Effect of COVID-19 over FTSE-MIB financial indicator

Francesco Maria Calistroni
Lorenzo Tausani

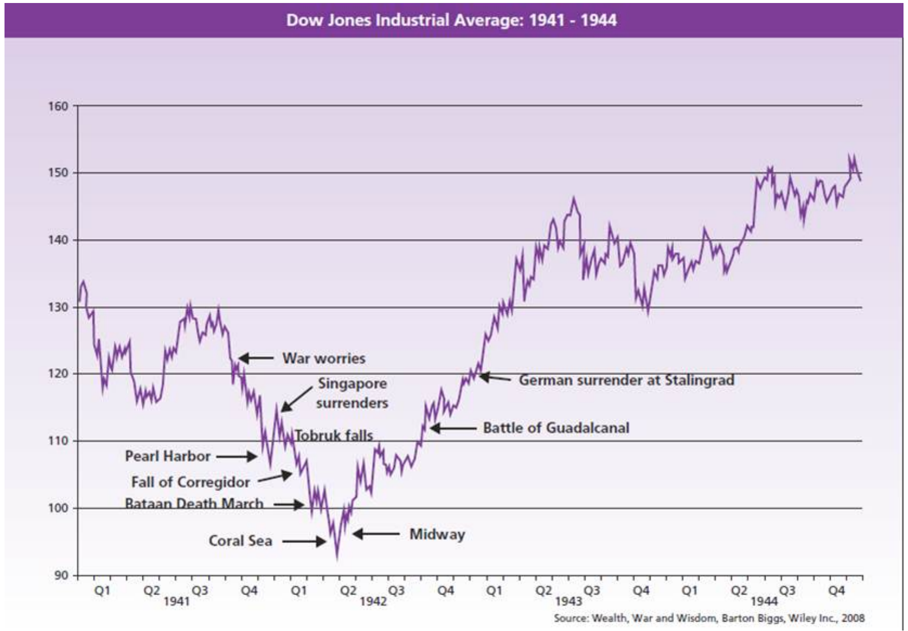
Business, economic and financial data – Final project
Master degree in Data Science
2022/01/18

DATA SCIENCE
UNIVERSITY OF PADOVA

Introduction

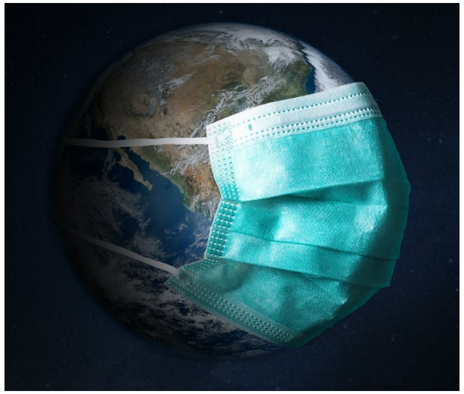
- It is well known that natural and artificial disasters can affect severily economic and financial indicators of countries and communities involved

E.g. World War 2



COVID-19: a contemporary catastrophy

November 2019



20 february 2020

Death toll (updated at 29 december 2021)
World: 5,41 Mln
Italy: 137000

→In this work we will analyze only the first wave of the covid pandemic in Italy (20 february 2020 – 03 june 2020)

For reference…

Total military deaths (WW1): ~ 10 Mln

Civil victims in Italy (WW2): 153147

FTSE-MIB

- ***FTSE-MIB*** → ***F**inancial **T**imes **S**tock **E**xchange **M**ilano **I**ndice di **B**orsa*
- (weighted) average stock price of the 40 biggest companies listed in the italian stock exchange
- Measure of the general trend of the italian economy

Aim of the work

- Understand the relationship between pandemic-related indicators and FTSE MIB daily open prices (***Apertura*** in short)
- Predict *Apertura* using pandemic-related predictors in a way useful for economic and governmental organizations

↓

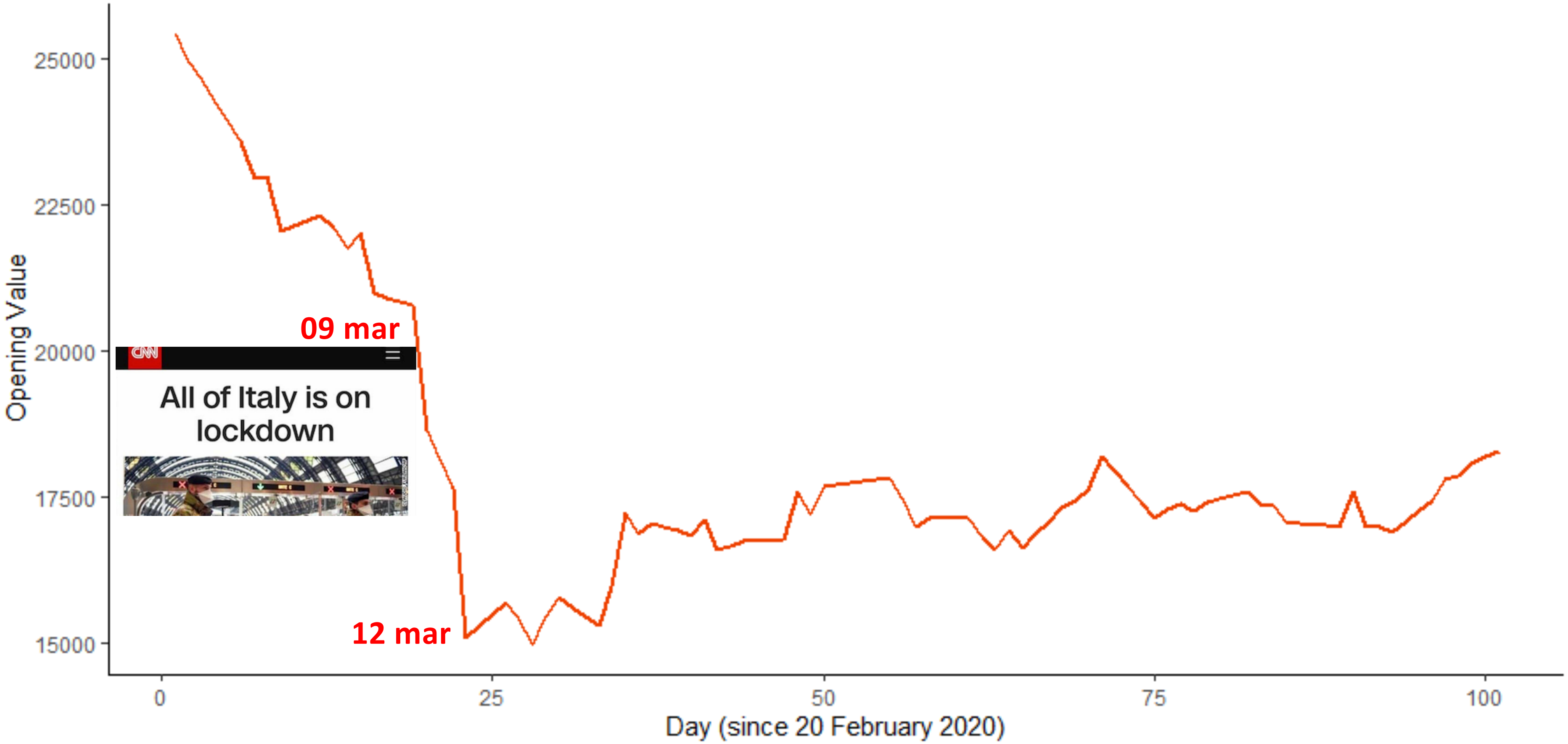
Knowing in advance the market response to the pandemic will lead governmental and economic organizations to take the better decisions w.r.t. their scopes!

Dataset structure

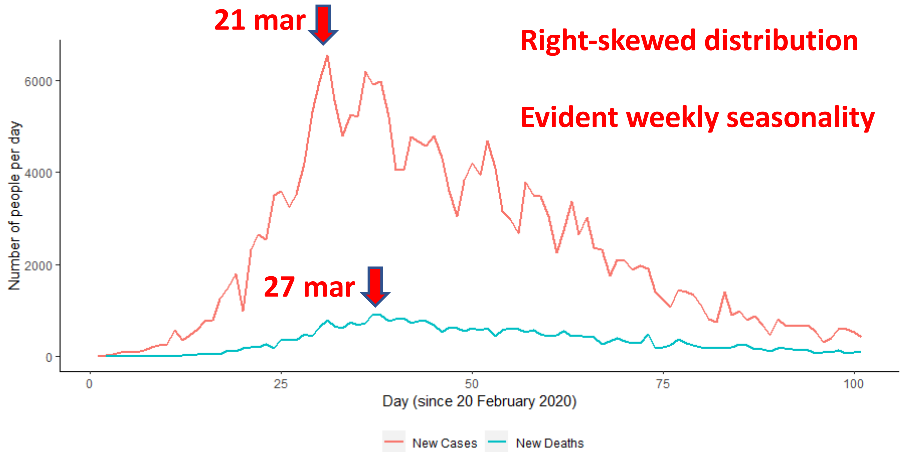
All timeseries are from 20 february 2020 to 03 june 2020 → 105 days

