# SMDS Final Project

## Covid-19 case study, Veneto's ICUs

Lorenzo Cabriel, Mattia Pividori, Renzo Testa

27 January 2022

Università degli Studi di Trieste

# Introduction

- This project aims to study the trend of the intensive care units (ICUs) in Veneto during the Covid-19 Second Wave, in particular during the period from October 1$^{st}$ 2020 to February 1$^{st}$ 2021, in order to obtain a model able to reproduce the dynamics of the number of ICUs.
- The intended outcome is to predict the total number of occupied ICUs over a 1-2 week period to ideally try to provide a useful tool for the healthcare professionals to optimally allocate the available units.
- In this process, we have deployed different techniques learned during the course and tried different approaches to the problem from standard linear models to generalized linear models (GLM) and time-series models like ARIMA, an Auto-Regression(AR) and Moving Average(MA) model.

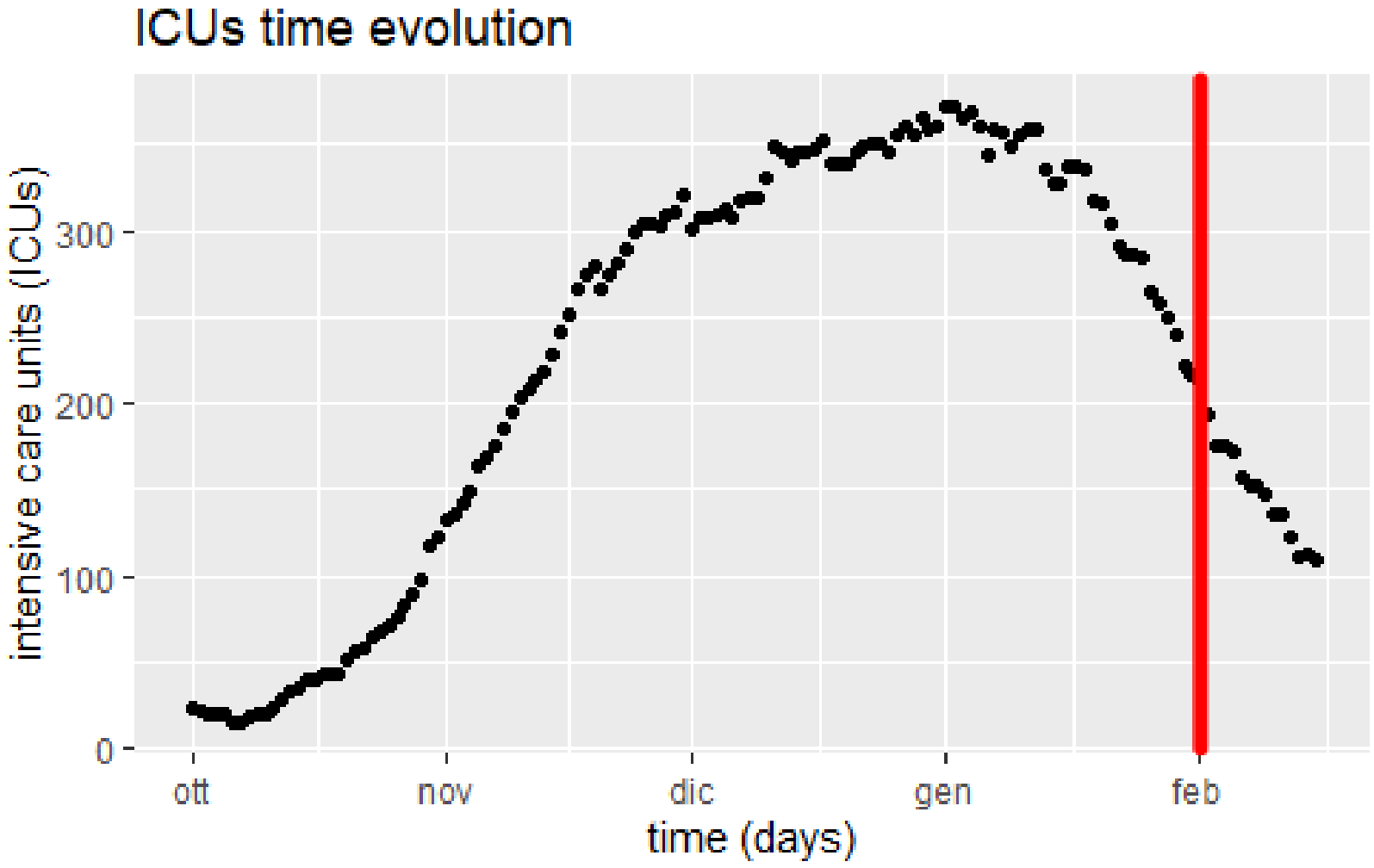
# The Dataset

We merged different datasets coming from three different sources. We filtered it for the period of interest, we dropped columns carrying many missing values and we took also the differences of each variable, to allow us modelling them in addition to their levels.

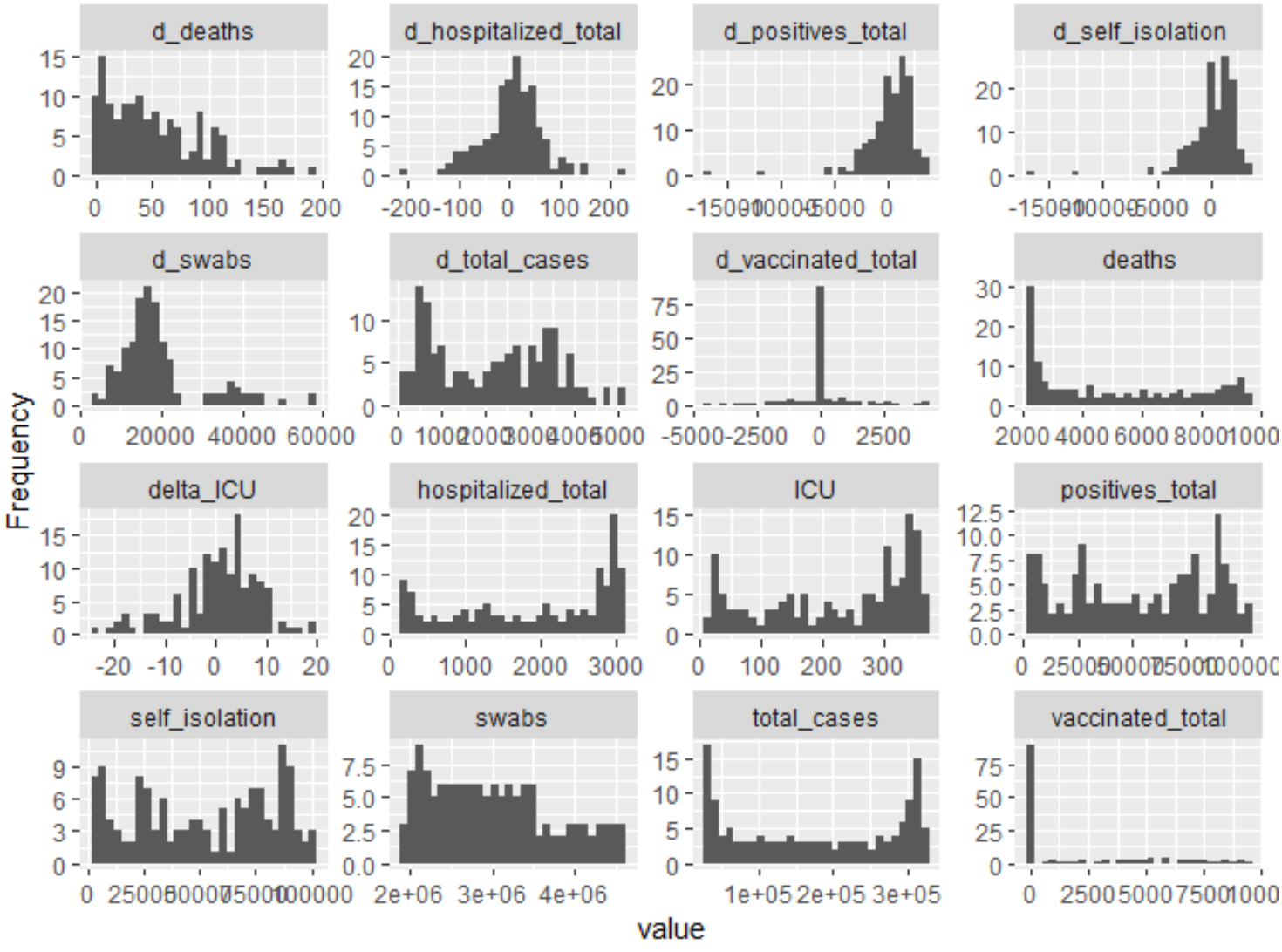
```
tibble [696 x 19] (S3: tbl_df/tbl/data.frame)
 $ date                : Date[1:696], format: "2020-02-24" "2020-02-25" "2020-02-26" "2020-02-27" ...
 $ ICU                 : num [1:696] 4 7 8 8 9 11 13 14 19 23 ...
 $ hospitalized_total  : num [1:696] 16 19 24 27 33 35 64 67 68 99 ...
 $ self_isolation      : num [1:696] 16 23 45 82 116 154 197 204 229 246 ...
 $ positives_total     : num [1:696] 32 42 69 109 149 189 261 271 297 345 ...
 $ deaths              : num [1:696] 1 1 2 2 2 2 2 2 3 6 ...
 $ total_cases         : num [1:696] 33 43 71 111 151 191 263 273 307 360 ...
 $ swabs               : num [1:696] 2200 3780 4900 6164 7414 ...
 $ vaccinated_total    : num [1:696] 0 0 0 0 0 0 0 0 0 0 ...
 $ residents           : num [1:696] 4879133 4879133 4879133 4879133 4879133 ...
 $ days                : int [1:696] 1 2 3 4 5 6 7 8 9 10 ...
 $ delta_ICU           : num [1:696] 4 3 1 0 1 2 2 1 5 4 ...
 $ d_hospitalized_total: num [1:696] 16 3 5 3 6 2 29 3 1 31 ...
 $ d_self_isolation    : num [1:696] 16 7 22 37 34 38 43 7 25 17 ...
 $ d_positives_total   : num [1:696] 32 10 27 40 40 40 72 10 26 48 ...
 $ d_total_cases       : num [1:696] 33 10 28 40 40 40 72 10 34 53 ...
 $ d_swabs             : num [1:696] 2200 1580 1120 1264 1250 ...
 $ d_vaccinated_total  : num [1:696] 0 0 0 0 0 0 0 0 0 0 ...
 $ d_deaths            : num [1:696] 1 0 1 0 0 0 0 0 1 3 ...
```

Sources: • COVID-19 Data • Vaccines data • ISTAT Data

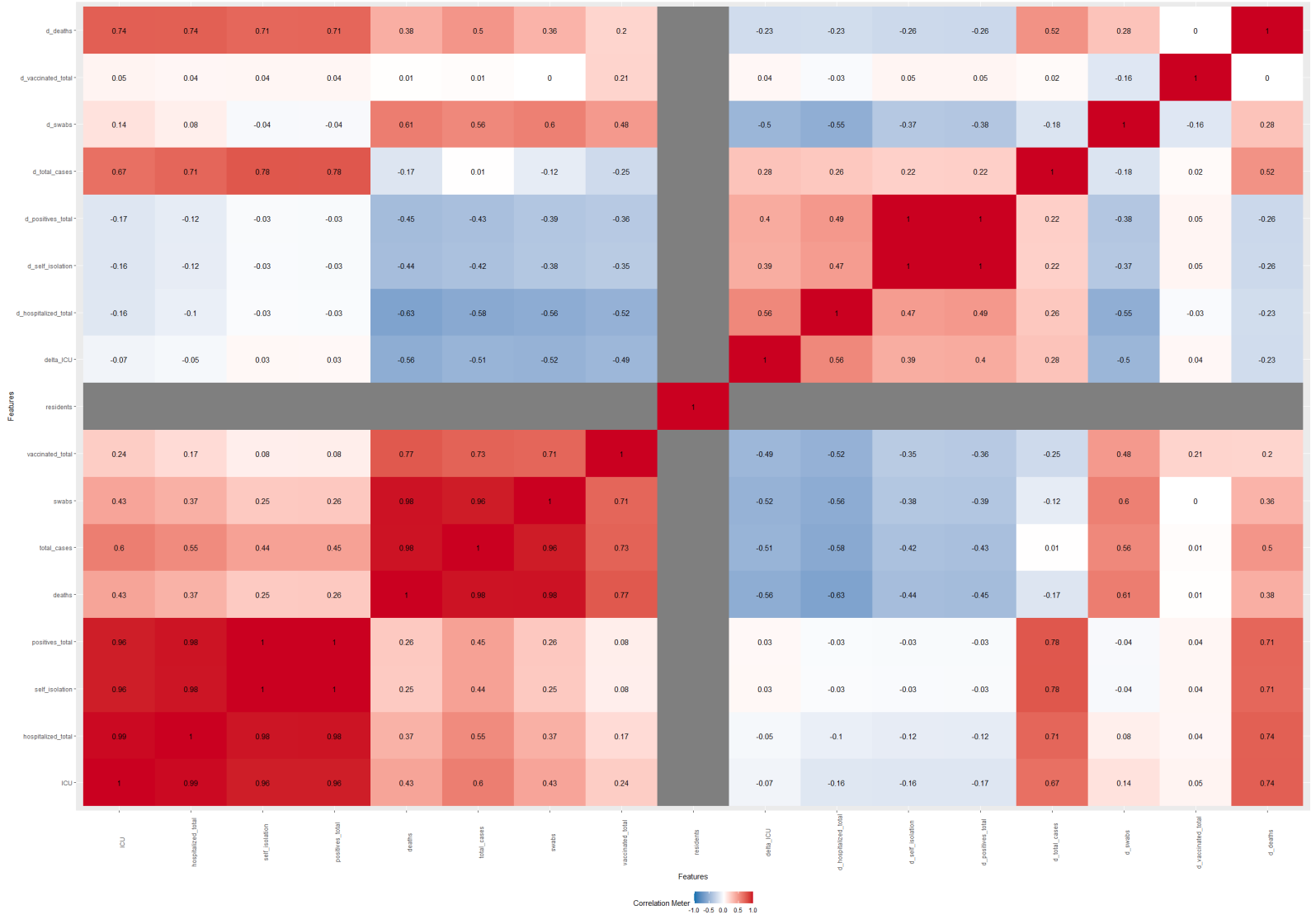
# Exploratory Data Analysis (EDA)



