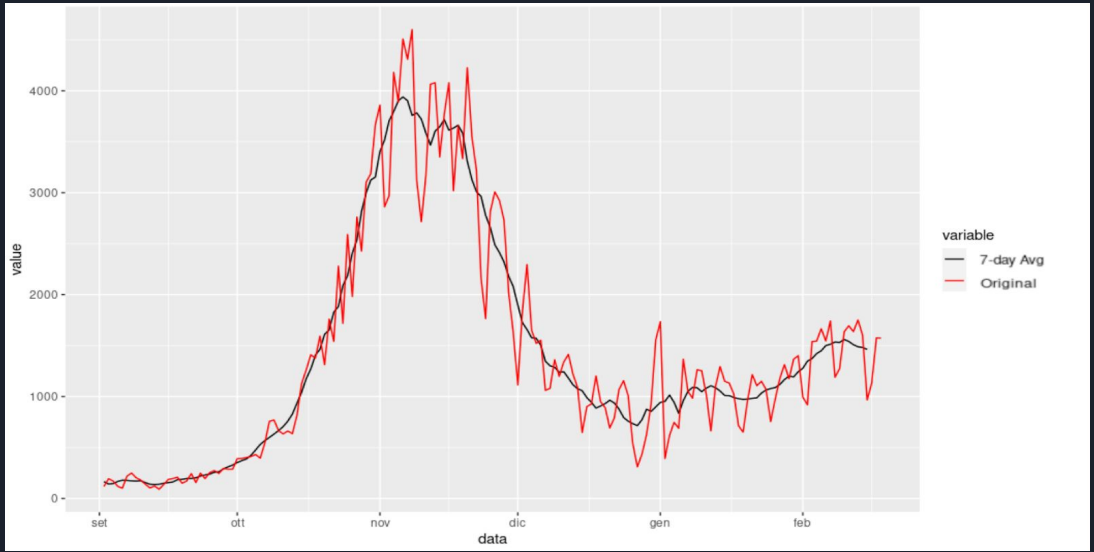
SMDS Final Project: Covid in Campania

Berti, Guglielmo, Tosato, Zinna

Problem Statement: Covid-19 in Campania

- Aim: build a model to predict Covid-19 spread in Campania
- Datasets: Protezione Civile's GitHub data [1], Harvard's dataverse [2], Apple Mobility Report [3] & Ansa [4]
- Response Variable: `nuovi_positivi`

Explanatory analysis



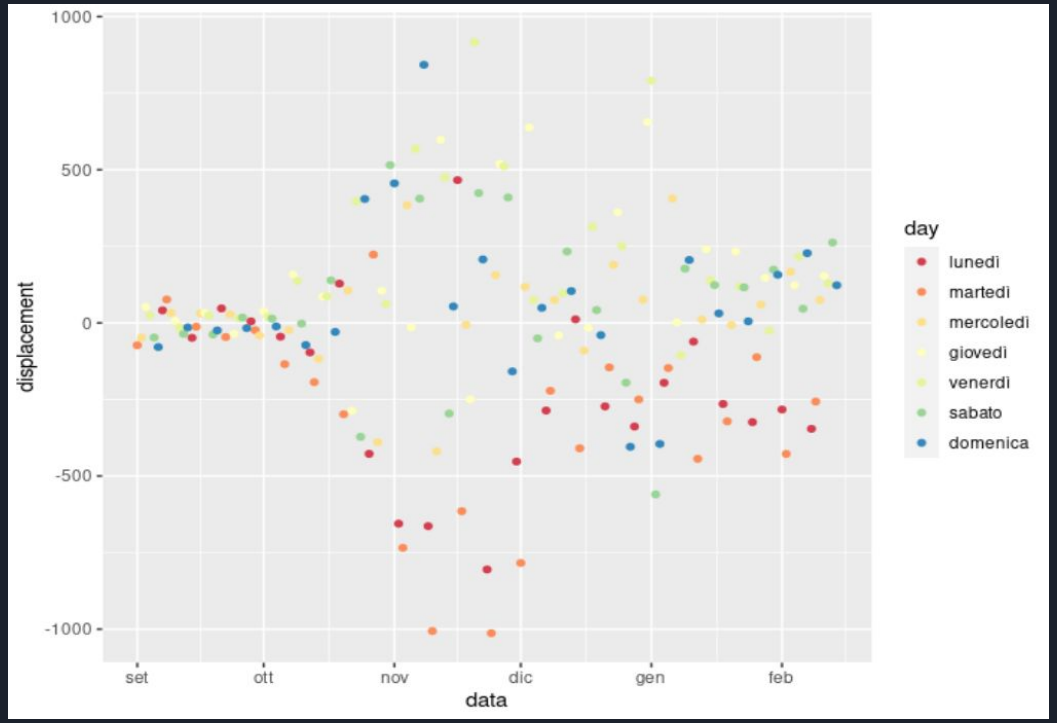
Nuovi Positivi & NP 7-day avg

Peak of cases mid-November

Notable events:

- 15 Oct: schools closed
- 6 Nov: “colors” system introduced
- 23 Dec - 6 Jan: Christmas break

Data Quality



Displacement from 7-day average

Weekday influenced response variable

7-day moving average was used

Data Quality

data	stato	codice_regione	denominazione_regione	lat	long	ricoverati_con_sintomi	terapia_intensiva	totale_ospedalizzati	isolamento_domiciliare
2021-02-16T17:00:00	ITA	13	Abruzzo	42.35122196	13.39843823	535	64	599	11705
2021-02-16T17:00:00	ITA	17	Basilicata	40.63947052	15.80514834	72	7	79	3381
2021-02-16T17:00:00	ITA	18	Calabria	38.90597598	16.59440194	197	20	217	6443
2021-02-16T17:00:00	ITA	15	Campania	40.83956555	14.25084984	1284	106	1390	67590
2021-02-16T17:00:00	ITA	08	Emilia-Romagna	44.49436681	11.341720800000001	1975	175	2150	34887
2021-02-16T17:00:00	ITA	06	Friuli Venezia Giulia	45.6494354	13.76813649	396	60	456	9015
2021-02-16T17:00:00	ITA	12	Lazio	41.89277044	12.48366722	2065	253	2318	34943
2021-02-16T17:00:00	ITA	07	Liguria	44.41149315	8.9326992	587	59	646	4314
2021-02-16T17:00:00	ITA	03	Lombardia	45.46679409	9.190347404	3693	373	4066	44607
2021-02-16T17:00:00	ITA	11	Marche	43.61675973	13.5188753	535	80	615	7390
2021-02-16T17:00:00	ITA	14	Molise	41.55774754	14.65916051	80	11	91	1423
2021-02-16T17:00:00	ITA	21	P.A. Bolzano	46.49933453	11.35662422	254	39	293	7656
2021-02-16T17:00:00	ITA	22	P.A. Trento	46.06893511	11.12123097	168	27	195	2522
2021-02-16T17:00:00	ITA	01	Piemonte	45.0732745	7.680687483	1900	141	2041	10185
2021-02-16T17:00:00	ITA	16	Puglia	41.12559576	16.86736689	1324	156	1480	37708
2021-02-16T17:00:00	ITA	20	Sardegna	39.21531192	9.110616306	311	24	335	13491
2021-02-16T17:00:00	ITA	19	Sicilia	38.11569725	13.362356699999998	1005	158	1163	33317
2021-02-16T17:00:00	ITA	09	Toscana	43.76923077	11.25588885	718	133	851	11460

Screenshot from Protezione Civile's GitHub

- Datasets updated daily
- Some variables are deprecated (e.g. `casi_da_sospetto_diagnostico`), some are newly added (e.g. `tamponi_test_molecolare`)
- RT and "color" missing from PC GitHub
- Transformed cumulative variables into daily variables

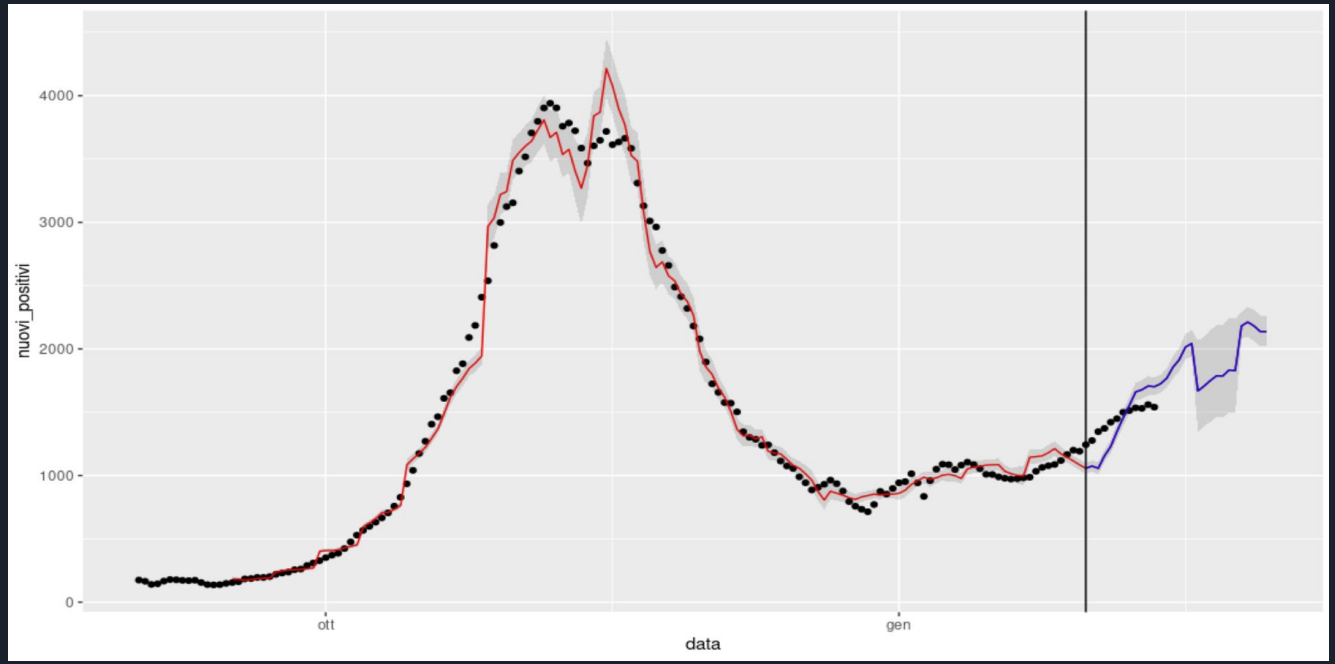
First model:

