Corona COVID-19 Analysis: Switzerland and Europe

Mario V. Wüthrich
RiskLab, ETH Zurich

Version of April 14, 2020

We provide an update of our corona COVID-19 analyses as of April 11, 5 and 1, 2020.

Disclaimer. There are a couple of disclaimers to be made. Firstly, and most importantly, this is work in progress with daily arriving data updates, and conclusions may change at any time. Secondly, the data we use is quite coarse and refinements can and should be done. Many thanks to everyone for making this data timely available, you are doing a great job! Thirdly, the models used are rather crude, to not say ad-hoc, and they can substantially be improved and refined. In particular, we did not perform *any* model validation. Fourth, we do not have any medical nor epidemic background.

Main conclusions as of April 14, 2020.

- Continuation of the "soft lockdown" in Switzerland (we refer to Figures 2 and 3, below):
  - it is likely that the propagation of the virus will continue until the end of May;
  - we expect between 1'800 and 2'600 fatality cases in Switzerland, more than 95% of these cases being among people above age 60 (and 88% above age 70);
  - the economic damage of a lasting lockdown will be huge, also with long-term negative impacts on the health care system.
- The examples of China and Austria show that under a hard lockdown the propagation lasts roughly 4 weeks less than in Switzerland. A drawback of a hard lockdown is that the population is vulnerable to a second wave, which may also have severe long-term economic impacts, in particular, if the development of vaccine is not successful very soon. Pros are that even if a second wave comes the health care system will be better prepared and, likely, the worldwide shortage of medical equipment will be resolved by then.
- Alternatively, people below 50 may return to work, and a hard lockdown for people above 60 is executed (including high-risk groups such as high blood pressure, cardiovascular problems, diabetes, etc.). Mostly the age group above 60 is at risk, and for younger people an infection is in most cases unproblematic. Also under this alternative social distancing is crucial, in particular, between different age generations. A slow propagation among younger and healthy people may also prevent from a potential second wave. Moreover, if schools and universities go back to normal classroom teaching, the working class is relieved from parenting (and they may support the health care system also in case of increasing numbers of hospitalizations). Politically and socially this may not be the most practical proposal, however, it is probably the solution that minimizes negative long-term consequences.

1 Data sources

The data has been compiled from the daily WHO situation reports:

`https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports`

The Swiss data is taken from the website:

`http://www.corona-data.ch`

Age related Swiss information is taken from:

`https://www.bag.admin.ch/bag/de/home/krankheiten/ausbrueche-epidemien-pandemien/aktuelle-ausbrueche-epidemien/novel-cov/situation-schweiz-und-international.html`

2 Observed data as of April 13, 2020

Figure 1 illustrates the number of confirmed cases (lhs) and the number of fatalities (rhs) per 100'000 residents for different countries. For China we have adjusted the Hubei province population size proportionally to the confirmed cases outside of Hubei province. Note that the ratio of *confirmed* cases depends on the testing intensity which differs between the countries.

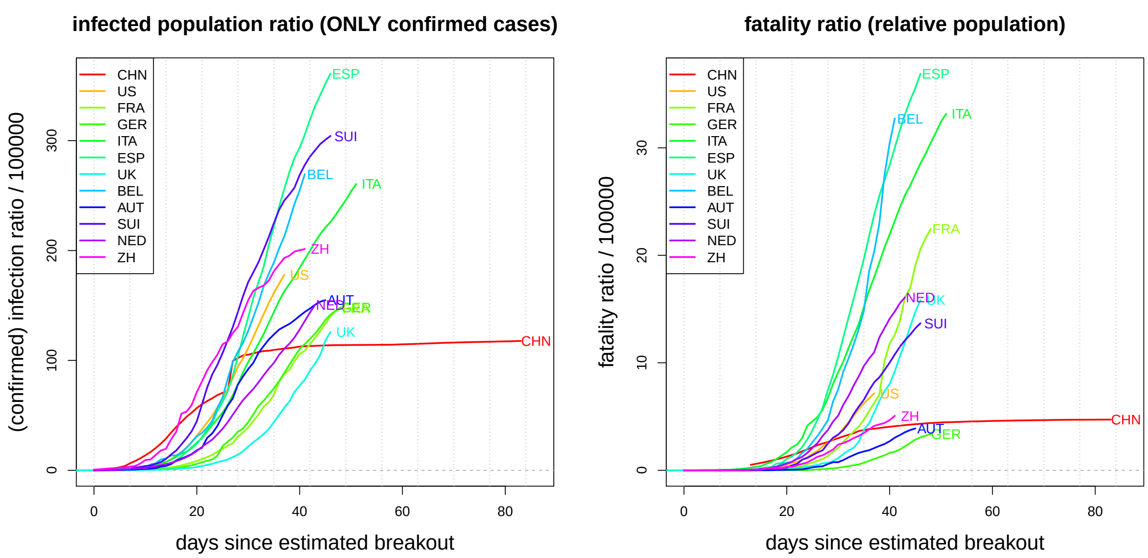

Figure 1: (lhs) Confirmed infections per 100'000 residents (this only considers *confirmed* cases which depends on the testing intensity in individual countries, tests per 100'000 residents: Switzerland 580, Italy 345 (date March 22, 2020)); (rhs) observed fatality ratios per 100'000 residents (this data is not complete due to missing numbers from elderly care homes in some countries).

The breakout of the disease in the different countries has been set manually to the time point when infection rates started to increase due to local transmissions, see Figure 1 (lhs). Of course, these choices are subjective and, e.g., for UK it could also have been shifted to a later date.

Remark on testing intensities: Main differences in Figure 1 (lhs) are explained by different testing intensities in individual countries; this may be caused by shortage of test equipment, medical staff, etc. Nevertheless, we expect the unreported number of cases being very large also due to other reasons. For instance, a common situation is that a family with two children gets infected. For this family it is sufficient to test one of the parents to have certainty that the family has been infected, thus, the statistics only shows one confirmed case for the whole family.

3 Analysis and prediction as of April 13, 2020

We provide a data analysis on selected European countries, additionally we consider the data of Kanton Zurich and the US.

Preliminary remark. State-of-the-art mathematical modeling of propagation of diseases uses susceptible-exposed-infectious-recovered (SEIR) models. These models involve a couple of parameters, the reproductive number being a crucial one. We could debate to which extent the available data allows one to reasonably fit such an epidemic propagation model, also because one has to account for many complex effects like mobility behavior of population, government intervention, etc. We take a different approach here by relying on very aggregated data, i.e. not aiming at modeling the micro-structure of the propagation of the disease, but rather providing a large-scale prediction. To this end, we consider a regression model on aggregate data, allowing us to analyze the trend of the propagation dynamics (we do this on a double-logarithmic scale receiving a sub-exponential growth rate, basically, we use Gompertz' law [1]). The prediction is then simply received by extrapolating this trend. This only involves two parameters per country, and it may be considered as being a crude method. However, working on aggregated data has often the advantage of receiving more robustness due to the fact that different effects may average out by aggregation, and as long as there is no structural break in the dynamics, aggregated methods often perform well for prediction. Another advantage is that our prediction only uses directly observable (and available) quantities.