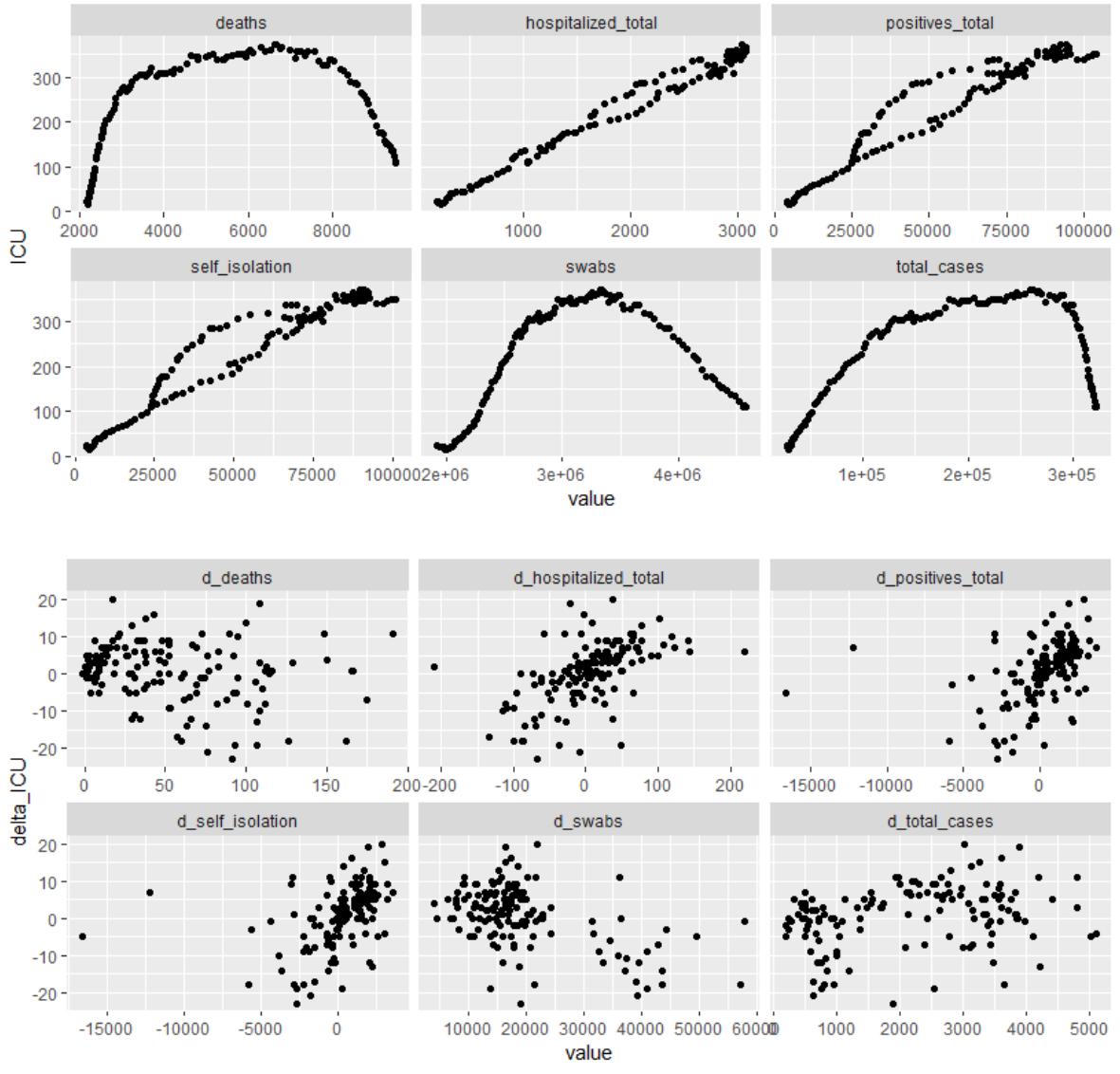
# Exploratory Data Analysis (EDA)



# Exploratory Data Analysis (EDA)



# Exploratory Data Analysis (EDA)



# Feature engineering

- We split the dataset into two chunks in order to train our different models on the first one and to test the prediction on the second to evaluate which are the pros and the cons of each model.
- Secondly, we set up also leave-one-out cross-validation (LOOCV) on the train data in order to make models' estimates more robust.
- After these passages, we tried different transformations for the three families of models and their supposed predictors such as:
    - Different lags (e.g. from 1 to 15 days) on some regressors.
    - Different degrees of orthogonal polynomials (poly function).
    - Natural logarithm and base 10-logarithm.
    - Centering and rescaling of some/all variables.
    - Normalization of some covariates.
    - Other...
- Moreover, we evaluated the inclusion of the interaction terms between the different variables.

# Linear models

We started our modeling part from the simplest family of models: the linear regression models (LMs). We performed a series of differences regressions, using mainly two approaches:

- A "direct modeling", which means we tried to formulate the most parsimonious model possible via a top-down approach including all the variables which are deemed to be the fundamental ones, by common sense.
- A "data-driven" one, where the discovery of the predictors to include is fully left to a process which encompasses first a manual backward/forward selection of single predictors based on ANOVA F-tests and the p-values of each predictor; then a second step aiming at selecting the best combination of predictors (i.e. a model) structured on another automatic backward/forward and bi-directional selection, grounded in their AIC value.

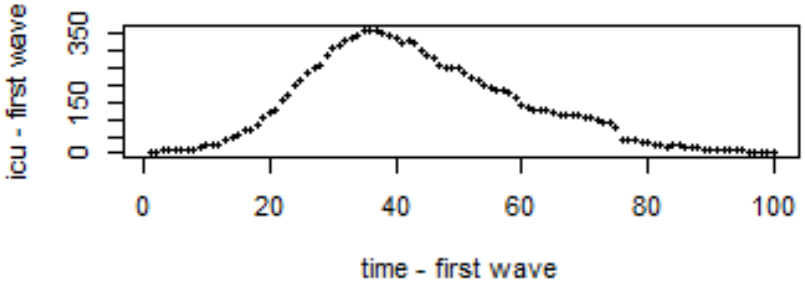
# Linear models (Direct modeling approach)

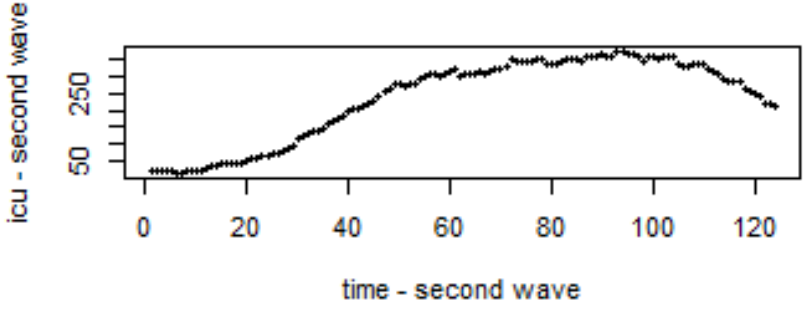
We consider a variable transformation, given by the daily difference of the ICU patients. We analyze:

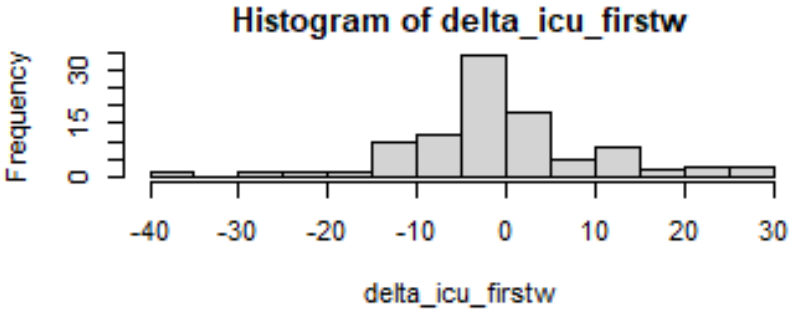
- the first wave to identify characteristics of the phenomenon that support the modeling of the second wave.
- the trend of the variables that can reasonably be considered as lagged predictors of the evolution of the ICU curve. The biological phenomenon evolves in fact, from the onset of symptoms to hospitalization and finally to critical ICU conditions (at least for most part of ICU patients). The number of new positives provides then information on the assumption of stationarity.

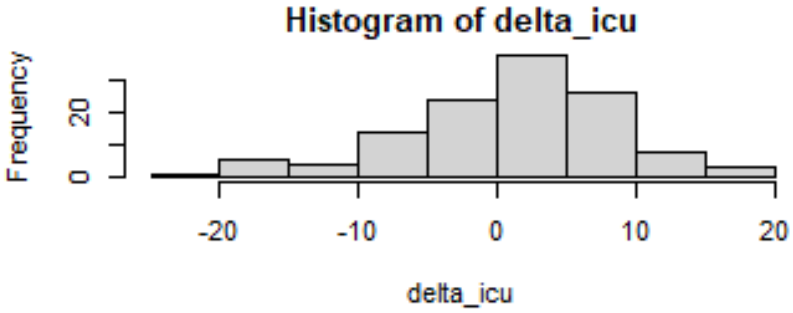
What we want to do is basically to model a stationary phenomenon for the period of the extrapolation.

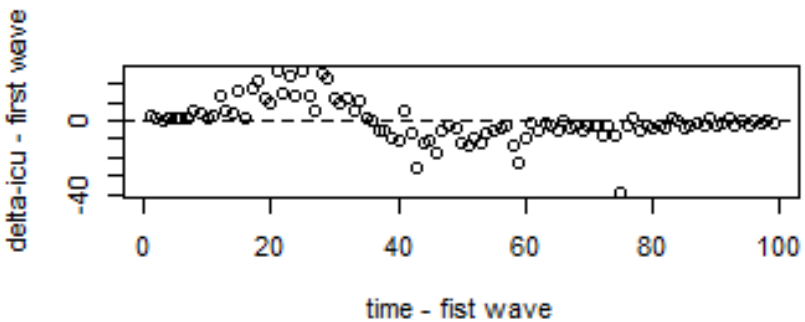
# Linear models (Direct modeling approach)