- Model for “nuovi positivi” from 15 days prior covariates

First Model

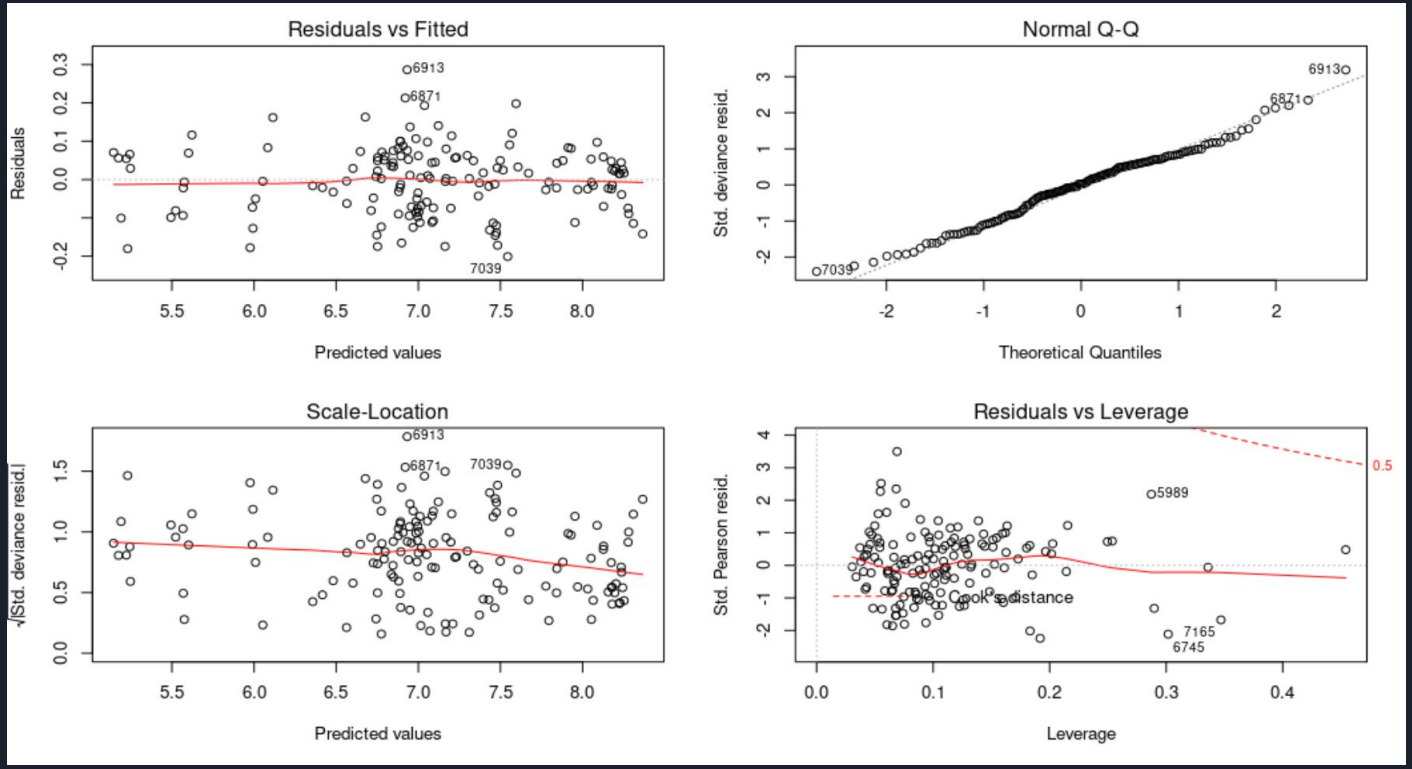
- Nuovi Positivi predicted using 15-day prior covariates
- Gamma GLM with log link
- Covariates used:
  - rt
  - color
  - tamp_rate
  - terapia_intensiva
  - isolamento_domiciliare
- Fitting from 01/09/2020 to 31/01/2021
- Standard errors extracted using Delta method

Model plot



First model's predictions on historical data

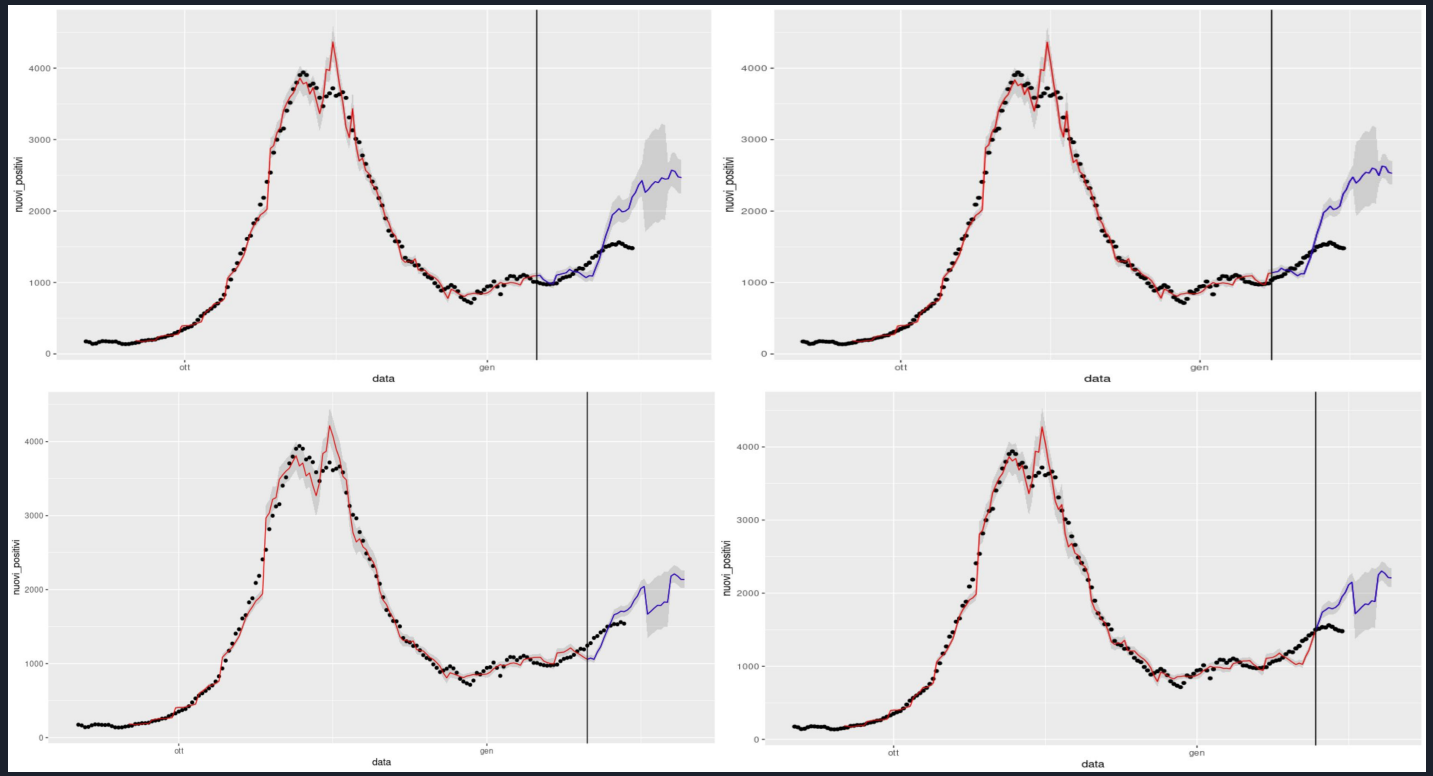
First Model's diagnostic



Summary

```
Deviance Residuals:
      Min         1Q     Median         3Q        Max
-0.264142  -0.064350  -0.003574   0.056068   0.284846

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                8.631e+00  5.659e-01  15.251  < 2e-16 ***
I(tamp_rate^2)            -3.656e+01  5.033e+00  -7.264 2.45e-11 ***
terapia_intensiva          1.659e-02  9.463e-04  17.528  < 2e-16 ***
I(isolamento_domiciliare) -1.967e-05  1.974e-06  -9.962  < 2e-16 ***
I(rt^2)                    5.631e+00  6.006e-01   9.376  < 2e-16 ***
tamp_rate:colorOrange      9.625e+00  1.997e+00   4.819 3.73e-06 ***
tamp_rate:colorRed         1.235e+01  2.073e+00   5.959 1.98e-08 ***
tamp_rate:colorWhite       1.441e+01  1.616e+00   8.916 2.47e-15 ***
tamp_rate:colorYellow      9.810e+00  1.886e+00   5.202 6.91e-07 ***
colorOrange:rt            -8.328e+00  1.132e+00  -7.357 1.49e-11 ***
colorRed:rt               -9.014e+00  1.087e+00  -8.290 8.69e-14 ***
colorWhite:rt             -9.490e+00  1.159e+00  -8.188 1.54e-13 ***
colorYellow:rt            -8.230e+00  1.102e+00  -7.466 8.23e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Gamma family taken to be 0.008987275)

    Null deviance: 80.4340  on 151  degrees of freedom
Residual deviance:  1.2458  on 139  degrees of freedom

AIC: 1878.8
```

Robustness



Second model:

- Model for “nuovi positivi” for the same day
- 14 - days predictions for the covariates

1 day model for “Nuovi positivi”

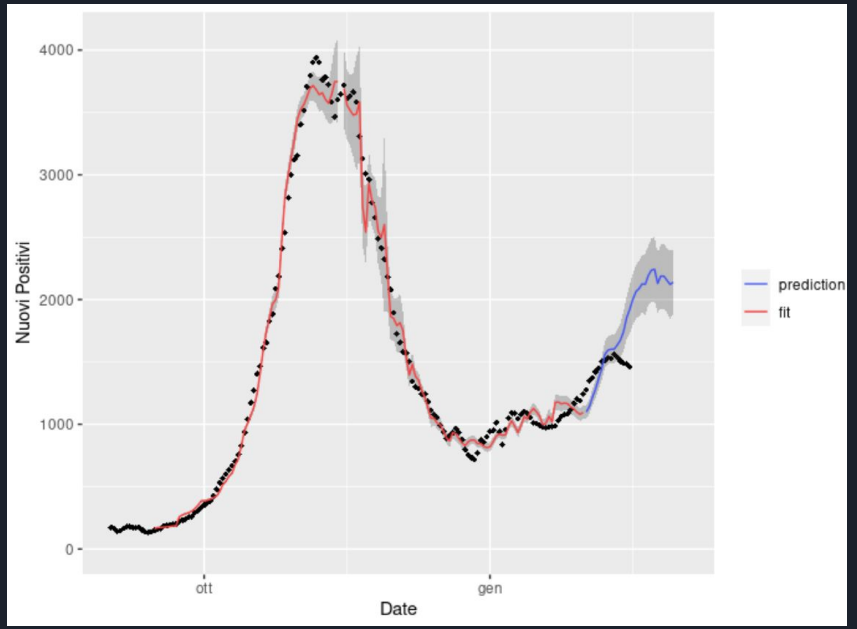
It’s aim is to predict the expected value for nuovi positivi, given the value of:

