A multivariate statistical approach to predict COVID-19 count data with epidemiological interpretation and uncertainty quantification

Fulvia Pennoni

Department of Statistics and Quantitative Methods - University of Milano-Bicocca (IT)

https://sites.google.com/view/fulviapennoni/home

fulvia.pennoni@unimib.it

Francesco Bartolucci

University of Perugia (IT)

Antonietta Mira

Università della Svizzera Italiana (CH) and University of Insubria (IT)

Outline

- We propose *statistical autoregressive models to analyze* the observed time series of count data referred to different categories
- We apply the approach to *Italian COVID-19 data* (at national level and for Lombardy) considering different categories of patients further to susceptible individuals and deceased
- For the *COVID-19*, $K = 6$ categories are considered:
  1. *susceptible* not previously ill (S)
  2. *recovered* (R)
  3. positive cases in *quarantine* (Q)
  4. *hospitalized* in regular wards (H)
  5. patients in *intensive care units* (ICU)
  6. *deceased* (D)
- The main assumption is that observed frequencies correspond to margins of a sequence of *unobserved contingency tables*

Model assumptions

- We observe *counts* for K categories over T time occasions, which are denoted by

$$y_{tk}, \quad t \in \mathcal{T} = \{1, \ldots, T\},\ k \in \mathcal{K} = \{1, \ldots, K\},$$

and are realizations of the random variables Y_{tk} collected in the vectors $\boldsymbol{Y}_t = (Y_{t1}, \ldots, T_{tK})'$

- The proposed approach is based on *three main assumptions*

- The *1st assumption* is that for $t \in \mathcal{T}' = \{2, \ldots, T\}$,

$$Y_{tk} = \sum_{j \in \mathcal{K}} X_{tjk}, \quad k \in \mathcal{K},$$

under the constraint

$$\sum_{k \in \mathcal{K}} X_{tjk} = Y_{t-1,j}, \quad j \in \mathcal{K}$$

- The X_{tjk} are frequencies of a *"transition table"* with row totals equal to $Y_{t-1,k}$ and column totals Y_{tk}, which are collected in the vectors $\boldsymbol{X}_{tj} = (X_{tj1}, \ldots, X_{tjK})'$

- The transition tables have *structural zeros* from any category different from S to S and from D to any category different from D

	S	R	Q	H	ICU	D	Total
S	X_{t11}	X_{t12}	X_{t13}	X_{t14}	X_{t15}	X_{t16}	$Y_{t-1,1}$
R	0	X_{t22}	X_{t23}	X_{t24}	X_{t25}	X_{t26}	$Y_{t-1,2}$
Q	0	X_{t32}	X_{t33}	X_{t34}	X_{t35}	X_{t36}	$Y_{t-1,3}$
H	0	X_{t42}	X_{t43}	X_{t44}	X_{t45}	X_{t46}	$Y_{t-1,4}$
ICU	0	X_{t52}	X_{t53}	X_{t54}	X_{t55}	X_{t56}	$Y_{t-1,5}$
D	0	0	0	0	0	X_{t66}	$Y_{t-1,6}$
Total	Y_{t1}	Y_{t2}	Y_{t3}	Y_{t4}	Y_{t5}	Y_{t6}	N

- X_{t35} corresponds to the number of individuals who moved from category Q at time $t-1$ into category ICU at occasion t

- The *overall frequency* N is kept fixed across time

- The *2nd assumption* concerns the distribution of every random vector $\boldsymbol{X}_{tj}$; there are two options:
  1. Multinomial distribution
  2. Dirichlet-Multinomial distribution

- *Multinomial formulation*:

$$\boldsymbol{X}_{tj}|\boldsymbol{Y}_{t-1} = \boldsymbol{y}_{t-1} \sim \text{Mult}(y_{t-1,j}; \boldsymbol{p}_{tj}),$$

  where $\boldsymbol{p}_{tj} = (p_{tj1}, \ldots, p_{tjK})'$ is a vector of "transition probabilities" from category j to the other categories

- The first two *moments* are:

$$\text{E}(\boldsymbol{X}_{tj}|\boldsymbol{Y}_{t-1} = \boldsymbol{y}_{t-1}) = y_{t-1,j}\boldsymbol{p}_{tj},$$

$$\text{Var}(\boldsymbol{X}_{tj}|\boldsymbol{Y}_{t-1} = \boldsymbol{y}_{t-1}) = y_{t-1,j}[\text{diag}(\boldsymbol{p}_{tj}) - \boldsymbol{p}_{tj}\boldsymbol{p}'_{tj}]$$

