ANALYSIS OF EMERGENCY DEPARTMENT DATA: EFFECT OF COVID-19

Business Economic and Financial Data Project

Dacia Braca   Michele M. Crudele   Simone Mistrali

January 28, 2022

UNIVERSITÀ DEGLI STUDI DI PADOVA

Outline

A legitimate question arises. . .

Why are we considering emergency department data for a business and economic course?

Introduction

- All *human activities* have significant economic implication on the market;
- to guarantee quality in health services, the administrations have to deal with the costs of managing and maintaining the hospital facilities.

Introduction

- The appearance of *Covid-19* has redefined our priorities and underlined the importance of the issue;
- both the virus and the government decisions to deal with it are having a major impact on global society, from sociology to economics.

- In this project we have focused on Emergency Room data of the *Policlinico di Bari* hospital;
- *objectives* are to draw a descriptive framework on the hospital management in times of normality and make a comparison with the Covid-19 period.

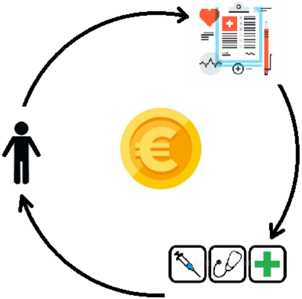
An economic perspective

Let us see more in detail these economic aspects. . .

There are several economic aspects we can directly relate to the hospitalization:

- essential public services covered by taxes;
- additional services that require a contribution from the individual patient.

This is needed to cover costs of medical machinery, maintenance of hospital structures and to guarantee an adequate health system.



Hospital-Related Business

There are economies and single people that depend on the hospital system, e.g.:

- the staff who works within the facilities;
- patients get sick and withdraws from work, causing an economic loss;
- pharmacies, catering, hotel or cleaning associated businesses.

Extra Costs due to Covid

The spread of the Covid-19 virus has deeply influenced the business related to hospitals:

- new and dedicated machinery and materials;
- increase of the salaries of medical staff due to more required hours of work per day.

On the other hand, one should take into account the closure of companies, shops, restaurants and any other kind of activity due to *lockdown*.

The dataset

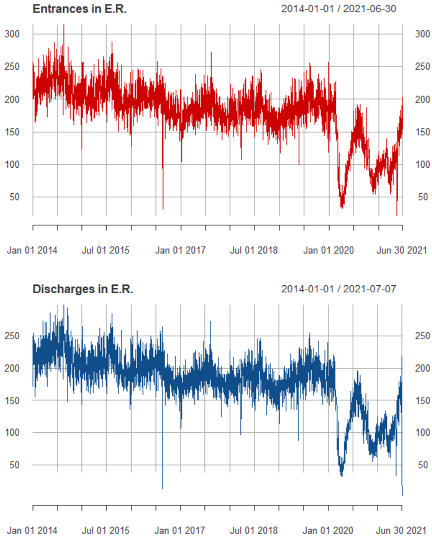
The data we have collect admissions in the E.D. of Bari from January 2014 to June 2021.
For each admission to the emergency room, lots of information about the patient are collected, e.g.:

- name, sex, id code, birth, residence and nationality;
- reason of recovery, admission and discharge date, type of discharge.

Using these data, we were able to build two time series: one for the admissions and one for the discharges.

Università degli Studi di Padova

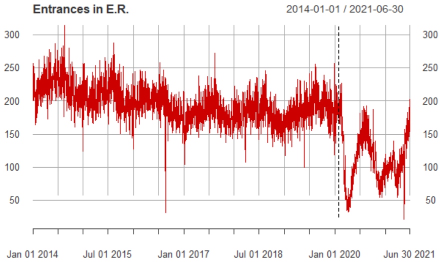
- They are pretty similar: same trend and seasonality, only shifted of a few days;
- seasonal harmonic behavior with a peak in the summertime and a drop during Christmas holidays;
- minor peaks due to weekly seasonality.

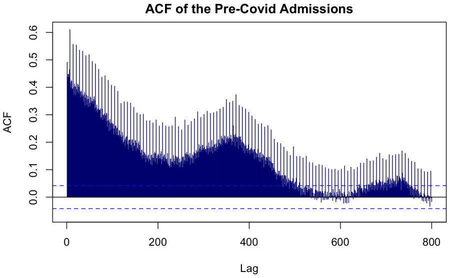


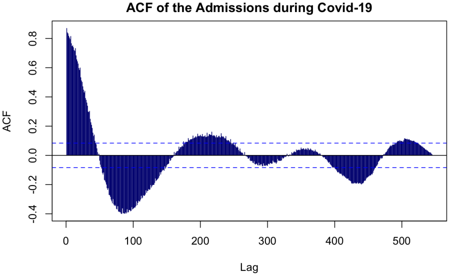
Università degli Studi di Padova

Covid-19 dramatically changed their behavior:

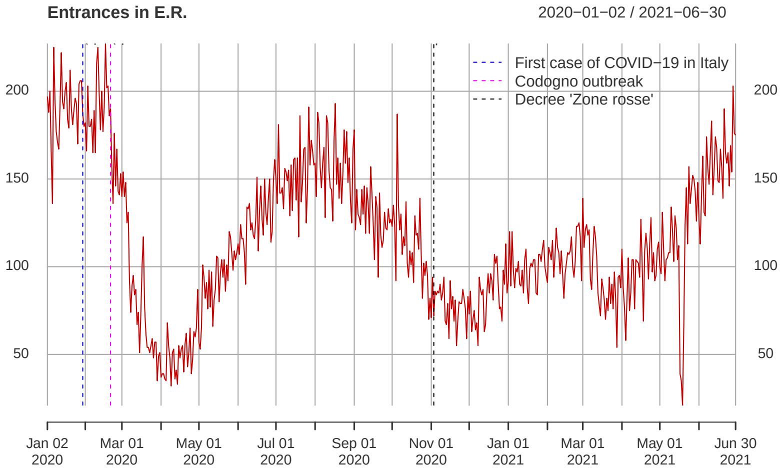
- no more linear trend
- no more seasonality
- two strong negative shocks due to two periods of *lockdown*.







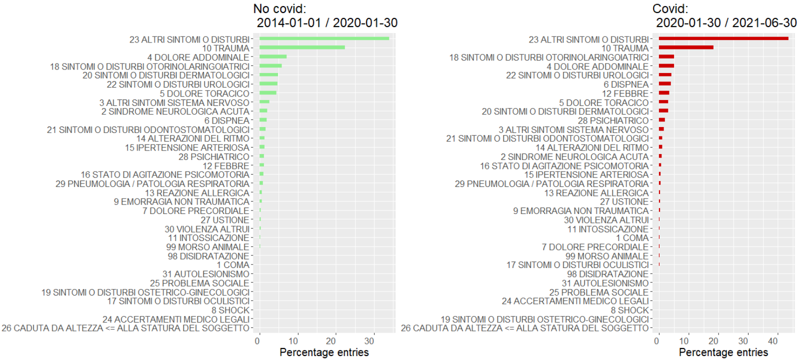
Covid-19 Hot Dates in Italy



EDA: reason for entry

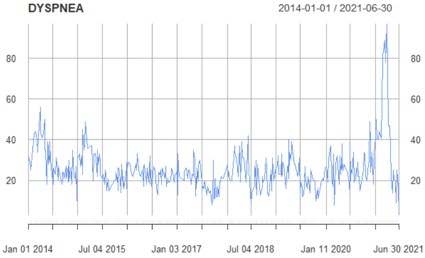
Università degli Studi di Padova

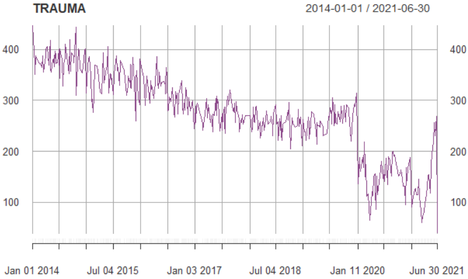
The hospital admits 33 possible reasons for entering the emergency room.

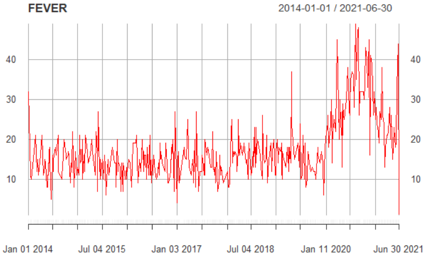


EDA: reason of recovery

- the entries curve for *trauma* suffers a violent negative shock due to Covid restrictions;
- differently, the entries for *dyspnea* and *fever* present a peak.