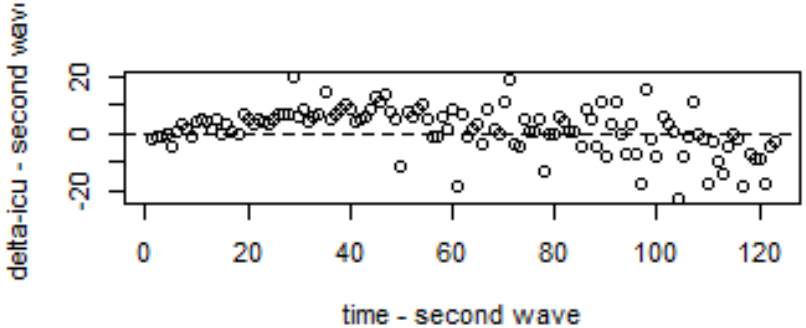


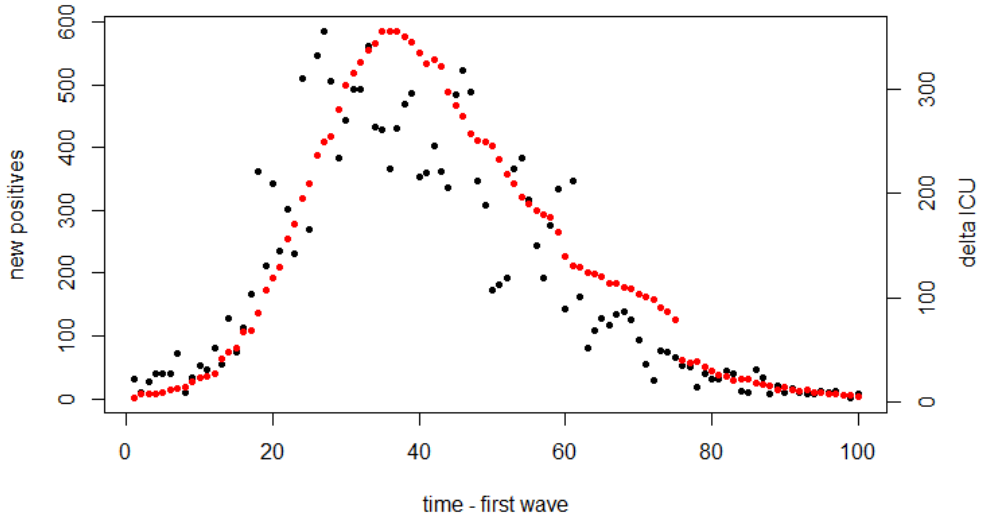



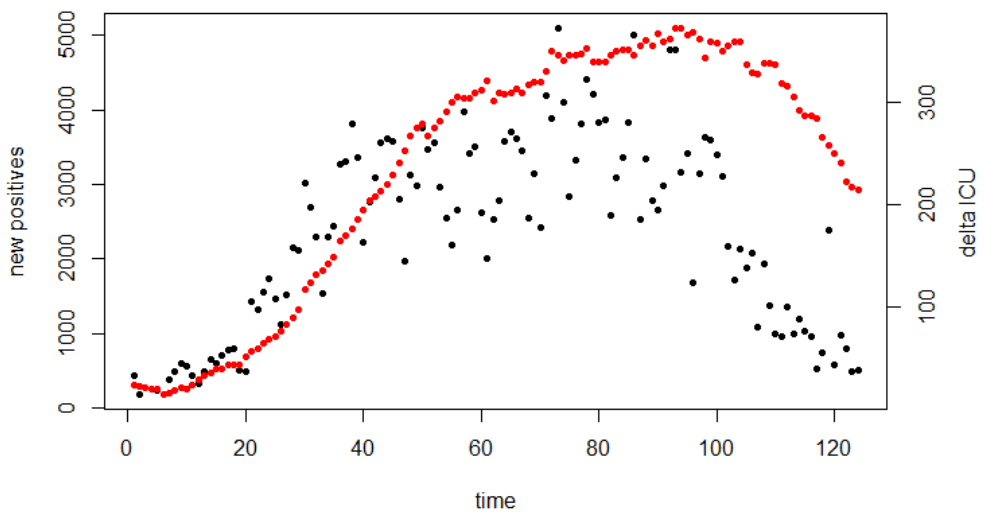



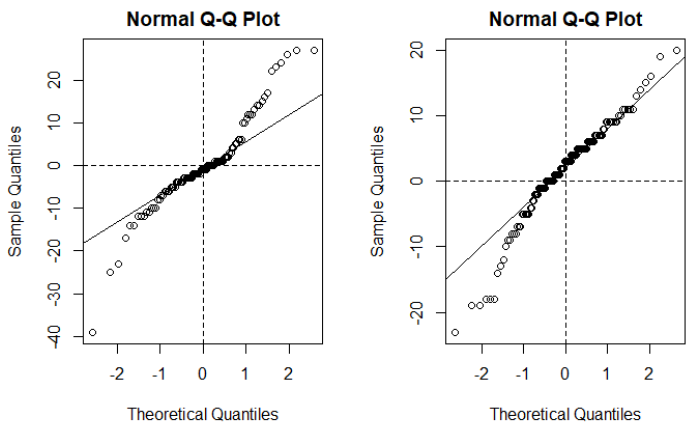
# Linear models (Direct modeling approach)

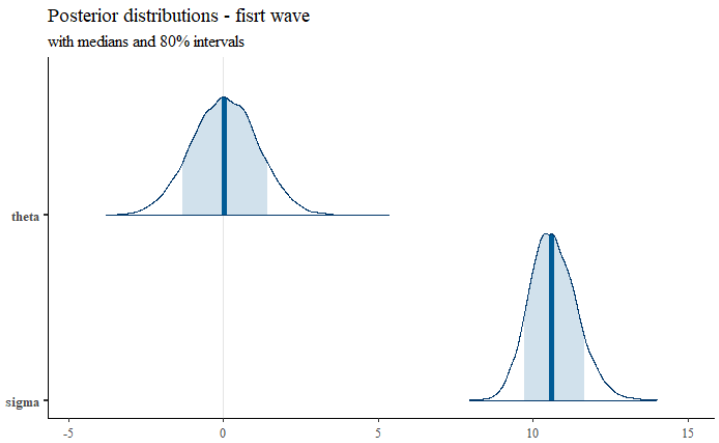


# Linear models (Direct modeling approach)

The first wave supports the idea of using the new positives to control for the stationarity of the ICU curve. The data also show similarities between the first and the second wave.
We also tried to model $\Delta$ICU with a probability distribution. The normal approximation provided reasonable results except for the tails. We estimated a posterior distribution using Rstan with a normal distribution, which suggests a mean approximately equal to zero.



# Linear models (Direct modeling approach)

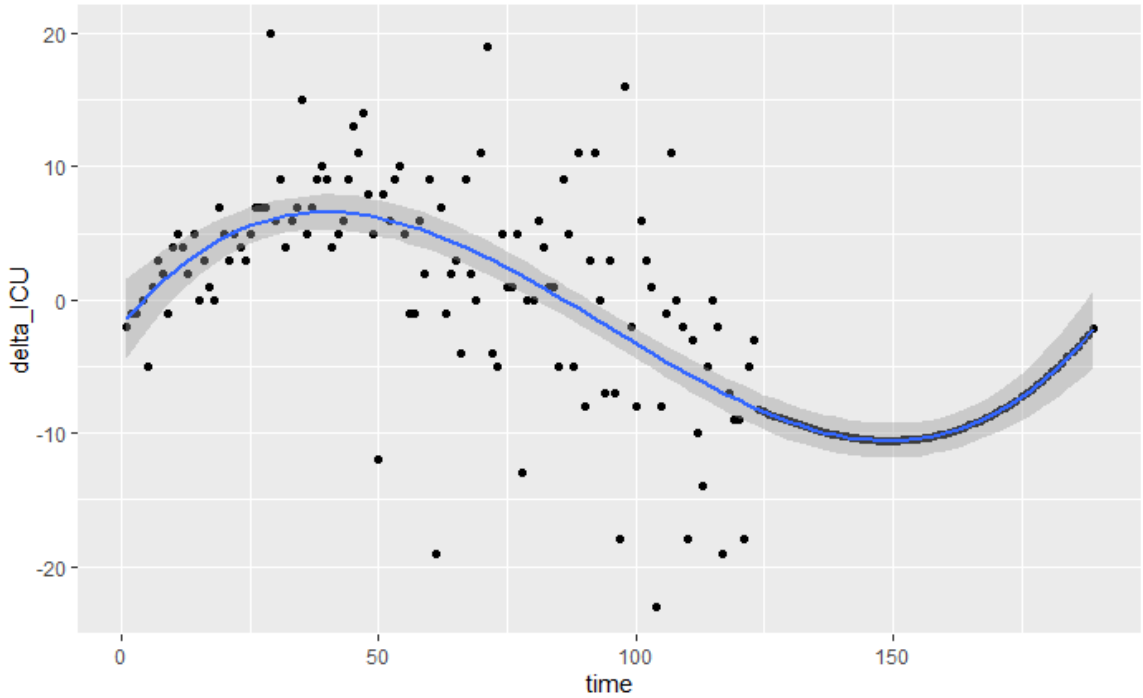
Supported by the experimental findings, we decided to model the stationary wave using a polynomial of order 3:

- the 3 roots allow to model the basic wave behaviour but also to control a potentially asymmetric shape of the descending phase of the curve.
- higher order degrees didn't provide better results.

In order to model the entire wave, we need to estimate the end of the descending phase. Data are especially noisy around the peak, so we use a moving average to estimate the peak which is set at $\Delta t = 95$ from the start of the period of interest. The end of the descending phase consequently falls at t_peak+95 days.

# Linear models (Direct modeling approach)

The graphical representation of the stationary second wave is reported below

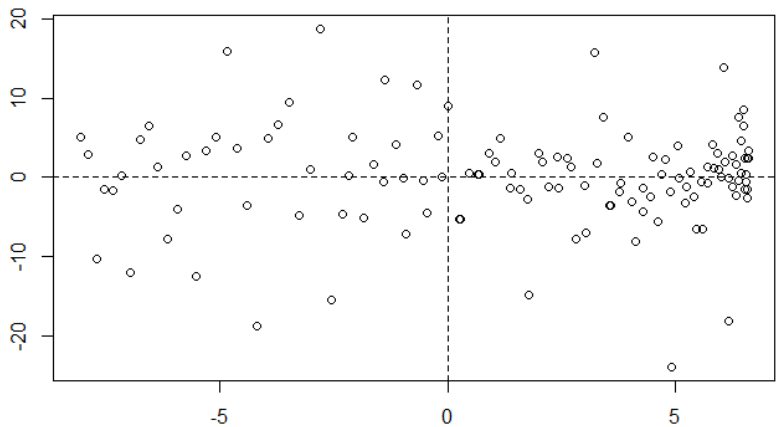



Despite the simplicity of the model, the residuals plot and the summary of the model provide indications of reasonable results. The results will be commented further at the end of the presentation, in comparison with the other models. Here we provide just a quick overview.

# Linear models (Direct modeling approach)

The experiment with the direct modeling, which used just the time variable, provided also evidence of the importance of this variable for the more complex models, as well as for first difference of the ICU data.

```
Call:
lm(formula = delta_icu_extr ~ poly(index_extr, 3, raw = TRUE),
    data = delta_icu_extr_df)

Residuals:
     Min       1Q   Median       3Q      Max
-23.9174  -2.4877   0.2343   3.0176  18.7902

Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                       -1.876e+00  2.067e+00  -0.907    0.366
poly(index_extr, 3, raw = TRUE)1  4.667e-01  1.059e-01   4.407 2.30e-05 ***
poly(index_extr, 3, raw = TRUE)2 -7.420e-03  1.441e-03  -5.149 1.04e-06 ***
poly(index_extr, 3, raw = TRUE)3  2.615e-05  5.546e-06   4.716 6.55e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.511 on 120 degrees of freedom
Multiple R-squared:  0.3264,    Adjusted R-squared:  0.3096
F-statistic: 19.39 on 3 and 120 DF,  p-value: 2.574e-10
```