- Terapia intensiva
- Isolamento domiciliare
- Ricoverati con sintomi
- Tamponi
- R_t (with a 15 days lag)
- Color of the region (with a 15 day lag)

The model

To describe the data a semiparametric glm with the following properties has been used:

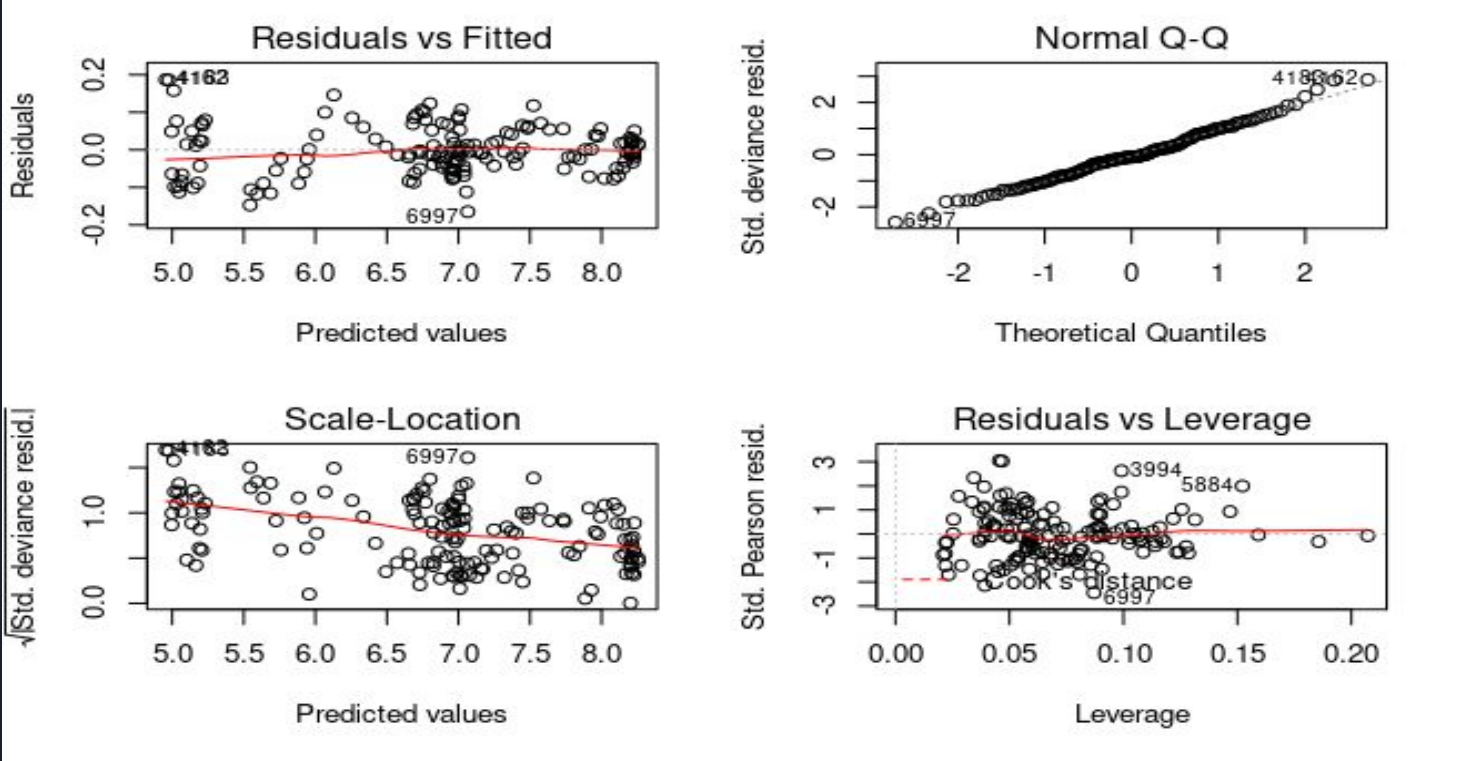
- Gamma glm family with log link function
- Polynomial terms of some regressors.
- Lagged R_t and lagged Color have been used only as interaction terms.



Summary

```
Deviance Residuals:
      Min         1Q     Median         3Q        Max
-0.165094  -0.046992  -0.005952   0.039366   0.187269

Coefficients:
                                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                                        4.459e+00  2.364e-02 188.635  < 2e-16 ***
poly(terapia_intensiva, 2, raw = TRUE)1            1.356e-02  1.919e-03   7.068 6.54e-11 ***
poly(terapia_intensiva, 2, raw = TRUE)2           -4.360e-05  6.525e-06  -6.682 4.97e-10 ***
poly(isolamento_domiciliare, 1, raw = TRUE)       -1.095e-05  1.280e-06  -8.554 1.73e-14 ***
I(isolamento_domiciliare^3)                        1.331e-15  9.740e-17  13.663  < 2e-16 ***
poly(ricoverati_con_sintomi, 2, raw = TRUE)1:old_rt 1.395e-03  1.904e-04   7.327 1.63e-11 ***
poly(ricoverati_con_sintomi, 2, raw = TRUE)2:old_rt -3.502e-07  5.850e-08  -5.987 1.67e-08 ***
old_rt:tamponi:old_colorOrange                     3.938e-05  7.617e-06   5.170 7.81e-07 ***
old_rt:tamponi:old_colorRed                        3.653e-05  6.755e-06   5.408 2.64e-07 ***
old_rt:tamponi:old_colorWhite                      5.406e-05  6.445e-06   8.388 4.47e-14 ***
old_rt:tamponi:old_colorYellow                     4.820e-05  6.758e-06   7.132 4.62e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for Gamma family taken to be 0.004454598)

    Null deviance: 120.46369  on 152  degrees of freedom
Residual deviance:   0.62514  on 142  degrees of freedom
AIC: 1714
```

Diagnostic Plots



Confidence Intervals

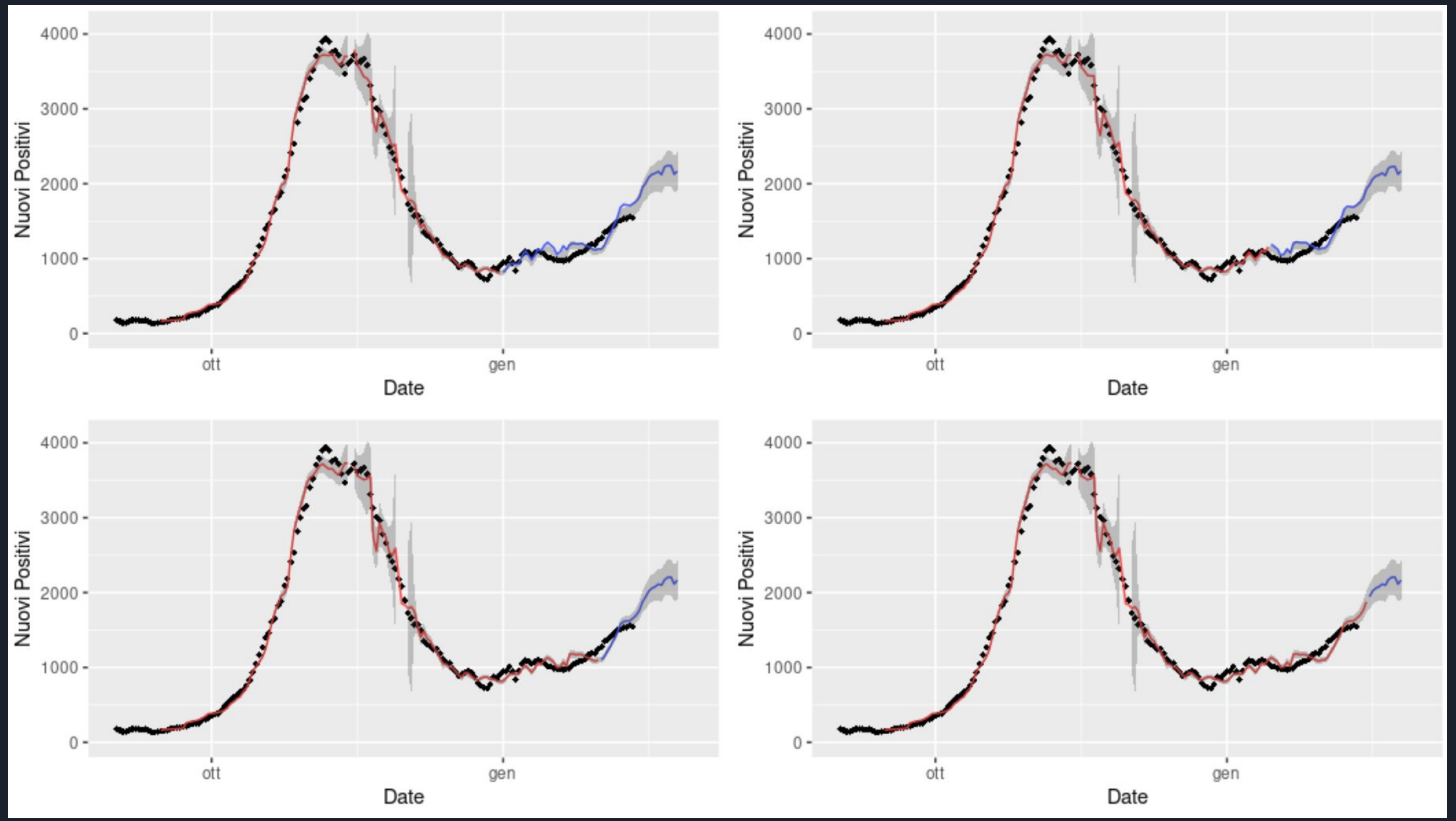
Confidence interval on this model is not trivial to estimate.

Standard error given by R prediction doesn't take into account covariates error, and his propagation is analytically hard to find.

