

COVID19 application

Ana Xiangning Pereira Ezquerro

Versión 02 noviembre, 2022

Índice

1	COVID-19 evolution, influenza impact and vaccination progress	3
1.1	Study on global data	3
1.2	Study on vaccination data aggregated by group of ages	4

In this document we evaluate our selection method in the IRAS dataset, aiming to study the impact of aggregated data (by age and demographical variables) of acute respiratory infections (ARIs) and vaccination progress to the COVID19 evolution in Catalonia (Spain).

The original IRAS dataset contains individual weekly data about ARIs cases. In order to preserve confidentiality and privacy of the patients, this dataset was aggregated by week, filtered and parsed to obtain a new dataset with the following variables:

- `fecha [char]`: Time variable which indicates the week of the aggregated data (format YYYY-WW).
- `sdgripal [int]`: Total number of influenza cases in Catalonia.
- `sarscov2 [int]`: Total number of COVID-9 cases in Catalonia.
- `edad04.sdgripal [int]`: Number of influenza cases in the age range from 0 up to 4 years.
- `edad04.sarscov2 [int]`: Number of COVID-19 cases in the age range from 0 up to 4 years.
- `edad514.sdgripal [int]`: Number of influenza cases in the age range from 5 up to 14 years.
- `edad514.sarscov2 [int]`: Number of COVID-19 cases in the age range from 5 up to 14 years.
- `edad1544.sdgripal [int]`: Number of influenza cases in the age range from 15 up to 44 years.
- `edad1544.sarscov2 [int]`: Number of COVID-19 cases in the age range from 15 up to 44 years.
- `edad4565.sdgripal [int]`: Number of influenza cases in the age range from 45 up to 64 years.
- `edad4565.sarscov2 [int]`: Number of COVID-19 cases in the age range from 45 up to 64 years.
- `edad65.sdgripal [int]`: Number of influenza cases in the age range from 65.
- `edad65.sarscov2 [int]`: Number of COVID-19 cases in the age range from 65.
- `pob04 [int]`: Scaled population in Catalonia in the age range from 0 up to 4 years.

- pob514 [int]: Scaled population in Catalonia in the age range from 5 up to 14 years.
- pob1544 [int]: Scaled population in Catalonia in the age range from 15 up to 44 years.
- pob4565 [int]: Scaled population in Catalonia in the age range from 45 up to 64 years.
- pob56 [int]: Scaled population in Catalonia in the age range from 65.
- PIRINEU.sdgripal [int]: Number of influenza cases in the *Alt Pirineu i Aran* Health Area.
- PIRINEU.sarscov2 [int]: Number of COVID-19 cases in the *Alt Pirineu i Aran* Health Area.
- BARCELONA.sdgripal [int]: Number of influenza cases in *Barcelona Ciutat* Health Area.
- BARCELONA.sarscov2 [int]: Number of COVID-19 cases in *Barcelona Ciutat* Health Area.
- CATALUNYA.sdgripal [int]: Number of influenza cases in *Catalunya Central* Health Area.
- CATALUNYA.sarscov2 [int]: Number of COVID-19 cases in *Catalunya Central* Health Area.
- GIRONA.sdgripal [int]: Number of influenza cases in *Girona* Health Area.
- GIRONA.sarscov2 [int]: Number of COVID-19 cases in *Girona* Health Area.
- LLEIDA.sdgripal [int]: Number of influenza cases in *Lleida* Health Area.
- LLEIDA.sarscov2 [int]: Number of COVID-19 cases in *Lleida* Health Area.
- METR_NORD.sdgripal [int]: Number of influenza cases in *Metropolitan Nord* Health Area of Barcelona.
- METR_NORD.sarscov2 [int]: Number of COVID-19 cases in *Metropolitan Nord* Health Area of Barcelona.
- METR_SUD.sdgripal [int]: Number of influenza cases in *Metropolitan Sud* Health Area of Barcelona.
- METR_SUD.sarscov2 [int]: Number of COVID-19 cases in *Metropolitan Sud* Health Area of Barcelona.
- TARRAGONA.sdgripal [int]: Number of influenza cases in *Camp of Tarragona* Health Area.
- TARRAGONA.sarscov2 [int]: Number of COVID-19 cases in *Camp of Tarragona* Health Area.
- TERRES_EBRE.sdgripal [int]: Number of influenza cases in *Terres de l'Ebre* Health Area.
- TERRES_EBRE.sarscov2 [int]: Number of COVID-19 cases in *Terres de l'Ebre* Health Area.
- VALLES.sdgripal [int]: Number of influenza cases in *Valles* Health Area.
- VALLES.sarscov2 [int]: Number of COVID-19 cases in *Valles* Health Area.
- vac1218[int]: Number of vaccinations in the age range from 12 up to 18 years (exclusive).
- vac1845[int]: Number of vaccinations in the age range from 18 up to 45 years (exclusive).
- vac4565[int]: Number of vaccinations in the age range from 45 up to 65 years (exclusive).
- vac6580[int]: Number of vaccinations in the age range from 65 up to 80 years (exclusive).
- vac80[int]: Number of vaccinations in the age range from 80 years.
- vactotal[int]: Number of total vaccinations in Catalonia.