- To account for overdispersion, we can alternatively assume a Dirichlet-Multinomial distribution:

$$\boldsymbol{X}_{tj}|\boldsymbol{Y}_{t-1} = \boldsymbol{y}_{t-1} \sim \text{Dir} - \text{Mult}(y_{t-1,j}; \boldsymbol{\alpha}_{tj}),$$

where $\boldsymbol{\alpha}_{tj}$ is a vector of K positive parameters α_{tjk}

- The first two *moments* are:

$$\text{E}(\boldsymbol{X}_{tj}|\boldsymbol{Y}_{t-1} = \boldsymbol{y}_{t-1}) = y_{t-1,j}\frac{\boldsymbol{\alpha}_{tj}}{\alpha_{tj+}},$$

$$\text{Var}(\boldsymbol{X}_{tj}|\boldsymbol{Y}_{t-1} = \boldsymbol{y}_{t-1}) = y_{t-1,j}\left[\text{diag}\left(\frac{\boldsymbol{\alpha}_{tj}}{\alpha_{tj+}}\right) - \frac{\boldsymbol{\alpha}_{tj}}{\alpha_{tj+}}\frac{\boldsymbol{\alpha}'_{tj}}{\alpha_{tj+}}\right]\frac{y_{t-1,j}+\alpha_{tj+}}{1+\alpha_{tj+}},$$

with $\alpha_{tj+} = \sum_{k\in\mathcal{K}}\alpha_{tjk}$

- Letting $p_{tjk} = \alpha_{tjk}/\alpha_{tj+}$, the expected value is the same as the Multinomial one; the variance terms *tend to the Multinomial ones* as $\alpha_{tj+} \to \infty$

- The *3rd assumption* concerns the parametrization of the assumed distribution

- Under the *Multinomial model*, we assume that

$$p_{tjk} = \frac{\exp(\boldsymbol{f}'_{tjk}\boldsymbol{\beta}_{jk})}{\sum_{l\in\mathcal{D}_j}\exp(\boldsymbol{f}'_{tjl}\boldsymbol{\beta}_{jl})}, \quad t \in \mathcal{T}',\ j \in \mathcal{K},\ k \in \mathcal{D}_j,$$

  where $\mathcal{D}_j$ is the *set of non-zero cells* in the j-th row of each "transition table"

- For *model identifiability* we constrain $\boldsymbol{\beta}_{jj} \equiv 0$ for each j

- The *design column vectors* $\boldsymbol{f}_{tjk}$ contain the terms of a polynomial (or spline) of time t of a suitable order and may include indicator variables for interventions (e.g., $\boldsymbol{f}_{tjk} = (1, t, t^2, t^3)'$ when 3rd order polynomials are adopted)

- Under the *Dirichlet-Multinomial parametrization*, we directly assume

$$\alpha_{tjk} = \exp(\boldsymbol{f}'_{tjk}\boldsymbol{\beta}_{jk}), \quad t \in \mathcal{T}', \; j \in \mathcal{K}, \; k \in \mathcal{D}_j,$$

  without constraining any regression vector $\boldsymbol{\beta}_{jk}$ to 0

- The resulting model has a straightforward interpretation, but the *distribution of the frequencies* Y_{tk} is difficult to deal with as it derives from the convolution of

$$\prod_{j \in \mathcal{K}} p(\boldsymbol{X}_{tj} = \boldsymbol{x}_{tj} | \boldsymbol{Y}_{t-1} = \boldsymbol{y}_{t-1})$$

- The proposed approach may be seen as an *extension* of that for 2×2 contingency tables proposed in Eleftheraki et al. (2009); a related model is also described in Zhang et al. (2020) and Whiteley & Rimella (2021)

Bayesian inference

- The $\boldsymbol{\beta}_{jk}$ *parameters* are assumed to be *a priori* independent with distribution

$$\boldsymbol{\beta}_{jk} \sim N(0, \sigma^2 \boldsymbol{I}), \quad j \in \mathcal{K},\ k \in \mathcal{D}_j,$$

where σ^2 is a large value (*diffuse prior distributions*)

- To incorporate specific *a prior* hypotheses and for stability reasons, *we also assume constrains* of type

$$a_{jk} \leq o_{tjk} \leq b_{jk}, \quad j, k \in \mathcal{K},\ t \in \mathcal{T}^* = \{2, \ldots, T^*\},\ a_{jk}, b_{jk} \in R^+,$$

where $o_{tjk} = p_{tjk}/p_{tjj}$ is the odds referred to category k with respect to category j at time occasion t

- *Informative priors* may alternatively be considered by suitably choosing the hyperparameters of the prior distributions

- The model is estimated through a *data augmentation* (Tanner and Wong, 1987) MCMC algorithm based on a Metropolis sampler repeating two steps:

  1. for all $t > 1$ update every contingency table with elements x_{tjk} given the observe margins y_{tk} and the current parameter vectors $\boldsymbol{\beta}_{jk}$
  2. draw the model parameters $\boldsymbol{\beta}_{jk}$ given the current values of the count variables X_{tjk}

- The *algebraic algorithm* of Diaconis (1998) is employed to sample tables with fixed margins, whereas the model parameters are drawn by a series of Metropolis-Hastings moves

- *Updating "transition tables"*:

  1. randomly select (several times) two rows and two columns of the current table so that a 2×2 subtable is identified
  2. propose a switch by adding (or subtracting) to the two cells in the main diagonal of the subtable a random integer number, which is subtracted (or added) to the off-diagonal cells

$$\begin{pmatrix} + & - \\ - & + \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} - & + \\ + & - \end{pmatrix} \quad \text{with probability } 1/2$$