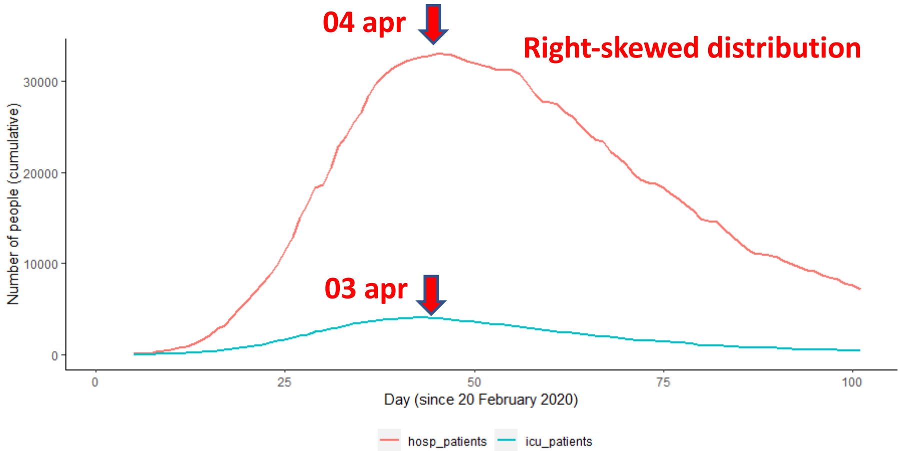
- ***Response variable (@Investing.com):*** FTSE MIB open price (Apertura) (interpolated for WEs)
- **Predictors (@OWID):**
  - **Covid19-pandemic indicators:** New cases, New deaths, reproduction rate, icu patients, hospitalized patients, new tests, positive rate
  - **Governnment measures:** containment index
  - **Google mobility measures:** retail_and_recreation, grocery_and_pharmacy, residential, transit_stations, parks, workplaces

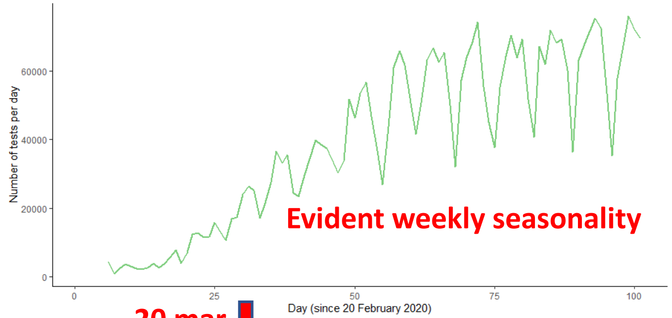
Exploratory data analysis (EDA) - Apertura

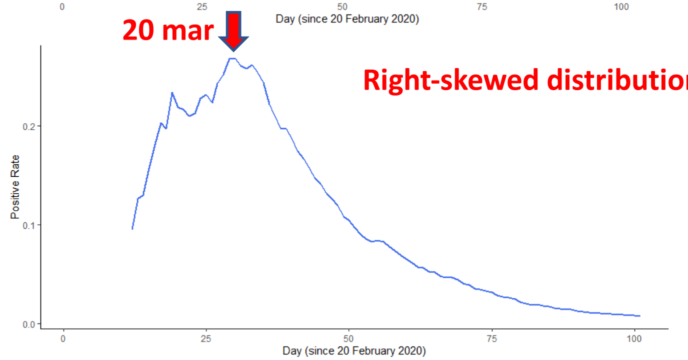


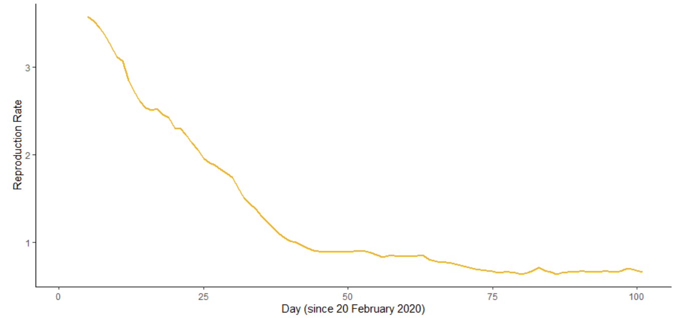
EDA – Pandemic indicators



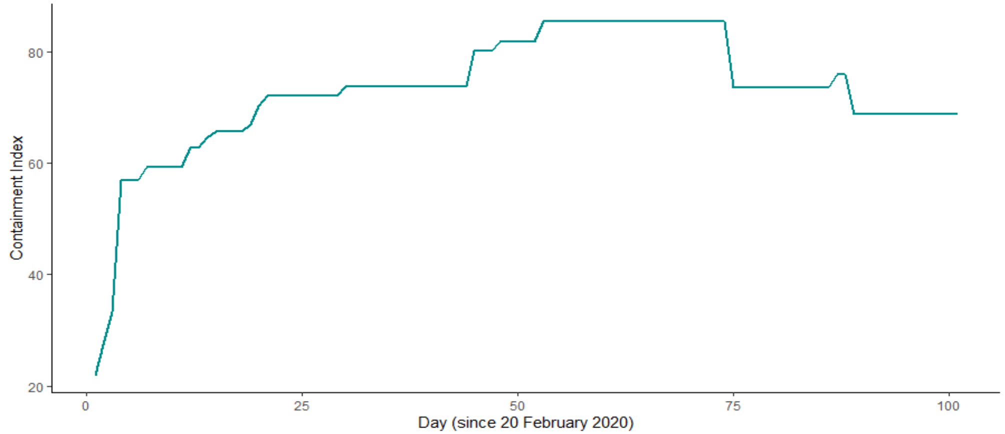






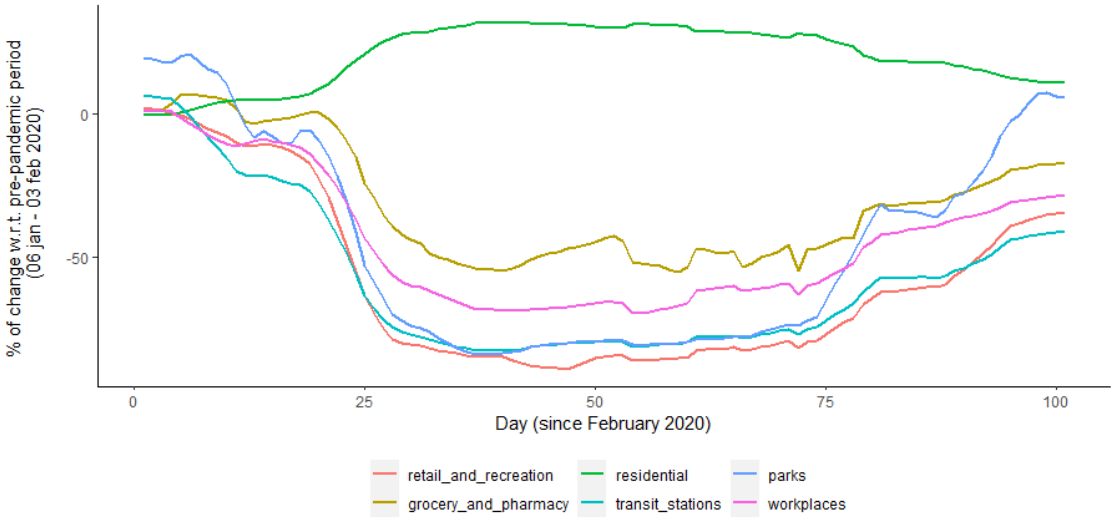


EDA – Government and mobility measures



Containment index (@Oxford university)

