

Decomposing Time Series Data

by Kevin Kotzé

Most economic time series exhibit behaviour that is repeated over time. This allows for these processes to be modelled with the aid of techniques that consider the evolution of the process over a period of time. The application of this methodology would usually refer to the use of techniques that were developed in the time domain and include the general specification of ARIMA and state-space models. Another important phenomena of time series variables is that they may be decomposed into different periodic variations.

For example, we may wish to extract the cyclical component of the time series, which may be regarded as the part that exhibits higher periodic variation than the trend. When we are looking to identify the cyclical component of economic output, we are able to conduct an investigation into the behaviour of the *business cycle*, which is the topic of a number of empirical macroeconomic research investigations. These decompositions may be used to describe the stylized facts of the business cycle literature, such as the persistence of economic fluctuations and correlations (or lack thereof) between the cyclical components of economic aggregates in different countries.¹

It is important to note that when seeking to derive a measure of the business cycle, there are no unique periodicities that may be used to identify this particular part of the process. This is partly due to the fact that it may have a relatively vague definition (from a purely statistical perspective) and may be considerably different when comparing the exact definition of the business cycle in a selection of countries. In addition, most economic time series are both fluctuating and growing, which makes these types of decompositions quite challenging, particularly when trying to perform this decomposition on the data of an emerging market economy.² In addition, these procedures may also be used to identify the output gap, which is the difference between potential and actual output.

When seeking to decompose a time series into different periodic variations, we could imagine that the process is responding to various driving frequencies that are produced from linear combinations of sine and cosine functions, where each of these sine and cosine functions would represent a different frequency or amplitude. Expressed in these terms, the application of frequency domain techniques may be regarded as a regression of periodic sine and cosines on the respective values of the time series.³

Therefore, this chapter considers the application of a few widely used methods that are used to decompose economic time series. It includes a comparison of several detrending methods that are used to extract the business cycle, including deterministic detrending, stochastic detrending and frequency filtering techniques. The appendix includes a discussion of techniques that have been developed more recently, which exist in the time-frequency domain.

1 Spectral analysis

Consider an example where we are given three quarterly time series variables, y_t , x_t and v_t , and the term $v_t \sim \text{i.i.d.}\mathcal{N}(0, \sigma)$. If it is the case that $x_t = y_t + v_t$, 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.

As noted in the introduction, an observed time series could be viewed as the weighted sum of a number of underlying series that have different periodic behaviour. Hence, the cumulative variation in an observed time series will then be the sum of the contributions of these underlying series, which may vary in different *frequencies*. Spectral analysis is a tool that can be used to decompose a time series into different frequency components, where we may be interested in the respective contribution that has been made by various periodic components. When compared to the initial example that we used, a frequency domain analysis would involve a regression of a time series variable, x_t , on a number of periodic components that have different frequencies to investigate whether or not information relating to each of the frequency components is contained in x_t .

The general notion of periodicity can be made more precise by introducing some terminology. In order to define the rate at which a series oscillates, we first define a cycle as one complete period of a sine or cosine function defined over a unit time interval. We can then consider the periodic process

$$y_t = A \cos(2\pi\omega t + \phi) \quad (1.1)$$

for $t = 0, \pm 1, \pm 2, \dots$, where ω is a frequency index, defined in cycles per unit time, while A determines the height or amplitude of the function. The starting point of the cosine function is termed the phase and is denoted ϕ . We could then introduce random variation in this time series by allowing the amplitude and phase to vary randomly. 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,⁴

$$y_t = U_1 \cos(2\pi\omega t) + U_2 \sin(2\pi\omega t) \quad (1.2)$$

where $U_1 = A \cos \phi$ and $U_2 = -A \sin \phi$ are often taken to be normally distributed random variables. In this case, the amplitude is $A = \sqrt{U_1^2 + U_2^2}$ and the phase is $\phi = \tan^{-1}(-U_2/U_1)$. From these facts we can show that if, and only if, in (1.1), A and ϕ are independent random variables, where A^2 is chi-squared with 2 degrees of freedom, and ϕ is uniformly distributed on $[-\pi, \pi]$, then U_1 and U_2 are independent, standard normal 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 time, or in cycles per point in the above illustration. For $\omega = 1$, the series makes one cycle per time unit; for $\omega = .50$, the series makes a cycle every two time units; for $\omega = .25$, every four units, and so on. In general, for data that occur at discrete time points will need at least two points to determine a cycle, so the highest frequency of interest is $\frac{1}{2}$ cycles per point. This frequency is called the folding frequency and defines the highest frequency that can be seen in discrete sampling. Higher frequencies sampled this way will appear at lower frequencies, called *aliases*; an example is the way a camera samples a rotating wheel on a moving vehicle in a movie, in which the wheel appears to be rotating at a different rate.

Consider a generalization of (1.2) that allows mixtures of periodic series with multiple frequencies and amplitudes,

$$y_t = \sum_{k=1}^q [U_{k1} \cos(2\pi\omega_k t) + U_{k2} \sin(2\pi\omega_k t)], \quad (1.3)$$

where U_{k1}, U_{k2} , for $k = 1, 2, \dots, q$, are independent zero-mean random variables with variances σ_k^2 , and the ω_k are distinct frequencies. Notice that (1.3) exhibits the process as a sum of independent components, with variance σ_k^2 for frequency ω_k . Using the independence of the U_s and the

trigonometry identities, it is easy to show that the autocovariance function of the process is

$$\gamma(h) = \sum_{k=1}^q \sigma_k^2 \cos(2\pi\omega_k h), \quad (1.4)$$

and we note the autocovariance function is the sum of periodic components with weights proportional to the variances σ_k^2 . Hence, y_t is a mean-zero stationary processes with variance which exhibits the overall variance as a sum of variances of each of the component parts.

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

$$\begin{aligned} x_{1,t} &= 2 \cos(2\pi t/100) + 3 \sin(2\pi t/100) \\ x_{2,t} &= 4 \cos(2\pi t/30) + 5 \sin(2\pi t/30) \\ x_{3,t} &= 6 \cos(2\pi t/40) + 7 \sin(2\pi t/40) \\ y_t &= x_{1,t} + x_{2,t} + x_{3,t} \end{aligned}$$

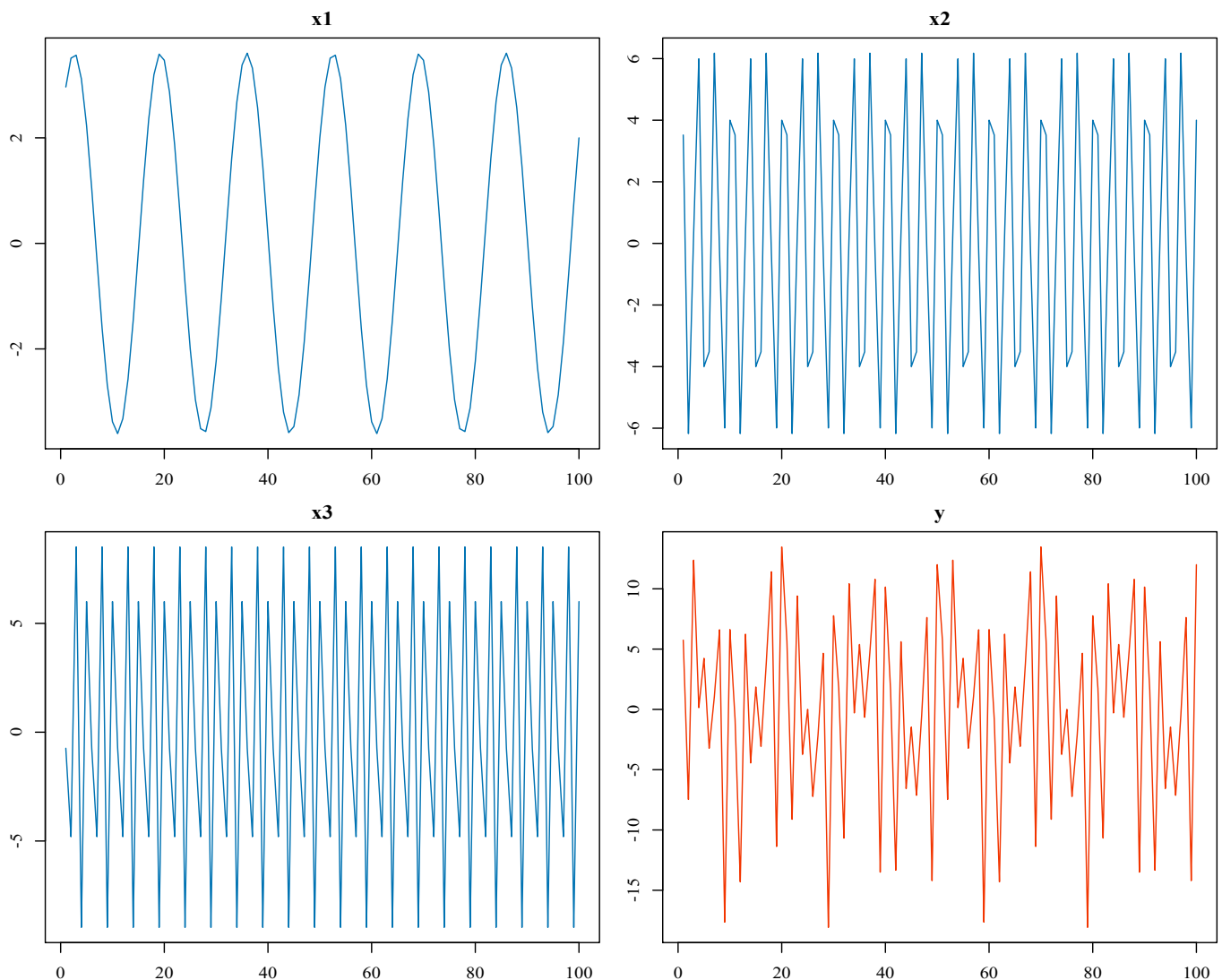


Figure 1: **Different frequency components**

Where the first three series $\{x_{1,t}, x_{2,t}, x_{3,t}\}$, are generated with different periodicities and amplitude, and the fourth series, y_t , is the sum of the three others. The series are plotted in Figure 1, where we display the actual series over time. These graphs clearly see how x_1 displays the longer waves (as it

has the low frequency component), followed by x_2 and x_3 . The y series is the sum of the other three series, and will therefore *inherit* the properties of the $\{x_{1,t}, x_{2,t}, x_{3,t}\}$ series. Note that it is not easy to visualize the periodicities from this time series plot.

The systematic 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,

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

and 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. The periodogram may be computed quickly using the fast Fourier transform (FFT), and there is no need to run repeated regressions. An example of a periodogram is provided in Figure 2, where x_1 has a peak in the periodogram to the left, followed by x_2 and x_3 , which has a peak that is to the right of the other two. Hence, x_3 is the most high frequent of the three series. Note that in this case, it is easy to visualize the three periodic components that are in y_t , as each corresponds to the original time series component. Hence, we confirm that the y -series has inherited the periodicity of the three other x -series.

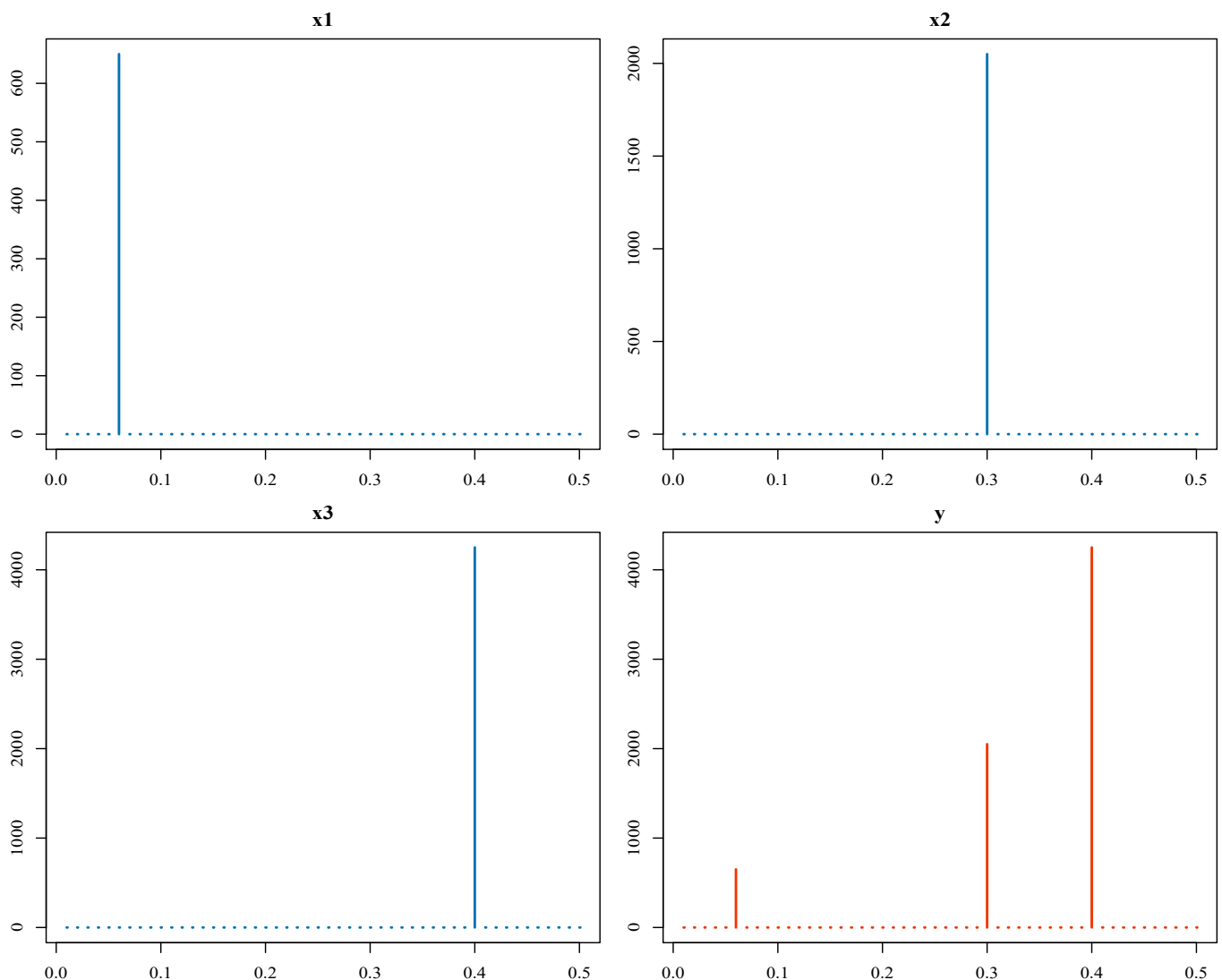


Figure 2: **Periodogram for frequency components**

Note that the horizontal scale of the periodogram represents $P(j/n)$ for $j = 0, 1, \dots, n - 1$. These values are related to the size of ω , since

$$P(6/100) = 0.06, P(30/100) = 0.3, P(40/100) = 0.4,$$

Hence, where the specific frequency is present in a time series, the value for $P(j/n) \neq 0$ in all other cases, $P(j/n) = 0$. Note, that in this example, these values for the respective frequencies of the components are displayed in Figure 1, which suggests that the periodogram may provide some insight into the variance components of any data. In addition, the vertical scale of the periodogram is a function of the amplitude, A , and may be used to show the relative strength of cosine-sine pairs at various frequencies in the overall behaviour of the time series.

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 that 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).

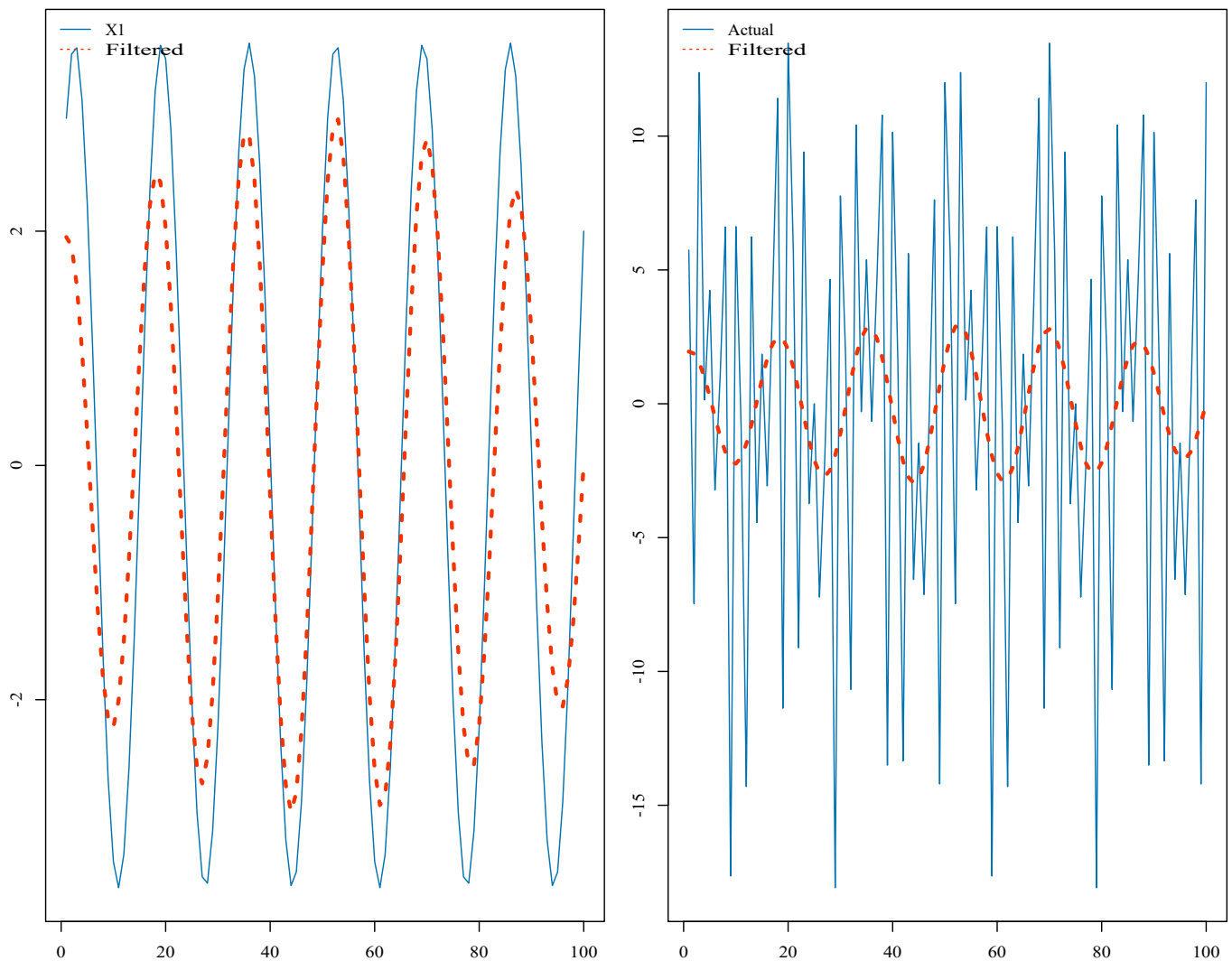


Figure 3: **Filtered result for frequency components**

Figure 3 displays the original series y and the new filtered series. From the periodogram we see that the filtered series is almost equal to the original x_1 series. Hence, the filter removes the high frequent components (corresponding to x_2 and x_3), and we are left with the periodic component that

corresponds to the original series x_1 .

While this example is particularly intuitive, it is worth bearing in mind that the simulated series were constructed from known cycles with a particular frequency. However, most of the data that we will encounter will consist of many frequencies, making interpretation more difficult. Still one or two frequencies usually dominate most economic time series, and this is typically what we are looking to identify.

2 Statistical detrending methods

The various detrending methods that are used on economic data provide different estimates of the cycle, where the most appropriate transformation should be determined by the underlying dynamic properties of the variable that is being decomposed.

If we assume that an economic time series can be decomposed into a trend g_t and a cycle c_t , then we could use the following expression for a time series,

$$y_t = g_t + c_t \quad (2.1)$$

where we abstract from the (irregular) noise and seasonal component. The objective of decomposing a time series would be to obtain estimates of g_t and c_t , from the respective detrending methods.

It is worth noting that different methods that may be used to decompose a time series would usually produce different estimates of the cycle.⁵ The ultimate choice of the appropriate transformation of the data should therefore depend on the nature of the underlying dynamic properties of the time series. For example, one may wish to make use of a unit root tests to determine whether the trend in the data is either stochastic or a deterministic. However, such initial testing might not make it obvious which detrending or transformation method should be used, and as such, one way also wish to consider the generally accepted practices that are currently applied in the literature.

In the subsequent sections of this chapter we consider the use of various techniques that may be used to obtain estimates of g_t and c_t , which may be used to extract a deterministic or stochastic trend. Many of these detrending methods are related, where by way of example the Hodrick-Prescott filter with a high smoothing value could be used to identify a linear trend. Similarly, the use of the Hodrick-Prescott filter with a low smoothing value, would provide an estimate of the trend that is largely equivalent to the result that would have been obtained after performing a stochastic Beveridge-Nelson decomposition.

3 Deterministic trends and filters

The traditional methodology for identifying the business cycle, was developed along the premise that there is a natural growth path for the economy, which is perturbed by cyclical fluctuations that are transitory in nature. This definition suggests that the trend could be described by deterministic influences, where the trend and cycle take following form,

$$\begin{aligned} y_t &= g_t + c_t \\ \hat{g}_t &= \hat{\alpha}_0 + \hat{\alpha}_1 t + \hat{\alpha}_2 t^2 + \dots \\ \hat{c}_t &= y_t - \hat{g}_t \end{aligned} \quad (3.1)$$

In this example, the estimated trend is defined by \hat{g}_t , which could be derived with the aid of a linear regression. The cycle would then correspond to the residual that is provided by this model. If the trend is linear, then the estimated coefficients should be, $|\alpha_1| > 0$ and $\alpha_2 = 0$. For a quadratic trend, the estimated coefficients would be $|\alpha_1| > 0$ and $|\alpha_2| > 0$.

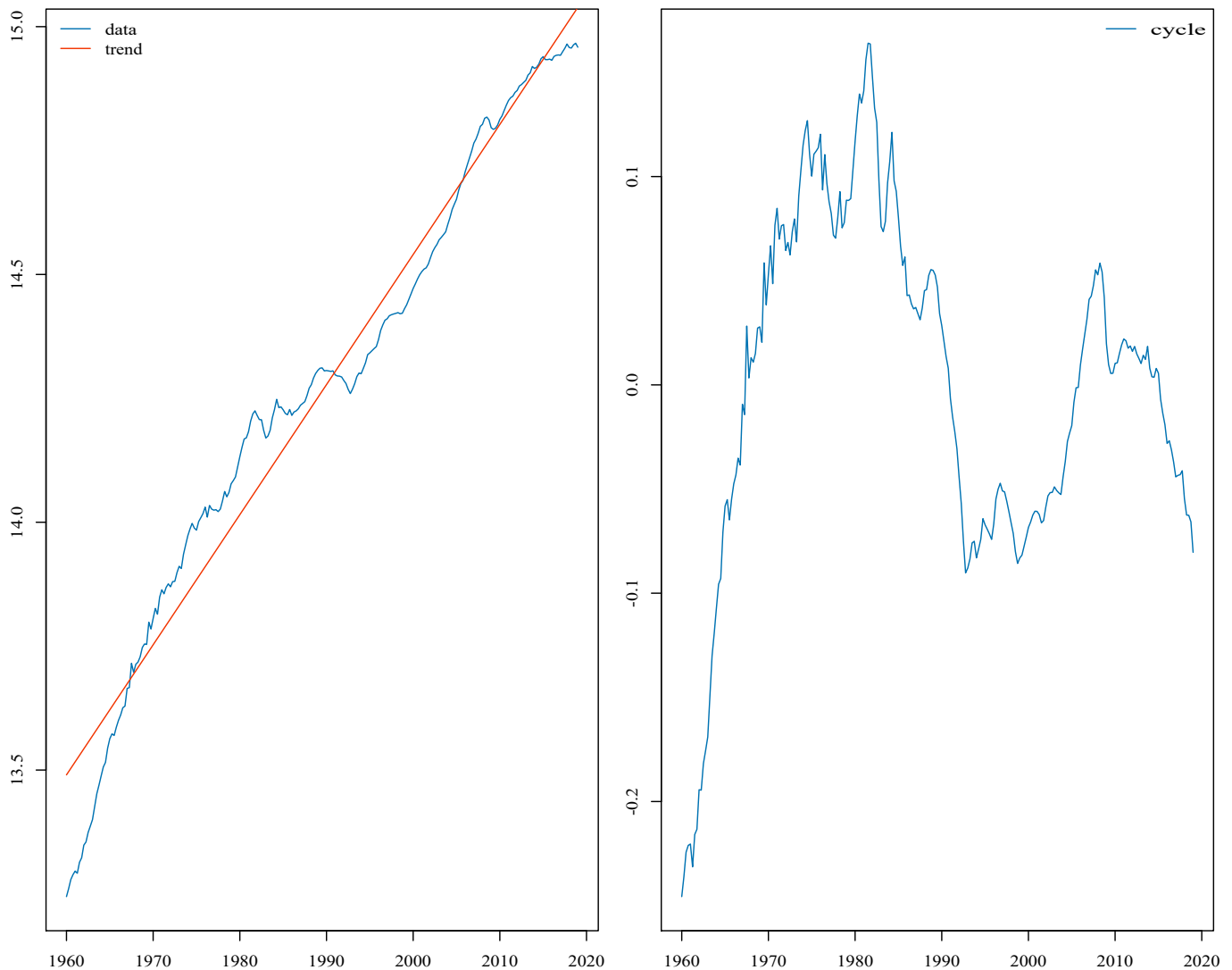


Figure 4: **Linear decomposition - SA output (1960Q1-2015Q1)**

Figure 4 displays the natural logarithm of South African output, together with a linear trend in the left frame. In addition, the deviation from the linear trend, which would represent the cycle is provided in the right frame. The graph of the cycle would suggest that during the 1970s and 1980s, South Africa experienced a protracted expansionary period, which is not necessarily consistent with economic events. It is also worth noting that productivity growth has not been perfectly log-linear (i.e. constant growth rate) and far from smooth, which would imply that the use of a linear trend may be inappropriate (from a theoretic perspective). In addition, there are several structural breaks such as the oil price shock in 1973/1974 and the more recent Global Financial Crisis, which would influence the slope and level of the linear regression.

To allow for a possible structural break in the trend, we could estimate,

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

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 , respectively. These dummy variables would then be constructed as $DS_t(j) = t - j$ and $DL_t(k) = 1$, if $t > j$ or $t > k$, while it would be zero otherwise. In both of these

cases one would need to have *a priori* knowledge about the date for such a break.

In this case the identification of the date of the structural break in the component of the time series that is yet to be identified could be problematic, particularly where the series has more than one date for a structural break. In addition, if the time series is integrated of an order that is greater than zero, the application of a deterministic detrending procedure would introduce a spurious cycle, which would make it susceptible to the critique of Nelson and King (1981).

4 Stochastic trends and filters

Filters may be used to transform a particular time series into various other time series. In this sense, we could use a linear filter to derive a new variables that may represent the trend and the cycle. Using a simple example, where g_t is the result of a moving-average filter for the trend, we can apply the moving-average filter to the observed variable, y_t ,

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

where m and n are positive integers and ω_j are the weights that are applied to past and/or future values of y_t . Alternatively, we could make use of the $G(L)$ polynomial in the expression

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

where L is defined so that $L^j y_t = y_{t-j}$ for positive and negative values for j . In most economic applications we focus our attention on the use of symmetric moving averages, where the weights are such that $\omega_j = \omega_{-j}$.

After defining the trend component with the aid of the above expressions, the cyclical component is then determined by taking the difference of the observed value of y_t from the trend component,

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

where $C(L)$ and $G(L)$ may be termed linear filters. In most instances, the weights are chosen to add up to one, $\sum_{j=-m}^n \omega_j = 1$, which would ensure that it would be possible to reconstitute the observed values of y_t by combining the trend and the cyclical components.

For example, the moving-average filter with a weight of $1/5$ for each component in the *moving window* of five observations, may be derived with the aid of the expression,

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

4.1 Hodrick-Prescott filter

Hodrick and Prescott (1980) provide details of a filter that constitutes the most widely used technique for extracting business cycles from economic variables.⁶ This filter identifies an estimate of the stochastic trend, \hat{g}_t , that is not correlated with the cycle. In essence, this technique seeks to identify the stochastic trend as the components of the time series that exhibit behaviour that are

below that of the business cycle frequency. The cycle would then include all the information that relates to those components that are of a higher frequency. The filter for the identification of the trend can be obtained as the solution to the following minimisation problem,

$$\min_{g_t} \sum_{t=1}^T \left[(y_t - g_t)^2 + \lambda \left\{ (g_{t+1} - g_t) - (g_t - g_{t-1}) \right\}^2 \right] \quad (4.1)$$

The first term $(y_t - g_t)$, is a penalty that is imposed for deviations in the trend from the observed time series. The second component $(g_{t+1} - g_t) - (g_t - g_{t-1})$ is the acceleration in the growth component and is minimised when there is no variability in the trend.

The λ parameter is treated as a constant and is called the smoothing parameter, which increases the *penalty* for the acceleration in the growth component. Hence, if λ approaches infinity, the minima is achieved when the variability in the trend is zero, and the trend is perfectly log-linear. On the other hand, a small value for λ will allow significant variation in the trend, such that if $\lambda = 0$, there will be no cycle as the trend will match the observed time series.

Since the smoothness of the trend, will be sensitive to the value of λ , a justification for the choice of λ should be made. Hodrick and Prescott (1980) argued that $\lambda = 1600$ is a reasonable choice for quarterly data given the characteristics of the U.S. data, and many subsequent studies have used this value. It may be shown that this value is equivalent to a cycle length of about 9.8 years. The choice of this value is the subject of much critique, as the results that are produced from filters that use different values of λ would produce results that may differ. In addition, while this may (or may not) be an appropriate value for the decomposition of output, it will not necessarily be a reasonable choice for other variables, or for decomposing output of other countries.

Another concern with the use of this technique (which is common to most filters) is the end-of-sample problem, as the estimates of the trend will converge on the first and last observations of the time series. This would imply that the filter produces relatively small values for the cycle at the beginning and end of the estimation period. Therefore, the trend would be more responsive to transitory shocks at the end of sample, which may be problematic during periods where the economy is at the peak or trough of a cycle.⁷

King and Rebelo (1993) note that this particular filter contains both forward and backward differences and as a result, the end of sample properties are poor when you do not have an observation for $t + 1$ or $t - 1$. In addition, they also note that these forward and backward looking components would ensure that the Hodrick-Prescott (HP) filter is able to render a stationary process from any integrated series up to fourth order. Furthermore, although the method is stochastic in nature, one should note that the smoothness of the stochastic trend component (i.e. λ) has to be specified *a priori*.

In contrast with these critiques, the advantage of this filter is that the minimization problem has a unique solution, and the filtered series, g_t , has the same length as the original series, y_t . These are considered to be relatively important considerations when deciding upon an appropriate filter. In addition, this technique is easy to apply although one should always note that when applied to a random walk (or any integrated series), it can generate business cycle periodicity, even if none is present in the original data (Harvey and Jaeger (1993) and King and Rebelo (1993)).

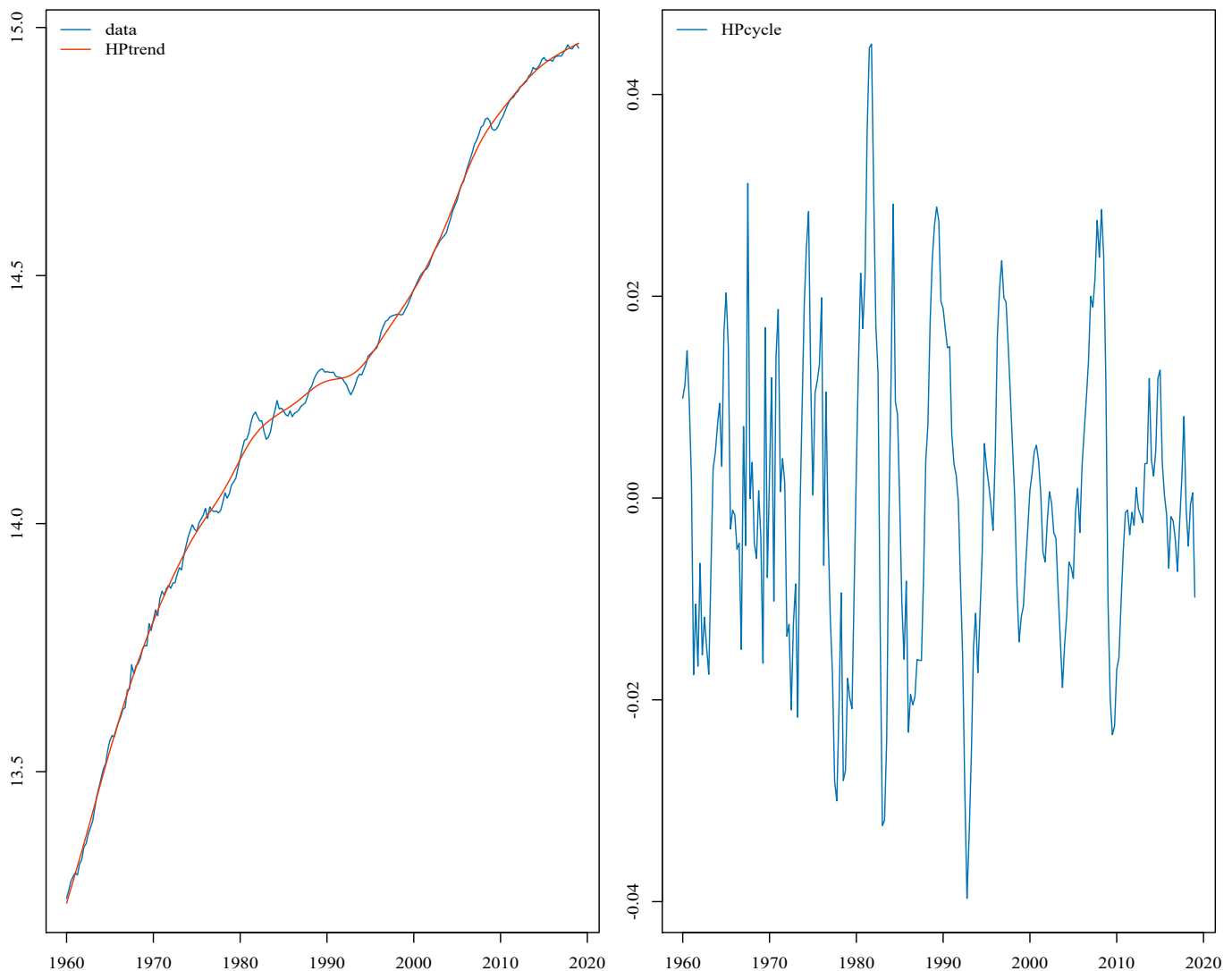
Figure 5: **HP filter - SA output (1960Q1-2015Q1)**

Figure 5 displays the natural logarithm of South African output along with the fitted trend that makes use of the HP-method with $\lambda = 1600$, in the left frame. The right frame displays the cycle constructed as the deviation from the HP-trend. The cycle would appear to resemble what we know of the South African business cycle, with the period of protracted growth during the “great moderation” and the recession during the global financial crisis that started towards the end of 2007.

4.2 Band pass filters

Another popular method that has been used to measure business cycles is the band pass (BP) filter, which were introduced by Baxter and King (1999) and Christiano and Fitzgerald (2003). The filter differs to those that have been discussed in that it seeks to identify specific (business cycle) components that correspond to the chosen frequency band that has an upper and lower limit. When applying this filter to the data, one has to determine the periodicity of the business cycles that one wants to extract. As such, the band pass filter is usually expressed within the frequency domain approach to time series analysis.

The basic idea behind the frequency domain approach is that any time series could be represented as a combination of sine and cosine functions. Therefore, consider a time series that may be generated from the function,

$$y_t = A \cos(2\pi\omega t) \quad (4.2)$$

Where A is the amplitude (height) of the cycle, ω is the frequency of oscillation (the number of occurrences of a repeating event per time) and t is usual the time. The value 2π is a constant that measures the period of the cycles. Hence, if $y_t = A \cos(2\pi t)$, then we will observe one cycle over the sample period that is under investigation. After increasing ω , we will increase the number of cycles that may be observed.

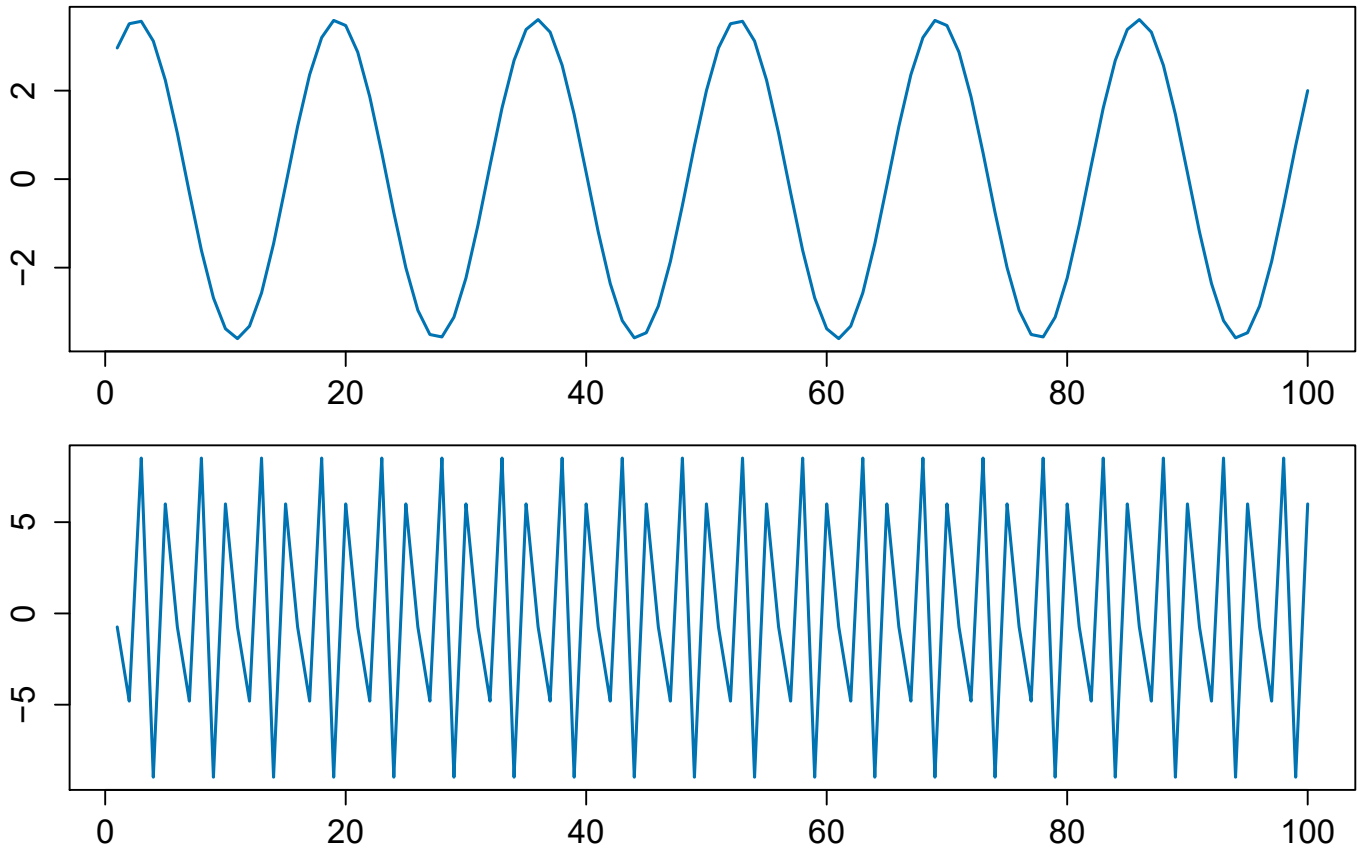


Figure 6: **Artificial Data**

Figure 6 provides an example of two frequency components that were produced for 100 observations. The top frame was produced with $2 \cos(2\pi \frac{6}{100}t)$ and the bottom frame was produced with $2 \cos(6\pi \frac{40}{100}t)$. When comparing these results, we note that after increasing the amplitude that is provided by A , the height of the cycles increases by three.

There is also a large difference regarding the frequency of oscillations (i.e. how often the periodic cycles repeat themselves). In the top frame (where $\omega = 6$), we can count six cycles over the time span of 100 (i.e. six cycles between 0 and 2π), while there are forty cycles over a time span of 100 in the bottom panel (where $\omega = 40$). We say that a cycle which oscillates more exhibits a higher frequency, while a cycle that oscillates less has lower frequency.

Hence, an intuitive measure of frequency is the amount of time that elapses per cycle, which we will call λ . It may be calculated as,

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

If we are working with quarterly data, then to find ω that corresponds to a cycle length of 1.5 years (that is, a high frequency cycle), we set $\lambda = 6$ quarters per cycle and solve for $6 = 2\pi/\omega_H$, such that

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

Similarly, the frequency corresponding to a low-frequency cycle length of 8 years would be given by

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

Inspired by the National Bureau of Economic Research (NBER) business cycle chronology, Baxter and King (1999) wanted to decompose a time series into three periodic components, which comprised of the 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 that were longer than eight years were identified with the trend, and those that had periodic components of less than one and a half years were identified with the irregular component.

To construct the cyclical component we need to weight the periodic components according to the Baxter and King definition, and then integrate across all frequencies. If we define the band pass filter as $B(\omega)$ then the chosen frequencies will imply the following restrictions on the band filter,

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

Hence, the interval $B(\omega) = [\pi/16, \pi/3]$ can be interpreted as *business cycle* frequencies. Any periodic components with a frequency within this interval can pass through the filter unchanged, as the filter multiplies them with one. The interval $[0, \pi/16]$ would then correspond to the trend and $[\pi/3, \pi]$ would define the irregular fluctuations. Periodic components with a frequency that lies within these intervals are eliminated, since they are multiplied by zero.

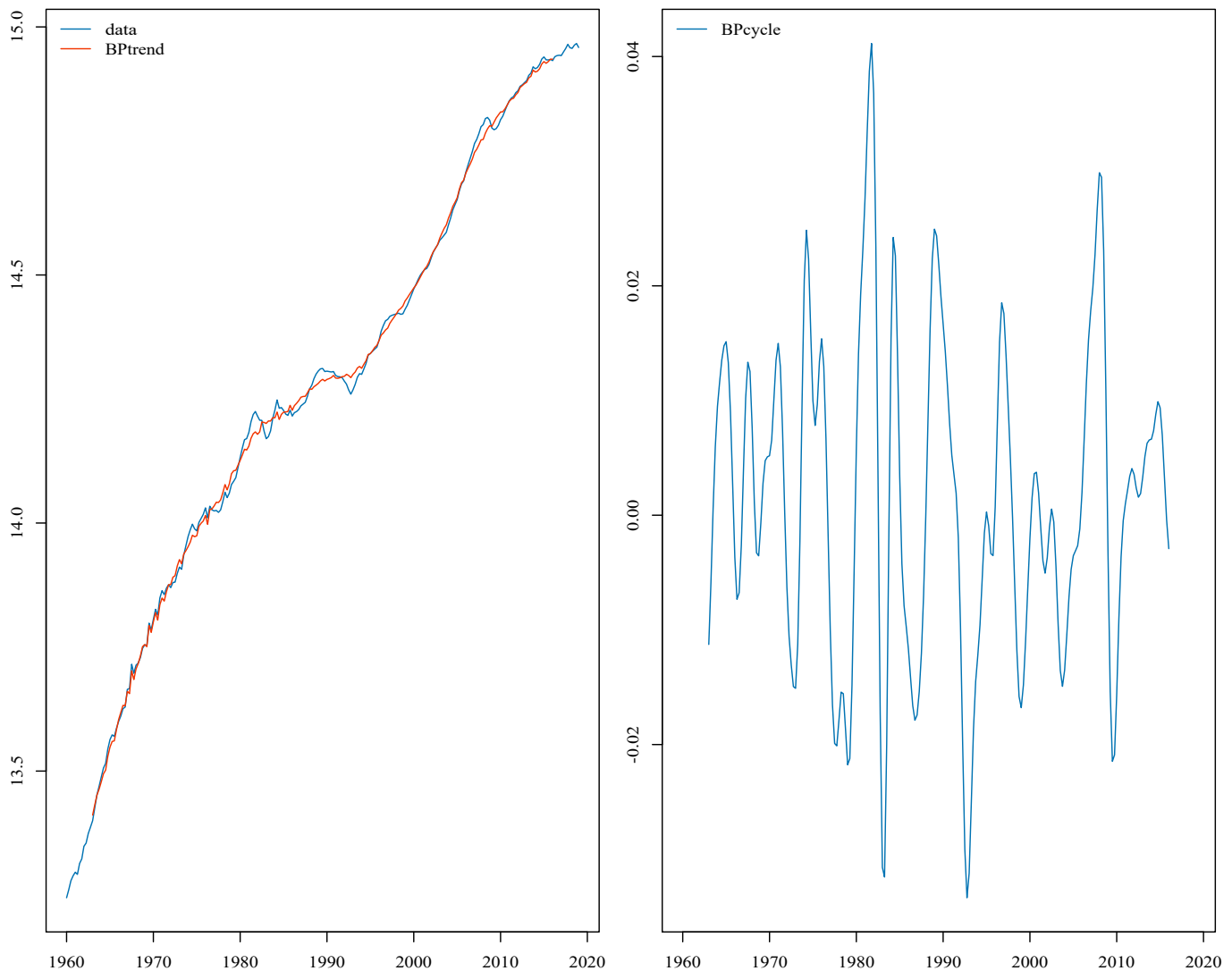


Figure 7: BP filter - SA output (1960Q1-2015Q1)

Figure 7 displays the filtered South African output using the Baxter and King (1993) band pass filter. Note that the cyclical resembles the HP cycle, except that it is much smoother. This smoothness follows from the fact that we have filtered out the high frequency noise. That is, cycles of a periodicity of less than one and a half years have been removed. A useful feature of the Baxter and King band pass filter is that it can easily be changed to accommodate data sampled at different frequencies (say monthly or yearly), by changing ω_L and ω_H .

While Baxter and King (1993) favour a three-part decomposition, other economists prefer a two-part classification in which the highest frequencies also count as part of the business cycle. Consider the following filter:

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

The trend component is still defined in terms of fluctuations lasting more than eight years, but the cyclical component now consists of all oscillation lasting eight years or less. This is known as a *high-pass* filter because it passes all of the frequencies that are higher than some pre-specified value, while it eliminates everything else.

A drawback of this filter is that one has to decide on the preferred frequencies for the cycles. This may not always be known in advance and it may not be consistent over the entire sample. Furthermore, the filter is subject to the end-of-sample problem as was the case with the HP-filter.

5 Beveridge-Nelson Decomposition

Many economic time series are integrated of the first-order, where the first difference of a process can be represented as a stationary process that has an autoregressive moving-average form. To decompose such a nonstationary series into a permanent and a transitory component one could make use of the decomposition, that is due to Beveridge and Nelson (1981). In this case, the permanent component is assumed to follow a random walk with drift, and the transitory component is a stationary process with zero mean (that is perfectly correlated with the permanent component). If we let y_t be integrated of first-order, then the first difference, Δy_t , is stationary. Where we assume that the stationary component has the following moving-average representation,

$$(1 - L)y_t = \Delta y_t = \mu + B(L)\varepsilon_t \quad (5.1)$$

To derive estimates for this decomposition, we would firstly 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 $B(L)$, provides

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

and substituting into (5.1) yields the following expression,

$$\Delta y_t = \mu + B(L)\varepsilon_t = \mu + [B(1) + (1 - L)B^*(L)]\varepsilon_t \quad (5.2)$$

For the decomposition, $y_t = g_t + c_t$, it follows that $\Delta y_t = \Delta g_t + \Delta c_t$. Using this together with equation (5.2), provides an expression for the change in the trend component of y_t ,

$$\Delta g_t = \mu + B(1)\varepsilon_t \quad (5.3)$$

and the change in the cyclical component of y_t is then provided by

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

From equation (5.3) we see that the trend follows a random walk with drift. This expression can be solved to yield,

$$g_t = g_0 + \mu t + B(1) \sum_{s=1}^t \varepsilon_s \quad (5.4)$$

As such, the trend consist 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 . The cyclical component is stationary and is given by

$$c_t = B^*(L)\varepsilon_t = (1 - L)^{-1}[B(L) - B(1)]\varepsilon_t \quad (5.5)$$

Beveridge and Nelson (1981) are able to show that the stochastic trend defined in equation (5.4) could also be interpreted as the long-term forecast of the series adjusted for the mean rate of change. In addition, the cycle that is defined in equation (5.5) as the stationary process that reflects the deviations of the trend from the observed series.

This decomposition implies that an innovation in g_t and c_t are proportional to ε_t . Hence they will be perfectly correlated and the permanent component will have the same drift, μ , as the observed series. Further, the variance of the innovations in the permanent component, $(B(L)\sigma^2)$ will be larger (smaller) than the variance of the innovations in the observed data y_t , if $B(1)$ is larger (smaller) than one. Note also that when the permanent component is restricted to be a random walk with drift, $B_0 = 1$ and all the $B_i = 0$ for $i > 0$, the variance of the permanent component equals the variance in the observed series, and the cyclical component will be zero for all t .

To be able to identify the cyclical and permanent component, one must specify models that can be written as the stationary moving average processes. There are two stages involved in this trend-cycle decomposition. First an ARIMA model (p, d, q) has to be estimated to the series y_t where p is the number of AR lags, d is the number of differencing and q is the number of MA lags. Then c_t has to be numerically estimated.

To see how one can estimate the Beveridge-Nelson (BN) decomposition in practice, assume that an AR(1) process is responsible for the growth rate of output, $\Delta y_t = \phi \Delta y_{t-1} + \varepsilon_t$, where we have ignored the constant term. Assuming $\phi < 1$, the AR(1) process can be written in terms of the infinite MA(∞) process where we find $B(L)$, $B(1)$ and $B^*(L)$ from,

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

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

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

Solving in terms of y_t , using equation (5.2), but now without a constant, provides us with

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

This 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

$$y_t = g_t + c_t$$

$$\Downarrow$$

$$y_t = \frac{1}{1 - \phi}(1 - L)^{-1}\varepsilon_t + \frac{-\phi}{(1 - \phi L)(1 - \phi)}\varepsilon_t$$

A quick computational approach was suggested by Cuddington and Winters (1987), where g_t is calculated directly from the expression in (5.4) by estimating $B(1)$ from a truncated Wold representation of Δy_t . The obvious difficulty is that the initial value of g_t in (5.4) is known, so the procedure is only correct up to an additive factor.

The advantage of Beveridge-Nelson is that it is appropriate for the extraction of cycles when a series is difference-stationary. Moreover, it may be used on those series that not only contain a unit root, but are also highly volatile.

One disadvantage, however, is that it is rather time-consuming, as one has to choose between different ARMA models. In addition, Cochrane (1988) notes that these different ARMA specifications may give very different trend-cycle decompositions, where low-order ARMA models will systematically overestimate the random walk component in the trend. Hence, the parameter estimates will match the short-run behaviour and misrepresent the long-run behaviour. As the innovative variance of the random walk is a property of the very long-run behaviour of the series, one should therefore estimate higher-order models that adequately capture this long-run behaviour. Finally, misrepresenting an $I(2)$ process as an $I(1)$ may be problematic in this setting.

6 Stylized facts of business cycles

The description of the aggregate fluctuations in a set of economic variables is an important exercise as it establishes the stylized facts of the business cycles in a particular country. However, as the cycles will not be invariant to the method that is used to describe them, the results should be tested against alternative trend specifications.

Figure 8 compares the cycles that are derived from gross domestic product in South Africa. These were obtained from the different de-trending methods that were presented above. The results suggest that a linear filter provides peculiar results that do not agree with what we know of economic events. The stochastic filters suggest a fairly similar pattern for the business cycles although there are parts

where the differences are more pronounced. The Beveridge-Nelson decomposition would appear to provide slightly divergent results and as such it would be important to compare how these results compare to the official dating procedure for recessions, which is provided by the South African Reserve Bank.

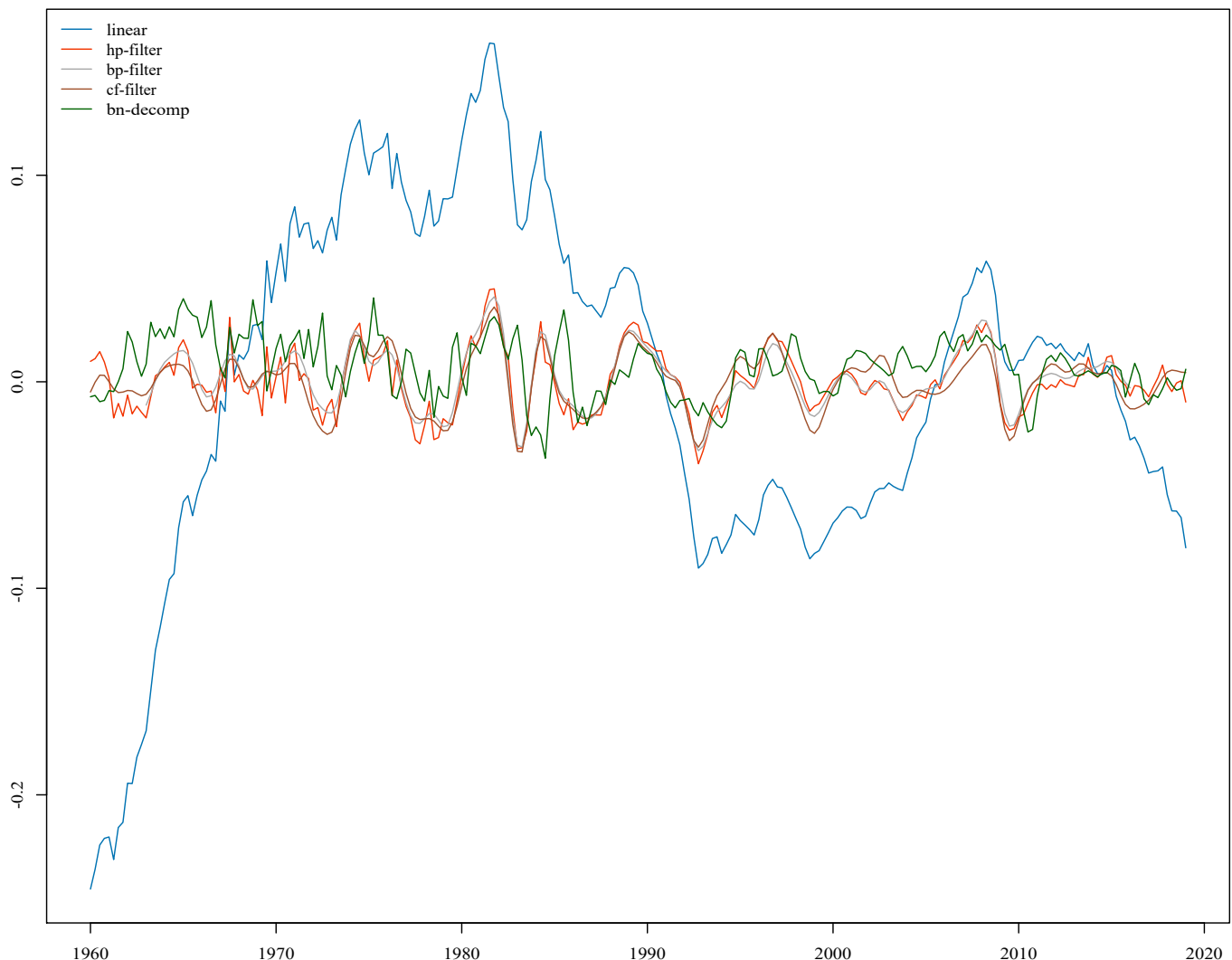


Figure 8: **Business Cycles - SA output (1960Q1-2015Q1)**

In addition to this comparison, it is usually a good idea to consider the stylized facts that relate to the cycles of other macroeconomic aggregates, such as: GDP, consumption (Cons), exports (Exp), import (Imp), productivity (Prod), investment (Invest) and employment (EMP). These should be compared to our knowledge of economic events and where relevant, we should also consider the correlation between these measures, as well as whether or not they are leading or lagging one another. This will help us to judge the plausibility of using a particular technique to decompose the business cycle.

7 Conclusion

Many economic and financial applications make use of decompositions for nonstationary time series, which are transformed into a permanent and a transitory component. To complete this task one could make use of a linear filter for the trend, which may be perturbed by transitory cyclical fluctuations. Alternatively one could make use of a Hodrick-Prescott filter, which is the most popular method for extracting business cycles. The filter extracts a stochastic trend which for a given value of the parameter λ . This trend moves smoothly over time and is uncorrelated with the cycle. It is worth noting that the results from this filter are not robust to changes in the value of the

smoothness parameter. Another popular method that is used to measure the business cycle is that of the band pass filter. The filter removes all the components in a series, except for those that correspond to the chosen frequency band. In the Beveridge-Nelson decomposition, the permanent component is shown to be a random walk with drift, and the transitory component is stationary process with zero mean, which is perfectly correlated with the permanent component. When applying these methods, one should interrogate the results for robustness by applying many alternative filters when analysing business cycles.

Then lastly, it is perhaps worth noting that a filtered time series is usually stationary but somewhat persistent; which is the case for the HP filtered measure of the output-gap. Given these properties of the data, one would usually be able to generate reasonable forecasts for output-gap (particularly when compared to the forecast for output growth, which is less persistent). However, one may wish to suggest that a forecast for the output-gap is less useful than a forecast for output growth and the end-of-sample problem that is encountered with the HP filter would detract further from the usefulness of the forecast for the output-gap. However, this would in no way prevent one from using the measure of the output gap in a multivariate model that is concerned with forecasting some other variable (such as inflation).

8 Appendix: Wavelet decompositions

Most of the commonly used decompositions in economics, such as those that were designed by Hodrick and Prescott (1997), Baxter and King (1999) and Christiano and Fitzgerald (2003), seek to approximate ideal filters, where one is able to identify the trend, cycle and noise components that are located at different periodicities. When applying these techniques within the frequency domain, one would effectively decompose a time series with a number of sine and cosine functions, which define the rate at which the time series oscillates. It is important to note, that this transformation results in the loss of all time-based information, where it is assumed that the periodicity of all the components are consistent throughout the entire sample.

To allow for changes in the periodicity of the respective components, Gabor (1946) developed the Short-Time Fourier Transform (STFT) technique, which involves applying a number of Fourier transforms to different subsamples of the data. Gencay, Selcuk, and Whitcher (2010) refer to the subsample as a data *window*, where the technique involves sliding the window across the time series and taking a Fourier transform of each subsample. Although this technique would provide potentially useful information on the timing of an event that may have arisen at a particular frequency, it is limited in that the precision of the analysis is affected by the size of the subsample. For instance, one would need a large subsample to identify changes that arise at a low frequency, and small subsamples to identify changes in the higher frequency components.

To overcome the limitations of the above frequency domain techniques, wavelet transformations were developed to capture features of time-series data across a wide range of frequencies that may arise at different points in time. This technique makes use of a number of wavelet functions that are stretched and shifted to describe features that are localised in frequency and time. For example, the wavelet function would be expanded over a relatively long period of time when identifying low-frequency events, and it would be relatively narrow when describing high frequency events.⁸ After shifting all of these wavelet functions that have different amplitudes over the entire sample of data, one is able to associate the components with specific time horizons that occur at different locations in time.

Early work with wavelet functions dates back to Haar (1910), who used a number of square-wave functions to decompose time-series data. Unfortunately, the properties of square-wave functions were found to be limited, and as such, a number of alternatives were developed, including those that are discussed in Grossmann and Morlet (1984) and Daubechies (1992).⁹ For the computation of these transformations, which make use of various wavelet functions at different scales, most studies currently employ the multiresolution decomposition of Mallat (1989) and Strang and Nguyen (1996).

To describe the use of this technique, one could allow for the case where a variable is composed of a trend and a number of higher-frequency components. In this instance, the trend may be represented by a father wavelet, $\phi(t)$, while the mother wavelets, $\psi(t)$, are used to describe information at lower scales (i.e. higher frequencies). Using an orthogonal wavelet transformation, one could then describe variable y_t as

$$y_t = \sum_k s_{0,k} \phi_{0,k}(t) + \sum_{j=0}^J \sum_k d_{j,k} \psi_{j,k}(t) , \quad (1)$$

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.

The mother wavelet functions, $\psi_{1,k}(t), \dots, \psi_{J,k}(t)$, are then generated by shifts in the location of the wavelet in time and scale, such that

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

where the shift parameter is represented by $2^j k$ and the scale parameter is 2^j . This choice of dyadic scaling factors is arbitrary but efficient (Daubechies 1992). As depicted in the daublet wavelet functions in Figure 9, smaller values of j (which produce a smaller scale parameter 2^j), would provide the relatively tall and narrow wavelet function on the left. For larger values of j , the wavelet function is more spread out and of lower amplitude. In addition, after shifting this function by one period, we produce the function that is depicted on the right of Figure 9.

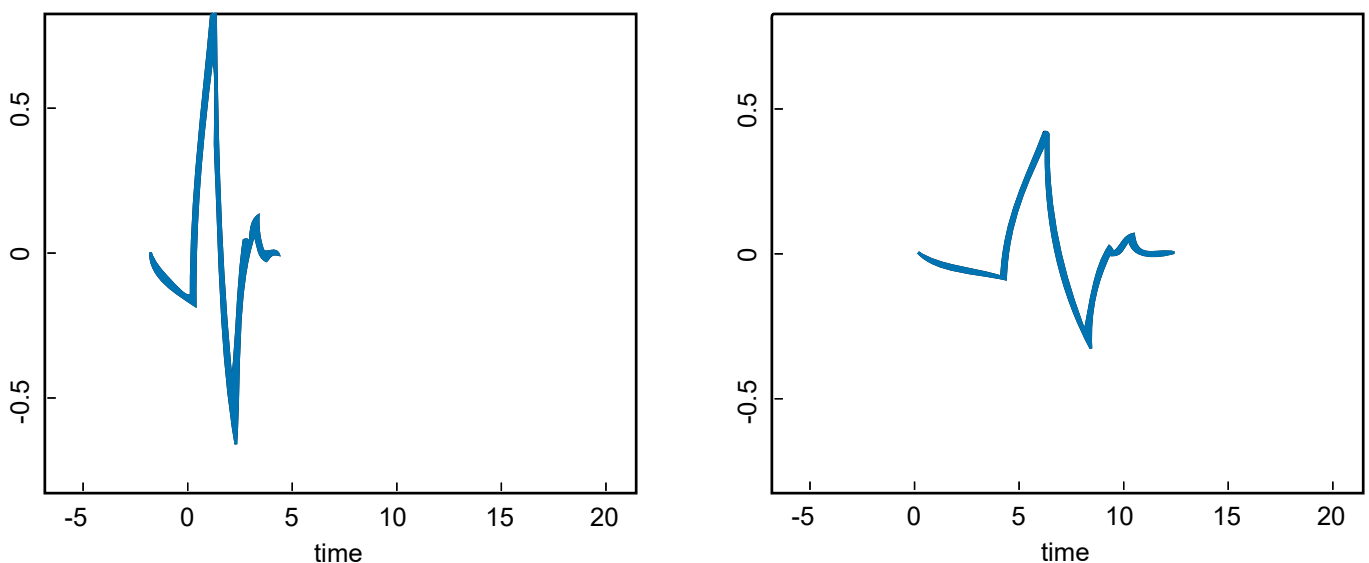


Figure 9: **Daublet (4) wavelet functions** - $\psi_{1,0}(t)$ and $\psi_{2,1}(t)$

Early applications of wavelet methods in economics include the work of Ramsey and Zhang (1997), which made use of a wavelet decomposition of exchange rate data to describe the distribution of this data at different frequencies. In addition, Ramsey and Lampart (1998a) made use of a decomposition of money and income data to describe the relationship between these variables at different frequencies, while Ramsey and Lampart (1998b) considered the relationship between income and expenditure (i.e. permanent income hypothesis) at different time scales.¹⁰

Modern wavelet functions may take various forms that could be summarized by *smoothed functions*, that may be used to decompose a series into trend and cycle, *peaked functions*, that may be used to identify the peak of cycle, or *square functions*, that are used to identify structural breaks. For the purposes of identifying the business cycle one would use smoothed functions that include daubelets, coiflets and symlets. There are also a number of transformations that may be used. Many studies make use of a maximum overlap discrete wavelet transform (MODWT), which does not restrict the sample size to a multiple of 2^j . In addition, this technique is also able to preserve the phase properties of the data, where it can match the smoothed terms to the underlying data.

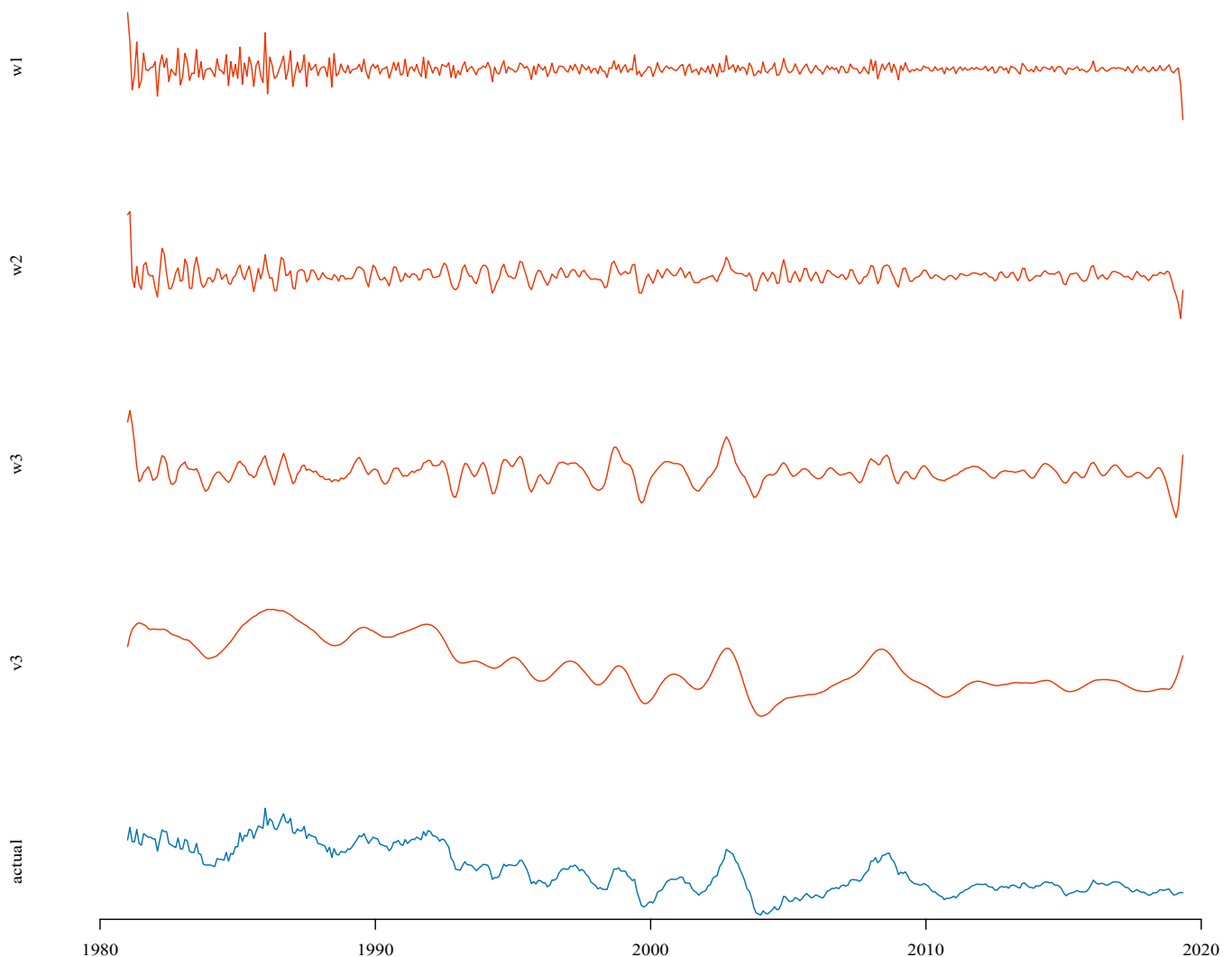


Figure 10: **Daubelet (4) wavelet decomposition - South African inflation**

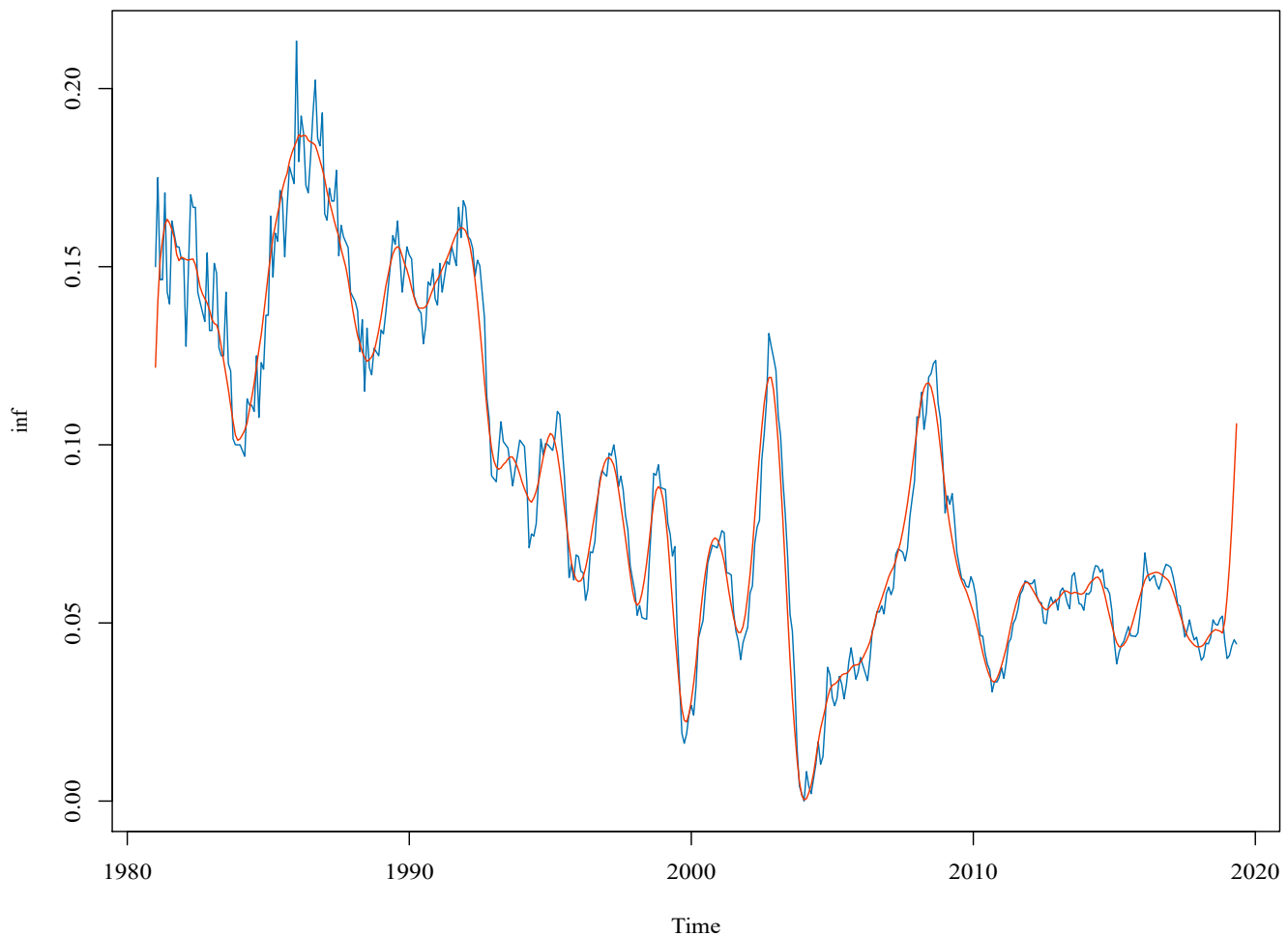


Figure 11: **Daublet (4) wavelet decomposition - South African inflation**

Figure 10 and 11 contain the results of a decomposition that was applied to South African Consumer Price inflation, where we are interested in removing the noise from the data. In this example, we make use of various smoothed wavelet functions that include daubelets 3-4. We also make use of both three scales, where J is set at 3.¹¹

The decompositions at various different scales are contained in Figure 10, where we note that there is significant change in the periodicity of the variables over time. The results of the filtered trend are contained in Figure 11, which could be used as a measure of core inflation, as in Du Plessis, Du Rand, and Kotzé (2015).

To conclude this section, there are a number of advantages that are inherent in the application of wavelet decompositions, as they can be applied to data of any integration order and allow for changes in the distribution of the frequency over time. In addition, they have many of the benefits that are associated with spectral techniques, but they do not lose the time support, which is useful when seeking to identify changes in the process at different frequencies. Then lastly, as one is able to include a number of bands, which are additive, one is able to focus attention on many possible periodic components.

9 References

- Baxter, Marianne, and Robert G. King. 1993. "Fiscal Policy in General Equilibrium." *American Economic Review* 83 (3): 315–34.
- . 1999. "Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series." *Review of Economics and Statistics* 81: 575–93.
- Beveridge, S., and C. R. Nelson. 1981. "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the Business Cycle." *Journal of Monetary Economics* 7 (2): 151–74.
- Burns, Arthur F., and Wesley C. Mitchell. 1947. *Measuring Business Cycles*. New York: National Bureau of Economic Research.
- Canova, Fabio. 1998. "Detrending and Business Cycle Facts: A Users Guide." *Journal of Monetary Economics* 41 (3): 475–512.
- Christiano, L., and T. Fitzgerald. 2003. "The Band Pass Filter." *International Economic Review* 44 (435–465).
- Cochrane, John H. 1988. "How Big Is the Random Walk in GNP?" *The Journal of Political Economy*, 893–920.
- Crowley, Patrick M. 2007. "A Guide to Wavelets for Economists." *Journal of Economic Surveys* 21 (2): 207–67.
- Cuddington, J. T., and L. A. Winters. 1987. "The Beveridge-Nelson Decomposition of Economic Time Series: A Quick Computational Method." *Journal of Monetary Economics* 19 (1): 125–27.
- Daubechies, Ingrid. 1992. *Ten Lectures on Wavelets*. Vol. 61. SIAM.
- Du Plessis, Stan, Gideon Du Rand, and Kevin Kotzé. 2015. "Measuring Core Inflation in South Africa." *South African Journal of Economics* 83 (4): 527–48.
- Gabor, Dennis. 1946. "Theory of Communication." *Journal of the I.E.E* 93(3): 429–57.
- Gencay, Ramazan, Faruk Selcuk, and Brandon Whitcher. 2010. *An Introduction to Wavelet and Other Filtering Methods in Finance and Economics*. San Diego: Academic Press.
- Grossmann, Alexander, and Jean Morlet. 1984. "Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape." *SIAM Journal on Mathematical Analysis* 15 (4). SIAM: 723–36.
- Haar, Alfred. 1910. "Zur Theorie Der Orthogonalen Funktionensysteme." *Mathematische Annalen* 69 (3): 331–71.
- Harvey, Andrew C., and A. Jaeger. 1993. "Detrending, Stylised Facts and the Business Cycle." *Journal of Applied Econometrics* 8: 231–47.
- Heil, C., and D. F. Walnut. 2006. *Fundamental Papers in Wavelet Theory*. Princeton, New Jersey: Princeton University Press.
- Hodrick, Robert J., and Edward C. Prescott. 1980. "Postwar U.S. Business Cycles: An Empirical Investigation." *Carnegie Mellon University Discussion Paper* 451.
- . 1997. "Postwar U.S. Business Cycles: An Empirical Investigation." *Journal of Money, Credit and Banking* 29 (1): 1–16.

- Hubbard, Barbara B. 1998. *The World According to Wavelets: The Story of a Mathematical Technique in the Making*. Second. A K Peters/CRC Press.
- King, Robert G., and Sergio T. Rebelo. 1993. "Low Frequency Filtering and Real Business Cycles." *Journal of Economic Dynamics and Control* 17 (1-2): 207–31.
- Mallat, Stephane G. 1989. "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 11 (7). Ieee: 674–93.
- Nelson, C. R., and H. King. 1981. "Spurious Periodicity in Anappropriately Detrended Time Series." *Econometrica*, 741–51.
- Ramsey, James B. 2002. "Wavelets in Economics and Finance: Past and Future." *Studies in Nonlinear Dynamics & Econometrics* 6 (3).
- Ramsey, James B., and Camille Lampart. 1998a. "Decomposition of Economic Relationships by Timescale Using Wavelets." *Macroeconomic Dynamics* 2 (1). Cambridge Univ Press: 49–71.
- . 1998b. "The Decomposition of Economic Relationships by Time Scale Using Wavelets: Expenditure and Income." *Studies in Nonlinear Dynamics & Econometrics* 3 (1): 23–42.
- Ramsey, James B., and Zhifeng Zhang. 1997. "The Analysis of Foreign Exchange Data Using Waveform Dictionaries." *Journal of Empirical Finance* 4 (4). Elsevier: 341–72.
- Schleicher, Christoph. 2002. "An Introduction to Wavelets for Economists." Working Paper 2002-3. Bank of Canada.
- Shumway, R. H., and D. S. Stoffer. 2011. *Time Series Analysis and Its Applications: With R Examples*. Edited by 3rd. New York: Springer.
- Strang, Gilbert, and Truong Nguyen. 1996. *Wavelets and Filter Banks*. Wellesley Cambridge Press.

-
1. Burns and Mitchell (1947) conducted an early empirical study into the behaviour of business cycles in the United States, with the aid of a systematic approach where they observed that a cycle would consist of expansions occurring at about the same time in many economic activities, followed by a contraction in many of the variables. They suggested that this sequence of changes is recurrent but not periodic. Similar work is currently performed by the South African Reserve Bank, which seeks to establish the dates for economic expansions and contractions.↵
 2. In terms of a formal definition, the business cycle refers to the regular periods of expansion and contraction in major economic aggregate variables (Burns and Mitchell 1947). We would infer that a turning point occurs when the business cycle reaches local maximum (peak) or local minimum (trough).↵
 3. Shumway and Stoffer (2011) suggest that most investigations into the cyclical component of a time series should be expressed with the aid of frequency domain techniques, which employ Fourier transformations that are driven by sine and cosine functions.↵

4. The general rules of trigonometry ensure that $\cos(0) = 1$, $\sin(0) = 0$, $\cos(-x) = \cos(x)$ and $\sin(-x) = -\sin(x)$.↵
5. See Canova (1998) amongst others.↵
6. This article was published as Hodrick and Prescott (1997).↵
7. To address the end-of-sample problem, certain researchers have sought to augment the dataset of observed time series with forecasts of the respective variable.↵
8. The wavelets literature refers to the use of scales rather than frequency bands, where the highest scale refers to the lowest frequency and *vice versa*.↵
9. See, Hubbard (1998) and Heil and Walnut (2006) for a detailed account of the history of wavelet analysis.↵
10. See Ramsey (2002), Schleicher (2002) and Crowley (2007) for a more general overview of the use of these methods in economics.↵
11. In this study we perform a simple wavelet analysis that seeks to identify the trend, or father wavelet. Note that these methods could also be used to remove noise from each of the respective scales, should they extend over a particular threshold, before each of the signals is combined to represent the de-noised signal.↵