The following bootstrap approach has been used:

- Resample dataset
- Build covariates models on resampled dataset
- Predict all covariates
- Build "one day model" using the predicted values of the covariates
- Predict "nuovi positivi"

Robustness of the model



Covariates models

In order to use the “one day” model and predict our main outcome variable, we need the 14 days prediction of following covariates:

- Terapia intensiva
- Isolamento domiciliare
- Ricoverati con sintomi
- Tamponi

Covariates models

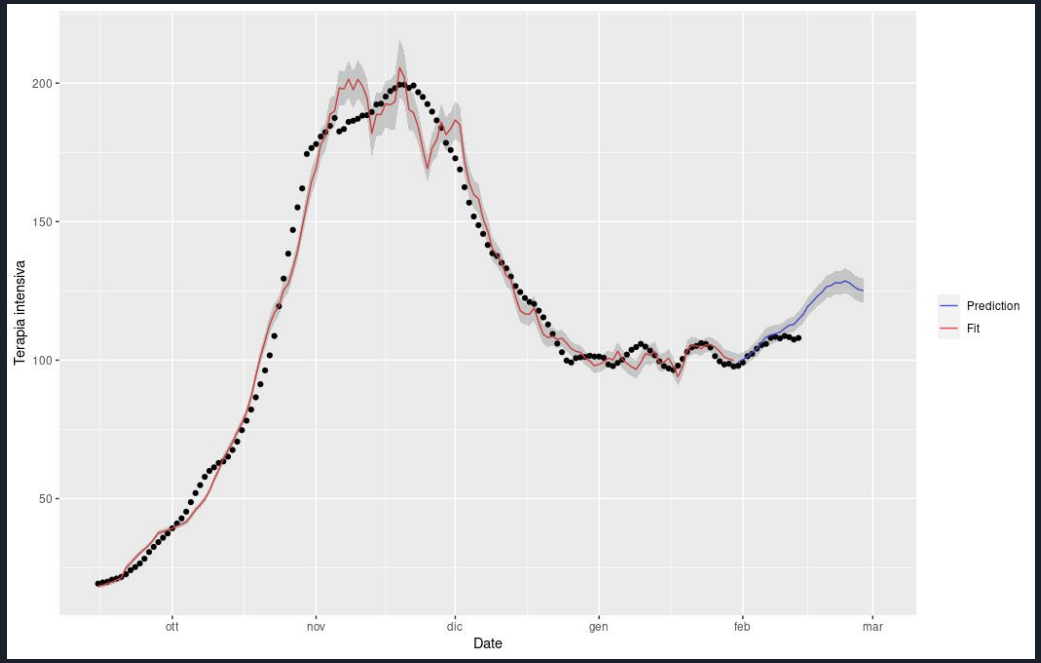
All covariates model are GLM, with log as link function and gamma distribution.

Model predictors:

- First order and/or higher order numerical predictors
- Categorical predictors
- Interaction terms, especially between categorical and numerical

Terapia intensiva

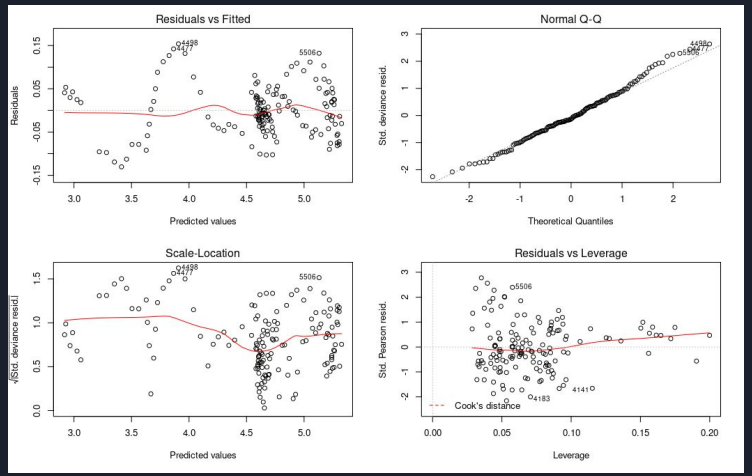
The model



Terapia intensiva

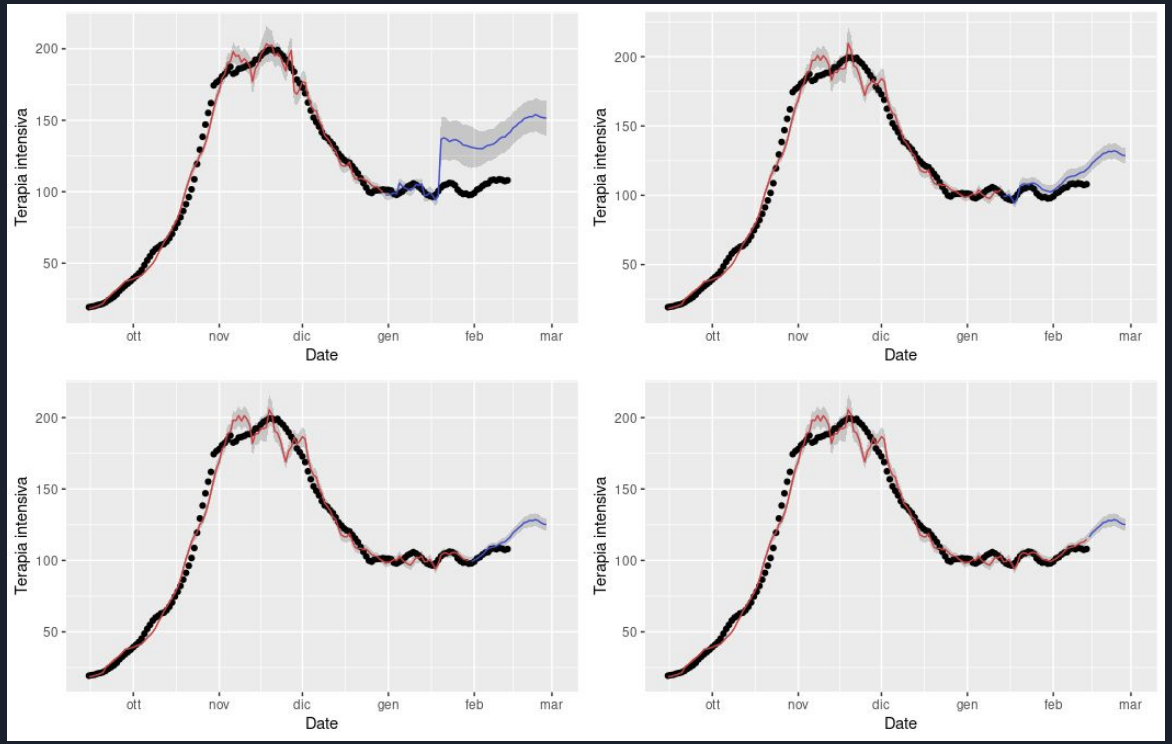
Summary and diagnostic plot

```
Coefficients:
                                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                                2.508e+00  7.570e-02  33.136  < 2e-16 ***
poly(isolamento_domiciliare, 2, raw = TRUE)1 -2.293e-05  2.331e-06  -9.838  < 2e-16 ***
poly(isolamento_domiciliare, 2, raw = TRUE)2  1.606e-10  1.734e-11   9.259 3.09e-16 ***
poly(terapia_intensiva, 3, raw = TRUE)1       4.855e-02  1.998e-03  24.293  < 2e-16 ***
poly(terapia_intensiva, 3, raw = TRUE)2      -3.300e-04  1.961e-05 -16.824  < 2e-16 ***
poly(terapia_intensiva, 3, raw = TRUE)3       6.919e-07  6.117e-08  11.312  < 2e-16 ***
poly(nuovi_positivi, 1, raw = TRUE)           2.999e-04  3.411e-05   8.793 4.57e-15 ***
rt:colorOrange                                3.443e-01  1.004e-01   3.428 0.000797 ***
rt:colorRed                                   3.930e-01  1.021e-01   3.851 0.000178 ***
rt:colorWhite                                 2.962e-01  8.871e-02   3.339 0.001077 **
rt:colorYellow                                3.741e-01  1.000e-01   3.740 0.000267 ***
---
```



