



Scoring Six Detrending Methods on Timing, Lead-Lag Relations, and Cycle Periods: An Empirical Study of US and UK Recessions 1977–2020

Knut Lehre Seip¹ · Dan Zhang¹

Accepted: 8 January 2024
© The Author(s) 2024

Abstract

This study evaluates six commonly used detrending methods and discuss how detrending may change the timing of events, the identification of lead-lag relations between GDP and employment, and the identification of cycle periods. The detrending methods examined includes linear detrending, polynomial detrending, the first-order differencing, locally weighted scatterplot smoothing (LOESS), Hodrick–Prescott filter, and the Hamilton filter. We apply the detrending methods to the United States and United Kingdom gross domestic product (GDP) from 1977 to 2020. We find that for the GDP series the first-order differencing score best on all three criteria, however, it also shows more false recessions than the other detrending methods. A linear, a polynomial, and a LOESS trend all scored well. The three methods miss-specified the timing of the recessions with less than one quarter and all three gave results that would comply with stylized facts in macroeconomics. The Hodrick–Prescott (HP) filter and Hamilton filter did not achieve high scores on one or two of the criteria and scored worst on average performance.

Keywords Detrending methods · Forecasting · Recessions · LOESS filter · HP-filter · Hamilton-filter

JEL Classification E32 · E37 · C10 · C82 · G01

1 Introduction

In empirical business-cycle analysis, it is widespread practice to decompose the trending variables, such as real output, into a secular (or trend) component and a stationary component that shows cyclical behavior. Various detrending methods

✉ Dan Zhang
danzhang@oslomet.no

¹ Oslo Metropolitan University: OsloMet – storbyuniversitetet, Oslo, Norway

are used to remove the effects of a trend and to identify potentially important cyclical patterns. However, there is an adage in economics: “One scientist’s time trend may be part of another scientist’s cycle”. Thus, detrending may in a sense be the partitioning of a cyclic series into cyclic components that each represent a distinct mechanism. In addition, some peaks and troughs in a cyclic series may be due to dynamic chaos (Sugihara & May, 1990; Tømte et al., 1998), or they may be due to interactions between two oscillating mechanisms, (Seip & Pley, 2000) and not relate to one single driving mechanism. Thus, with respect to business cycle series, there is no “ground truth”.

Detrending methods may affect the raw time series in several ways. Different methods usually lead to a different trend-cycle decomposition that can be compared to theoretical macroeconomic models (see, e.g., Canova, 1998, 1999). There are two groups of rationales for detrending (i) to get focus on the objective of the study or to disentangle component series that are generated by different processes and (ii) to get the data into a format that allows common statistical methods to be applied correctly to the data.

Within the first group are the study of fluctuations around the trend of a series, for example to examine business cycles or growth cycles. (Canova, 1999). Another objective is to distinguish short term movements in the economy caused by e.g., rapid movements in the stock market, from longer movements caused by e.g., changes in tax policies (Mountford and Uhlig 2009). Stylized facts about economic issues may help setting parameter values for detrending algorithms. One issue is the dating of business cycle turning points, or the length of business cycles or growth cycles (Burnside, 1998; Canova, 1999; Pollock, 2016). Canova (1998) examined the US Gross national product (GNP) series 1955–1973 using the Hodrick Prescott (HP) filter with its parameter λ set to 4 (high frequency variability) and 1600 (business cycle frequency), but other values for λ have also been used. Correct timing is also important for the dating of recessions, e.g., as defined by the National Bureau of Economic Research (NBER). Detrending must not distort the time series so that dates are significantly changed and associations between patterns in the time series and candidate causal events become flawed. A second set of dates, or tie-points, are structural breaks in the economy. A third issue are lead-lag relations between indexes that are used to forecast future development in a macroeconomic variable and their target variable (Christiansen et al., 2014; Seip et al., 2019, Krüger, 2021). To compare two detrended series, one should probably apply the same detrending procedure to both series (Burnside, 1998; Enders., 2010, p 256).

Within the second group are the question related to cointegration of time series to avoid spurious regressions. Differencing the series may help avoid spurious results, but a first differencing will show a peak when the raw series has its steepest slope, and thus shift peaks (and troughs) relative to the raw series. Second, the statistical parameters of ordinary least square (OLS) regression applied to series with descending or increasing trends will characterize the trends, and not variabilities around the trends. Third, cross correlation tests are often applied to series that are candidate cause and effect variables, but since the technique is to shift one series relative to another to see if one get better overlap between the two series by shifting them

(measured by the explained variance, r^2 versus time shifts $t_{-2}, t_{-1}, t, t_{+1}, t_{+2}$..), the series has to be detrended.

Many studies compare detrending methods. For example, Canova (1998) using six different detrending methods, examines the business cycle properties of seven real US macroeconomic time series including the GNP. The results suggests that the HP(1600) filter of Hodrick and Prescott (1980) and the segmented polynomial time function method (his SEGM) are those that come closest in reproducing standard dating and business cycle features. For other series, amplitude and duration of cycles are sensitive to detrending method. Bjørnland (2000) finds for variables like real wage and prices that five different detrending methods with selections of parameters suggested very different cyclical behaviors. In addition, detrending without appropriately adjusting for the structural break in the trend could severely distort the results.

In this study we compare six detrending methods and discuss how detrending may change (i) the timing of events, (ii) the identification of high-resolution lead-lag relations between GDP and employment and (iii) the identification of cycle periods. The detrending methods we examine in this paper include the subtraction of a linear or second order polynomial trend, a segment detrending method (we use LOESS, a locally weighted scatterplot smoothing algorithm), the first order differencing procedure, the HP-filter from Hodrick and Prescott (1997), and the Hamilton-filter developed by Hamilton (2018). Except for the linear and second order polynomial detrending, all methods require some parameters to be determined that depend on characteristics of the time series studied. Lastly, although not a detrending method, we study the effects of stabilizing the variance of the US GDP by taking the logarithm of the data. Characteristics and source information of the alternative detrending methods are summarized in Table 1.

We hypothesize that linear detrending is sufficient for time series that have a persistent upward sloping trend like the US GDP (or alternatively a persistent downward trend). The rationale is that more complex detrending will only give minor changes in the timing of important event which has little improvement of the overall performance.

Our contribution is to examine six commonly use detrending methods and evaluate their performance with respect to the GDP of the United States (US) and the United Kingdom (UK) on three criteria, the timing of recessions, lead-lag relations to employment and cycles in the US and UK GDP time series. We find that the first difference gave the best score on the three criteria, whereas the linear detrending, the second order polynomial detrending, and the LOESS methods all satisfied criteria values for reasonable detrending results. The HP-filter and the Hamilton-filter methods failed on one or two of the three criteria.

Our study differs from other studies in five ways. First, we interpolate the GDP data to monthly data to be able to determine the timing of recessions within a month (the NBER recession dating gives the timing in months). Second, we identify recessions by calculating the slope of a moving seven-month regression applied to both raw and detrended data. Third, we measure the skill of detrending on three criteria: the dating of a proxy to the NBER recessions, high resolution lead-lag relations to employment, and cycle periods. In addition, we interpret the results in a richer

Table 1 Characteristics and source for detrending methods

Method	Acronym	Parameter characteristics	Purpose	Source
1	Ordinary linear regression	LIN $N=1$, none	Simplistic	
2	Polynomial regression	POL $N=2$;	Progressively increasing trend	Canova (1998)
3	LOESS	LOESS f = fraction of series used as moving window; p =polynomial interpolation	Remove multidecadal and decennial trends, $f \approx 0.8$. Remove noise, $f \approx 0.05$, identify cycle periods, $f \approx 0.2-0.4$	Cleveland & Devlin, (1988), Cleveland and Grosse (1991)
4	First difference	DIF “ h ” in $y_{t+h} - y_t$	Avoid spurious correlations	
5	Hodrick Prescott	HP λ = frequency	$\lambda = 48$ low, $\lambda = 1600$, business cycles, quarterly, $\lambda = 10,000$ = trend	Hodrick and Prescott (1997)
6	Hamilton	HAM $p=4$, $h=8$	Business cycle periods	Hamilton (2018)
7	Logarithm	LOG $N=1$, none,	Stabilizes variance	

The table summarizes the characteristics and source for the detrending methods covered in this study

macroeconomic context that normally associated with the term “stylized facts”. We can do this because the lead-lag relations we use are calculated over very short time intervals and over series that are not detrended ($n=3$; $n=9$ allows calculation of confidence interval). Fourth, we use a version of a multiple window spectrum (MWS) method, that is common in climate research for analyzing power spectral densities (PSD) (Johnson, et al., 1996). Last, we use principal component analysis (PCA) to compare trends and cyclic components to reference trends and reference cycles.

The remainder of the paper is organized as follows. The next section presents a literature survey. Descriptions on the data and methods are shown in Sects 3 and 4, respectively. In Sect. 5, we present the results for the detrended series and in Sect. 6 we discuss the results in an economic context. Section 7 concludes.

2 Literature Review

Several detrending methods are studied in the literature. We here outline (i) the methods examined, (ii) the tests that are applied to the detrending results, and (iii) the goodness of the detrending results.