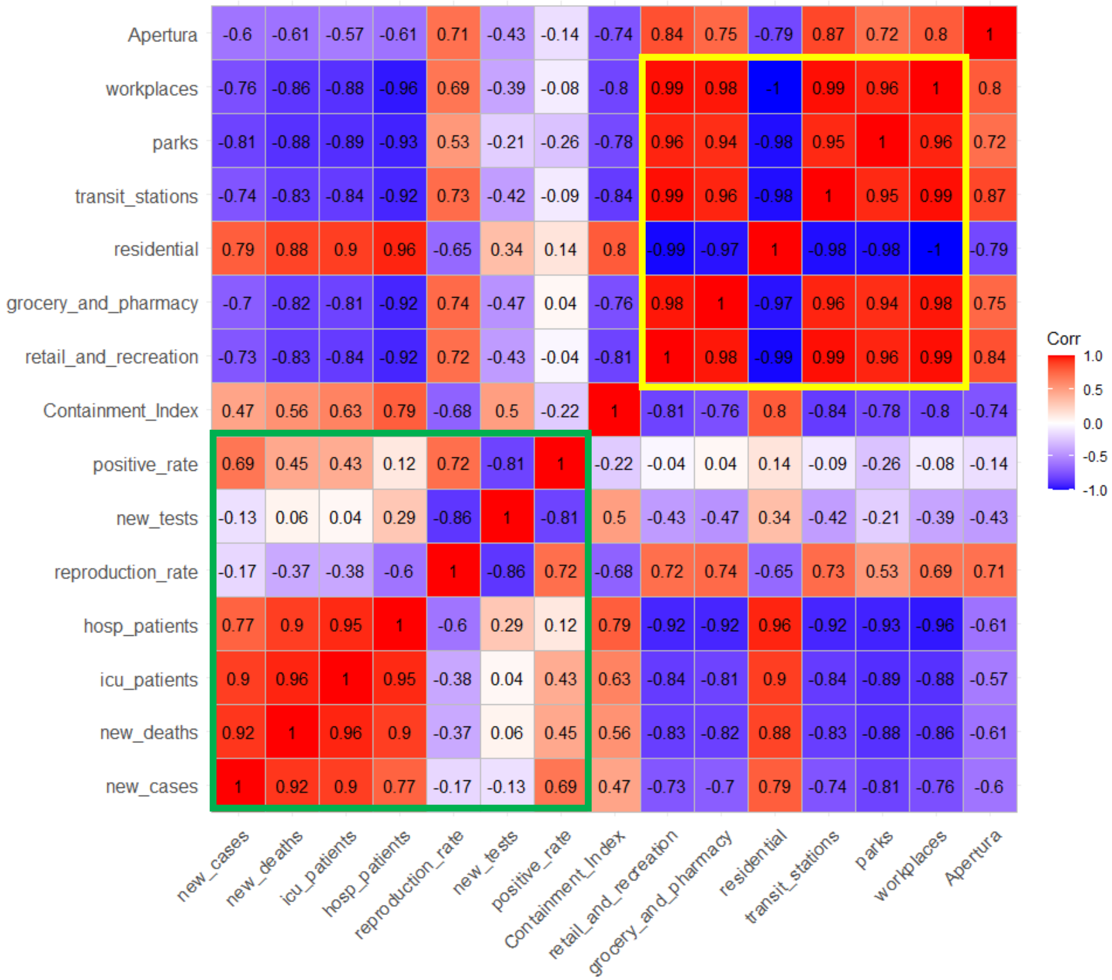
- Mean of 14 categorical variables
- Summarizes all the measures implemented by the italian government to avoid the spreading of the pandemic (e.g movement restrictions, wearing masks, school closing, etc..)



Google mobility measures

- Google mobility reports
- Shows how people movements have changed during the pandemic
- Measuring number of visitor to specific categories of locations

EDA: Correlation analysis



- High correlation between predictors
  - ➢ **Need to counter collinearity problems**
  - ➢ **Variable selection (next…)**

Forecast

Modelling approach

- 4 different models:
  - Linear model → simple, interpretable
  - GAM → generalization of linear model, models non-linearity
    - Nonlinearity=splines (df=1:4); stepwise-selection
  - ARIMA → classic approach to financial timeseries, though less easy to interpret
    - Auto-arima
  - ARIMAX → generalization of ARIMA, external regressors (i.e. the same used for linear regr.)

- Train-Test splitting → 2 splittings
  → Train:
  1. **2020-02-20** to **2020-04-19** → *hard* lockdown
  2. **2020-02-20** to **2020-05-20** → *relaxed* lockdown

  → Testing (2 weeks period):
  1. **2020-04-20** to **2020-05-04**
  2. **2020-05-21** to **2020-06-03**

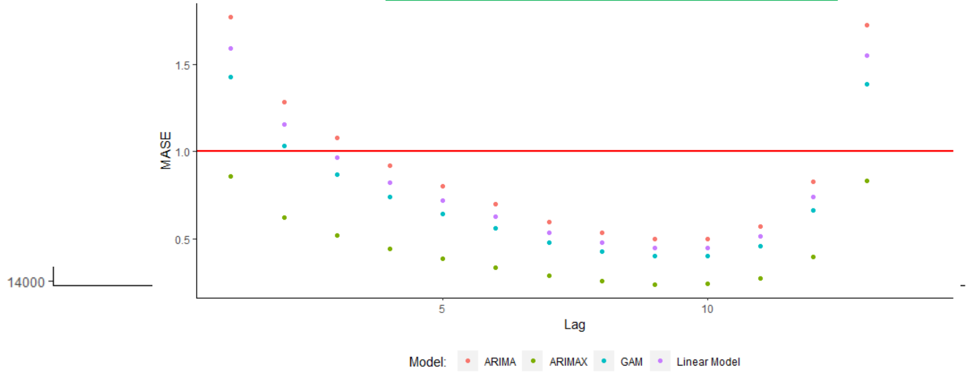
Performance measures

- **MAPE** → Forecasting accuracy measure independent from the scale of the data

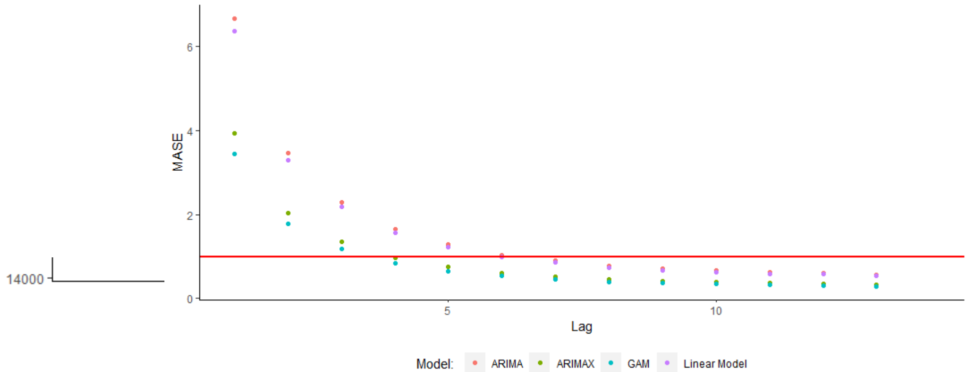
- **MASE** → Mean Absolute Scaled Error: Comparing models' predictions with predictions of ***Naive forecasting***
  - ***Naive forecasting:*** the last period's sales are used for the next period's forecast (e.g. Naïve forecasting, Lag=1: tomorrow Apertura will be the same as today)

Forecast: all predictors

	ARIMA		ARIMAX	GAM	Linear Model
MAPE	2.71	>	1.33	2.18	2.43

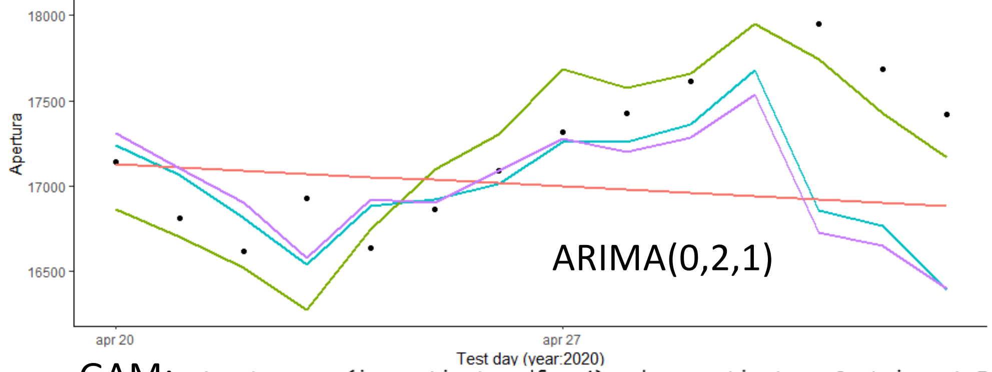


	ARIMA		ARIMAX	GAM	Linear Model
MAPE	5.24	>	3.22	2.82	5.16



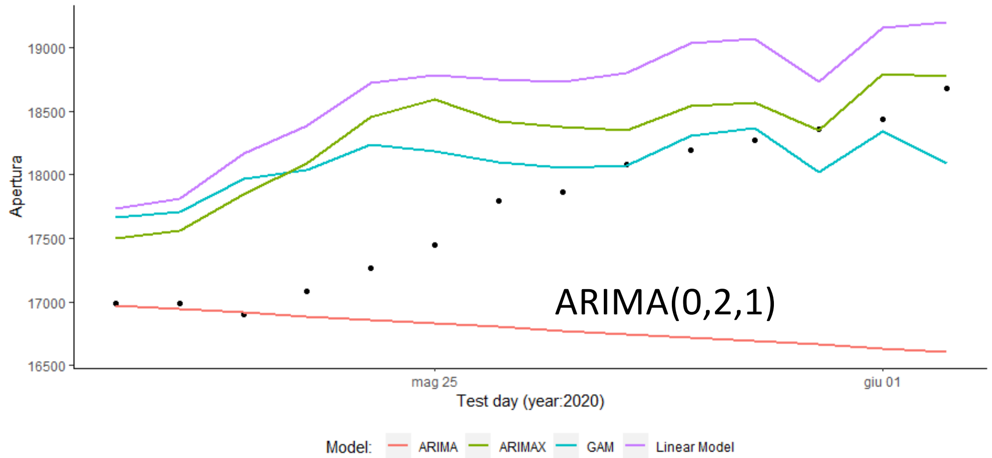
GAM: `Apertura ~ new_cases + new_tests + hosp_patients + positive_rate + Containment_Index + retail_and_recreation + residential + parks + workplaces + transit_stations`

