Decomposing Time Series

Kevin Kotzé

Contents

Introduction to decompositions

- Most time series exhibit repetitive or regular behaviour over time
- Hence a great part of the study of this data is conducted in the time domain, with ARIMA or state-space models
- Another important phenomena of time series is that they may be decomposed into periodic variations of the underlying phenomenon
- Shumway & Stoffer (2011) suggest these frequency decompositions should be expressed as Fourier frequencies that are driven by sines and cosines
- This follows the tradition of Joseph Fourier

Identifying the business cycle

- These decompositions may be used to describe the stylised facts of the business cycle:
- Business cycle refers to regular periods of expansion and contraction in major economic aggregate variables (Burns and Mitchell, 1946)
  - i.e. persistence of economic fluctuations and correlations (or lack thereof) across economic aggregates
- We say that a turning point occurs when the business cycle reaches local maximum (peak) or local minimum (trough)
- May also be used to identify the output gap (difference between potential and actual output)
- It is important to note that when seeking to measure the business cycle, there are no unique periodicities that are of relevance
- In addition, most economic time series are both fluctuating and growing, which makes the decomposition quite difficult

Introduction to decompositions

- We may imagine a system responding to various driving frequencies by producing linear combinations of sine and cosine functions
- The frequency domain may be considered as a regression of a time series on periodic sines and cosines
- This lecture considers a few widely used methods that are used to decompose economic time series

Spectral Analysis

- A time series could be considered as a weighted sum of underlying series that have different cyclical patterns
- The total variation of an observed time series will be the sum of each underlying series, which may vary in different *frequencies*
- Spectral analysis is a tool that can be used to decompose the variation of a time series into different frequency components

Spectral Analysis

- Consider an example of three quarterly time series variables, y_t, x_t and v_t
- Where the term $v_t \sim \mathrm{i.i.d.}\mathcal{N}(0, \sigma)$
- If $x_t = y_t^\top \beta + v_t$, then after regressing y_t on x_t, we would expect to find that the coefficient would be large and significant, provided that σ is not too large
- The rationale for this is that x_t contains information about y_t, which is reflected by the coefficient value for β
- From an intuitive perspective, frequency domain analysis involves regressing a time series variable, x_t, on a number of different periodic frequency components
- This would allow us to identify which frequency component is contained in x_t

Spectral Analysis

- To define the rate at which a series oscillates, we define a cycle as one completed period of a sine or cosine functions

$$y_t = A\cos(2\pi\omega t + \phi)$$

- for $t = 0, \pm 1, \pm 2, \ldots$, where ω is a frequency index, defined in cycles per unit of time, which is an expression of T
- A determines the height or amplitude of the function and the starting point of the cosine function is termed the phase, ϕ

Spectral Analysis

- When seeking to conduct some form of data analysis, it is usually easier to use a trigonometric identity of this expression which may be written as,

$$y_t = U_1 \cos(2\pi\omega t) + U_2 \sin(2\pi\omega t)$$

- where $U_1 = A \cos \phi$ and $U_2 = -A \sin \phi$ are often taken to be normally distributed random variables
- The above random process is also a function of its frequency, defined by the parameter ω
- The frequency is measured in cycles per unit of time
- For $\omega = 1$, the series makes one cycle per time unit
- For $\omega = .50$, the series makes two cycles per time unit

Spectral Analysis

- To see how the spectral techniques can be used to interpret the regular frequencies in the series, consider the following four periodic time series

$$
\begin{aligned}
x_{1,t} &= 2\cos(2\pi t6/100) + 3\sin(2\pi t6/100) \\
x_{2,t} &= 4\cos(2\pi t30/100) + 5\sin(2\pi t30/100) \\
x_{3,t} &= 6\cos(2\pi t40/100) + 7\sin(2\pi t40/100) \\
y_t &= x_{1,t} + x_{2,t} + x_{3,t}
\end{aligned}
$$

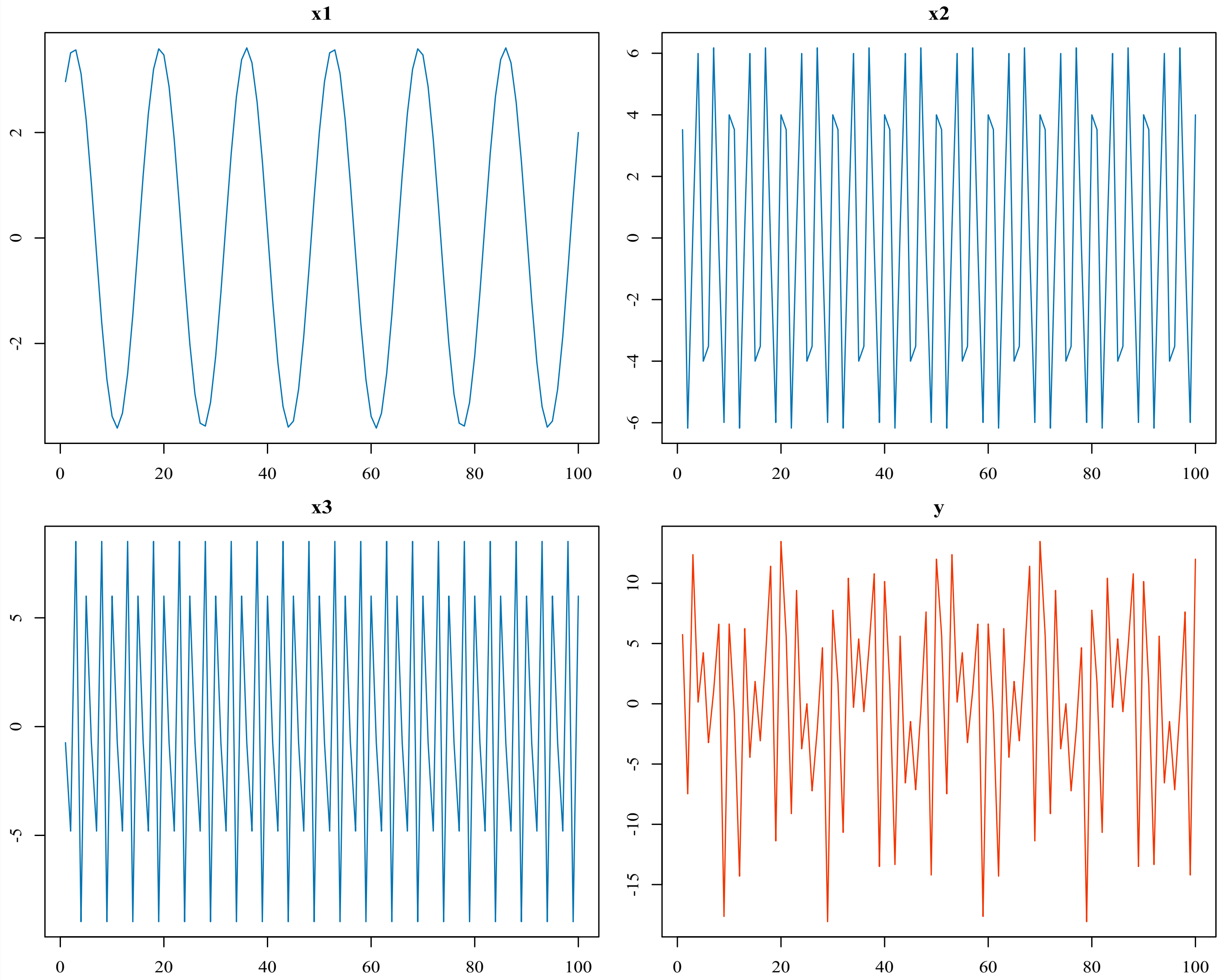

Figure : Different frequency components

Spectral Analysis

- Sorting out of the essential frequency components in a time series, including their relative contributions, constitutes one of the main objectives of spectral analysis
- One way to accomplish this objective is to regress sinusoids that vary at the different fundamental frequencies on the data
- This is represented by the periodogram (or sample spectral density) and may be expressed as,

$$P(j/n) = \frac{2}{n}\sum_{t=1}^{n} y_t \cos\left(2\pi t\ j/n\right)^2 + \frac{2}{n}\sum_{t=1}^{n} y_t \sin\left(2\pi t\ j/n\right)^2$$