  3. accept the new table with probability

$$\min\left(1, \prod_{j \in \mathcal{K}} \frac{p(\boldsymbol{X}_{tj} = \boldsymbol{x}^*_{tj} | \boldsymbol{Y}_{t-1} = \boldsymbol{y}_{t-1}, \boldsymbol{\beta}_j)}{p(\boldsymbol{X}_{tj} = \boldsymbol{x}_{tj} | \boldsymbol{Y}_{t-1} = \boldsymbol{y}_{t-1}, \boldsymbol{\beta}_j)}\right),$$

  where $\boldsymbol{x}_{tj}$ is the vector of the frequencies in the j-th row of the current table, $\boldsymbol{x}^*_{tj}$ is that of the proposed table, and $\boldsymbol{\beta}_j$ is the matrix containing all current regression vectors $\boldsymbol{\beta}_{jk}$, $k \in \mathcal{D}_j$

- *Drawing new parameter vectors*:
  1. for all j and $k \in \mathcal{D}_j$ a new value of $\boldsymbol{\beta}_{jk}$, denote by $\boldsymbol{\beta}^*_{jk}$, is drawn from the proposal distribution $N(\boldsymbol{\beta}_{jk}, \tau^2 \boldsymbol{I})$
  2. the proposed vector is accepted with probability

$$\min\left(1, \frac{\prod_{t\in\mathcal{T}'} p(\boldsymbol{X}_{tj} = \boldsymbol{x}_{tj} | \boldsymbol{Y}_{t-1} = \boldsymbol{y}_{t-1}, \boldsymbol{\beta}^{\dagger}_{jk})}{\prod_{t\in\mathcal{T}'} p(\boldsymbol{X}_{tj} = \boldsymbol{x}_{tj} | \boldsymbol{Y}_{t-1} = \boldsymbol{y}_{t-1}, \boldsymbol{\beta}_{j})} \frac{\pi(\boldsymbol{\beta}^*_{jk})}{\pi(\boldsymbol{\beta}_{jk})}\right),$$

  where $\boldsymbol{\beta}^{\dagger}_{jk}$ is the same matrix as $\boldsymbol{\beta}_j$ with $\boldsymbol{\beta}_{jk}$ substituted with $\boldsymbol{\beta}^*_{jk}$, and $\pi(\boldsymbol{\beta}_{jk})$ is the prior density of the regression parameters

- The simulated posterior distribution of the parameters and tables is *summarized* in the usual way also providing variability measures in order to quantify the uncertainty

- At each step, the algorithm also performs *in-sample and out-sample predictions*
- For $t \in \mathcal{T}$, *(in-sample) predictions* of the frequencies y_{tk} at step s of the algorithm are computed as

$$\hat{y}_{tk}^{(s)} = \sum_{j \in \mathcal{K}} y_{t-1,j} p_{tjk}^{(s)}$$

- For $t > T$, (out-sample) predictions are based on the recursive rule

$$\hat{y}_{tk}^{(s)} = \sum_{j \in \mathcal{K}} \hat{y}_{t-1,j}^{(s)} p_{tjk}^{(s)},$$

initialized with $\hat{y}_{Tj}^{(s)} = y_{Tj}$

- For the COVID-19 application, at each step of the MCMC algorithm, the *net reproduction number* R_t is predicted as

$$\widehat{R}_t^{(s)} = \frac{\widehat{\Delta I}_t^{(s)}}{\sum_{r=1}^{t-1} \omega_{s,t-1} \widehat{\Delta I}_{t-r}^{(s)}},$$

  - $\omega_{r,t-1}$ is a weight obtained by normalizing the density of the Gamma distribution with parameters 1.87 and 0.28
  - $\widehat{\Delta I}_t^{(s)}$ is the number of new positive individuals predicted by the model for day t

- This method *directly derives* from Riccardo et al. (2020) for the Italian context

Model checking

- The *goodness-of-fit* of the model is assessed by a discrepancy measure between observed counts and in-sample predictions

$$\widehat{\mathrm{Dist}}^{(s)} = \sum_{t \in \mathcal{T}'} \sum_{k \in \mathcal{K}} \frac{(y_{tk} - \hat{y}_{tk}^{(s)})^2}{\hat{y}_{tk}^{(s)}}$$

- When data are available, the quality of *(out-sample) predictions* is assessed by

$$\widehat{\mathrm{Dist}}_t^{(s)} = \sum_{k \in \mathcal{K}} \frac{(y_{tk} - \hat{y}_{tk}^{(s)})^2}{\hat{y}_{tk}^{(s)}}, \quad t > T$$

- A similar discrepancy measure is used to check the *prediction power* for each specific category and denoted by $\widehat{\mathrm{Dist}}_k^{*(s)}$

- The discrepancy measures computed across iterations are *summarized* by simple means obtaining $\widehat{\mathrm{Dist}}$, $\widehat{\mathrm{Dist}}_t$, and $\widehat{\mathrm{Dist}}_k^*$