mean(VIF)=989



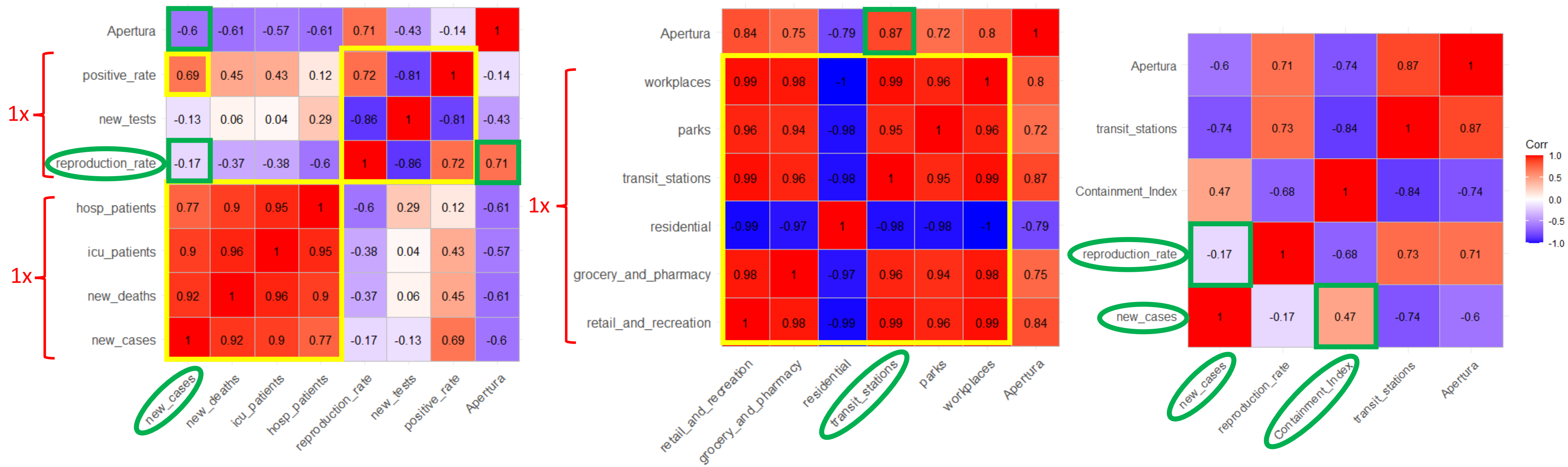
GAM: `Apertura ~ s(icu_patients, df = 4) + hosp_patients + Containment_Index + retail_and_recreation + grocery_and_pharmacy + residential + parks + workplaces + transit_stations`

mean(VIF)=436



Collinearity problem

- By examining the correlation plots and the VIF of linear models and GAMs it emerges that the full model has a serious collinearity problem, which in turn points to low interpretability of model results
    - ➢ **Variable selection!**

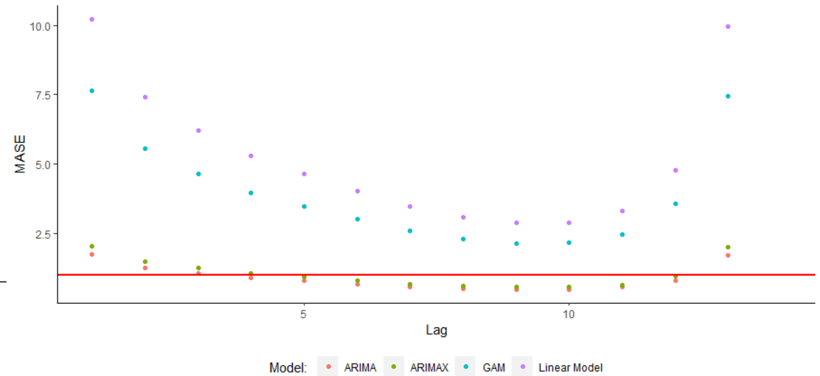


Interpretable models

- **Model 1: Pandemic indicators only →** *Reproduction rate, new cases*
- **Model 2: Containment index only**
- **Model 3: Transit_station only**
- **Model 4: Containment index + new_cases**

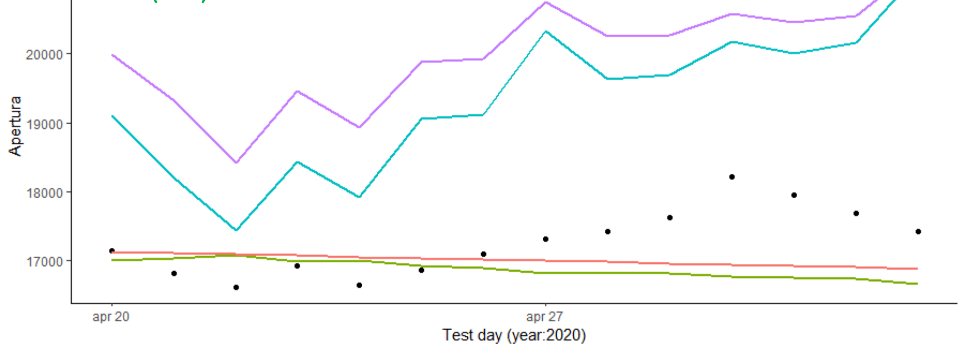
Model 1 - Pandemic indicators only

	ARIMA		ARIMAX	GAM	Linear Model
MAPE	2.71	<	3.15	11.83	15.85

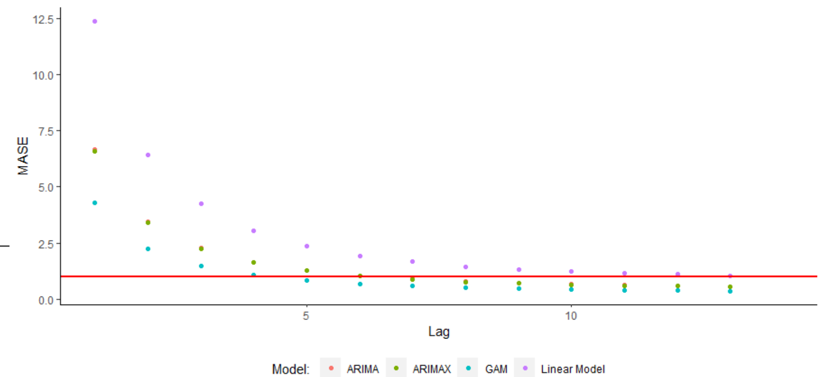


GAM: `Apertura ~ s(new_cases, df = 4) + s(reproduction_rate, df = 4)`

mean(VIF)=1.27

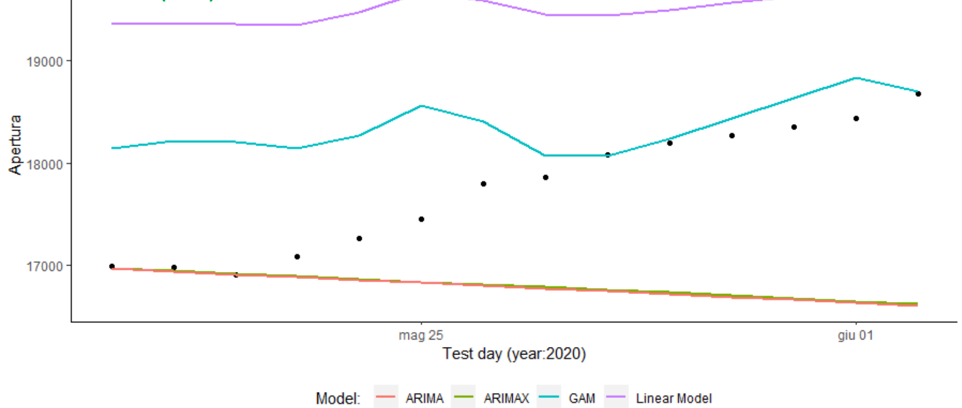


	ARIMA		ARIMAX	GAM	Linear Model
MAPE	5.24	≈	5.17	3.56	10.07



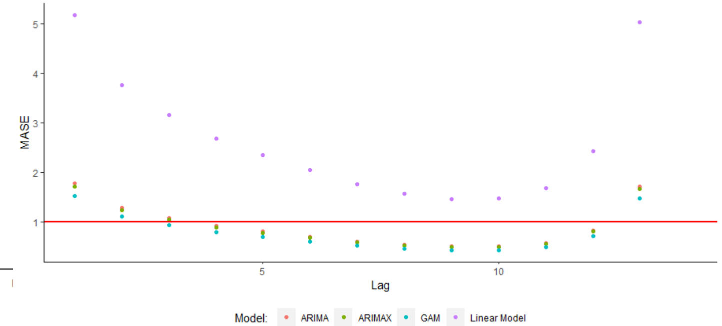
GAM: `Apertura ~ s(new_cases, df = 4) + s(reproduction_rate, df = 4)`