We describe our analysis on the Swiss data in Figure 2, and the plots of other selected countries are provided in Figures 5-16 in the appendix below.

We introduce some notation. We denoted by C_t the total number of confirmed cases at day t of a given population, for $1 \leq t \leq \tau =$ April 13, 2020; time is initialized manually for each country as described above, and $\tau =$ April 13, 2020, denotes the latest observation point. The number of newly confirmed cases on day t of the chosen population is denoted by $Y_t = C_t - C_{t-1} \geq 0$. The blue dots in Figure 2 (lhs) show the totally observed confirmed cases on the log-scale, i.e. $(\log C_t)_{1\leq t\leq\tau}$, and the blue dots in Figure 2 (rhs) show the newly confirmed cases $(Y_t)_{1\leq t\leq\tau}$ for each day.

We perform two linear regressions (with square loss functions) on the logged cumulative cases

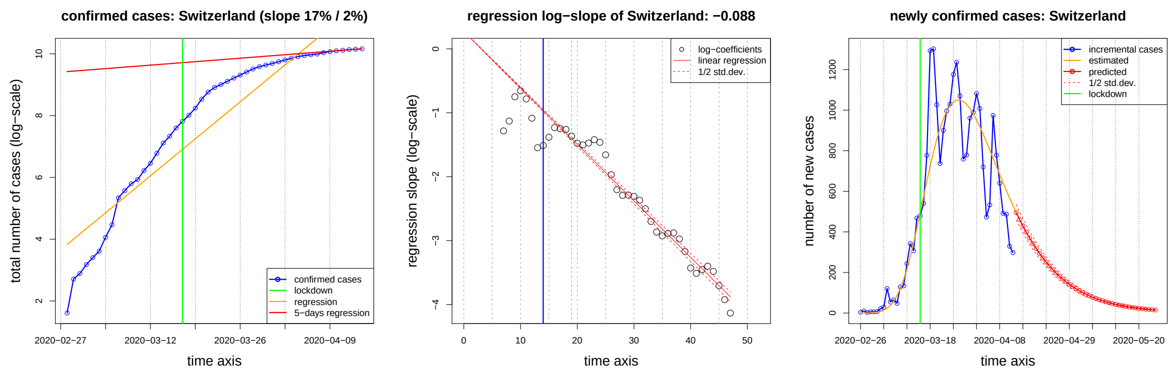

Figure 2: Switzerland: (lhs) confirmed cases on log-scale (in blue) and linear regression (entire time range in orange, last 5 days in red); (middle) logarithm of 5-days regression coefficients (black) and linear regression (red); (rhs) newly confirmed cases (blue) and predicted new cases (red); as of April 13, 2020.

$(\log C_t)_{1\leq t\leq \tau}$. The first linear regression considers all cases $(\log C_t)_{1\leq t\leq \tau}$ over the entire observation period $1 \leq t \leq \tau$; it is illustrated in orange color in Figure 2 (lhs). The second linear regression only considers the latest 5 days $(\log C_t)_{\tau-4\leq t\leq \tau}$; it is illustrated in red color in Figure 2 (lhs). The title of the plot states (in brackets) the slopes of these two regression lines. These slopes correspond to exponential growth rates, and the smaller this slope the smaller the number of new cases. Therefore, we hope that these exponential growth rates decrease over time, and this is exactly the basic idea of the following study and prediction analysis.[1]

We consider successive linear regressions always based on 5 days of observations $(\log C_t)_{T-4\leq t\leq T}$ for time points $5 \leq T \leq \tau$. Denote the resulting estimated regression parameters (only slope parameters) by β_T, $5 \leq T \leq \tau$. The latest one $\beta_{T=\tau}$ exactly corresponds to the slope of the red line in Figure 2 (lhs). The black dots in Figure 2 (middle) give the logarithms of these regression slopes $(\log \beta_T)_{5\leq T\leq \tau}$. The key observation is that these black dots (surprisingly) almost form a straight line (with negative slope). This observation is quite consistent across different countries, see middle plots in Figures 5-16, below. This motivates to approximate the decay of the exponential growth rate by an exponential function, and its parameter can be estimated by a linear regression on the log-scale, as illustrated in Figure 2 (middle). We choose country specific initial time points τ_0 for this regression estimation (illustrated by the vertical blue lines in the corresponding plots). Then, we fit a linear regression to $(\log \beta_T)_{\tau_0\leq T\leq \tau}$ (with square loss function). This regression line is illustrated in red color in Figure 2 (middle), and the dotted red lines give estimated 1/2 standard deviations around the central regression line. This linear regression function allows us to extrapolate β_T beyond the latest observation point $T > \tau =$ April 13, 2020. Back-transformation of this extrapolation then provides the predictions in Figure 2 (rhs). The dotted lines again give the estimated 1/2 standard deviation predictions from the central regression line. Note that in the back-transformation we exponentiate the regression lines, therefore, the dotted confidence bounds are not symmetric around the central prediction. Moreover, under the assumption that the exponential growth rate is decreasing this

[1]More mathematically speaking, if the growth rate decreases it means that we have sub-exponential growth and we try to determine this sub-exponential growth behavior in this study.

extrapolation is conservative, i.e. the true observations will likely be below the central prediction. In Appendix B, below, we show that the resulting functional form corresponds to Gompertz' law [1].

Results as of April 14, 2020 (see Figures 6-16, below).

- Austria, Italy, Spain and Switzerland (including Kanton Zurich) have reached the peak of new infections, and over the next weeks the numbers of new infections should decrease. However, the decrease of new infections may be slow and take until the end of May in Switzerland, Austria being clearly ahead in this development.
- Germany has probably left the peak behind, and France, The Netherlands and Belgium are currently around the peak of new infections. Note that in these statements quite some uncertainty is involved.
- The UK and the US have not yet reached the peak of new infections, but we expect this to happen within, say, the next week.

Remarks.

- The above predictions and results are based on the assumption that there is no second wave, for instance, in a different region of an affected country. The latter is currently observed in Iran or Japan, and also China may be vulnerable w.r.t. a second wave being imported by emigrants coming home from abroad.
- We did not explore differences in testing intensities in the different countries. These differences should not affect our analysis too much as long as philosophies within countries do not drastically change. This is due to the fact that we work on the log-scale where different testing intensities just lead to shifts that should not affect changes in speeds.

In Figures 17 and 18 we compare our predictions as of today to the ones made as of April 4, 2020, and April 10, 2020. The red lines show the predictions based on the data from April 13, 2020, and the magenta lines rely on the data from April 4, 2020, and April 10, 2020, respectively.

- In general, our earlier predictions as of April 4, 2020, have been too conservative, except for The Netherlands. However, for Switzerland (including Kanton Zurich), Italy, Austria and the US they did not substantially change.
- For Spain, France, Germany and Belgium the peak is reached sooner than predicted in the earlier analysis as of April 4, 2020. Moreover, the UK has not had a reasonable prediction in our previous analysis as of April 4, 2020.
- The analysis as of April 10, 2020, seems quite accurate.

