Journal of Econometrics and Statistics
Vol. 2, Issue 2, 2022, pp. 253-266
https://doi.org/10.47509/JES.2022.v02i02.07

Adjusting for Trend Removal in the Frequency Domain

Erhard Reschenhofer

Article History
Received : 29 October 2022; Revised : 19 November 2022; Accepted : 26 November 2022; Published : 15 December 2022

Abstract

In this paper, it is argued that for the detection of a stochastic trend in a time series it is advisable to use the detrended series rather than the original series or the differenced series. While the examination of the original series is clearly impaired by the possible presence of a deterministic trend and dealing with the differenced series comes along with an increased variability, trend removal is a data dependent transformation and is prone to overfitting. However, it is shown that the latter disadvantage can be overcome by simply omitting the lowest Fourier frequency when the analysis is carried out in the frequency domain. This issue is illustrated using climatological, macroeconomic and financial time series. The results of an extensive simulation corroborate the usefulness of this approach for different sample sizes and different types of long-term dependence and short-term dependence.

Keywords: Nonstationarity, stochastic trend, global surface temperature, GDP per capita, stock market index.

1. Introduction

Depending on the type of a time series, there are different questions that may arise when the series exhibits an apparent trend. In the case of a climatological time series (such as a series of temperature measurements taken over a long period of time), the key question is whether this trend is genuine or not. The nature of the trend (stochastic or deterministic) is of minor importance. Any trend is evidence of climate change (see, e.g., Mangat and Reschenhofer, 2020). In contrast, for a macroeconomic time series (such as the GDP) it makes a huge difference whether its long-term behavior is governed by a random walk or by a simple linear trend. In the former case, the effect of a shock on the future level of the series remains constant and does not decrease as the time horizon increases, whereas in the latter case, the effect is only temporary and vanishes eventually (see, e.g., Christiano and Eichenbaum, 1990; Hauser et al., 1999). Of course, it is also possible that a trend can best be described by a combination of a stochastic and a deterministic trend. There are two standard options to produce evidence of the existence of a stochastic trend when the possibility of a deterministic trend cannot be ruled out. We can either fit a deterministic trend to the time series and search for remaining traces of the stochastic trend in the trend residuals or we can take first differences and take a closer look at the differenced series. Inference on the first differences is generally considered safer than inference on the trend residuals because differencing is a

* *Department of Statistics and Operations Research, University of Vienna Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria. E-mail: erhard.reschenhofer@univie.ac.at*

To cite this paper
Erhard Reschenhofer (2022). Adjusting for Trend Removal in the Frequency Domain. *Journal of Econometrics and Statistics.* 2(2), 253-266.

simple, fully deterministic transformation whereas the way how the original observations are transformed into trend residuals depends on the data.

In this paper, it is argued that trend removal may still be the better option because it preserves more of the original information and the negative effect of possible overfitting can easily be controlled in the frequency domain by omitting the lowest Fourier frequency. Section 2 discusses this issue using a climatological as well as a macroeconomic time series. For illustration, a financial time series is also included because we can assume that it has only a stochastic trend. The knowledge of the nature of the trend of this time series allows us to study the effect of the respective transformations without interference from an actual deterministic trend. In Section 3, the results of an extensive simulation study are presented. Section 4 concludes.

2. Empirical study

We investigate three long annual time series, the Earth's global surface temperature from 1850 to 2020 ($n = 171$), the UK's GDP per capita from 1252 to 2018 ($n = 767$), and the Standard and Poor's 500 index from 1927 to 2020 ($n = 94$). The first time series (HadCRUT5 dataset; see Morice et al., 2021) was downloaded from the website https://sites.uea.ac.uk/cru/data of the Climatic Research Unit (CRU) of the University of East Anglia (UEA), the second from the Maddison Project Database (see Bolt and van Zanden, 2020; Scheidel and Friesen, 2009; Stohr, 2016), and the third from Yahoo Finance. All computations were carried out with the free statistical software R (R Core Team, 2018).

Assuming that a possible deterministic trend in any of these time series is quite smooth and can adequately be described by a cubic polynomial (see the first row of Figure 1), we first remove this trend and then examine the trend residuals (see the second row of Figure 1) to check whether there is still a trend left. If so, we interpret this (positive) finding as evidence of a stochastic trend. Analogously, we check whether there is any indication of fractional over-differencing in the first differences (see the third row of Figure 1). If not, this (negative) finding is interpreted as evidence of a stochastic trend because taking first differences usually ensures that any smooth deterministic trend is eliminated. These checks will later be explained in more detail. But before we get there we carry out some basic analyses.

For each of the three time series, sample autocorrelations are shown in Figure 2 and periodograms in Figure 3. Not surprisingly, the sample autocorrelations of the original series decay much slower than those of the trend residuals. In the differenced series, there is either no significant autocorrelation left (financial time series) or only short-term negative auto-correlation (climatological and macroeconomic series). Accordingly, there is always a steep increase in the periodogram of the original series (as the frequency approaches zero), which may be an indication of a pole in the underlying spectral density. This increase is much less pronounced in the case of the trend residuals. Especially remarkable is the downsizing of the first periodogram ordinate. The periodogram of the differenced series is either flat (financial series), which is consistent with white noise, or has a power deficiency in the low frequency range (climatological and macroeconomic time series). In the case of the climatological time series, we might even suspect that the underlying spectral density vanishes at frequency zero.

