Catholic
University
of Leuven

Faculty
of Science

Capita Selecta in Statistics: Report

Ishika Jain, r0915387
Axl Wynants, r0861234
Academic year 2022–2023

1 Exploratory Analysis

This report focuses on the analysis and modeling of a patient's systolic blood pressure signal (in mm Hg). A total of 300 measurements was recorded at evenly spaced intervals of 0.2 seconds over a duration of 60 seconds, resulting in a sampling frequency of 5 Hz. It should be noted that the patient's blood pressure was not measured while at rest. Additionally, it will be explored whether the influence of the patient's heart rate and respiratory rate can be detected within this signal.

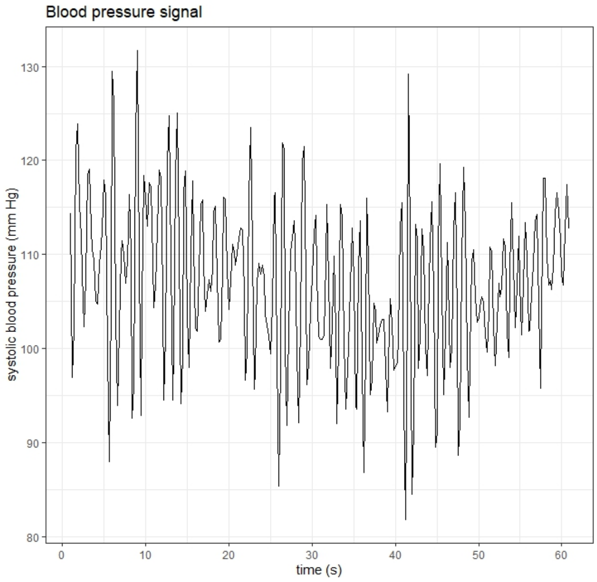

Figure 1: Blood pressure signal

A logical first step is to plot the measured signal, as depicted in Figure 1. At first sight, the mean of the series appears to depend on time. We also notice periodic volatility clustering. Observing the autocorrelation there seems to be periodicity at lag 5. It is unclear whether this series would satisfy the weak stationarity assumption, hence we try applying a differentiator.

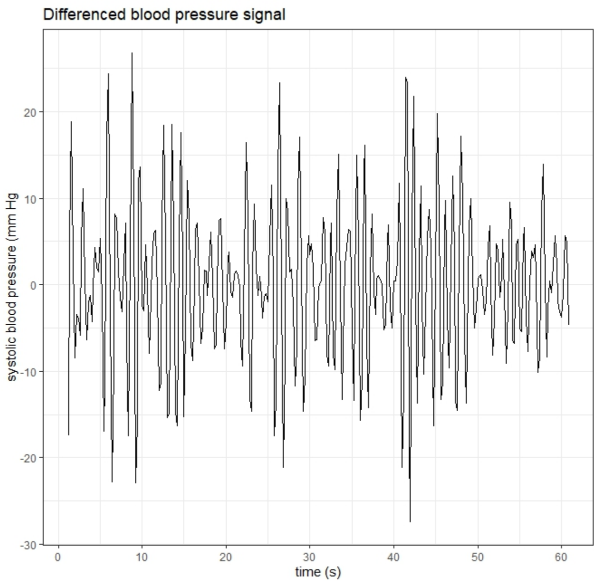

Figure 2: Differenced blood pressure signal

The differencing has clearly improved the mean stationarity of the series, depicted in Figure 2. However, the series still exhibits some volatility clustering, hence we check for seasonal effects in the autocorrelation and partial autocorrelation functions to see if seasonally differencing at the aforementioned lag 5 diminishes the observed periodicity.

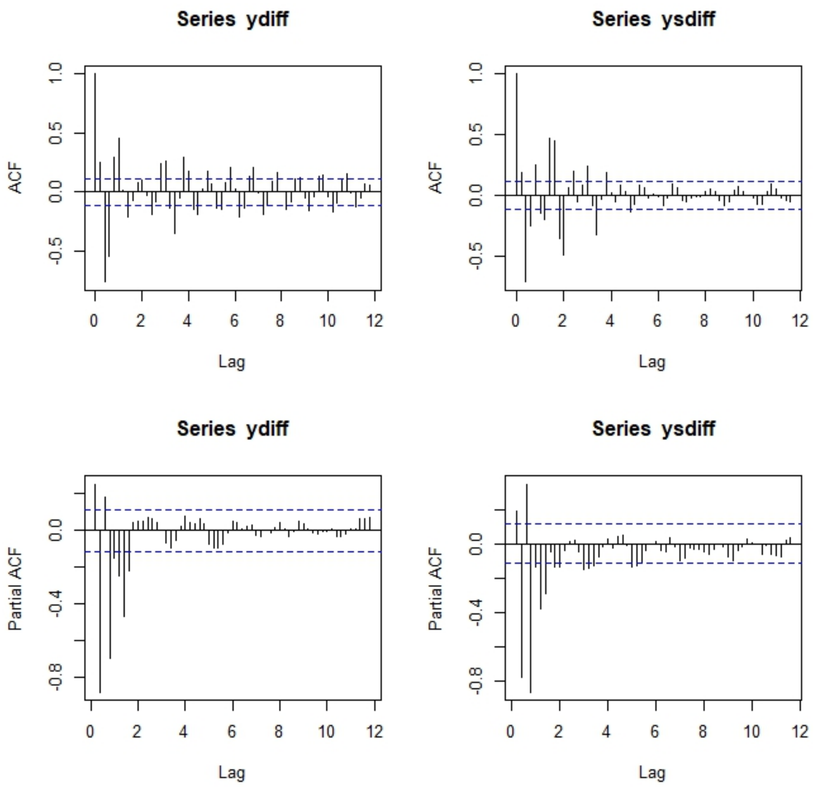

Figure 3: ACF and PACF of (seasonally) differenced blood pressure signal

The ACF shows that the differenced series still contains seasonality at lag 5, though less pronounced than in the raw series. After seasonally differencing the ACF looks much better in the sense that the autocorrelation decays more quickly. The PACF looks very similar to before, though after seasonally differencing it shows marginally more significant periodicity. Thus we proceed with the differenced and seasonally differenced series.

2 ARIMA Model Configuration

Under the assumption that the preprocessed series is now weakly stationary, it may now be modeled with (S)ARIMA class models.

Through analysis of the spectrum of the series, dominant frequencies can be identified and mapped to AR and MA terms in the generating process of the series. This spectrum is depicted in Figure 4. It shows 2 clear peaks, each of which associated to 2 AR terms due to the cyclical nature of the periodogram (each root has a complex conjugate). The frequency band is also attenuated between these two peaks, which hints at the presence of 2 MA terms, because of similar reasoning.

This means that the non-seasonal part of the model is ARIMA(4, 1, 2). Upon inspection of the roots of this model, their angles and radii match the extrema observed in the spectrum.

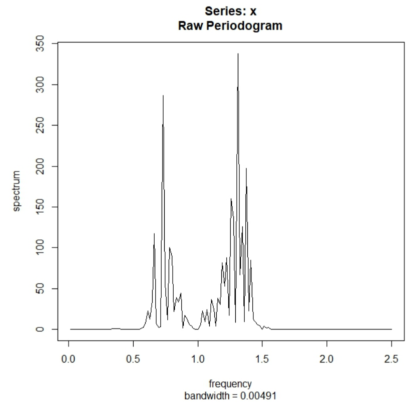

Figure 4: Periodogram of preprocessed blood pressure signal

To identify the seasonal part of the model, the (partial) autocorrelation can aid in detecting the presence of seasonal AR and MA terms. On the seasonally differenced ACF in Figure 3, the autocorrelation pattern is repeated once before becoming insignificantly small, pointing to the presence of 1 seasonal MA term. Similarly, on the PACF the pattern is seen repeating twice at multiples of the base periodicity, suggesting overtones. We choose 2 seasonal AR terms.
The model configuration chosen through our exploration of the signal is thus ARIMA(4, 1, 2),(2, 1, 1)[5]. This result will be compared to an automated procedure based on information criteria and unit root tests.

3 Automated ARIMA Modeling

We conducted an automated ARIMA modeling using the R function `auto.arima` from the package `forecast`. The function suggested the following model:

```
> autoARIMA <- auto.arima(y)
> autoARIMA
Series: y
ARIMA(4,1,4)(2,0,1)[5]

 Coefficients:
          ar1       ar2     ar3      ar4      ma1      ma2     ma3     ma4
       1.2937   -2.0679  1.1344  -0.7489  -0.0056  -0.8027  0.1341  0.7357
s.e.   0.0591    0.0845  0.0809   0.0498   0.0555   0.0557  0.1146  0.0514
         sar1      sar2    sma1
      -0.9602   -0.5802  0.4924
s.e.   0.1073    0.0547  0.1879

sigma^2 = 1.6:   log-likelihood = -498.33
AIC=1020.67    AICc=1021.76    BIC=1065.07
```

We can note that the highest order terms AR4, MA4, and SAR2 are all significant as twice the standard error is much smaller than the value of coefficients(except SMA1, which has a relatively high standard error though still significant). The model acknowledges the periodicity at lag 5, but does not apply a seasonal difference since $D = 0$. We study the roots of the AR and MA polynomials of both the non-seasonal components as well as seasonal components by calculating all the roots of different polynomials in the time series model. We report their angle and radius in the table below.
We can note that the value of radii for the SAR, and SMA roots are less than 0.8, thus suggesting a weak seasonal effect. Whereas, the roots of classical AR and MA polynomials indicate two maxima and minima in the spectrum respectively at the given angles.

AR	Rootsfp	Radius	Angle
3	0.18023+0.9265i	0.9438645	1.0970599
2	0.1803-0.9265i	0.9438645	-1.0970599
1	0.4666-0.7892i	0.9168368	-0.8251292
4	0.4666+0.7892i	0.9168368	0.8251292

MA	Rootsfp	Radius	Angle
1	0.7952-0.5199i	0.9500498	-0.4607784
4	0.7952+0.5199i	0.9500498	0.4607784
2	-0.7924-0.4327i	0.9028469	-2.102242
3	-0.7924+0.4327i	0.9028469	2.102242

(a) Non-seasonal AR and MA

SAR	Rootsfp	Radius	Angle
2	-0.4801-0.5913i	0.7616883	1.792699
1	-0.4801+0.5913i	0.7616883	-1.792699

SMA	Rootsfp	Radius	Angle
1	-0.4924+0i	0.4924131	-2.5

(b) Seasonal AR and MA

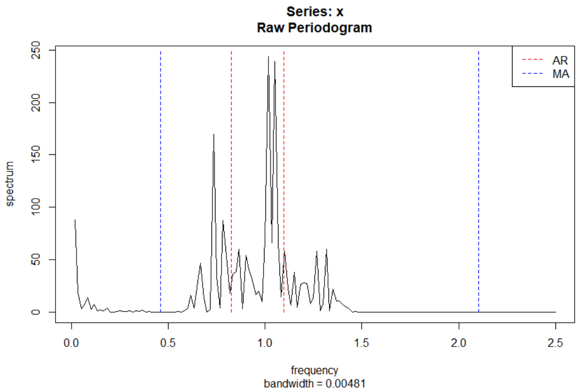

Figure 5: Raw Periodogram with ARIMA zeroes

Above, we plot the roots obtained on the periodogram in order to inspect whether the roots match the peaks and troughs of the spectrum.
We know that the seasonal roots cannot be captured by the spectrum. Whereas, the classical AR roots are not able to detect the peaks accurately, as we can see the red lines occur a little later than the spectrum peaks.
Theoretically, the spectrum seems to be split by two frequency bands, one near 0.5 and the other after 1.5, where the spectral behavior is attenuated towards zero. This indicates that there should be two moving average components. And, indeed the two zeroes of the classical MA polynomials fall under the stopband regions.
Thus, there is a possibility to get a better fitting model as the ARIMA model obtained by the `auto.arima` function is not able to detect the accurate frequencies of the dominant peaks, and the seasonal roots do not provide any significant interpretation.
The low frequency peak appears to be a result of the circulatory system's baroreflex, a low frequency regulatory mechanism that helps maintain blood pressure stability. Our model did not identify it as a true peak, potentially because the time series is relatively short compared to the period of the baroreflex.

4 The effect of the Heart Rate and Breathing Rate on the BP Signal

The given BP signal is affected by a mixture of two biological processes: heart rate and respiratory rate. Heart rate typically exhibits a periodic pattern in the BP signal. Thus the dominant peak in the frequency spectrum could correspond to the heart rate frequency. Breathing results in amplitude modulation of the sinusoidal heart rate, thus the respiratory rate may contribute to additional frequency components or sidebands around the dominant frequency.

If we assume a sinusoidal heart rate,

$$x(t) = A * sin(2\pi f_c t) \tag{1}$$

and a sinusoidal respiratory rate:

$$m(t) = B * sin(2\pi f_m t) \tag{2}$$

then the BP Signal could be written as (k belongs to [0,1])

$$y(t) = x(t) * [1 + k * m(t)] \tag{3}$$

so that the respiratory rate modulates the amplitude of the heart rate.

$$y(t) = A * sin(2\pi f_c t) * [1 + k * B * sin(2\pi f_m t)] \tag{4}$$

$$y(t) = A * sin(2\pi f_c t) + k * A * B * sin(2\pi f_c t) * sin(2\pi f_m t) \tag{5}$$

Using the trigonometric identity $2sin(a)sin(b) = cos(a - b) - cos(a + b)$

$$y(t) = A * sin(2\pi f_c t) + k/2 * A * B * [cos(2\pi f_c t - 2\pi f_m t) - cos(2\pi f_c t + 2\pi f_m t)] \tag{6}$$

According to the above calculations, we can see that the resulting BP Signal should contain 3 dominant frequency peaks in the spectrum corresponding to the frequency f_c, $f_c + f_m$, and $f_c - f_m$.
In the provided ARIMA model, we can only observe two dominant peaks at the following frequencies which are the roots of the classical AR model:

- $f_c * T_s$=1.097 Hz, 1.097 beats per second or 65.82 beats per minute.
- $(f_c - f_m) * T_s$=0.825Hz, of $f_m * T_s$=1.097-0.825=0.272 breaths per second or 16.32 breaths per minute.

The third peak at $(f_c + f_m) * T_s = 1.097 + 0.272 = 1.369$ or 82.14 beats per minute is not detected by the AR roots, but a small peak can be seen near 1.35 in the spectrum.
Graphically, this means that the periodicity in the series $y(t)$ is due to the heart rate, with a frequency of f_c. This makes sense because, during the heartbeat, pressure on the arterial walls increases, while in between heartbeats it diminishes again. Whereas, the respiratory rate is visible in the periodic volatility clustering in $y(t)$ with frequency f_m. Additionally, the frequency of heart rate f_c should depend on the frequency of respiration f_m as the heart rate fluctuates in synchrony with breathing cycles, an effect called the RSA (respiratory sinus arrhythmia).

5 Prediction of the Series

The model discovered via the automated procedure can also be used to forecast the patient's future evolution in blood pressure. We forecast the next minute.

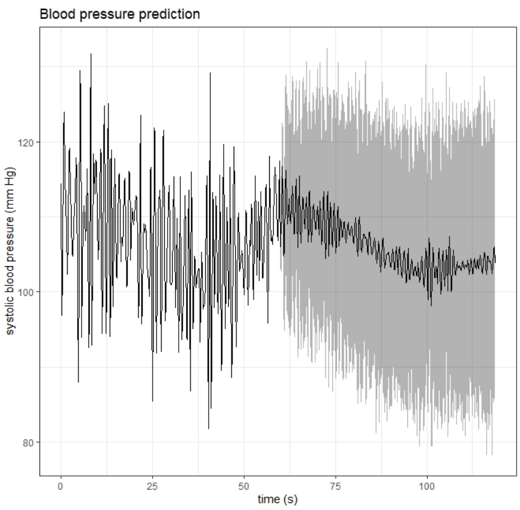

Figure 6: ARIMA prediction with 95% prediction interval