4 Prediction of fatality ratios in Switzerland

We first allocate the predicted numbers of confirmed cases of Figure 2 (rhs) to different age classes: ages 0-49, ages 50-59, ages 60-69, ages 70-79 and ages 80+. To this end we choose estimated static partition ratios based on the latest observations.[2] These estimates are presented on line (b) of Table 1.

	ages 0-49	ages 50-59	ages 60-69	ages 70-79	ages 80+	total
(a) population size	5'204K	1'288K	936K	709K	445K	8'582K
– in % of total population	60.6%	15.0%	10.9%	8.3%	5.2%	
(b) estimated partition confirmed cases	43.9%	20.9%	12.6%	10.0%	12.6%	100.0%
(c) partition fatalities across ages	0.6%	2.2%	7.8%	22.1%	67.2%	100.0%

Table 1: Switzerland: partition across age classes; (b)-(c) are estimates as of April 14, 2020.

Using these estimated (static) partition ratios for confirmed cases on line (b) of Table 1, we receive the predicted development of confirmed cases per age class given in Figure 3 (lhs). For the 3 oldest age classes (which are most vulnerable to the corona disease) we predict roughly 3'500 to 4'500 confirmed positive cases for each of these age classes.[3]

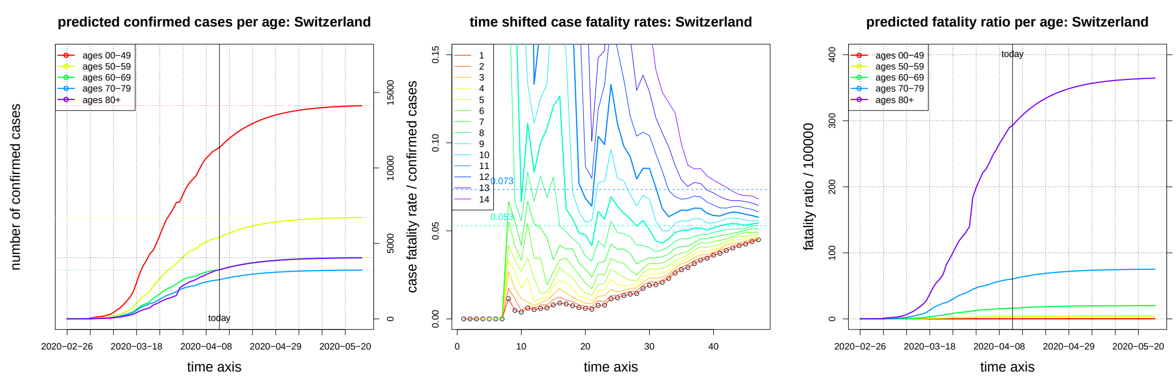

Figure 3: Switzerland: (lhs) Predicted number of confirmed cases per age class, (middle) observed time shifted case fatality rates D_t/C_{t-s} for time lags $s \geq 1$ and time points $t \geq s$, and (rhs) predicted fatality ratio per age class (per 100'000 people).

We relate the number of fatalities to the number of confirmed cases, this is the so-called case fatality rate (CFR). We denote the total number of people who die until time t by D_t. Naively, the case fatality rate is calculated by considering the ratios D_t/C_t. These ratios are illustrated by the black dots in Figure 3 (middle). We observe that these values are increasing over time, the reason being that deaths occur with a time lag after infection. Therefore, we plot time shifted ratios D_t/C_{t-s}, for time shifts $s = 1, \ldots, 14$ days, in different colors in Figure 3 (middle). A crude analysis of that plot shows that a time lag of 11 days may lead to a stationary pattern (this is a very rough estimate). This suggests that roughly 5.5% to 7.5% of the confirmed cases have a fatal outcome.[4] Supposed that this crude case fatality rate estimate is not completely off, we expected between 1'800 and 2'600 fatal outcomes up to the end of May (everything

[2] We note that the partition of confirmed cases across age classes in Switzerland is rather stable over time, maybe with a slight increase for age class 80+.

[3] This prediction is based on the assumption that no essential factors change, and that the static estimated partition ratios are appropriate.

[4] Franco Moriconi [2] also performed an estimate for Switzerland, and he proposes a case fatality rate of 7.8% and a time lag of 12 days.

unchanged). If we partition these predicted fatal outcomes to age classes according to line (c) of Table 1, we receive the fatality ratios per 100'000 residents presented in Figure 3. Roughly speaking, we would not be surprised to observe close to 400 death cases per 100'000 residents for persons in age class 80+, in age class 70-79 we predict almost 80 cases per 100'000 residents to occur, and in age class 60-69 roughly 20 cases per 100'000, and for younger ages we expect around 1 case per 100'000 residents.

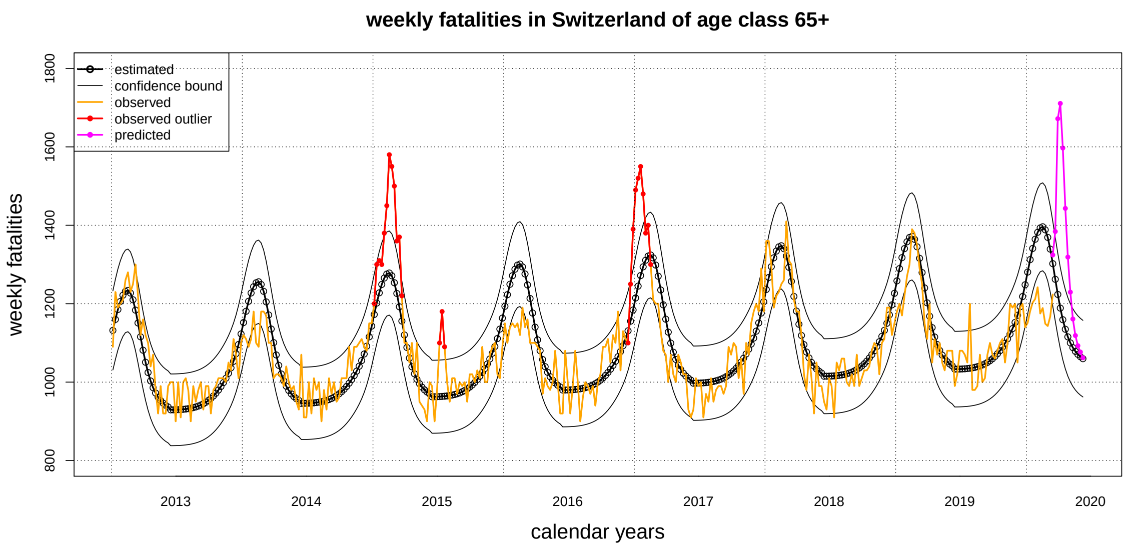

Figure 4: Switzerland: weekly fatality cases of age class 65+.