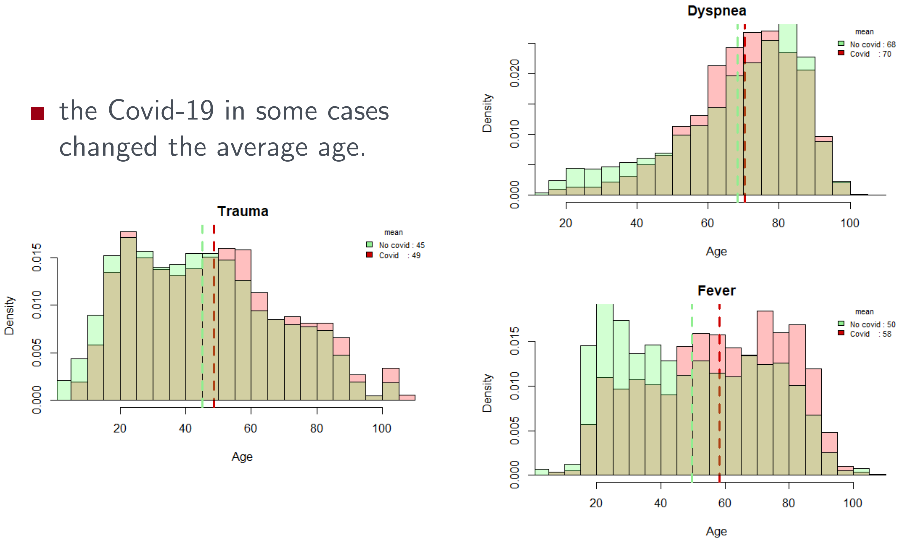






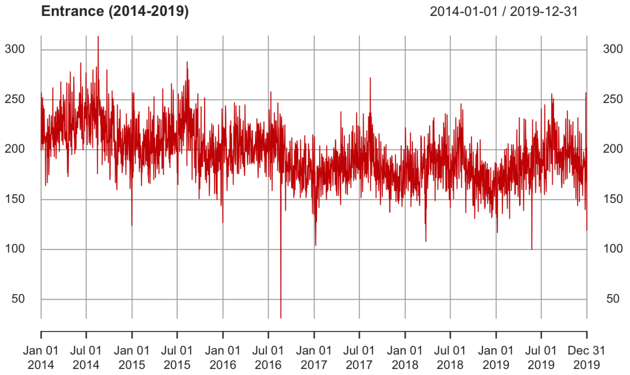
EDA: reason for entry vs age

- the Covid-19 in some cases changed the average age.



Pre-Covid Time Series - Fitting

We first tested Linear Models, then we try to improve the results using Generalized Additive Models.



Linear Models

Università degli Studi di Padova

Different Linear Models were tested using different combinations of trend, seasonality and external regressor as explanatory variables. The monthly seasonality was found to be never significant.

- AIC(lm(entrance $\sim$ trend + sy)) = 20146.6
- AIC(lm(entrance $\sim$ trend + sw + sy)) = 19586.93

- AIC(lm(entrance $\sim$ trend + disch.)) = 17694.57
- AIC(lm(entrance $\sim$ trend + sw + disch.)) = 17043.15
- AIC(lm(entrance $\sim$ trend + sw + sy + disch.)) = 17303.74

Best Linear Model

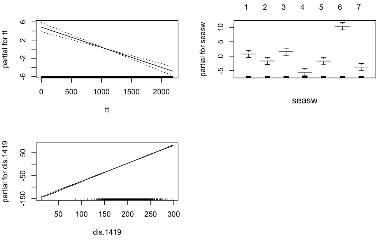
The best model (good fit, significance of explanatory variable and lowest AIC) was:

- lm(entrance $\sim$ trend $+$ sw $+$ discharges)

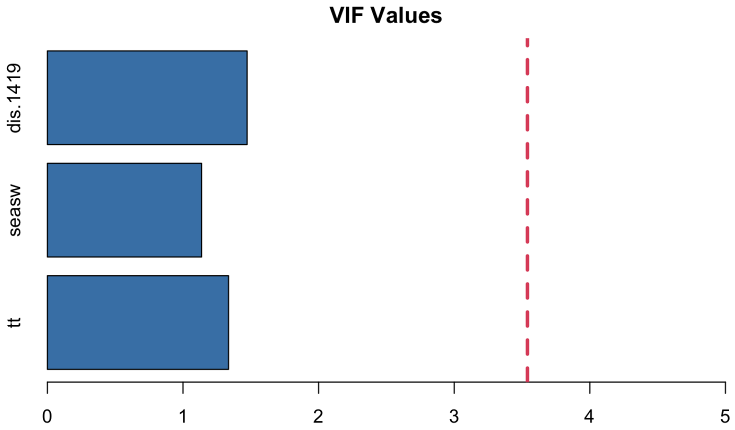
Best Linear Model - Summary

- lm(entrance ∼ trend + sw + discharges)

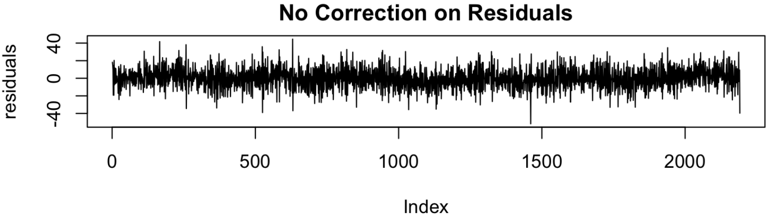
```
[1] 17043.15 = AIC

Call:
lm(formula = entrance ~ tt + seasw + dis.1419)

Residuals:
    Min      1Q  Median      3Q     Max
-51.759  -7.516  -0.320   7.456  44.471

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 46.0649187  2.5318262  18.194  < 2e-16 ***
tt          -0.0044186  0.0004606  -9.593  < 2e-16 ***
seasw2      -2.4227253  0.9436451  -2.567   0.0103 *
seasw3       0.7915665  0.9432309   0.839   0.4014
seasw4      -6.3245889  0.9435271  -6.703 2.59e-11 ***
seasw5      -2.4563708  0.9566528  -2.568   0.0103 *
seasw6       9.6244255  0.9546484  10.082  < 2e-16 ***
seasw7      -4.5289829  0.9494023  -4.770 1.96e-06 ***
dis.1419     0.7925524  0.0110858  71.492  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.8 on 2182 degrees of freedom
Multiple R-squared:  0.8117,    Adjusted R-squared:  0.811
F-statistic:  1176 on 8 and 2182 DF,  p-value: < 2.2e-16


        Durbin-Watson test

data:  best.ent
DW = 2.7475, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is not 0
```