- For $\widehat{\mathrm{Dist}}$, a *posterior predictive (PP) p-value* is also obtained; it is computed as the proportion of iterations for which $\widetilde{\mathrm{Dist}}^{(s)}$ is greater than $\widehat{\mathrm{Dist}}^{(s)}$, where $\widetilde{\mathrm{Dist}}^{(s)}$ is obtained by substituting each observed frequency y_{tk} with a simulated frequency

- Particular care is necessary to *assess the PP p-values*; for in-sample predictions we expect a value close to 0.5 when the model has an adequate fit (Gelman, 2013)

Application: Italian COVID-19 data at the beginning of the pandemic

- We examined the *daily Italian data collected from February 24 until April 24, 2020* (61 days)

- We considered *different models* based on:
  - Multinomial or Dirichlet-Multinomial distribution
  - polynomials of 2nd or 3rd order of the time and intervention dummies
  - with or without constraints on the odds:

	S	R	Q	H	ICU	D
S	-	10^{-7}	0.001	10^{-4}	10^{-6}	10^{-7}
R	-	-	0.001	10^{-4}	10^{-6}	10^{-7}
Q	-	0.1	-	0.1	10^{-5}	10^{-6}
H	-	0.1	0.1	-	0.1	0.01
ICU	-	10^{-7}	10^{-7}	0.25	-	0.25
D	-	-	-	-	-	-

- All the models also included two dummy variables to account for *the effect of for public health non-pharmaceutical interventions* enforced on February 24th and on March 8th 2020 in Italy

- *Goodness-of-fit* of the estimated models:

Multinomial	$\widehat{\text{Dist}}$	$\widetilde{\text{Dist}}$	p-value
Model 1 (2nd order, without constraints)	1,658.011	124.670	0.000
Model 2 (2nd order, with constraints)	2,347.274	68.474	0.000
Model 3 (3rd order, without constraints)	1,565.587	122.793	0.000
Model 4 (3rd order, with constraints)	2,203.832	70.512	0.000
Dirichlet-Multinomial	$\widehat{\text{Dist}}$	$\widetilde{\text{Dist}}$	p-value
Model 5 (2nd order, without constraints)	2,608.502	3,060.236	0.679
Model 6 (2nd order, with constraints)	2,992.213	3,629.419	0.750
Model 7 (3rd order, without constraints)	2,414.970	2,811.524	0.536
Model 8 (3rd order, with constraints)	2,915.772	3,344.208	0.661

- We considered in particular *Models 7 and 8*

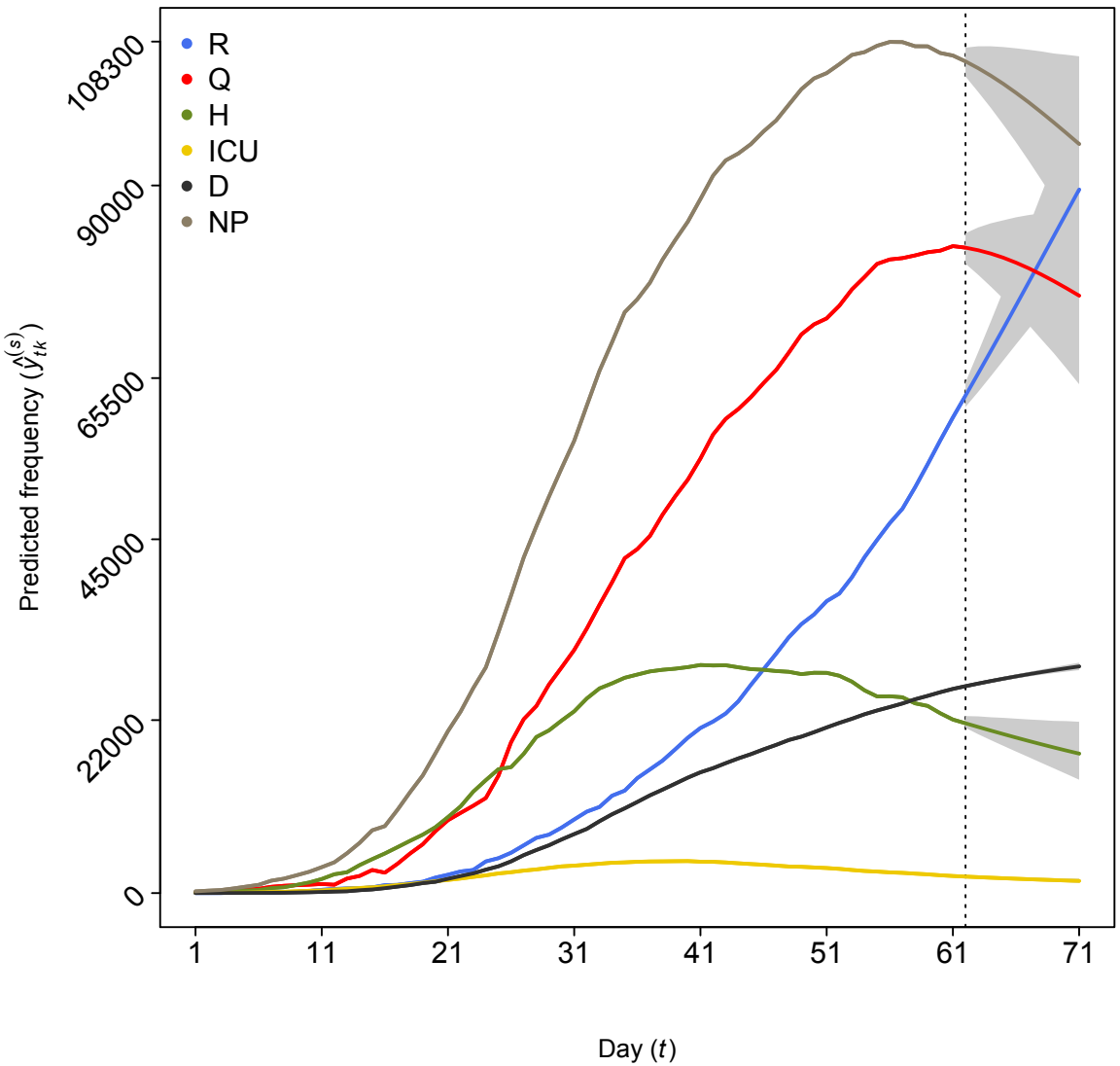
- Discrepancy measures for the *forecasted cases* (Model 8, 3rd order with constraints) according to the posterior predictive distribution:

Day	$\widehat{\text{Dist}}_t$	$\widetilde{\text{Dist}}_t$	p-value
25th April	3,231.755	24.523	0.769
26th April	3,347.780	36.457	0.403
27th April	2,976.716	19.313	0.198
28th April	3,105.249	26.695	0.161
29th April	3,216.649	31.738	0.137
30th April	3,095.463	31.599	0.164
1st May	2,979.734	37.135	0.118
2nd May	3,169.230	47.058	0.103
3rd May	3,223.772	58.826	0.095
4th May	3,112.596	44.670	0.069

- The *best predicted counts* are for categories ICU and D:

	S	R	Q	H	ICU	D	Total
$\widehat{\text{Dist}}_k^*$	0.000	1,409	1,397	372	31	12	3,220

- Daily *observed and predicted counts* for each category with a time horizon of 10 days and estimated 95% prediction intervals (in grey):

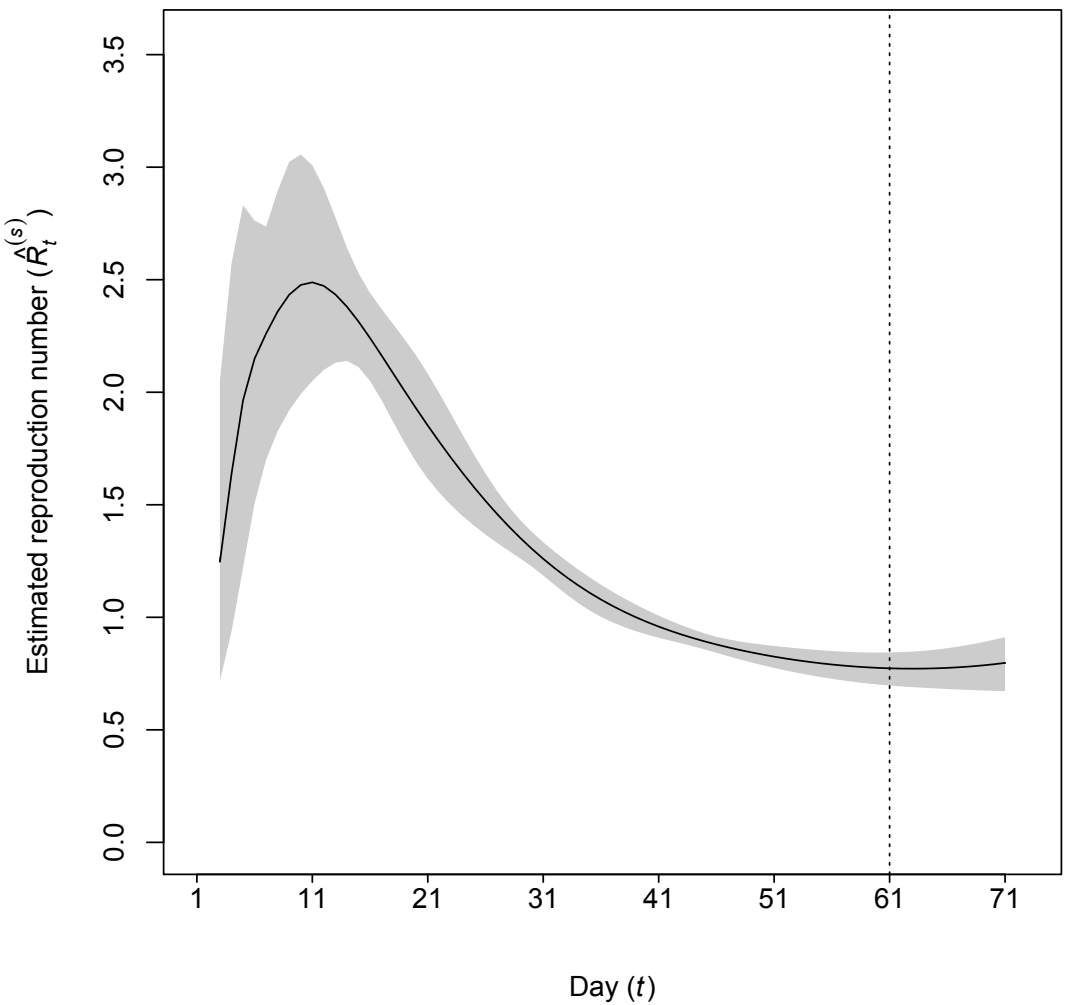


- Estimated posterior means of the *predicted transitions* between categories from 25th to 26th of April, 2020 (from the 61st to the 62nd day) and 95% prediction upper and lower bounds:

	S	R	Q	H	ICU	D
S	60,121,632	0	2,219	154	1	0
R	0	60,489	9	0	0	0
Q	0	2,665	79,105	516	0	0
H	0	116	757	20,925	73	197
ICU	0	0	0	0	2,023	149
D	0	0	0	0	0	25,969

	S	R	Q	H	ICU	D
S	-	(0, 0)	(1,217, 3,188)	(0, 718)	(0, 2)	(0, 0)
R	-	(60,471, 60,498)	(0, 26)	(0, 0)	(0, 0)	(0, 0)
Q	-	(1,269, 4,357)	(77,182, 80,672)	(32, 1,479)	(0, 0)	(0, 0)