We summarize the literature on detrending methods in business cycle research in Table 2. We have not included studies that attempt to determine the “best” parameters for a specific detrending method, e.g., as Franke & Kukacka (2020) and Ravn & Uhlig (2002) for the HP-filter. Most studies include linear detrending (abbreviated as LIN in the table), as one option. The HP- filter is also used abundantly in macroeconomics, often with its standard parameter $\lambda = 1600$ for quarterly data (Hamilton, 2018). Other detrending methods listed in the table include polynomial functions of time (abbreviated as POL), LOESS filter, first-order differencing (abbreviated as DIF), Beveridge & Nelson (1981) method (abbreviated as BN), frequency domain filtering (abbreviated as FR), unobservable component method (abbreviated as UCM) and Hamilton filter (abbreviated as HAM) using the approach as in Hamilton (2018).

Apart from linear detrending and a second order polynomial function, most methods require judgements on which parameters to choose for the detrending method, (Canova, 1999; p. 130). To assess the goodness of the detrending algorithms the most common tests are comparison with recessions (see, for example, Canova, 1999), examination of volatilities (e.g., Bjørnland, 2000; Park, 1996) and cycle period characteristics like amplitudes, durations, and persistence (Canova, 1999; pp. 142–144). However, there are also comparisons with “stylized facts”, which most often refer to relations that are commonly observed in economics (Brault & Khan, 2020). Bjørnland (2000; p. 381) establish “facts” in terms of correlations and lead-lag relation between macroeconomic series and GDP, and in terms of volatility of the detrended series. Recessions are determined in numerous ways. Canova (1999) gives two definitions, e.g., a peak is defined by two consecutive increases (quarters) followed by a decline. This study defines the beginning of a recession for GDP data that first are normalized to unit standard deviation as a negative regression coefficient over 7 months (about two quarters) that exceeds -1.5 or -2.0 (values

Table 2 A literature survey of detrending studies

Studies	Methods	Pretreatment	Series	Tests	Recommendations
Baxter (1991)	LIN, DIF, HP	Logarithm	GNP, WH, PRO, C, IN, EXP	Volatility, Stylized facts	All methods not robust
Park (1996) USA 1957–1986	BN, HP LIN,	Unit-root test	GDP, US-coincident index	GDP, Simulated data, volatility	No recommendation
Canova (1999) USA 1955–1986	LIN, DIF, FR, BN, UCM, HP, others	Logarithm, Comparing GDP_t to GDP_{t+h} to determine tie-points	GDP	Dating of NBER recessions or DOC	HP, FR
Bjørnland (2000) Norway 1971–1990	BN, HP, UCM, others	logarithm, unit root test	GDP, C, I, EXP, M, PR, W, RWG, UE, CPI, M2	Spectral density, Volatility, correlation with GDP, Stylized facts	No recommendation
Zarnowitz and Ozylidirim (2006) US 1948–2000	LIN, HP, FR	12, 75 months averages to determine tie-points	GDP, US-Coincident index	NBER recessions; peaks troughs in GDP growth	LOESS, FR, HP
Hall et al., (2017, p. 212)	HP, LOESS, others	Logarithm	New Zealand GDP, CPI, others	Stylized facts, volatility, persistence	HP ($\lambda = 1600$)
Hamilton (2018) US 1950–2016	HP, HAM	Logarithm	GDP, UE	Stylized facts	HAM
Our study US 1977–2019	LIN, POL, LOESS, DIF, HP, HAM	Normalizing to unit standard deviation	GDP, UE, EM	Dating NBER recessions, Lead-lag relations, FR	DIF, LIN, POL, LOESS

The list of methods used, and the series included in the analyses for some studies are not complete. LIN, Linear regression; POL, Polynomial regression; LOESS, LOESS filter; DIF, 1st order differencing; BN, Beveridge and Nelson (1981) method, HP, Hodrick–Prescott filter; FR, frequency domain filtering; UCM, Unobservable component method; HAM, Hamilton filter (Hamilton, 2018); GDP, Gross domestic product, C, consumption; I = Investment, EXP, Export, M, Import, PR, productivity, RWG, real wage, WH, Working hours; UE, unemployment, CPI, Consumer price index; M2, Money supply; CS, Capital stock; EM, Employment, DOC, Department of commerce recessions, persistence is measured as the lag 1 autocorrelation

Table 3 Recessions in the USA 1970 to 2020

No	Recession	Key factors	Seriousness	Period
1	The 1980 recession	The Volcker inflation targeting	1	Jan. 1980–July 1980
2	The 1981–1982 recession	The 1979 energy crisis	3	July 1981–Nov. 1982
3	Early 1990s recession	1990 oil price shock	4	July 1990–Mar. 1991
4	The 2001 recession	The dotcom bubble, the 9/11 attacks and accounting scandals at major US firms	2	Mar. 2001–Nov. 2001
5	The great recession	The subprime mortgage crisis	5	Dec. 2007–June 2009
6	The Covid-19 recession	Pandemic	6	Feb. 2020

Recession seriousness is calculated as the sum of deviation from a linear trend and its duration, both variables normalized to unit standard deviation. The period is peak through trough. Source: NBER, National bureau of economic research. <https://www.nber.org/research/data/us-business-cycle-expansions-and-contractions>, visited 11–11-21

normalized to unit standard deviation rang between about -3 and $+3$). Several studies make recommendations after testing detrending methods, but not always with respect to which detrending methods that have highest skill in reproducing a cyclic component that are “best” or “satisfactorily”.

3 Data

Our sample period is from 1977 to 2020. For part of our analysis, we use economic data at monthly frequency, from 1977M1 to 2019M5 for the US. The reason is that the pandemic in 2020 showed a marked decrease in GDP that tended to distort the detrending processes abnormally. In an additional analysis, we also evaluate our findings by applying the same detrending methods to the UK data.

We use the real GDP as a proxy for real economic growth and identified recession periods using National Bureau of Economic Research (NBER) definitions. We study the US GDP for the period 1977–2020 because the NBER recessions are well-defined events during this period, and they are fixed in time to certain dates. We have collected all the data from the Federal Reserve Bank of St. Louis. The GDP is available at the quarterly frequency and is linearly interpolated to monthly data to match the frequency of employment data. We compare the monthly GDP data with corresponding monthly data supplied by the Research and Analysis IHS Markit,¹ see robustness section.

The US employment (EM) is a measure of the number of US workers (in thousands of persons) in the economy that excludes proprietors, private household employees, unpaid volunteers, farm employees, also known as total nonfarm payroll.² It is monthly and seasonally adjusted. The data sets contain 520 entries.

¹ The data is retrieved from <https://ihsmarkit.com/research-analysis/index.html>

² <https://fred.stlouisfed.org/series/PAYEMS>

Table 3 lists all the NBER recession dates for recessions in USA during the 1970–2020 period. The NBER dating of recessions uses as a “rule-of-thumb” (ROT), a decrease in GDP for two consecutive quarters. However, there are three additional criteria for identifying a NBER recession: it should have a certain depth, it should be widespread across US (diffusion), and it should have a certain duration. The three criteria combine are used to define a recession, but one criterion may compensate for another in identifying a recession. In Table 2 we have added a measure of seriousness as the sum of its deviation from a linear trend and its duration over the period defined by NBER, both measures normalized to unit standard deviation. The data were retrieved from St. Louis Fed.³

4 Methods

This section first briefly describes the seven procedures used to extract trends from the observable time series. A survey of the methods examined and references to their first use were given in Table 1. Then we explain how we determine the reference tie-points, which are the proxy for the NBER recessions, as well as two empirical tests for the resulting cyclic series.

4.1 Alternative detrending methods

In the following, we denote the original time series by $y(t)$, its trend by $y^*(t)$ and the resulting series by $Y(t)$. The trend series $y^*(t)$ is the difference between the original series, $y(t)$ and the detrended series $Y(t)$.

Linear and polynomial trends. Linear and polynomial detrending assume that $y^*(t)$ is a deterministic process which can be approximated by polynomial functions of time. With linear and second order polynomial detrending we used the residuals after subtracting the data corresponding to the regression line from the raw data $y(t)$ where $y^*(t)$ is obtained by regressing the raw series, $y(t)$, on time, t .

$$y^*(t) = \beta_1 t + \beta_2 t^2 + \gamma \quad (1)$$

The detrended series is then.

$$Y(t) = y(t) - y^*(t). \quad (2)$$

The LOESS smoothing algorithm. LOESS, originally proposed by Cleveland (1979) and further developed by Cleveland and Devlin (1988) and Cleveland and Grosse (1991), is also known as local regression. It is a method that smooths a time series piecewise by fitting a smooth curve to a set of data points with weighted linear regression. For each value of t , an estimated value of $f(t)$ is found by using its neighboring sampled values within a running window. The length

³ GDP is retrieved from <https://fred.stlouisfed.org/series/UKNGDP> and unemployment from <https://fred.stlouisfed.org/series/AURUKM>

of the running window is defined by a factor (f) that determines the fraction of the time series that is used as a running window, that is, it determines n in windows $t_i - t_{i+n}$, $i < t_{\max-n}$. A factor (p) that determines the polynomial degree used to interpolate within the window. To detrend the GDP series in our example, we used $f=0.8$ and $p=2$ to identify the trend $y^*(t)$. Since we always use $p=2$, we use the acronym LOESS(f) to show the LOESS parameters used. The LOESS algorithm is available in many statistical packages, and we use the program package SigmaPlot for the LOESS smoothing. The LOESS algorithm may perform fairly similar to the Detrended Fluctuation Analysis (DFA) as it is discussed in Bashan et al., (2008 p. 5082). However, here we avoid calculating the difference between maximal and minimal values.

The first derivative. There are two major techniques for taking the first derivative. One method is to subtract observations over a certain interval, h .

$$Y(t) = y_{t+h} - y_t \tag{3}$$

We follow Estrella and Hardouvelis (1991) by making the subtraction $y_{t+4} - y_t$. Using this method, the unit root is eliminated.

A second method is to calculate an ordinary linear regression, (OLR), over a moving time window and then use the β_1 —value of the regression as the new detrended time series. The β values will replace $\Delta_n = Y_{t+h} - Y_t$, in the traditional first difference method, but damp extreme values that could occur if Y_{t+h} and Y_t , should happen to be extreme in opposite directions, e.g., Seip and Wang (2023).

$$Y(t) = \beta_1(t), y^*(t) = Y(t) - y(t). \tag{4}$$

The HP-filter. The Hodrick–Prescott, HP, high-pass filter separates a time series into trend and cyclical components. The HP filter is frequently used in economics. For example, Bjørnland et al. (2008) apply it among several other methods to study the effect of output gaps on forecasting. The HP filter extracts a stochastic trend which for a given value of λ moves smoothly over time and is uncorrelated with the cycle. Kydland & Prescott (1990) argue that $\lambda = 1600$ is a reasonable choice for quarterly data and many subsequent studies have used this value. As we have monthly data, we use $\lambda = 3 \times 1600 = 4800$. The HP-filter is implemented in several statistical packages, e.g., Stata© and R©.

The Hamilton- filter. The HP filtering technique is said to introduce spurious dynamics that have no basis in underlying data-generating processes (Hamilton, 2018). Hamilton therefore introduces a new technique that estimates an OLR regression of y_{t+h} on a constant and the p most recent values of y as of date t . The description follows closely that of Hamilton (2018).

$$Y_{t+h} = \beta_0 + \beta_1 y_t + \beta_2 y_{t-1} + \beta_3 y_{t-2} + \beta_4 y_{t-3} + \gamma_{t-h} \tag{5}$$

The residuals are

$$\hat{\gamma}_{t+h} = \gamma_{t-h} - \hat{\beta}_0 - \hat{\beta}_1 y_t - \hat{\beta}_2 y_{t-1} - \hat{\beta}_3 y_{t-2} - \hat{\beta}_4 y_{t-3} \tag{6}$$

This gives a way to construct the transient response component.

For quarterly GDP data Hamilton (2018) recommend $p = 4$ and $h = 8$, which for monthly data would translate into $p = 12$ and $h = 24$. The parameter values for p and h refer to cyclical factors for the business cycle movements (Hamilton, 2018).

The logarithm. The logarithm is often applied to time series to stabilize the variance (see Table 1), but the series will not be detrended.

$$Y(t) = \log_{10}(1 + y(t)) \quad (7)$$

4.2 Comparing Detrending Methods

We first explain how we determine the reference tie-point, which is the proxy for the NBER recessions. Thereafter we describe two empirical tests for the resulting cyclic series, their lead-lag relations to employment (EM), and their cycle periods.

4.2.1 Determining the Reference Dates

To determine recessions in the GDP data, we examine if there are periods in the data where GDP decreases over two consecutive quarters. A negative trend will correspond to an approximation to the NBER definition of a recession. (The NBER definition has three additional criteria). To the raw GDP data, we first apply a moving ordinary linear regression, OLR, over 7 months, about two quarters, but with an additional month to get an odd number of months. Our “rule of thumb” (ROT) is that there is an overall decrease in GDP for seven months. The β -coefficient of the OLR is then negative, and thus signaling a possible recession. We do not stabilize the variance of the GDP data since the NBER 2-quarter rule is applied to the raw data. We identify the mid-point month when a negative β -value is encountered. Since the slope may continue to be negative for several months, we discard the dates that correspond to consecutive negative trends. The rationale is that after the first negative GDP change, the following negative coefficients describe the continuation of a deep recession. The dates we identify with the negative slope may correspond either to a NBER recession or to an event with a negative trend over 7 months but not assigned as a NBER recession. Such events that are more than 4 months apart from a NBER recession will be ranked as false. If no signal appears within 4 months interval around the ROT recession, the detrending method will be ranked as missing that ROT recession.

To compare the results for the detrended series, we normalized all series to unit standard deviation. As the year reported is the midpoint year for the slope, the negative trend may start 3 months ahead of the reported date. To make it easier to evaluate the comparisons of dates obtained with the different detrending methods, we report the number of the month since January 1977 (1977M1) when a recession or a ROT event occurred.

For the detrended data, we do not know with any precision what a negative slope in the raw data would correspond to in the detrended data. We therefore divided the β -values into compartments separated by the values: 0, -0.5 , -1.0 , -1.5 , -2 , -2.5 , and -3.0 and record β -coefficients within each compartment. Naturally, there

will be more dates with negative β -coefficients with values in the compartments closest to zero and a negative sign far from zero would suggest a deep recession. We chose to record dates that showed β -coefficients steeper than -1.5 and -2 . The values are a tradeoff between identifying all official recessions and not obtaining too many false recessions.

4.2.2 Lead-Lag Relations

One objective for detrending may be to identify leading indexes for prediction of future movements in GDP. To compare a leading index to GDP, one must normally detrend the GDP since the leading indexes often vary between set limits, e.g., percentages. During cross-correlation studies, series are shifted forward and backward in time to see overlapping patterns. Cross correlation procedure is used to establish if “stylized facts” are supported by macroeconomic time series that are assumed to have a certain relation to GDP, e.g., as in Bjørnland (2000).

We examine the relation between the raw GDP and EM series (both with long-term trends). The lead-lag method calculates an angle, $\theta(3)$, for two paired series $x(t)$ and $y(t)$ based on three consecutive paired observations in the phase diagram for the series ($x = x(t)$, $y = y(t)$). The angle, $\theta(3)$, gives information on the lead-lag relations for three consecutive paired observations in the time series $x(t)$ and $y(t)$. The lead-lag analysis is related to the Lissajous representation of cyclic curves, e.g., Seip & Gron (2017).

A lead- lag method that also uses the dual presentation of $x(t)$ and $y(t)$ as time series and as phase plot is described in Krüger (2021). The rotational direction represented by the angle θ between two successive vectors v_1 and v_2 through three consecutive observations in the trajectory is calculated with Eq. (2)⁴:

$$\theta = \text{sign}(v_1 \times v_2) \cdot \text{Arccos} \left(\frac{v_1 \cdot v_2}{|v_1| |v_2|} \right) \tag{8}$$

The vectors are calculated as $(y_i - y_{i-1}) / (x_i - x_{i-1})$ with $i = 2, 3, \dots$

Since the lead-lag method uses a moving window of three time steps, we can apply the method to time series that are not detrended and not stationary. The lead-lag relations implicitly assumes that a peak in the leading series is followed less than $1/2$ of a common cycle period, λ , of a peak in the target series.

A measure of the persistence of a leading relation, the *LL strength*, is defined as

$$LL\text{-strength} = (N_+ - N_-) / (N_+ + N_-) \tag{9}$$

where N_+ is the number of leading relations for two time series x and y , $x \rightarrow y$, and N_- is the number of lagging relations between the time series, $x \leftarrow y$. Thus, if there is

⁴ The equation can be implemented in a spread sheet (e.g., Excel format): With $v_1 = (A1, A2, A3)$ and $v_2 = (B1, B2, B3)$ in an Excel spread sheet, the angle is calculated by pasting the following Excel expression into C2: $=\text{SIGN}((A2 - A1) * (B3 - B2) - (B2 - B1) * (A3 - A2)) * \text{ACOS}(((A2 - A1) * (A3 - A2) + (B2 - B1) * (B3 - B2)) / (\text{SQRT}((A2 - A1)^2 + (B2 - B1)^2) * \text{SQRT}((A3 - A2)^2 + (B3 - B2)^2)))$.

a persistent positive lead-lag relation over nine consecutive observations, we get $N_+ = 9$, $N_- = 0$, and $LL\ strength = (9-0)/(9+0) = 1$. The number 9 is a tradeoff between measuring lead-lag relations over short periods and the possibility to establish significance. The 95% confidence interval (CI) for the $LL\ strength$ for two raw uniformly stochastic series is ± 0.32 . Frequently, we will smooth time series, partly to avoid high frequency noise and partly to identify cycle variabilities that are of interest, e.g., business cycles. When we calculate $LL\ strength(9)$ for the smoothed series, the probability to obtain sequential angles, $\theta(3)$, that have the same sign increases, so the CI does not strictly apply. Since we want to report the results for the smoothed versions, we use the term “pseudo-significant” when $Abs(LL\ strength) > 0.32$. We screened for noise by applying increasing LOESS smoothing to the GDP and EM series and found that significant (positive or negative) LL values stabilized after LOESS (0.15) were applied to the series. To score the lead-lag relations for US data, we calculated the distances in a PCA loading plot between the lead-lag results for paired detrended data series, GDP and EM and the lead-lag result for the paired raw series GDP and EM. The lead-lag method is illustrated with an example in Appendix B and the spreadsheet with the essential calculations are available upon request.

4.2.3 Cycle Periods

Power spectral density analysis, PSD estimates the magnitudes of the frequency components that combine to make up the variability of a (semi) cyclic time series. By applying PSD, we identify cycle periods in the series. Its result is a graph that shows the density (or strength) of component sine functions of a given cycle period, λ . We stack the PSD graphs, normalized to unit standard deviation, obtained with all six detrending methods and examine if there are cycle periods (peaks) that do not cancel out across detrending methods (Johnson et al., 1996). We examine if the detrended series identify similar cyclic periods, and if the cycles we identify can be related to stylized facts in the macro economy.

4.2.4 Principal Component, PCA, Plots, and Tabulation Of Scores

We examine similarities between the six trends and the six detrended series by comparing them in PCA loading plots. We calculate the effects of detrending on lead-lag relations between GDP, EM pairs and we calculate the effects on cycle period identification. The end-result is a set of scores for the six detrending method on three criteria related to the timing of events, the lead-lag relations to EM, and the identification of cycle periods.

5 Results

First, we examine how our ROT definition of recessions, that is, a negative trend that prevails over 7 months, corresponds with the recessions defined by NBER. The reason we do not use the NBER dates directly for comparison with our detrending results is that NBER has a wider definition than we use, and thus may differ slightly

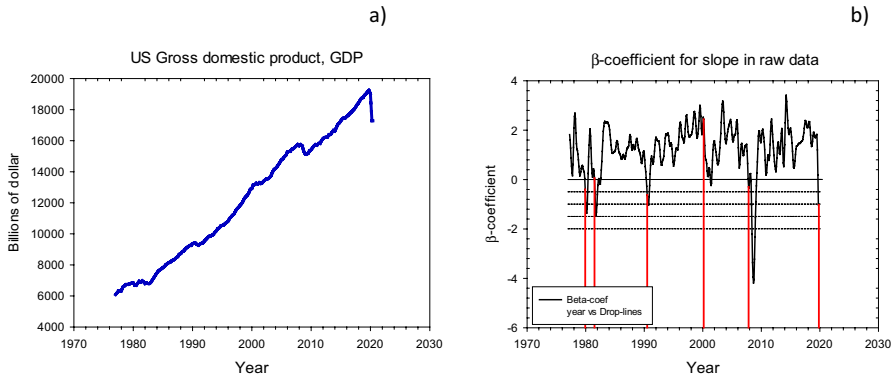


Fig. 1 Raw data for US GDP 1977 to 2020, monthly data. **a** The raw GDP time series including the Feb. 2020 pandemic, **b** Running average regression coefficient (7 months). Dots and drop lines show dates for NBER recessions

Table 4 Time between the US NBER recession dating and negative regression coefficient over 7 months (β -coefficient)

	Recession						Difference from NBER	Average difference	False
	1980	1981–1982	1990	2001	2008	2020			
Month since Jan 1977	37	55	163	279	372	518	n.a	n.a	
Dates identified by negative β -coefficient	36	56	162	293	371	514			1
Difference, months	-1	+1	-1	+14	-1	-4	8 (ex 2001)	1.6	

The number in the “False” column show that negative β -values are registered, but the dates are more than 4 months from the dates for any of the six NBER recessions. n.a.= not applicable

from the dates found by the ROT method. We first present the detrending methods that is close to being generic and last the detrending methods where the parameters are set depending upon characteristics of the time series to be detrended.

All our results in the main text apply to the GDP series for USA. We use that series because the well-defined NBER recessions (tie-points) allows us to identify errors in timing caused by detrending. However, to secure that our test for the skill of detrending methods are robust, we applied the tests also to UK data and report the results as an additional analysis in the Appendix A.

5.1 Dating Proxies for NBER—Recessions

We compare the timing of the NBER recessions with the recessions found by applying our ROT criterion. Figure 1a shows the raw series, not detrended for the GDP

1977–2020 and Fig. 1b shows the 7 months running OLR β -coefficient for GDP, after the GDP has been normalized to unit standard deviation. The NBER recessions are shown with drop lines from the dates of the recession peaks. Dotted lines show incremental β -coefficients values 0, -0.5 , -1.0 , -1.5 , -2 , -2.5 , -3.0 .

Table 4 summarizes the time difference (in months) between the NBER recession start dates and those identified using the negative β -coefficient over 7 months. (The comparison shows that, apart from the 2001 recession, the average difference between the ROT recessions and the NBER recessions is 1.6 months, Table 4.

5.2 The Trends and the Detrended Series

The trends for the six detrending method were shown in Fig. 2a, c, and their similarity in the PCA loading plot in Fig. 2e. (We removed a common average trend because the slopes became identical down to the third decimal without removing it). The linear, LIN and the second order polynomial, POL trends are similar to each other, and the LOESS (0.8) trend is also similar to the LIN and POL trends. (In the figure, they are at about the same position on the PC1 axis that counts for 71% of the variance). The Hamilton-filter, HAM, and the first order differencing, DIF, trends are quite similar, and they are similar to the raw (RAW) data. The detrended series were shown in Figs. 2b, d. Their PCA loading plot shows that the cyclic components for the LIN, POL, and the LOESS time series are similar. In addition, the Hodrick–Prescott filter, HP, and the HAM series are a somewhat similar (they score similarly on PC1), however, the DIF time series are different, Fig. 2f.

5.3 The Three Tests

To evaluate the results, we constructed three criteria for acceptable performance of the detrended GDP time series. (i) The deviation in time of the recessions from the ROT dates should be less than three months (one quarter). (ii) The lead-lag relations $LL(\text{GDP}, \text{EM})$ detrended should not be far from the $LL(\text{GDP}, \text{EM})$ raw. (iii) It should be possible to identify the same cycle periods for the detrended GDP series as a cycle that corresponds to a prominent peak in the stacked series.

5.3.1 Dating Recessions

The results for all methods are calculated in the same way as explained for Fig. 1b. Numerical results are shown in Table 5. For each alternative detrending method, we report two sets of “ROT recession” dates. The first alternative shows slopes (β -coefficients) that are steeper than (minus) -1.5 . The second alternative shows slopes that are steeper than -2.0 . The first alternative identifies more dates for negative β -coefficient as possible recessions, but also more “false” recessions. The right-most column shows the number of “false” recessions. The COVID-19 pandemic that started in 2019 caused an abnormal V-shaped pattern in GDP, which leads to extreme values in the detrended GDP. These extreme values are problematic for some detrending methods that give unreasonable results at the extreme ends of the

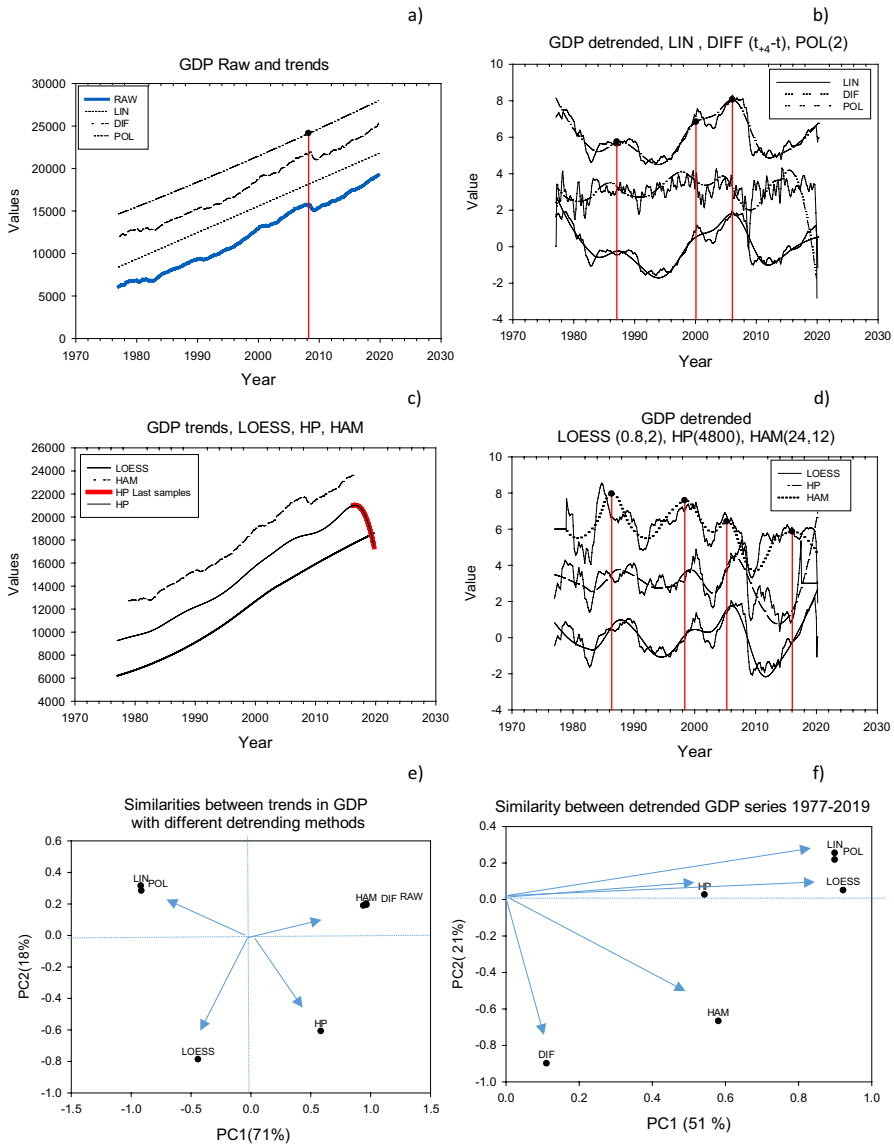


Fig. 2 Trends and detrended series. **a** Gross domestic product for the US 1977 to 2019, and trends in the series obtained with linear (LIN) and polynomial (POL) regression and by subtracting a series obtained by first difference of the GDP series. The trends are shifted upward 3 units to better distinguish among them. **b** Detrended series corresponding to the trends in figure (a). Detrended series are normalized to unit standard deviation. Droplines show peak (or bend) values for HAM detrended series. **c** Similar to (a), but the trends are obtained by LOESS smoothing GDP with parameters $f=0.8$ and $p=2$ (See text), and detrended by the Hodrick–Prescott, HP, filter and the Hamilton method, HAM. **d** Similar to (b) but with LOESS, HP, and HAM detrending methods. Droplines show peak (or bend) values for HAM detrended series. **e** PCA loading plot for the trends. **f** PCA loading plot for the detrended series

Table 5 Time between reported beginning of a US recession and observed beginning in detrended versions

Method	Neg slope	Recession				Difference from NBER			
		1980	1981–1982	1990	2001	2008	2020	Diff/res	False
<i>Month since Jan 1977</i>									
Recessions		37	55	163	279	372	518	0	
ROT		36	56	162		371	514	8	1.6
LIN	-1.5	36	55		280	375	511	12	2.4
LIN	-2.0	37	57	164		377	514	12	2.4
POL	-1.5	35	56	163		371	514	8	1.6
POL	-2.0	36	57	164		377	514	13	2.6
LOESS	-1.5	37	57	163		371	514	7	1.4
LOESS	-2.0	38	58	164		377	514	14	2.8
DIF	-1.5	35	56	163		375	512	12	2.4
DIF	-2.0	36	57	164		377	514	13	2.6
HP-	-1.5	37	57	163	287	377	511	22	3.7
HP	-2.0	39	58	165		378	512	19	3.8
HAM	-1.5	38	57	160	282	370	513	16	2.7
HAM	-2.0			165	287	377	514	103	4.8
Log		38	59			379		12	4
Average		36.8	56.8	163.3	283.0	374.9	513.5	18.1	4.3
Difference Aver-NBER		-0.2	1.8	0.3	4.0	2.9	-4.5		
Difference Aver-ROT		0.8	0.8	1.3	3.0	3.9	-0.5		

ROT, the Rule-of-Thumb value, that is, the β -coefficient for running average regression over 7 months. LIN, residuals from a linear regression, POL, residuals from a 2nd order polynomial function, LOESS, residuals from a LOESS smoothing with parameters f , 0.8 and p , 2; HP, the series after HP filtering, Hamilton, the series after Hamilton filtering. Log, the log 10 values

time series. Therefore, we deleted 3 months at the beginning and 6 months at the end of the time series.

Linear (LIN) and polynomial (POL) trends. The GDP linearly and polynomial detrended is shown in Fig. 2a. The corresponding detrended series normalized to unit standard deviation are shown in Fig. 2b. In this latter figure, we have added a LOESS (0.3) smoothed thin line to better distinguish common peaks and troughs in the series. The droplines identify peaks in the uppermost series. We applied a 7-month moving OLR to the detrended data as in Fig. 1b. We found that the average difference from the ROT dates were 2.4 months for linear detrending and 1.6 months for the second-order polynomial function. (We use the cut-off value that gives the shortest average difference from the ROT recessions. Data for recessions that are not identified are not included in the average.) The average differences were less than a quarter. There is no false recession for linearly detrended series. For polynomial detrended series, the number of false recessions is 3 at negative slope -1.5 and 1 at negative slope -2.0 .

The LOESS smoothing algorithm, LOESS. The time series are first smoothed with the LOESS algorithm and then the residuals after smoothing are calculated. The degree of smoothing may be chosen to disentangle two cyclic processes that together give their imprint on the studied series. This algorithm thus presupposes that the degree of smoothing is evaluated a priori and then implemented in the LOESS algorithm by determining the values of the method parameters f and p . For the GDP series we chose LOESS(0.8), that is, a high degree of smoothing. We will discuss this choice in the discussion section. The trend and the detrended series after the LOESS smoothing are shown in Figs. 2c, d. The average difference from the ROT dates were 1.4–2.8 months, that is less than a quarter. However, the date for the 2001 recession is not identified (and therefore not included in the average). There is 1 false recession at -1.5 and zero at -2 .

The first difference (DIF). The detrended series is shown in Fig. 2b. The droplines allow us to compare peaks in the original series and the first difference series. Differentiating shifts the original series backward. The date for the 2001 recession were not identified. The average difference from the ROT dates is 2.4 and 2.6 months, that is less than a quarter, Table 5. There are 8 false recessions at -1.5 and 1 at -2.0 .

The HP-filter (HP). The trend and the detrended series using the HP filter with its standard parameter $\lambda = 4500$ for monthly series are shown in Fig. 2c, d. The average difference from the ROT dates is 3.7 to 3.8 months, which is more than a quarter off the ROT values. However, the date for the 2001 recession is identified at the cut-off value of -1.5 . According to Table 5, there are two false recessions at cut-off value -1.5 and zero at cut-off value -2 .

The Hamilton filter (HAM). The trend and the detrended series using the Hamilton-filter identifies all 6 recessions with a cut off value at -1.5 and 4 recessions with a cut-off value of -2.0 . However, the average differences from the ROT dates are 2.7 and 4.75 months, the latter being greater than one quarter. There are zero false recessions.

The logarithm (LOG). For the GDP series, the average difference from the ROT dates are 4 months. However, the date for three of the recession is not identified (and therefore not included in the average), Table 5. Many of the studies in

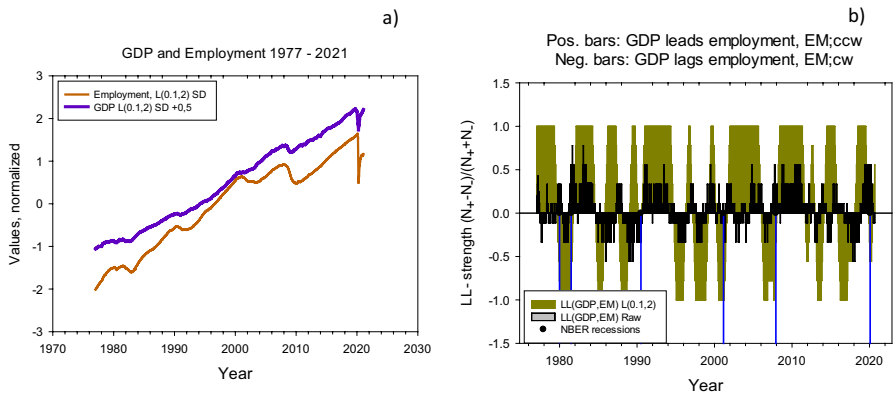


Fig. 3 Lead-lag relations between GDP, Employment, EM. **a** GDP and EM, both LOESS(0.2) smoothed. (text removed). **b** LL relations between GDP and EM, both LOESS(0.2) smoothed (green) and raw, unsmoothed (black). (Text removed). Drop-lines show the beginning of recessions. OLR between green and black bars give $R=0.30$, $p < 0.001$

Table 1 first log-transform the data, but as far as we can see they do not comment on possible shift in timing caused by the log-transformation.

5.3.2 Lead-Lag Relations

Figure 3a shows the GDP and EM not detrended. The time series are displaced relative to each other for clarity. Figure 3b shows their lead-lag relations. The green bars show results for the LOESS(0.2) smoothed, but not detrended, series and the black bars show the results for raw unsmoothed series. An OLR between the two series gives $R=0.30$ and $p < 0.001$. Here GDP leads EM 63% of the time and GDP lags EM 35% of the time.

We applied the lead-lag method to the six set of detrended series for GDP and EM and made a PCA plot for the six lead-lag series to see similarities between them, Fig. 4a. The LIN, POL and LOESS give similar $LL(GDP, EM)$ series. The DIF and the HP detrended series show lead-lag relations that are similar to the lead-lag relations based on the RAW series shown in Fig. 3b. The HAM detrending method is different from the RAW series and from the other detrended series.

5.3.3 Cycle Periods

The stacked power spectra for the six time series is shown in Fig. 4b. The shortest cycles have been removed. The resulting PSD graph shows that there are peaks for the stacked series at 10, 14, 16, 24, 33, 37, 39, and 72 months.

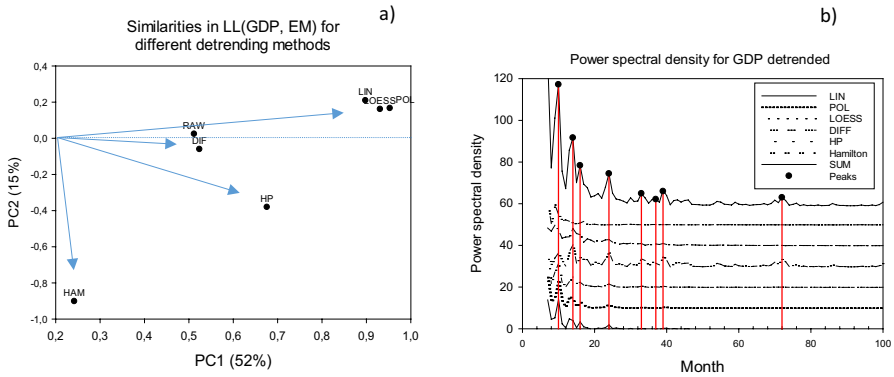


Fig. 4 Empirical tests. **a** Principal component plot for LL(GDP, EM) both series either raw or detrended with different detrending methods, and then LOESS (0.1) smoothed. Letters designate detrending methods. **b** Power spectral density for GDP time series detrended with different detrending methods. Peaks for the stacked series at 10, 14, 16, 24, 33, 37, 39 and 72 months. Acronyms: LIN, linearly detrended, POL polynomial detrended, LOESS, LOESS detrended, DIF, first difference detrended, HP, HP filter, HAM, Hamilton detrending method

Table 6 Three criteria for good detrending (US)

Method	Recessions		LL(GDP, EM)*		Contributing to peak 25 months		Average	
	Score, months	Best-worst	Score	Best-worst	Score, months	Score	Best-worst	
LIN	2.4	5	0.29	3	1.54	2	2.50	
POL	1.6	2	0.32	5	1.38	3	4.00	
LOESS	<i>1.4</i>	1	0.31	4	1.02	5	3.80	
DIF	2.0	3	0.03	1	4.96	1	<i>1.20</i>	
HP	3.7	6	0.2	2	1.04	4	2.47	
HAM	2.3	4	0.41	6	0.01	6	4.67	
Average	2.23	3.50	0.26	3.50	1.66	3.50	3.11	
Criteria	<3		<3.9		>1			

Recessions zero best; LL: zero best; Peak: large value best. *Italic*: outside and better than the 95% confidence interval for the six methods. Numbers marked in bold are outside and worse than the 95% confidence interval

Criteria: The dating of recessions <3 months wrong, LL(GDP,EM)_{detrended} close to LL(GDP,EM)_{raw}, identifying 24 months cycle period with peak >1.0, Scores outside acceptable levels in bold. Average: lowest number is best

*For all detrending methods, except HAM, GDP leads EM more than 50% of the time. Raw (57), LIN(66), POL(63), LOESS (62), DIF (66), HP (69), HAM (77)

5.3.4 Summary of Results

We summarize the results by reporting: (i) the average dating difference for each of the detrending methods to the ROT dating, Table 6. (ii) The difference in LL relations between the LL value for each method and the LL relations found for the raw

Table 7 Three criteria for good detrending (UK)

Method	Recessions		LL(GDP, EM)*		Contributing to peak 25		Average	Sum US and UK
	Score months	Best-worst	Score	Best-worst	Score, months	Best-worst	Best-worst	
LIN	<i>0.50</i>	1	0	1	0.86	5	2.33	4.83
POL	0.75	3	0.002	4	2.12	3	3.33	7.33
LOESS	0.75	3	0.038	3	2.12	3	3.00	6.80
DIF	0.50	2	0.074	1	2.41	2	<i>1.67</i>	2.87
HP	1.25	6	0.204	5	3.70	1	4.00	6.47
HAM	1.00	5	0.182		0.18	6	5.50	10.17
Average	0.79	3.33	0.08	2.80	1.90	3.33	3.31	
Criteria	< 3		< 1.0		> 1			

Recessions: zero best; LL: zero best; Peak: large value best. *Italic*: outside and better than the 95% confidence interval for the six methods. Numbers marked in bold are outside and worse than the 95% confidence interval. Last column shows the scores for both US and UK

series. (iii) A measure of the contribution to the peak for cycle periods at 24 months (calculated as peak value at 24 months minus the average of PSD values for the two previous and the two following months), and the method's skill for each of the three tests are then ranked 1–6. The average of their ranks then expresses their overall skill, all three tests contributing equally.

5.3.5 Additional Analysis Using the UK Data

To verify that our test for the skill of detrending methods are robust, we apply the same tests also to UK data for the period 1977 to 2021. The Bank of England reported four recessions during the period 1977–2021: 1980Q1–1991Q1, 1990Q3–1991Q3, 2008Q2–2009Q2, and 2020Q1–2020 Q2. For the sake of space, the tables are included in Appendix A. Table A1 summarizes the time difference (in months) between the NBER recession start dates and those identified using the negative β -coefficient over 7 months. The table shows that the average difference between the estimated recessions with the rule-of-thumb and the actual recessions is 4.5 months.

Table 7 provides a summary of the three criteria for evaluation the detrending methods. Overall, the detrended GDP for UK gave better results than the detrending for US. For example, the four recessions were determined with an average of 0.79 months precision in the UK (Table 7), whereas it was 2.23 months precision for the US (Table 6). We calculated 95% confidence intervals for the recession scores and for the average scores. The first difference method gave the best detrending results both in the US and the UK, but the method also gave most false recessions. The LOESS detrending method, US, and the LIN detrending, UK, scored better than the averages (averages are scores within the 95% confidence band) on recession dating. The HP and the HAM detrending methods both scored worse than average, either on the recession scores or on the average scores.

6 Discussion

Our results are closely related to our choice of test series, the GDP and EM, series for the US the UK. If we replace EM with unemployment, the two series are related through Okun's law. For GDP it is often assumed that there is steady trend that last for the whole study period (Hamilton, 2011; Franke & Kukacka, 2020, p. 1 and 11). The UK data showed much of the same characteristics as the US data.

6.1 Detrending Methods

We use three criteria for assessing the skill of the detrending method. The scores on the criteria we chose for acceptable results are subjective but are based on the distribution of scores for all six detrending methods. However, depending on the purpose of the detrending, one of the three tests may be more relevant than the others.

Our numerical results show that the first order differencing (DIF) method gave the best overall result, which were quite surprising since differentiating a cyclic series would shift peaks and troughs in the series. On the second place comes linear detrending (LIN). The polynomial detrending (POL), LOESS filter and HP filter are all within the 95% confidence band. On the last places came the HAM method. Complex detrending with the HP and the HAM method using the parameters recommended for GDP gave results that were worse on one or two of the criteria for acceptable results (Tables 6 and 7). However, better determination of the parameters may increase the skill of the two methods.

Our results can be compared to results in other studies, although test criteria are different. Hall et al., (2017, p. 212) found that the HP- filter ($\lambda=1600$; quarterly values) had best fit to New Zealand stylized facts, whereas LOESS (≈ 0.5) fit the facts "to a lesser extent". Hall and Thomson (2021) found that the HP- filter performed poorly without extensions at the extremes, but that the HAM ($h=8$) performed worse.

6.2 The Three Tests

We first discuss the dating of recessions, then the lead-lag relations, and last the cycle periods. The discussion relates to the US time series. The test results using UK data are similar.

6.2.1 Skill in the Dating of Six Recessions

Except for the 2001(2) recessions, all recessions were identified by the ROT method (the number in parentheses is the seriousness of the recessions measured as deepness plus duration, high numbers give the most serious recessions). The recessions were identified with various timing errors. The recessions in 1980 (1) and 1990 (4) was best identified giving an average of $1/3-1/2$ month in error relative to the NBER identification. The recessions in 1981–82 (3) was also well identified by all detrending

methods, and with an average of 2 months difference from the NBER recessions. The two recessions, 2008 (5) and 2020 (6) had an average error of 3–5 months. The recession in 2001 was not identified by the ROT method, but a ROT recession was identified in May 2002, 14 months later. The LIN, the HP, and the HAM methods all detected the 2001 recession. All methods, except for the HP filter, gave dates that were shifted less than a quarter.

For most detrending purposes, the correct dating of events in the series may be most important, for example when events that are assumed to be predictive or causal for a recession are compared to the date of the recession. For detrending the LOESS method scored best with an average of 1.4 months difference from the ROT dates, but with the POL method close (1.6 months). The trends that are removed from the raw data are similar for LIN, POL and LOESS (Fig. 2e, PC1 counts most), and they result in the detrended series for LIN, POL and LOESS to be quite similar (Fig. 2f).

6.2.2 Skill in Determining Lead-Lag Relations

A leading role for a causal effect is a prerequisite, but not a sufficient criterion for causation. However, the leading role is often offered as a strong argument for a causal effect, (Sugihara et al., 2012). Thus, it is important that lead-lag relations are preserved. We examined the lead-lag relations between: Real GDP and employment. The results for EM support Hamilton (2018; p. 838)'s finding that the cyclical component of EM starts to decline significantly before the NBER business cycle peak for essentially every recession. (Here before 1980, 1990, 2008 and 2020 recessions). However, although EM leads GDP before a recession, it lags GDP after the recession.

Since the lead-lag method can be applied to raw, not detrended series, we compared the results obtained with the different detrending methods to the result for the raw series (slightly smoothed). We found that the DIF method gave lead-lag relations that were closest to the results for the RAW series. The LIN, POL and LOESS methods came out similarly, whereas the HAM method came out worst.

6.2.3 Skill in Determining Cycle Periods

Strong peaks, apart from those at very short cycle periods, are at 24 months, (2 years) and at 72 months, (6 years). Both cycle periods can be related to the in-between elections to the congress and to the senate and less directly to presidential tenures of 4 and 8 years. The cycle periods (24 and 72 months) are close to the domain for business cycles identified by Burns & Mitchell (1946) who suggest that business cycles are between 18 and 96 months. To our surprise, the DIF method gave the far best result in identifying cycle period of 24 months and it also identified a cycle period of 72 months. The LIN and POL methods came on the next places and LOESS, HP, and HAM (worst) came on the last places. Hallett & Richter (2006) use a short-time Fourier transform for the UK GDP 1981Q4 to 2003Q1 and find the US and the UK have cycle periods that are common in the long run. (There is a principal cycle period of 62 quarters, 15 years).

6.3 Robustness

We have applied the detrending methods to monthly US GDP from 1977 to 2019. To obtain monthly values, we interpolated linearly the quarterly data reported by US Bureau of Economic Analysis and retrieved from the Federal Reserve Bank of St. Louis. However, the IHS Markit reports monthly values from 1992 to present. Our interpolated monthly GDP data compare well with the monthly data supplied by the IHS for the period 1992 to 2020 ($R^2=0,999$, $p < 0.001$.) Furthermore, using quartile data would not have given us the resolution we obtained, with many recession dating events less than a quarter.

It is a concern that smoothing, e.g., LOESS (0.1) smoothing gives a moving window of ≈ 53 months for our data set, and the LL strength period (9 months) may impact the variations in LL strength and the cycle periods we identify. However, the lead-lag measure for the “not detrended and not smoothed” data shows a recession pattern that coincide with the NBER recession pattern. In addition, a cycle period of 24 months was found by the PSD for almost all detrending methods (no smoothing). Thus, we believe that the advantage of smoothing the time series and using a short period for the LL strength measure outweighs the possible disadvantage that an imprint of our periods (≈ 53 and 9 months) may have on the results.

Statistical confidence intervals. There are no “ground truth” to what the trend for economic time series, like the GDP really is. Thus, we cannot compare our results to the “real” trend. We calculated the 95% confidence interval for the average scores in Table 6. In theory one could construct a trend by incorporating in a model the variables that are assumed to cause a trend, like population, capital stock and technology (Hamilton, 2018, p. 1006). A second approach, as used here, was to apply the detrending methods to the GDP of two “unions”, the US and the UK. Both the US and the UK experienced similar recessions, although the UK only recorded four recessions and the dating was with respect to quarters.

For the GDP of both countries the “simple” detrending methods, LOESS detrending scored well and met our criteria values. Furthermore, determining the factor, f , for smoothing with the LOESS algorithm is visually simple and the LOESS smoothing algorithm is available in most statistical packages.

6.4 Further Studies

Several studies explore the effects of parameter settings for different detrending methods, e.g., Ravn & Uhlig (2002), Hamilton (2018), and Hall & Thomson (2021) on the HP filter. Wills et al. (2018) try to disentangle time series to identify component series that are due to specific mechanisms in climate science. Thus, further studies could explore how generic algorithms for the choice of parameters for each filter could be established and how each choice of parameters would help identify the mechanisms that are acting to create component series in the observed superimposed series. Several detrending methods, e.g., the HP filter, the LOESS smoothing algorithm would allow a stepwise use of detrending from low frequency to high

frequency variabilities. Thus, criteria that would signal that an economically significant cycle component is identified would be especially useful. One criterion could be to stop detrending when stylized facts suggest that a cause/target relationship between two series actually show persistent leading relations and persistent common cycle periods. A second criteria could be that visual inspection of a series show reasonable reproduction of peaks and troughs. A third criterion could be that forecasting based on a training/ test set gives a minimum for root mean square error values.

7 Conclusion

We applied the six detrending methods to the gross domestic product, GDP, for US and the UK 1977 to 2019, and found that first differencing the series gave a detrended series that on the average scored best on three tests: (i) Small shifts in the date for six recessions during the period. (ii) Good and reasonable reproduction of lead-lag relations between GDP and EM, and (iii) identification of cycle periods at 24 and 72 months. However, methods that detrended by subtracting a linear, a second order polynomial, or a LOESS smoothed trend also performed well and satisfied our criteria for adequate detrending. Two common detrending methods, the Hodrick–Prescott, HP-filter and a detrending method developed by Hamilton (2018) obtained worst overall score. However, better judgements of the two last method's parameters may improve their detrending skills.

Appendix A

Detrending of United Kingdom Gross Domestic Product, GDP

Estimating recession dates for United Kingdom recessions for the period 1977–2021. There were four recessions: 1980Q1 to 1991Q1, 1990Q3 to 1991Q3, 2008Q2 to 2009Q2, and 2020Q1–2020 Q2 (Fig. 5).

The calculations differ a little from the calculations for the US recessions. The GDP time series shows an extreme outlier after 2019:11 (due to the Covid-19 pandemic and probably also to the Ukraine war), so we used the linear trend 1977–2019:11 for the whole period 1977:1 to 2021:7 (Tables 8, 9).

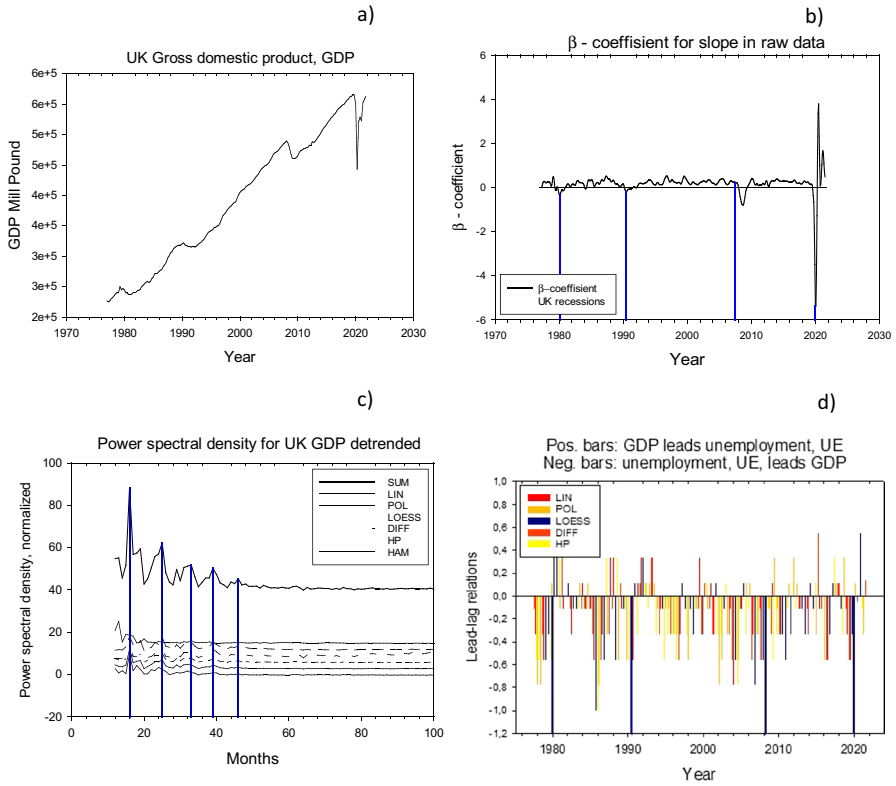


Fig. 5 UK GDP. **a** UK- GDP raw format. **b** β - coefficient for slopes in Raw data. **c** Powe spectral density for the detrended UK-GDP data. There are peaks at 16,25,33, 39 and 46 months, that is with about 7 months intervals. **d** LL relations for five detrending methods

Table 8 Time between Recession dating and negative regression coefficients over 7 months.

	Recessions				Diff from official recession	Diff/res	False
	1980Q1-1981Q1	1990Q3-1991Q3	2008Q2-2009Q2	2020Q1-2020Q2			
Months since Jan 1977	37-38	163-175	376-388	517-520			
Dates identified by negative β -coefficient	29	160	180-182	512			
Difference months	+8	+3	+2	+5	18	4.5	

The difference is our date minus the official first date. We have used the first month in a quarter as the month for the beginning of the recession

Table 9 Time between reported beginning of recession and observed beginning detrended versions. False is slopes that score at -1,5 or -2,0 but do not correspond to an observed recession. "Not detected" are correspond to UK recessions (slopes) but that are not detected

	Neg slope	1980Q1-1981Q1	1990Q3-1991Q3	2008Q2-2009Q2	2020Q1-2020Q2	Diff from ROT	False	Not detected
Recessions		37-38	163-175	376-388	517-520			
ROT		35	160	374	512	0	3	0
LIN	-1.5	35	161	374	513	2	1	0
LIN	-2.0	36	162	375	513	5	0	0
POL	-1.5	36	161	374	513	3	1	0
POL	-2.0	36	-	375	513	3	0	1
LOESS	-1.5	36	161	374	512	2	0	0
LOESS	-2.0	37	162	375	512	5	0	0
DIF	-1.5	36	160	373	512	2	9	0
DIF	-2.0	36	160	374	512	1	4	0
HP	-1.5	37	161	375	513	5	1	0
HP	-2.0	37	-	376	513	5	0	1
HAM	-1.5	36	161	375	513	4	1	0
HAM	-2.0	37	-	376	513	5	0	1

Appendix B

Economy Example

US GDP 1947Q1-1965Q2 We use GDP on itself. The GDP raw would be an idealized leading index for GDP + 3. An independent leading index would be similar to the raw GDP (Fig. 6).

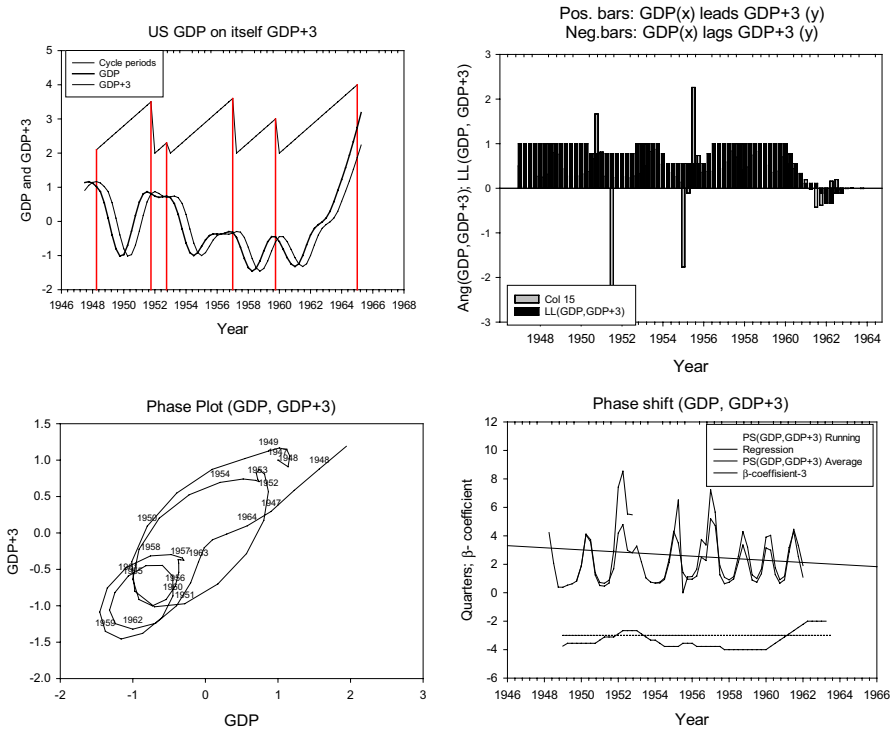


Fig. 6 Example: GDP LOESS (0.2) smoothed (x) and shifted forward three quarters (y). **a** GDP time series, original (blue) and shifted 3 quarters forward (red). The zigzag curve indicates the length of cycle periods found by the cumulative angle method. **b** Lead-lag relations between GDP original and GDP shifted 3 quarters forward, $LL(GDP, GDP + 3)$. The black bars show $\theta(3)$; that is, the lead-lag relation over three consecutive observations in the paired time series. The grey bars show $LL(9)$; that is, the relation between positive and negative angles over 9 consecutive observations. **c** Phase plot for the pairs GDP and GDP + 3. Note that most rotations are counterclockwise (positive, +) showing that GDP leads GDP + 3. **d** Phase shifts calculated relative to running cycle period and relative to an average cycle period. The phase shift corresponds to the three quarters design phase shift. Average cycle period is about 13–14 quarters, lower curve shows running β -coefficient. Number for cycle periods are divided by 10 in a)

Funding Open access funding provided by OsloMet - Oslo Metropolitan University. Knut L. Seip and Dan Zhang are funded by Oslo Metropolitan University. No funds and grants were received during the preparation of this manuscript.

Declarations

Conflict of interest The authors have not disclosed any competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bashan, A., Bartsch, R., Kantelhardt, J. W., & Havlin, S. (2008). Comparison of detrending methods for fluctuation analysis. *Physica a: Statistical Mechanics and Its Applications*, 387(21), 5080–5090.
- Baxter, M. (1991). Business cycles, stylized facts, and the exchange-rate regime—Evidence from the United States. *Journal of International Money and Finance*, 10(1), 71–88.
- Beveridge, S., & Nelson, C. R. (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurements of “business cycles.” *Journal of Monetary Economics*, 7, 151–174.
- Bjørnland, H. C. (2000). Detrending methods and stylized facts of business cycles in Norway—an international comparison. *Empirical Economics*, 25, 369–392.
- Bjørnland, H. C., Brubakk, L., & Jore, A. (2008). Forecasting inflation with an uncertain output gap. *Empirical Economics*, 35(3), 413–436.
- Brault, J., & Khan, H. (2020). The shifts in lead-lag properties of the US Business cycle. *Economic Inquiry*, 58(1), 319–334.
- Burns, G. W., & Mitchell, W. C. (1946). *Measuring business cycles*. National Bureau of Economic Research.
- Burnside, C. (1998). Detrending and business cycle facts: A comment. *Journal of Monetary Economics*, 41(3), 513–532.
- Canova, F. (1998). Detrending and business cycle facts. *Journal of Monetary Economics*, 41(3), 475–512.
- Canova, F. (1999). Does detrending matter for the determination of the reference cycle and the selection of turning points? *Economic Journal*, 109(452), 126–150.
- Christiansen, C., Eriksen, J. N., & Møller, S. V. (2014). Forecasting US recessions: The role of sentiment. *Journal of Banking & Finance*, 49, 459–468.
- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74, 829–836.
- Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83, 596–610.
- Cleveland, W. S., & Grosse, E. (1991). Computational methods for local regression. *Statistics and Computing*, 1, 47–62.
- Enders, W. (2010). *Applied econometric time series* (3rd ed.). Wiley.
- Estrella, A., & Hardouvelis, G. A. (1991). The term structure as a predictor of real economic activity. *Journal of Finance*, 46(2), 555–576.
- Franke, R., & Kukacka, J. (2020). *Notes on the neglected premises of the Hodrick–Prescott detrending and Hamilton regression filter*. University of Kiel.

- Hall, V. B., & Thomson, P. (2021). Does Hamilton's OLS regression provide a "better alternative" to the Hodrick–Prescott filter? A New Zealand business cycle perspective. *Journal of Business Cycle Research*, 17, 151–183.
- Hall, V. B., Thomson, P., & McKelvie, S. (2017). On the robustness of stylised business cycle facts for contemporary New Zealand. *New Zealand Economic Papers*, 51(3), 193–216.
- Hallett, A. H., & Richter, C. (2006). Measuring the degree of convergence among European business cycles. *Computational Economics*, 27(2), 229–259.
- Hamilton, J. D. (2011). Calling recessions in real time. *International Journal of Forecasting*, 27(4), 1006–1026.
- Hamilton, J. D. (2018). Why you should never use the Hodrick–Prescott filter. *Review of Economics and Statistics*, 100(5), 831–843.
- Hodrick, R., & Prescott, E. (1997). Post-war U.S. business cycles: an empirical investigation. *Journal of Money Credit and Banking*, 29(1), 1–16.
- Johnson, G., Thomson, D. J., Wu, E. X., & Williams, S. C. R. (1996). Multiple-window spectrum estimation applied to in vivo NMR spectroscopy. *Journal of Magnetic Resonance, Series B*, 110(2), 138–149.
- Krüger, J. J. (2021). A wavelet evaluation of some leading business cycle indicators for the German economy. *Journal of Business Cycle Research*, 17(3), 293–319.
- Kydland, F. E., & Prescott, E. C. (1990). Business cycles: Real facts and a monetary myth. *Federal Reserve Bank of Minneapolis Quarterly Review*, 14(2), 3–18.
- Park, G. Y. (1996). The role of detrending methods in a model of real business cycles. *Journal of Macroeconomics*, 18(3), 479–501.
- Pollock, D. S. G. (2016). Econometric filters. *Computational Economics*, 48(4), 669–691.
- Ravn, M. O., & Uhlig, H. (2002). On adjusting the Hodrick–Prescott filter for the frequency of observations. *Review of Economics and Statistics*, 84(2), 371–376.
- Seip, K. L., & Gron, O. (2017). A New method for identifying possible causal relationships between CO₂, total solar irradiance and global temperature change. *Theoretical and Applied Climatology*, 127(3–4), 923–938.
- Seip, K. L., & Pleym, H. (2000). Competition and predation in a seasonal world. *Internationale Vereinigung Für Theoretische Und Angewandte Limnologie: Verhandlungen*, 27(2), 823–827.
- Seip, K. L., & Wang, H. (2023). Maximum Northern Hemisphere warming rates before and after 1880 during the Common Era. *Theoretical and Applied Climatology*. <https://doi.org/10.1007/s00704-023-04398-0>
- Seip, K. L., Yilmaz, Y., & Schroder, M. (2019). Comparing sentiment- and behavioral-based leading indexes for industrial production: When does each fail? *Economies*, 7(4), 104.
- Sugihara, G., & May, R. M. (1990). Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, 344(6268), 734–741.
- Sugihara, G., May, R., Ye, H., Hsieh, C. H., Deyle, E., Fogarty, M., & Munch, S. (2012). Detecting causality in complex ecosystems. *Science*, 338(6106), 496–500.
- Tømte, O. T., Seip, K. L., & Christophersen, N. (1998). Evidence that loss in predictability increases with weakening of (metabolic) links to physical forcing functions in aquatic ecosystems. *Oikos*, 82, 325–332.
- Wills, R. C., Schneider, T., Wallace, J. M., Battisti, D. S., & Hartmann, D. L. (2018). Disentangling global warming, multidecadal variability, and El Niño in Pacific temperatures. *Geophysical Research Letters*, 45(5), 2487–2496.
- Zarnowitz, V., & Ozyildirim, A. (2006). Time series decomposition and measurement of business cycles, trends and growth cycles. *Journal of Monetary Economics*, 53(7), 1717–1739.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.