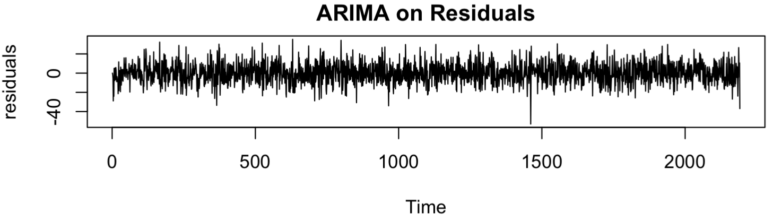


Best Linear Model - VIF

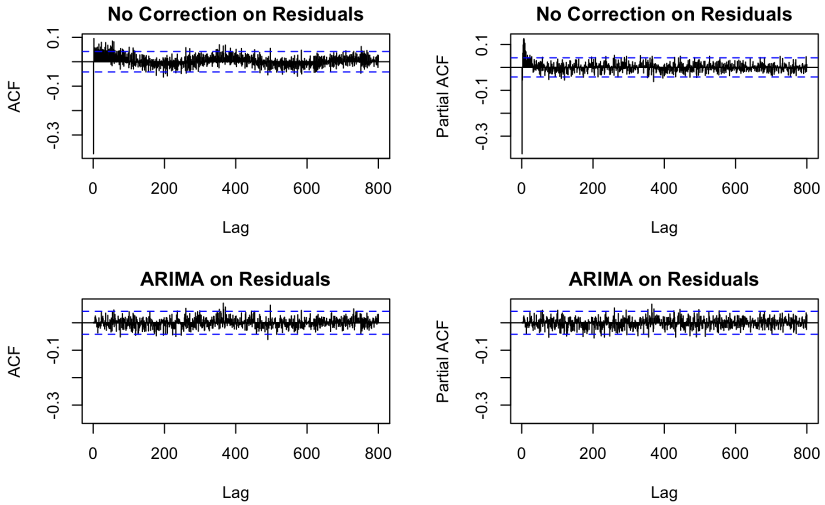


Best Linear Model - Residuals

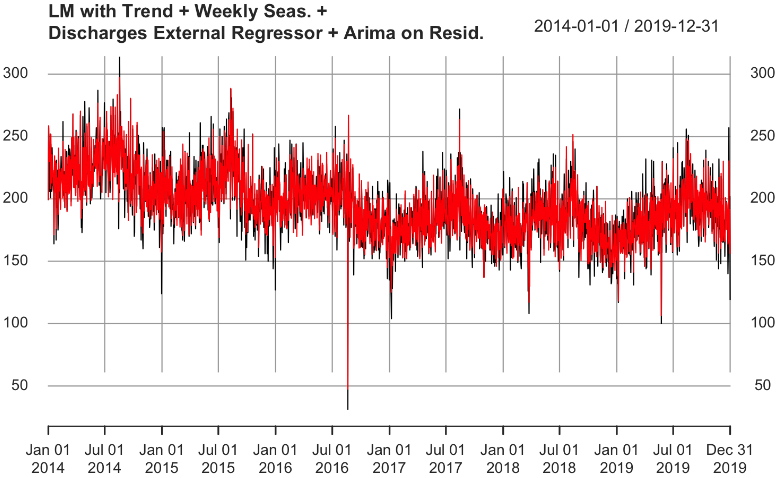




Best Linear Model - ACF of Residuals



Best Linear Model - Fit



Generalized Additive Models (GAMs)

Università degli Studi di Padova

Different GAMs were tested to improve results. Also in this case, the monthly seasonality was found to be never significant.

- $\text{AIC}(\text{gam}(\text{entrance} \sim \text{lo}(\text{trend}))) = 19884$
- $\text{AIC}(\text{gam}(\text{entrance} \sim \text{lo}(\text{trend}) + \text{sy})) = 19825$
- $\text{AIC}(\text{gam}(\text{entrance} \sim \text{lo}(\text{trend}) + \text{sw} + \text{sy})) = 19159$
- $\text{AIC}(\text{gam}(\text{entrance} \sim \text{lo}(\text{trend}) + \text{lo}(\text{dis.}))) = 17341$
- $\text{AIC}(\text{gam}(\text{entrance} \sim \text{lo}(\text{trend}) + \text{sw} + \text{lo}(\text{dis.}))) = 17010$
- $\text{AIC}(\text{gam}(\text{entrance} \sim \text{lo}(\text{trend}) + \text{sy} + \text{lo}(\text{dis.}))) = 17644$
- $\text{AIC}(\text{gam}(\text{entrance} \sim \text{lo}(\text{trend}) + \text{sw} + \text{sy} + \text{lo}(\text{dis.})))=17203$

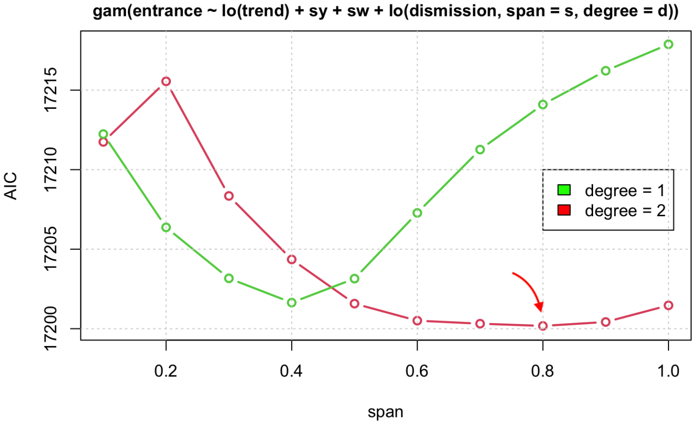
Generalized Additive Models

The two most promising models (good fit and low AIC) were:

- gam(entrance $\sim$ lo(trend) $+$ sw $+$ lo(discharges))
- gam(entrance $\sim$ lo(trend) $+$ sy $+$ sw $+$ lo(discharges))

For each of those models, a grid search on the parameters of the smoothing terms was then implemented.