- It may be regarded as a measure of the squared correlation of the data with sinusoids oscillating at a frequency of $\omega_j = j/n$, or j cycles in n time points

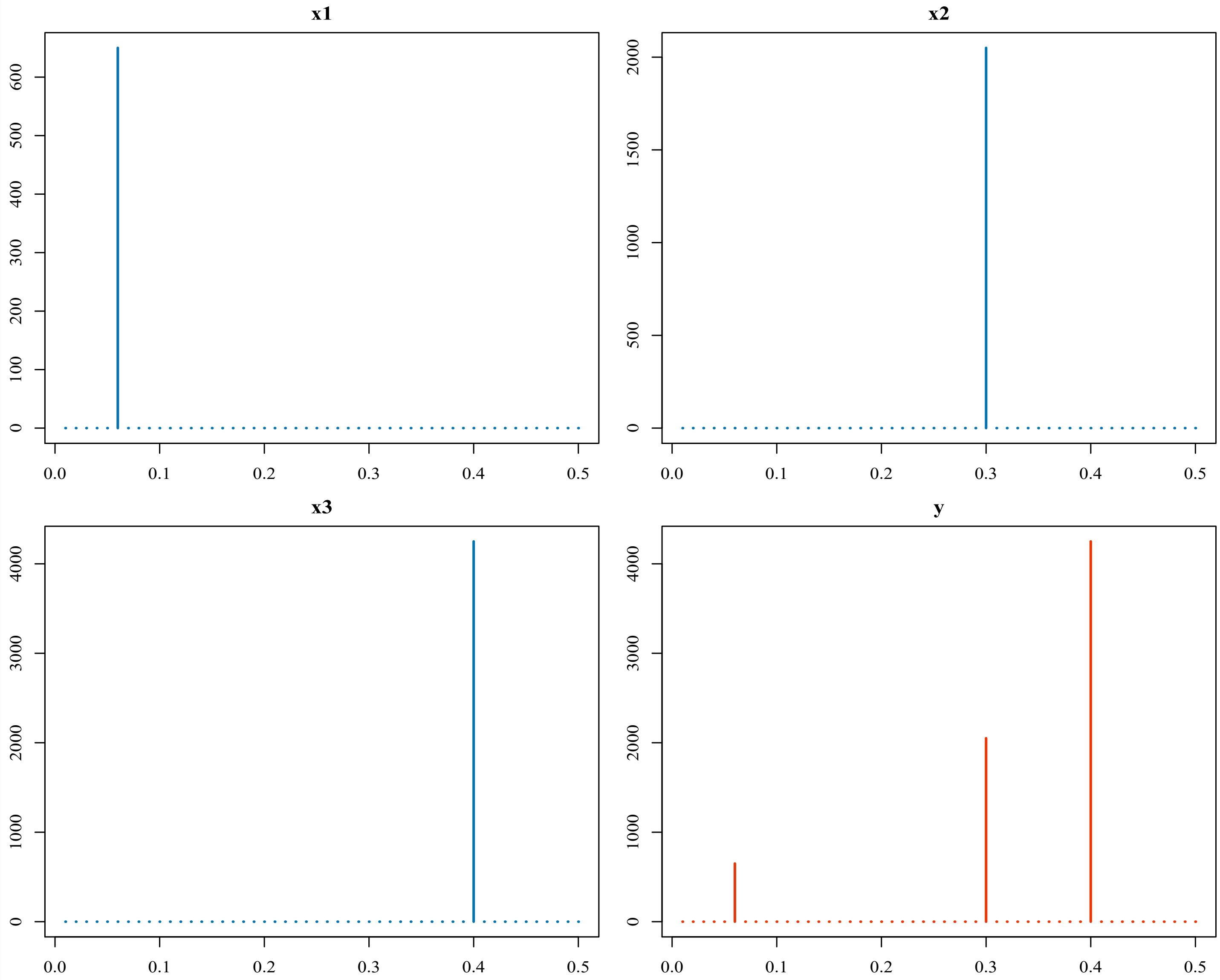

Figure : Periodogram for frequency components

Spectral Analysis

- An interesting exercise would be to construct the x_1 series from y_t, which may be regarded as actual data
- To do so we need to filter out all components that lie outside the chosen frequency band of x_1
- Such a filter could operate with the aid of a regression model that contains the information that relates to a particular frequency (although there are more convenient ways of going about this)
- Hence, cycles with frequencies corresponding to x_2 and x_3 would be excluded, while cycles with frequency corresponding to x_1 will be maintained (i.e. can pass through the filter)

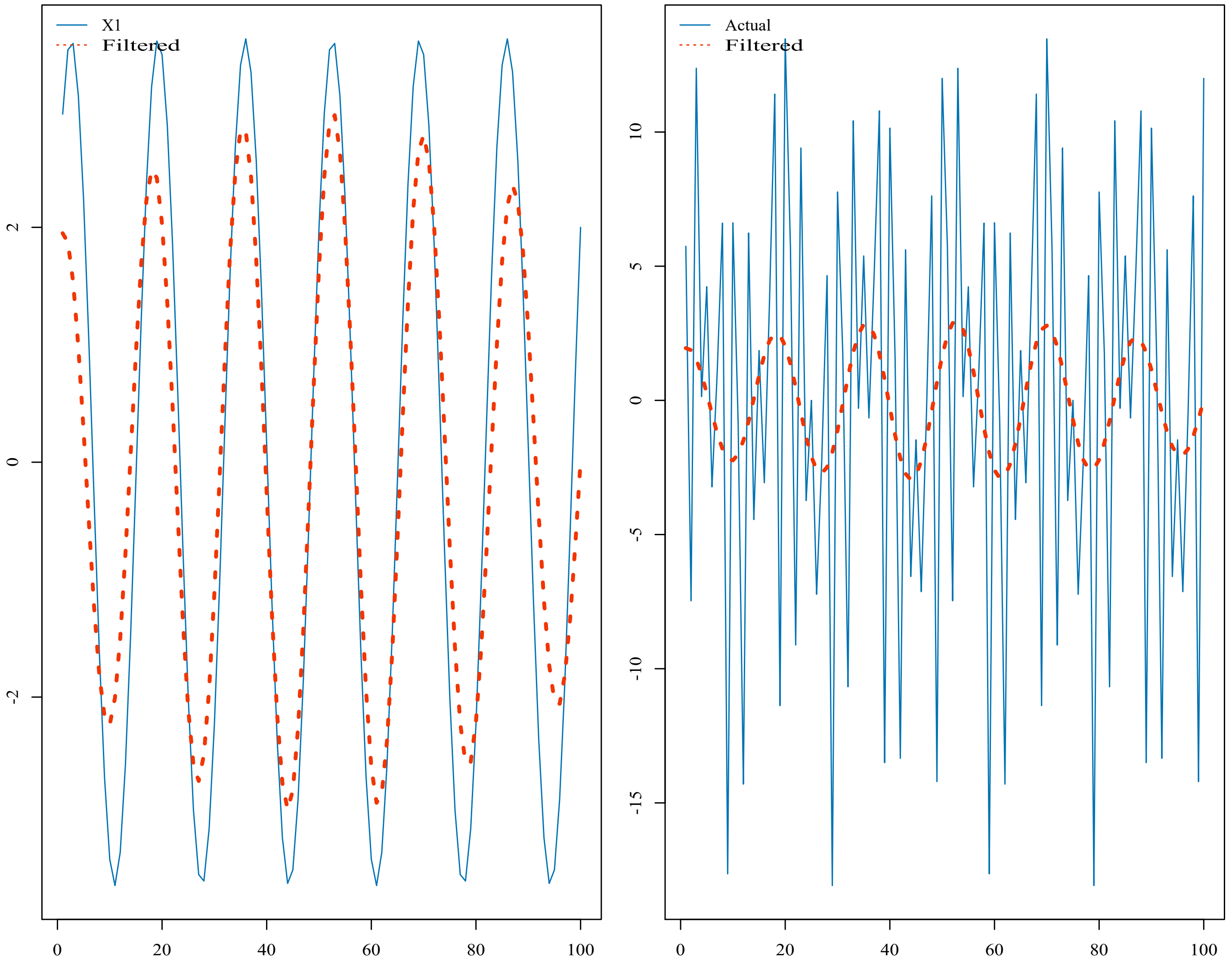

Figure : Filtered result for frequency components

Methods for decomposing a time series

- Various detrending methods provide different estimates of the cycle

- Appropriate transformation should depend on the underlying dynamic properties
  - Usually a good idea to consider whether a series has a stochastic trend (i.e. unit root)