H	-	(0, 506)	(463, 1,129)	(20,438, 21,321)	(25, 137)	(123, 282)
ICU	-	(0, 0)	(0, 0)	(0, 40)	(1,963, 2,075)	(98, 210)
D	-	-	-	-	-	-

- Estimated and predicted (from the vertical line) *reproduction number* R_t (61 observed days, prediction from 25th of April to 4th of May). Estimated 95% credibility and prediction intervals (in grey):

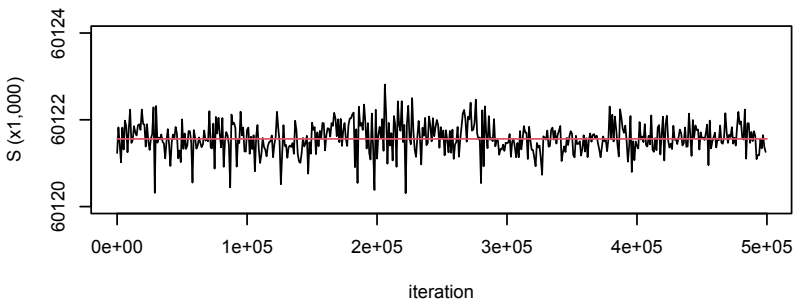


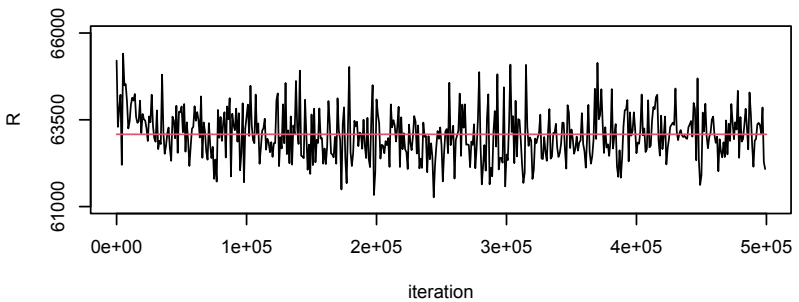
- We repeated the same analysis with Model 7 on Italian data and with Models 7 and 8 on data referred to the *Lombardy region*, obtaining similar results from several points of view
- The MCMC algorithms were run for *500,000 iterations* after a burnin of 100,000 iterations and a thinning of 10 iterations
- Diagnostics of the MCMC output reveals that the *effective sample size (ESS)* for the forecasted frequencies is satisfactory:

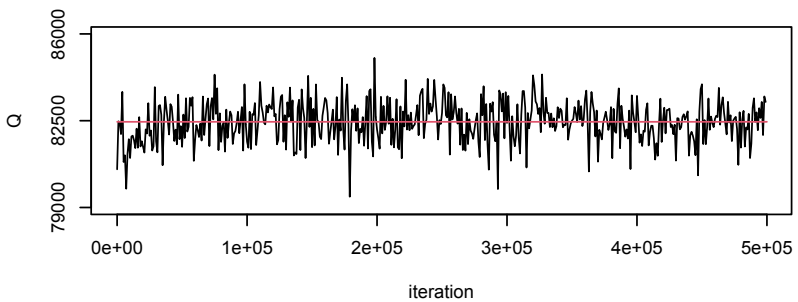
	Model 8			Model 7		
	Day 1	Day 2	Day 3	Day 1	Day 2	Day 3
S	12,893	5,641	3,677	6,911	2,049	897
R	11,605	4,611	2,865	4,768	2,129	1,603
Q	12,257	4,288	3,672	4,660	2,731	1,046
H	20,548	3,968	2,892	3,928	2,459	1,546
ICU	16,892	4,067	2,914	14,014	3,280	1,767
D	16,512	6,712	3,447	3,757	2,463	1,538