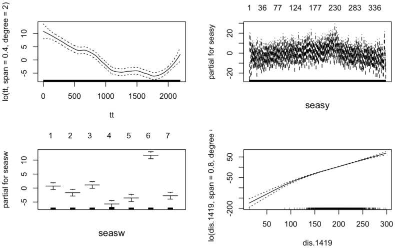
GAMs - Grid Search



GAMs - Grid Search

$$gam(entrance \sim lo(trend, span = 0.4, degree = 2) + sy + sw + lo(dismission, span = 0.8, degree = 2))$$

```
AIC: 17183.13

Number of Local Scoring Iterations: NA

Anova for Parametric Effects
                                      Df Sum Sq Mean Sq   F value    Pr(>F)
lo(tt, span = 0.4, degree = 2)       2.0 486662  243331 1911.3698 < 2.2e-16 ***
seasy                              364.0 338958     931    7.3146 < 2.2e-16 ***
seasw                                6.0 207474   34579  271.6188 < 2.2e-16 ***
lo(dis.1419, span = 0.8, degree = 2)   2.0 330983  165491 1299.9386 < 2.2e-16 ***
Residuals                         1808.7 230266     127
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Anova for Nonparametric Effects
                                  Npar Df Npar F     Pr(F)
(Intercept)
lo(tt, span = 0.4, degree = 2)        5.3 5.8581 1.348e-05 ***
seasy
seasw
lo(dis.1419, span = 0.8, degree = 2)   1.9 7.4559 0.0006957 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```



Università degli Studi di Padova

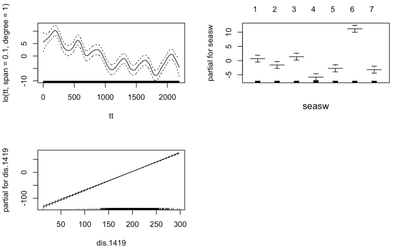
summary(gam(entrance $\sim$ *lo(trend)* $+$ *sw* $+$ *lo(dismission))*

```
Anova for Parametric Effects
                              Df Sum Sq Mean Sq F value    Pr(>F)    
lo(tt)                       1.0 384957  384957 2813.75 < 2.2e-16 ***
seasw                        6.0 215178   35863  262.13 < 2.2e-16 ***
lo(dis1419$dismission)       1.0 673021  673021 4919.28 < 2.2e-16 ***
Residuals                 2176.9 297828     137                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Anova for Nonparametric Effects
                          Npar Df  Npar F     Pr(F)    
(Intercept)                                            
lo(tt)                        2.3 17.2313 5.636e-09 ***
seasw                                                  
lo(dis1419$dismission)        2.8  1.5567    0.2006    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