# Linear models (Data-driven method)

Coming to the second approach, the first step was defining which would be the probable predictors of our linear model starting from a wide list composed by the original predictors, their transformed versions and all the possible interactions among them (many hundreds). We report here just an example of the many different takes we tried, performing an ANOVA(F-Test)-based backward/forward selection of single predictors, dropping those with lesser significant p-value.

```
Model:
delta_ICU ~ residents + d_hospitalized_total + d_self_isolation +
    d_positives_total + d_total_cases + d_swabs + d_vaccinated_total +
    lag_1_d_swabs + days_ns_1 + days_ns_2 + days_ns_3
                   Df Sum of Sq    RSS    AIC F value   Pr(>F)
<none>                          4561.0 464.41
residents           0     0.000 4561.0 464.41
d_hospitalized_total 0     0.000 4561.0 464.41
d_self_isolation    0     0.000 4561.0 464.41
d_positives_total   0     0.000 4561.0 464.41
d_total_cases       1     0.775 4561.8 462.43  0.0192 0.890069
d_swabs             1    76.557 4637.6 464.46  1.8967 0.171169
d_vaccinated_total  1     0.249 4561.3 462.42  0.0062 0.937496
lag_1_d_swabs       1   171.975 4733.0 466.97  4.2607 0.041293 *
days_ns_1           1     0.330 4561.3 462.42  0.0082 0.928128
days_ns_2           1     2.026 4563.0 462.47  0.0502 0.823120
days_ns_3           1   314.935 4876.0 470.63  7.8026 0.006129 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Single term deletions

Model:
delta_ICU ~ lag_1_d_swabs + days_ns_1 + days_ns_2 + days_ns_3
            Df Sum of Sq    RSS    AIC F value    Pr(>F)
<none>                   5033.2 466.53
lag_1_d_swabs  1    78.99 5112.2 466.45  1.8517    0.1762
days_ns_1      1    10.94 5044.2 464.80  0.2564    0.6136
days_ns_2      1     3.59 5036.8 464.62  0.0842    0.7722
days_ns_3      1  1395.42 6428.7 494.63 32.7145 8.212e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

# Linear models (Data-driven method)

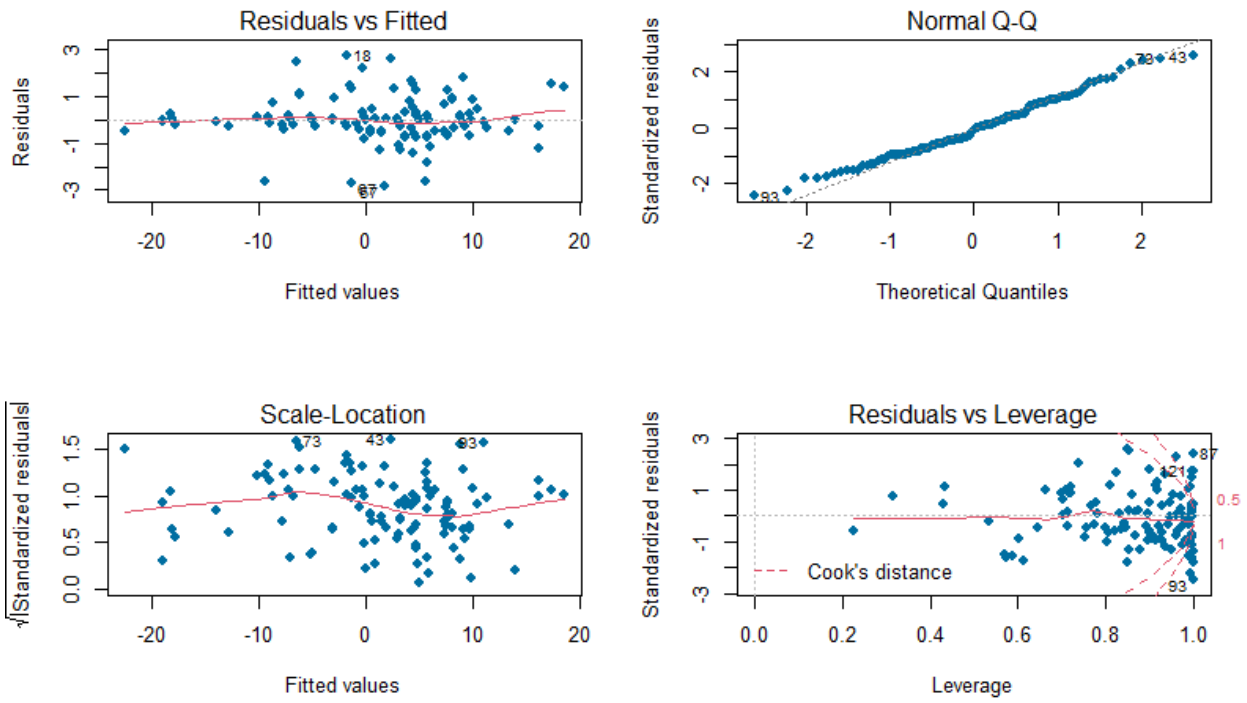
Once selected some promising predictor-candidates, a second layer has been the adoption of an automatic algorithm for model backward / forward / bidirectional selection, based on the Akaike information criterion (AIC): the routine stopped when finding the sequence of predictors with the lowest AIC among all possible combinations, here below again a glimpse into our attempts:

```
Start:  AIC=469.66
delta_ICU ~ residents + d_total_cases + d_vaccinated_total +
    lag_1_d_swabs + lag_1_d_self_isolation + lag_1_d_positives_total +
    lag_2_d_swabs + lag_2_d_self_isolation + lag_2_d_positives_total +
    lag_3_d_swabs + lag_3_d_self_isolation + lag_3_d_positives_total +
    lag_4_d_swabs + lag_4_d_self_isolation + lag_4_d_positives_total +
    lag_5_d_swabs + lag_5_d_self_isolation + lag_5_d_positives_total +
    days_ns_1 + days_ns_2 + days_ns_3


Step:  AIC=469.66
delta_ICU ~ d_total_cases + d_vaccinated_total + lag_1_d_swabs +
    lag_1_d_self_isolation + lag_1_d_positives_total + lag_2_d_swabs +
    lag_2_d_self_isolation + lag_2_d_positives_total + lag_3_d_swabs +
    lag_3_d_self_isolation + lag_3_d_positives_total + lag_4_d_swabs +
    lag_4_d_self_isolation + lag_4_d_positives_total + lag_5_d_swabs +
    lag_5_d_self_isolation + lag_5_d_positives_total + days_ns_1 +
    days_ns_2 + days_ns_3

                          Df Sum of Sq    RSS    AIC
- lag_4_d_positives_total  1     0.059 4328.3 467.66
- lag_4_d_self_isolation   1     0.102 4328.3 467.66
- days_ns_1                1     0.385 4328.6 467.67
- lag_5_d_swabs            1     0.686 4328.9 467.68
- d_vaccinated_total       1     0.710 4328.9 467.68
- days_ns_2                1     3.346 4331.5 467.75
- d_total_cases            1     3.926 4332.1 467.77
- lag_2_d_swabs            1    12.302 4340.5 468.00
- lag_1_d_self_isolation   1    30.490 4358.7 468.50
- lag_1_d_positives_total  1    30.942 4359.1 468.51
- lag_2_d_positives_total  1    32.977 4361.2 468.56
- lag_4_d_swabs            1    33.258 4361.4 468.57
- lag_2_d_self_isolation   1    33.394 4361.6 468.57
- lag_3_d_swabs            1    53.182 4381.4 469.11
- lag_1_d_swabs            1    53.201 4381.4 469.11
<none>                               4328.2 469.66
- lag_5_d_self_isolation   1   124.079 4452.3 471.02
- lag_5_d_positives_total  1   125.451 4453.6 471.06
- days_ns_3                1   203.178 4531.4 473.12
- lag_3_d_positives_total  1   230.348 4558.5 473.83
- lag_3_d_self_isolation   1   233.296 4561.5 473.91
```

```
Step:  AIC=448.93
delta_ICU ~ lag_1_d_swabs + lag_3_d_self_isolation + lag_3_d_positives_total +
    lag_5_d_self_isolation + lag_5_d_positives_total + days_ns_3

                          Df Sum of Sq    RSS    AIC
<none>                                4601.0 448.93
- lag_1_d_swabs            1     99.40 4700.4 449.48
- lag_5_d_self_isolation   1    147.18 4748.1 450.68
- lag_5_d_positives_total  1    149.24 4750.2 450.73
- lag_3_d_positives_total  1    267.95 4868.9 453.67
- lag_3_d_self_isolation   1    271.44 4872.4 453.75
- days_ns_3                1    875.48 5476.5 467.66

Call:
stats::lm(formula = delta_ICU ~ lag_1_d_swabs + lag_3_d_self_isolation +
    lag_3_d_positives_total + lag_5_d_self_isolation + lag_5_d_positives_total +
    days_ns_3, data = veneto_train_preprocessed_lm)

Residuals:
     Min       1Q   Median       3Q      Max
-20.3164  -2.7793   0.6095   2.9447  18.7701

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)              -1.575e+00  1.628e+00  -0.968   0.3353
lag_1_d_swabs             1.214e-04  7.807e-05   1.556   0.1226
lag_3_d_self_isolation    3.501e-02  1.362e-02   2.571   0.0115 *
lag_3_d_positives_total -3.467e-02  1.358e-02  -2.554   0.0120 *
lag_5_d_self_isolation  -2.593e-02  1.370e-02  -1.893   0.0610 .
lag_5_d_positives_total  2.602e-02  1.365e-02   1.906   0.0592 .
days_ns_3                -1.620e+01  3.508e+00  -4.616 1.05e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.409 on 112 degrees of freedom
  (5 osservazioni eliminate a causa di valori mancanti)
Multiple R-squared:  0.3875,    Adjusted R-squared:  0.3547
F-statistic: 11.81 on 6 and 112 DF,  p-value: 3.14e-10
```

# Linear models (Data-driven method)

Eventually, we settled for this model given the overall very good level of fit testified by the low AIC, the high adjusted R-squared and the significance of the F-test. Anyway are already apparent: a possible problem with the interpretation of predictors (in particular the interactions), the very high number of predictors which poses many challenges in the prediction phase, even when using a simple polynomial fit, and the high risk of overfitting.



# Linear models (Data-driven method)

```
Call:
stats::lm(formula = delta_ICU ~ lag_5_d_swabs + lag_6_d_swabs +
    lag_8_d_swabs + lag_9_d_swabs + lag_10_d_swabs + lag_6_d_self_isolation +
    lag_7_d_self_isolation + lag_2_d_hospitalized_total + lag_1_d_positives_total +
    lag_1_d_total_cases + lag_1_d_vaccinated_total + days_poly_1 +
    days_poly_3 + lag_5_d_swabs_x_lag_6_d_swabs + lag_5_d_swabs_x_lag_7_d_swabs +
    lag_5_d_swabs_x_lag_8_d_swabs + lag_5_d_swabs_x_lag_9_d_swabs +
    lag_5_d_swabs_x_lag_10_d_swabs + lag_5_d_swabs_x_lag_6_d_self_isolation +
    lag_5_d_swabs_x_lag_7_d_self_isolation + lag_5_d_swabs_x_lag_2_d_hospitalized_total +
    lag_5_d_swabs_x_lag_1_d_total_cases + lag_5_d_swabs_x_days_poly_1 +
    lag_5_d_swabs_x_days_poly_3 + lag_6_d_swabs_x_lag_7_d_swabs +
    lag_6_d_swabs_x_lag_8_d_swabs + lag_6_d_swabs_x_lag_9_d_swabs +
    lag_6_d_swabs_x_lag_10_d_swabs + lag_6_d_swabs_x_lag_6_d_self_isolation +
    lag_6_d_swabs_x_lag_7_d_self_isolation + lag_6_d_swabs_x_lag_2_d_hospitalized_total +
    lag_6_d_swabs_x_lag_1_d_positives_total + lag_6_d_swabs_x_lag_1_d_total_cases +
    lag_6_d_swabs_x_lag_1_d_vaccinated_total + lag_6_d_swabs_x_days_poly_1 +
    lag_6_d_swabs_x_days_poly_3 + lag_7_d_swabs_x_lag_8_d_swabs +
    lag_7_d_swabs_x_lag_9_d_swabs + lag_7_d_swabs_x_lag_10_d_swabs +
    lag_7_d_swabs_x_lag_7_d_self_isolation + lag_7_d_swabs_x_lag_2_d_hospitalized_total +
    lag_7_d_swabs_x_lag_1_d_positives_total + lag_7_d_swabs_x_lag_1_d_vaccinated_total +
    lag_7_d_swabs_x_days_poly_1 + lag_8_d_swabs_x_lag_9_d_swabs +
    lag_8_d_swabs_x_lag_10_d_swabs + lag_8_d_swabs_x_lag_6_d_self_isolation +
    lag_8_d_swabs_x_lag_7_d_self_isolation + lag_8_d_swabs_x_lag_2_d_hospitalized_total +
    lag_8_d_swabs_x_lag_1_d_positives_total + lag_8_d_swabs_x_lag_1_d_total_cases +
    lag_8_d_swabs_x_lag_1_d_vaccinated_total + lag_8_d_swabs_x_days_poly_1 +
    lag_8_d_swabs_x_days_poly_3 + lag_9_d_swabs_x_lag_10_d_swabs +
    lag_9_d_swabs_x_lag_6_d_self_isolation + lag_9_d_swabs_x_lag_2_d_hospitalized_total +
    lag_9_d_swabs_x_lag_1_d_positives_total + lag_9_d_swabs_x_lag_1_d_total_cases +
    lag_9_d_swabs_x_lag_1_d_vaccinated_total + lag_9_d_swabs_x_days_poly_1 +
    lag_9_d_swabs_x_days_poly_3 + lag_10_d_swabs_x_lag_6_d_self_isolation +
    lag_10_d_swabs_x_lag_7_d_self_isolation + lag_10_d_swabs_x_lag_2_d_hospitalized_total +
    lag_10_d_swabs_x_lag_1_d_positives_total + lag_10_d_swabs_x_lag_1_d_total_cases +
    lag_10_d_swabs_x_lag_1_d_vaccinated_total + lag_10_d_swabs_x_days_poly_1 +
    lag_10_d_swabs_x_days_poly_3 + lag_6_d_self_isolation_x_lag_7_d_self_isolation +
    lag_6_d_self_isolation_x_lag_2_d_hospitalized_total + lag_6_d_self_isolation_x_lag_1_d_positives_total +
    lag_6_d_self_isolation_x_lag_1_d_total_cases + lag_6_d_self_isolation_x_lag_1_d_vaccinated_total +
    lag_6_d_self_isolation_x_days_poly_1 + lag_6_d_self_isolation_x_days_poly_3 +
    lag_7_d_self_isolation_x_lag_2_d_hospitalized_total + lag_7_d_self_isolation_x_lag_1_d_positives_total +
    lag_7_d_self_isolation_x_lag_1_d_total_cases + lag_7_d_self_isolation_x_lag_1_d_vaccinated_total +
    lag_7_d_self_isolation_x_days_poly_1 + lag_7_d_self_isolation_x_days_poly_3 +
    lag_2_d_hospitalized_total_x_lag_1_d_positives_total + lag_2_d_hospitalized_total_x_lag_1_d_total_cases +
    lag_2_d_hospitalized_total_x_lag_1_d_vaccinated_total + lag_2_d_hospitalized_total_x_days_poly_1 +
    lag_2_d_hospitalized_total_x_days_poly_3 + lag_1_d_positives_total_x_lag_1_d_total_cases +
    lag_1_d_positives_total_x_lag_1_d_vaccinated_total + lag_1_d_positives_total_x_days_poly_1 +
    lag_1_d_positives_total_x_days_poly_3 + lag_1_d_total_cases_x_days_poly_1 +
    lag_1_d_total_cases_x_days_poly_3 + lag_1_d_vaccinated_total_x_days_poly_1 +
    lag_1_d_vaccinated_total_x_days_poly_3 + days_poly_1_x_days_poly_3,
    data = veneto_train_preprocessed_lm)