mean(VIF)=1.04

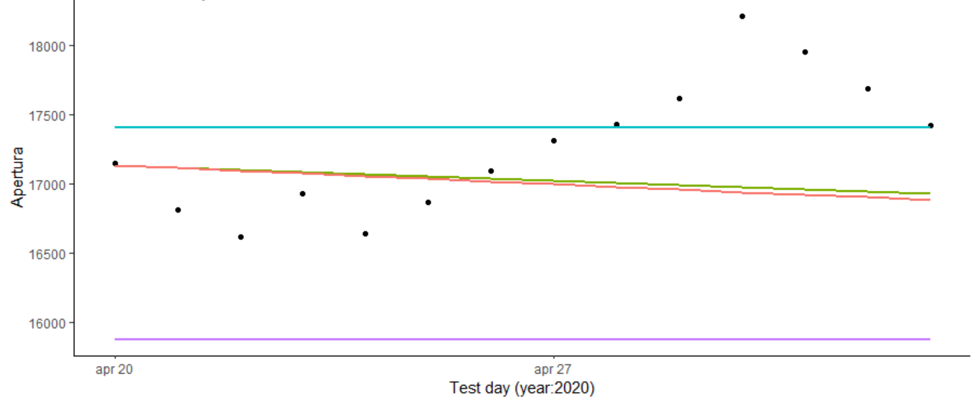


Model 2- Containment index only

	ARIMA		ARIMAX	GAM	Linear Model
MAPE	2.71	≈	2.62	2.38	7.97

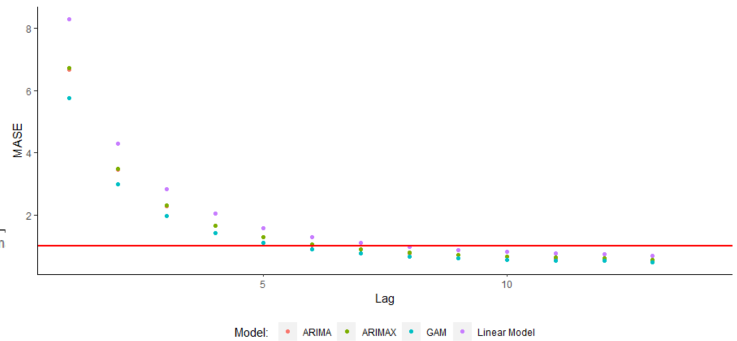


GAM: `Apertura ~ s(Containment_Index, df = 4)`

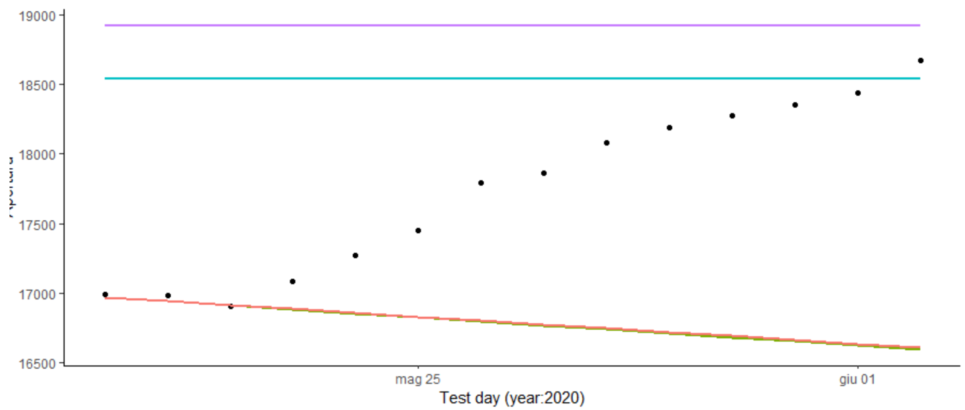


Model: ARIMA ARIMAX GAM Linear Model

	ARIMA		ARIMAX	GAM	Linear Model
MAPE	5.24	≈	5.28	4.75	6.8

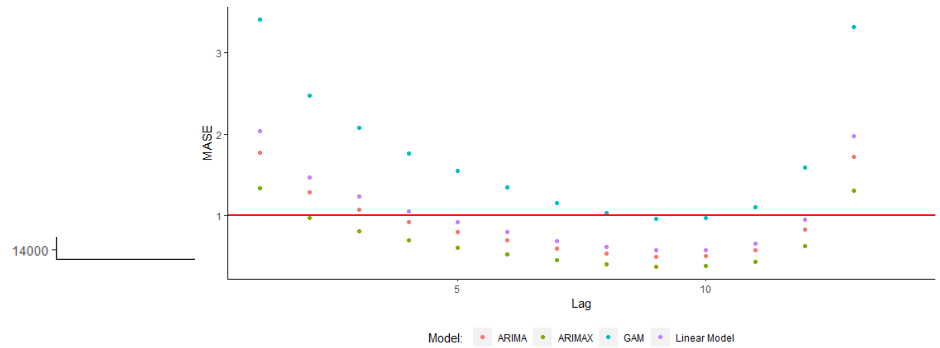


GAM: `Apertura ~ s(Containment_Index, df = 4)`

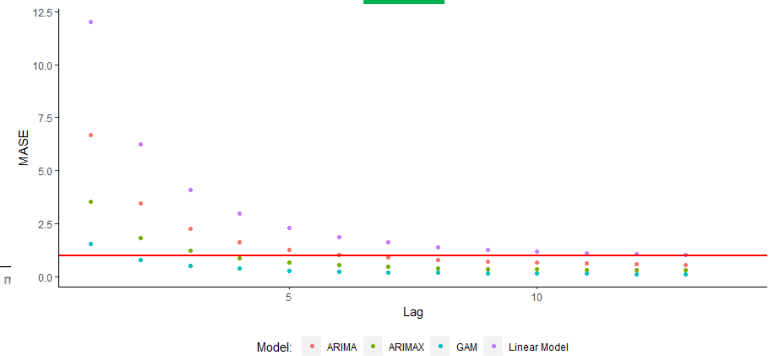


Model 3- Transport index only

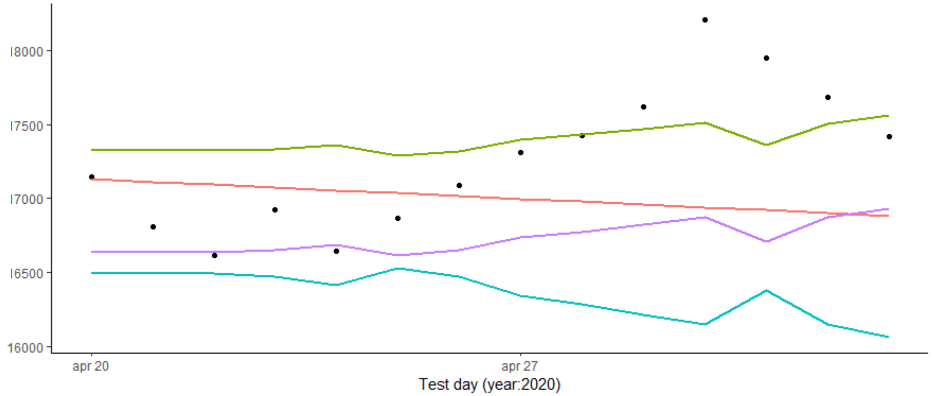
	ARIMA		ARIMAX	GAM	Linear Model
MAPE	2.71	>	2.09	5.2	3.1



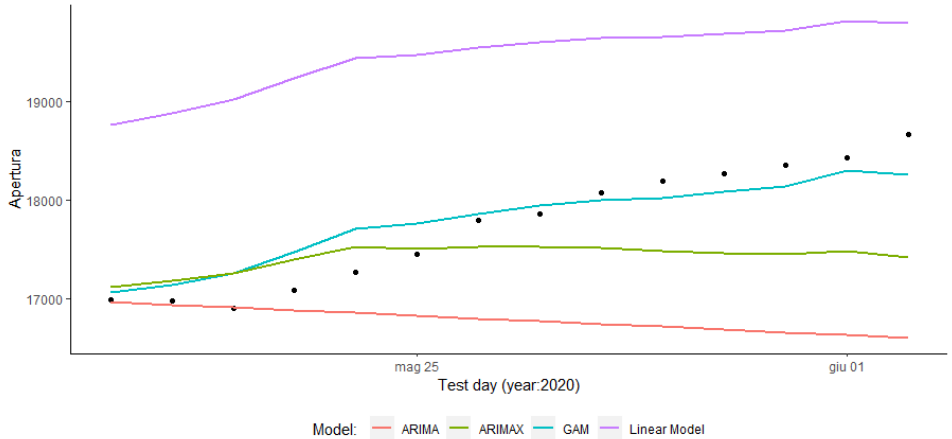
	ARIMA		ARIMAX	GAM	Linear Model
MAPE	5.24	>	2.81	1.25	9.74