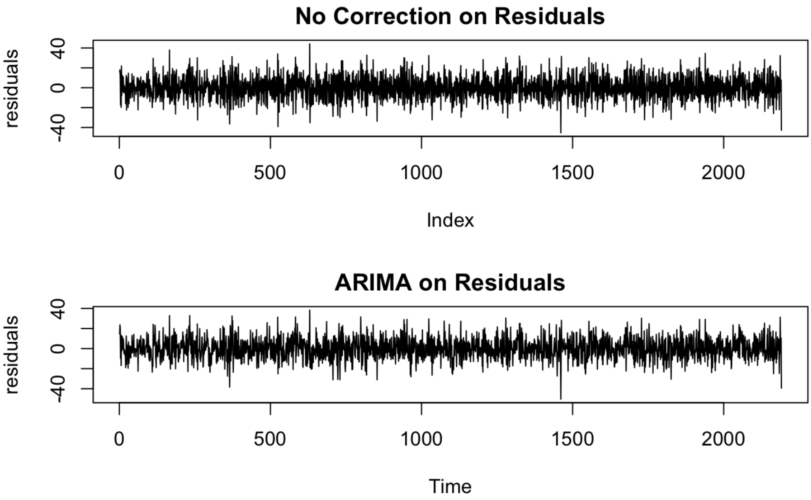
Best GAM - Summary

$$gam(entrance \sim lo(trend, span{=}0.1, deg{=}1) + sw + discharges)$$

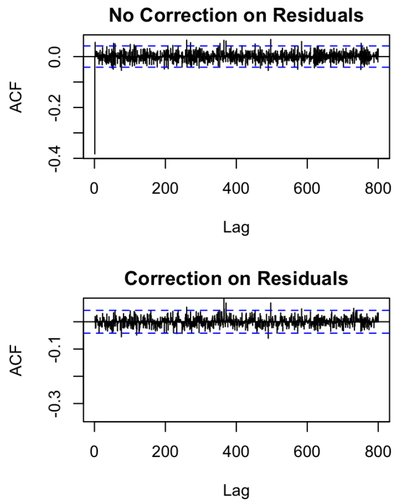
```
Call: gam(formula = entrance ~ lo(tt, span = 0.1, degree = 1) + seasw +
    dis.1419)
Deviance Residuals:
     Min       1Q   Median       3Q      Max
-45.4092  -7.2998  -0.2977   7.3777  44.1508

(Dispersion Parameter for gaussian family taken to be 131.8966)

    Null Deviance: 1613193 on 2190 degrees of freedom
Residual Deviance: 285590.2 on 2165.259 degrees of freedom
AIC: 16941.88

Number of Local Scoring Iterations: NA

Anova for Parametric Effects
                                   Df Sum Sq Mean Sq F value    Pr(>F)
lo(tt, span = 0.1, degree = 1)    1.0 383522  383522 2907.74 < 2.2e-16 ***
seasw                             6.0 214533   35756  271.09 < 2.2e-16 ***
dis.1419                          1.0 598539  598539 4537.94 < 2.2e-16 ***
Residuals                      2165.3 285590     132
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Anova for Nonparametric Effects
                               Npar Df Npar F     Pr(F)
(Intercept)
lo(tt, span = 0.1, degree = 1)    16.7 8.2041 < 2.2e-16 ***
seasw
dis.1419
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

        Durbin-Watson test

data:  best.ent
DW = 2.7475, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is not 0
```



Best GAM - Residuals

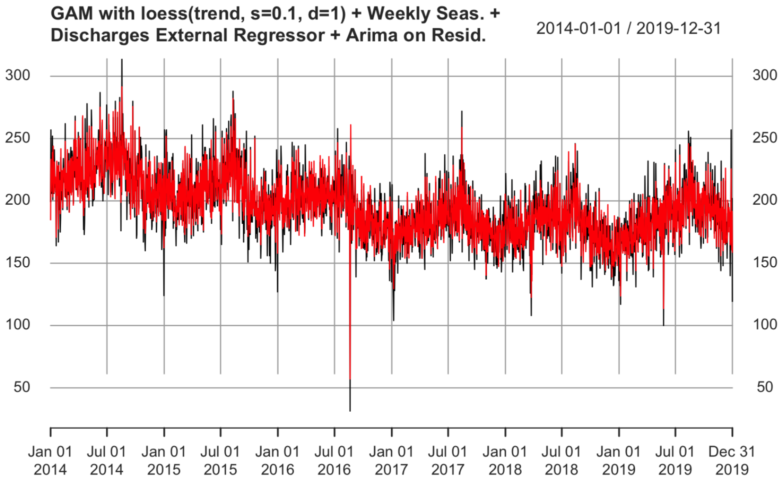


Best GAM - ACF of Residuals



Best GAM - Fit

Università degli Studi di Padova



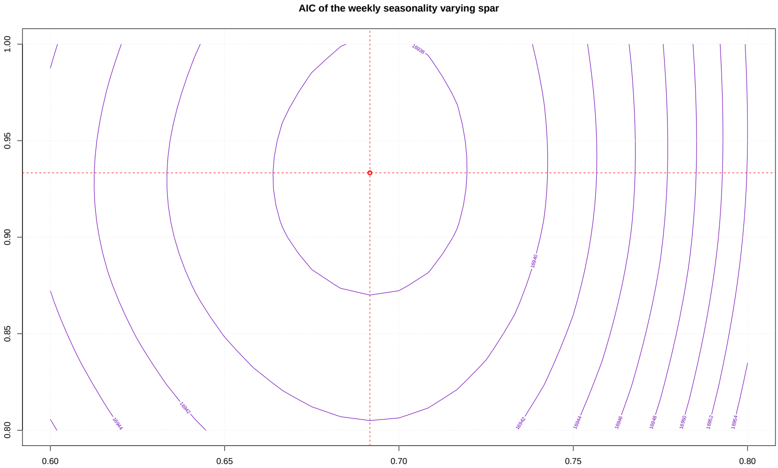
GAMs with Smoothing Splines

The two most promising models are, as in the previous case:

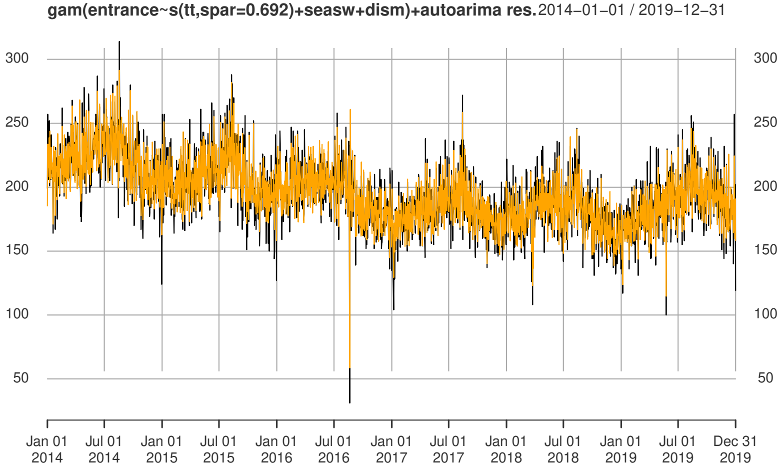
- `gam(entrance ~ s(trend) + sw + discharges)`
- `gam(entrance ~ s(trend) + sw +sy + s(discharges))`

The smoother `s()` has only one parameter to be set: `spar`, that typically assumes values in $(0, 1]$.