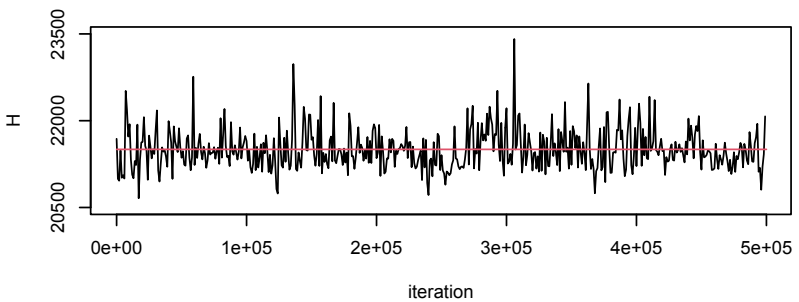
- The *ESS computed for the parameters* in β_{jk} are much lower and overall not completely satisfactory

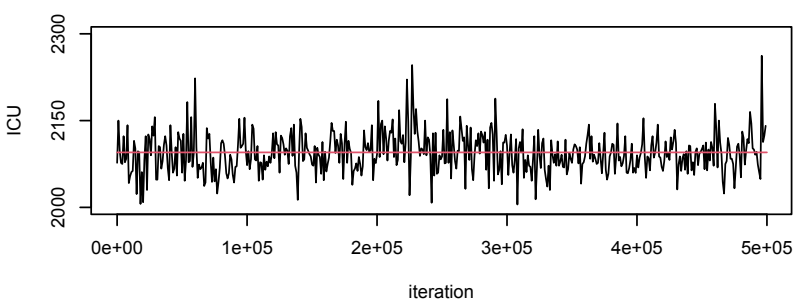
- *Trace plots* for 1-day ahead forecasts (one iteration every 1,000):