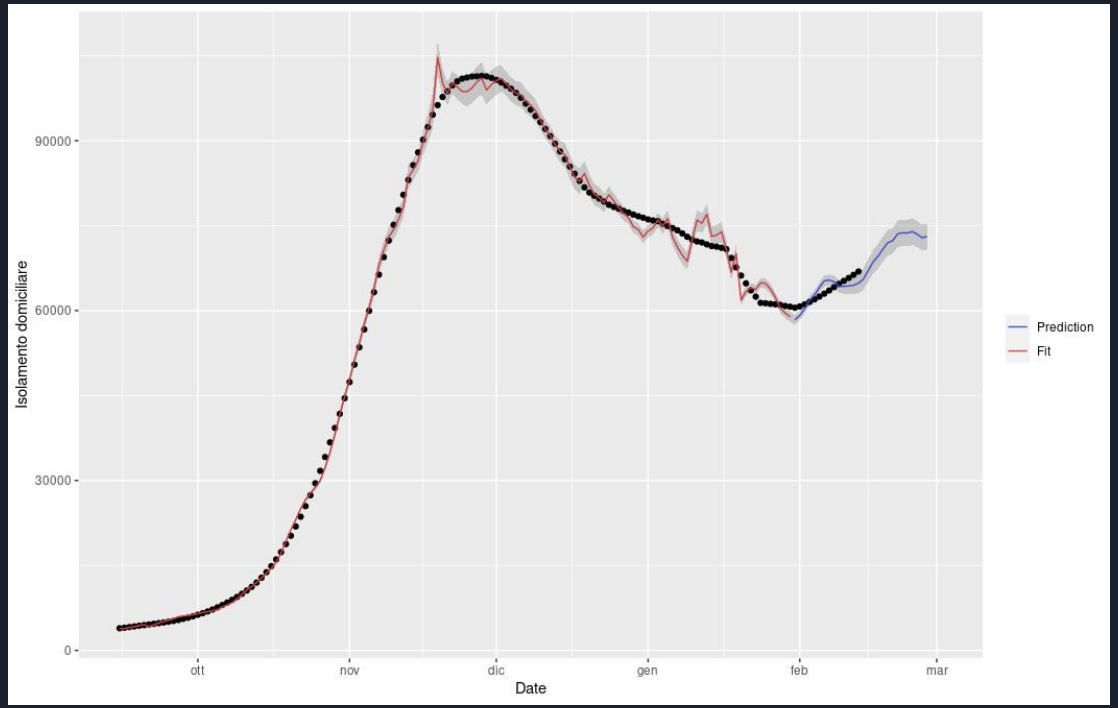
Terapia intensiva

Robustness towards time



Isolamento domiciliare

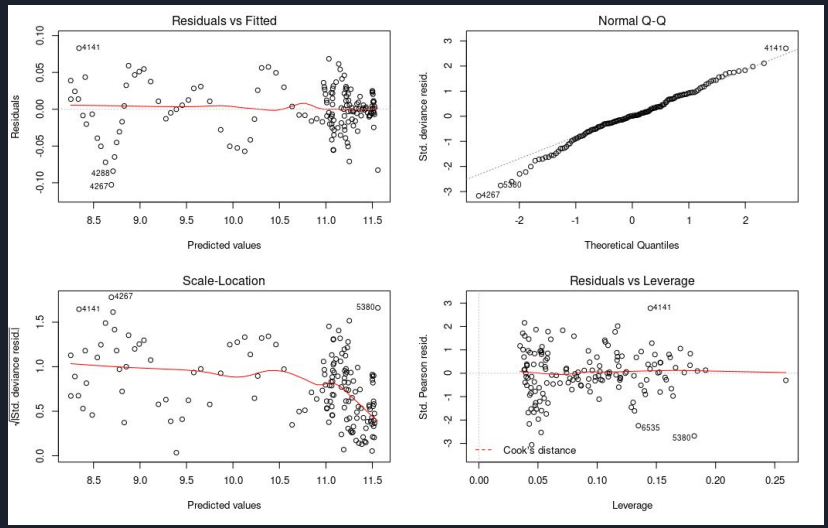
The model



Isolamento domiciliare

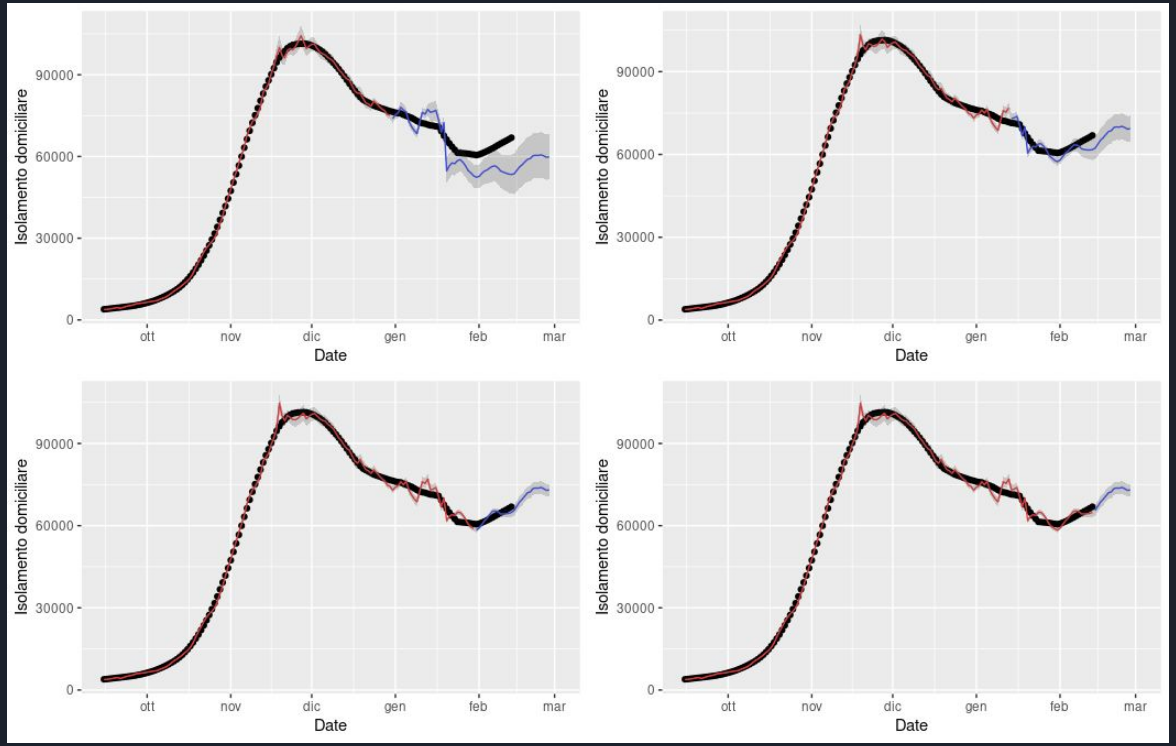
Summary and diagnostic plot

```
Coefficients:
                                      Estimate Std. Error t value Pr(>|t|)
(Intercept)                          6.998e+00  1.278e-01  54.743  < 2e-16 ***
poly(terapia_intensiva, 3, raw = TRUE)1  4.938e-02  1.531e-03  32.257  < 2e-16 ***
poly(terapia_intensiva, 3, raw = TRUE)2 -3.164e-04  1.358e-05 -23.292  < 2e-16 ***
poly(terapia_intensiva, 3, raw = TRUE)3  7.279e-07  3.817e-08  19.071  < 2e-16 ***
poly(nuovi_positivi, 3, raw = TRUE)1     9.911e-04  7.406e-05  13.382  < 2e-16 ***
poly(nuovi_positivi, 3, raw = TRUE)2    -4.035e-07  3.767e-08 -10.713  < 2e-16 ***
poly(nuovi_positivi, 3, raw = TRUE)3     5.100e-11  5.442e-12   9.370  < 2e-16 ***
colorRed                            -3.806e-01  1.271e-01  -2.993  0.00327 **
colorWhite                           1.129e+00  1.225e-01   9.214 4.66e-16 ***
colorYellow                          7.001e-01  1.490e-01   4.700 6.23e-06 ***
rt:colorOrange                       1.001e+00  1.307e-01   7.659 2.97e-12 ***
rt:colorRed                          1.419e+00  1.432e-01   9.910  < 2e-16 ***
rt:colorWhite                       -3.233e-01  4.996e-02  -6.471 1.57e-09 ***
rt:colorYellow                       1.694e-01  1.246e-01   1.359  0.17624
```