GAMs with s() - Grid Search on s

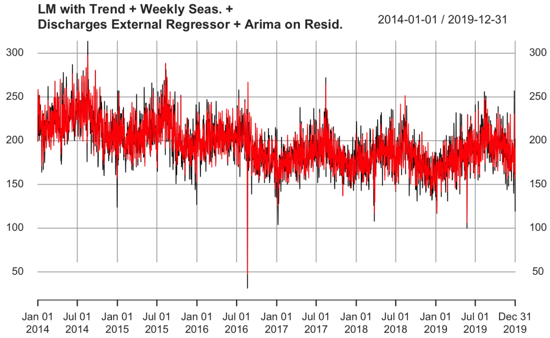


GAMs with s() - Best Fit



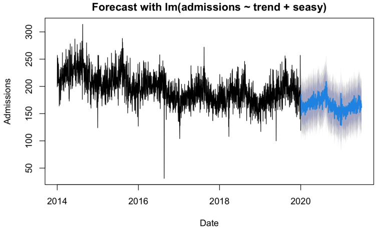
Pre-Covid Fitting Results

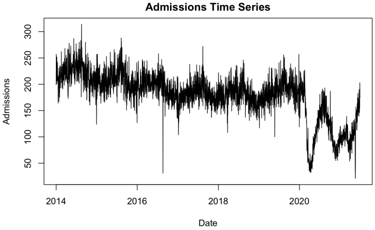
Model	Best AIC
LM	17043
GAM `lo`	16941
GAM `s`	16941



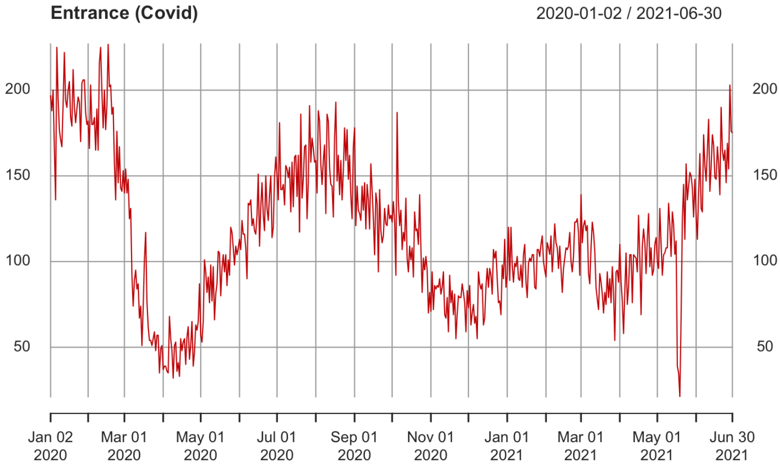
Università degli Studi di Padova

We eventually compared the forecast obtained with a linear model with the real data.