Finally, we set the predicted number of fatality cases in context to past observed mortality events in Switzerland.[5] Figure 4 shows the weekly numbers of fatalities of age class 65+ from calendar year 2013 until today. The black line shows the expected number of cases[6] (including a confidence bound of 3 standard deviations based on a Poisson assumption), and the orange line shows the effectively observed numbers. In red we highlight the outliers, being an influenza in 2015 and one in 2017. These two influenza caused 2'000 and 1'500 extra deaths among people above age 65. In magenta color we plot the current corona prediction. It is clear that the corona crisis is an outlier, but not a very extreme one in view of 2015 and 2017. However, we need to do an important remark concerning the corona situation: for sure the corona virus is more dangerous than past viruses, and it propagates faster than what has been experienced in past flues. The magenta prediction has been achieved under the current lockdown situation, and the picture would look rather differently if we would not have such strong measures by the Swiss government trying to flatten the curve.

[5]We compiled the data from `https://www.bfs.admin.ch/bfs/de/home/statistiken/bevoelkerung/geburten-todesfaelle.html`

[6]We received this smoothed expectation using a neural network approach.

Closure.
We have predicted duration and peak of the propagation phase, and we have partitioned predicted confirmed cases to different age classes. Moreover, we have estimated the case fatality rate which has allowed us to predict the fatality ratio per age class in Switzerland.
It is very clear that we have gone very far to receive these predictions. They are based on limited data and we have used coarse methods. We are fully aware of these deficiencies and we are open to any criticism that helps us to improve this study.
Nevertheless, we believe that our thoughts in this direction are important, and hopefully helpful for government, politicians, scientists and the public. It is very difficult to take the right measures and actions in this time of uncertainty where one must carefully weight all interests. Moreover, from the media we often hear about single cases that have deteriorated, we aim at providing a global picture here that does not focus on individuals, but considers the population as a whole.

A Supplementary data

Listing 1: Supplementary information used (compiled from internet)

```
  Country Population AreaCode  Size   LockDown Dens
      CHN       71.0     Asia 185.0 2020-01-23  384
      ITA       60.0   Europe 294.0 2020-03-09  204
      ESP       46.0   Europe 499.0 2020-03-14   92
      FRA       64.0   Europe 640.0 2020-03-17  100
      GER       83.0   Europe 348.0       <NA>  239
      SUI        8.5   Europe  40.0 2020-03-16  212
       UK       67.0       UK 241.0 2020-03-20  278
      NED       17.0   Europe  34.0       <NA>  500
      BEL       11.0   Europe  30.0 2020-03-18  367
      AUT        9.0   Europe  82.0 2020-03-15  110
       ZH        1.5   Europe   1.7 2020-03-16  882
```

B Gompertz' law

In this section we mathematically describe the model used in Section 3 to predict the newly confirmed cases. Starting point is the following recursive form

$$\log C_t = \log C_{t-1} + a \exp\{-bt\}. \tag{B.1}$$

For $b = 0$ and $a > 0$ we receive exponential growth, and for $b > 0$ and $a > 0$ we receive sub-exponential growth. In particular, on the original scale we have

$$C_t = C_{t-1} \exp\left\{ a e^{-bt} \right\}.$$

Using iteration and the fact that we have a geometric series, we can reformulate the previous identity as follows

$$C_t = C_0 \exp\left\{ \sum_{s=1}^{t} a e^{-bs} \right\} = C_0 \exp\left\{ \frac{a e^{-b}}{1 - e^{-b}} \left(1 - e^{-bt}\right) \right\}.$$

This is exactly the continuous time model derived in Moriconi [2]. Reparametrization of this function shows that we exactly use Gompertz' law [1].
We have limiting behavior for $b > 0$, see also formula (6) in Moriconi [2],

$$\lim_{t\to\infty} C_t = C_0 \exp\left\{\frac{ae^{-b}}{1-e^{-b}}\right\}.$$

The newly confirmed cases can be calculated as

$$Y_t = C_t - C_{t-1} = C_0 \exp\left\{\frac{ae^{-b}}{1-e^{-b}}\left(1-e^{-bt}\right)\right\}\left(1-\exp\left\{-ae^{-bt}\right\}\right) \;>\; 0.$$

Similarly to Moriconi [2] we determine the peak of the newly confirmed cases Y_t by calculating the maximum in t of Y_t. Expanding our discrete time function to continuous time we receive requirement

$$\frac{d}{dt}Y_t = C_0 \exp\left\{\frac{ae^{-b}}{1-e^{-b}}\left(1-e^{-bt}\right)\right\} abe^{-bt}\left[\frac{1}{e^b-1} - \frac{e^b}{e^b-1}\exp\left\{-ae^{-bt}\right\}\right] \;\stackrel{!}{=}\; 0.$$

This provides critical (continuous) time of the peak

$$t^* = -\frac{1}{b}\log(b/a),$$

and a peak height of

$$Y_{t^*} = C_0 \exp\left\{\frac{a-b}{e^b-1}\right\}\left(1-e^{-b}\right).$$

There remains the discussion of parameter estimation. To this end we come back to (B.1)

$$\log C_t = \log C_{t-1} + a\exp\{-bt\} = \log C_0 + \sum_{s=1}^{t}\exp\{\log a - bs\} = \log C_0 + \sum_{s=1}^{t}\beta_s,$$

where we set $\log\beta_s = \log a - bs$. The latter exactly reflects the linear regression with intercept $\log a$ and slope $-b$ studied in Figure 2 (middle), and $(\beta_T)_T$ are approximated by 5-days regression slopes on $(\log C_t)_{T-4\le t\le T}$.

Acknowledgment. I would like to thank Hans Bühlmann, Paul Embrechts, Franco Moriconi and Daniel Meier for very useful comments that have substantially improved my analysis.

References

[1] Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society* **115**, 513-585.

[2] Moriconi, F. (2020). A model with exponentially decreasing intensity for COVID-19 epidemic outbreak. Preprint of Version April 13, 2020.

C Plots of selected European countries and Kanton Zurich

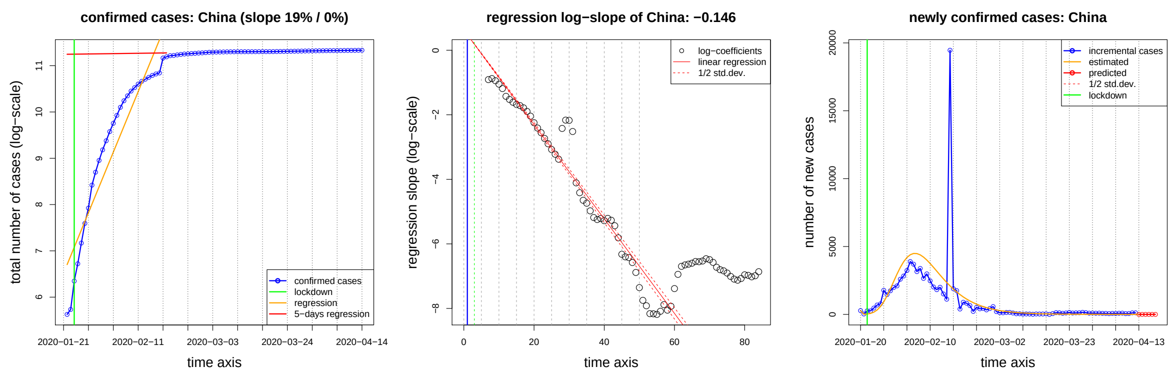

Figure 5: China: (lhs) confirmed cases on log-scale (in blue) and linear regression (entire time range in orange, last 5 days in red); (middle) logarithm of 5-days regression coefficients (black) and linear regression (red); (rhs) newly confirmed cases (blue) and predicted new cases (red); as of April 13, 2020.