- Assume that an economic time series can be decomposed into trend, g_t, and cycle, c_t:

$$y_t = g_t + c_t$$

- where we abstract from a noise and seasonal component

- Estimates of g_t and c_t may be obtained from various univariate detrending methods

Deterministic trends & filters

- Early methods to decompose economic variables assumed that the (natural) growth path for the economy was largely deterministic

- Therefore, the trend cycle decomposition was described as,

$$
\begin{aligned}
y_t &= g_t + c_t \\
\hat{g}_t &= \widehat{\alpha}_0 + \widehat{\alpha}_1 t + \widehat{\alpha}_2 t^2 + \ldots \\
\hat{c}_t &= y_t - \hat{g}_t
\end{aligned}
$$

  - where the trend, $\hat{g}_t$, is found by simple estimation techniques
  - cycle corresponds to the residual in the series

- When we assume a linear trend, $|\alpha_1| > 0$ and $\alpha_2 = 0$

- For a quadratic trend, $|\alpha_1| > 0$ and $|\alpha_2| > 0$

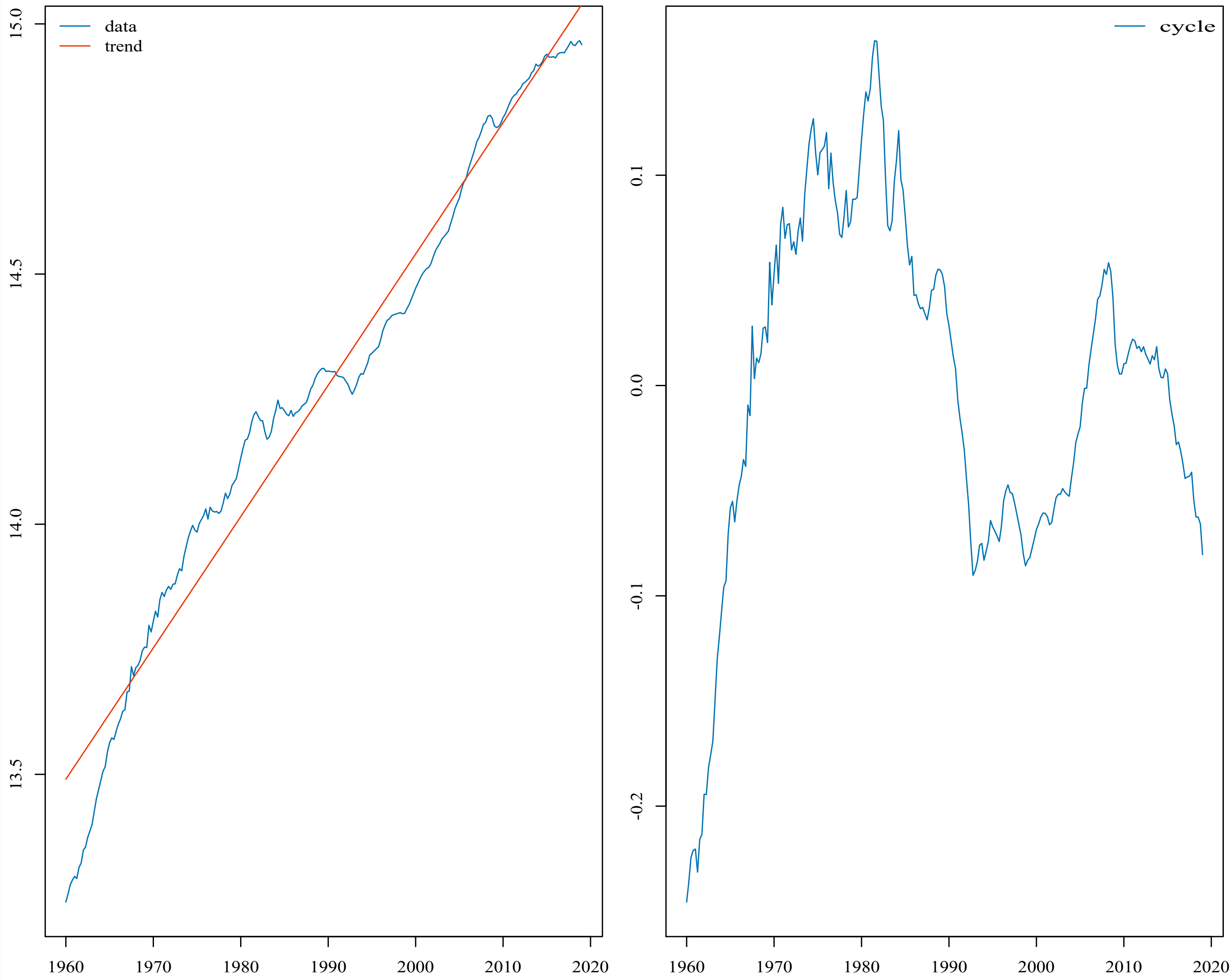

Figure : Linear decomposition - SA output (1960Q1-2018Q4)

Deterministic trends & filters

- Previous graph displays the logarithm of South Africa GDP with linear trend and cycle
- However, productivity growth has not been perfectly log-linear (i.e. constant growth rate) and far from smooth
- In addition, there are several structural breaks such as the oil price shock in 1973/1974 & recent GFC
- To allow for a possible structural break in the trend, we could estimate,

$$\hat{g}_t = \widehat{\alpha}_0 + \widehat{\alpha}_1 t + \widehat{\alpha}_2 DS_t(j) + \widehat{\alpha}_3 DL_t(k) + \ldots$$

- where $DS_t(j)$ and $DL_t(k)$ are dummy variables that capture the change in the slope or the level of the trend in periods j and k
- Hence, $DS_t(j) = t - j$ and $DL_t(k) = 1$, if $t > j$ or $t > k$, while it would be zero otherwise

Deterministic trends & filters

- Identifying structural breaks could be problematic
- Detrending an integrated process with a deterministic trend may result in the introduction of a spurious cycle
- See Nelson-King (1981) for details

Stochastic trends & filters

- Let g_t represent a moving average of observed y_t
- We can extract the trend component, g_t, by applying

$$g_t = \sum_{j=-m}^{n} \omega_j y_{t-j}$$

- where m and n are positive integers and ω_j are weights in the $G(L)$ polynomial

$$G(L) = \sum_{j=-m}^{n} \omega_j L^j$$

- where L is defined so that $L^j y_t = y_{t-j}$ for positive and negative values for j

Stochastic trends & filters

- The cyclical component is the difference between y_t and g_t

$$c_t = [1 - G(L)]\, y_t \equiv C(L)\, y_t$$

- where $C(L)$ and $G(L)$ are linear filters
- Weights are chosen to add up to one, $\sum_{j=-m}^{n} c_j = 1$, so that the level of the series is maintained
- The moving-average filter with weight $1/5$, may be obtained by filtering over the {moving window} of five observations

$$g_t = \frac{1}{5} \sum_{j=-2}^{2} y_t = \frac{1}{5} \left(y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2} \right)$$