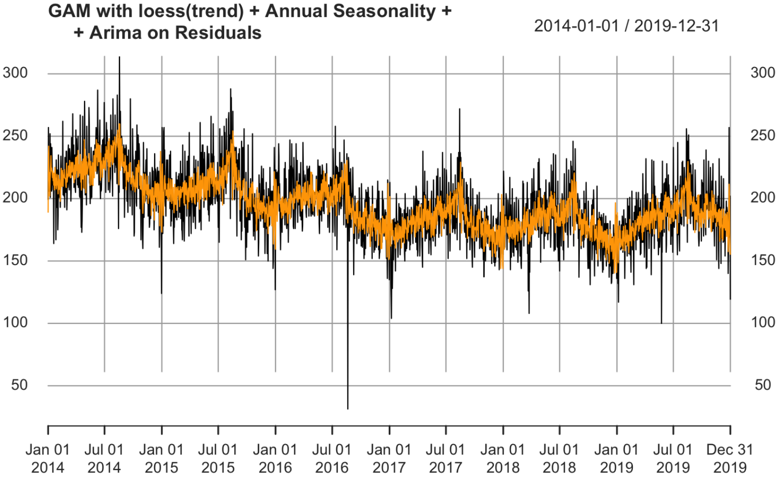




Fitting of Covid Time Series



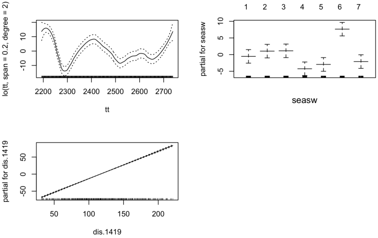
Significance of Seasonality



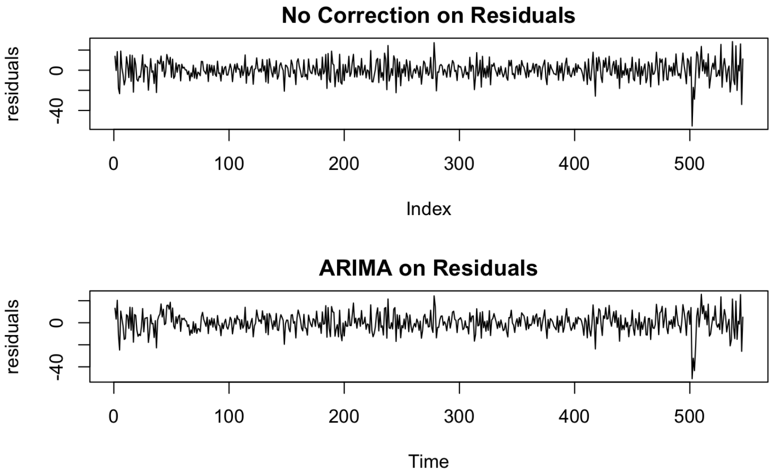
Covid - Best Model

- gam(entrance $\sim$ lo(trend, span $=$ 0.2, deg $=$ 2) $+$ sw $+$ dis)

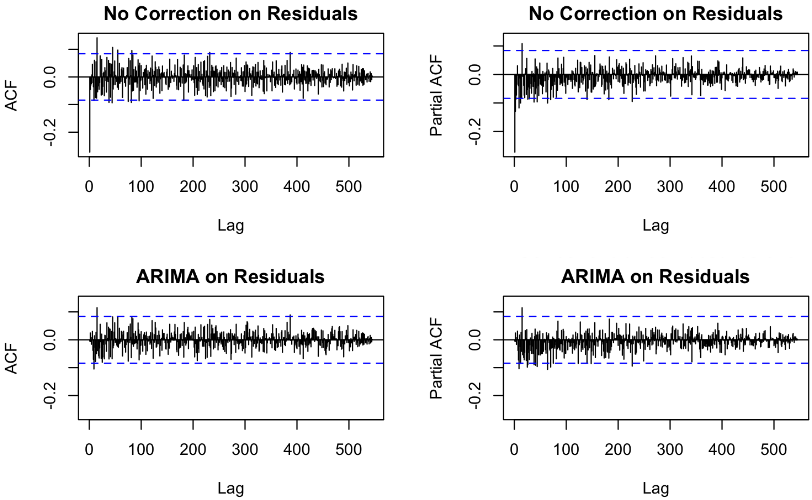
```
AIC: 4055.071

Number of Local Scoring Iterations: NA

Anova for Parametric Effects
                                 Df Sum Sq Mean Sq  F value    Pr(>F)
lo(tt, span = 0.2, degree = 2)   2.00 131476   65738  699.057 < 2.2e-16 ***
seasw                            6.00  25403    4234   45.023 < 2.2e-16 ***
dis.1419                         1.00 504161  504161 5361.225 < 2.2e-16 ***
Residuals                      522.76  49160      94
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Anova for Nonparametric Effects
                               Npar Df Npar F     Pr(F)
(Intercept)
lo(tt, span = 0.2, degree = 2)    13.2  4.196 9.416e-07 ***
seasw
dis.1419
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

        Durbin-Watson test

data:  best.ent
DW = 2.6733, p-value = 9.179e-15
alternative hypothesis: true autocorrelation is not 0
```



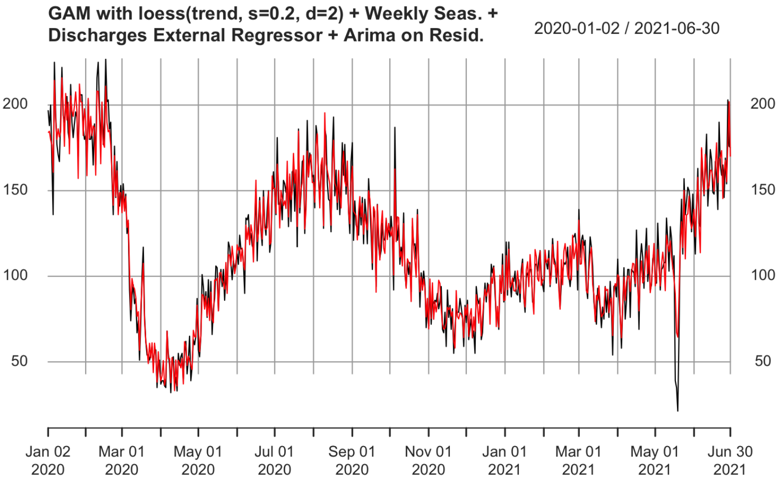
Covid - Best Model - Residuals



Covid - Best Model - ACF of Residuals



Covid - Best Fit



Conclusions

- we fit the pre-covid time series using linear models and we have seen that the GAMs didn't improve results that much;
- the covid time series is less regular and predictable: we didn't manage to fit it using linear models and GAMs weren't able to fit those data by using only trend and seasonality;
- we have quantitatively confirmed what we have seen in a qualitative way in the beginning, that is the fact that covid-19 caused a dramatic shock in the time series;