Note that age ranges in vaccination and IRAS data do not match. This is due to the different sources of each dataset.

```
cataluna <- read.csv("../data/evolucion_gripe_covid.csv")
```

1 COVID-19 evolution, influenza impact and vaccination progress

1.1 Study on global data

In this section we analyze the impact of the influenza cases and vaccination progress to the COVID-19 evolution. Let's assume we want to model, if it exists, the correlation between the COVID-19 evolution (dependent variable) and influenza and vaccination progress (candidate covariates).

We construct a new time series matrix containing only the necessary variables:

```
sarscov2 <- ts(cataluna$sarscov2)
xregs <- ts(cataluna[, c('sdgripal', 'vactotal')])
```

And use the function `auto.fit.arima.regression()` with the following configuration:

- The information criterion used is the AICc (`ic="aicc"`).
- The significant level used in the hypothesis tests is 0.05 (`alpha=0.05`).
- The stationary property is checked by adjusting an ARIMA(p,d,q) and analyzing the *d* order (`stationary_method="auto.arima"`).

```
adjust <- drm.select(sarscov2, xregs,
                    ic='aicc', alpha=0.05, st_method='auto.arima', show_info=T)
```

Covariate sdgripal has been tested [ic=896.9828870118, lag=0]

Covariate vactotal has been tested [ic=997.623845904429, lag=-3]

Covariate sdgripal has been added [aicc=896.9828870118, lag=0]

Series: serie

Regression with ARIMA(2,0,1) errors

Coefficients:

	ar1	ar2	ma1	intercept	xreg
	1.7816	-0.8531	-0.6866	-134.4028	0.7563
s.e.	0.0863	0.0773	0.1382	49.3195	0.0476

sigma² = 6818: log likelihood = -441.88

AIC=895.77 AICc=896.98 BIC=909.75

Covariate vactotal has been tested [ic=880.988687960867, lag=-3]

Covariate vactotal has been added [aicc=880.988687960867, lag=-3]

Series: serie

Regression with ARIMA(2,0,1) errors

Coefficients:

	ar1	ar2	ma1	sdgripal	vactotal
	1.8208	-0.8982	-0.8046	0.7062	-0.0063
s.e.	0.0622	0.0557	0.1209	0.0323	0.0012

sigma² = 5511: log likelihood = -433.89

AIC=879.77 AICc=880.99 BIC=893.76

No more variables will be added

Historical of added covariates to the model (ndiff=0)				
var	lag	ic		
sdgripal	0	896.9828870118		
vactotal	-3	880.988687960867		

Series: serie

Regression with ARIMA(2,0,1) errors

Coefficients:

	ar1	ar2	ma1	sdgripal	vactotal
	1.8208	-0.8982	-0.8046	0.7062	-0.0063
s.e.	0.0622	0.0557	0.1209	0.0323	0.0012

sigma^2 = 5511: log likelihood = -433.89

AIC=879.77 AICc=880.99 BIC=893.76

Our selection method states that there is a significant correlation between the influenza cases and vaccination progress of the Catalan population and the COVID-19 evolution in this region. This relation can be defined as a dynamic linear regression model with the form:

$$\text{sarscov2}_t = 0.7062 \cdot \text{sdgripal}_t - 0.0063 \cdot \text{vactotal}_{t-9} + \eta_t, \quad \eta_t \sim \text{ARIMA}(2,0,1)$$

where $\phi_1 = 1.8208$, $\phi_2 = -0.8982$ and $\theta_1 = -0.8046$ are the AR and MA coefficients, respectively, of η_t .

Thus, our selection approach confirms the existence of correlation between the current influenza cases and COVID-19 evolution, and the negative impact (with approximately 3 weeks delay) of the vaccination progress in the virus confirmed cases.

1.2 Study on vaccination data aggregated by group of ages

Once we have confirmed the existence of relation between the COVID-19 confirmed cases and vaccination progress in Catalonia, it might be relevant to investigate if the vaccination of some group of ages have more or less impact in the COVID-19 evolution. Since our database provides us of grouped vaccination data, we can launch again the selection function with the same configuration and interpret its result.

```
sarscov2 <- ts(cataluna$sarscov2)
xregs <- ts(cataluna[, c('vac1218', 'vac1845', 'vac4565', 'vac6580', 'vac80')])
adjust <- drm.select(sarscov2, xregs,
                    ic='aicc', alpha=0.05, st_method='auto.arima', show_info=T)
```

Covariate vac1218 has been tested [ic=902.564676702562, lag=-6]

Covariate vac1845 has been tested [ic=961.007698785186, lag=-3]

Covariate vac4565 has been tested [ic=961.424068569701, lag=-3]