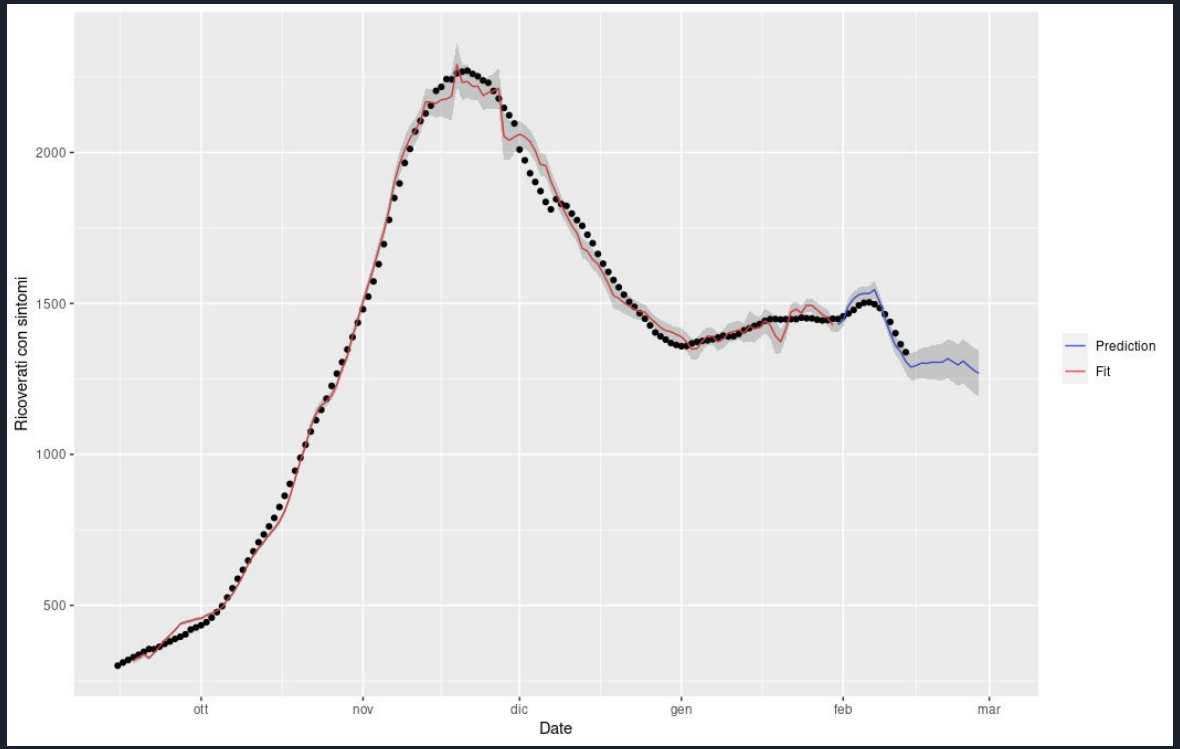
Isolamento domiciliare

Robustness towards time



Ricoverati con sintomi

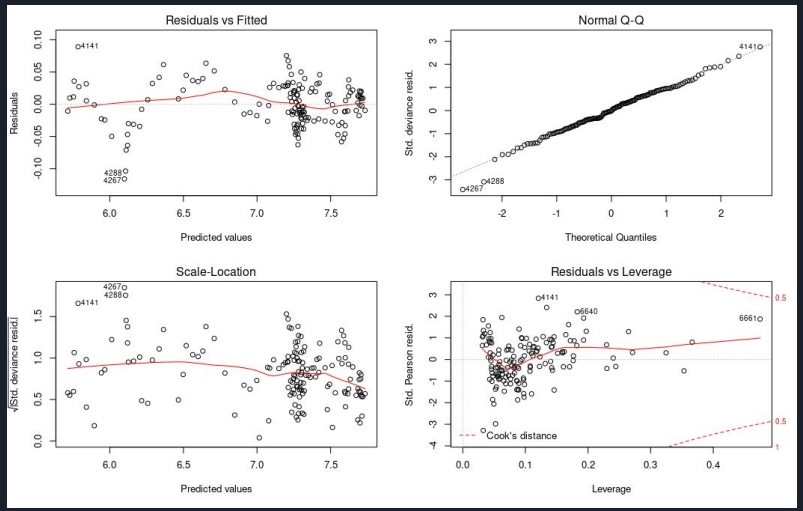
The model



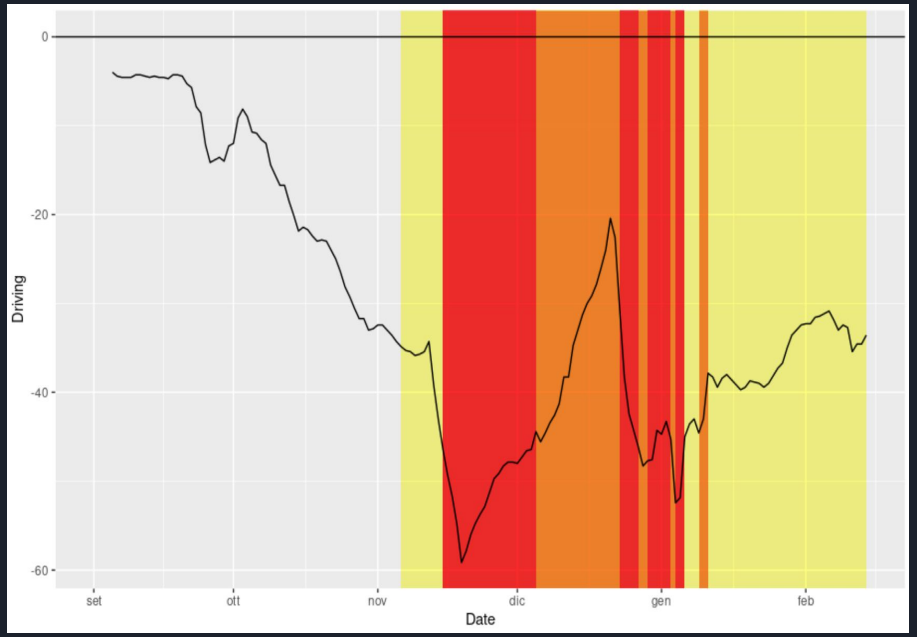
Ricoverati con sintomi

Summary and diagnostic plot

```
Coefficients:
                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                              4.914e+00  1.325e+00   3.709 0.000304 ***
poly(terapia_intensiva, 3, raw = TRUE)1  3.947e-02  1.201e-03  32.862  < 2e-16 ***
poly(terapia_intensiva, 3, raw = TRUE)2 -2.531e-04  1.273e-05 -19.886  < 2e-16 ***
poly(terapia_intensiva, 3, raw = TRUE)3  5.719e-07  3.841e-08  14.887  < 2e-16 ***
colorRed                                 2.080e+00  1.418e+00   1.467 0.144868
colorWhite                               7.718e-01  1.327e+00   0.582 0.561867
colorYellow                              2.997e+00  1.416e+00   2.116 0.036203 *
isolamento_domiciliare:colorOrange      -6.043e-07  7.276e-06  -0.083 0.933930
isolamento_domiciliare:colorRed         -1.147e-05  2.379e-06  -4.824 3.79e-06 ***
isolamento_domiciliare:colorWhite       -3.019e-06  1.307e-06  -2.309 0.022472 *
isolamento_domiciliare:colorYellow      -1.219e-05  2.130e-06  -5.723 6.63e-08 ***
colorOrange:rt                           1.940e-01  7.370e-01   0.263 0.792737
colorRed:rt                             -9.866e-01  2.786e-01  -3.541 0.000550 ***
colorWhite:rt                           -2.198e-01  5.658e-02  -3.885 0.000161 ***
colorYellow:rt                          -1.872e+00  3.170e-01  -5.904 2.80e-08 ***
rt:driving                              -3.993e-03  5.695e-04  -7.012 1.07e-10 ***
---
```



Apple’s Mobility Trends

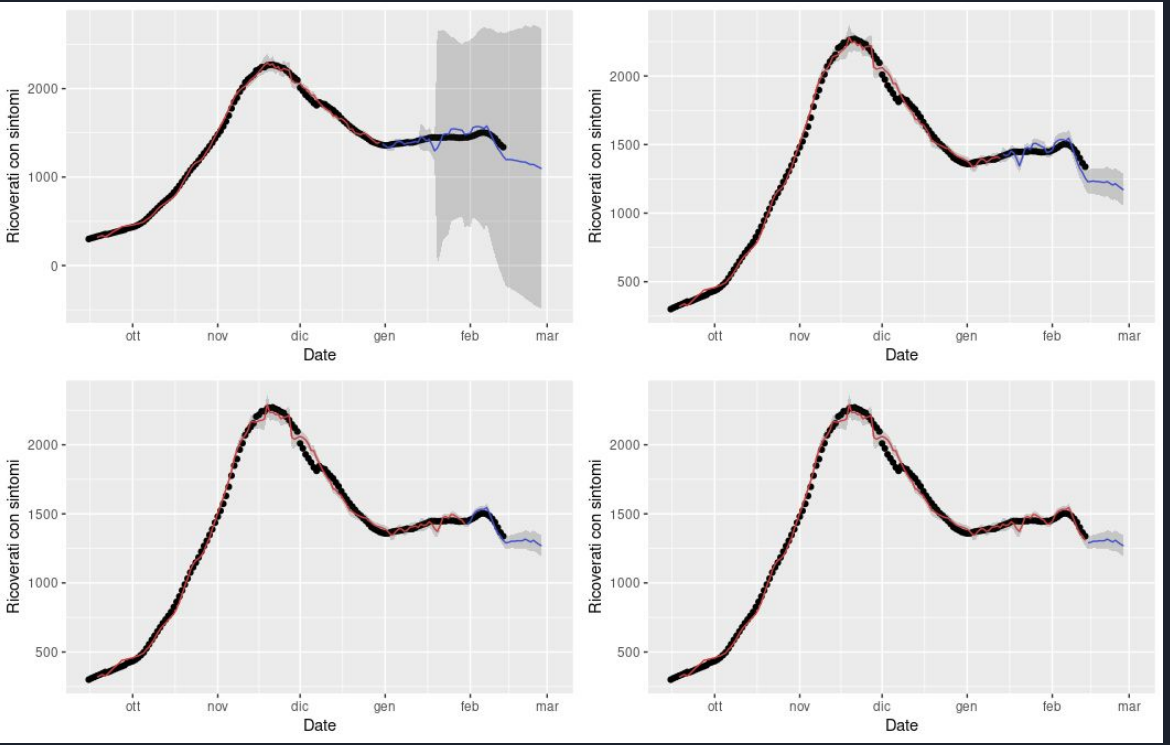


Mobily Trends in Campania, apple.com

Data generated by counting the number of requests to Apple Maps over the last year

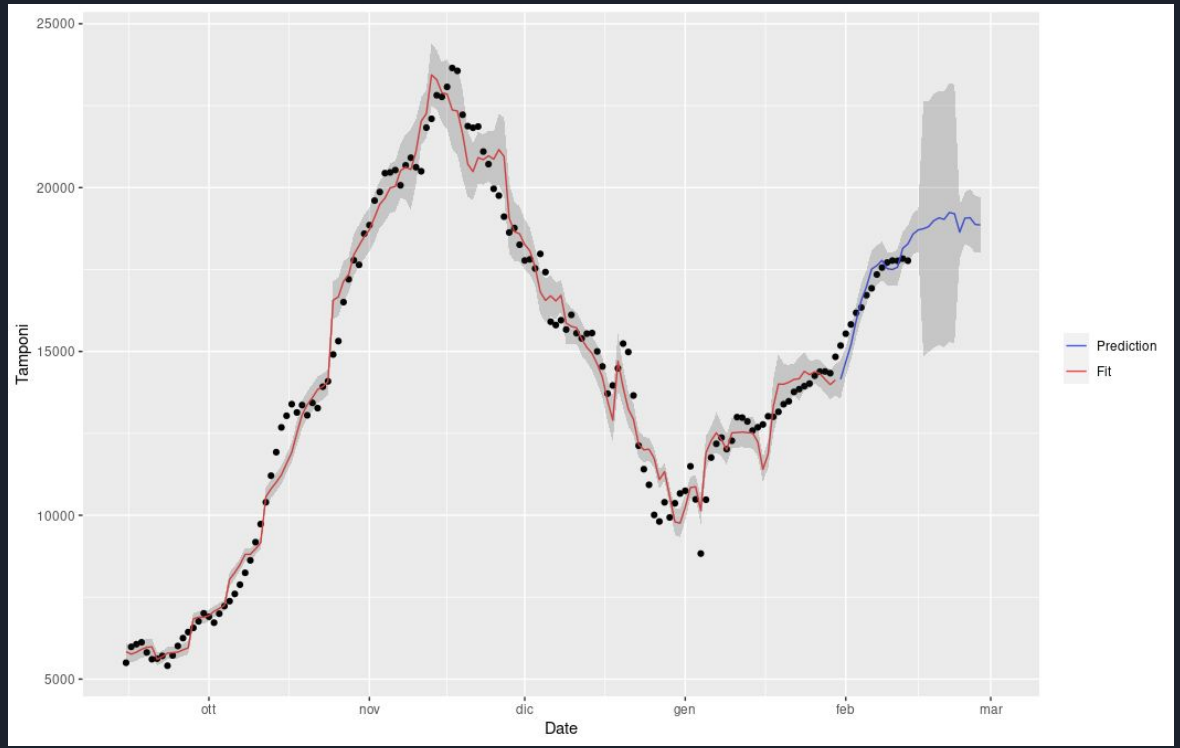
Ricoverati con sintomi

Robustness towards time



Tamponi

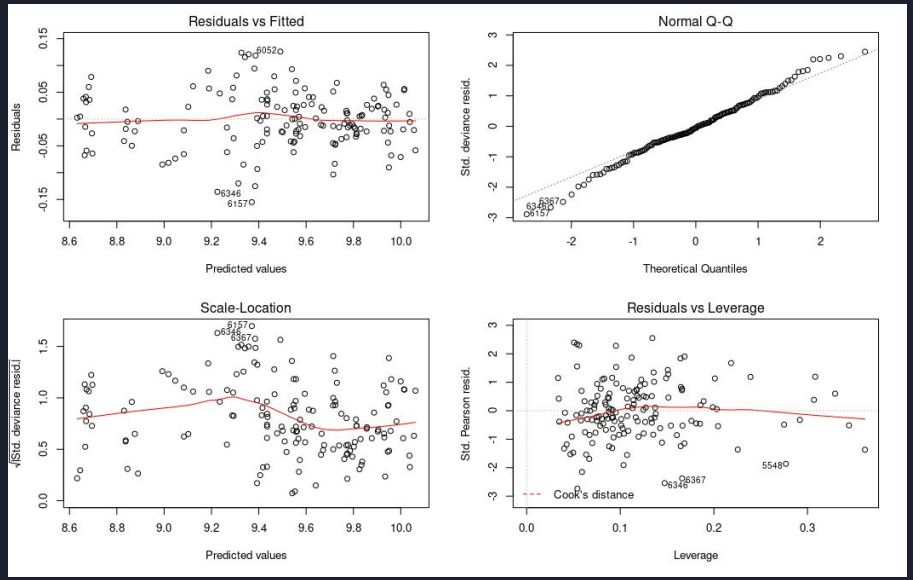
The model



Tamponi

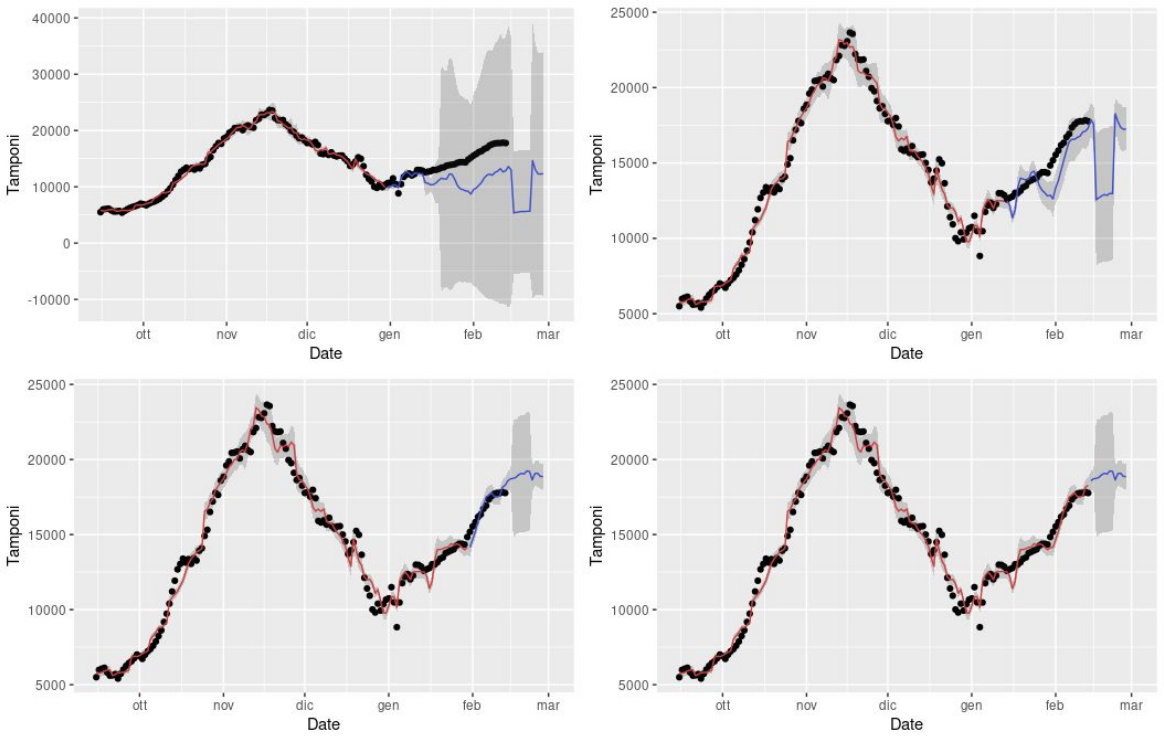
Summary and diagnostic plot

```
Coefficients:
                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                             1.118e+01  3.511e-01  31.833  < 2e-16 ***
I(tamp_rate^2)                         -1.224e+01  2.713e+00  -4.513 1.39e-05 ***
terapia_intensiva                       7.947e-03  9.878e-04   8.045 4.07e-13 ***
I(rt^2)                                 3.156e+00  4.082e-01   7.731 2.26e-12 ***
isolamento_domiciliare                 -9.333e-06  1.688e-06  -5.528 1.63e-07 ***
colorOrange:rt                         -6.385e+00  7.255e-01  -8.801 5.98e-15 ***
colorRed:rt                            -5.436e+00  6.758e-01  -8.043 4.10e-13 ***
colorWhite:rt                          -5.959e+00  7.595e-01  -7.846 1.21e-12 ***
colorYellow:rt                         -4.982e+00  7.240e-01  -6.881 2.07e-10 ***
rt:nuovi_positivi                      -3.560e-04  1.301e-04  -2.736 0.007058 **
tamp_rate:colorOrange:rt                1.060e+01  1.255e+00   8.445 4.42e-14 ***
tamp_rate:colorRed:rt                   4.701e+00  1.483e+00   3.170 0.001889 **
tamp_rate:colorWhite:rt                 4.495e+00  8.637e-01   5.204 7.14e-07 ***
tamp_rate:colorYellow:rt                2.164e+00  1.265e+00   1.710 0.089560 .
tamp_rate:colorOrange:casi_testati      5.068e-04  1.382e-04   3.666 0.000354 ***
tamp_rate:colorRed:casi_testati         2.981e-04  1.172e-04   2.545 0.012073 *
tamp_rate:colorWhite:casi_testati       4.440e-04  1.304e-04   3.405 0.000874 ***
tamp_rate:colorYellow:casi_testati      3.414e-04  1.163e-04   2.936 0.003912 **
---
```



Tamponi

Robustness towards time



Third model:

- 14 - days predictions for the number of tested cases and the rate of positive cases.
- Product between the two predictions to evaluate “nuovi_positivi”

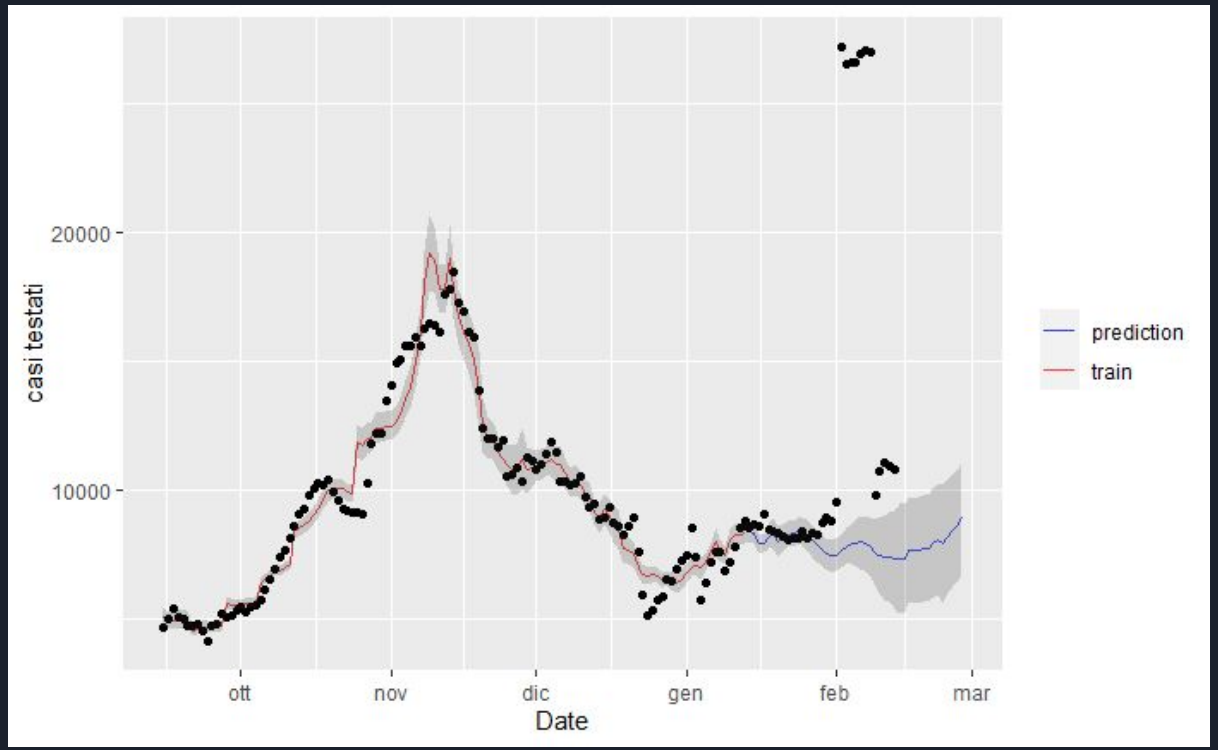
The model:

The prediction for “nuovi_positivi” is obtained by the product of the 14 days prediction on:

- Number of tested cases
- Positive rate of tests

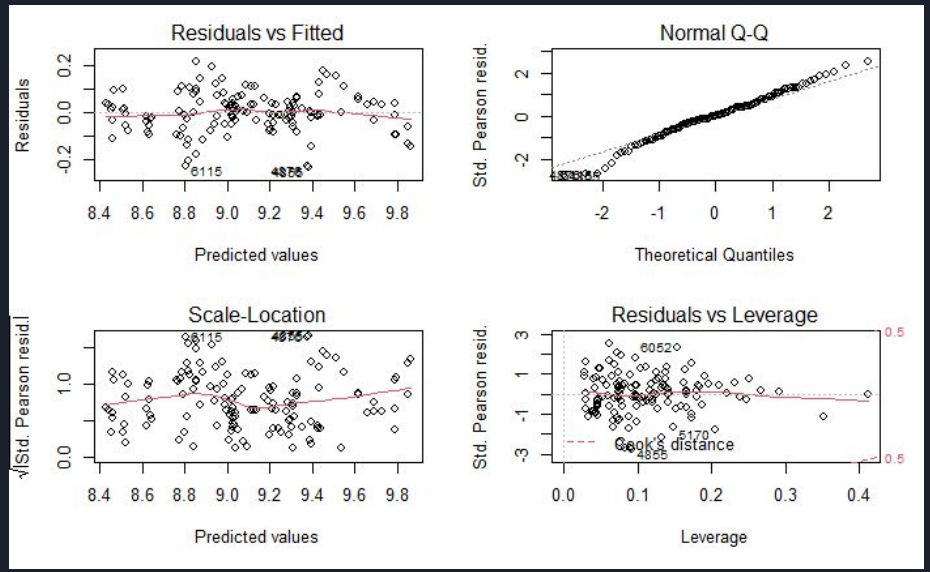
Booth the models are GLM with gamma distribution and log as link function.

Casi Testati

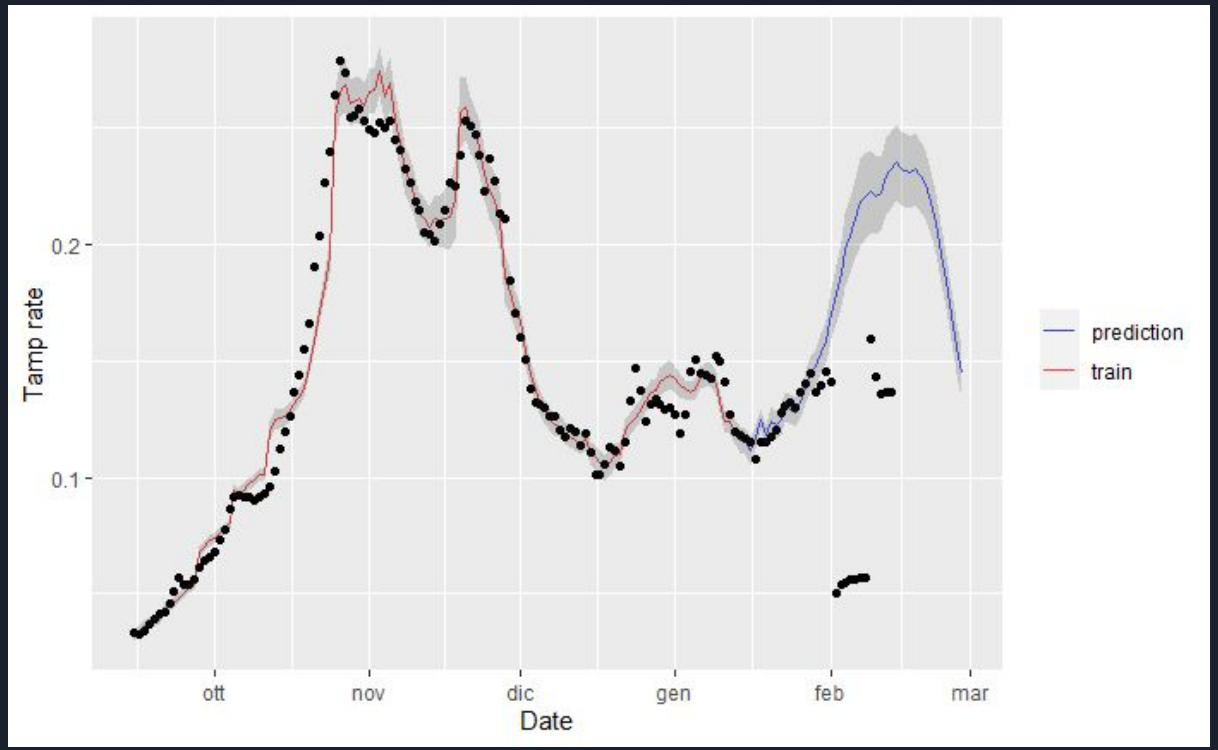


Casi Testati

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             11.8676318  0.3763384  31.534  < 2e-16 ***
ricoverati_con_sintomi  -0.0008686  0.0002186  -3.973 0.000121 ***
terapia_intensiva        0.0095947  0.0016426   5.841 4.53e-08 ***
rt:colorOrange          -7.1953795  1.0484772  -6.863 3.17e-10 ***
rt:colorRed             -8.4494043  1.5841115  -5.334 4.60e-07 ***
rt:colorWhite           -7.6420146  0.7459168 -10.245  < 2e-16 ***
rt:colorYellow          -4.7066147  1.4482243  -3.250 0.001498 **
colorOrange:I(rt^2)      4.0580379  0.6605778   6.143 1.09e-08 ***
colorRed:I(rt^2)         5.9443558  1.5873687   3.745 0.000279 ***
colorWhite:I(rt^2)       4.1000572  0.3625158  11.310  < 2e-16 ***
colorYellow:I(rt^2)      2.1242219  1.1262616   1.886 0.061701 .
colorOrange:tamp_rate    3.3189307  1.3074041   2.539 0.012411 *
colorRed:tamp_rate       1.8574187  0.6765002   2.746 0.006969 **
colorWhite:tamp_rate     3.7281026  0.5970125   6.245 6.67e-09 ***
colorYellow:tamp_rate   -0.2990992  0.9728111  -0.307 0.759027
```

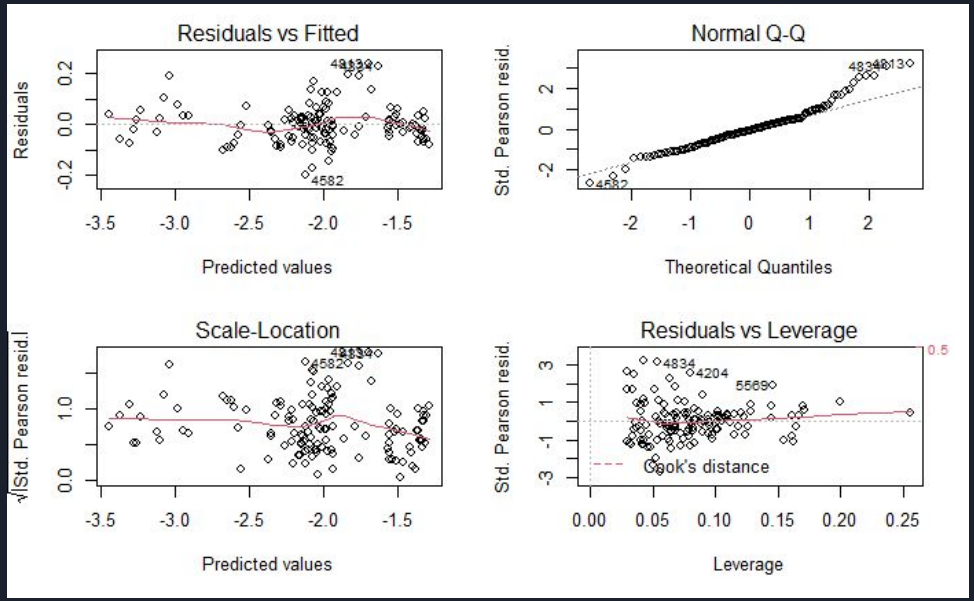


Test rate

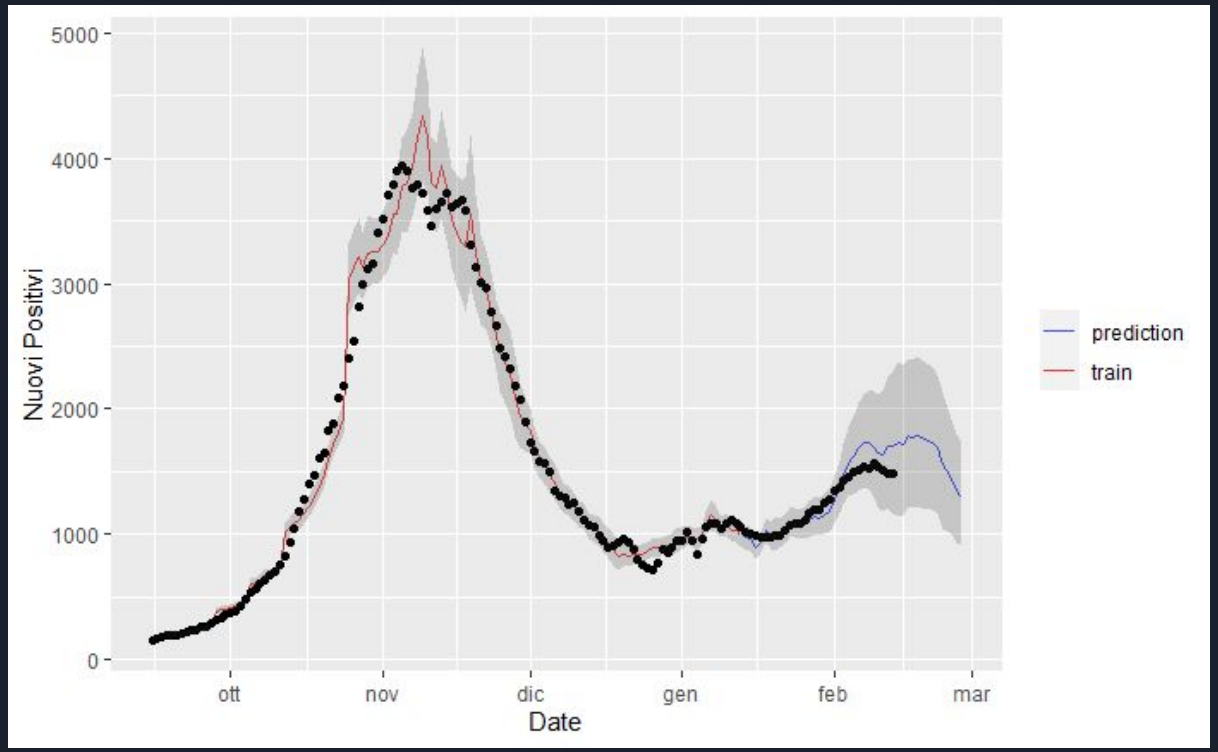


Test rate

```
Coefficients:
                                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                                   -9.037e-01  3.635e-01  -2.486   0.0142 *
poly(nuovi_positivi, 2, raw = TRUE)1          -6.898e-04  9.655e-05  -7.144 6.78e-11
***
poly(nuovi_positivi, 2, raw = TRUE)2           1.220e-07  1.618e-08   7.544 8.44e-12
***
poly(ricoverati_con_sintomi, 2, raw = TRUE)1   4.182e-03  2.885e-04  14.496  < 2e-16
***
poly(ricoverati_con_sintomi, 2, raw = TRUE)2  -7.355e-07  5.720e-08 -12.858  < 2e-16
***
I(rt^2)                                        4.058e+00  3.876e-01  10.470  < 2e-16
***
isolamento_domiciliare                        -2.818e-05  2.941e-06  -9.582  < 2e-16
***
rt:colorOrange                                -7.040e+00  7.650e-01  -9.203 1.07e-15
***
rt:colorRed                                   -7.077e+00  7.796e-01  -9.078 2.13e-15
***
rt:colorWhite                                 -7.117e+00  7.532e-01  -9.449 2.72e-16
***
rt:colorYellow                                -6.969e+00  7.608e-01  -9.159 1.36e-15
***
```



Model results



Models comparison

	RMSE	Maximum error
Model 1	390.8	741.3
Model 2	164.8	373.3
Model 3	146.4	220.1

Bibliography

[1] Protezione Civile. (2020). pcm-dpc/COVID-19. GitHub. https://github.com/pcm-dpc/COVID-19

[2] Harvard Dataverse. (2021, February 14). National and Subnational Estimates of the Covid 19 Reproduction Number (R) for Italy Based on Test Results. https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/8DUSHZ

[3] Apple. (2020). COVID-19 - Mobility Trends Reports. https://covid19.apple.com/mobility

[4] Ansa. (2020). ANSA.it. ANSA.It. https://www.ansa.it/