over the past half millennium in the key new duration-matched r-g series, relatively short periods where all country-series and global constructions record break dates – dates that are also relatively consistent across different test specifications. In addition to these "clusters", there are country-specific break dates which we can assess for historical plausibility. We begin with a historical assessment of the first major r-g inflection, as demarcated by the 1564-1729 era:

- First, prior to the inception point of this era, 1564, global r-g trends downwards from initial levels of around 12-13% – this is a period marked by the Age of Discoveries (setting off a large-scale bullion influx and the "Price Revolution"), by the Habsburg-Valois clash in the Italian Wars, and by the rise of internationally operating lenders such as the Fuggers. On the growth side, European economies' income levels are shrinking in real per capita terms in Germany, are stagnant in France, and rise moderately in England: generally, moderately positive levels of population growth account for structurally positive aggregate g growth.
- But around 1564, r-g switches signs and begins to rise moderately, from levels just above 5% almost returning to double-digit levels. Most aggressively, this turnaround is being felt in France, which is the first country to record a robust break in the 1564-1629 cluster. France is shaken by the final defeat in the Italian Wars (1559), and soon engulfed in domestic instabilities, aggravated by religious strife. Elsewhere, too, the decades around 1564 are fiscally particularly volatile, with the "debtor from hell" reigning in Spain and the German Habsburgs (lacking a regular central government tax capacity) perpetually teetering on the brink of default, too (Rauscher, 2004).
- This public finance turbulence also includes England (U.K.), where r-g reaches more than 20% by 1555, amid Henry VIII's currency experiments and the excessively costly wars against France and in Ireland. The fiscal situation worsens gradually – from the reign of Queen Mary I few contemporary fiscal sources fail to lament the the "heavy incubus of debt"(Dietz, 1921).[11] In response, Queen Elizabeth begins with the approval of sizable forced loans, including a major GBP 99,999 forced loan in 1597 (0.5% of English GDP) which was never repaid. The fiscal situation deteriorates further from here – with English r-g reaching 11% by 1610 – and the damning verdict on her successor James I. was that "neither [in 1602] nor at any subsequent time is there any indication that James had any sense of the value of money or of the meaning of the balance of credit and debit. He seems to have been incapable of understanding the fact that a considerable income and even the occasional presence of large amounts of coined money in the hands of the tellers of the exchequer were entirely compatible with a condition verging on bankruptcy". Formally, the year 1729 is then identified via Table 2 as a key break for the U.K. – this date, marking the early reign

[11] By the 1580s, the war in Ireland alone cost the Treasury about 17% of GDP over the decade 1588-1599; another 8% of GDP combined was spent on subsidies for Dutch and French Protestant allies, such as the Duke of Alençon, see Dietz (1921), who remains the best resource for English fiscal history during these centuries.

of the Hanoverians, is finally associated with a soothing of the fiscal volatility, and sees the inception of a downward trend in r-g continuing until the eve of World War I.

- Interestingly, the year 1629 – the year of our German r-g inflection – is indeed often associated with a key inflection point of the Thirty Years War, by that time the most destructive international conflict since at least ancient times. In the historiography of the Habsburg Monarchy, the period of 1628-31 has been identified as *en gran viraje* – "the great turning point of its history" (Stradling, 1990). After initial successes over 1620-1627, the year 1627 saw the end of the siege of La Rochelle – which finally opened up a critical second front for the Habsburg Empire via France, quickly followed by the outbreak of the disastrous Mantuan War. But most importantly in fiscal terms, however, was the spectacular *coup* by the Dutch rebels in September 1628 to capture the Spanish treasure fleet docking in the harbour of Matanzas – an event wreaking "havoc" on Habsburg finances: bullion worth more around 10 million ducats was lost (close to a full year of war expenses), in a "major disaster" for the entire Habsburg Empire, including its German constituents.[12]

- Especially from the 1640s, the increasingly international dimension of the Thirty Years War – by then financially involving all relevant European powers, including the U.K. – raises the spectre of spectacular sovereign defaults. Indeed, the Spanish sovereign decline now spreads through the entire Habsburg realm, and from there beyond: the proud city of Milan enters bankruptcy in 1635 after excessive debt monetization through its state bank *Banco Ambrogio*; upon the cessation of hostilities, France immediately defaults on at least 146M Livres Tournois in debts in 1648 (equivalent to 9% of GDP); in the Holy Roman Empire, a wave of defaults and restructurings ripples through the key principalities and dukedoms, including large defaults by Brandenburg from 1642, Württemberg in 1652, Saxony in 1656, the Palatine in 1669 – at the Reichstag in 1654, the Emperor announces a general debt moratorium.[13] Our observation of both highly volatile and elevated global r-g levels by the 1650s (reaching its second highest level ever on record by 1659) are thus very well backed up by the historical record.

- Around 1665, after drawn-out attempts to resolve the sovereign debt crises, the upwards trend in r-g appears to reverse, and enters a renewed downward trend that will eventually last until the early 20th century. Despite geopolitical risk remaining elevated – with sharp spikes around the wars of Louis XIV and the Napoleonic Wars – real interest rates fall sharply over the subsequent 250 years, with a clear acceleration of real aggregate GDP growth in all

[12]Spain was at the time almost fully bankrolling its Habsburg-Bavarian Catholic allies during the Thirty Years War: for details from the Spanish policy perspective over 1628-9, a premier resource remains Elliott (1986, 362ff.). For a new German perspective, see Münkler (2017, esp. chapter 4).

[13]On this default wave, recent literature is relatively oblivious. See in some detail (in German) Kaphahn (1912), and on Brandenburg Krug (1861). On the French default, see Bonney (1981), on Milan Cova (1972), and on the Spanish defaults in 1607, 1627, and 1647, see Boyajian (1978), and Kirk (2005) – with Casoni (1800) providing useful details on restructurings for the Genoese creditors.

constituent economies even pre-1800.[14]

- In sum, the cluster of trend breaks that we identify during 1907-1932 appears to be have its most important precursors in period during 1564-1729: these years saw key safe asset providers ultimately collapse under the weight of "fiscal overstretch" and years of (military-driven) deficit financing – with the period between 1729 and the first third of the 20th century being associated with a continuous, broad-based, global improvement in public debt sustainability.

This result of course naturally puts the focus now at this second, more recent and broader break cluster: we recall that all country level and global r-g series showed evidence of a break in the 1907-1932 period, across a variety of test specifications that incorporated both trend and intercept breaks.

5.2 The 1907-1932 trend breaks

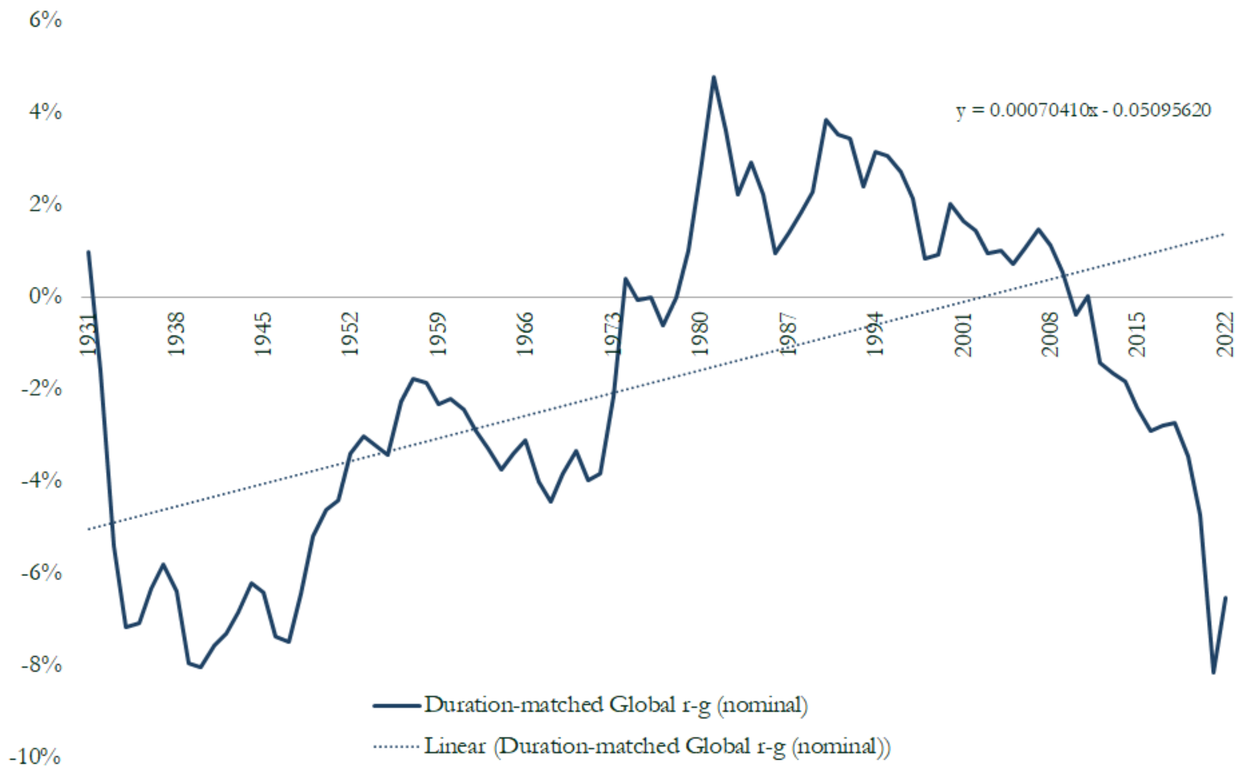
Even more pronounced than the first cluster– and much more condensed – the key cluster of suggested breaks in intercept and trend across our new r-g series concerns the 1907-1932 era: not least, despite some plausibility that individual countries could have seen a break in the 1980s, the accumulation of evidence across series for a broad international break the interwar era appears dominant over other dates.

To get closer to understanding the underlying context, we investigate the sub periods separately, following the (intercept and trend) break points obtained via Bai and Perron (1998) and Perron and Yabu (2009). Figure 4, which displays the "Global AW" duration-matched r-g data over 1931-2023, suggests that the key structural break during this period reversed the previous multi-century *downward* trend in r-g, which then came to an end. Over the observation period of 1931-2023, "Global AW" duration-matched r-g increases by an annual average of 7.0bps when measured on a linear basis, but the periods exhibits substantial volatility preventing conclusive econometric statements. What is clear is that the post-1932 era represents a substantial qualitative change compared to the downward trend of 2.7bps observed for the same series over 1507-1931. If one looks at the longer sample, the idea that the world today can likely count on favorable trends in r-g and public debt sustainability dynamics, as much recent literature has done, becomes questionable.