Covariate vac6580 has been tested [ic=961.155786056672, lag=-3]

Covariate vac80 has been tested [ic=961.612340826851, lag=-3]

Covariate vac1218 has been added [aicc=902.564676702562, lag=-6]

Series: serie
Regression with ARIMA(2,1,0) errors

Coefficients:

	ar1	ar2	xreg
	1.1832	-0.5241	-0.2086
s.e.	0.0989	0.0979	0.0131

sigma² = 14752: log likelihood = -446.98
AIC=901.97 AICc=902.56 BIC=911.07

Covariate vac1845 has been tested [ic=884.279707733966, lag=-3]
Covariate vac4565 has been tested [ic=883.685557201231, lag=-3]
Covariate vac6580 has been tested [ic=884.591633342795, lag=-3]
Covariate vac80 has been tested [ic=884.030149535582, lag=-3]
Covariate vac4565 has been added [aicc=883.685557201231, lag=-3]

Series: serie
Regression with ARIMA(2,1,0) errors

Coefficients:

	ar1	ar2	vac1218	vac4565
	1.2517	-0.5490	-0.2098	-0.0233
s.e.	0.0974	0.0979	0.0109	0.0047

sigma² = 11120: log likelihood = -436.39
AIC=882.78 AICc=883.69 BIC=894.16

Covariate vac1845 has been tested [ic=880.629524560978, lag=-6]
Covariate vac6580 has been tested [ic=877.400381043013, lag=-2]
Covariate vac80 has been tested [ic=880.408412615052, lag=-2]
Covariate vac6580 has been added [aicc=877.400381043013, lag=-2]

Series: serie
Regression with ARIMA(2,1,0) errors

Coefficients:

	ar1	ar2	vac1218	vac4565	vac6580
	1.2273	-0.4932	-0.2093	-0.0361	-0.0388
s.e.	0.1057	0.1088	0.0105	0.0063	0.0132

sigma² = 10019: log likelihood = -432.05
AIC=876.11 AICc=877.4 BIC=889.77

Covariate vac1845 has been tested [ic=867.034569176019, lag=-6]
Covariate vac80 has been tested [ic=875.77069864328, lag=-6]
Covariate vac1845 has been added [aicc=867.034569176019, lag=-6]

Series: serie
Regression with ARIMA(5,1,0) errors

Coefficients:

	ar1	ar2	ar3	ar4	ar5	vac1218	vac4565	vac6580	vac1845
	0.9694	0	0	-0.5038	0.2884	0	-0.0493	-0.0650	-0.0864
s.e.	0.0649	0	0	0.1149	0.1097	0	0.0051	0.0112	0.0048

sigma² = 8466: log likelihood = -425.64

AIC=865.28 AICc=867.03 BIC=881.22

Covariate vac80 has been tested [ic=867.034569177984, lag=-6]

No more variables will be added

The global model does not have stationary errors

Trying to adjust a model that do have stationary errors

```
-----
|                               Historical of added covariates to the model (ndiff=0)                               |
-----
      var lag              ic
vac1218  -6 902.564676702562
vac4565  -3 883.685557201231
vac6580  -2 877.400381043013
vac1845  -6 867.034569176019
-----
```

Series: serie

Regression with ARIMA(4,0,0) errors

Coefficients:

	ar1	ar2	ar3	ar4	vac1218	vac4565	vac6580	vac1845
	2.0816	-1.2837	0	0.1919	0	-0.0410	-0.0468	-0.0901
s.e.	0.0810	0.1152	0	0.0432	0	0.0057	0.0120	0.0047

sigma^2 = 8767: log likelihood = -435.27

AIC=884.53 AICc=886.26 BIC=900.57

As our natural reasoning might suspect, by separating the vaccinations by group of ages, our selection method states that:

- There is a lagged negative correlation between the vaccination data and COVID-19 evolution.
- The vaccination of the population between 65 and 80 years has a greater impact in the COVID19 confirmed cases than the rest of population (the coefficient magnitude is greater and the impact delay is smaller).
- The vaccination of the youngest population (from 12 up to 18 years old) has some impact in the COVID19 evolution, but the standard deviation of the estimated coefficient is greater.

In addition, our implemented function `forecast_model()` makes predictions at horizon h by introducing the original data and the result of the function `auto.fit.arima.regression()`.

```
preds <- forecast_model(sarscov2, xregs, adjust, h=10,
                        mode='bootstrap', levels=c(50, 75, 90))
fig <- plot_forecast(preds)

# change x axis
time_vec <- window(cataluna$fecha, start=start(preds$x))
fig <- fig %>% layout(
  xaxis = list(
    ticktext = c('2021', '2022'),
    tickvals = time(time_vec)[1] +
      c(
```

```

        which(str_detect(time_vec, '2021'))[1],
        which(str_detect(time_vec, '2022'))[1]
    ),
    tickmode = 'array'
)
)

display(fig, name='preds_covid19_vac')

```