Residuals:
     Min       1Q   Median       3Q      Max
-2.83003 -0.49048 -0.01672  0.31285  2.74314
```

```
Coefficients:
                                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                                        2.839      0.996   2.851 0.011567 *
lag_5_d_swabs                                     -7.230      2.663  -2.715 0.015292 *
lag_6_d_swabs                                    -22.868      3.461  -6.607 6.03e-06 ***
lag_8_d_swabs                                    -15.646      3.123  -5.010 0.000128 ***
lag_9_d_swabs                                      9.647      3.103   3.109 0.006749 **
lag_10_d_swabs                                    -2.039      2.864  -0.712 0.486799
lag_6_d_self_isolation                             9.486      6.905   1.374 0.188458
lag_7_d_self_isolation                           -41.527      7.134  -5.821 2.60e-05 ***
lag_2_d_hospitalized_total                        23.847      4.260   5.598 4.01e-05 ***
lag_1_d_positives_total                          -51.658      7.720  -6.691 5.18e-06 ***
lag_1_d_total_cases                               42.286      5.154   8.204 3.99e-07 ***
lag_1_d_vaccinated_total                        -298.235     42.456  -7.025 2.87e-06 ***
days_poly_1                                      -10.883      7.284  -1.494 0.154604
days_poly_3                                      -26.972      3.752  -7.189 2.16e-06 ***
lag_5_d_swabs_x_lag_6_d_swabs                    -30.220      5.040  -5.996 1.87e-05 ***
lag_5_d_swabs_x_lag_7_d_swabs                    -12.180      4.542  -2.681 0.016384 *
lag_5_d_swabs_x_lag_8_d_swabs                    -27.183      6.273  -4.333 0.000514 ***
lag_5_d_swabs_x_lag_9_d_swabs                     41.456      6.762   6.131 1.45e-05 ***
lag_5_d_swabs_x_lag_10_d_swabs                    53.032     10.128   5.236 8.15e-05 ***
lag_5_d_swabs_x_lag_6_d_self_isolation            18.773      9.466   1.983 0.064778 .
lag_5_d_swabs_x_lag_7_d_self_isolation            17.770     11.524   1.542 0.142618
lag_5_d_swabs_x_lag_2_d_hospitalized_total        12.504      3.393   3.686 0.002003 **
lag_5_d_swabs_x_lag_1_d_total_cases              -20.688      4.386  -4.717 0.000233 ***
lag_5_d_swabs_x_days_poly_1                       41.702      6.826   6.109 1.51e-05 ***
lag_5_d_swabs_x_days_poly_3                       -3.146      3.254  -0.967 0.347971
lag_6_d_swabs_x_lag_7_d_swabs                    -71.923      8.197  -8.775 1.64e-07 ***
lag_6_d_swabs_x_lag_8_d_swabs                     12.156      5.470   2.222 0.041017 *
lag_6_d_swabs_x_lag_9_d_swabs                    -21.021      5.956  -3.530 0.002784 **
lag_6_d_swabs_x_lag_10_d_swabs                    14.687      7.311   2.009 0.061743 .
lag_6_d_swabs_x_lag_6_d_self_isolation            20.072      9.789   2.051 0.057065 .
lag_6_d_swabs_x_lag_7_d_self_isolation            -7.916      9.484  -0.835 0.416201
lag_6_d_swabs_x_lag_2_d_hospitalized_total       -15.744      3.928  -4.008 0.001014 **
lag_6_d_swabs_x_lag_1_d_positives_total          -13.520      7.825  -1.728 0.103265
lag_6_d_swabs_x_lag_1_d_total_cases                7.966      5.358   1.487 0.156537
lag_6_d_swabs_x_lag_1_d_vaccinated_total           5.771      7.755   0.744 0.467579
lag_6_d_swabs_x_days_poly_1                       11.735      8.967   1.309 0.209141
lag_6_d_swabs_x_days_poly_3                       20.884      4.676   4.466 0.000390 ***
lag_7_d_swabs_x_lag_8_d_swabs                    -80.088     11.682  -6.856 3.87e-06 ***
lag_7_d_swabs_x_lag_9_d_swabs                     66.361      7.560   8.778 1.63e-07 ***
lag_7_d_swabs_x_lag_10_d_swabs                    10.992      4.372   2.514 0.023014 *
lag_7_d_swabs_x_lag_7_d_self_isolation           -38.191      9.045  -4.222 0.000644 ***
lag_7_d_swabs_x_lag_2_d_hospitalized_total        40.120      6.959   5.765 2.90e-05 ***
lag_7_d_swabs_x_lag_1_d_positives_total          -32.349      5.325  -6.075 1.61e-05 ***
lag_7_d_swabs_x_lag_1_d_vaccinated_total         -26.508      6.259  -4.235 0.000630 ***
lag_7_d_swabs_x_days_poly_1                       37.724      5.798   6.506 7.24e-06 ***
lag_8_d_swabs_x_lag_9_d_swabs                    -18.890      9.802  -1.927 0.071920 .
lag_8_d_swabs_x_lag_10_d_swabs                    15.072      6.404   2.354 0.031712 *
lag_8_d_swabs_x_lag_6_d_self_isolation            43.347      5.465   7.931 6.20e-07 ***
lag_8_d_swabs_x_lag_7_d_self_isolation           -25.852      9.297  -2.781 0.013363 *
lag_8_d_swabs_x_lag_2_d_hospitalized_total        22.403      6.620   3.384 0.003783 **
lag_8_d_swabs_x_lag_1_d_positives_total           38.618      5.258   7.344 1.65e-06 ***
lag_8_d_swabs_x_lag_1_d_total_cases              -54.073      5.875  -9.204 8.61e-08 ***
lag_8_d_swabs_x_lag_1_d_vaccinated_total         -16.958      7.288  -2.327 0.033427 *
lag_8_d_swabs_x_days_poly_1                       97.418     11.642   8.368 3.08e-07 ***
lag_8_d_swabs_x_days_poly_3                      -19.212      4.895  -3.925 0.001210 **
lag_9_d_swabs_x_lag_10_d_swabs                    14.507      7.872   1.843 0.083968 .
lag_9_d_swabs_x_lag_6_d_self_isolation            -4.426      6.282  -0.704 0.491286
lag_9_d_swabs_x_lag_2_d_hospitalized_total        15.618      6.183   2.526 0.022472 *
lag_9_d_swabs_x_lag_1_d_positives_total          -44.054      6.431  -6.850 3.90e-06 ***
lag_9_d_swabs_x_lag_1_d_total_cases              39.113      5.329   7.340 1.66e-06 ***
lag_9_d_swabs_x_lag_1_d_vaccinated_total         40.879      6.896   5.928 2.12e-05 ***
lag_9_d_swabs_x_days_poly_1                     -55.600      8.550  -6.503 7.29e-06 ***
lag_9_d_swabs_x_days_poly_3                      -9.134      4.353  -2.098 0.052103 .
lag_10_d_swabs_x_lag_6_d_self_isolation          56.770      7.864   7.219 2.05e-06 ***
lag_10_d_swabs_x_lag_7_d_self_isolation         -30.442      6.178  -4.927 0.000152 ***
lag_10_d_swabs_x_lag_2_d_hospitalized_total      25.977      4.538   5.725 3.13e-05 ***
lag_10_d_swabs_x_lag_1_d_positives_total          5.420      3.178   1.705 0.107448
lag_10_d_swabs_x_lag_1_d_total_cases            -23.106      3.565  -6.482 7.57e-06 ***
lag_10_d_swabs_x_lag_1_d_vaccinated_total       -38.321     10.411  -3.681 0.002022 **
lag_10_d_swabs_x_days_poly_1                     16.302     10.017   1.628 0.123157
lag_10_d_swabs_x_days_poly_3                     -8.363      4.429  -1.888 0.077257 .
lag_6_d_self_isolation_x_lag_7_d_self_isolation  52.698     10.521   5.009 0.000129 ***
lag_6_d_self_isolation_x_lag_2_d_hospitalized_total  -8.776  2.965  -2.960 0.009215 **
lag_6_d_self_isolation_x_lag_1_d_positives_total 46.249     9.091   5.087 0.000110 ***
lag_6_d_self_isolation_x_lag_1_d_total_cases     -54.688     6.638  -8.238 3.78e-07 ***
lag_6_d_self_isolation_x_lag_1_d_vaccinated_total  52.447   8.190   6.404 8.73e-06 ***
lag_6_d_self_isolation_x_days_poly_1             -40.433     14.306  -2.826 0.012162 *
lag_6_d_self_isolation_x_days_poly_3             -53.431     10.339  -5.168 9.34e-05 ***
lag_7_d_self_isolation_x_lag_2_d_hospitalized_total  20.330  5.026   4.045 0.000938 ***
lag_7_d_self_isolation_x_lag_1_d_positives_total -42.997     4.990  -8.616 2.09e-07 ***
lag_7_d_self_isolation_x_lag_1_d_total_cases      33.441     8.471   3.948 0.001152 **
lag_7_d_self_isolation_x_lag_1_d_vaccinated_total  22.735   7.054   3.223 0.005315 **
lag_7_d_self_isolation_x_days_poly_1             -44.475     14.367  -3.096 0.006942 **
lag_7_d_self_isolation_x_days_poly_3              -8.856     10.563  -0.838 0.414142
lag_2_d_hospitalized_total_x_lag_1_d_positives_total  17.662  3.865   4.569 0.000315 ***
lag_2_d_hospitalized_total_x_lag_1_d_total_cases  -27.317     3.917  -6.973 3.14e-06 ***
lag_2_d_hospitalized_total_x_lag_1_d_vaccinated_total  24.561  4.791   5.126 0.000102 ***
lag_2_d_hospitalized_total_x_days_poly_1          -4.808      6.574  -0.731 0.475114
lag_2_d_hospitalized_total_x_days_poly_3           3.123      3.627   0.861 0.401902
lag_1_d_positives_total_x_lag_1_d_total_cases     10.434      4.705   2.218 0.041401 *
lag_1_d_positives_total_x_lag_1_d_vaccinated_total  22.741   7.690   2.957 0.009274 **
lag_1_d_positives_total_x_days_poly_1             78.622     12.800   6.142 1.42e-05 ***
lag_1_d_positives_total_x_days_poly_3             36.306      7.727   4.699 0.000241 ***
lag_1_d_total_cases_x_days_poly_1                 -5.897      4.730  -1.247 0.230424
lag_1_d_total_cases_x_days_poly_3                 21.081      4.928   4.278 0.000577 ***
lag_1_d_vaccinated_total_x_days_poly_1           354.841     52.740   6.728 4.85e-06 ***
lag_1_d_vaccinated_total_x_days_poly_3           -52.875     12.674  -4.172 0.000719 ***
days_poly_1_x_days_poly_3                        -22.877      6.095  -3.754 0.001735 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.621 on 16 degrees of freedom
  (10 osservazioni eliminate a causa di valori mancanti)
Multiple R-squared:  0.9853,    Adjusted R-squared:  0.8959
F-statistic: 11.03 on 97 and 16 DF,  p-value: 1.568e-06

> generics::glance(fit1)
# A tibble: 1 x 12
  r.squared adj.r.squared sigma statistic    p.value    df logLik   AIC   BIC deviance df.residual  nobs
      <dbl>         <dbl> <dbl>     <dbl>      <dbl> <dbl> <dbl> <dbl> <dbl>    <dbl>       <int> <int>
1     0.985         0.896  2.62      11.0 0.00000157    97  -[illegible]  517.  788.      110.          16   114
```

# GLM Poisson

- The second family of models we decided to adopt are the Poisson Generalized Linear Models (GLMs) with canonical (log) link: now our response variable (ICUs) is taken in levels as this kind of model is often used to represent count data.
- This time, the choice of predictors was based on an ANOVA step-wise selection based on the drop in deviance by means of a Chi-square test. Chi-square distribution was used here, since regular Poisson regression does not require the estimation of a dispersion parameter.
- The parameters selected were the natural cubic spline of time, the natural log of cumulative number of lagged (1 day) swabs and their interactions plus the log of lagged (1 day) self-isolation, so to keep a quite parsimonious approach, in line with the results of [1].

# GLM Poisson

We settled for this model, which is a compromise between the use of few reasonable predictors and a quite good fit even if with a slightly higher AIC compared with other models tried, to try avoiding the risk of overfitting.

```
Call:
stats::glm(formula = ..y ~ ., family = ~stats::poisson(link = "log"),
    data = data)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.24174  -0.30504  -0.00793   0.33501   1.15213

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)           -2.59537    2.81438  -0.922  0.35643
lag_1_self_isolation   0.13344    0.09007   1.481  0.13848
lag_1_swabs           -3.23493    1.58453  -2.042  0.04119 *
days_ns_1              5.39451    2.28079   2.365  0.01802 *
days_ns_2             15.38331    5.21555   2.950  0.00318 **
days_ns_3              2.65274    1.34376   1.974  0.04837 *
lag_1_swabs_x_days_ns_1  3.59382    1.33073   2.701  0.00692 **
lag_1_swabs_x_days_ns_2  5.48245    2.25363   2.433  0.01499 *
lag_1_swabs_x_days_ns_3  2.15425    1.21019   1.780  0.07506 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 9846.924  on 122  degrees of freedom
Residual deviance:   25.525  on 114  degrees of freedom
  (1 osservazione eliminata a causa di un valore mancante)
AIC: 907.76

Number of Fisher Scoring iterations: 4

> generics::glance(glm_pois_train_fit)
# A tibble: 1 x 8
  null.deviance df.null logLik   AIC   BIC deviance df.residual  nobs
          <dbl>   <int>  <dbl> <dbl> <dbl>    <dbl>       <int> <int>
1         9847.     122  -448.  908.  933.     25.5         114   123
```

# GLM Poisson with offset

Trying to generalize the results of this model along possibly diverse regions and time frames (i.e. at different times, population count is different), we introduced as an offset term the natural log of the most recent number of Veneto's residents available [Istat], with respect to our period of interest. Consequently, there is a change in the intercept coefficient here reported:

```
Call:
stats::glm(formula = ICU ~ . - residents, family = stats::poisson(link = "log"),
    data = veneto_train_preprocessed_glm_pois, offset = log(residents))

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.24174  -0.30504  -0.00793   0.33501   1.15213

Coefficients:
                       Estimate Std. Error z value Pr(>|z|)
(Intercept)           -17.99585    2.81438  -6.394 1.61e-10 ***
lag_1_self_isolation    0.13344    0.09007   1.481  0.13848
lag_1_swabs            -3.23493    1.58453  -2.042  0.04119 *
days_ns_1               5.39451    2.28079   2.365  0.01802 *
days_ns_2              15.38331    5.21555   2.950  0.00318 **
days_ns_3               2.65274    1.34376   1.974  0.04837 *
lag_1_swabs_x_days_ns_1 3.59382    1.33073   2.701  0.00692 **
lag_1_swabs_x_days_ns_2 5.48245    2.25363   2.433  0.01499 *
lag_1_swabs_x_days_ns_3 2.15425    1.21019   1.780  0.07506 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 9846.924  on 122  degrees of freedom
Residual deviance:   25.525  on 114  degrees of freedom
  (1 osservazione eliminata a causa di un valore mancante)
AIC: 907.76

Number of Fisher Scoring iterations: 4

> generics::glance(fit2.1)
# A tibble: 1 x 8
  null.deviance df.null logLik   AIC   BIC deviance df.residual  nobs
          <dbl>   <int>  <dbl> <dbl> <dbl>    <dbl>       <int> <int>
1         9847.     122  -447.  908.  933.     25.5         114   123
```

# GLM Quasi-Poisson with offset

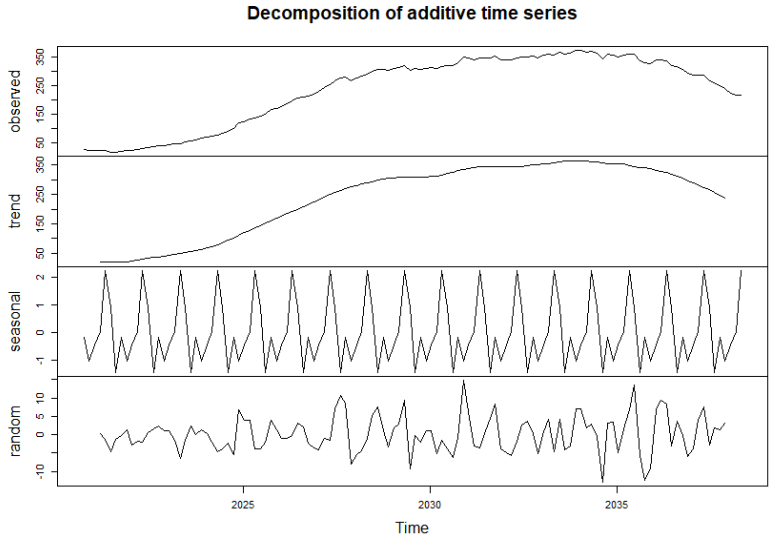
As a final improvement to our GLM, we opted for a Quasi-Poisson version of the model, keeping the same predictors and the offset as before. This is a remedy to the underdispersion we noticed in the previous models' summaries. The overall outcome is a better fit along all predictors.

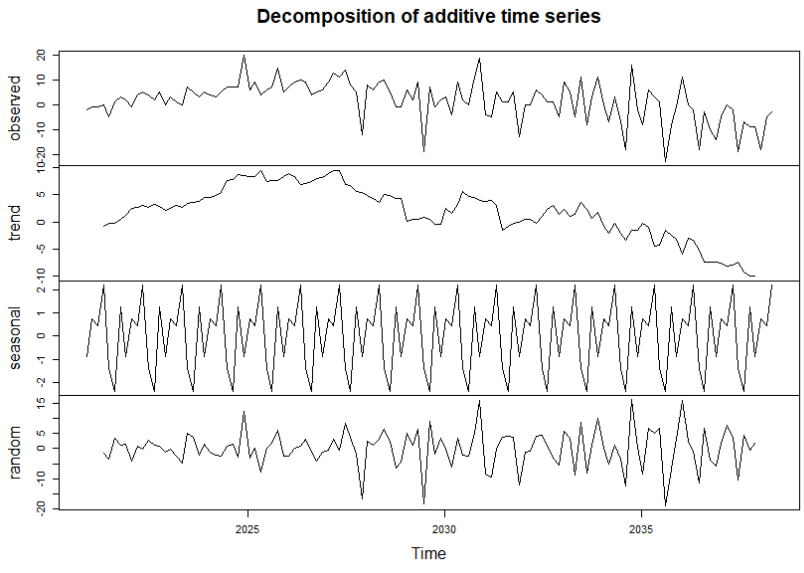
```
Call:
stats::glm(formula = ICU ~ . - residents, family = stats::quasipoisson(link = "log"),
    data = veneto_train_preprocessed_glm_pois, offset = log(residents))

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.24174  -0.30504  -0.00793   0.33501   1.15213

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)            -17.99585    1.32775 -13.554  < 2e-16 ***
lag_1_self_isolation     0.13344    0.04249   3.140 0.002150 **
lag_1_swabs             -3.23493    0.74754  -4.327 3.25e-05 ***
days_ns_1                5.39451    1.07601   5.013 1.98e-06 ***
days_ns_2               15.38331    2.46055   6.252 7.26e-09 ***
days_ns_3                2.65274    0.63395   4.184 5.64e-05 ***
lag_1_swabs_x_days_ns_1  3.59382    0.62780   5.724 8.54e-08 ***
lag_1_swabs_x_days_ns_2  5.48245    1.06320   5.157 1.07e-06 ***
lag_1_swabs_x_days_ns_3  2.15425    0.57094   3.773 0.000257 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 0.2225689)

    Null deviance: 9846.924  on 122  degrees of freedom
Residual deviance:   25.525  on 114  degrees of freedom
  (1 osservazione eliminata a causa di un valore mancante)
AIC: NA

Number of Fisher Scoring iterations: 4

> generics::glance(fit3.1)
# A tibble: 1 x 8
  null.deviance df.null logLik   AIC   BIC deviance df.residual  nobs
          <dbl>   <int>  <dbl> <dbl> <dbl>    <dbl>       <int> <int>
1         9847.     122     NA    NA    NA     25.5         114   123
```

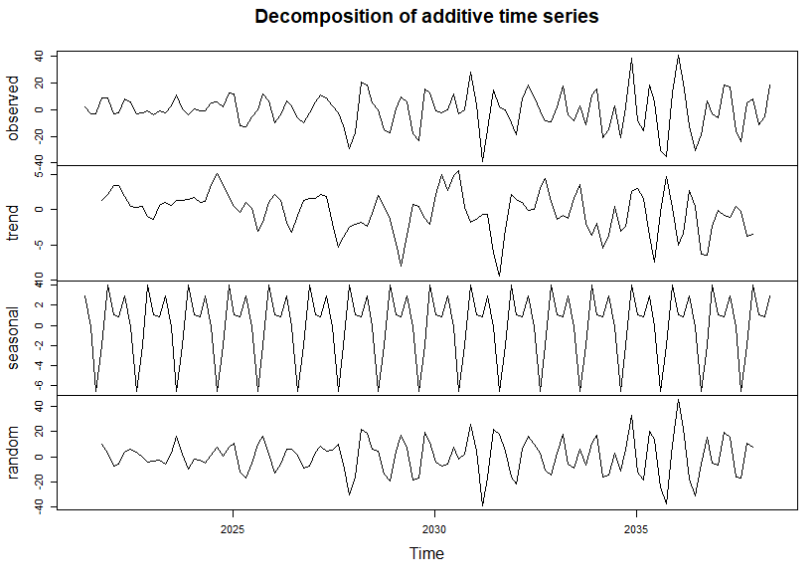
# ARIMA

- The last family of models we used is the ARIMA one, a special type of regression model in which the dependent variable has been stationarized and the independent variables are all lags of the dependent variable and/or lags of the errors.
- This seems to us a natural solution to address the problem we are investigating, as its nature is that of a typical time-series and this method allows us to be both effective and more parsimonious in our choice of covariates, as just the response variable and time are strictly necessary.
- As a first step, we need to plot the characteristics of our time series along three different dimensions which are namely trend, seasonality and random behaviour to check for potential non-stationarity of our time-series.

# ARIMA


Decomposition of additive time series
observed
trend
seasonal
random
2025
2030
2035
Time


Decomposition of additive time series
observed
trend
seasonal
random
2025
2030
2035
Time


Decomposition of additive time series
observed
trend
seasonal
random
2025
2030
2035
Time

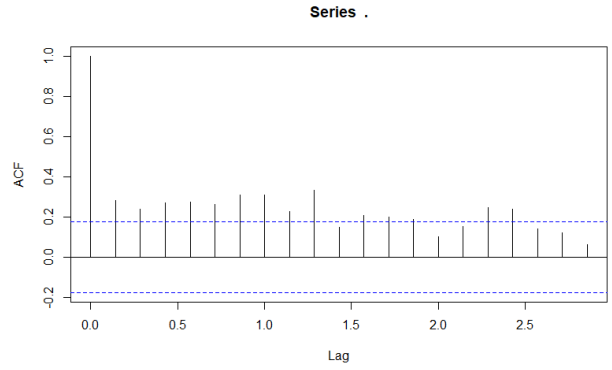
# ARIMA

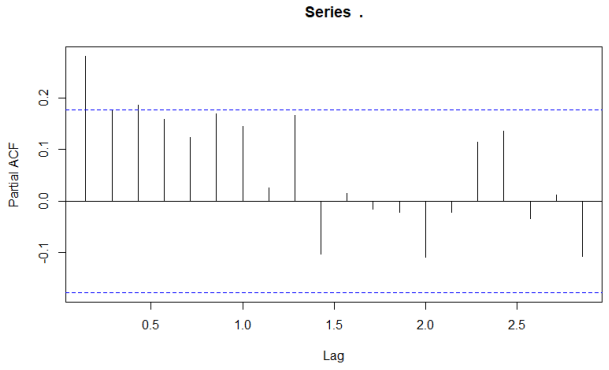
- As you can see from the previous slide, we need to differentiate the series at least two times to have an approximately stationary series.

```
        Augmented Dickey-Fuller Test

data:  .
Dickey-Fuller = -8.5345, Lag order = 0, p-value = 0.01
alternative hypothesis: stationary
```

```
        KPSS Test for Level Stationarity

data:  .
KPSS Level = 0.092732, Truncation lag parameter = 4, p-value = 0.1

Warning message:
In tseries::kpss.test(.) : p-value greater than printed p-value
```

- Now we can focus on understanding which are the best parameters to feed our model, in particular the p-order and q-order of non seasonal auto-regressive (NSAR) and moving average (NSMA) respectively.




- On the left side we plot the auto-correlation function to help us choosing the q-parameter, while on the right side we have partial auto-correlation to provide hints on the value of p.

# ARIMA: 2 different approaches

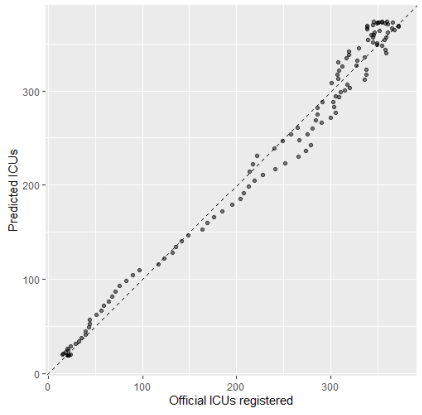
- From the previous qualitative considerations, we ended up deciding to set a first ARIMA model with p=6, d=2, q=2, in short ARIMA(6,2,2). Here below the results of model fitting:

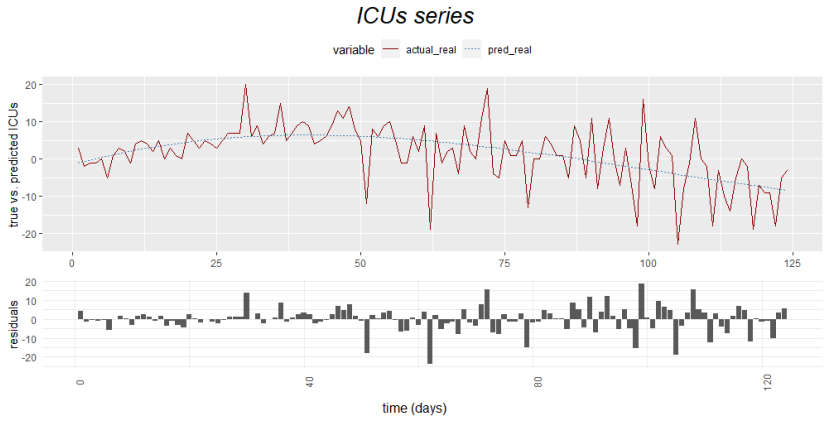
```
Series: outcome
ARIMA(6,2,2)

Coefficients:
          ar1      ar2      ar3      ar4      ar5      ar6     ma1      ma2
      -1.1574  -0.6120  -0.5555  -0.4382  -0.3527  -0.2171  0.2103  -0.5511
s.e.   0.1812   0.2491   0.2335   0.2107   0.1798   0.1046  0.1745   0.1630

sigma^2 estimated as 47.36:  log likelihood=-405.28
AIC=828.55   AICc=830.16   BIC=853.79
```

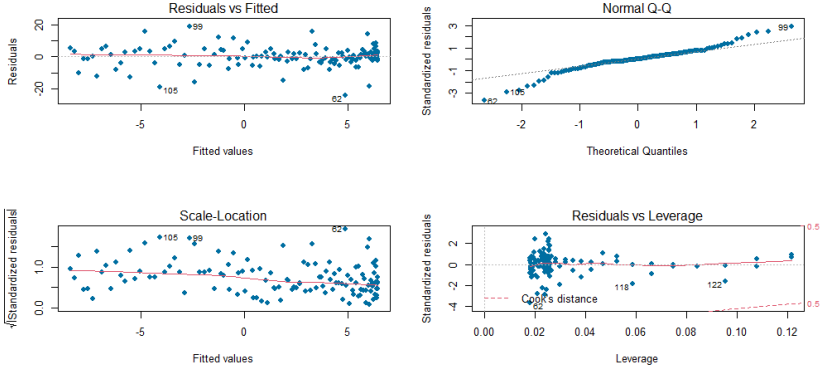
- To compare and contrast the previous findings, we also adopted a parameters' automatic tuning approach, by means of the auto.arima function. The ending result is the following ARIMA(1,2,2) model, which slightly improves the AIC of the fit:

```
Series: outcome
ARIMA(1,2,2)

Coefficients:
         ar1      ma1     ma2
      0.8492  -1.8772  0.9127
s.e.  0.0753   0.0564  0.0529