The phenomenon that the downward r-g trend came to an end and was replaced by an era of volatility since the interwar period also holds for all individual country series. For the U.K., taking the break date of 1931 in Table 3 as the relevant inception point, we observe an annual increase in duration-matched r-g, which stands at 5.4bps p.a. over the period 1931-2023. For Germany, where Table 3 suggests 1907 as the decisive break in intercept and trend, the 1907-2023 upwards slope

[14] For relevant general surveys covering this period for our key economies and the global level, see on the fiscal side Dincecco (2011) and on the growth side Bolt and van Zanden (2024).

Figure 4: Duration-matched Global AW r-g, 1931-2023.



Notes: Annual data based on arithmetically weighted German, French, U.S., and U.K. duration-matched series ("Global AW"), as described above. Numerical label represents linear slope formula.

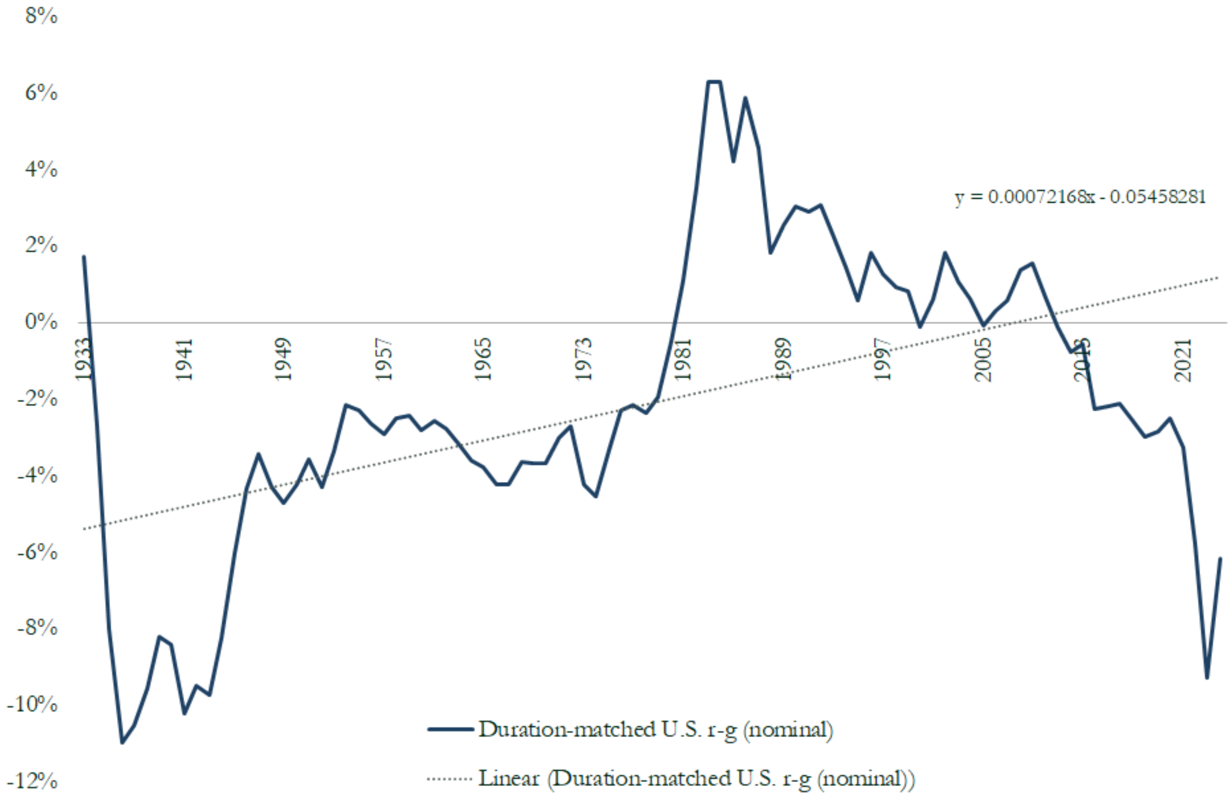
stands at an even steeper 7.4bps per annum.[15] For the U.S., where 1931 marks a key mean break in Table 2, the 1931-2023 upwards slope in r-g stands at 7.2bps p.a. Figure 5 displays the U.S. series visually, illustrating the degree to which post-COVID r-g level may be unusually depressed relative to the secular trendline: relative to the trend-implied 2023 value of 1.7%, the actual 2022 figure (at -6.1%) thus stands at a meaningful spread to this trend (even though it is too early to definitively say whether COVID has created a new structural break, given the limitation of post-2021 data).

We recall that the immediate pre-1945 era, which records the unusually elevated share of negative r-g observations, also saw an unusual degree of financial repression in advanced economies related to war-era policies, as noted earlier. All else equal, the lifting of yield curve controls and other liberalization measures should be associated with rising interest-output differentials (Reinhart and Sbrancia, 2015; Mauro and Zhou, 2021).

However, timing-wise the turnaround in advanced economy r-g differentials appears not to be a function of the lifting of such policies after the conclusion of the war – we note that caps on long-maturity interest rates in the U.S., for one, were only lifted by 1951, the time of the Treasury-Fed accord. And of course, financial repression appears unable to explain why r-g trends did not

[15] Per the underlying Schmelzing (2025) data set, the German hyperinflation years are excluded from the interest rate series and instead linearly interpolated, following similar approaches in the literature.

Figure 5: Duration-matched U.S. r-g, 1931-2023.



Notes: Annual duration-matched data based for U.S. with GDP growth figures based on Sutch (2006) and Bureau of Economic Analysis (BEA). Numerical label represents linear slope formula.

re-set to the downward trend after its phasing-out. It appears more promising therefore, to search for deeper structural drivers. In this spirit, we turn to a closer look at the (aggregate) growth component underpinning our debt sustainability constructions.

5.3 Growth inflections

We depart from the observation in Rogoff et al. (2024a) that evidence for a "recent" structural break in sovereign real interest rates in the modern era is not compelling, when utilizing econometrically powerful multi-century data. So by inference, this suggests that the evidence of a secular inflection for r-g may could plausibly by driven by changes in output growth rates. We test this idea by analyzing separately aggregate output growth rate series.

Indeed, via Table 4, we test the most recent output growth rate series separately for structural breaks Bai and Perron (1998).[16] While we are not the first of course to test output growth series for stationarity features, to our knowledge we are the first to use the "new generation" of post-Maddison national accounting data for such exercises – which include Broadberry et al. (2015)

[16]See Appendix section 9 for ADF-GLS results, where among other exercises we confirm all output growth series to exhibit trend stationarity, at the very least at a 5% significance level.

for the U.K., Ridolfi and Nuvolari (2021) for France, and Pfister (2022) for Germany. Among such relevant existing econometric analyses, see for instance Antolin-Diaz et al. (2017), who are representative in using post-war U.S. data (1947-), which leads them to posit a structural GDP growth break around the year 2000.

Looking at our long-run results, it is striking that a clear cluster of growth breaks is once more confirmed for the 20th century interwar period, with no less than six out of eight breaks in this era (across all the combined series). The only outlier: Germany, which experiences an unparalleled post-war boom in the aftermath of its near-extinction shows an additional 1975 break when its "growth miracle" ("Wirtschaftswunder") comes to an end.[17]

Against the backdrop of a scarcity of real interest rate breaks in this period (Rogoff et al., 2024a), it therefore appears that indeed, an aggregate growth inflection took place in the interwar period which mainly accounts for the detrimental evolution of r-g trends since then. In all cases, the output growth rate is on a downwards-sloping trajectory since the most recent structural break as identified by Bai and Perron (1998): that is not least true for the U.S., where real aggregate output growth rates since 1945 roughly halved.

Figure 6 displays the data for the U.S., the Global AW level, and Germany visually: the respective series are initiated at the date of the most recent structural mean break as reported in Table 4 on the basis of Bai and Perron (1998), and a simple linear trend is imposed. Strikingly, all three series point downwards – the real growth rate basis that is relevant for public debt sustainability measures has become consistently more adverse for several decades.

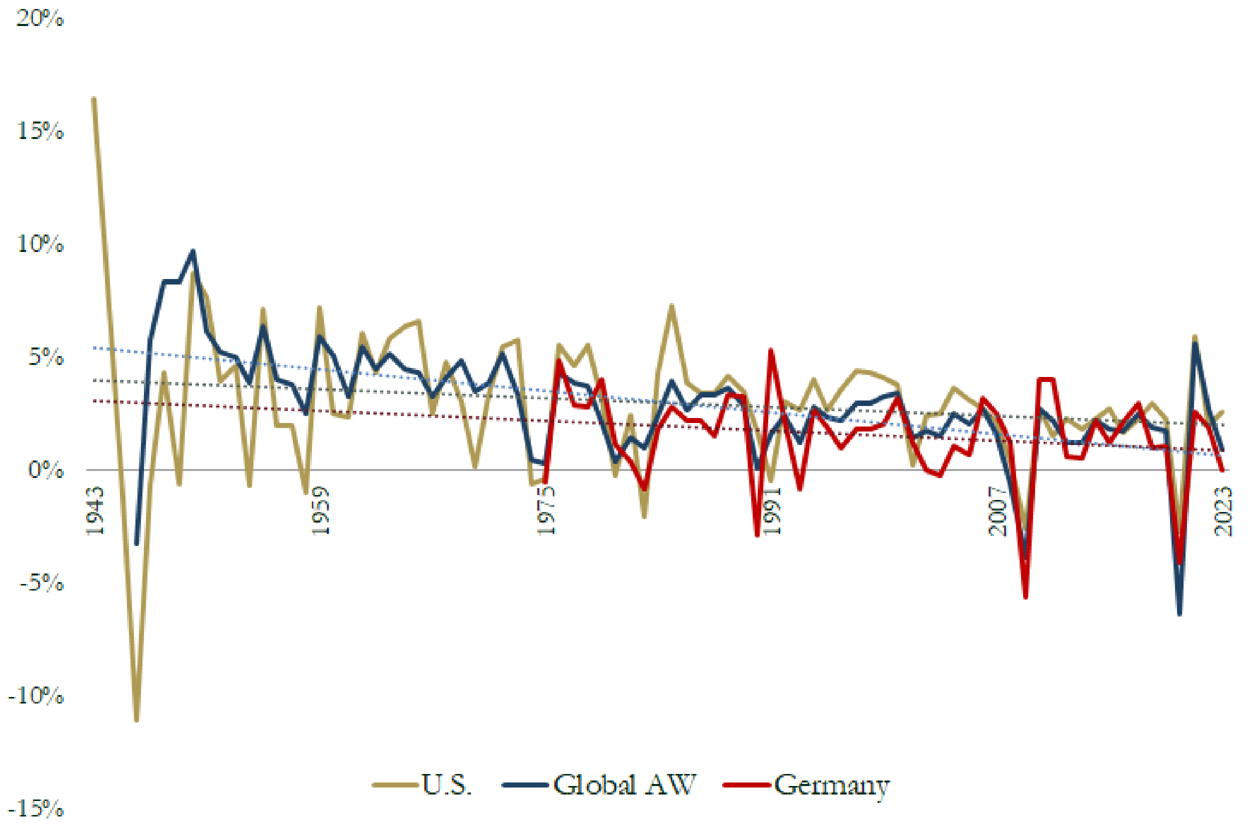
[17] In the Online Appendix, we also show results for Perron and Yabu (2009), which also confirm joint breaks in the intercept and trend being concentrated in the 20th century interwar period.

Table 4:		
Bai-Perron Test, real aggregate GDP growth rates.		
Series	Break year	Criterion
U.K.	–	–
France	1944	Seq
Germany	(1919), 1946, 1975	BIC, Seq
U.S.	1929, 1943	BIC
Global AW	1817, 1946	Seq

Notes: The test follows Bai and Perron (1998), assuming a time-trend, and allows for break in both trend and level. Maximum number of breaks allowed is 5, results reported at 5% significance level. "Criterion" indicates the optimal number of lags according to the sequential procedure ("Seq"), the Bayesian Information Criterion (SIC), or the Modified Information Criterion (MAIC). The test rejects when the test statistic is negative and larger (in absolute value) than the critical value. The result in brackets (Germany 1919) indicates significance only on Seq basis. GDP growth sources per text and Appendix section 1.

Often, observers may have in mind the generally favorable *per capita* growth trajectory over the 20th century, and have (against slowing population growth) therefore all too lightly neglected that the secular trends on the aggregate growth basis – the basis that matters for aggregate public debt sustainability – have become more and more adverse well before 1970. Indeed, what the growth picture therefore suggests is that aggregate growth may have positively underpinned public debt sustainability in the early modern (pre-1914) period, and in some countries (such as the U.S.) into the first decades of the 20th century. But since then, the contribution from aggregate output growth has been decidedly *unfavorable* for public debt sustainability, starting well before the productivity slowdown sometimes pinpointed to 1970 (Gordon, 2016) or 2000 (Antolin-Diaz et al., 2017): all else equal, falling output growth rates for the past decades have in fact mechanically pushed up r-g measures in all advanced economies under long-run investigation here. Absent any countervailing evidence from real interest rates – where evidence of 1914 or 1980s breaks is decidedly weak (Rogoff et al., 2024a) – this puts considerable doubt on the popular notion that advanced economy debt sustainability has been secularly improving over recent years. Much, in fact, supports the notion of structural trends in the opposite direction, namely upwards-sloping r-g trajectories. Of course, it is possible that the upward trend break starting in the 1930s will prove ephemeral, and that the post-1980 period reflects only a restoration of the long-term trend downwards instead of upwards. The data do not yet support this thesis, but nor is possible to reject if one has strong priors, as results from Chow tests in Appendix section 3 show. In the next section, we argue, however, that a closer look at historical changes that happened in the 1930s and that persist today, should give one pause in ignoring pre-1980s data.

Figure 6: Aggregate real growth rates, U.S., Germany, and Global AW, most recent structural mean break to 2023.



Notes: Series displays real aggregate GDP growth rate from the year of the most recent structural break via Bai and Perron (1998), to the year 2023. Dotted lines impose linear trend. For all GDP growth sources, see Appendix section 1.

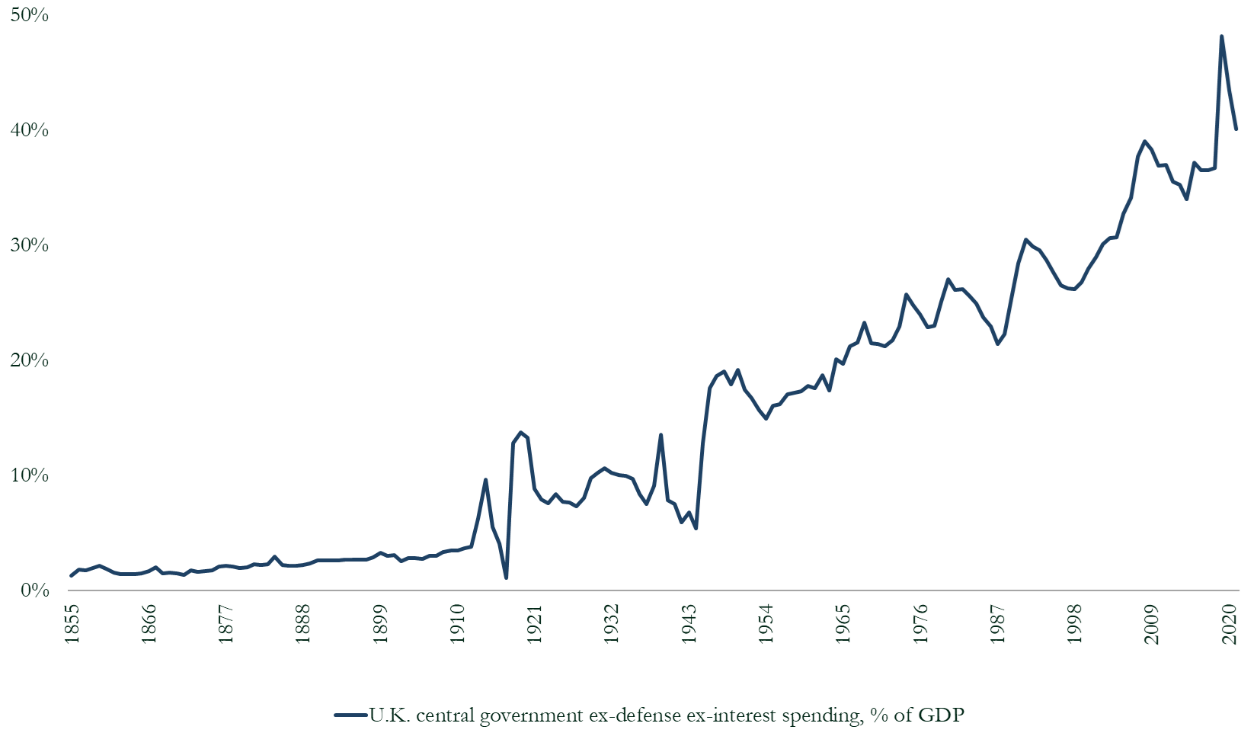
5.4 The rise of welfare states, and the 1930s inflection

Prior to the 20th century, advanced economies spent most of their public expenditures on defense and interest expenses. Incipient poor relief spending did occur through public budgets, but recent evidence suggests these factors did not assume relevant levels relative to GDP, nor do they display pre-1850 time trends.[18] Often, the Bismarckian reforms in the late 19th century are seen as a watershed, and indeed German social transfer payments to GDP from 1850 can be reconstructed on the basis of Hoffmann (1965, table 231).

More granularly, using the data in Dimsdale and Thomas (2016), we can reconstruct Figure 7, which displays U.K. central government non-defense non-interest expenditures relative to GDP on an annual level over 1855-2016. A visible qualitative change occurs over the first few decades of the 20th century. Within two generations, non-defense, non-interest expenditures surge from levels below 5% to reach more than 20% by the 1960s. The residual in public spending – defense

[18] Per van Bavel and Rijpma (2016), who study data for Northern Italy, the U.K., and the Netherlands, poor relief stayed at a relatively stable 2-5% of GDP in these societies from the earliest pre-1850 data points – however, numbers include institutional spending, notably by religious orders.

Figure 7: U.K. central government non-defense, non-interest expenditures to GDP, 1855-2022.



Notes: Annual data based on Dimsdale and Thomas (2016) and sources therein to 2016, and afterwards via U.K. Treasury (PESA accounts).

and interest spending – meanwhile show large spikes during World War I, and World War II, but if anything secularly display moderate downward trends: in fact, British defense spending as a percentage of GDP stands at 2.96% over 1855-1911, but at only 2.85% over 1980-2022. When testing these expenditure series for trend and mean breaks, the visual impression of a sharp inflection in the 1940s is confirmed: per table 7, the matching Perron and Yabu (2009) approach (analogous to its application to our r-g bases) reveals 1945 as the single trend break in U.K. non-defense non-interest spending.

Not least, this chronology is also well supported by historical evidence. Historians have referred to the 1942 Beveridge Report and the subsequent 1945 Labor election victory as the "birth" of the modern welfare state in the British context (Fraser, 2017). From 1945, the Attlee government then implements key pillars of the Beveridge report, including the National Health Service, national insurance, and unemployment programs, with the Dalton and Cripps Treasuries highly influenced by Keynesian ideas (Peden, 2000, chapter 8) – in 1948, for the first time in British history, non-defense non-interest spending hit 15% of GDP, a level below which it should never fall again ever since.[19]

Figure 8 displays equivalent trends for the United States and Germany, showing non-interest

[19]There are, of course, other factors that may have especially impacted growth after World War I, in particular the rise of unions.

Table 5: Testing for Shifts in Trend with an Integrated or Stationary Noise Component					
Series	BreakDate	test statistic	critical values		
			10%	5%	1%
U.K. non-defense non-interest expenditures/GDP	1945	22.11	2.48	3.12	4.47
U.S. non-defense non-interest expenditure share of total	1940	12.10	2.48	3.12	4.47

Table 1: Model: $y_t = \beta_{j0} + \beta_{j1}t + \varepsilon_t$, where j=0,1 is the number of breaks (trend). Trimming is set to 15%.

non-defense spending relative to GDP. The two series generally trend intriguingly similar over the long-run, beginning at levels below 5% pre-1914, before markedly rising to levels close to 10% by the interwar period. Afterwards, in the U.S., the New Deal typically marks the inception of qualitatively new welfare policies: from welfare spending below 0.1% of GDP through the 1920s, welfare spending had reached more than 2% by 1945. However, for the U.S., table 7 suggests that the most likely trend break – 1890 – does not assume relevant significance.[20] In both countries, we subsequently observe two periods of particularly sharp rises – in the late 1960s, and after 2008. Compared to 2007, U.S. non-defense non-interest spending rose by no less than 11.5% to 2020 (with the 2019-20 rise exceeding the rapid 1917-8 rise during World War I), and in Germany, it rose by just over 9% – in both cases ending the series at all-time records above 25%.

We emphasize that in all the non-interest non-defense spending series show, unfunded welfare liabilities (pensions) are excluded – integrating these spending commitments would further sharpen the described trends and could substantially raise absolute numbers relative to GDP, but require various assumptions about discount rate smoothing.[21]

One can confirm basic trends also via annual-level U.S. federal non-interest non-defense expenditures as a percentage of total U.S. federal expenditures, over the period 1791-2022, using the historical statistics in Wallis (2006). Notably, there one observes that – while the percentage has dropped sharply during war events, including the Napoleonic Wars, the U.S. Civil War of the 1860s, and World War II – there also appears to be a secular trend towards higher and higher shares of total federal expenditures allocated to non-interest non-defense sectors. Crucially, as of 2021, the non-interest non-defense share of total expenditures hit a new all-time record, at no less than 79.9%, surpassing the previous peak of 1936, when the acceleration of New Deal policies accounted for a share 77.5%.

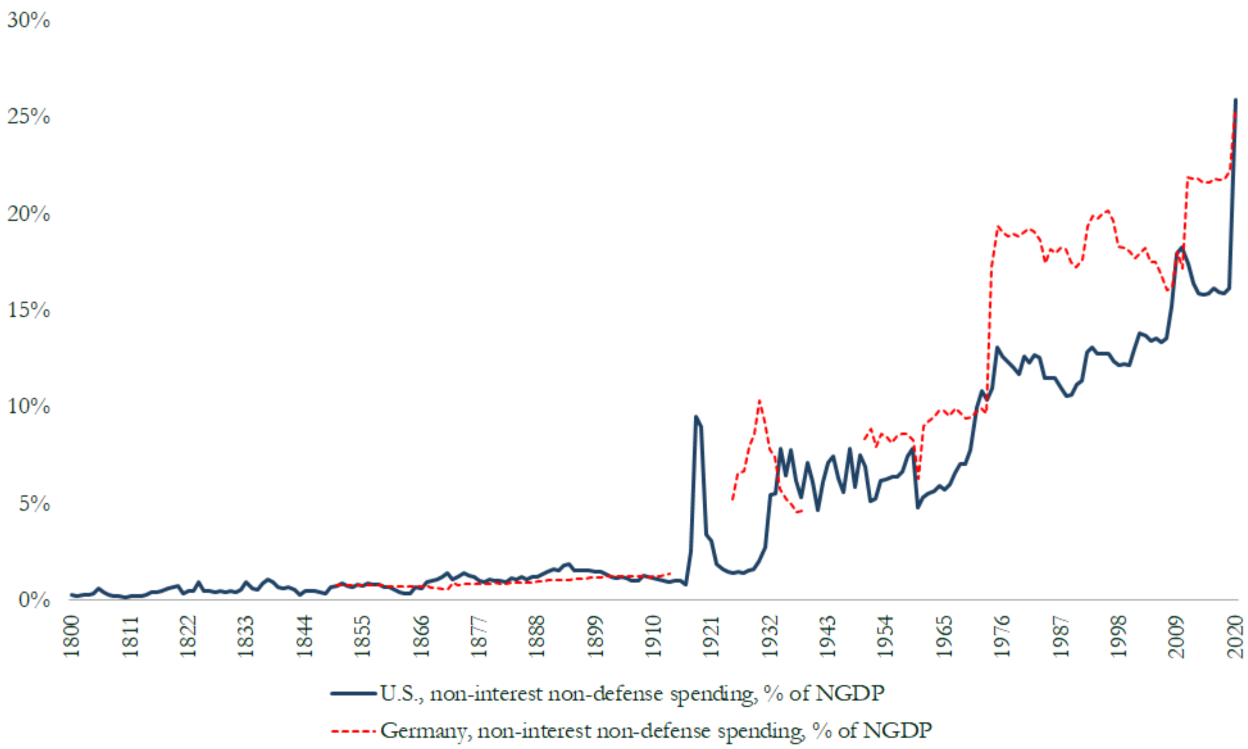
In other words, the COVID-related federal expenditure programs enacted over recent years have

[20]The 1890 date is plausibly related to the federal spending drive under the Harrison administration, which is typically credited as the key policy reason for the 1890 Republican election defeat (Socolofsky and Spetter, 1987).

[21]For the U.S., Giesecke and Rauh (2023) document a rise in unfunded pension liabilities from levels well below USD 500BN in the 1970s, to more than USD 6.5TN (market value) in 2021. Similar evidence exists for other advanced economies.

led to multi-centennial record of public funds allocated (predominantly) to welfare spending.

Figure 8: U.S. and German federal non-interest, non-defense expenditures, % of GDP, 1791-2022.



Notes: Annual data over 1791-1959 based on Historical Statistics of the U.S. Millennium Edition, chapter by Wallis (2006). From 1960, U.S. data via FRED. German data via Hoffmann (1965) for pre-1950 data, Rahlf (2016) for 1950-2011, and GENESIS/Deutscher Bundestag for 2011-2020. The latter two German bases calculate non-interest non-defense spending on the basis of departmental approaches, thus including some pension spending for defense personnel. West Germany is used over 1950-1990.

Finally, of course we acknowledge that the post 1980s period witnessed a number of extraordinary developments that may have raised growth relative to real interest rates, including the rise of independent central banks (Rogoff, 1985), the peace dividend resulting from the fall of the iron curtain and the rise of China. However, as Afrouzi et al. (2024) emphasize in a paper about long-term inflation trends, many of the these trends have gone into reverse along with a rise in populism, so our read is that it is by no means clear that factors pushing towards higher r-g in the post-pandemic era are likely to go away anytime soon, albeit it is beyond the scope of this paper to dissect these structural factors.

5.5 "Risk-adjusted" r-g approximations and trends: an extension

A long line of previous studies has emphasized the intuitive relationship between "r-g" in the context of debt sustainability debates, and its centrality as a concept in the influential debates on *dynamic efficiency*. While Diamond (1965) and others use government interest rates as identical to marginal product of capital, given capital risk, it is of course not clear that government interest rates track the aggregate return to capital well over time – while these early modern government interest rates may have been "safe" in a relative sense (vis-a-vis other early modern assets), they are likely to embed higher absolute risk premia than today. Several recent contributions have therefore emphasized the need to embed measures of "risky" capital returns in r-g series to closer approximate capital stock returns, whether for purposes of assessing dynamic inefficiency, or for public debt analyses (Reis, 2021; Mankiw, 2022). However, even when analyzing the welfare costs of debt rollovers – and general public debt sustainability features – risky capital returns are key to assess feasibility – if risky capital returns are significantly above output growth rates, public debt expansions may have meaningful welfare costs, even if the safe interest rate is below the output growth rate (Blanchard, 2019).

Recently, Reis (2022) has aggregated private and public capital returns for the U.S. and other advanced economies – arguing that the wedge between public and private capital returns may qualify arguments on r-g trends – an approach we build on.

Measuring corporate returns over centuries is fraught with much greater measurement issues compared to long-lived governments. To empirically approximate "risky" elements of the capital stock, the literature has typically turned to corporate profits or equity returns (Abel et al., 1989).

Now, to go one step further, we can contemplate briefly how the separate "safe" and "risky" components could be integrated over time, to yield a closer approximation of aggregate MPK trends, thus aggregate r-g trends, over time. To to this, we require some general idea about modern capital stock evolution and respective shares of risky and safe assets, to attach general weights to them. Various data points are available to give us an idea: Goldsmith (1985) remains the source for many early modern studies, reconstructing national balance sheets for twenty advanced economies from 1688. Recent updates to his series are available from 1870 via Jordà et al. (2019) and, for 18th century Netherlands, Korevaar (2023).

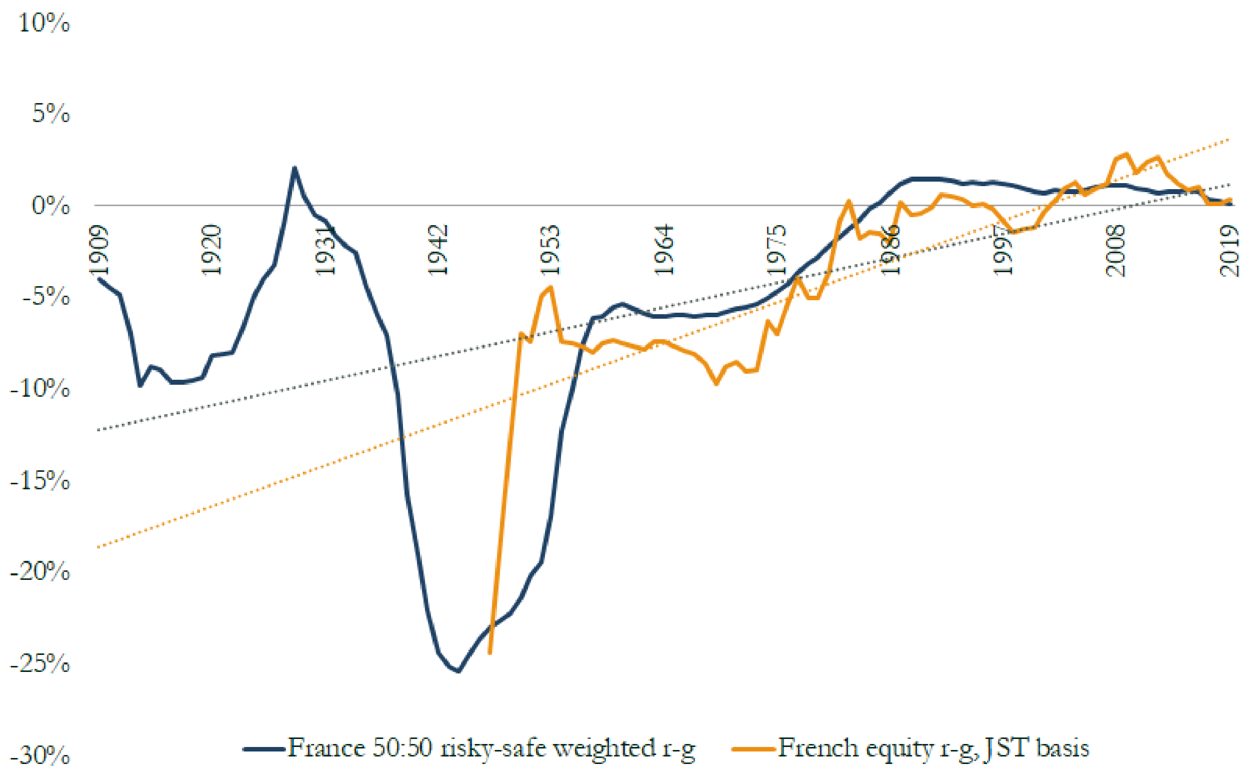
Generally, these balance sheet decompositions – for France, and for other advanced economies – suggest equity shares in the pre-1800 era below 20%, but rising with p.c. income levels. French-specific top-income asset compositions for earlier periods can also be derived from Forster (1971) and Dewald (1980). While we abstract from land for the moment – for which separate time series are available – equity and public debt assets constitute the key financial assets in the capital stock.[22]

Importantly, however, there is evidence across alternative procedures that the French r-g series displays a break in 1905 or 1909 (via Tables 2 and 3, with risk-adjusted breaks occurring in 1933-4), against a backdrop of clear trend stationarity (Table 6). Figure 9 therefore just displays the sub-period of 1909-2020 for our French "risk-adjusted r-g". Importantly, it is evident on this basis that even "risk-adjusted" r-g – including a basis that uses aggregate French dividend yields from 1947 (orange line), rather than just the Bazacle basis – appear to shows a clear upward bias since the 1909 break. On this basis, it appears once more that negative r-g observations are increasingly rare as economies are approaching the 21st century, and current implied r-g trends are *above* realized values. That is despite the lack of fully definitive econometric proof – mainly on the basis of high

[22]Reconstructions of housing returns are discussed in Schmelzing (2024) who presents multi-century series for these, with accompanying mortgage yields.

volatility around any post-1914 trendline, an observation that also holds for the "risk-adjusted" r-g data analyzed here.

Figure 9: Duration-matched "risk-adjusted" French r-g, 1909-2020.



Notes: Figure displays the weighted return-output differential for France, using two different share compositions between "risky" and "safe" capital stock components. The dividend yield data via le Bris et al. (2020) is used for the "risky r" component, and French government interest rates are used for the "safe r" component. French per capita *g* is duration-matched on a 10-year average basis. Over 1947-2020, data is using "JST" for French dividend yields, and Schmelzing (2025) for government interest rates.

6 Conclusion

Recent years have seen major advances in long-run reconstructions of both "r" and "g" trends, which allow new perspectives on fundamental debates on debt sustainability dynamics in the 21st century. This allows the full exploitation of the statistical power such samples provide, and the potential revisions of stylized empirical and econometric facts. This paper has provided such a first long-sample investigation of advanced economy r-g differentials, investigated its econometric properties, and identified major turning points.

Drawing on a nuanced reconstruction of long horizon duration-matched r-g spreads for key Western economies, we found overwhelming evidence for trend stationarity, and relatively short half lives, which make the findings economically relevant. We related the secular fluctuations in r-g to the historical record, finding strong plausibility in fiscal histories for general trends, local peaks

Table 6: ADF-GLS / Elliot, Rothenberg and Stock Unit Root Test, risk-adjusted r-g bases					
Series	Time-period	test statistic	critical values		
		(t-value for α_1)	1%	5%	10%
"Risk-adjusted" r-g, 50:50	1441-2020	-4.5154	-3.48	-2.89	-2.57
"Risk-adjusted" r-g, 20:80	1441-2020	-4.8984	-3.48	-2.89	-2.57

Notes: The table reports the results of the ADF-GLS test as per Elliott et al. (1996). All r-g series refer to aggregate measures of g. For a lag of 2 time periods with a trend and intercept. H_0 = series is non-stationary. H_A = Series is stationary. For t-value $<$ critical value, series is stationary. Model: $y_t = \alpha_1 y_{t-1} + \alpha_2 \Delta y_{t-1} + \alpha_3 \Delta y_{t-2} + \epsilon_t$. The two series refer to alternative weightings of the "risky" and "safe" components of the overall capital return (r) figure: the first series weights these two arithmetically in the overall r, while the second (80:20) assigns a 20% weight to the risky component, the remainder being the "safe" government interest rate component.

and temporary reversals in the newly reconstructed r-g series – such as during the key 1564-1729 era, which was beset by sharply worsening dynamics in debt sustainability, fiscal overstretch in the safe asset providers, and escalating financial repression – a period, we showed, that could have some precedents to our own.

Perhaps most importantly, we find that the *sign* of the r-g differential reverses during this period, across all time series reconstructed: while a centuries-long downward trend in r-g persisted prior to World War I, it came to an end in the interwar period of the 20th century. Much – including aggregate output growth evolution – suggests that from the 1930s, r-g differentials across advanced economies appear to be secularly rising – though there is high volatility since, which necessitates caution on the existence of a trend. If anything, in the full context of our trend and intercept evidence, it appears that benchmark base case projections of r-g over the coming decades remain clearly biased to the downside.[23]

These results suggest that – while it can be confirmed from a multi-century perspective that negative r-g spells for marginal rates are indeed becoming more frequent, as posited by Mauro and Zhou (2021) – negative r-g spells could mainly be a function of the interwar era. But since the 1945, the trend appears to be reversing and negative r-g spells have actually become less frequent: on some bases, it even appears that the international financial system is returning to a positive interest rate-output differential, a state consistent with economic theory. These secular trends unfold despite the temporary reprieve that may have been gained through a combination of post-war growth recovery, financial repression, and primary surpluses in the 1950s and 1960s. Overall, the analyses stand in sharp contrast with the vast majority of public finance literature, and points towards a potentially much more adverse legacy of the interwar years than so far acknowledged.

In the second part of this paper, we hypothesized about the drivers of this historical turnaround in the interwar period. We showed that aggregate real growth rates in advanced economies appear to have inflected and trended lower since this period, therefore mechanically underpinning higher r-g

[23]For instance, even the Congressional Budget Office in its pessimistic projections for U.S. r-g (CBO, 2024) suggests a moderately positive r-g environment in the run-up to the year 2054 – one that remains well below the linear historical slope for the U.S. since our most recent trend break in 1932.

rates. This may sounds counter-intuitive against literature that pinpoints productivity slowdowns to the 1970s (Gordon, 2016) and has generally painted a more benign narrative of rising real (per capita) income: for debt sustainability analyses, however, aggregate real growth rates are the imperative basis to analyze. We argued that the public sector underwent a major qualitative change during this exact period around 1930, marked most notably by the widespread introduction of welfare state policies, which underpinned a surge in expenditures relative to output across all advanced economies, as well as a surge in non-interest non-defense spending as a percentage of all public expenditures.

Meanwhile, in an extension that presented plausible approximations of "risk-adjusted" r-g series, we argued that on these measures, too, structural trends – at the center of recent contributions emphasizing *r-m* as in Reis (2021) or Barro (2023) – point towards the same general observations and conclusions compared to traditional r-g concepts based on "safe" government interest rates, as used in Diamond (1965) and elsewhere. These results echo the results from the government interest rate r-g series, and underscore that while there may have been structural trends towards "dynamic inefficiency" in advanced economies prior to the first third of the twentieth century, since then differentials between risk-adjusted returns to investments and output growth are growing – with the frequency of negative r-g observations actually *decreasing* since.

References

Abel, Andrew B., N. Gregory Mankiw, Lawrence H. Summers, and Richard J. Zeckhauser. 1989. Assessing dynamic efficiency: Theory and evidence. *Review of Economic Studies* 56(1): 1–19.

Acalin, Julien, and Lawrence Ball. 2023. Did the U.S. really grow its way out of its WWII debt? *Working Paper* .

Afrouzi, Hassan, Marina Halac, Kenneth S. Rogoff, and Pierre Yared. 2024. Changing Central Bank Pressures and Inflation. *NBER Working Paper* (32308).

Antolin-Diaz, Juan, Thomas Drechsel, and Ivan Petrella. 2017. Tracking the Slowdown in Long-Run GDP growth. *Review of Economics and Statistics* 99(2): 343–356.

Bai, Jushan, and Pierre Perron. 1998. Estimating and Testing Linear Models with Multiple Structural Breaks. *Econometrica* 66(1): 47–78.

Barro, Robert J. 2023. r Minus g. *Review of Economic Dynamics* 48: 1–17.

Blanchard, Olivier. 2019. Public Debt and Low Interest Rates. *American Economic Review* 109(4): 1197–1229.

Blanchard, Olivier. 2022. *Fiscal Policy under Low Interest Rates*. MIT Press.

Bohn, Henning. 1995. The Sustainability of Budget Deficits in a Stochastic Economy. *Journal of Money, Credit, and Banking* 27(1): 257–271.

Bohn, Henning. 1998. The Behavior of U.S. debt and deficits. *Quarterly Journal of Economics* 113(3): 949–963.

Bolt, Jutta, and Jan-Luiten van Zanden. 2024. Maddison-style estimates of the evolution of the world economy: A new 2023 update. *Journal of Economic Surveys* forthcoming.

Bonney, Richard. 1981. *The King's Debts. Finance and Politics in France, 1589-1661*. Clarendon Press, Oxford.

Boyajian, James Charles. 1978. *The Portuguese Bankers and the International Payments Mechanism*. University of California Berkeley, Dissertation.

Broadberry, Stephen, Bruce Campbell, Alexander Klein, Mark Overton, and Bas van Leeuwen. 2015. *British economic growth: 1270-1870*. Cambridge University Press.

Cao, Yongquan, Vitor Gaspar, and Adrian Peralta-Alva. 2024. Costly Increases in Public Debt when r < g. *IMF Working Paper* .

Casoni, Filippo. 1800. *Annali della Repubblica di Genova del secolo decimo settimo – Vol. V*. Stamperia Casamara.

CBO. 2024. The Budget and Economic Outlook: 2024 to 2034 .

Chow, Gregory C. 1960. Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica* 28(3): 591–605.

Cova, Alberto. 1972. *Il Banco di S.Ambrogio nell'economia milanese dei secoli XVII e XVIII.* A. Giuffre.

Dewald, Jonathan. 1980. *Formation of a Provincial Nobility: The Magistrates of the Parlement of Rouen*. Princeton Unviversity Press.

Diamond, Peter A. 1965. National Debt in a Neoclassical Growth Model. *American Economic Review* 55(5): 1126–1150.

Dietz, Frederick. 1921. *English Government Finance, 1485-1558*. University of Illinois Publishing.

Dimsdale, Nicholas, and Ryland Thomas. 2016. A Millennium of Macroeconomic Data, Version 3.1. *Bank of England Data Repository* .

Dincecco, Mark. 2011. *Political transformations and public finances: Europe, 1650–1913*. Cambridge Unversity Press.

Dupaquier, Jacques. 1988. *Histoire de la population française, 4 vols.* Presses Universitaires de France.

Elliott, Graham, Thomas J. Rothenberg, and James H. Stock. 1996. Efficient Tests for an Autoregressive Unit Root . *Econometrica* 64(4): 813–836.

Elliott, John. 1986. *The Count-Duke of Olivares. Statesman in the Age of Decline*.

Ellison, Martin, and Andrew Scott. 2020. Managing the UK National Debt 1694–2018 . *American Economic Journal: Macro* 12(3): 227–257.

Elmendorf, Douglas, and Gregory Mankiw. 1999. Government Debt. In *Handbook of Macroeconomics, Volume 1C*, edited by Taylor, John B., and Michael Woodford, 1615–1669. North-Holland.

Forster, Robert. 1971. *The Nobility of Toulouse in the Eighteenth Century. A Social and Economic Study.* Johns Hopkins University Press.

Francese, Maura, and Angelo Pace. 2008. Italian Public debt since national unification. A reconstruction of the time series. *Banca d'Italia Occasional Papers (Questioni di economia e finanza)* (31).

Fraser, Derek. 2017. *The Evolution of the British Welfare State: A History of Social Policy since 1801. Fifth Edition.* Bloomsbury.

Funke, Manuel, Moritz Schularick, and Christoph Trebesch. 2023. Populist leaders and the economy. *American Economic Review* 113(12): 3249–3288.

Giesecke, Oliver, and Josh Rauh. 2023. Trends in State and Local Pension Funds. *Annual Review of Economics* 15: 221–238.

Goldsmith, Raymond W. 1985. *Comparative National Balance Sheets: A Study of Twenty Countries, 1688-1978*. University of Chicago Press.

Gordon, Robert. 2016. *The Rise and Fall of American Growth*. Princeton University Press.

Hall, George, and Thomas Sargent. 2011. Interest Rate Risk and Other Determinants of Post-WWII US Government Debt/GDP Dynamics. *AEJ: Macro* 3: 192–214.

Hoffmann, Walther G. 1965. *Das Wachstum der Deutschen Wirtschaft seit der Mitte des 19. Jahrhunderts*. Springer Verlag.

Jorda, Oscar, Moritz Schularick, and Alan M. Taylor. 2017. Macrofinancial History and the New Business Cycle Facts. In *NBER Macroeconomics Annual 2016*, edited by Eichenbaum, Martin, and Jonathan A. Parker. University of Chicago Press.

Jordà, Oscar, Katharina Knoll, Dmitry Kuvshinov, Moritz Schularick, and Alan M. Taylor. 2019. The Rate of Return on Everything, 1870-2015. *Quarterly Journal of Economics* 134(3): 1225–1298.

Kaphahn, Fritz. 1912. Der Zusammenbruch der deutschen Kreditwirtschaft im XVII Jahrhundert und der Dreissigjaehrige Krieg. *Deutsche Geschichtsblaetter* XIII(6-7): 139–162.

King, Mervyn, and David Low. 2014. Measuring the"world"real interest rate. *NBER Working Paper 19887* .

Kirk, Thomas. 2005. *Genoa and the Sea: Policy and Power in an Early Modern Maritime Republic, 1559–1684*. Johns Hopkins University Press.

Korevaar, Matthijs. 2023. Reaching for yield and the housing market: Evidence from 18th-century Amsterdam. *Journal of Financial Economics* 148: 273–296.

Krug, Leopold. 1861. *Geschichte der Preussischen Staatsschulden*. Eduard trewendt.

le Bris, David, William Goetzmann, and Sebastien Pouget. 2020. The present value relation over six centuries: The case of the Bazacle company. *Journal of Financial Economics* 132(1): 248–265.

Lian, Weicheng, Andrea Presbitero, and Ursula Wiriadinata. 2020. Public debt and R-G at risk. *IMF Working Paper* (20-137).

Mankiw, Greg. 2022. Government Debt and Capital Accumulation in an Era of Low Interest Rates. *NBER Working Paper 30024* .

Mauro, Paulo, and Jing Zhou. 2021. r g < 0: Can We Sleep More Soundly? *IMF Economic Review* 69: 197–229.

Müller, Ulrich, and Mark Watson. 2018. Long-run Covariability. *Econometrica* 86(3): 775–804.

Münkler, Herfried. 2017. *Der Dreißigjährige Krieg: Europäische Katastrophe, deutsches Trauma 1618–1648*. Rowohlt.

Obstfeld, Maurice. 2023. Natural and Neutral Real Interest Rates: Past and Future. *NBER Working Paper 31949* .

Peden, G.C. 2000. *The Treasury and British Public Policy, 1906–1959*. Oxford University Press.

Perron, Pierre, and Tomoyoshi Yabu. 2009. Testing for Shifts in Trend With an Integrated or Stationary Noise Component. *Journal of Business Economic Statistics* 27(3): 369–396.

Pfister, Ulrich. 2022. Economic growth in Germany, 1500–1850. *Journal of Economic History* 82(4): 1071–1107.

Phelps, Edmund. 1961. The golden rule of accumulation: a fable for growthmen. *American Economic Review* 51(4): 638–643.

Piketty, Thomas. 2014. *Capital in the Twenty-First Century*. Harvard University Press.

Rahlf, Thomas. 2016. The German Time Series Dataset 1834-2012. *Jahrbücher für Nationalökonomie und Statistik / Journal of Economics and Statistics* 236(1): 1–36.

Rauscher, Peter. 2004. *Zwischen Ständen und Gläubigern: die kaiserlichen Finanzen unter Ferdinand I. und Maximilian II.(1556-1576)*. Oldenbourg.

Reinhart, Carmen, and Belen Sbrancia. 2015. The liquidation of government debt. *Economic Policy* 30(82): 291–333.

Reis, Ricardo. 2021. The constraint on public debt when r< g but g< m. *LSE Working Paper* .

Reis, Ricardo. 2022. Which r-star, public bonds or private investment? Measurement and policy implications. *LSE Working Paper* .

Ridolfi, Leonardo. 2019. Six Centuries of Real Wages in France from Louis IX to Napoleon III: 1250–1860. *Journal of Economic History* 79(3): 589–627.

Ridolfi, Leonardo, and Alessandro Nuvolari. 2021. L'histoire immobile? A reappraisal of French economic growth using the demand-side approach, 1280–1850. *Explorations in Economic History* 25(3): 402–428.

Rogoff, Kenneth S. 1985. The optimal degree of commitment to an intermediate monetary target. *Quarterly Journal of Economics* 100(4): 1169–1189.

Rogoff, Kenneth S., Barbara Rossi, and Paul Schmelzing. 2024a. Long-Run Trends in Long-Maturity Real Rates, 1311-2022. *American Economic Review* 114(8): 2271–2307.

Rogoff, Kenneth S., Barbara Rossi, and Paul Schmelzing. 2024b. Rethinking short-term real interest rates and term spreads using very long-run data. *NBER Working Paper* (33079).

Schmelzing, Paul. 2022. Eight centuries of global real rates, R-G, and the 'suprasecular' decline, 1311-2021. *mimeo* .

Schmelzing, Paul. 2024. Housing and the emergence of the safe asset, 1465-2024. *Working Paper* .

Schmelzing, Paul. 2025. *Eight Centuries of Global Real Rates, 1311-2024*. Yale University Press, forthcoming.

Socolofsky, Homer E., and Allan B. Spetter. 1987. *The Presidency of Benjamin Harrison*. The University of Kansas Press.

Steinsson, Jón. 2008. The Dynamic Behavior of the Real Exchange Rate in Sticky Price Models. *American Economic Review* 98: 519–33.

Stradling, R.A. 1990. Prelude to Disaster; the Precipitation of the War of the Mantuan Succession, 1627-29. *The Historical Journal* 33(4): 769–785.

Sutch, Richard. 2006. National Income and Product. In *Historical Statistics of the United States: Millennial Edition*, edited by Carter, Susan B., and Richard Sutch. Cambridge University Press.

van Bavel, Bas, and Auke Rijpma. 2016. How important were formalized charity and social spending before the rise of the welfare state? A long-run analysis of selected western European cases, 1400–1850. *Economic History Review* 61(1): 159–187.

Wallis, John Joseph. 2006. Chapter Ea – Government Finance and Employment. In *Historical Statistics of the U.S. – Millennium Edition*. Cambridge University Press.

Zivot, Eric, and Donald W.K. Andrews. 1992. Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis. *Journal of Business Economic Statistics* 10(3): 251–270.

APPENDIX

1. r and g data

Here we describe the output growth data used to construct the headline r-g series. We draw on new output constructions from economic historians over recent years. Unless noted otherwise, the sources are identical for aggregate and per capita g bases.

- for the U.K., we rely on Dimsdale and Thomas (2016), with their figures being based on the underlying Broadberry et al. (2015), beginning in 1525. GDP growth data after 2016 is based on Office for National Statistics, ONS (2024).
- for France, up to 1860 we rely on Ridolfi and Nuvolari (2021), and the population figures in Dupaquier (1988) as applied by Ridolfi (2019). After 1860, we use the GDP growth data in "JST" (Jorda et al., 2017).
- for Germany, up to 1850 we rely on Pfister (2022), who constructs aggregate g figures starting in 1500. His population figures are interpolated wherever necessary. German growth figures even prior to 1805 de facto refer to Germany in current (2024) borders. After 1850, we rely on "JST".
- for the U.S., we rely on Sutch (2006), starting in 1801. Afterwards, we use Bureau of Commerce data, as retrieved via FRED (2024).

Except for Germany, these polities have seen relatively limited political and geographical changes over long periods of time. For Germany specifically, it should be noted that Pfister (2022) uses a geographically consistent definition of Germany that aligns closely with the current borders of the Federal Republic of Germany.

For interest rates, we rely on marginal nominal long-maturity interest rates, as constructed and described by Schmelzing (2022). All underlying sources are therein and in Schmelzing (2025). Importantly, for the U.K. we also check results for the independently constructed data in Dimsdale and Thomas (2016).

On the dividend yield side discussed in section 3.2, we rely on annual observations of the Bazacle company over 1532-1946 as reported in le Bris et al. (2020). For the series incorporating 1947-2020 dividend yield data for France, we rely on Jordà et al. (2019). See section 7 below for a visual representation.

2. Assuming no deterministic trend, results

Our main results assumed a deterministic trend. Here we test the key econometric properties assuming no such deterministic trend, using our standard aggregate r-g duration-matched basis throughout. Table A.1 begins with ADF-GLS assuming no deterministic trend.

Here, we observe that for the majority of country series, stationarity is still confirmed at the 1% level (France, U.S., and Germany) – except for the U.K., which fails to reach 10% significance.

[Table A.1]					
ADF-GLS / Elliot, Rothenberg, and Stock Unit Root Test					
Series	Time-period	Test Statistic	Critical Values		
		(t-value for α_1)	1%	5%	10%
U.K. r-g	1525-2022	-0.3353	-2.57	-1.94	-1.62
Germany r-g	1501-2022	-3.0809	-2.57	-1.94	-1.62
France r-g	1331-2022	-3.0537	-2.57	-1.94	-1.62
U.S. r-g	1801-2022	-3.3449	-2.57	-1.94	-1.62
Global r-g	1501-2022	-1.5898	-2.57	-1.94	-1.62

Test conducted with a lag of 4 periods including intercept and trend. H_0: Series is non-stationary. H_A: Series is stationary. Using aggregate duration-matched r-g throughout.

For test statistic $<$ critical value, the series is stationary.

Model: $y_t = \beta_0 + \beta_1 y_{t-1} + \sum_{i=1}^{4} \alpha_i \Delta y_{t-i} + \epsilon_t$

3. Chow test results

Tables A.2 and A.3 below displays Chow test results, as per the methodology of Chow (1960). We jointly test trend and mean breaks based on existing literature that has posited strong evidence for these four inflection points in particular. See in particular Rogoff et al. (2024a) for further discussion.

We report details for the Global r-g and the U.K. series below. For the global, our results can be summarized as follows: we cannot reject the null (no break) for 1914 in either trend or mean. We reject the null for all other break points. For the U.K., we cannot reject the null (no break) for 1557 and 1694. We reject the null at 1914, 1981 in both trend and mean.

These results appear to suggest some plausibility for 1914 and 1981 as relevant breaks – which prompts us to retain caution when it comes to definitively characterizing the post-1914 trajectory of r-g as described in the main paper. Despite other structural break tests rejecting the idea of a 1981 break, we cannot unanimously rule out that the direction and intensity of the r-g trend may have changed in the early 1980s. Though not reported separately here, the Chow results are consistent for the other country series France, U.S., and Germany.

[Table A.2 – Chow results for Global r-g]

	Coefficient	Std. error	t	p-value
trendbreak_1557	.2339757	.0924606	2.53	0.012
trendbreak_1694	.1507752	.0717608	2.10	0.036
trendbreak_1914	.3408038	.2339087	1.46	0.146
trendbreak_1981	-.0311331	.0101898	-3.06	0.002
meanbreak_1557	-.0816334	.0419701	-1.95	0.052
meanbreak_1694	-.098457	.0435489	-2.26	0.024
meanbreak_1914	-.3824712	.2733333	-1.40	0.162
meanbreak_1981	-.0662549	.0126423	-5.24	0.000
trend	-.4828772	.2195043	-2.20	0.028
Intercept	.6298874	.2797905	2.25	0.025

Table A.3: The model includes a constant and a deterministic trend (as a fraction of the total sample size). The standard errors are based on Newey and West's (1987) HAC estimator where the lag length is chosen according to Lazarus et al. (2018). The critical value of the (absolute value of the) t-statistic is 2.2847 at the 5 percent significance level and 3.1728 at the 1 percent. The coefficient associated with "trendbreak 1981" denotes the difference of the estimated trend coefficients before and after a break in 1981 (allowing for breaks in all the other potential break dates); hence, the associated t-statistic is the Chow test for the absence of a structural break in trend in 1981. Similarly, "meanbreak 1981" refers to a break in the mean in 1981.

[Table A.3 – Chow results for U.K. r-g]

	Coefficient	Std. error	t	p-value
trendbreak_1557	-.7115069	.5114966	-1.39	0.165
trendbreak_1694	-.0427865	.0463662	-0.92	0.357
trendbreak_1914	.5702282	.1507531	3.78	0.000
trendbreak_1981	-.012286	.0044703	-2.75	0.006
meanbreak_1557	.3102941	.2174704	1.43	0.154
meanbreak_1694	.0333896	.0343622	0.97	0.332
meanbreak_1914	-.7049071	.1873275	-3.76	0.000
meanbreak_1981	-.0859541	.0146446	-5.87	0.000
trend	-.6870603	.1504248	-4.57	0.000
Intercept	.9331398	.2013591	4.63	0.000

Table A.4: The model includes a constant and a deterministic trend (as a fraction of the total sample size). The standard errors are based on Newey and West's (1987) HAC estimator where the lag length is chosen according to Lazarus et al. (2018). The critical value of the (absolute value of the) t-statistic is 2.2847 at the 5 percent significance level and 3.1728 at the 1 percent. The coefficient associated with "trendbreak 1981" denotes the difference of the estimated trend coefficients before and after a break in 1981 (allowing for breaks in all the other potential break dates); hence, the associated t-statistic is the Chow test for the absence of a structural break in trend in 1981. Similarly, "meanbreak 1981" refers to a break in the mean in 1981.

4. Half-lives

Finally, we test half-lives of r-g series. We follow the econometric approach in Rogoff et al. (2024a). If r-g trend stationarity exhibits a high degree of persistence (large half-lives of, say, 20 or 30 years), then –

[Table A.4]				
Half-Live Tests				
Series	Time-period	Half-Life	90% Confidence Interval	
			5%	95%
U.K. Aggregate r-g	1525-2023	4.82	3.40	7.16
Germany Aggregate r-g	1501-2023	16.67	11.05	27.76
France Aggregate r-g	1331-2023	9.84	6.78	14.49
U.S. Aggregate r-g	1801-2023	9.87	5.66	20.67
Global Aggregate r-g	1501-2023	13.64	8.16	25.89

Test conducted with an AR(2) model.

Table A.4: the half-life tests as outlined above except here the series are broken into two by the Perron and Yabu (2009) test with the intercept and trend model. All series refer to aggregate g bases. Note: The table reports median unbiased estimates and 90% confidence intervals of the half life (h) based on Steinsson (2008). The regression is: $y_t = \mu_0 + \mu_1 t + \alpha y_{t-1} + \sum_{j=1}^{p} \gamma_j \Delta y_{t-j} + \varepsilon_t$ where α is the largest root. The row with the country name reports the full sample estimates, while the rows with sub-samples report the sub-sample estimates.

while being of econometric importance – the economic relevance of our results would be less applicable for practical purposes.

However, the results in Table A.4 suggest that much shorter half-lives are at play, with early modern half-lives generally standing at between 1-4 years, and full sample values between 5-17 years. In the 20th century, half-lives are then higher, broadly ranging from 5-12 years across series. These ranges are generally consistent with earlier ranges for early modern variables as discussed in Rogoff et al. (2024a) for the case of long-maturity real interest rates, and elsewhere. Mean-reversions in r-g following shocks can take one or two decades, and therefore analyzing just multi-year windows around the GFC or COVID may mask such dynamics – but evidently the adjustments are not "generational" processes, which underscores the relevance of incorporating their stationarity features.

5. Perron and Yabu (2009) – additional results

Here we separately report the results for the breaks in intercept and trend using the methodology of Perron and Yabu (2009). First, table A.6 displays the results for the break in trend only, allowing for either zero or one break(s). We observe that five of the seven duration-matched r-g series display no break at the 10% significance: German r-g, U.S. r-g, Global r-g, and the risk-adjusted variations. One additional series displays one break at the 5% significance level, the French r-g series; and one series displays a trend break at the 1% significance level, U.K.

Next, table A.6 displays the results for breaks in intercept only, again allowing for zero or one break(s). Here, two series display intercept breaks at the 5% significance level, the Global and U.K. series. The remaining five series all display breaks at the 1% significance level.

All breaks in intercept are clustered around 1506-1977. Again, the U.S. appears to be an outlier in showing a

Table A.5: Testing for Shifts in trend with an Integrated or Stationary Noise Component					
Series	BreakDate	test statistic	critical values		
			10%	5%	1%
U.K. r-g	1945	5.7377	1.1300	1.6700	3.0600
GER r-g	1942	-0.2706	1.1300	1.6700	3.0600
FRA r-g	1787	1.6765	1.1300	1.6700	3.0600
U.S. r-g	1944	-0.2537	1.1300	1.6700	3.0600
Global r-g	1942	1.0433	1.1300	1.6700	3.0600
Risk-adjusted r-g, 50:50	1532	-0.11	1.1300	1.6700	3.0600
Risk-adjusted r-g, 20:80	1933	-0.24	1.1300	1.6700	3.0600

Table A.6: Model: $y_t = \beta_{j0} + \beta_{j1} t + \varepsilon_t$, Where j=0,1 is the number of breaks (trend). Trimming is set to 15%.

post-1945 break in intercept.

6. Zivot-Andrews

Here we report results for the test in Zivot and Andrews (1992), which tests for stationarity in the presence of a structural break. We observe that overwhelmingly, our results remain intact at the 1%, except for the shortest time series across the sample, that of the U.S., where stationarity is rejected.

Table A.6: Testing for Shifts in intercept with an Integrated or Stationary Noise Component					
Series	BreakDate	test statistic	critical values		
			10%	5%	1%
UK r-g (nominal)	1670	3.0230	1.2600	1.7400	3.1200
GER r-g (nominal)	1844	8.7229	1.2600	1.7400	3.1200
FRA r-g (nominal)	1901	16.9850	1.2600	1.7400	3.1200
US r-g (nominal)	1977	19.6449	1.2600	1.7400	3.1200
Global r-g (nominal)	1942	2.8728	1.2600	1.7400	3.1200
Risk-adjusted r-g, 50:50	1907	7.26	1.2600	1.7400	3.1200
Risk-adjusted r-g, 20:80	1903	11.81	1.2600	1.7400	3.1200

Model: $y_t = \beta_{j0} + \beta_1 t + \varepsilon_t$, Where j=0,1 is the number of breaks (intercept)
Trimming is set to 15%.

Table A.7: Zivot-Andrews test		
Series	N. breaks	Estimated break dates
GER r-g	-6.675	1906
FRA r-g	-6.853	1928
U.K. r-g	-6.215	1972
U.S. r-g	-4.879	1978
Global r-g	-5.443	1930

Table A.10: The table reports the Zivot and Andrews (1992) unit root test statistic, which allows for a break in both the mean and the trend under the alternative. The critical values at the 1, 5 and 10 percent significance levels are the following for all observations: -5.57 (1%); -5.08 (5%); -4.82 (10%). The trimming parameter is 0.10. The test rejects when the test statistic is negative and larger (in absolute value) than the critical value. When the test rejects the unit root, the column labeled "break date" reports the estimated break date.