GAM: `Apertura ~ s(transit_stations, df = 4)`

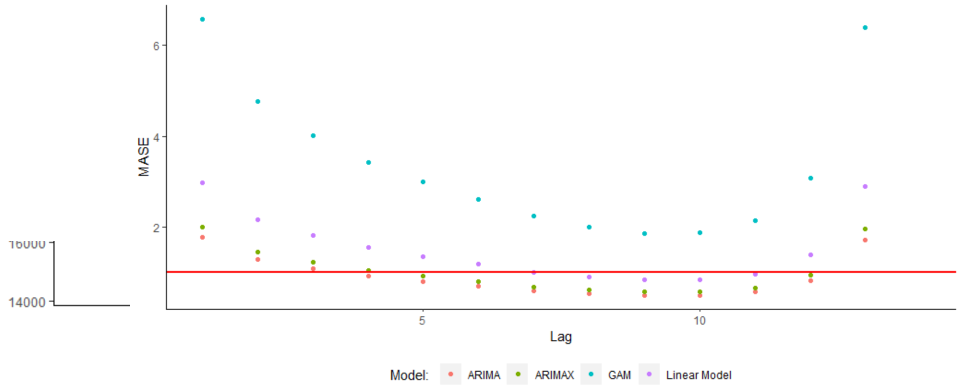


GAM: `Apertura ~ s(transit_stations, df = 4)`

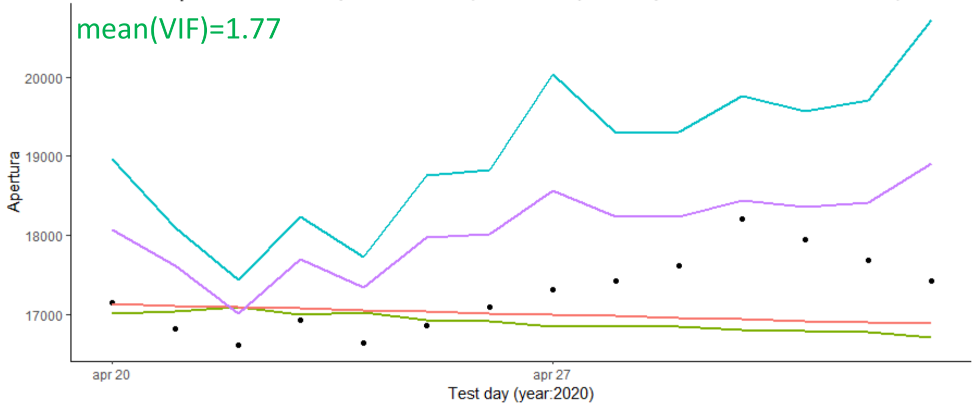


Model 4 - Containment index + new_cases

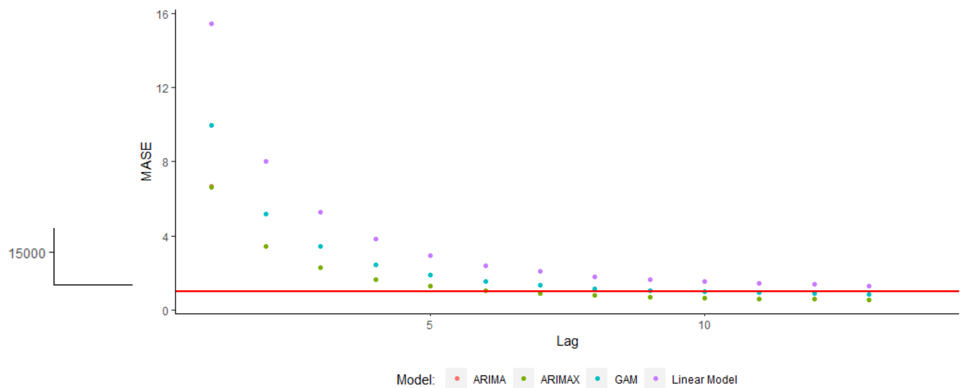
	ARIMA		ARIMAX	GAM	Linear Model
MAPE	2.71	<	3.07	10.19	4.64



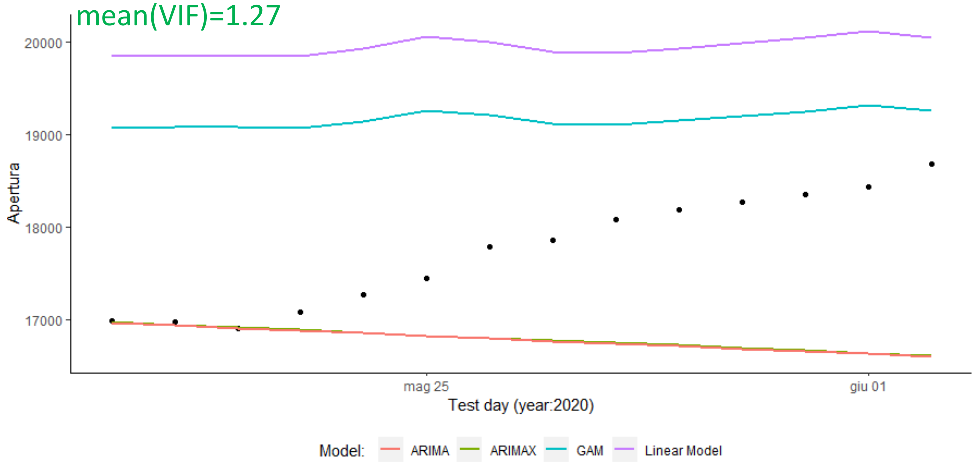
GAM: `Apertura ~ s(new_cases, df = 4) + s(Containment_Index, df = 4)`



	ARIMA		ARIMAX	GAM	Linear Model
MAPE	5.24	≈	5.21	8.17	12.55



GAM: `Apertura ~ s(new_cases, df = 4) + s(Containment_Index, df = 4)`



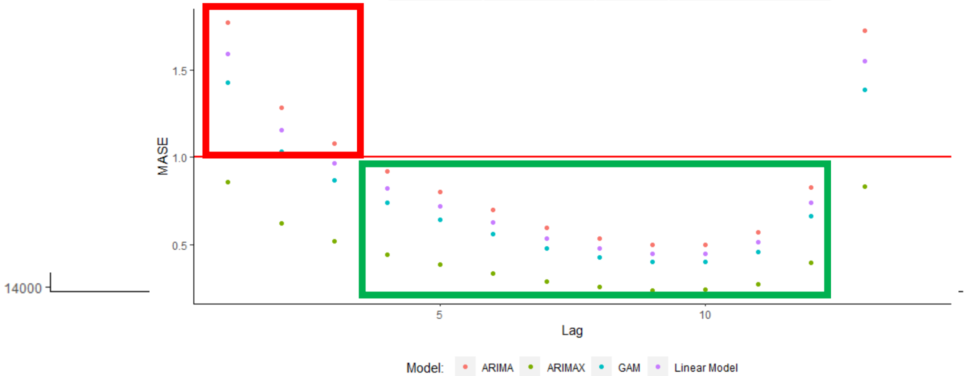
But… utility of these predictions?

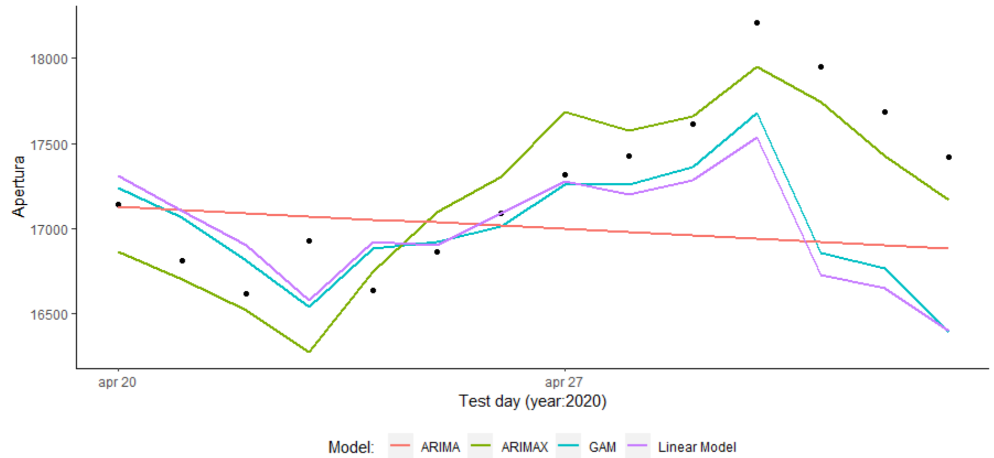
- Many models perform worse than naive forecasting at LAG=1 or close to 1, but outperform naive forecasts at higher time lags
  - ➢ **PROBLEM: in order to have business value, models need to outperform naive predictions**
  - ➢ **Our independent variables data are taken day by day, so a model that performs worse than naive forcasting at Lag=1 is useless**

➢ ***IDEA: using predictor's forecasts as input to our forecasting models, outperforming naive forecasts at higher timelags!***

E.g. Model with all predictors, *hard lockdown*

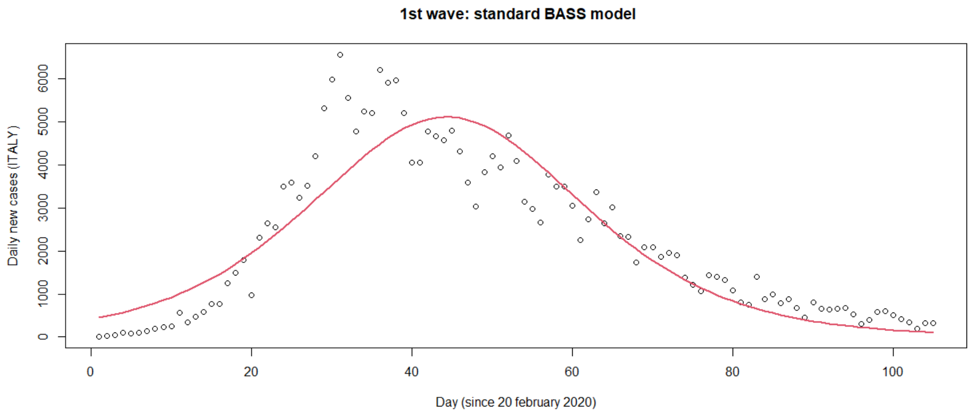
	ARIMA	ARIMAX	GAM	Linear Model
MAPE	2.71	1.33	2.18	2.43

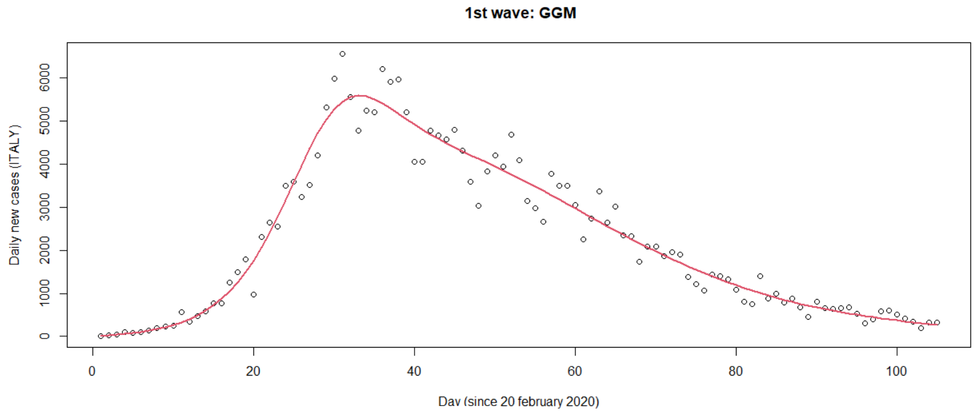


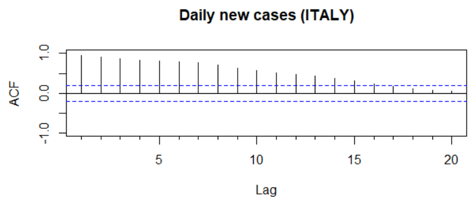


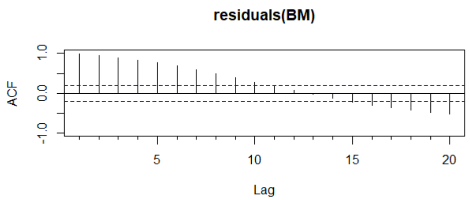
Bass_based forecasts - intro

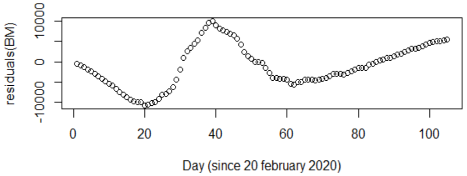
- Pandemic spreading can be modelled using nonlinear regression models for new product growth (i.e. **Bass models**):



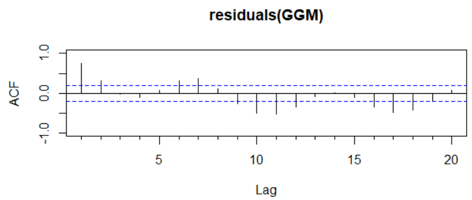








Harmonic behavior of residuals
(DW test = 0.02)
→ Positive autocorrelation



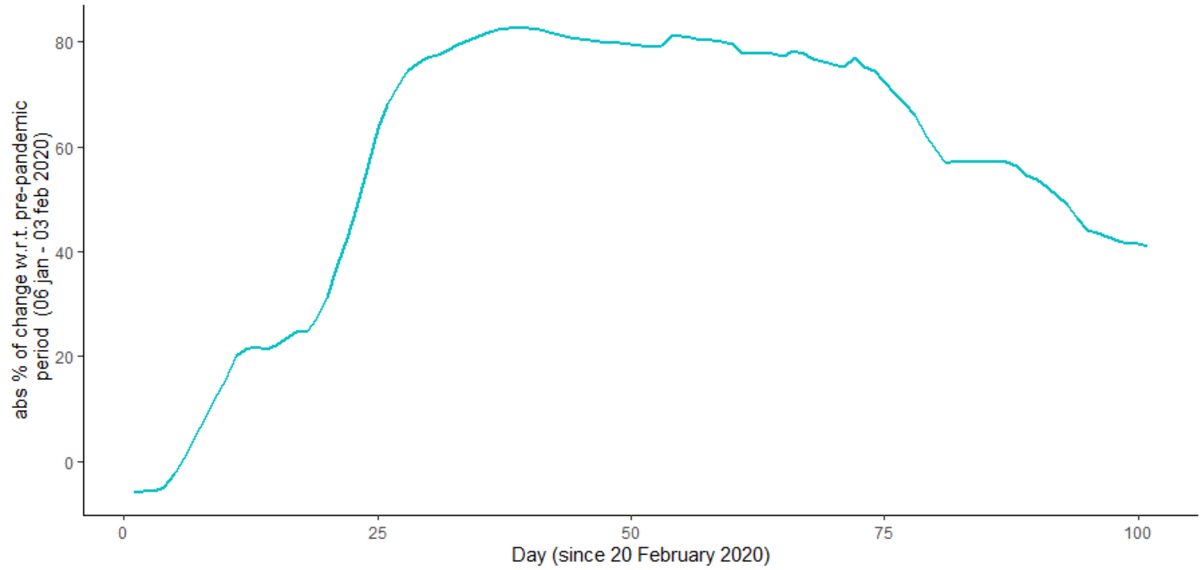
- Still tendency towards positive autocorrelation (DW test =0.51)
- Improvement w.r.t. standard bass model ($\tilde{R}^2 = 0.98$)

```
Coefficients:
          Estimate     Std.Error         Lower         Upper    p-value
k     2.377382e+05 2.291523e+02 2.372891e+05 2.381874e+05 2.00e-203 ***
pc    1.749580e-04 1.361290e-05 1.482772e-04 2.016388e-04  6.66e-23 ***
qc    2.429567e-01 3.655420e-03 2.357922e-01 2.501212e-01  1.45e-84 ***
ps    4.217857e-03 6.179420e-05 4.096743e-03 4.338971e-03  1.07e-85 ***
qs    6.009159e-02 4.699943e-04 5.917041e-02 6.101276e-02 1.25e-112 ***
```

What about transit mobility?

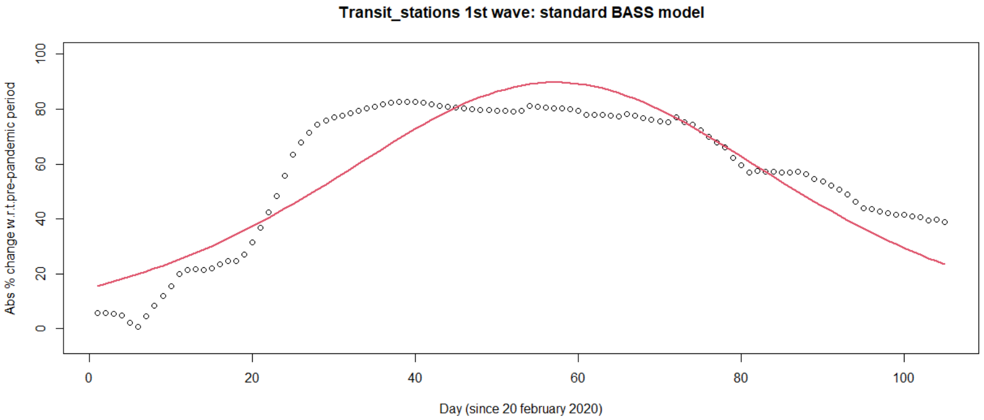
- Bass-type diffusion models assume the presence of four distinct phases
    1. Introduction
    2. Growth
    3. Maturity
    4. Decline

➢ The variable *transit_station* seems indeed to follow a similar dynamic

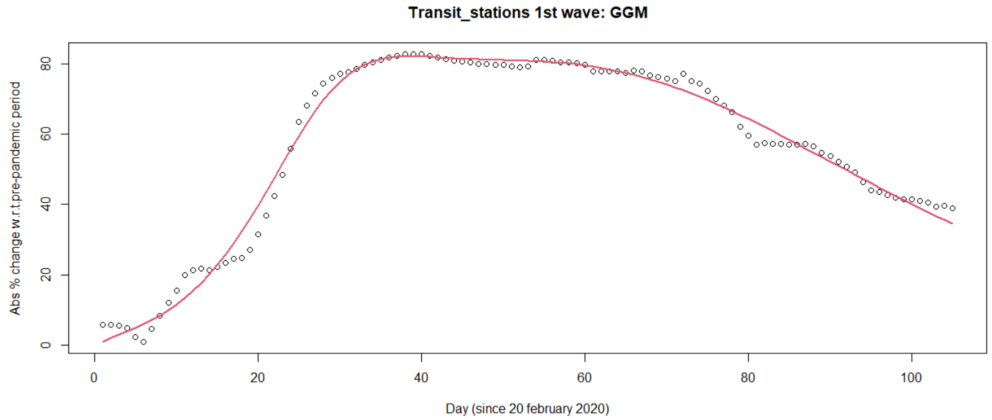


Parameter	Classic interpretation	Transit mobility interpretation
m	Market potential	Maximal alteration of people's movement causable by the pandemic
p	Innovation	First people that stop moving
q	Imitation	Speed at which people reduce their moving habits