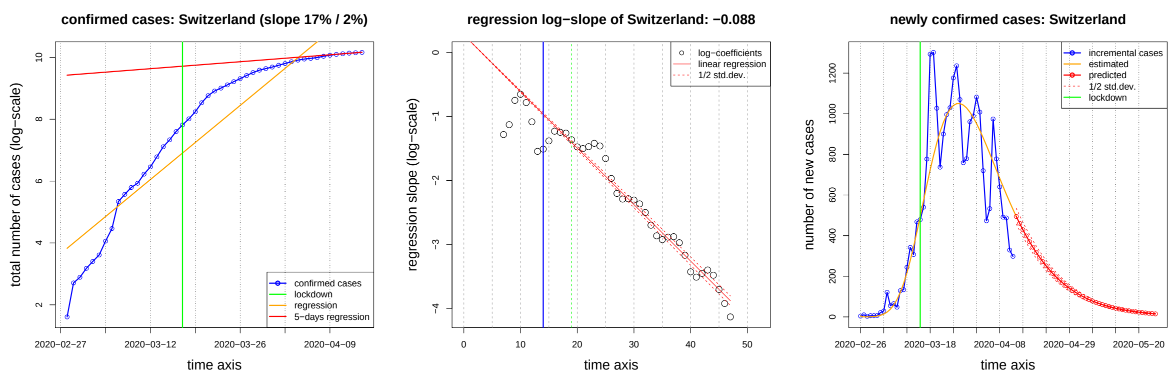

Figure 6: Switzerland: (lhs) confirmed cases on log-scale (in blue) and linear regression (entire time range in orange, last 5 days in red); (middle) logarithm of 5-days regression coefficients (black) and linear regression (red); (rhs) newly confirmed cases (blue) and predicted new cases (red); as of April 13, 2020.

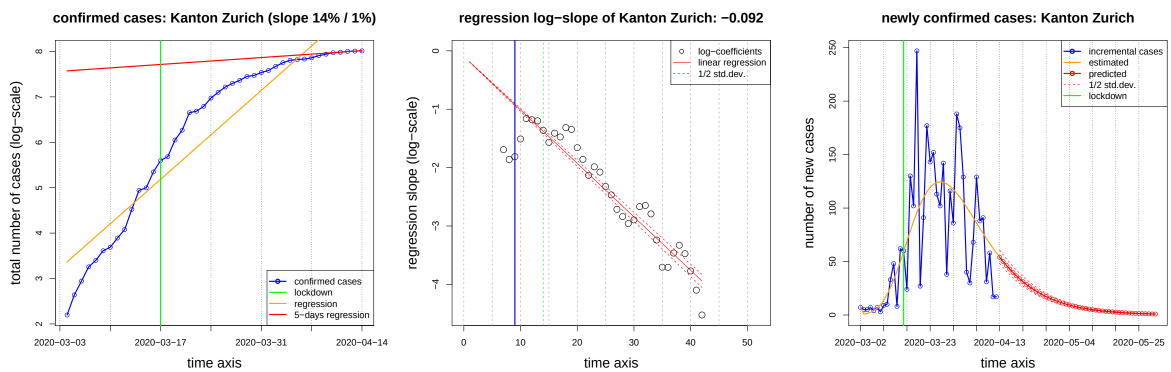

Figure 7: Kanton Zurich: (lhs) confirmed cases on log-scale (in blue) and linear regression (entire time range in orange, last 5 days in red); (middle) logarithm of 5-days regression coefficients (black) and linear regression (red); (rhs) newly confirmed cases (blue) and predicted new cases (red); as of April 13, 2020.

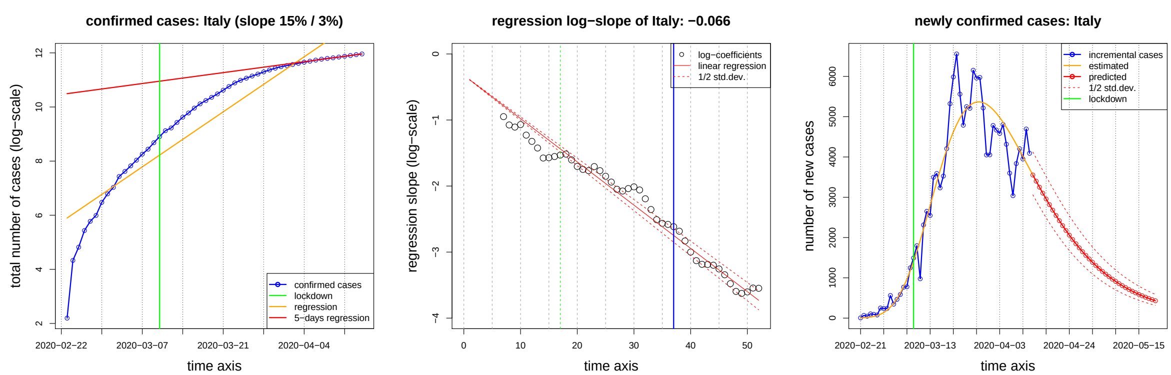

Figure 8: Italy: (lhs) confirmed cases on log-scale (in blue) and linear regression (entire time range in orange, last 5 days in red); (middle) logarithm of 5-days regression coefficients (black) and linear regression (red); (rhs) newly confirmed cases (blue) and predicted new cases (red); as of April 13, 2020.

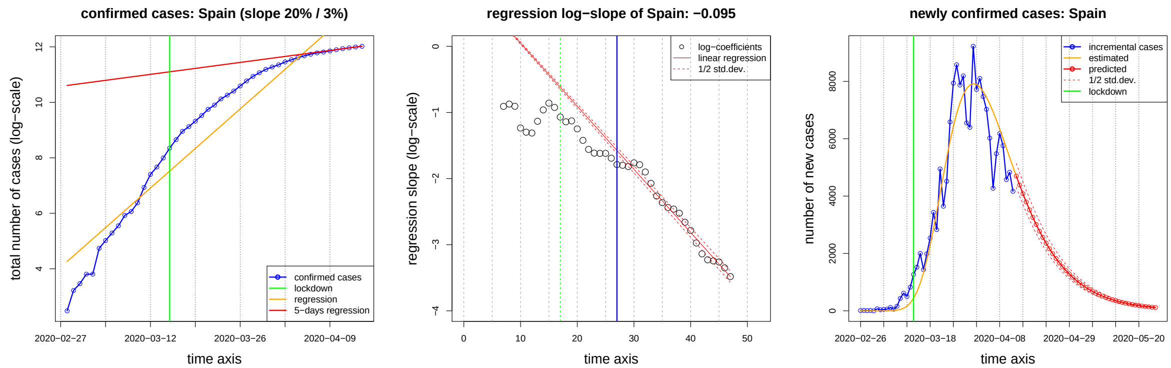

Figure 9: Spain: (lhs) confirmed cases on log-scale (in blue) and linear regression (entire time range in orange, last 5 days in red); (middle) logarithm of 5-days regression coefficients (black) and linear regression (red); (rhs) newly confirmed cases (blue) and predicted new cases (red); as of April 13, 2020.

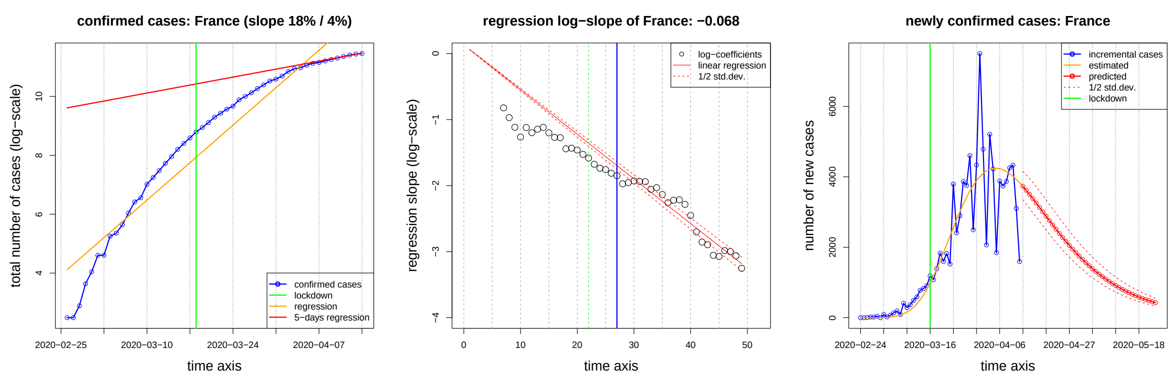

Figure 10: France: (lhs) confirmed cases on log-scale (in blue) and linear regression (entire time range in orange, last 5 days in red); (middle) logarithm of 5-days regression coefficients (black) and linear regression (red); (rhs) newly confirmed cases (blue) and predicted new cases (red); as of April 13, 2020.

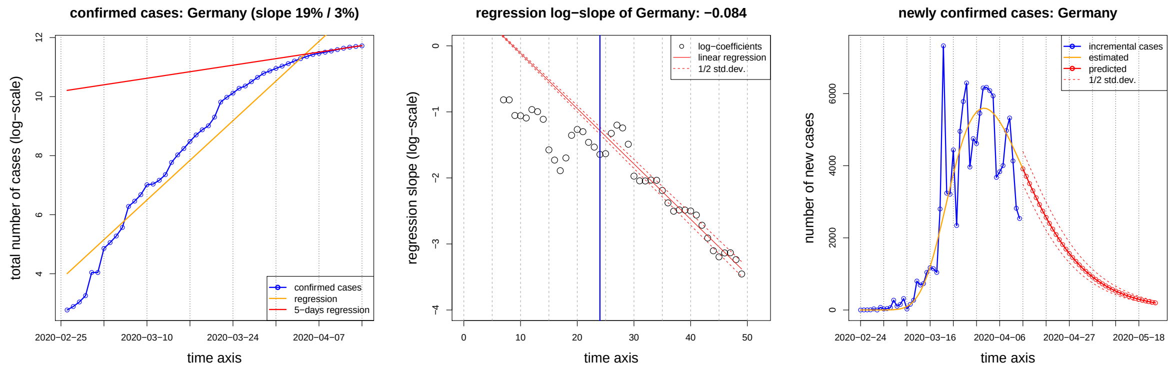

Figure 11: Germany: (lhs) confirmed cases on log-scale (in blue) and linear regression (entire time range in orange, last 5 days in red); (middle) logarithm of 5-days regression coefficients (black) and linear regression (red); (rhs) newly confirmed cases (blue) and predicted new cases (red); as of April 13, 2020.

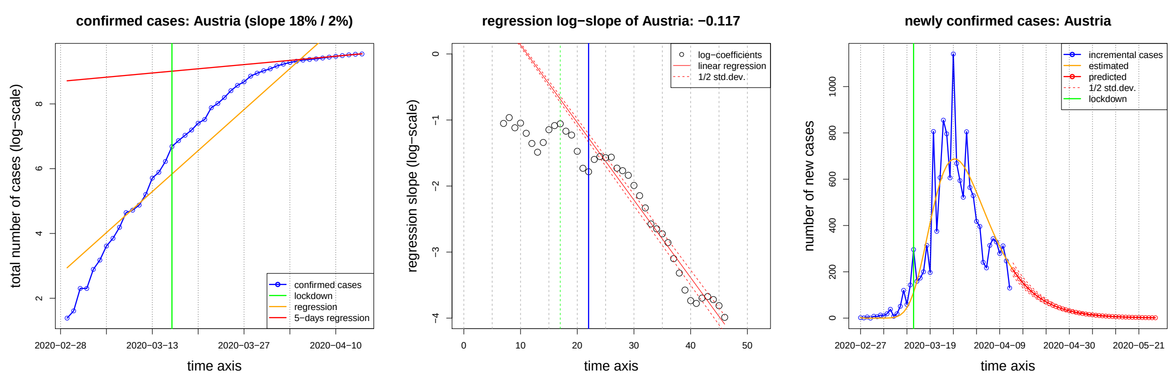

Figure 12: Austria: (lhs) confirmed cases on log-scale (in blue) and linear regression (entire time range in orange, last 5 days in red); (middle) logarithm of 5-days regression coefficients (black) and linear regression (red); (rhs) newly confirmed cases (blue) and predicted new cases (red); as of April 13, 2020.

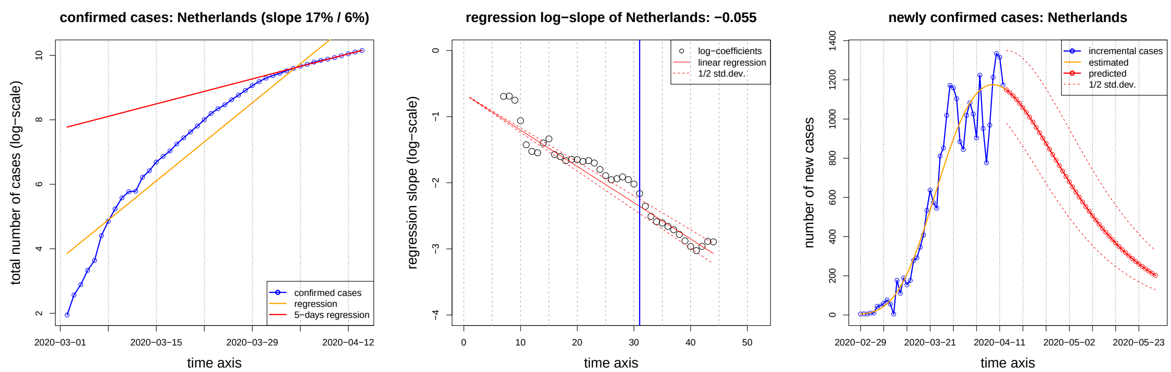

Figure 13: The Netherlands: (lhs) confirmed cases on log-scale (in blue) and linear regression (entire time range in orange, last 5 days in red); (middle) logarithm of 5-days regression coefficients (black) and linear regression (red); (rhs) newly confirmed cases (blue) and predicted new cases (red); as of April 13, 2020.

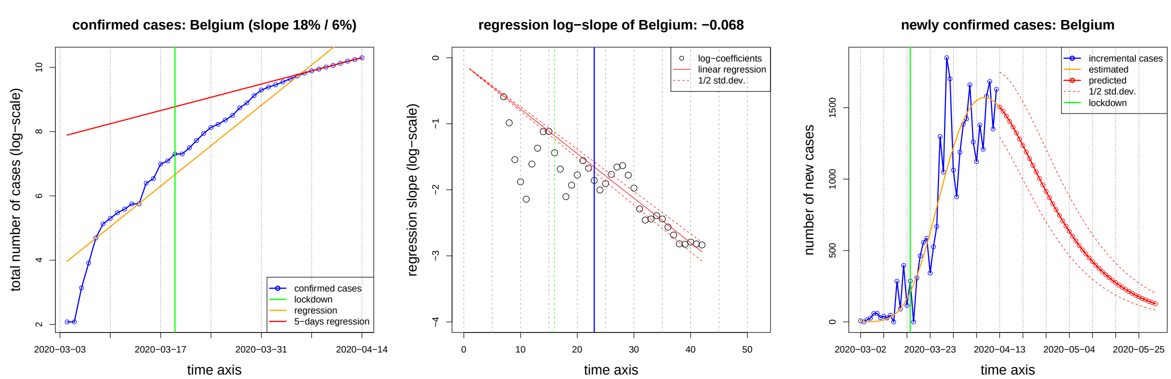

Figure 14: Belgium: (lhs) confirmed cases on log-scale (in blue) and linear regression (entire time range in orange, last 5 days in red); (middle) logarithm of 5-days regression coefficients (black) and linear regression (red); (rhs) newly confirmed cases (blue) and predicted new cases (red); as of April 13, 2020.

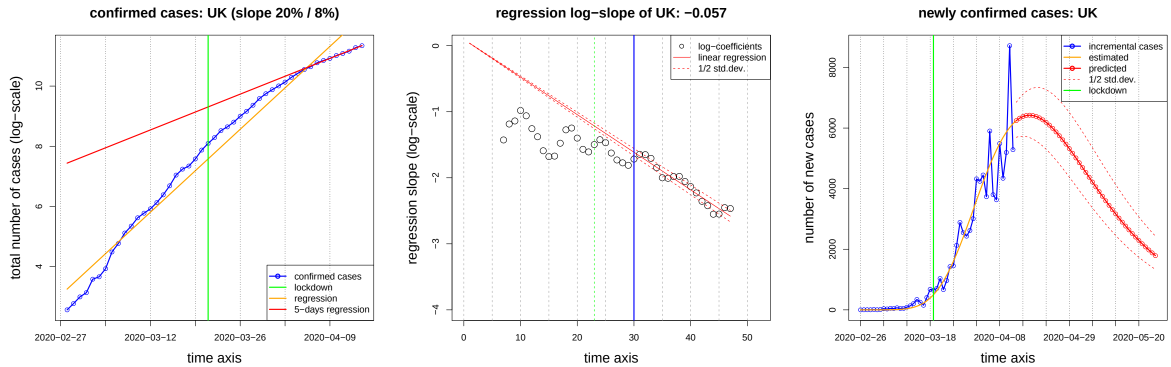

Figure 15: United Kingdom: (lhs) confirmed cases on log-scale (in blue) and linear regression (entire time range in orange, last 5 days in red); (middle) logarithm of 5-days regression coefficients (black) and linear regression (red); (rhs) newly confirmed cases (blue) and predicted new cases (red); as of April 13, 2020.

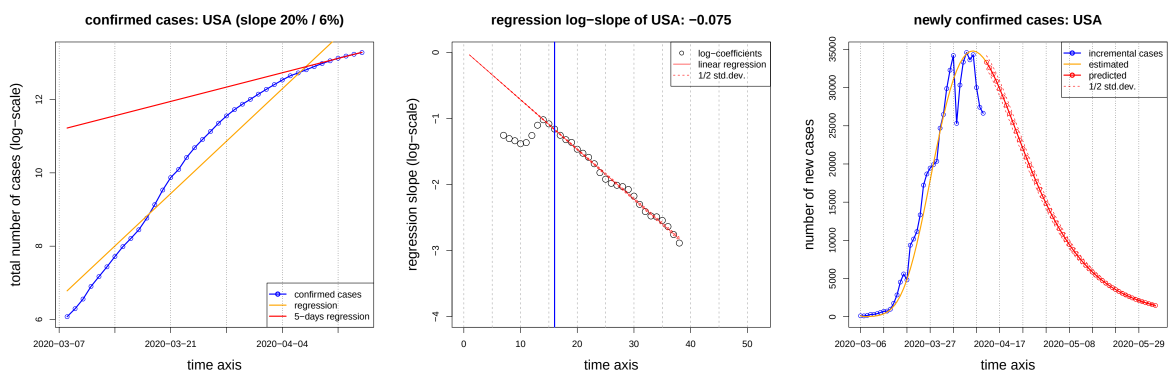

Figure 16: United States of America: (lhs) confirmed cases on log-scale (in blue) and linear regression (entire time range in orange, last 5 days in red); (middle) logarithm of 5-days regression coefficients (black) and linear regression (red); (rhs) newly confirmed cases (blue) and predicted new cases (red); as of April 13, 2020; the data is taken from `https://www.worldometers.info/coronavirus/country/us/`.