For a more thorough examination of the long-term behavior of the three time series, we carry out a log periodogram analysis, which is based on the spectral representation

$$\sigma_y^2 = \int_{-\pi}^{\pi} f(\omega) d\omega \tag{1}$$

of the variance of a stationary process y. A pole of the spectral density f at frequency zero can be compatible with stationarity provided that

$$\int_0^{\varepsilon} f(\omega) d\omega < \infty \tag{2}$$

for some small value of ε. To assess the integrability of f, we use simple functions of the form

$$g_\alpha(\omega) = \omega^\alpha \tag{3}$$

as benchmarks, which are integrable if $\alpha > -1$. Plotting $\log(\omega^\alpha)$ against $\log(\omega)$, we obtain a straight line with slope α, which we can compare with the slope of the scatter plot obtained by plotting log periodogram ordinates against log Fourier frequencies in the neighborhood of frequency zero. As expected, the latter slope is less than -1 for all three time series (see the first row of Figure 4), which indicates nonstationarity of the original series. However, the same is also true for the trend residuals (see the second row of Figure 4). Nonstationarity after the removal of a possible deterministic trend is an indication of the presence of a stochastic trend. A slope less than 1 in the differenced series corresponds to a slope less than -1 before differencing, hence the third row of Figure 4 corroborates our suspicion of a stochastic trend. Remember, taking first differences is a proven method to eliminate a deterministic trend. Concluding we may state that all three times series are integrated of order $d \geq 0.5$, the climatological series with $0.5 < d < 1$, the macroeconomic series with $d > 1$, and the financial series with $d = 1$.

A noticeable anomaly in the log periodograms of the trend residuals is the small size of the first ordinate. This finding is particularly severe in the case of the financial time series (see Figure 4.h), where we can be quite sure that this anomaly is solely due to an unnecessary effort to remove a non-existing deterministic trend. However, there seems to be a simple remedy. All we have to do is omit the first Fourier frequency from the log periodogram analysis. Figures 4.b and 4.e show that this measure works also for much larger sample sizes. To investigate this matter thoroughly, we will carry out an extensive simulation study in the next section.

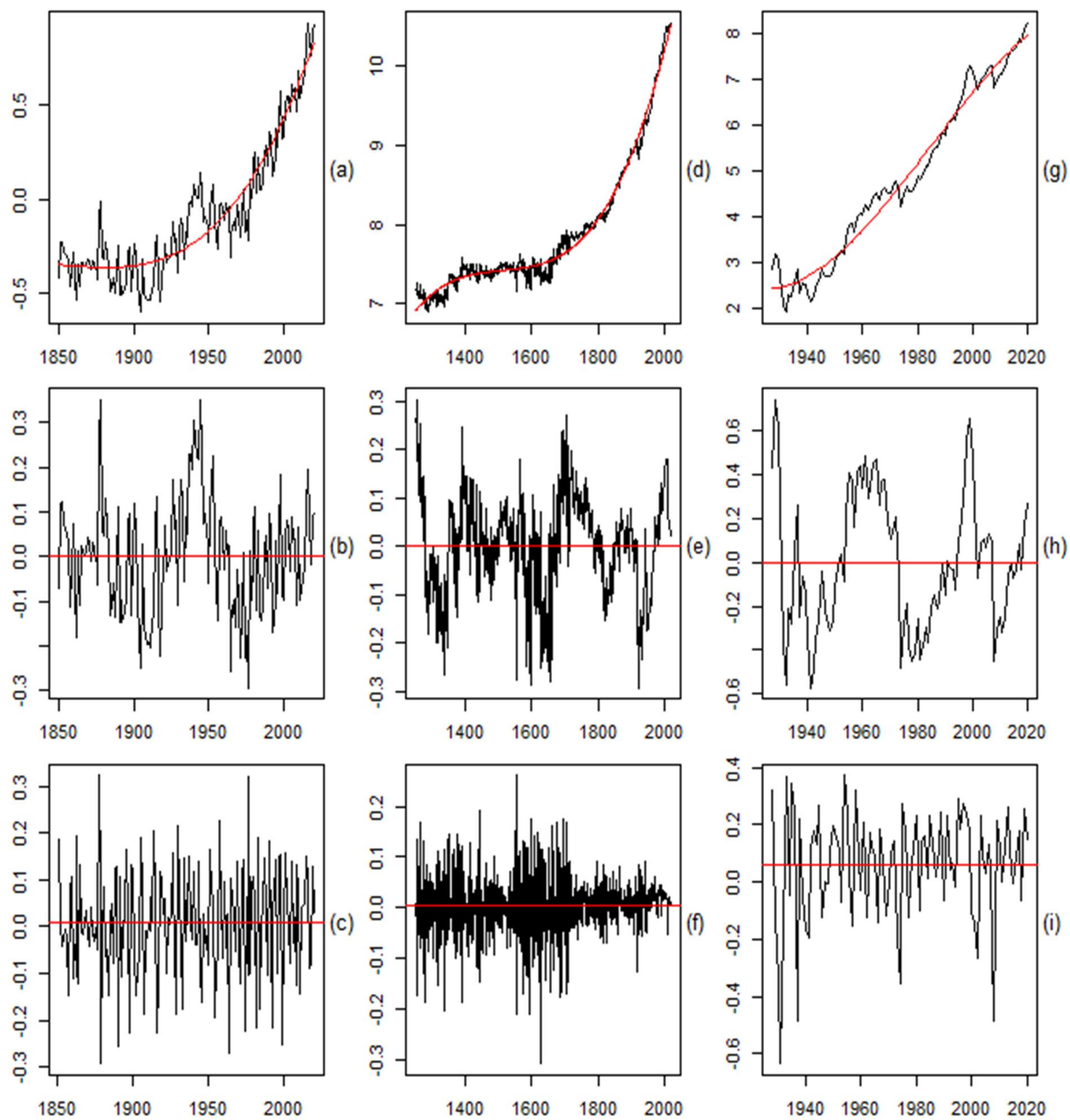

Figure 1: Climatological, macroeconomic, and financial time series
1st column: Global surface temperature from 1850 to 2020 (HadCRUT5)
2nd column: log UK GDP per capita from 1252 to 2018 (Maddison Project Database)
3rd column: log Standard and Poor's 500 index from 1927 to 2020 (source: Yahoo Finance)
1st row: Fitted cubic trend (red line)
2nd row: Deviations from cubic trend (the red line represents the mean)
3rd row: First differences (the red line represents the mean)

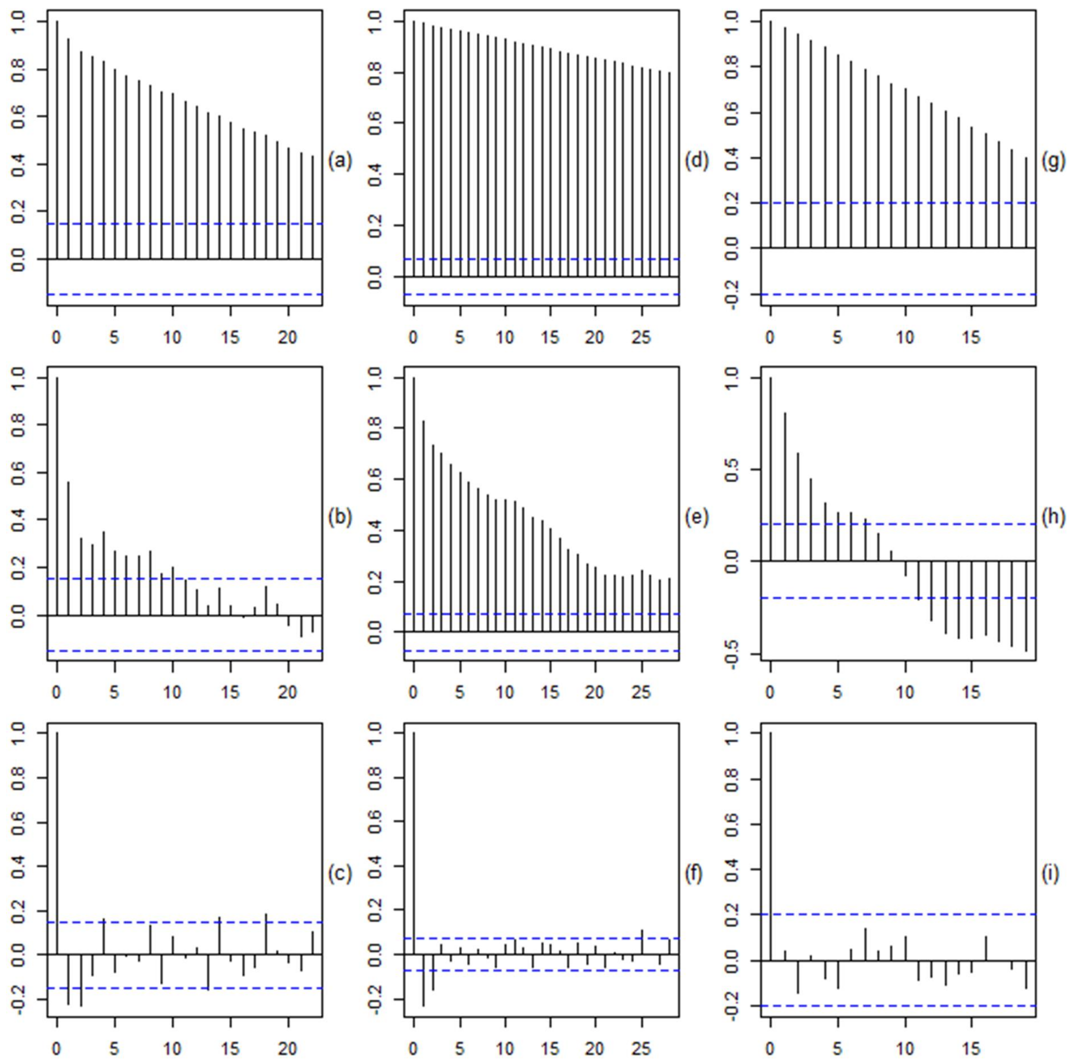

Figure 2: Sample autocorrelations
1st column: Global surface temperature from 1850 to 2020 (HadCRUT5)
2nd column: log UK GDP per capita from 1252 to 2018 (Maddison Project Database)
3rd column: log Standard and Poor's 500 index from 1927 to 2020 (source: Yahoo Finance)
1st row: Sample autocorrelations of original time series
2nd row: Sample autocorrelations of deviations from cubic trend
3rd row: Sample autocorrelations of first differences

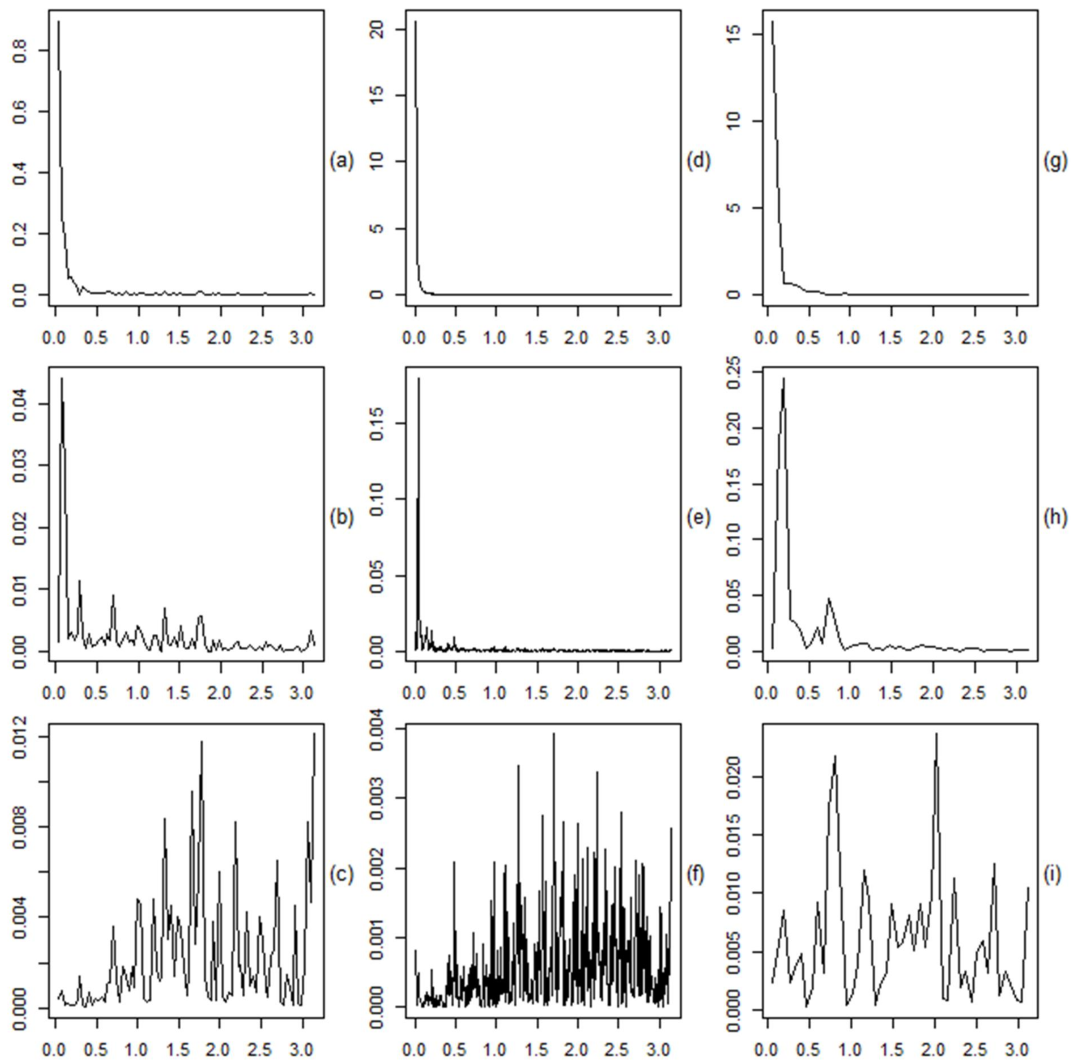

Figure 3: Periodograms
1st column: Global surface temperature from 1850 to 2020 (HadCRUT5)
2nd column: log UK GDP per capita from 1252 to 2018 (Maddison Project Database)
3rd column: log Standard and Poor's 500 index from 1927 to 2020 (source: Yahoo Finance)
1st row: Peridograms of original time series
2nd row: Periodograms of deviations from cubic trend
3rd row: Periodograms of first differences

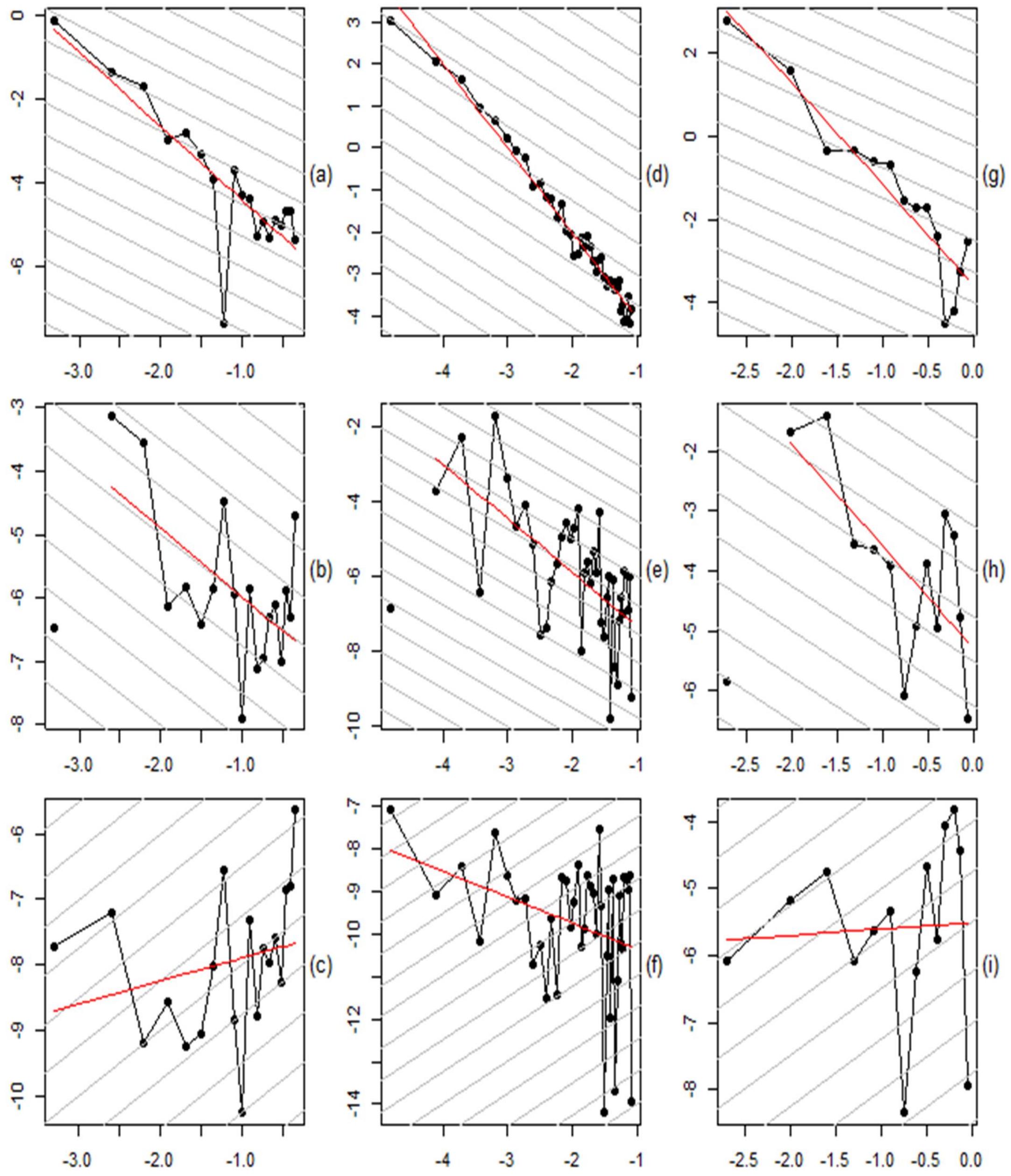

Figure 4: Log periodogram plotted against log frequency

1st column: Global surface temperature from 1850 to 2020 (HadCRUT5)

2nd column: log UK GDP per capita from 1252 to 2018 (Maddison Project Database)

3rd column: log Standard and Poor's 500 index from 1927 to 2020 (source: Yahoo Finance)

1st row: Log periodogram regression for original time series (slope of gray baselines: -1)

2nd row: Log periodogram regression for deviations from cubic trend (slope of lines: -1)

3rd row: Log periodogram regression for first differences (slope of lines: 1)

3. Simulations

In our simulation study, we use fractionally integrated AR(1) processes

$$y_t = (1 - \phi L)^{-1}(1 - L)^{-d} u_t \tag{4}$$

(Granger and Joyeux, 1980; Hosking, 1981) with $\phi = -0.3$, 0, 0.3, $d = 0.2$, 0.4, 0.6, 0.8, 1, and $\sigma_u^2 = 1$. In each case, we generate 1000 realizations of length $n = 100,\ 500$. For each realization, the periodogram

$$I(\omega_k) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} y_t e^{-i\omega_k t}\right|^2 \tag{5}$$

of the residuals obtained by fitting a cubic trend is calculated at the first $K = \lceil 1.5\sqrt{n} \rceil$ Fourier frequencies $\omega_k = 2k\pi/n$, $k = 1, \dots, K$. In Figures 5 and 7, the log means over all realizations as well as the logs of the first and third quartiles are plotted against the log frequencies for the sample sizes 100 and 500, respectively. Analogous figures are produced for the first differences (see Figures 6 and 8). Straight lines with slope $-2d$ ($-2d + 2$ in the case of first differences) are used as benchmarks because the spectral density

$$f(\omega) = \frac{\sigma^2}{2\pi}\left|1 - e^{-i\omega}\right|^{-2d}\left|1 - \phi e^{-i\omega}\right|^{-2} \tag{6}$$

of the process (4) can in the neighborhood of frequency zero be approximated by

$$f(\omega) \approx C\left|1 - e^{-i\omega}\right|^{-2d} = C\left|2\sin\left(\frac{\omega}{2}\right)\right|^{-2d} \approx C|\omega|^{-2d}, \tag{7}$$

where

$$C = \frac{\sigma^2}{2\pi}|1 - \phi|^{-2}. \tag{8}$$

The slopes of the scatterplots of the log means are, by and large, consistent with the values $-2d$ and $-2d + 2$, respectively, expected for the respective data generating processes. However, in the case of the trend residuals, this is only true when the first Fourier frequency is omitted. Minor distortions at this frequency occur also in the case of first differences. Fortunately, they are a lot less severe. But this does not mean that taking first differences is the better option. Looking at the first and third quartiles, we see that the spread relative to the slope is much larger in the case of the first differences. Furthermore, the omission of (at least) the first Fourier frequency can also be justified by distributional considerations. Because of the irregular behavior of the spectral density in the neighborhood of frequency zero, the standard asymptotical results for the periodogram can only be obtained when frequencies too close to zero are excluded. Indeed, Künsch (1986) showed that the ratios $I(\omega_k)/f(\omega_k)$, $k = H + 1, \dots, H + K$ are asymptotically i.i.d. standard exponential provided that $(H + 1)/\sqrt{n} \to \infty$ and $(H + K)/n \to 0$.

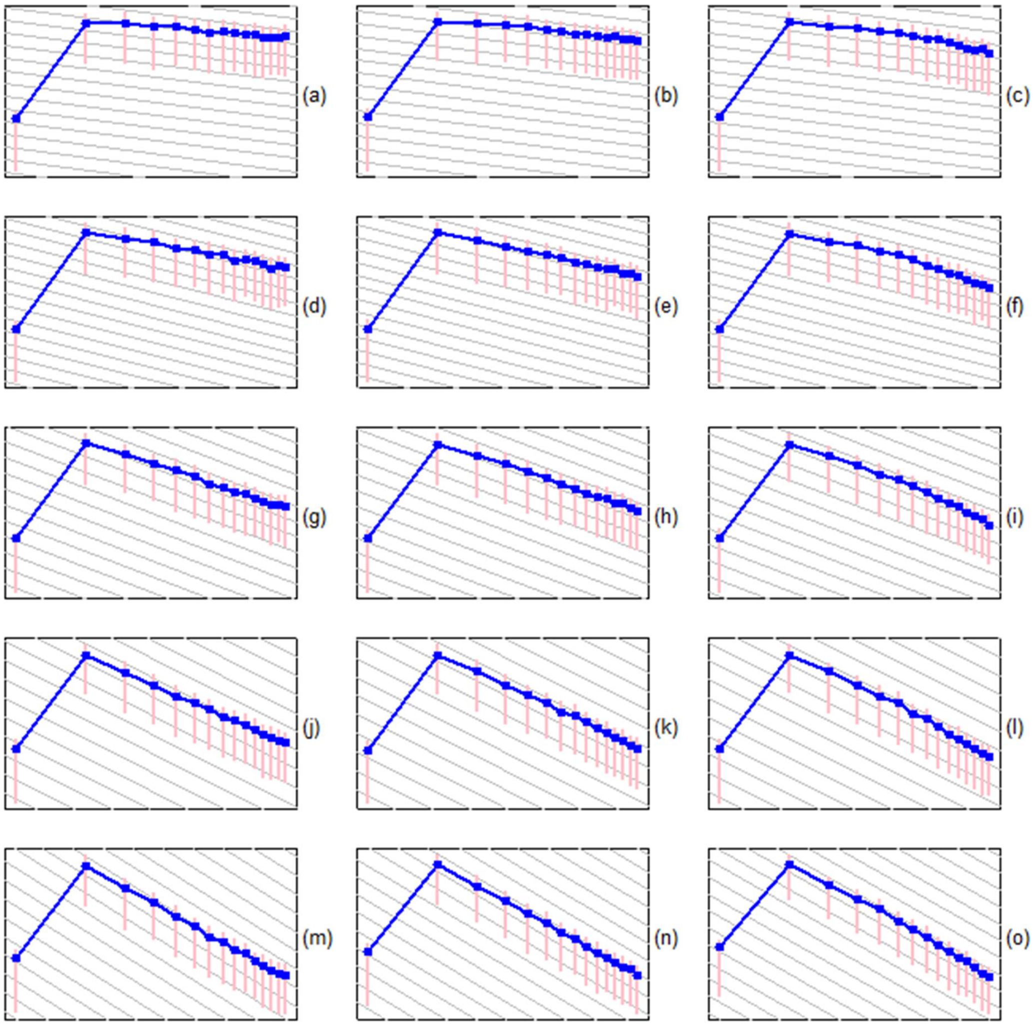

Figure 5: Log periodogram analysis of the trend residuals obtained by fitting a cubic trend to a sample of size $n = 100$ from an integrated AR(1) process with $\phi = -0.3$ (1^{st} column), $\phi = 0$ (2^{nd} column), $\phi = 0.3$ (3^{rd} column) and with $d = 0.2$ (1^{st} row), $d = 0.4$ (2^{nd} row), $d = 0.6$ (3^{rd} row), $d = 0.8$ (4^{th} row), $d = 1$ (5^{th} row). The blue lines represent means over 1000 realizations. The pink lines connect first and third quartiles. The gray baselines have slope $-2d$.

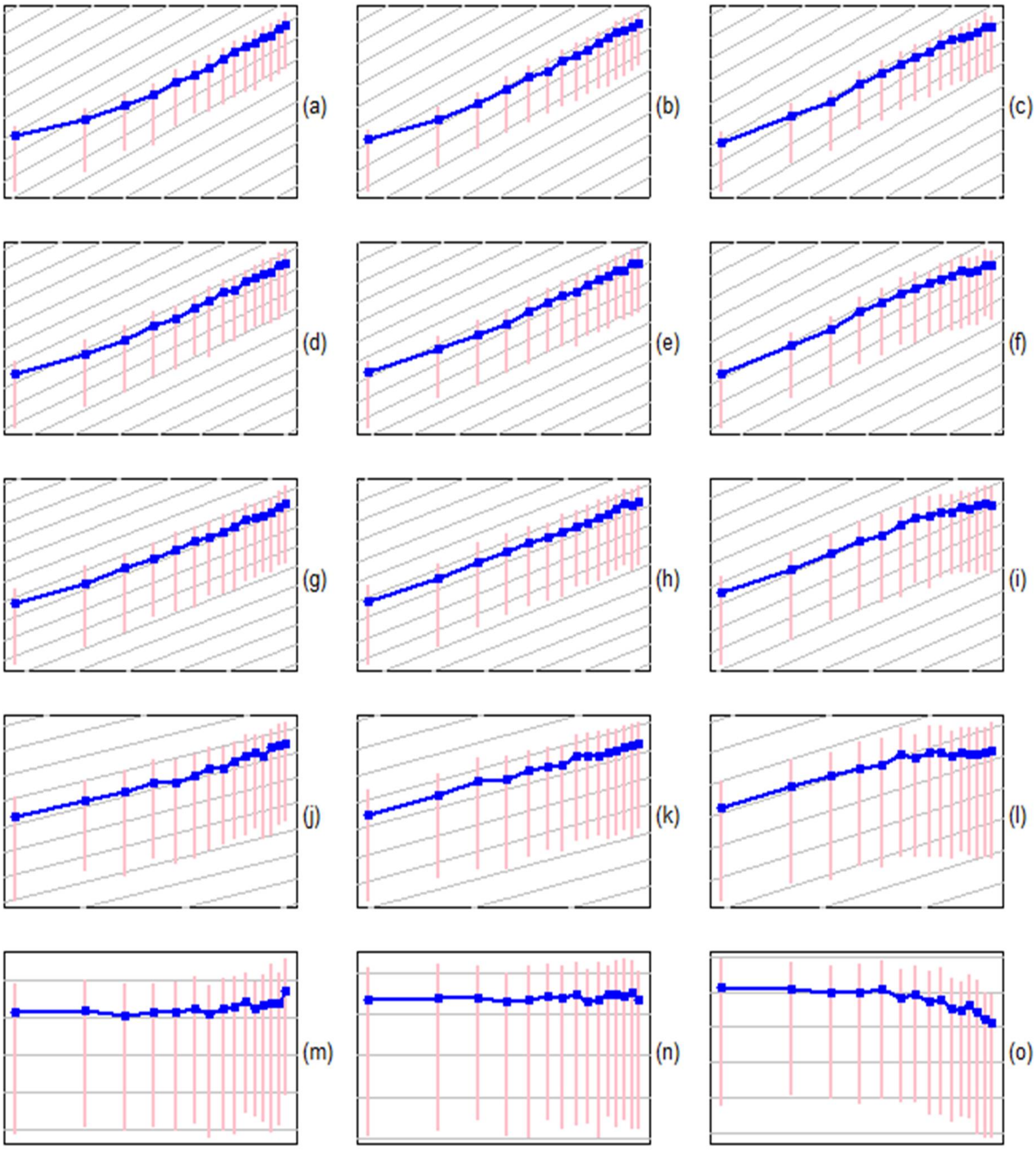

Figure 6: Log periodogram analysis of the first differences of a sample of size $n = 100$ from an integrated AR(1) process with $\phi = -0.3$ (1st column), $\phi = 0$ (2nd column), $\phi = 0.3$ (3rd column) and with $d = 0.2$ (1st row), $d = 0.4$ (2nd row), $d = 0.6$ (3rd row), $d = 0.8$ (4th row), $d = 1$ (5th row). The blue lines represent means over 1000 realizations. The pink lines connect first and third quartiles. The gray baselines have slope $-2d + 2$.

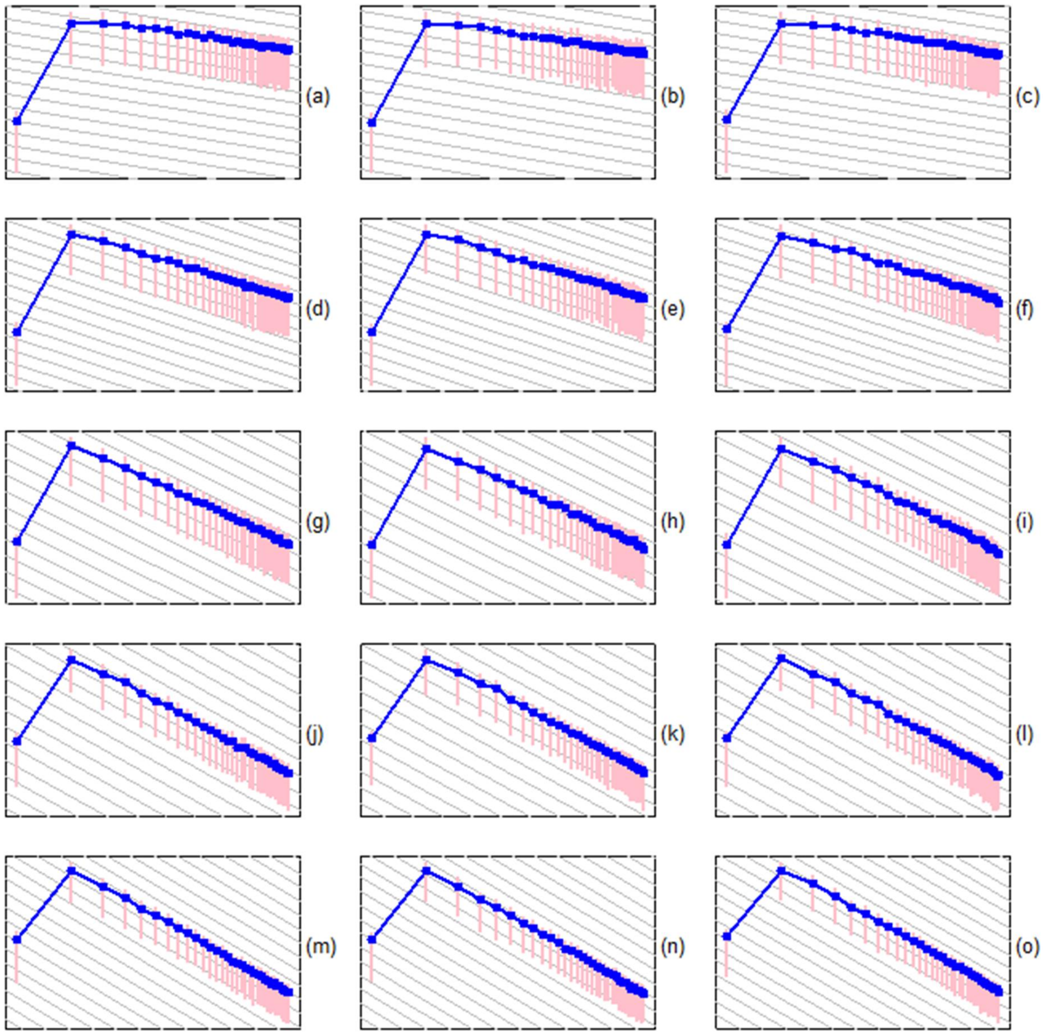

Figure 7: Log periodogram analysis of the trend residuals obtained by fitting a cubic trend to a sample of size $n = 500$ from an integrated AR(1) process with $\phi = -0.3$ (1^{st} column), $\phi = 0$ (2^{nd} column), $\phi = 0.3$ (3^{rd} column) and with $d = 0.2$ (1^{st} row), $d = 0.4$ (2^{nd} row), $d = 0.6$ (3^{rd} row), $d = 0.8$ (4^{th} row), $d = 1$ (5^{th} row). The blue lines represent means over 1000 realizations. The pink lines connect first and third quartiles. The gray baselines have slope $-2d$.