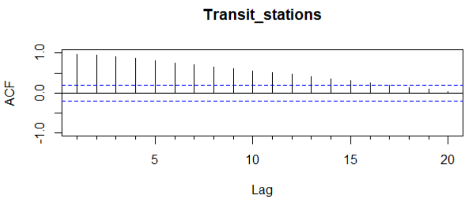
Transit_station: Bass modelling

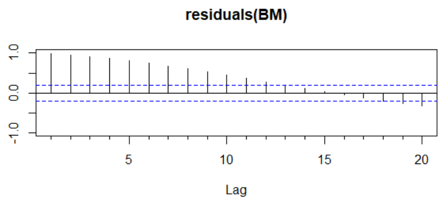



```
Coefficients:
     Estimate     Std.Error        Lower        Upper    p-value
m 6.344098e+03 6.036919e+01 6.225776e+03 6.462419e+03 8.61e-106 ***
p 2.404535e-03 7.906508e-05 2.249570e-03 2.559500e-03  5.88e-53 ***
q 5.166644e-02 1.231091e-03 4.925355e-02 5.407934e-02  3.65e-66 ***
```

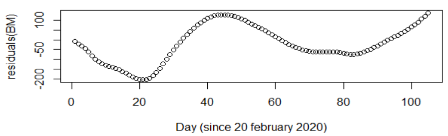


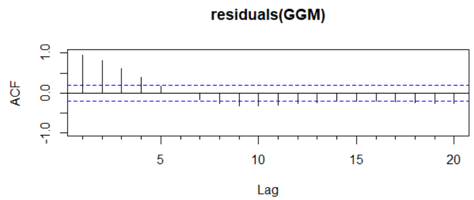

```
Coefficients:
      Estimate    Std.Error        Lower        Upper   p-value
k  6.953503e+03 1.790166e+01 6.918417e+03 6.988590e+03 9.03e-161 ***
pc 1.629010e-03 8.029940e-05 1.471626e-03 1.786394e-03  3.18e-37 ***
qc 1.630872e-01 2.778266e-03 1.576419e-01 1.685325e-01  2.64e-79 ***
ps 3.523455e-03 2.693719e-05 3.470659e-03 3.576251e-03 1.30e-113 ***
qs 3.768495e-02 3.002584e-04 3.709646e-02 3.827345e-02 7.90e-112 ***
```





Harmonic behavior of residuals
(DW test = 0.01)
→ Positive autocorrelation



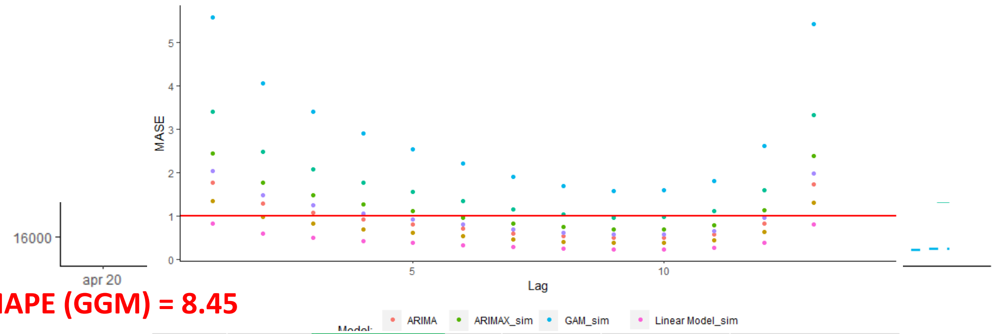


- Still tendency towards positive autocorrelation (DW test =0.09)
- Improvement w.r.t. standard bass model ($\tilde{R}^2 = 0.99$)

Bass_based forecasts - results

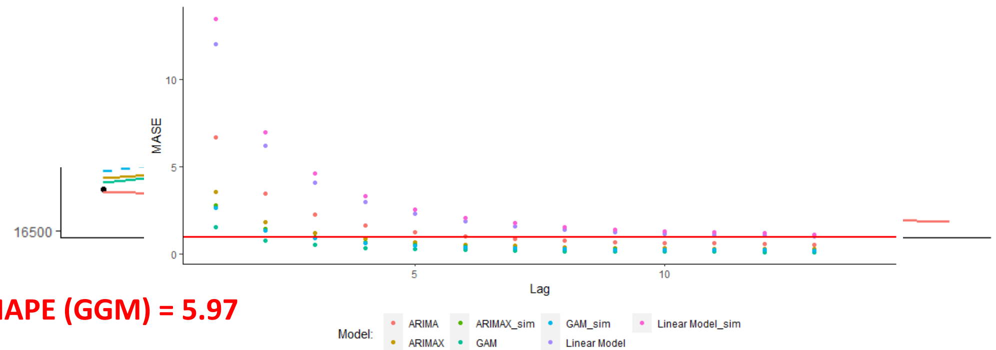
Transit_station only

	ARIMA	ARIMAX	ARIMAX_sim	GAM	GAM_sim	Linear Model	Linear Model_sim
MAPE	2.71	2.09	3.82	5.2	8.54	3.1	1.28



MAPE (GGM) = 8.45

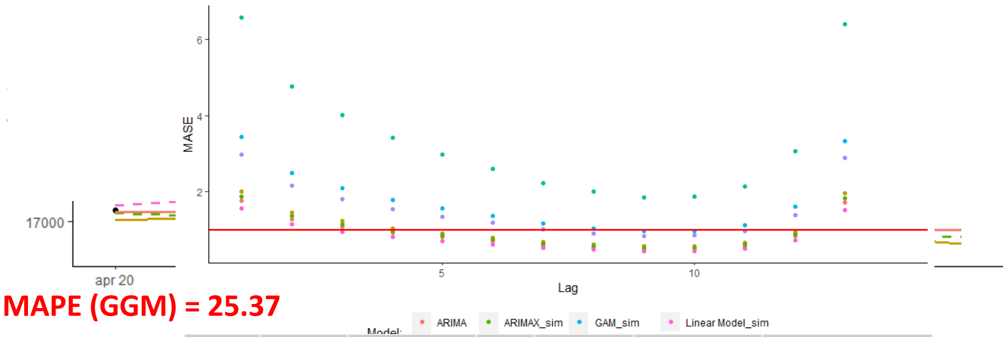
	ARIMA	ARIMAX	ARIMAX_sim	GAM	GAM_sim	Linear Model	Linear Model_sim
MAPE	5.24	2.81	2.25	1.25	2.15	9.74	10.87



MAPE (GGM) = 5.97

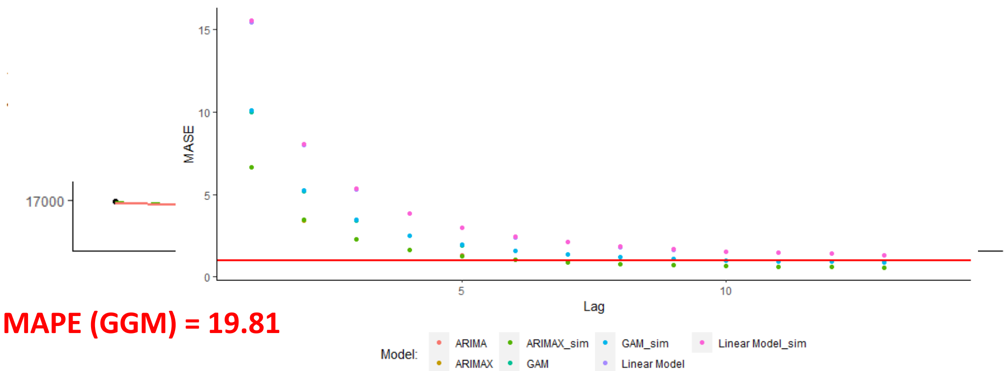
Containment index + new_cases

	ARIMA	ARIMAX	ARIMAX_sim	GAM	GAM_sim	Linear Model	Linear Model_sim
MAPE	2.71	3.07	2.87	10.19	5.35	4.64	2.46



MAPE (GGM) = 25.37

	ARIMA	ARIMAX	ARIMAX_sim	GAM	GAM_sim	Linear Model	Linear Model_sim
MAPE	5.24	5.21	5.21	8.17	8.25	12.55	12.65



MAPE (GGM) = 19.81

Conclusions

- In the 1st wave of covid pandemic in Italy, our models seem to yield good forecasting of Apertura
- Through variable selection we were able to implement different models with high interpretability
- Surprisingly, people's travel (i.e. *transit_station*) seems very predictive of Apertura
- Using bass-based forecasts for the variables *new cases* and *transit_stations* we are able to obtain forecasting results similar the ones obtained with true data

Future directions

- Study in a similar way the other pandemic waves, in order to check if also there similar dynamics unfold


Thanks for your attention

References

- Hannah Ritchie, Edouard Mathieu, Lucas Rodés-Guirao, Cameron Appel, Charlie Giattino, Esteban Ortiz-Ospina, Joe Hasell, Bobbie Macdonald, Diana Beltekian and Max Roser (2020) - "Coronavirus Pandemic (COVID-19)". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/coronavirus' [Online Resource]
- https://it.investing.com/indices/it-mib-40