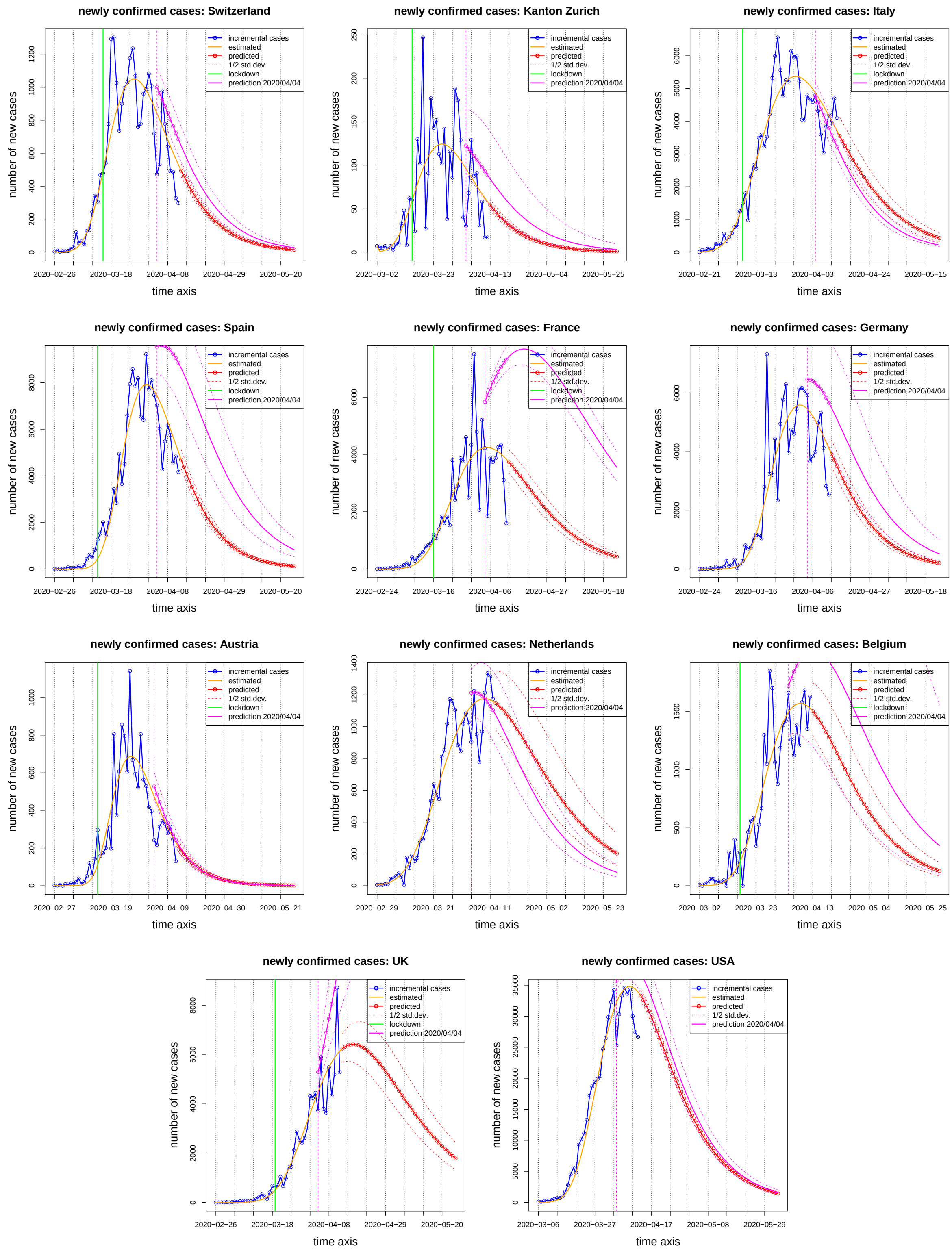

Figure 17: Comparison of predictions based on data of April 4, 2020 (magenta), and of April 13, 2020 (red).

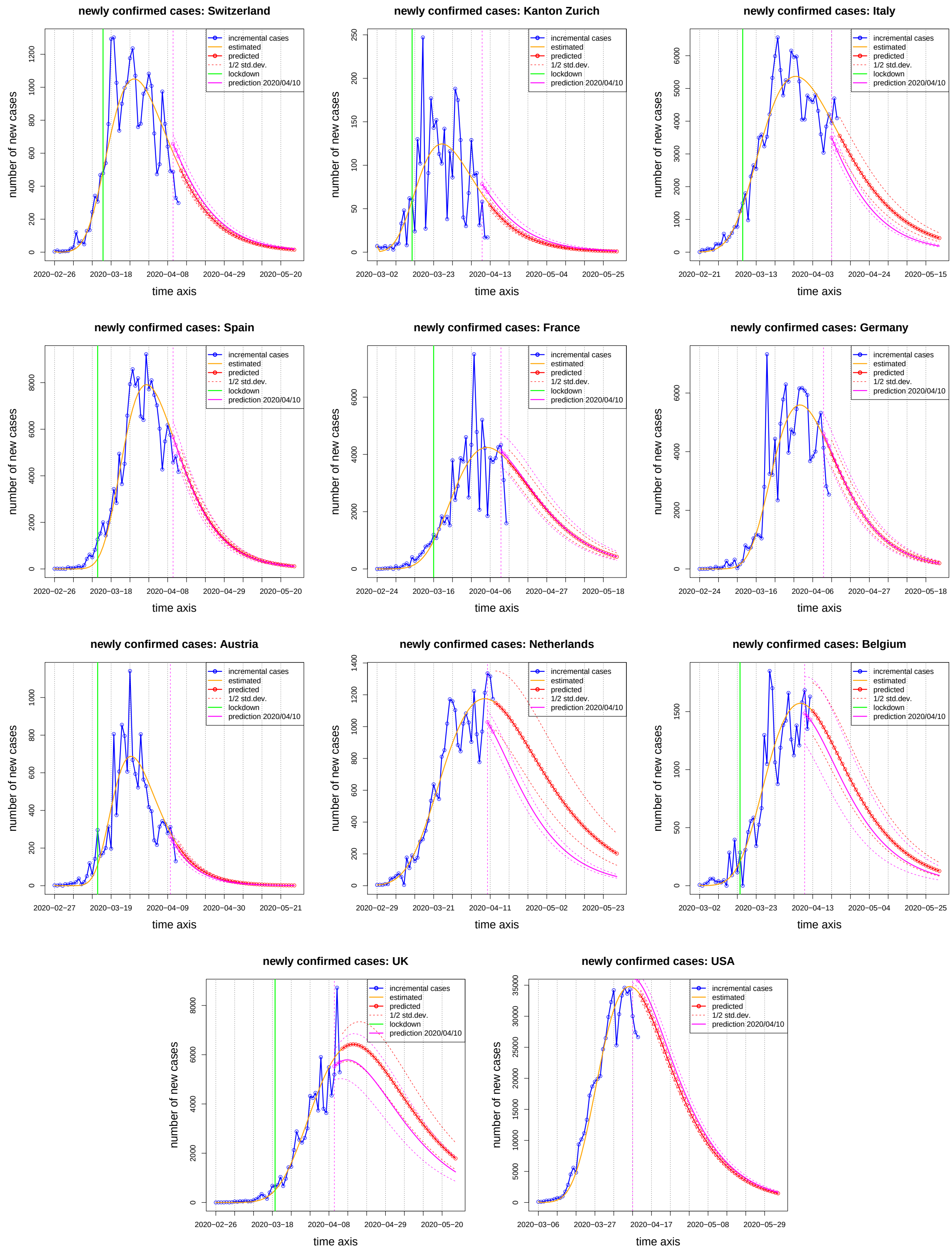

Figure 18: Comparison of predictions based on data of April 10, 2020 (magenta), and of April 13, 2020 (red).