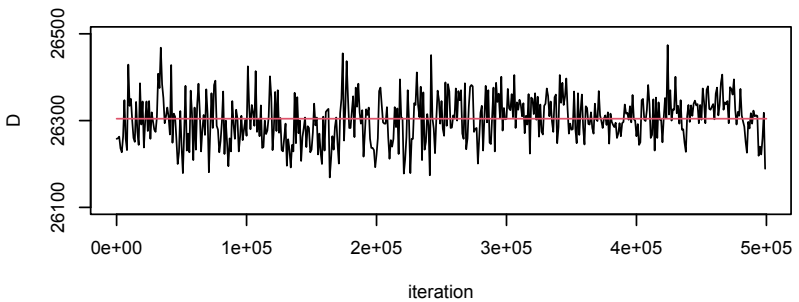








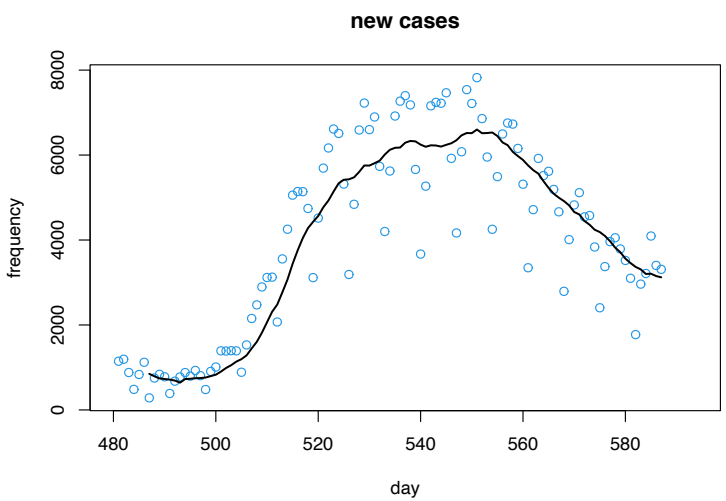


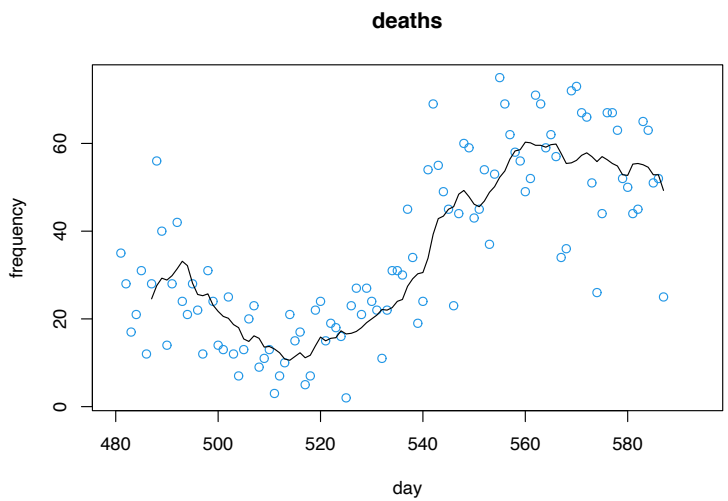


Weekly Italian data

- We use the proposed approach to perform *weekly forecasts* of the number of new cases and deaths for *Italy*, trying different model specifications
- These forecasts are published in the *"European Covid-19 Forecast Hub"* (`https://covid19forecasthub.eu/index.html`)
- We use the models based on the Dirichlet-Multinomial with a 3rd or 2nd order polynomial. We publish predictions of one of them or their averages
- The approach seems to perform *better for the weekly number of deaths* than for the number of new cases, even if the evaluated bias is particularly low with respect to that of the other research groups

- On the basis of the *observed data (from June 21 to October 3, 2021)* and with the mix of polynomial of 2nd and 3rd order, we predict 18227 (95% CI: 14212-22492) new cases and 299 (95% CI: 178-424) deaths for the week from October 04 to October 9





Main conclusions

- The approach allows us to *predict "transition tables"* on the basis of observed counts that may be useful in epidemiological contexts
- Being based on a *Bayesian approach*, it is possible to easily incorporate prior hypotheses on the basis of previous observations
- Despite the complexity of the distribution of the observed counts, *estimation is not particularly complex* by the MCMC algorithm that also allows us to easily perform predictions and quantify uncertainty
- We make our *implementation of the approach* available in `R` (`https://github.com/francescobartolucci/ARMultinomial`)
- This approach can also be used in several *other contexts*, whenever observed frequencies may be conceived as sums of "transition frequencies" (e.g., electoral flows)

Limits and possible developments

- The model is essentially *overparametrized* and the MCMC algorithm has a reduced ESS for the parameters $\rightarrow$ the parametrization of the transition probabilities p_{tjk} (or α_{tjik}) can be improved
- At the moment we do not use *covariates* apart from the temporal ones $\rightarrow$ we can easily include covariates (e.g., number of vaccinations)
- Under the Dirichlet-Multinomial formulation *prediction intervals* seem rather wide $\rightarrow$ explore restrictions on the parameters α_{tjk}
- In epidemiological contexts, the proposed model is closely related to models of type *Susceptible-Infected-Recovered* (SIR; e.g., Phenyo, 2006) $\rightarrow$ an accurate comparative analysis could be performed
- There are common points with *hidden Markov (HM) models* $\rightarrow$ try to cast the proposed model in the HM literature (Bartolucci et al., 2013; Zucchini, et al. 2017)

Main References

- Bartolucci, F., Farcomeni, A., and Pennoni, F. (2013), *Latent Markov Models for Longitudinal Data*. Boca Raton, FL: Chapman and Hall/CRC.
- Bartolucci, F., Pennoni F., and Mira A. (2021), A multivariate statistical approach to predict COVID-19 count data with epidemiological interpretation and uncertainty quantification, *Statistics in Medicine*, https://onlinelibrary.wiley.com/doi/10.1002/sim.9129.
- Diaconis, B. (1998), Algebraic algorithms for sampling from conditional distributions, *The Annals of Statistics*, **26**:363–397.
- Eleftheraki, A.G., Kateri, M., Ntzoufras, I. (2009), Bayesian analysis of two dependent 2×2 contingency tables, *Computational Statistics & Data Analysis*, **53**:2724–2732.
- Gelman, A. (2013), Two simple examples for understanding posterior p-values whose distributions are far from uniform, *Electronic Journal of Statistics*, **7**:2595–2602.

- Lekone, P.E. and Finkenstädt, B.F. (2006), Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study, *Biometrics*, **62**:1170–1177.
- Riccardo, F., Ajelli, M., Andrianou X., et al. (2020), aEpidemiological characteristics of COVID-19 cases and estimates of the reproductive numbers 1 month into the epidemic, Italy, 28 January to 31 March 2020 *Euro Surveill.*, **25**:1–11.
- Tanner, M.A. and Wong, W.H. (1987), The calculation of posterior distributions by data augmentation, *Journal of the American Statistical Association*, **82**:528–540.
- Whiteley, N. and Rimella, L. (2021), Inference in Stochastic Epidemic Models via Multinomial Approximations. In *International Conference on Artificial Intelligence and Statistics*, 1297–1305.
- Zhang, J., Wang, D., Yang, K., and Xu, Y. (2020), A multinomial autoregressive model for finite-range time series of counts, *Journal of Statistical Planning and Inference*, **207**:320–343.
- Zucchini, W., MacDonald, I.L., and Langrock, R. (2017), *Hidden Markov Models for Time Series: An Introduction Using R*, New York: Springer-Verlag.

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA