UCLouvain
Faculté des sciences

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
LOUVAIN SCHOOL OF STATISTICS

LSTAT2170 - Time series

Final Project

LIONEL LAMY
1294-1700

April 19, 2021

Contents

1 Introduction

In this project we will focus on the analysis of real data. These data are the U.S. Natural Gas State Data and concern the monthly quantity of natural gas delivered to residential and commercial consumers (excluding vehicle fuel) in Florida. Aggregated on a monthly basis, the data are presented in millions of cubic feet (MMcf[1]) and cover the period from January 1991 to August 1999.

We will proceed to the beginning of this report with a first visual discovery of the dataset. Then, we will eventually apply a series of transformations to stabilize the variance, remove possible trends and seasonality. Next, we will analyze the autocorrelation and partial autocorrelation functions in order to have a first intuition on the type of model to fit. Following this, we will establish which model would be the most appropriate for our data and verify the insight of our choice by several methods such as a significance test of the coefficients, an analysis of the residuals (by a Portmanteau test) or the evaluation of the predictive capacity "on sample". The final objective is to be able to give a prediction interval for future values over roughly one year.

2 Data discovery

By decomposing the time series into pieces we can have a first insight of the data.

The first obvious observation we can make by looking at just the first line is that we are dealing with seasonal data. Indeed, we can see in the third row an almost perfect seasonality with maximum values at the beginning of each year and minimum values towards the middle. These results are hardly surprising given the nature of the data. Indeed, it seems normal that a greater quantity of gas is used during winter and that gas consumption decreases during summer.

Decomposition of the deliveries of natural gas

---
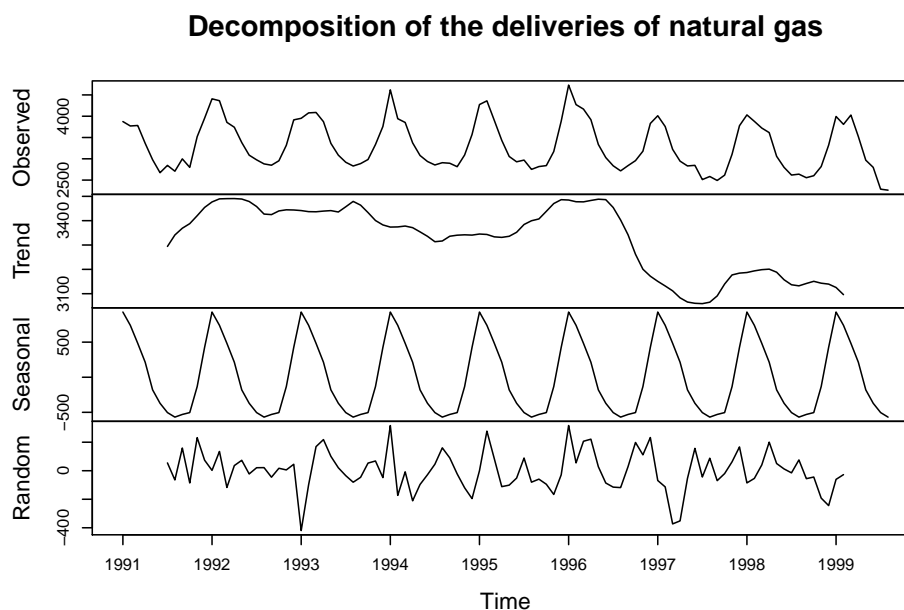
[1]   "Mcf" means 1,000 cubic feet of natural gas; "MMcf" means 1,000 Mcf.

Looking at the second line of the graph, we see that the data do not really seem to vary except in 1996-1997 when a decrease is noticeable. A closer look at the first line shows this phenomenon.

Another interesting way to present the data is to display the evolution of the deliveries by stacking the years line by line. The outcome is fortunately the same. We notice a strong seasonality (since the lines all follow the same pattern) as well as a small decrease that seems to start from 1997 (colored dashed lines). On the other hand, the lines seems to remain relatively close to each other and there is no drastic change in variance.

**Superposed monthly deliveries of gas**
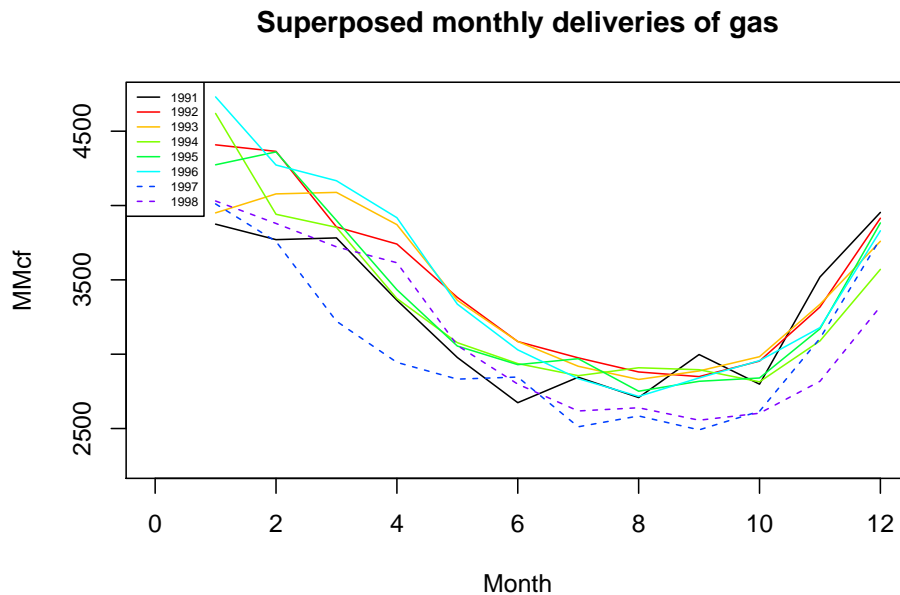
3   Data transformation

The previous statements make us decide not to apply a logarithmic transformation on the data. On the other hand, as we aim to examine the correlation structure of the residuals in the next section, we need to have stationary data. Meaning that it is necessary to deseasonalize and delinearize the data in order to go further with the visual analysis.

3.1   Deseasonalize and detrend

For this purpose, we will use the method of (iterated) differences. As a first step, we will first-differencing the time series to remove the linear trend. Next, we will repeat this process at a lag equal to the periodicity. That is, we will use the value 12 since the data is collected monthly and that the seasonality is annual.

Detrended time series

Time
Difference of order 1
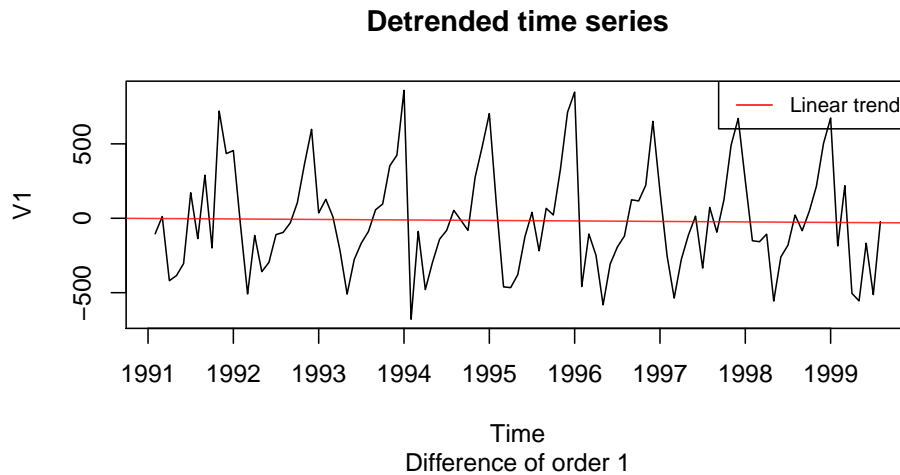
In the first figure above, we observe that the linear trend is well removed since the data are now centered around zero but we still have a visible periodicity. This particularity seems to be corrected on the second graph below and we now have a time series that appears to be random errors.

Detrended and deseasonalized time series

Time
Difference of order 1 at lag 12
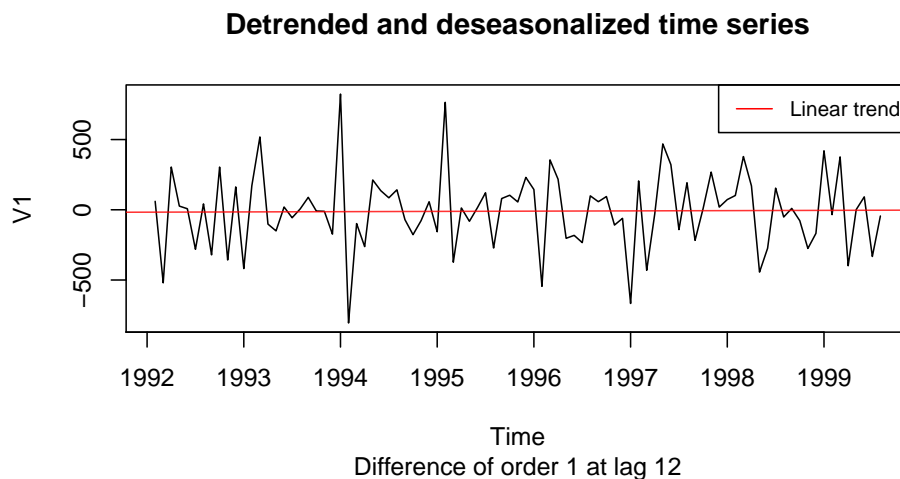
## 3.2   Modelling the stationary component

As the simultaneous look on the ACF and PACF does not suggest a pure moving-average (MA) or autoregressive (AR) model as we have oscillating autocorrelations. We might therefore move towards an ARMA model with values .

> **Note 1**
>
> Voir pages 135 à 140 pour la théorie concernant les processus SARIMA et les étapes à suivre
>
> Il faut regarder les lags 1,2,3,4 pour la dépendance annuelle (P,Q) mais seulement [0,1[(12-1 lags) pour la mensuelle (p,q)
>
> On a seulement enlevé un élément deterministique de saisonnalité, si correlation significative au lag 1, alors il pourrait rester une composante stochastique.. donc on fit un S-ARIMA !
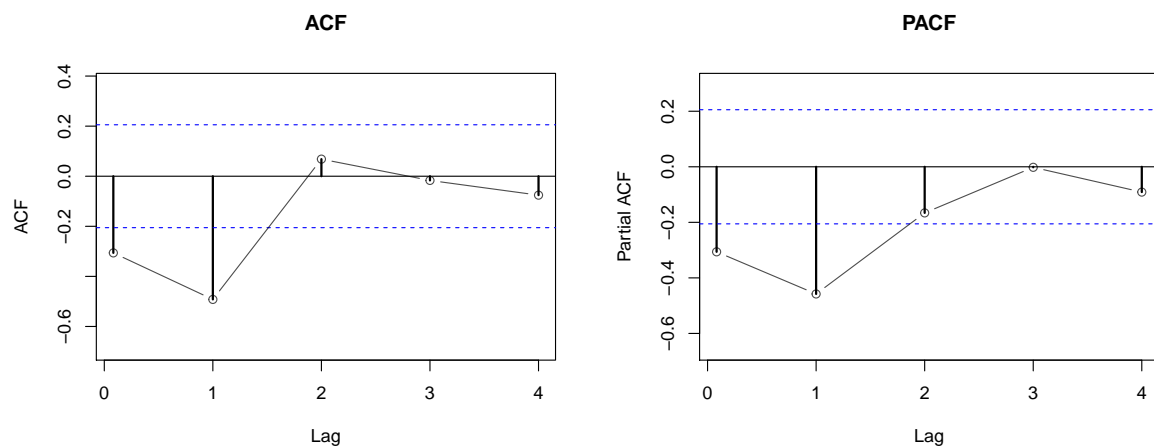
4 Models selection

4.1 Definition

4.2 Selection

4.3 Comparison

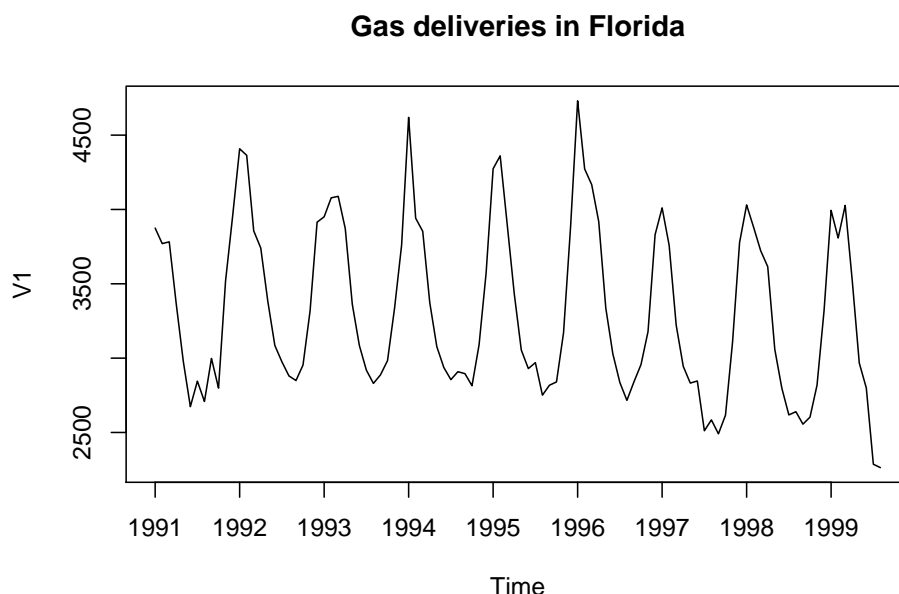## 5 Models validation

5.1 Coefficients

5.2 Portemanteau

5.3 Predictive power

6 Predictions

7 Conclusion

A Appendix

A.1 Figures

Gas deliveries in Florida

A.2 Code

> **Note**
>
> For reproducibility purposes, the complete R project containing the source code and the results is available on https://github.com/lamylio/LSTAT2170-Project.
>
> The below section is automatically generated and tidied.

```r
#' Import facility functions in a separate attached environment
#' To keep our global clean

facilities <- new.env()
source("./resources/scripts/fonctionsSeriesChrono.R", local = facilities)

# Please check the github repository
source("./resources/scripts/custom_plots.R", local = facilities)
source("./resources/scripts/sarima_model_selection.R", local = facilities)
source("./resources/scripts/ts_significance_test.R", local = facilities)

attach(facilities, name = "facilities")

# Import the gas dataset
gas <- read.table("./resources/data/gasflorida.txt", header = F)
gas <- ts(gas, start = 1991, frequency = 12)

# Define some useful variables
gas.start <- attributes(gas)$tsp[1]
```

```r
gas.end <- attributes(gas)$tsp[2]
gas.freq <- attributes(gas)$tsp[3]
gas.t <- seq(gas.start, to = gas.end, length = length(gas))
gas.xaxp <- c(floor(gas.start), floor(gas.end), floor(gas.end - gas.start))

# Plot the decomposition of the data (see custom_plot.R)
plot.decompose.ts(gas, main = "Decomposition of the deliveries of natural gas",
    cex.lab = 0.9, cex.axis = 0.95, xaxp = gas.xaxp)

# Plot the superposed view of the data (see custom_plot.R)
plot.superposed.ts(gas, title = "Superposed monthly deliveries of gas",
    xlab = "Month", ylab = "MMcf", dashed_thick_from = c(6, 8), xlim = c(0,
        12))

# Remove the seasonality using lag 12 as we have monthly data
gas.1 <- diff(gas, lag = 1, differences = 1)
# Remove global trend
gas.2 <- diff(gas.1, lag = 12, differences = 1)

# Removed trend
plot(gas.1, main = "Detrended time series", sub = "Difference of order 1",
    xaxp = gas.xaxp)
abline(reg = lm(gas.1 ~ tail(gas.t, -1)), col = rgb(1, 0, 0, 0.8),
    lty = 1)
legend("topright", legend = "Linear trend", col = 2, cex = 0.8, lty = 1)
# Removed trend and seasonality
plot(gas.2, main = "Detrended and deseasonalized time series", sub = "Difference of order 1 at
    xaxp = gas.xaxp)
abline(reg = lm(gas.2 ~ tail(gas.t, -12 - 1)), col = rgb(1, 0, 0,
    0.8), lty = 1)
legend("topright", legend = "Linear trend", col = 2, cex = 0.8, lty = 1)

# Plot the ACF and PACF
plot.acf.pacf(gas.2)

############
#' Appendix #

# Plot the complete dataset non decomposed
plot(gas, main = "Gas deliveries in Florida", xaxp = gas.xaxp)
```