7. Duration-matched "risky" French r-g, 1532-1946.

While long-run empirics for this basis are generally much rarer, le Bris et al. (2020) recently reconstructed the financial performance of one of the the longest corporate entities in continuous existence, the privately held French Bazacle company, a listed milling company ultimately dissolved in 1946. Over the period of 1532-1946, continuous dividend yield data was compiled in this case – the advantage being that we measure a consistent unit of analysis over long periods of time, avoiding survivorship and other equity measurement biases. Visually, the series trends down, and ADF-GLS results using the 1532-1946 period confirm trend stationarity at the 1% level.[24]

Matching these dividend yield observations with our French g data might represent a closer approximation of French "risky" r-g trends.[25] Indeed, while volatile the series also trends downwards over time. On average, this "risky" r-g series for France displays a downward trend of -2.3bps p.a., or -2.3% per century. This compares to -3.0bps p.a. for the non "risk-adjusted" r-g French basis in Figure 3 – the long-run equity return series itself shows a positive spread over the government interest rate as is generally assumed, including by contributions on recent trends (Reis, 2021; Barro, 2023). Risk-adjusted r-g shows sharp volatility towards the late 17th century, during the Napoleonic era, and during the two World Wars.

But on balance, as with government interest rates, it also at first appears that the frequency of negative return-output differentials is increasing over time: negative periods are getting longer – the 1906-1922 spell represents the longest on record – and negative annual observations are more frequent in absolute terms, with only 14% of annual observations being negative over the first century of the series, but 27% of years during the century prior to 1914. However, once again, as with the safe r–g series, the trend since 1914 appears to have, if anything, gone into reverse.

In Figure A.1, we now display the construction of the "risky" French r-g series, using the dividend yield data of le Bris et al. (2020) as a "risky" capital stock element – we recall that their dataset has the advantage of measuring a single company in continuous existence over this period, the Bazacle company, which therefore eliminates questions of survivorship biases. For duration-matched nominal French GDP growth, we use the new data of Ridolfi and Nuvolari (2021) to 1860 here, and afterwards the data in "JST" (Jorda et al., 2017) (the years 1861-1870 are linearly interpolated on the GDP growth side).

8. Monte-Carlo exercise

To investigate the use of long versus small sample sizes in testing for unit roots in our data, we performed a small Monte Carlo simulation exercise – with details per tables A 8.1-8.2. We generated 5000 time series that have the same features as our (stationary) global r-g series (such as the mean, the deterministic trend and the standard error) as estimated in the regressions underlying Tables 8.1-8.2, calibrated for the global r-g series.

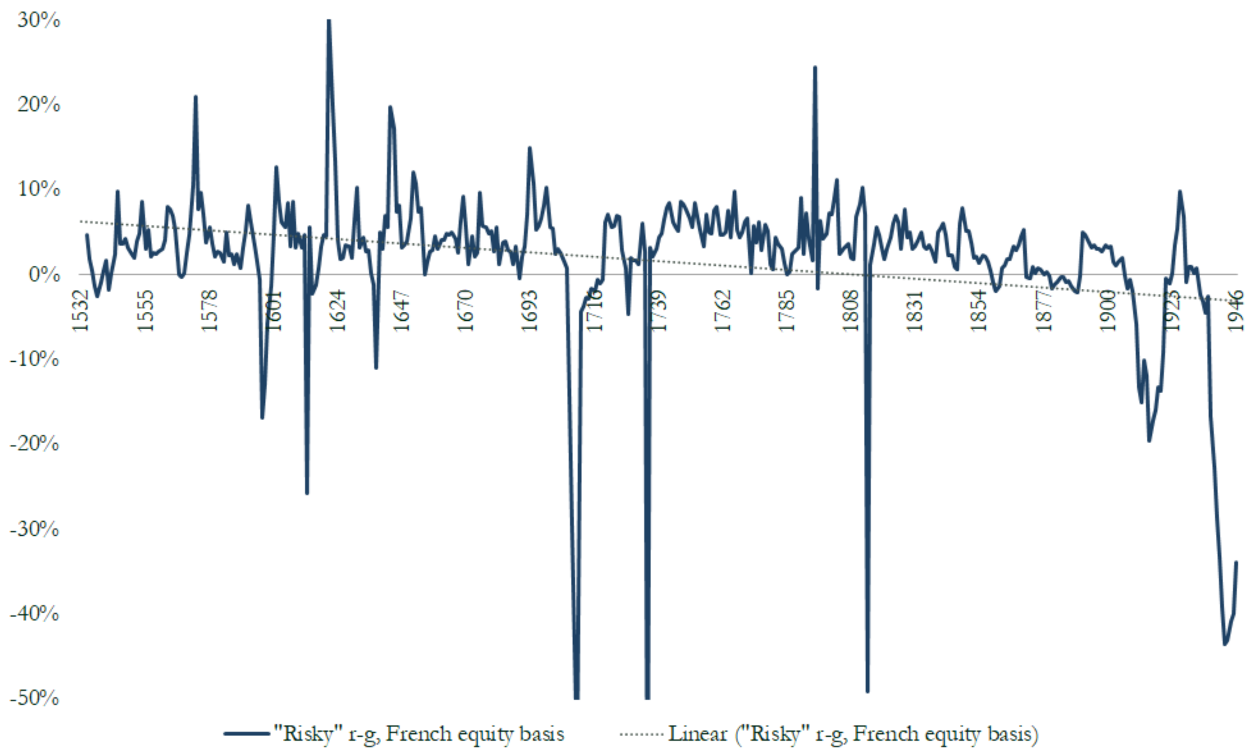
For each generated series, we calculated:

- (a) the ADF-GLS test statistic for a sample of 705 observations (the same size as our full sample);

[24] le Bris et al. (2020) start measuring corporate data for Bazacle starting with the company's establishment in 1372 – however, until 1532 there are frequent gaps in the dividend yield observations, so that we begin series at this later date. Occasional data gaps after 1532 are linearly interpolated.

[25] The Figure is displayed in the Appendix via Figure A.1, we display the resulting series, also using a 10-year average growth basis for the g.

Figure A.1: Duration-matched "risky" French r-g, 1532-1946.



Notes: Figure displays the spread between nominal dividend yields of the Bazacle company as measured in le Bris et al. (2020), and 10-year averaged nominal French GDP growth rate, per Ridolfi and Nuvolari (2021) to 1860, and "JST" from 1870 (Jorda et al., 2017).

- (b) the ADF-GLS test statistic for the sub-sample including only 150 observations;
- (c) the ADF-GLS test statistic for the sub-sample including 100 observations;
- (d) the ADF-GLS test statistic for the sub-sample including 75 observations;
- (e) the ADF-GLS test statistic for the sub-sample including 50 observations.

We then report the number of times the ADF-GLS test rejects a unit root at the 5 percent significance level across Monte Carlo simulations.

9. Growth series – additional tests

In tables 9.1-9.2, we present evidence of trend stationarity for our "new generation" output growth series, first confirming strongly trend stationarity on the 1% significance level across both real aggregate and real per capita bases (table 9.1). Afterwards, we record structural break dates using the methodology of Bai and Perron (1998), confirming the general pattern of an interwar cluster of such breaks, again for multiple bases of output growth series (table 9.2).

Table A.8.1: Monte-Carlo exercise

series	n=705	n=150	n=100	n=75	n=50	nLag (ADF)	half life
France	0.9904	0.2642	0.1674	0.1496	0.1194	1	5.55
U.K.	0.9970	0.3600	0.2122	0.1568	0.1242	2	2.11
U.S.	0.9994	0.5504	0.3262	0.2302	0.1920	1	9.7054
Germany	0.9926	0.2836	0.1808	0.1520	0.1354	1	16.4846
Global	0.9832	0.2182	0.1366	0.1150	0.0998	2	7.95

Table A.8.1: Rejection frequency of ADF test with different sample sizes. Notes. To construct this table, we first estimate an AR model for each time series, based on the MAIC selected number of lag in the previous ADF results. Half life is estimated for the models. Then, we use Monte-Carlo methods to generate a sample of 5000 periods for each model, and examine the power of the ADF test within subsamples of different sizes. In columns two to six, we present the frequency of rejection.

Table A.8.2: Monte-Carlo exercise, global series

n=705, trend	n=705, no trend	n=100, trend	n=100, no trend	trend value
0.97	0.42	0.15	0.28	-0.000008
0.97	0.21	0.14	0.26	-0.000010
0.97	0.08	0.14	0.26	-0.000012
0.97	0.02	0.14	0.25	-0.000014
0.97	0.00	0.14	0.24	-0.000016

Table A.8.2: Rejection frequency of ADF test specifications with and without trend. Notes. True model is estimated based on the Global series, with estimated trend value -1.4×10^{-5}. We show that in the large sample (n=705), the correctly specified model can reject the null with frequency 0.97, close to the nominal size of the test. Once sample size drops to n=100, rejection frequency decreases substantially. On the other hand, the mis-specified model without trend cannot reject the null even in large samples.

Table 9.1

ADF-GLS: Aggregate real output growth series, duration-weighted, deterministic trend					
Series	Time-period	Test Statistic	Critical Values		
		(t-value for α_1)	1%	5%	10%
U.K. output growth	1501-2022	-3.8970	-3.48	-2.89	-2.57
Germany output growth	1501-2022	-5.1314	-3.48	-2.89	-2.57
U.S. output growth	1801-2022	-4.1217	-3.48	-2.89	-2.57
France output growth	1501-2022	-5.0204	-3.48	-2.89	-2.57
U.K. p.c. output growth	1501-2020	-3.2923	-3.48	-2.89	-2.57
GER p.c. output growth	1501-2020	-7.5974	-3.48	-2.89	-2.57
U.S. p.c. output growth	1871-2020	-4.1451	-3.46	-2.93	-2.64
France p.c. output growth	1501-2020	-3.6395	-3.48	-2.89	-2.57

Methodology follows Elliott et al. (1996). Test conducted with a lag of 4 periods including intercept and trend. H_0: Series is non-stationary. H_A: Series is stationary.
For test statistic $<$ critical value, the series is stationary.
Model: $y_t = \beta_0 + \beta_1 t + \beta_2 y_{t-1} + \sum_{i=1}^{4} \alpha_i \Delta y_{t-i} + \epsilon_t$

Table 9.2

Bai-Perron Test Results (Deterministic Trend)		
Series	Breakpoint(s) (obs no.)	Breakpoint(s) (years)
U.K. output growth	219, 306, 431	1719, 1806, 1931
Germany output growth	121, 407	1621, 1907
U.S. output growth	63, 96, 131, 164	1863, 1896, 1931, 1964
France output growth	356, 435	1856, 1935
U.K. p.c. output growth	196, 306, 431	1696, 1806, 1931
Germany p.c. output growth	414	1914
U.S. p.c. output growth	41, 63	1911, 1933
France p.c. output growth	330, 409	1830, 1909

Model: $y_t = \beta_{j0} + \beta_{j1} t + \varepsilon_t$, Where j=1, . . . , 5 is the number of segments
Methodology follows Bai and Perron (1998). Trimming parameter set to 15%. Up to 5 dates could be selected by the model.