- Will produce a smooth stochastic trend if the underlying data has such a trend

Hodrick-Prescott (HP) filter

- The HP filter has been a widely used approach to extract cycles in economic data
- Extracts a stochastic trend, g_t, for a given value of λ, which is the smoothing parameter
- Seeks to emphasize true business cycle frequencies
- The filter can be obtained as the solution to the following problem:

$$\min_{g_t} \sum_{t=1}^{T} \left[(y_t - g_t)^2 + \lambda \Big\{ (g_{t+1} - g_t) - (g_t - g_{t-1}) \Big\}^2 \right]$$

- This minimization problem has a unique solution, and the filtered series, g_t has the same length as y_t
- Termed a low-pass filter as it only models low frequency data

Hodrick-Prescott (HP) filter

- The smoothness is determined by λ, which penalizes the acceleration in the growth component
- If $\lambda \rightarrow \infty$, the lowest minimum is achieved when variability in the trend is zero (as in the case of a linear trend)
- If $\lambda = 0$ there is more variation in the trend, such that there will be no cycle
- Hodrick and Prescott argue that $\lambda = 1600$ is a reasonable choice for U.S. quarterly data
- However, it is not necessarily the case that it can be universally applied, to other variables or output of other components
- Specifies a typical cycle of between eight to ten years when using traditional values for λ

Hodrick-Prescott (HP) filter

- Another concern is the end-of-sample problem:
  - trend is close to observed data at the beginning and end of the sample
  - problematic when we are at the peak of a cycle
  - some researchers use forecasts to generate additional data at the end of the series
- King & Rebelo (1993) note that the HP-filtered cyclical component contains both forward and backward differences
- As a result, the end of sample properties are poor when you do not have an observation for $t+1$ or $t-1$
- Method is also criticised on the basis that the smoothness of the stochastic trend component has to be determined *a priori*

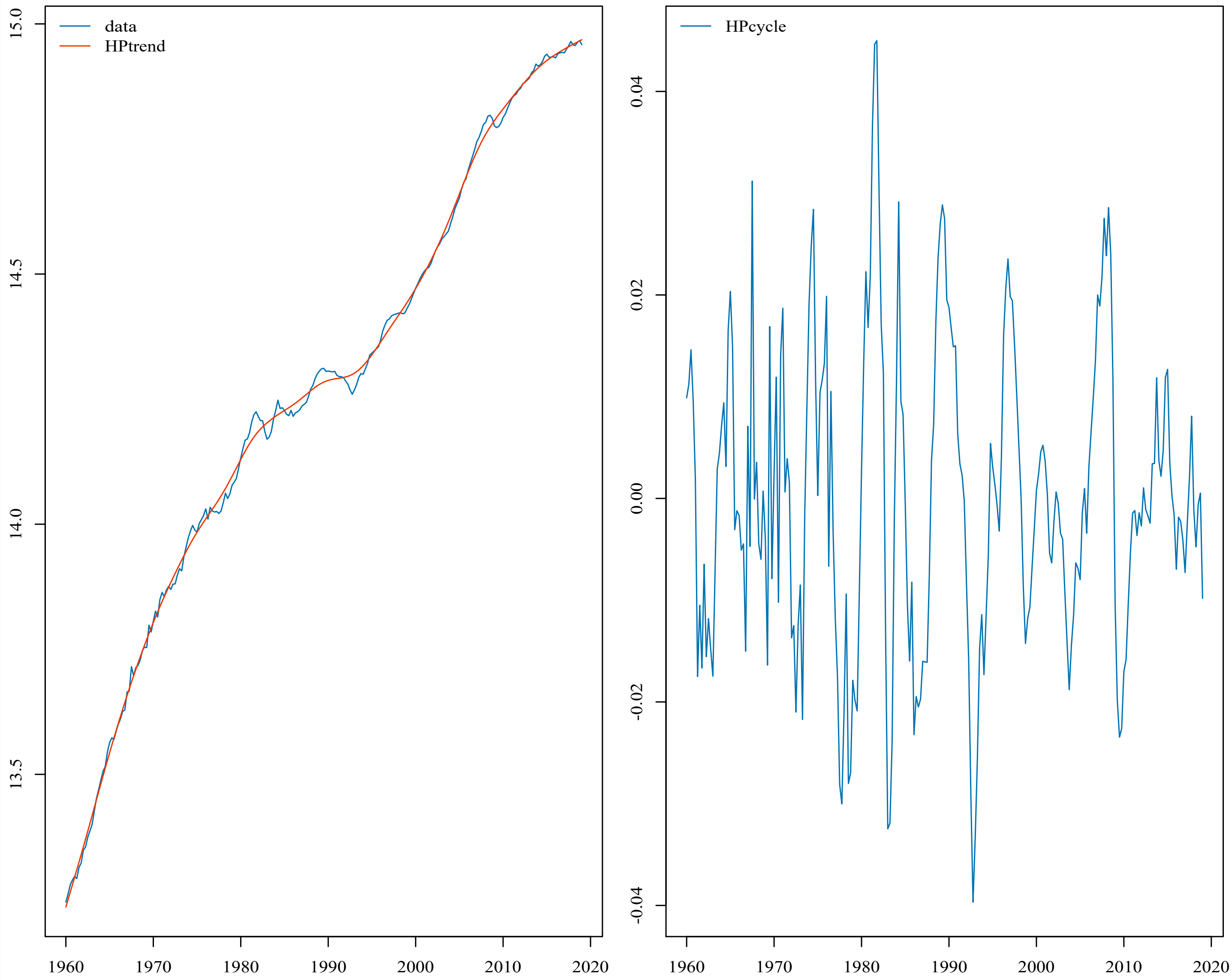

Figure : HP filter - SA output (1960Q1-2018Q4)

Band pass filters

- Band pass filters introduced to economic data by Baxter & King (1999) and Christiano & Fitzgerald (2003)
- Identify all components that correspond to the chosen frequency band that has an upper and lower limit
- Need to determine the periodicity of the business cycles one wants to extract
- This is usually expressed within the frequency domain

Frequency-domain

- Consider a time series,

$$y_t = A\cos(2\pi\omega t)$$

- where A is the amplitude (height) of the cycle
- ω is the frequency of oscillation (the number of occurrences of a repeating event per unit of time)
- 2π measures the period of the cycles
- t is the time
- Hence, if $y_t = A\cos(2\pi t)$, we will observe one cycle over the data sample
- By increasing ω, we increase the number of cycles

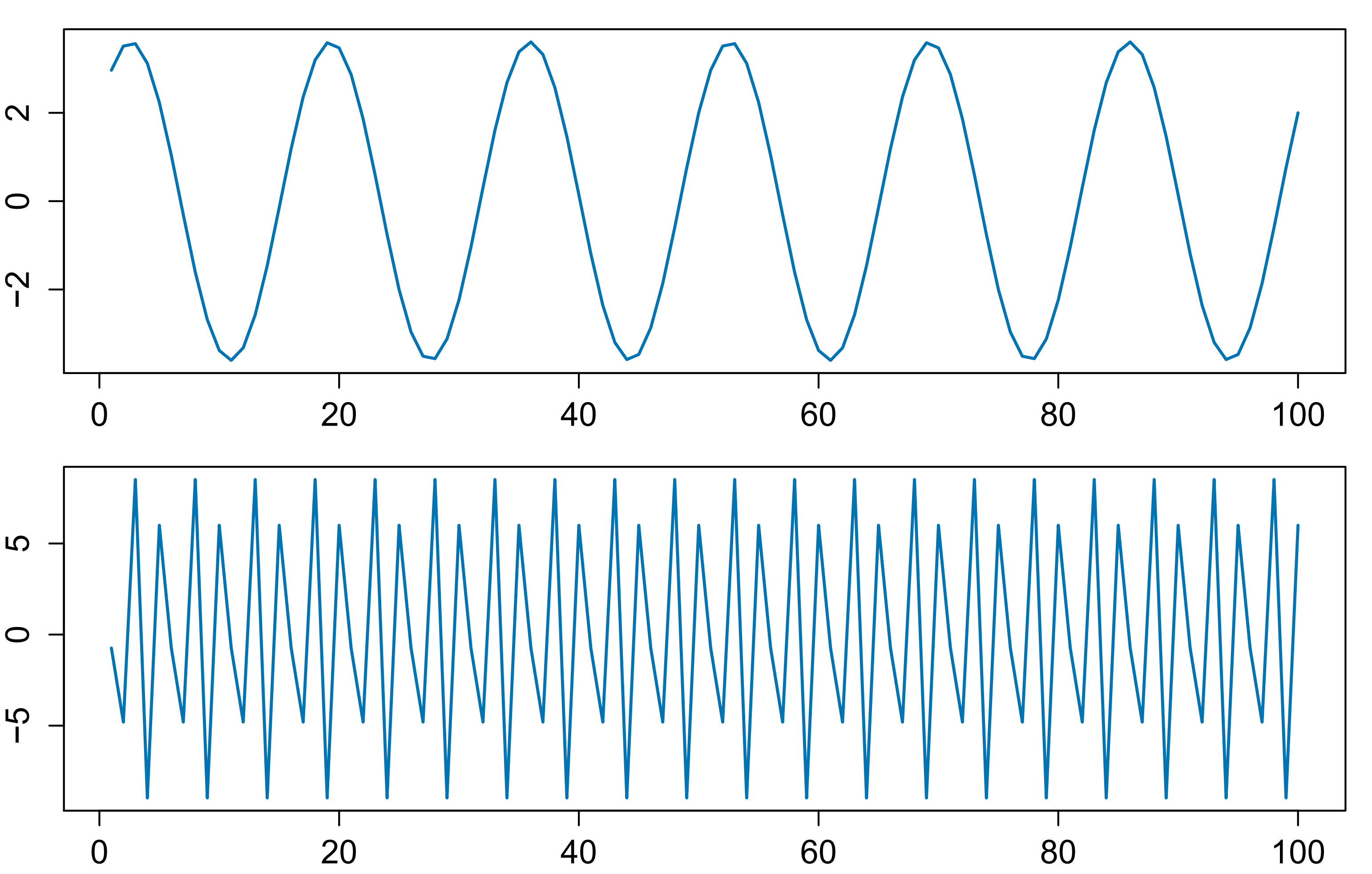

Figure : Artificial Data

Band Pass Filters

- An intuitive measure of frequency is the amount of time that elapses per cycle, λ

$$\lambda = 2\pi/\omega$$

- Where we have quarterly data, for ω corresponding to a cycle length of 1.5 years
- Set $\lambda = 6$ quarters per cycle and solve for $6 = 2\pi/\omega_h$

$$\omega_h = 2\pi/6 = \pi/3$$

- Similarly, the frequency corresponding to a low frequent cycle length of 8 years is:

$$\omega_1 = 2\pi/32 = \pi/16$$

Baxter and King (1999)

- Baxter and King (1999) decompose a time series into three periodic components: trend, cycle, and irregular fluctuations
- Business cycles were defined as periodic components whose frequencies lie between 1.5 and 8 years per cycle
- Periodic components with lengths longer than 8 years were identified with the trend
- Periodic components of less than 1.5 years were identified with the irregular component

$$
\begin{aligned}
B(\omega) &= 1 \text{ for } \omega \in [\pi/16, \pi/3] \text{ or } [-\pi/3, -\pi/16] \\
&= 0 \text{ otherwise}
\end{aligned}
$$

- Hence, the interval $B(\omega) = [\pi/16, \pi/3]$ can be interpreted as the business cycle frequency
- The interval $[0, \pi/16]$ corresponds to the trend and $[\pi/3, \pi]$ defines irregular fluctuations

Band Pass Filters

- While Baxter and King favour a 3-part decomposition, other economists prefer a two-part classification
- This may be incorporated in this setup, where

$$\begin{aligned} H(\omega) &= 1 \text{ for } \omega \in [\pi/16, \pi] \text{ or } [\text{-}\pi, -\pi/16] \\ &= 0 \text{ otherwise} \end{aligned}$$

- The trend component is still defined in terms of fluctuations lasting more than 8 years
- Cyclical component now consists of all oscillation lasting 8 years or less
- This is known as a high pass filter, as only higher frequency components are captured in $H(\omega)$
- As with the HP filter one has to decide on the preferred frequencies for the cycles *a priori*
- There are potential end-of-sample problems, but usually eliminate the estimates at the start and end

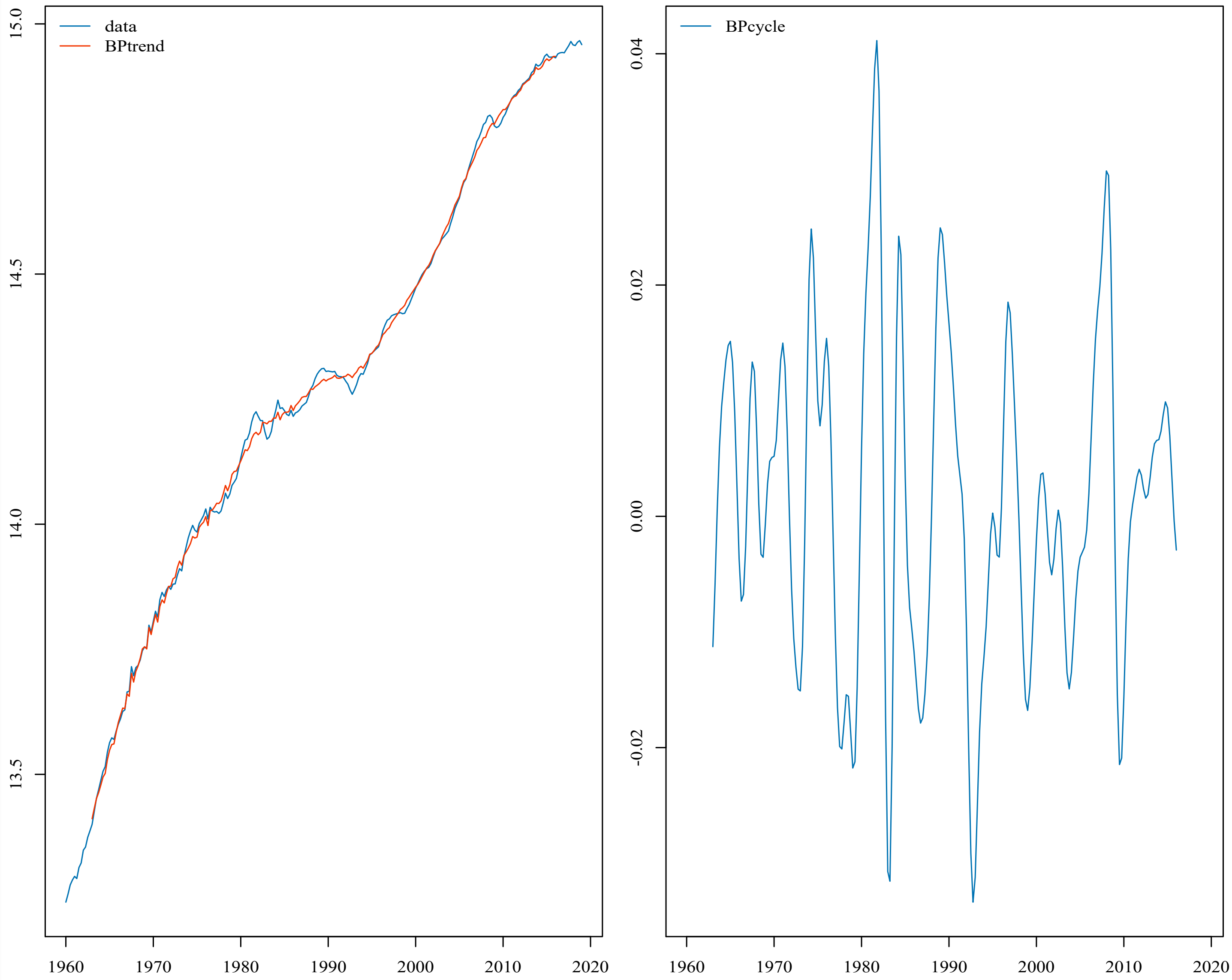

Figure : BP filter - SA output (1960Q1-2018Q4)

Beveridge-Nelson Decomposition

- Beveridge & Nelson (1981) model the trend as a random walk with drift, and the cycle is treated as stationary process with zero mean
- To perform this decomposition, let y_t be integrated of first order, so that its first difference, Δy_t are stationary
- Assume it has the following moving average representation

$$(1-L)y_t = \Delta y_t = \mu + B(L)\varepsilon_t$$

Beveridge-Nelson Decomposition

- The BN decomposition explores the following
- First, define the polynomial,

$$B^*(L) = (1 - L)^{-1}[B(L) - B(1)]$$

- where $B(1) = \sum_{s=0}^{\infty} B_s$
- Rewriting this polynomial in terms of $B(L)$, gives

$$B(L) = [B(1) + (1 - L)B^*(L)]$$

- and substituting into the above yields

$$\Delta y_t = \mu + B(L)\varepsilon_t = \mu + [B(1) + (1 - L)B^*(L)]\varepsilon_t$$

Beveridge-Nelson Decomposition

- For the decomposition, $y_t = g_t + c_t$, it follows that $\Delta y_t = \Delta g_t + \Delta c_t$
- Therefore a change in the trend component of y_t equals

$$\Delta g_t = \mu + B(1)\varepsilon_t$$

- and the change in the cyclical component is,

$$\Delta c_t = (1 - L)B^*(L)\varepsilon_t$$

- where we see that the trend follows a random walk with drift
- This expression can be solved to yield

$$g_t = g_0 + \mu t + B(1) \sum_{s=1}^{t} \varepsilon_s$$

Beveridge-Nelson Decomposition

- As such, the trend consists of both a deterministic term

$$g_0 + \mu t$$

  and a stochastic term

$$B(1)\sum_{s=1}^{t}\varepsilon_s$$

- For $B(1) = 0$, the trend reduces to a deterministic case
- where for $B(1) \neq 0$, the stochastic part indicates the long-run impact of a shock ε_t on the level of y_t

Beveridge-Nelson Decomposition

- The cyclical component is stationary and is given by

$$c_t = B^*(L)\varepsilon_t = (1-L)^{-1}[B(L) - B(1)]\varepsilon_t$$

- Beveridge & Nelson (1981) showed that the stochastic trend could also be interpreted as the long-term forecast from RW plus drift model
- Cycle is the stationary process that reflects the deviations from the trend

Beveridge-Nelson Decomposition

- To estimate the BN decomposition in practice, assume an $AR(1)$ process for the growth rate of output

$$\Delta y_t = \phi \Delta y_{t-1} + \varepsilon_t,$$

- where we ignore the constant term
- Assuming $\phi < 1$, the $AR(1)$ process can be written in terms of the infinite order $MA(q)$ process where we find $B(L)$, $B(1)$ and $B^*(L)$ as

$$B(L) = \frac{1}{1 - \phi L}$$

$$B(1) = \frac{1}{1 - \phi}$$

$$B^*(L) = (1 - L)^{-1}[B(L) - B(1)] = \frac{\phi}{(1 - \phi)(1 - \phi L)}$$

Beveridge-Nelson Decomposition

- Solving in terms of y_t

$$y_t = (1-L)^{-1}[B(1) + (1-L)B^*(L)]\varepsilon_t$$

- which can be rewritten as

$$y_t = B(1)(1-L)^{-1}\varepsilon_t + (1-L)^{-1}[B(L) - B(1)]\varepsilon_t$$

- Substituting in for the $AR(1)$ solution derived above, we have

$$\begin{aligned} y_t &= g_t + c_t \\ &\Downarrow \\ y_t &= \frac{1}{1-\phi}(1-L)^{-1}\varepsilon_t + \frac{-\phi}{(1-\phi L)(1-\phi)}\varepsilon_t \end{aligned}$$

Beveridge-Nelson Decomposition

- The advantage of Beveridge-Nelson method is that it is appropriate when a series is difference-stationary
- It also allows the series to contain a unit root that can be highly volatile
- However, it has the disadvantage of being rather time-consuming to compute
- In addition, one has to choose between different $ARMA(p, q)$ models that may give quite different results
- Misrepresenting an $I(2)$ process as an $I(1)$ process may generate excess volatility in the trend

Figure : Evaluation of decompositions - SA output (1960Q1-2018Q4)

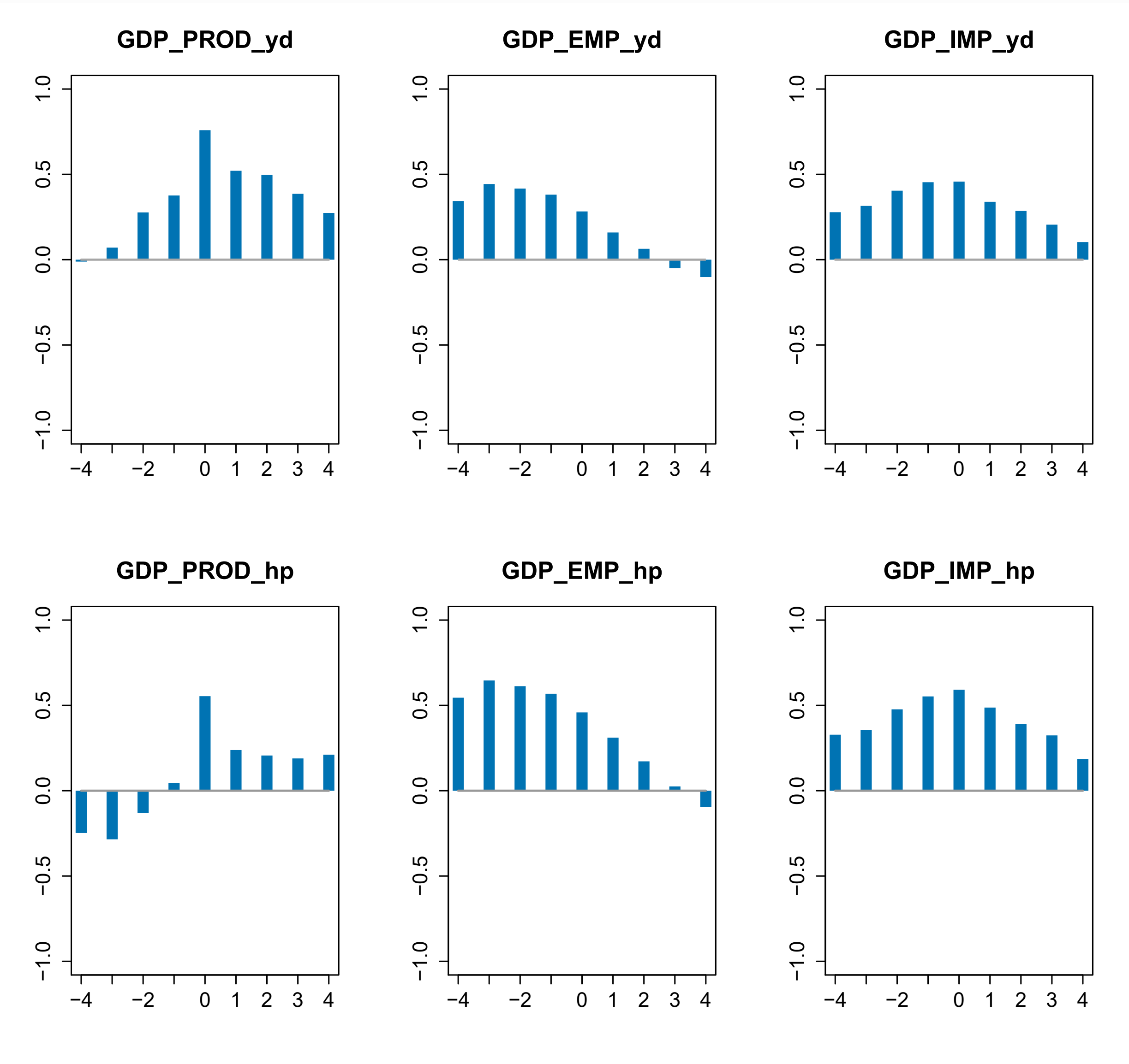

Figure : Leads and Lags - Correlation with GDP

Summary

- Many economic and financial applications make use of decompositions for nonstationary time series, which are transformed into a permanent and a transitory component
- Could use a linear filter where trend is perturbed by transitory cyclical fluctuations
- The Hodrick-Prescott (HP) filter is the most popular way to extract business cycles
- The HP filter extracts a stochastic trend for a given value of the parameter λ
  - Trend moves smoothly over time and is uncorrelated with the cycle
  - Results are not robust to the value of the smoothness parameter
- Another popular method used to measure the business cycle is the band pass (BP) filter
- The filter removes (filters out) all the components in a series except those that correspond to the chosen frequency band

Summary

- In the Beveridge and Nelson (BN) decomposition, the permanent component is shown to be a random walk with drift
- The transitory component is stationary process with zero mean, which is perfectly correlated with the permanent component
- Different decompositions provide different results and should be interpreted with caution
- Usually a good idea to consider different options before drawing conclusions

Wavelet transformations

- Spectral decompositions define the rate at which the time series oscillates
- Results in the loss of all time-based information
- Assumes that the periodicity of all the components is consistent throughout the entire sample
- This may not be the case:
  - Gabor (1946) developed the Short-Time Fourier Transform (STFT) technique
  - Involves the application of a number of Fourier transforms to different subsamples
  - Precision of the analysis is affected by the size of the subsample
- Large subsample to identify changes in low frequency
- Small subsamples to identify changes in higher frequency

Wavelet transformations

- Wavelet transformations capture features of time-series data across different frequencies that arise at different points in time
- Wavelet functions are stretched and shifted to describe features that are localised in frequency and time
  - Could be expanded over a relatively long period of time when identifying low-frequency events
  - Could be relatively narrow when describing high frequency events
- Involves shifting various wavelet functions with different amplitudes over the sample of data
- One is then able to associate the components with specific time horizons that occur at different locations in time
- Wavelets use scales rather than frequency bands, where the highest scale refers to the lowest frequency

Wavelet transformations

- Early work with wavelet functions dates back to Haar (1910)
- See, Hubbard (1998) and Heil (2006) for a detailed account of the history of wavelet analysis
- For computation most studies currently employ the multiresolution decomposition of Mallat (1989) and Strang (1996)
- Early applications of wavelet methods in economics include the work of Ramsey (1997), which made use of a wavelet decomposition of exchange rate data to describe the distribution of this data at different frequencies

Wavelet transformations

- To describe this technique, consider a variable that is composed of a trend and a number of higher-frequency components
- The trend may be represented by a father wavelet, $\phi(t)$
- The mother wavelets, $\psi(t)$, are used to describe information at lower scales (i.e. higher frequencies)
- One could then describe variable x_t as

$$x_t = \sum_k s_{0,k}\phi_{0,k}(t) + \sum_{j=0}^{J}\sum_k d_{J,k}\psi_{J,k}(t)$$

- where J refers to the number of scales, and k refers to the location of the wavelet in time
- The $s_{0,k}$ coefficients are termed smooth coefficients, since they represent the trend, and the $d_{J,k}$ coefficients are termed the detailed coefficients, since they represent finer details in the data

Wavelet transformations

- Mother wavelet functions, $\psi_{J,k}(t), \ldots, \psi_{1,k}(t)$, are then generated by shifts in the location of the wavelet in time and scale

$$\psi_{j,k}(t) = 2^{-j/2}\psi\left(\frac{t - 2^j k}{2^j}\right), \quad j = 1, \ldots, J$$

- where the shift parameter is represented by $2^j k$ and the scale parameter is 2^j
- As depicted in the daublet wavelet functions, smaller values of j (which produce a smaller scale parameter 2^j), would provide the relatively tall and narrow wavelet function on the left
- Larger values of j, the wavelet function is more spread out and of lower amplitude
- After shifting this function by one period, we produce the function that is depicted on the right

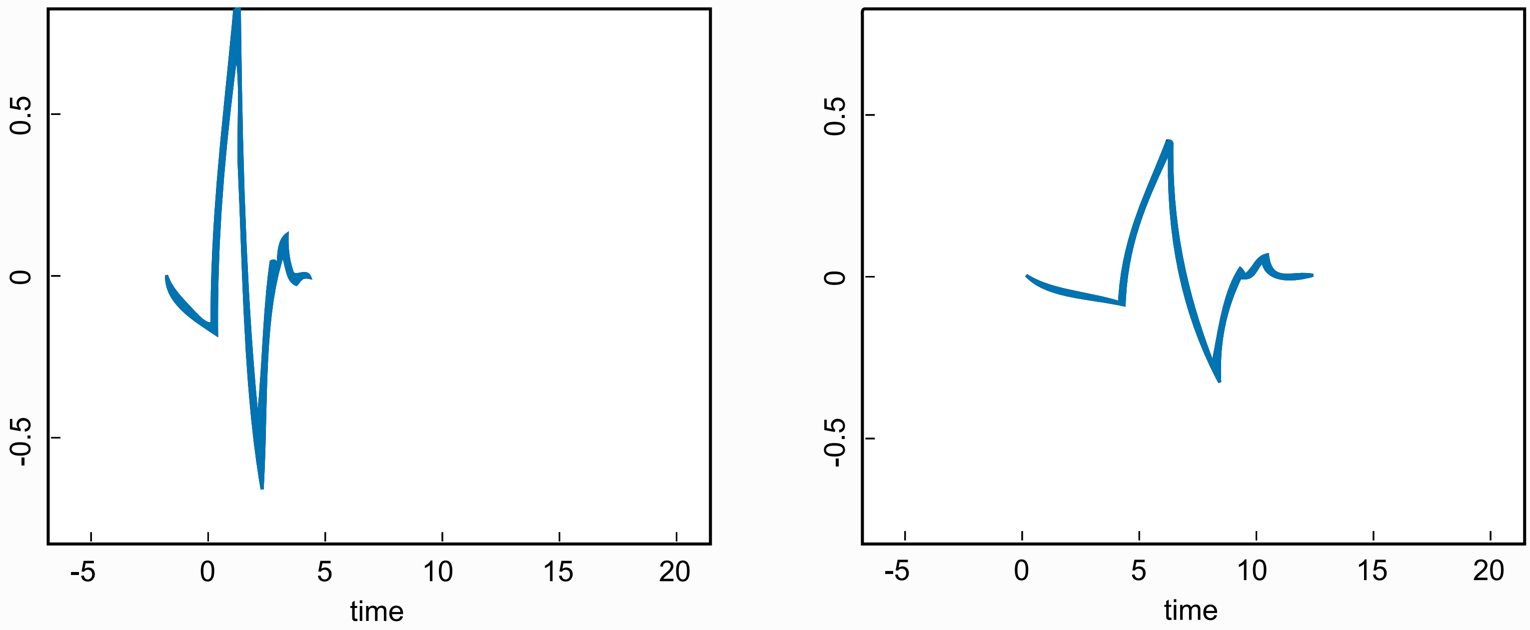

Figure : Daublet (4) wavelet functions - $\psi_{1,0}(t)$ and $\psi_{2,1}(t)$

Wavelet transformations

- Wavelet functions may be:
  - Smooth - decompose data into trend, cycle, noise (or various cycles)
  - Peaked - identify peak and trough of cycle
  - Square - identify structural breaks
- Use smooth functions that include daublets, coiflets and symlets
- Multiresolution techniques are used for computation, which includes the maximum overlap discrete wavelet transform (MODWT)

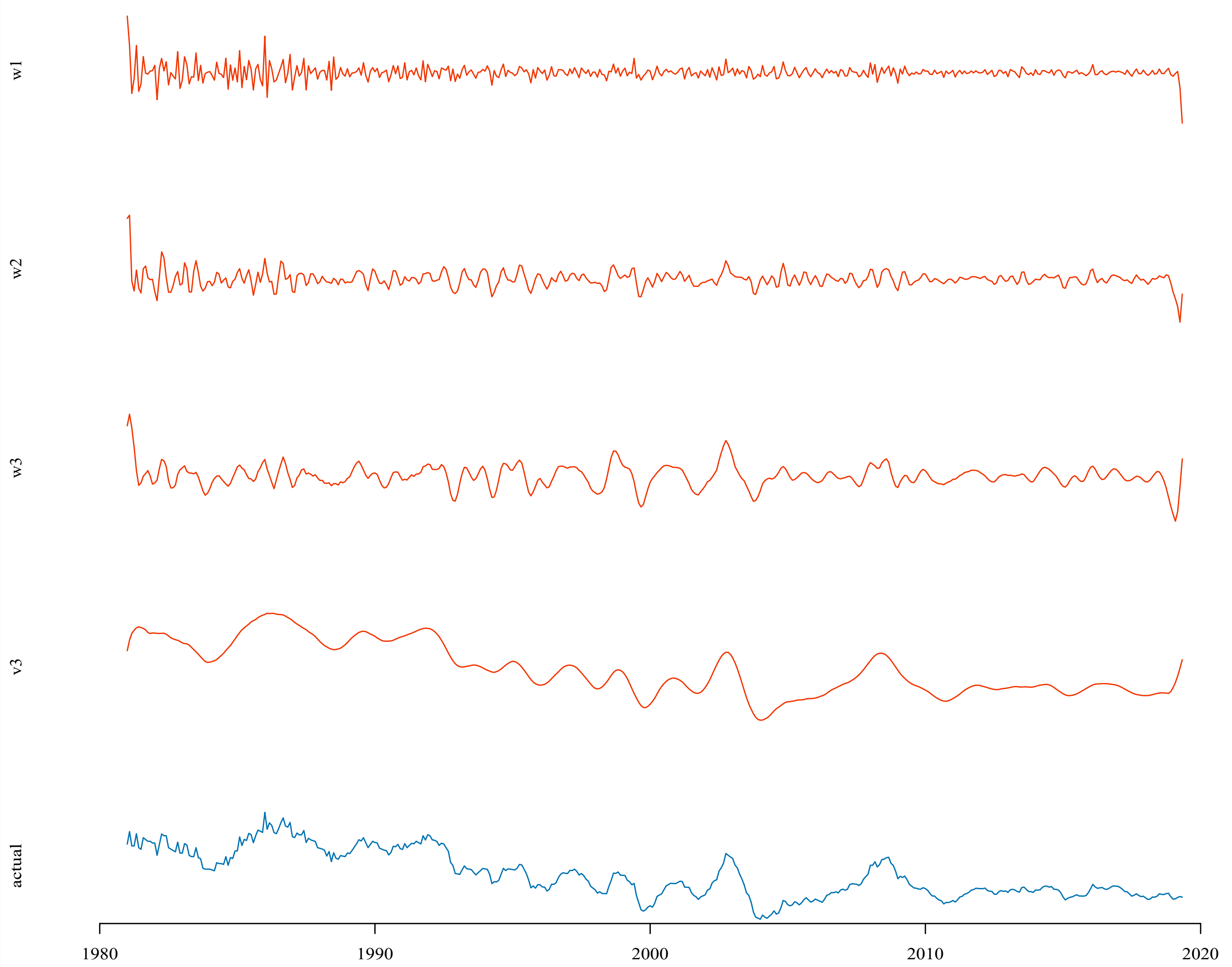

Figure : Daublet (4) wavelet decomposition - South African inflation

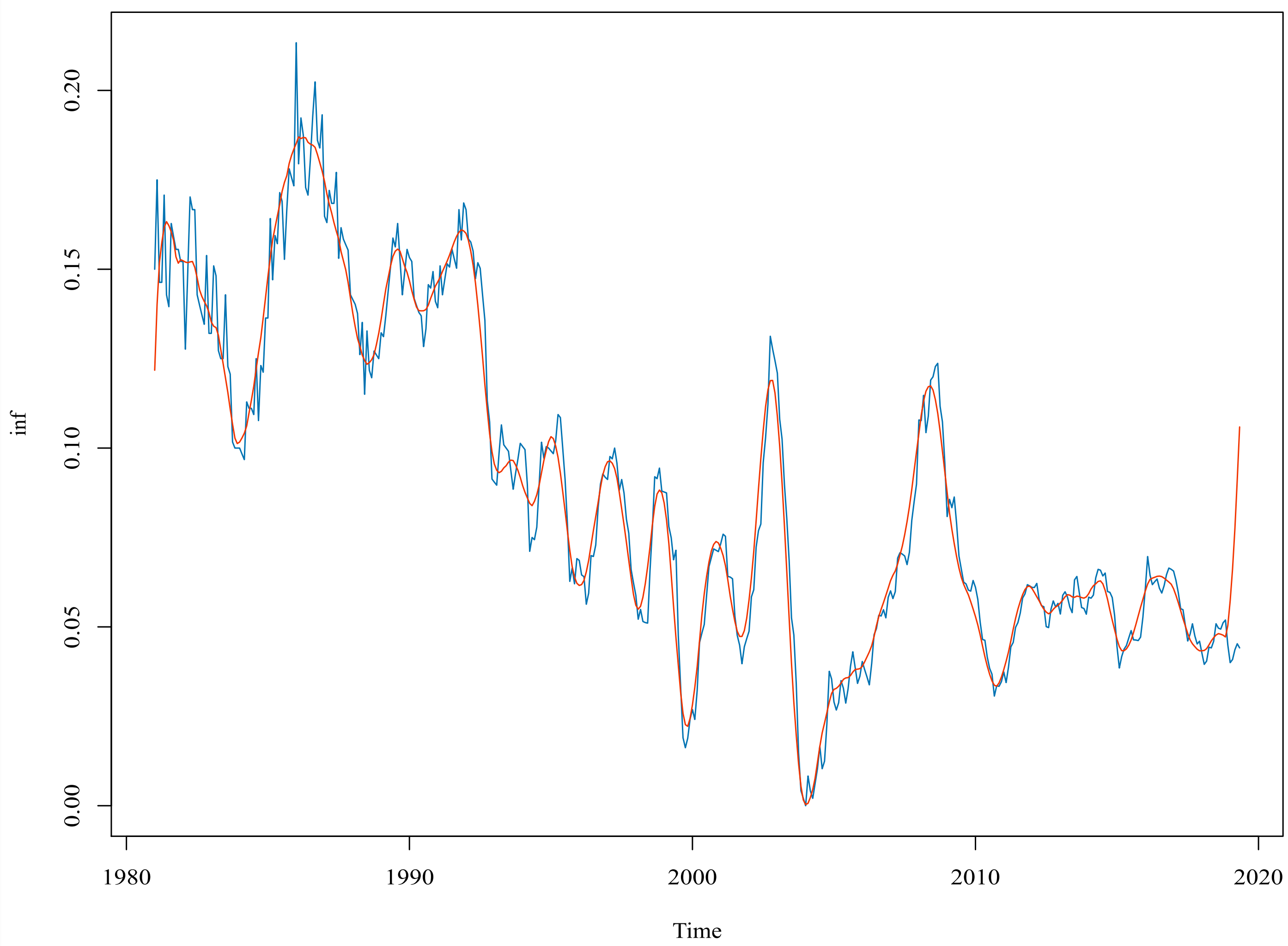

Figure : Daublet (4) wavelet decomposition - South African inflation

Wavelet transformations – Summary

- Advantages:
  - Can be applied to data of any integration order
  - Has the benefits of spectral techniques without losing time support (very useful when identifying changes in the process at different frequencies)
  - Can include a number of bands, which are additive