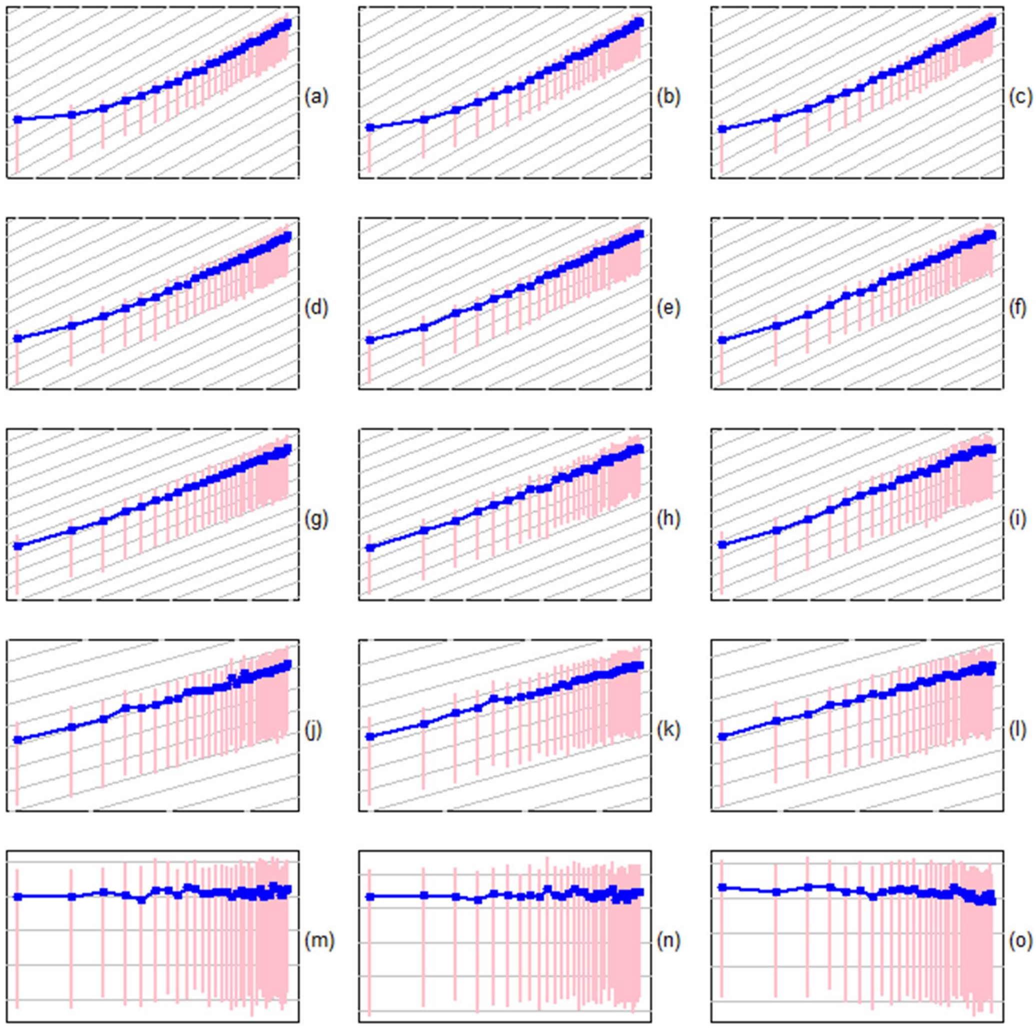

Figure 8: Log periodogram analysis of the first differences of a sample of size $n = 500$ from an integrated AR(1) process with $\phi = -0.3$ (1st column), $\phi = 0$ (2nd column), $\phi = 0.3$ (3rd column) and with $d = 0.2$ (1st row), $d = 0.4$ (2nd row), $d = 0.6$ (3rd row), $d = 0.8$ (4th row), $d = 1$ (5th row). The blue lines represent means over 1000 realizations. The pink lines connect first and third quartiles. The gray baselines have slope $-2d + 2$.

4. Discussion

The estimation of the fractional differencing parameter d can be seriously impaired by the presence of a deterministic trend. It may therefore be prudent to either remove a possible deterministic trend or take first differences before the estimation. The estimate obtained from the transformed series can then be used to draw conclusions about the size of the parameter for the original time series. An advantage of differencing is that this transformation does not depend on the data and a disadvantage is that the variation relative to the size of d is usually large in the case of the differenced series. In this paper, it is argued that trend removal is the better option because the potentially dangerous effects of fitting a deterministic trend to the original time series can easily be controlled in the frequency domain by omitting the first Fourier frequency. Examining the log periodograms of climatological, macroeconomic and financial time series graphically, we always conclude that the original time series is integrated of order $d > 0.5$. The results of an extensive simulation study show that the omission of only one Fourier frequency is sufficient for sample sizes typically occurring in practice and various types of long-term dependence and short-term dependence.

Instead of the graphical inspection of the log periodogram, we could clearly also carry out a log periodogram regression (Geweke and Porter-Hudak, 1983) and thereby obtain a unique estimate and possibly even a confidence interval. However, this may create a false sense of security because the estimation of d is an ill-posed estimation problem (see Pötscher, 2002). Inference in a strict sense is therefore not possible unless extremely strong restrictions are imposed which are implausible in most applications.

Our approach can easily be extended to the case of more than one time series. For example, in the case of two time series x and y, we could check whether there is a cointegrating relationship by looking at the log periodogram of the residuals obtained by regressing y_t on x_t and a deterministic trend (in the simplest case: $a + bt$). Of course, it would again be required to omit the first Fourier frequency in order to avoid possible distortions caused by the trend removal.

References

Bolt, J. and van Zanden, J.L. (2020). Maddison style estimates of the evolution of the world economy. A new 2020 update, Maddison Project Database (2020).

Christiano, L.J. and Eichenbaum, M.S. (1990). Unit roots in real GNP: do we know, and do we care? Carnegie-Rochester Conference Series on Public Policy, Elsevier, 7–61.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models, Journal of Time Series Analysis, 4, 221–38.

Granger, C.W.J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing, Journal of Time Series Analysis, 1, 15-29.

Hauser, M.A., Pötscher, B.M. and Reschenhofer, E. (1999) Measuring persistence in aggregate output: ARMA models, fractionally integrated ARMA models and nonparametric procedures, Empirical Economics, 24, 243-269.

Hosking, J.R.M. (1981). Fractional differencing, Biometrika, 68, 165-176.

Künsch, H.-R. (1986). Discrimination between monotonic trends and long-range dependence, Journal of Applied Probability, 23, 1025–30.

Mangat, M.K. and Reschenhofer, E. (2020). Frequency-domain evidence for climate change, Econometrics, 8, 28.

Morice, C.P., Kennedy, J.J., Rayner, N.A., Winn, J.P., Hogan, E., Killick, R.E., Dunn, R.J.H., Osborn, T.J., Jones, P.D. and Simpson, I.R. (2021). An updated assessment of near-surface temperature change from 1850: the HadCRUT5 dataset, Journal of Geophysical Research.

Pötscher, B.M. (2002). Lower risk bounds and properties of confidence sets for ill□posed estimation problems with applications to spectral density and persistence estimation, unit roots, and estimation of long memory parameters, Econometrica, 70, 1035-1065.

R Core Team (2018). R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna.

Scheidel W. and Friesen, S.J. (2009). The size of the economy and the distribution of income in the Roman Empire, Journal of Roman Studies, 99, 61–91.

Stohr, C. (2016). Trading gains: new estimates of Swiss GDP, 1851-2008, Economic History Working Papers, London School of Economics and Political Science, Economic History Department, 245.