sigma^2 estimated as 45.65:  log likelihood=-406.01
AIC=820.02   AICc=820.36   BIC=831.23
```

# Results: LM in sample

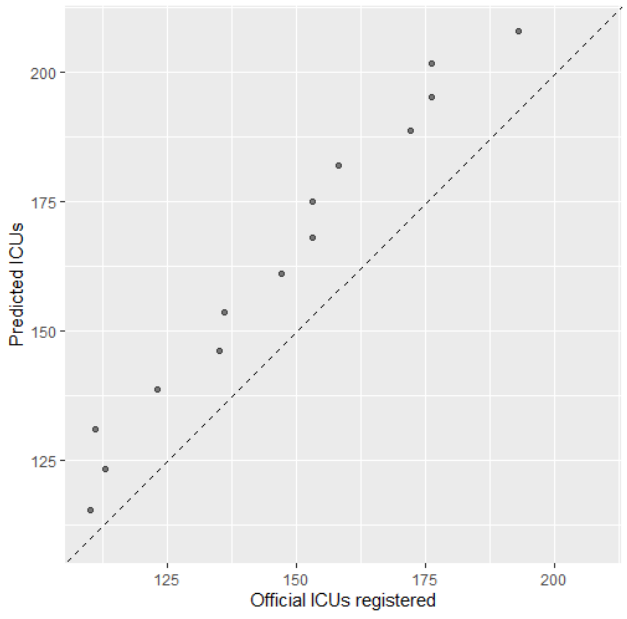


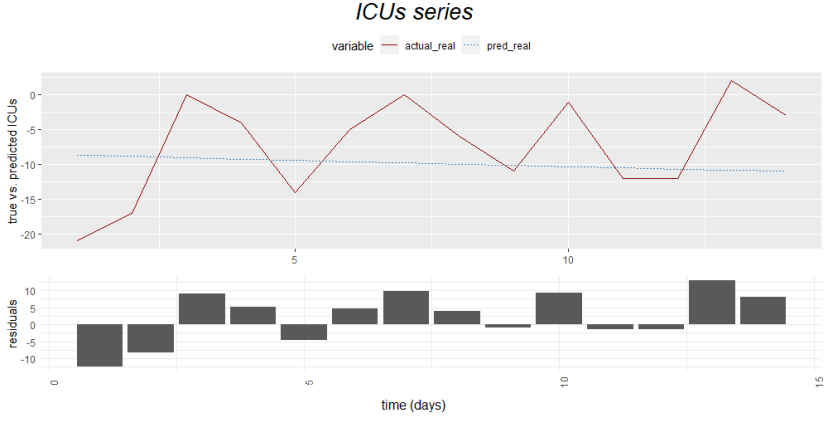




| LM Train Error Table | | |
|---|---|---|
| | transf_metrics | real_metrics |
| rmse | 6.42 | 14.98 |
| rsq | 0.323 | 0.985 |
| mae | 4.47 | 12.05 |
| mape | - | 7.55% |
| mase | 0.668 | 1.945 |
| smape | 96.66% | 7.28% |

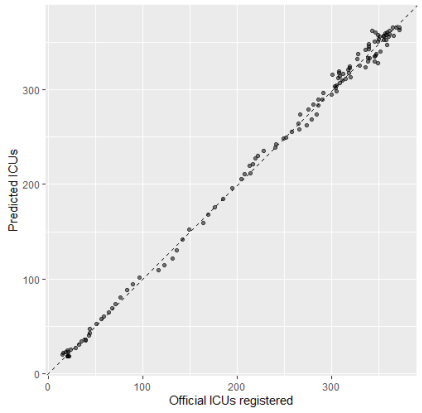
# Results: LM out-of-sample

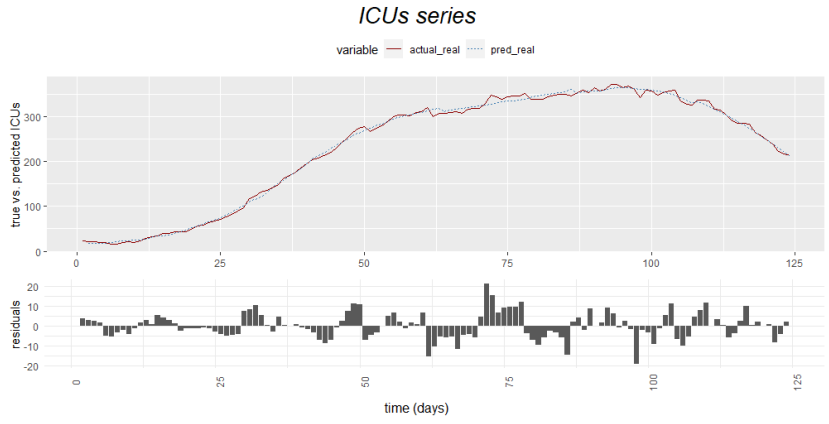


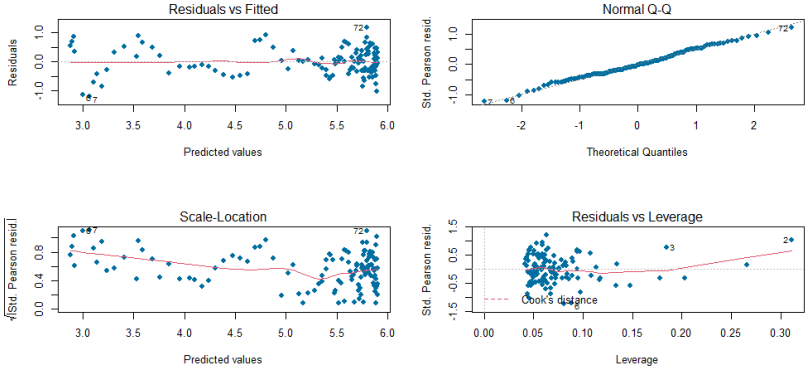




| LM Test Error Table | | |
|---|---|---|
| | transf_metrics | real_metrics |
| rmse | 6.87 | 16.89 |
| rsq | 0.152 | 0.977 |
| mae | 5.71 | 15.86 |
| mape | - | 11.31% |
| mase | 0.762 | 2.337 |
| smape | 89.69% | 10.30% |

# Results: GLM Poisson in sample

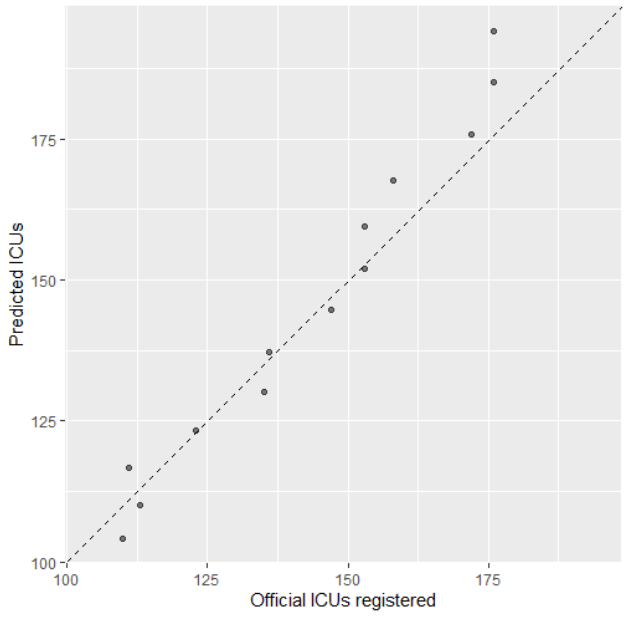


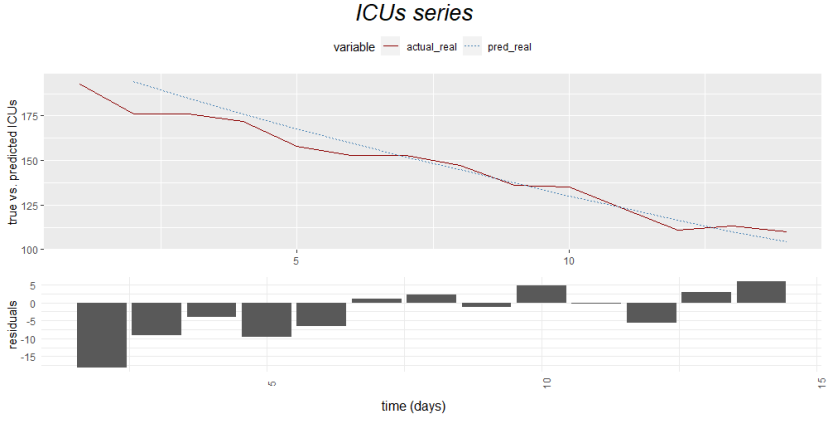




| GLM Poisson Train Error Table | | |
|---|---|---|
| | transf_metrics | real_metrics |
| rmse | 0.06 | 6.39 |
| rsq | 0.995 | 0.997 |
| mae | 0.04 | 4.95 |
| mape | 0.92% | 3.67% |
| mase | 0.888 | 0.795 |
| smape | 0.92% | 3.60% |

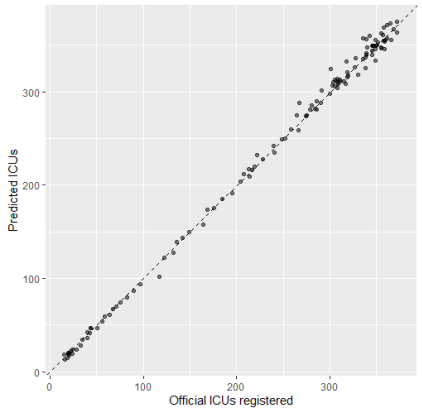
# Results: GLM Poisson out-of-sample

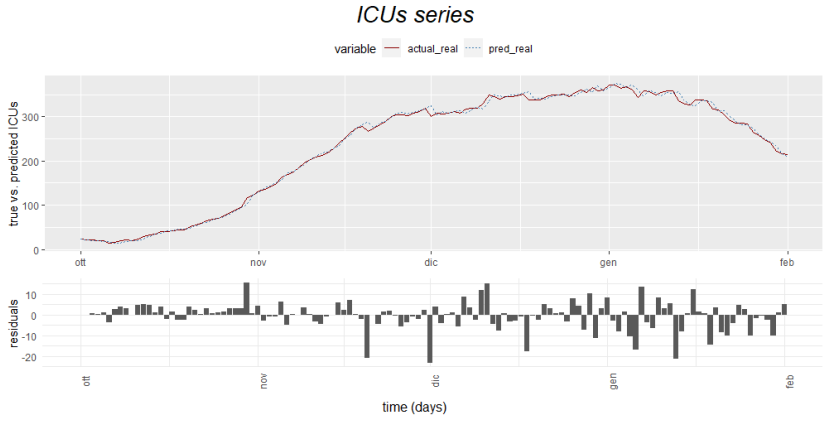


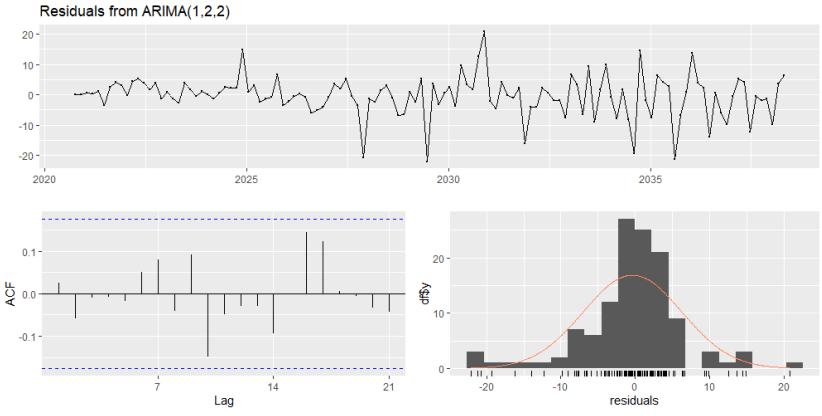




| GLM Poisson Test Error Table | | |
|---|---|---|
| | transf_metrics | real_metrics |
| rmse | 0.044 | 7.15 |
| rsq | 0.970 | 0.969 |
| mae | 0.04 | 5.48 |
| mape | 0.73% | 3.69% |
| mase | 0.861 | 0.940 |
| smape | 0.73% | 3.63% |

# Results: ARIMA in sample

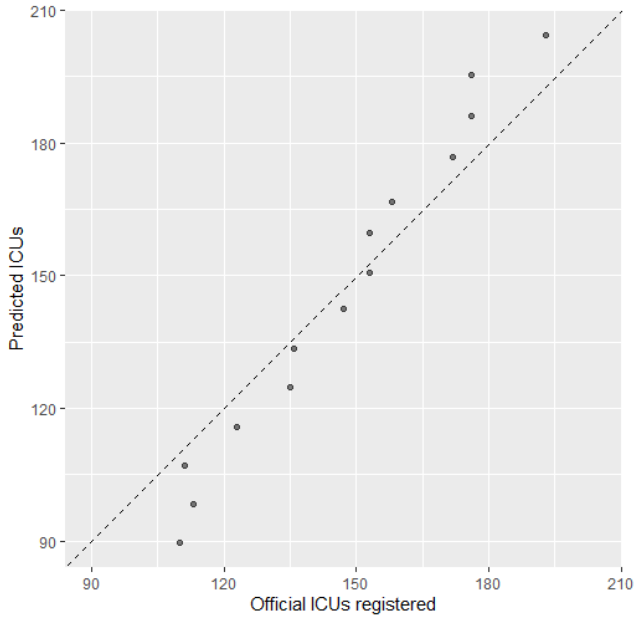


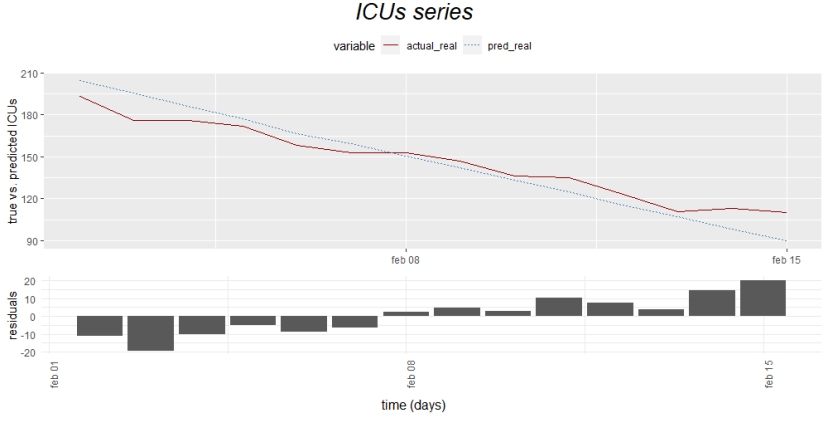




| ARIMA Train Error Table | | |
|---|---|---|
| transf_metrics | | real_metrics |
| rmse | - | 6.59 |
| rsq | - | 0.997 |
| mae | - | 4.59 |
| mape | - | 3.07% |
| mase | - | 0.740 |
| smape | - | 3.14% |

# Results: ARIMA out-of-sample




| ARIMA Test Error Table | | |
|---|---|---|
| | transf_metrics | real_metrics |
| rmse | - | 10.62 |
| rsq | - | 0.978 |
| mae | - | 9.05 |
| mape | - | 6.44% |
| mase | - | 1.353 |
| smape | - | 6.59% |

# Conclusions

As expected, one of the main results coming from this experiment is that each model offers some pros and some cons. In short, for LMs we appreciate the simplicity of their formulation which provides a deeper understanding of the model for the end user, and eases the communication of its results. On the downside, it yields to slightly higher errors of prediction producing less reliable estimates.

| | LM Test Error Table | GLM Test Error Table | ARIMA Test Error Table |
|---|---|---|---|
| | real_metrics | real_metrics | real_metrics |
| rmse | 16.89 | 7.15 | 10.62 |
| rsq | 0.977 | 0.969 | 0.978 |
| mae | 15.86 | 5.48 | 9.05 |
| mape | 11.314% | 3.691% | 6.442% |
| mase | 2.337 | 0.940 | 1.353 |
| smape | 10.299% | 3.629% | 6.600% |

For the GLMs here analysed, we value their overall higher quality of prediction out-of-sample, and their design which favours their re-usability in different contexts. On the other hand, the interpretation of some covariates' interactions may be more difficult to explain and, it seems to us that the benefit of adding further features may have a somewhat limited upside in performance. Finally, ARIMAs bring a different framework where we do not need to predict other covariates (or lag existing ones) to formulate a prediction, leading anyway to satisfying results; as for GLM, there may arise some issues related to its interpretability due to its intrinsic bootstrapping nature in prediction.

# Table of predictions - ICUs values

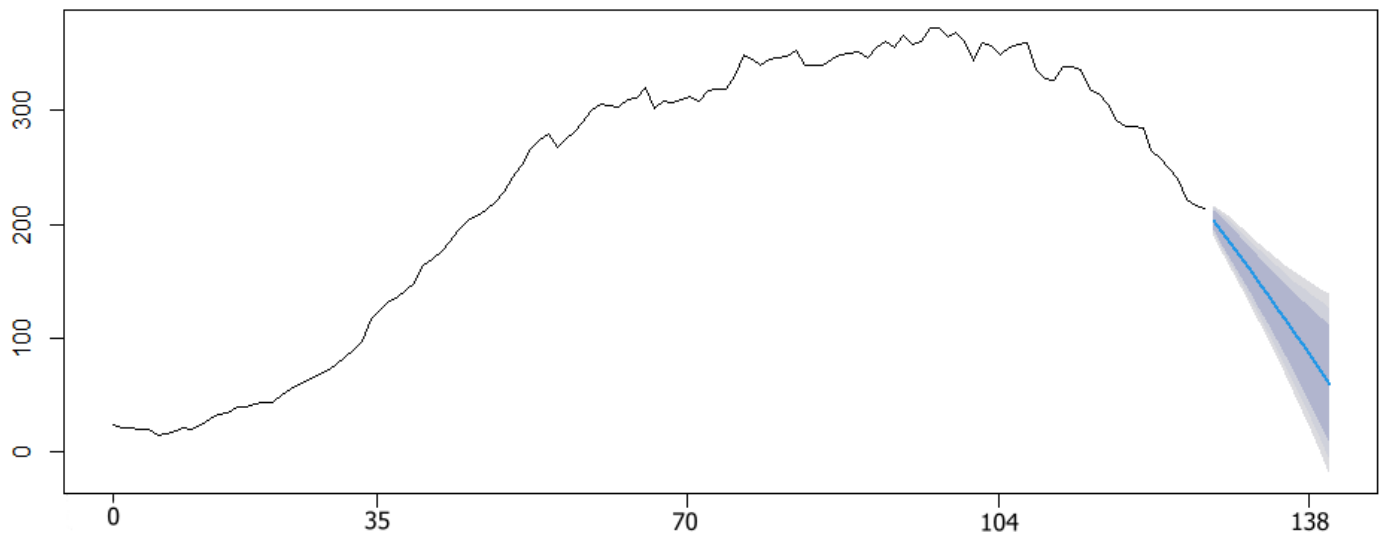
Below you can find the forward predictions for Veneto's ICUs on the period starting from February, 2nd to February 15, 2021: on the left we present the point estimates for one of our 2 models of choice (ARIMA) and, moving to the right, the respective upper and lower bounds for increasing confidence intervals (80%, 90%, 95%).

| Dates | Point Forecast | Lower 80% | Higher 80% | Lower 90% | Higher 90% | Lower 95% | Higher 95% |
|---|---|---|---|---|---|---|---|
| 02/02/2021 | 204 | 195 | 213 | 192 | 215 | 190 | 217 |
| 03/02/2021 | 193 | 180 | 206 | 177 | 209 | 174 | 212 |
| 04/02/2021 | 183 | 167 | 198 | 163 | 202 | 159 | 206 |
| 05/02/2021 | 171 | 154 | 189 | 149 | 194 | 144 | 199 |
| 06/02/2021 | 160 | 140 | 181 | 135 | 186 | 130 | 191 |
| 07/02/2021 | 150 | 127 | 172 | 120 | 179 | 115 | 184 |
| 08/02/2021 | 139 | 113 | 164 | 106 | 171 | 100 | 178 |
| 09/02/2021 | 127 | 99 | 156 | 91 | 164 | 84 | 171 |
| 10/02/2021 | 116 | 84 | 148 | 75 | 157 | 67 | 165 |
| 11/02/2021 | 105 | 69 | 141 | 59 | 151 | 50 | 160 |
| 12/02/2021 | 93 | 54 | 133 | 42 | 144 | 33 | 154 |
| 13/02/2021 | 82 | 38 | 126 | 25 | 139 | 14 | 149 |
| 14/02/2021 | 70 | 22 | 119 | 8 | 133 | 0 | 145 |
| 15/02/2021 | 59 | 5 | 112 | 0 | 128 | 0 | 141 |

Note that a floor of zero has been put for negative predictions (highlighted).

# Table of predictions - a visual interpretation

Finally, we provide a qualitative plot to allow appreciating the 3 different prediction-ranges along with point predictions: it is easy to notice that for increasing confidence levels (i.e. from 80% to 95%) the prediction bands become larger; secondly, the more time passes, the further we move out-of-sample so we get bands increasingly larger (less precise estimates). All this suggests that for a higher precision it would be better to re-fit model frequently and focus on shorter term predictions (e.g. 1-4 days).



# References


- [1] https://www.proquest.com/docview/2528274421
- [2] https://statgroup-19.blogspot.com/2020/03/why-go-for-poisson-regression-and-not.html?m=0
- [3] https://onlinelibrary.wiley.com/doi/10.1002/bimj.202000189
- [4] https://github.com/ImperialCollegeLondon/covid19model/blob/master/Technical_description_of_Imperial_COVID_19_Model.pdf
- [5] https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30627-9/fulltext
- [6] https://www.mdpi.com/2075-4426/11/5/343