Removing weekly administrative noise in the daily count of COVID-19 new cases.
Application to the computation of Rt

Luis Alvarez[1], Miguel Colom[2] and Jean-Michel Morel[2]

[1] CTIM. Departamento de Informática y Sistemas,
Universidad de Las Palmas de Gran Canaria. Spain

[2] Université Paris-Saclay, ENS Paris-Saclay, France

Abstract

The way each country counts and reports the incident cases of SARS-CoV-2 infections is strongly affected by the "weekend effect". During the weekend, fewer tests are carried out and there is a delay in the registration of cases. This introduces an "administrative noise" that can strongly disturb the calculation of trend estimators such as the effective reproduction number $R(t)$. In this work we propose a procedure to correct the incidence curve and obtain a better fit between the number of infected and the one expected using the renewal equation. The classic way to deal with the administrative noise is to invoke its weekly period and therefore to filter the incidence curve by a seven days sliding mean. Yet this has three drawbacks: the first one is a loss of resolution. The second one is that a 7-day mean filter hinders the estimate of the effective reproduction number $R(t)$ in the last three days before present. The third drawback of a mean filter is that it implicitly assumes the administrative noise to be additive and time invariant. The present study supports the idea that the administrative is better dealt with as being both periodic and multiplicative. The simple method that derives from these assumptions amount to multiplying the number of infected by a correcting factor which depends on the day of the week. This correcting factor is estimated from the incidence curve itself. The validity of the method is demonstrated by its positive impact on the accuracy of an the estimates of $R(t)$. To exemplify the advantages of the multiplicative periodic correction, we apply it to Sweden, Germany, France and Spain. We observe that the estimated administrative noise is country dependent, and that the proposed strategy manages to reduce it noise considerably. An implementation of this technique is available at `www.ipol.im/ern`, where it can be tested on the daily incidence curves of an extensive list of states and geographic areas provided by the European Centre for Disease Prevention and Control.

1

Keywords: COVID-19, Weekly administrative noise; Effective Reproduction number; reproduction rate; R0; Rt; SARS-CoV-2; Serial Interval.

Abbreviations:

EpiEstim : Software to compute the effective reproduction number proposed by Cori et al. in the paper: *A new framework and software to estimate time-varying reproduction numbers during epidemics* published in the American Journal of Epidemiology.

$R(t)$: Effective Reproduction Number. To differentiate between the continuous and discrete cases, we use the notation $R(t)$ in the continuous case and R_t in the discrete case.

$i(t)$: incidence curve, the number of daily tested positive registered. To differentiate between the continuous and discrete cases, we use the notation $i(t)$ in the continuous case and i_t in the discrete case.

Φ: serial interval.

The effective reproduction number $R(t)$ is one of the most important epidemiological characteristics of the COVID-19 pandemic. It is constantly invoked by politicians and their scientific advisers to steer the social distancing measures. $R(t)$ represents the expected number of secondary cases produced by a primary case at each time t. It can be computed from the incidence curve $i(t)$ and the serial interval Φ, which is the empirical probability distribution of the time between the onset of symptoms in a primary case and the onset of symptoms in secondary cases. There are different strategies to estimate $R(t)$ from $i(t)$ and Φ: EpiEstim, the method proposed in [4] is widely used. In the online interface available at `www.ipol.im/ern`, we compare our estimate of $R(t)$ with the one of EpiEstim. In [11] a technique to compute $R(t)$ separating local transmission and imported cases is proposed. In [6] the authors make a detailed comparison of the methods proposed in [4], [2] and [13]. A systematic review and analysis of the serial interval is presented in [12]. In [10], [5], [7] different statistical distributions of the serial interval Φ for the SARS-CoV-2 are proposed.

The so-called renewal equation (see [9]) is a key epidemiological model linking the daily count of new detected cases of infections, $i(t)$, with the reproductive power, $A(t, s)$, at time t and infection-age s at which an infected individual generates secondary cases. $A(t, s)$ depends on $R(t)$ and $\Phi(s)$ and the renewal equation can be expressed as

$$i(t) = F(i, R, \Phi, t) = \int_0^\infty i(t-s)A(t,s)ds \qquad \text{for} \ \ t = 0, .., t_c, \tag{1}$$

where t_c represents the current time (the last time at which $i(t)$ is available). For instance, in [4] and [3] the following formulation of the renewal equation is used to compute $R(t)$ from $i(t)$ and $\Phi(s)$:

$$i(t) = F_1(i, R, \Phi, t) = R(t) \int_0^\infty i(t-s)\Phi(s)ds. \tag{2}$$

This model is a simplification of the Nishiura equation [8]:

$$i(t) = F_2(i, R, \Phi, t) = \int_{-\infty}^\infty R(t-s)i(t-s)\Phi(s)ds. \tag{3}$$

In the original formulation of this model it was assumed that $\Phi(s) = 0$ for $s \le 0$. However, in the case of SARS-CoV-2 a patient can show symptoms before the patient who infected him/her shows symptoms him/herself. This means that, actually, $\Phi(s)$ can be positive for $s < 0$. In [1], the above model is used without any restriction on the support of $\Phi(s)$. Since measurements are generally made daily, a discrete formulation of model (3) is sound. It can be expressed as

$$i_t = \sum_{s=f_0}^f i_{t-s} R_{t-s} \Phi_s \tag{4}$$

where f_0 is, in general, negative.

In the technique we propose in this work, either of the above expressions of the renewal equation can be used. In the experiments presented, we use the model $F_2(i, R, \Phi, t)$, and the method developed in [1] to estimate R_t from i_t and Φ_s. This method is based on the the minimization of the following energy:

$$E(\{R_t\}) = \sum_{t=0}^{t_c} \left(\frac{i_t - \sum_{s=f_0}^f i_{t-s} R_{t-s} \Phi_s}{p_{90}(i)} \right)^2 + \sum_{t=1}^{t_c} w_t (R_t - R_{t-1})^2 + \sum_{m=0}^M \beta_m (R_{t_m} - \bar{R}_{t_m})^2, \tag{5}$$

where $p_{90}(i)$ is the 90th percentile of $\{i_t\}_{t=0,..,t_c}$ used to normalize the energy with respect to the size of i_t. The first term of E is a data adjustment term which forces the renewal equation (4) to be satisfied as much as possible. The second term forces R_t to be a smooth curve; $w_t \geq 0$ represents the weight of the regularization at each time t. The higher the value of w_t the smoother R_t. The last term of E forces R_{t_m} to be close to an initial estimate given by $\bar{R}_{t_m}$ for some particular times t_m. Roughly speaking, minimizing the energy E leads to satisfy approximately the renewal equation (4) with a reasonably smooth R_t and, optionally, prescribed initial values for some particular times t_m.

The daily number of new detected cases, i_t, is strongly affected by the "weekend effect". During the weekend, fewer tests are carried out and there is a delay in the registration of cases. It follows that the actual number of cases is systematically underestimated in some days of the week and overestimated in others. The usual way to deal with this weekly administrative noise is to use a 7-day moving average of i_t, but this procedure negatively affects the accuracy of the point estimate of the trend of i_t when approaching the current date and forces stopping the estimate three days before present.

The main assumption of this paper is that a significant part of the discrepancy between i_t and its expected value $F(i, R, \Phi, t)$ is given by the weekend effect. We therefore assume that the quotient

$$q(i, R, \Phi, t) = \frac{F(i, R, \Phi, t)}{i_t}, \qquad (6)$$

follows a 7-day periodic-dynamic, that is, $q(i, R, \Phi, t) \approx q(i, R, \Phi, t - 7)$. In other words, the ratio between $F(i, R, \Phi, t)$ and i_t depends mainly on the day of the week. In the experiments shown in this article, it will be observed that, indeed, $q(i, R, \Phi, t)$ follows accurately this periodic pattern in several countries[1]. For example, in Sweden every Monday the registered value i_t is systematically underestimated and on Saturdays the opposite effect occurs. So we propose to approximate $q(i, R, \Phi, t)$ using a 7-day periodic function given by the vector $\hat{\mathbf{q}}(i, R, \Phi) = (\hat{q}_0, \hat{q}_1, \hat{q}_2, \hat{q}_3, \hat{q}_4, \hat{q}_5, \hat{q}_6)$. Therefore $q(i, R, \Phi, t) \approx \hat{q}_{t\%7}$ where the symbol % represents the modulo operation, that is, the remainder of the division between two numbers.

To compute $\hat{\mathbf{q}}(i, R, \Phi)$ we proceed to a least square estimate

$$\hat{\mathbf{q}}(i, R, \Phi) = \underset{\mathbf{q}=(q_0,q_1,q_2,q_3,q_4,q_5,q_6)}{\arg\min} E(i, R, \Phi, \mathbf{q}) \equiv \sum_{t=t_c-T+1}^{t_c} \left(F(i, R, \Phi, t) - q_{t\%7} i_t \right)^2, \qquad (7)$$

where T represents the number of days used in the estimation (in our experiments we use $T = 56$, that is 8 weeks). We point out that the value $\hat{i}_t = \hat{q}_{t\%7} i_t$ can be considered as an update of i_t where we have removed the weekly administrative noise. To preserve the number of accumulated cases in the period of estimation, we add to the minimization problem (7) the constraint

$$\sum_{t=t_c-T+1}^{t_c} i_t = \sum_{t=t_c-T+1}^{t_c} \hat{i}_t. \qquad (8)$$

In that way, the multiplication by the factor $\hat{q}_{t\%7}$ produces a redistribution of the cases i_t during the period of estimation, but it does not change the global amount of cases. Notice that we could add weights to some particular days in the expression of $E(i, R, \Phi, \mathbf{q})$ in (7). For example,

---

[1]To test this assumption on many more countries, the reader is invited to the online demo `www.ipol.im/ern`

if a day is a holiday in the middle of the week, one might reduce, in the energy $E(i, R, \Phi, \mathbf{q})$, the weight of that day and the following ones.

Once i_t is updated using $\hat{i}_t = \hat{q}_{t\%7} i_t$, we can use $\hat{i}_t$ to recompute R_t using the renewal equation (1). We denote by $\hat{R}_t$ the updated version of R_t. This whole procedure is repeated to update iteratively $\hat{i}_t$ and $\hat{R}_t$ until convergence. The final value $\hat{i}_t$ and $\hat{R}_t$ provides a more realistic trend of the evolution of i_t and improves the estimation of R_t. The final vector $\hat{\mathbf{q}}(i, \hat{R}, \Phi)$ represents the set of multiplicative factors used to remove the weekly administrative noise. In Fig. 1 we present a flowchart of the whole procedure and in the appendix we give technical details.
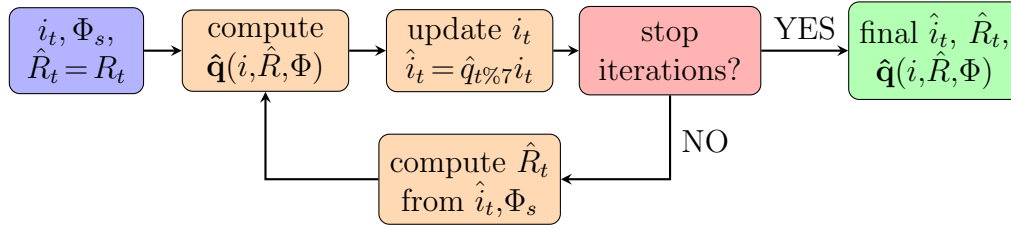

Figure 1: Flowchart of the estimation of $\hat{i}_t$, $\hat{R}_t$ and $\hat{\mathbf{q}}(i, \hat{R}, \Phi)$. First R_t is computed from i_t and Φ_s using the method proposed in [1] and $\hat{R}_t = R_t$ is initialized. Next, $\hat{\mathbf{q}}(i, \hat{R}, \Phi)$ is obtained by minimizing (7) with the constraint (8). Then i_t is updated as $\hat{i}_t = \hat{q}_{t\%7} i_t$. The iteration is stopped when the efficiency measure $\mathcal{I}$ defined in formula (9) does not improve in the current iteration. Otherwise, $\hat{R}_t$ is updated by the method proposed in [1] from $\hat{i}_t$ and Φ_s and the iteration goes on.

RESULTS

All of the experiments made here can be reproduced with the online interface available at `www.ipol.im/ern`. In the appendix, we have included some details about the use of this online demo. To measure how well the removal of the weekly administrative noise improves the explanation of i_t by the renewal equation $F(i, R, \Phi, t)$, we use the following efficiency measure:

$$\mathcal{I} = \sqrt{\frac{E(\hat{i}, \hat{R}, \Phi, \hat{\mathbf{q}}(\hat{i}, \hat{R}, \Phi))}{E(i, R, \Phi, \hat{\mathbf{q}}(i, R, \Phi))}}, \tag{9}$$

where $E(.)$ is defined in (7). $\mathcal{I}$ represents the reduction, after the removal of the weekly administrative noise, of the average distance between i_t and $F(i, R, \Phi, t)$. The smaller $\mathcal{I}$, the more efficient the noise reduction has been. In fact, the value of $\mathcal{I}$ can be used to assess whether it is worth applying the proposed method to a given country and in a given time interval.

In Fig. 2 we plot, for Sweden, Germany, France and Spain the values of the vector $\hat{\mathbf{q}}(i, \hat{R}, \Phi)$ obtained for each day of the week. We observe that in Sweden and Germany, Monday is the day of the week where the value of i_t is most underestimated (the higher $\hat{q}_k$ is, the more underestimated is i_t in the day k). However, in France, that day corresponds to Tuesday. This suggests that in France there is an additional delay of one day due to the way France records and reports the number of new cases. In these three countries, the effect of weekly administrative noise is mainly concentrated on Monday, Tuesday and Wednesday (with a 1-day
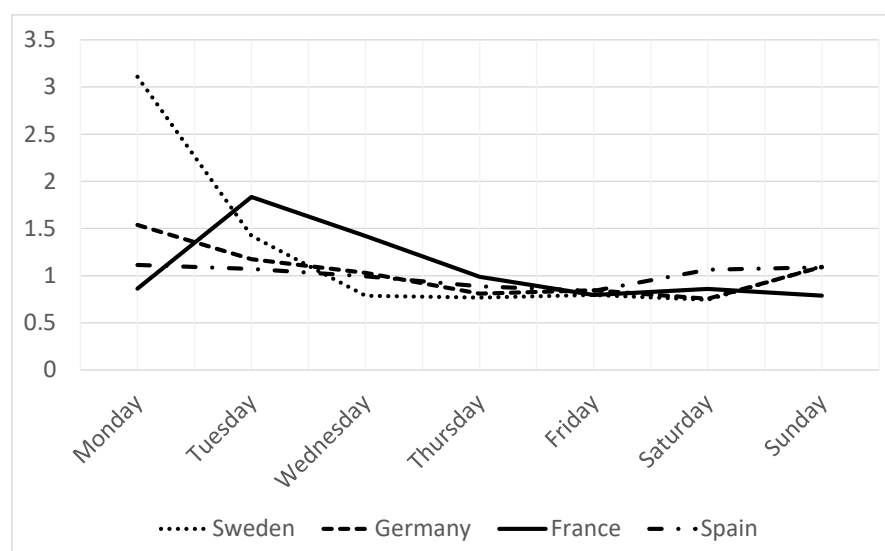
5

Figure 2: Plot of $\hat{\mathbf{q}}(i, \hat{R}, \Phi) = (\hat{q}_0, \hat{q}_1, \hat{q}_2, \hat{q}_3, \hat{q}_4, \hat{q}_5, \hat{q}_6)$ for different countries.

lag in the case of France). The case of Spain is special because it does not provide data on the weekend. In that case we have assumed that on Saturday, Sunday and Monday the number of cases is constant and equal to the accumulated number of cases on Saturday, Sunday and Monday divided by 3. Moreover, in Spain, some regions sometimes do not report data for one or several days, which produces an additional administrative noise different from the weekend effect. The more orderly the daily count of new infected, the better the result obtained by our method.

In table 1 we present the numerical values of the vector $\hat{\mathbf{q}}(i, \hat{R}, \Phi)$, the efficiency measure $\mathcal{I}$ and the effective reproduction number in the current time before and after the weekly administrative noise removal given by R_{t_c} and $\hat{R}_{t_c}$ respectively. Notably, and with the exception of Spain, the efficiency value $\mathcal{I}$ shows a high reduction of the administrative noise. At the end of the period chosen for the study, between September 9 and October 28, there is a strong expansion of the virus. This period ends on a Wednesday. In the previous days the number of cases was underestimated. It follows that the estimated effective reproduction number becomes considerably higher after eliminating the weekly administrative noise.

	Sweden	Germany	France	Spain
Monday $(\hat{q}_0)$	3.109167	1.537838	0.863379	1.113635
Tuesday $(\hat{q}_1)$	1.424005	1.175497	1.834212	1.071432
Wednesday $(\hat{q}_2)$	0.787374	1.029281	1.419457	0.993049
Thursday $(\hat{q}_3)$	0.766827	0.808254	0.989091	0.887398
Friday $(\hat{q}_4)$	0.795438	0.846798	0.794311	0.838609
Saturday $(\hat{q}_5)$	0.744447	0.754752	0.858157	1.063243
Sunday $(\hat{q}_6)$	1.100563	1.093202	0.788024	1.087866
$\mathcal{I}$	0.2526414	0.2717397	0.3725009	0.6439962
R_{t_c}	1.426663	1.470402	1.277142	1.225078
$\hat{R}_{t_c}$	1.509397	1.637230	1.563844	1.263821

Table 1: We present for Sweden, Germany, France and Spain the values of the vector $\hat{\mathbf{q}}(i, \hat{R}, \Phi)$ for each day of the week, the efficiency measure, $\mathcal{I}$, and the effective reproduction number in the current time before and after the administrative noise removal given by R_{t_c} and $\hat{R}_{t_c}$ respectively.

In Fig. 3 - 6 we plot on the one hand the quotient $q(i, \hat{R}, \Phi, t)$ and its periodic approximation $\hat{q}_{t\%7}$, and on the other hand the values of i_t, its update $\hat{i}_t$ after the removal of the weekly administrative noise and the expected value using $F_2(\hat{i}, \hat{R}, \Phi, t)$. For Sweden, Germany and France we observe a quite a good agreement between the quotient $q(i, \hat{R}, \Phi, t)$ and its periodic approximation $\hat{q}_{t\%7}$, which supports the validity of the proposed method and our assumption that the evolution of $q(i, \hat{R}, \Phi, t)$ follows a 7-day periodic dynamic. For these countries, $\hat{i}_t = \hat{q}_{7\%t} i_t$, the number of new cases after the removal of the administrative noise, is less oscillating than i_t and very close to its expected value following the renewal equation $F_2(\hat{i}, \hat{R}, \Phi, t)$. In the case of Spain, the obtained improvement is minor. On the one hand, Spain does not report data during the weekend, which introduces an additional disturbance in the data that hinders the application of the proposed method. In addition, in two of the weeks included in the period of the study, $q(i, \hat{R}, \Phi, t)$ diverges significantly from its periodic approximation $\hat{q}_{t\%7}$.
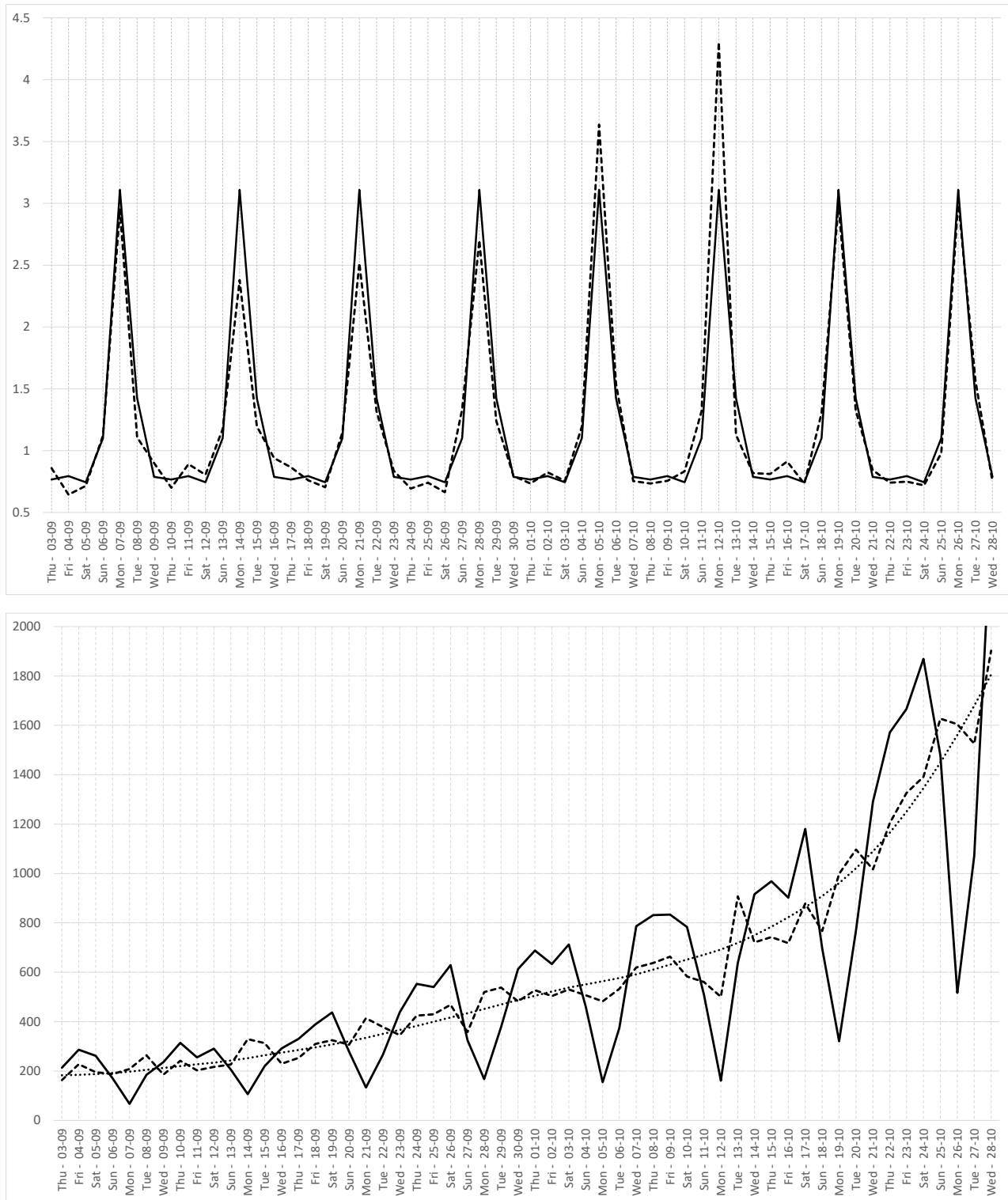
Figure 3: In the case of Sweden, we plot, between September 9 and October 28: **(up)** $q(i, \hat{R}, \Phi, t)$
(dotted line) and its periodic approximation $\hat{q}_{t\%7}$ (solid line). **(down):** i_t (solid line), its update
$\hat{i}_t$ (dashed line) removing the administrative noise and the expected value using $F_2(\hat{i}, \hat{R}, \Phi, t)$
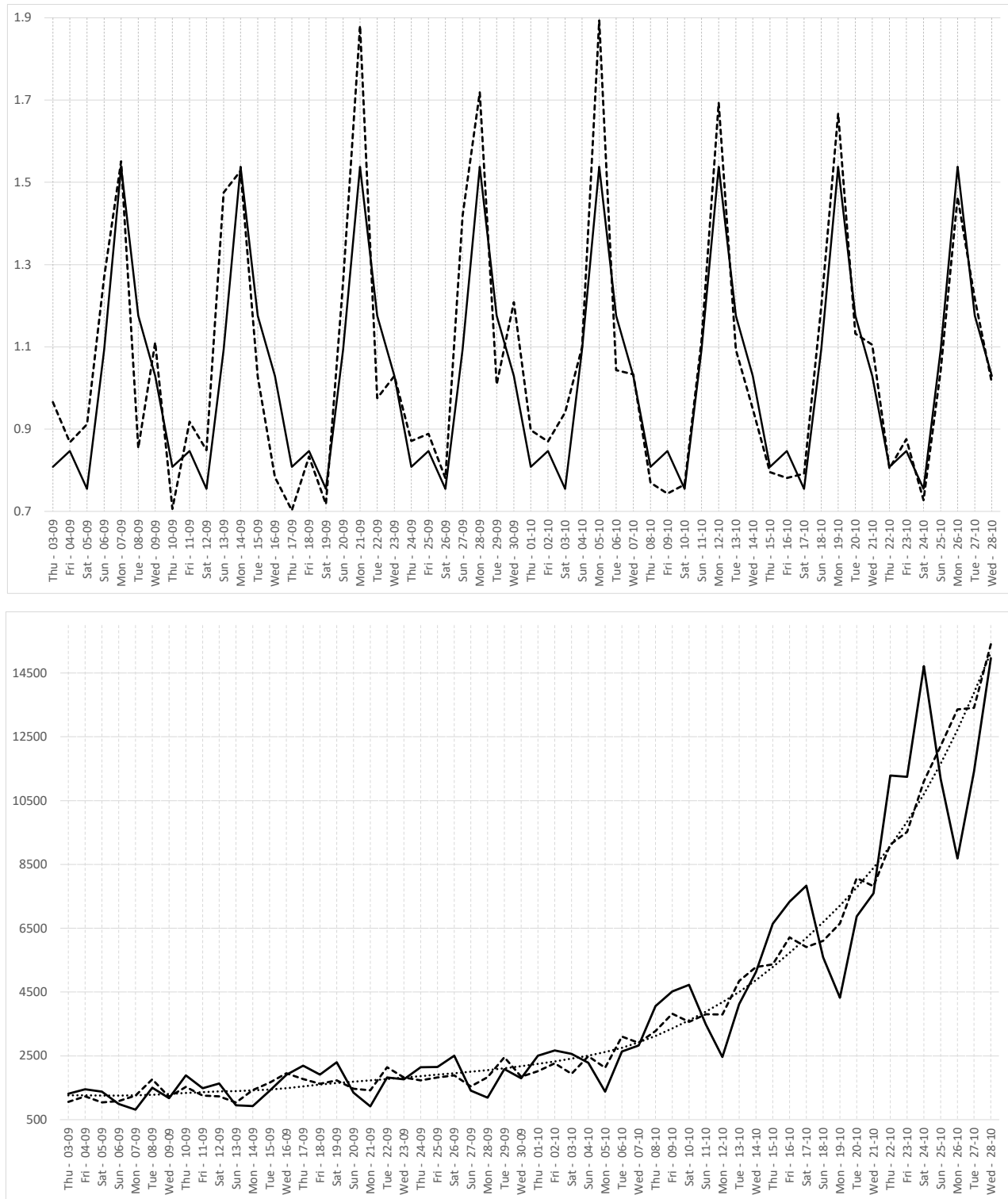(dotted line).

8

Figure 4: In the case of Germany, we plot, between September 9 and October 28: **(up)** $q(i, \hat{R}, \Phi, t)$ (dotted line) and its periodic approximation $\hat{q}_{t\%7}$ (solid line). **(down)**: i_t (solid line), its update $\hat{i}_t$ (dashed line) removing the administrative noise and the expected value using $F_2(\hat{i}, \hat{R}, \Phi, t)$ (dotted line).
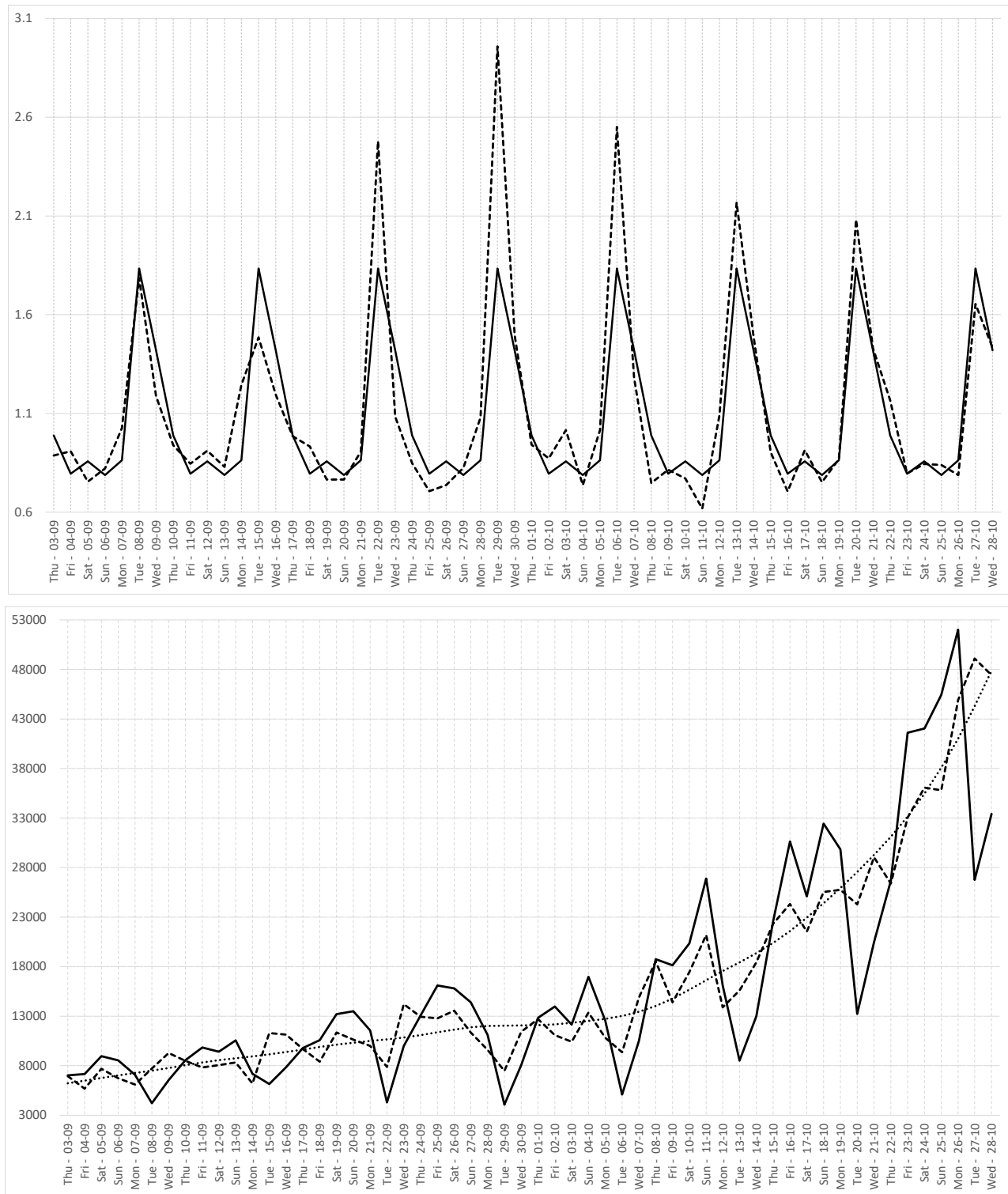
9

Figure 5: In the case of France, we plot, between September 9 and October 28: **(up)** $q(i, \hat{R}, \Phi, t)$ (dotted line) and its periodic approximation $\hat{q}_{t\%7}$ (solid line). **(down)**: i_t (solid line), its update $\hat{i}_t$ (dashed line) removing the administrative noise and the expected value using $F_2(\hat{i}, \hat{R}, \Phi, t)$ (dotted line).
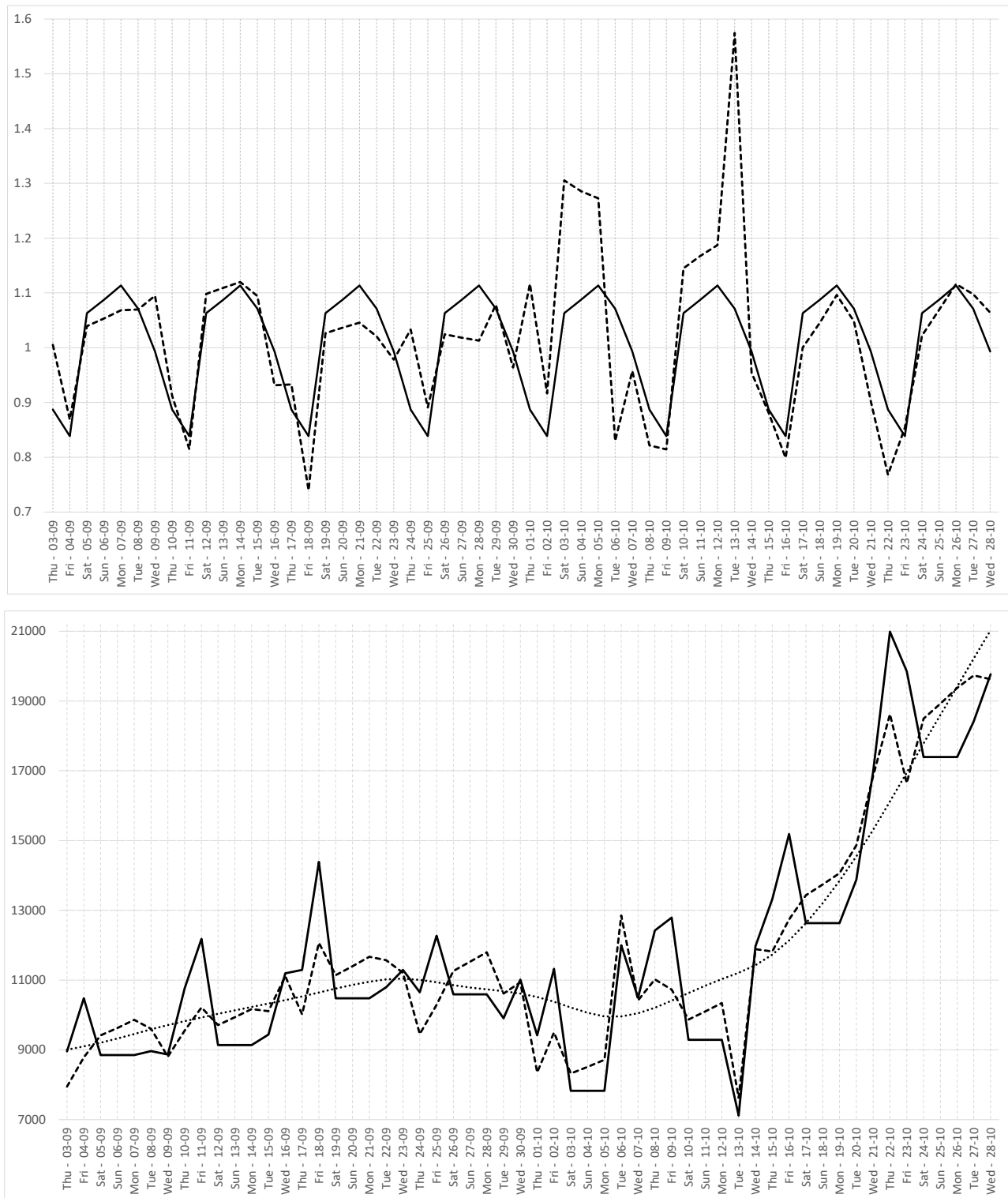
Figure 6: In the case of Spain, we plot, between September 9 and October 28: **(up)** $q(i, \hat{R}, \Phi, t)$ (dotted line) and its periodic approximation $\hat{q}_{t\%7}$ (solid line). **(down)**: i_t (solid line), its update $\hat{i}_t$ (dashed line) removing the administrative noise and the expected value using $F_2(\hat{i}, \hat{R}, \Phi, t)$ (dotted line).

11

1   CONCLUSION

We proposed a technique to remove the weekly administrative noise in the incidence curves of COVID-19, based on the improvement of the agreement between i_t and its expected value using the renewal equation $F(i, R, \Phi, t)$. The method boils down to multiplying i_t by a factor $\hat{q}_{t\%7}$ which depends on the day of the week. The main assumption supporting the validity of this approach is that the quotient of i_t and $F(i, R, \Phi, t)$ follows approximately a 7-day periodic dynamic. We verified this assumption on Sweden, Germany and France. In the case of Spain the method brings less improvement, due to the fact that Spain does not report data on weekends. The proposed method updates iteratively i_t and the effective reproduction number R_t. The number of cases after the removal of the weekly administrative noise $\hat{i}_t = \hat{q}_{t\%7} i_t$ is much less oscillating than i_t and very close to its expected value following the used renewal equation. From $\hat{i}_t$ we obtain a more accurate value of the effective reproduction number R_t. An implementation of this technique is available at `www.ipol.im/ern`. In this online interface we compare the final estimation of R_t with the estimate obtained by EpiEstim. As shown in [3], in practice, EpiEstim uses a 7-day moving average to remove the administrative weekly noise. As can be observed using the online interface, this 7-day moving average introduces a shift towards the past that is not observed in our estimate. Therefore our method seems to provide a more to date estimate of R_t than EpiEstim.

References

[1] L. ALVAREZ, M. COLOM, AND J.-M. MOREL, *A variational model for computing the effective reproduction number of SARS-CoV-2*, MedRxiv, (2020).

[2] L. BETTENCOURT AND R. RIBEIRO, *Real time bayesian estimation of the epidemic potential of emerging infectious diseases*, PLoS ONE, 3 (2008).

[3] T. Z. BOULMEZAOUD, L. ALVAREZ, M. COLOM, AND J.-M. MOREL, *A daily measure of the SARS-CoV-2 daily reproduction number for all countries*, IPOL Journal. Image Processing On Line, submitted, (2020).

[4] A. CORI, N. M. FERGUSON, C. FRASER, AND S. CAUCHEMEZ, *A new framework and software to estimate time-varying reproduction numbers during epidemics*, American journal of epidemiology, 178 (2013), pp. 1505–1512.

[5] Z. DU, X. XU, Y. WU, L. WANG, B. J. COWLING, AND L. A. MEYERS, *The serial interval of COVID-19 from publicly reported confirmed cases*, medRxiv, (2020).

[6] K. GOSTIC, L. MCGOUGH, E. BASKERVILLE, S. ABBOTT, K. JOSHI, C. TEDIJANTO, R. KAHN, R. NIEHUS, J. HAY, P. DE SALAZAR, J. HELLEWELL, S. MEAKIN, J. MUNDAY, N. BOSSE, K. SHERRATT, R. THOMPSON, L. WHITE, J. HUISMAN, J. SCIRE, S. BONHOEFFER, T. STADLER, J. WALLINGA, K. S. FUN, M. LIPSITCH, AND S. COBEY, *Practical considerations for measuring the effective reproductive number, Rt*, MedRxiv, (2020).

[7] S. MA, J. ZHANG, M. ZENG, Q. YUN, W. GUO, Y. ZHENG, S. ZHAO, M. H. WANG, AND Z. YANG, *Epidemiological parameters of coronavirus disease 2019: a pooled analysis of publicly reported individual data of 1155 cases from seven countries*, Medrxiv, (2020).

[8] H. NISHIURA, *Time variations in the transmissibility of pandemic influenza in Prussia, Germany, from 1918–19*, Theoretical Biology and Medical Modelling, 4 (2007), p. 20.

[9] H. NISHIURA AND G. CHOWELL, *The Effective Reproduction Number as a Prelude to Statistical Estimation of Time-Dependent Epidemic Trends*, Springer Netherlands, Dordrecht, 2009, pp. 103–121.

[10] H. NISHIURA, N. M. LINTON, AND A. R. AKHMETZHANOV, *Serial interval of novel coronavirus (covid-19) infections*, International journal of infectious diseases, (2020).

[11] R. THOMPSON, J. STOCKWIN, R. D. VAN GAALEN, J. POLONSKY, Z. KAMVAR, P. DEMARSH, E. DAHLQWIST, S. LI, E. MIGUEL, T. JOMBART, ET AL., *Improved inference of time-varying reproduction numbers during infectious disease outbreaks*, Epidemics, 29 (2019), p. 100356.

[12] M. A. VINK, M. C. J. BOOTSMA, AND J. WALLINGA, *Serial intervals of respiratory infectious diseases: a systematic review and analysis*, American journal of epidemiology, 180 (2014), pp. 865–875.

[13] J. WALLINGA AND P. TEUNIS, *Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures*, American Journal of epidemiology, 160 (2004), pp. 509–516.

Appendix

Technical issues

Here, we show some technical details of the proposed technique. First we notice that, in general, the only data we use is i_t and Φ_s. So initially, we do not know the actual day of the week which corresponds to each datum i_t. For the computation of $\hat{\mathbf{q}} = (\hat{q}_0, \hat{q}_1, \hat{q}_2, \hat{q}_3, \hat{q}_4, \hat{q}_5, \hat{q}_6)$, we will assume, without loss of generality, that the first day of the period we consider for the estimation, that is $t_c - T + 1$ corresponds to the time $t = 0$. This means that we are initially assuming that this day corresponds to a Monday, but in fact, this is not relevant because we can reorganize at any time the values of $\hat{\mathbf{q}}$ to fit the actual days of the week.

The minimization problem (7) given by the quadratic energy:

$$E(i, R, \Phi, \mathbf{q}) \equiv \sum_{t=0}^{t_c} \left(F(i, R, \Phi, t) - q_{t\%7} i_t \right)^2, \tag{10}$$

can be expressed in matrix form as

$$E(i, R, \Phi, \mathbf{q}) = \|A\mathbf{q} - \mathbf{b}\|^2, \tag{11}$$

13

where $\|.\|$ is the usual Euclidean norm, $A \in R^{T \times 7}$ is defined by

$$
A_{k,j} = \begin{cases} 0 & if \quad j \neq k\%7 \\ 1 & if \quad j = k\%7, \end{cases}
$$

$\mathbf{b} \in R^T$ is defined by $b_k = F(i, R, \Phi, k)$. It is well-known that the above minimization problem has a closed form solution given by

$$
\hat{\mathbf{q}} = (A^t A)^{-1} b^t A
$$

The constraint (8) given by

$$
\sum_{t=0}^{t_c} i_t = \sum_{t=0}^{t_c} \hat{i}_t = \sum_{t=0}^{t_c} q_{t\%7} i_t, \tag{12}
$$

can be expressed as the following additional linear equation:

$$
\mu_0 q_0 + \mu_1 q_1 + \mu_2 q_2 + \mu_3 q_3 + \mu_4 q_4 + \mu_5 q_5 + \mu_6 q_6 = \sum_{t=0}^{t_c} i_t
$$

where

$$
\mu_k = \sum_{t=0}^{k+7t \leq t_c} i_{k+7t},
$$

this constraint can be included in the minimization procedure by removing one of the unknowns using the above equation. However, we use another approach which consists of adding the above equation, multiplied by a certain weight w, to the expression (11). In this way we can control the weight that is assigned to this constraint. If w is large, we force the restriction to be fulfilled and if $w = 0$ we remove this constraint. In the experiments performed in this work we use $w = 10^5$, so we choose that the constraint is satisfied.

As explained in the Fig. 1, in the proposed method, we update iteratively i_t, R_t and $\hat{\mathbf{q}}(i, R, \Phi)$. Let us denote by i_t^n, R_t^n and $\hat{\mathbf{q}}^n(i, R, \Phi)$ these updates for each iteration n starting at $n = 0$. Following the flowchart of Fig. 1 these updates are computed using the following algorithm:

14

Data: i, Φ.
Result: $\hat{i}$, $\hat{R}$, $\hat{\mathbf{q}}(i,\hat{R},\Phi)$.
We fix $i^0 \equiv i$, $\mathcal{I}^0 = 1$, $\mathbf{q}^0 \equiv 1$, and compute R^0 from i^0 and Φ using the method proposed in [1];
for $n = 1, 2, .., MaxIter$ **do**
$\quad$ compute $\mathbf{q}^n(i, R^{n-1}, \Phi)$ using the method explained above;
$\quad$ compute $i_t^n = q_{t\%7}^n i_t$;
$\quad$ compute $\mathcal{I}^n = \sqrt{\frac{E(i^n, R^{n-1}, \Phi, \mathbf{q}^n)}{E(i^0, R^0, \Phi, \mathbf{q}^0)}}$;
$\quad$ **if** $\mathcal{I}^n > \mathcal{I}^{n-1}$ **then**
$\quad\quad$ stop iterations;
$\quad$ **else**
$\quad\quad$ $\hat{i} = i^n$;
$\quad\quad$ $\hat{\mathbf{q}}(i,\hat{R},\Phi) = \mathbf{q}^n(i, R^{n-1}, \Phi)$;
$\quad\quad$ compute R^n from i^n and Φ using the method proposed in [1];
$\quad\quad$ $\hat{R} = R^n$;
$\quad$ **end**
end

Algorithm 1: Algorithm to estimate $\hat{i}$, $\hat{R}$, $\hat{\mathbf{q}}(i,\hat{R},\Phi)$. *MaxIter* represents the maximum number of iterations allowed.

Using the online interface

In the online interface, available at `www.ipol.im/ern`, one can test the method proposed in [1] to estimate R_t as well as the method proposed here to remove the weekly administrative noise. When the DEMO is executed with the option "remove weekly administrative noise" activated, the method proposed in this paper is used to compute $\hat{i}_t$, $\hat{\mathbf{q}}(i, \hat{R}, \Phi)$ and $\hat{R}_t$. A comparison with the EpiEstim method is showed. We also show (in the plot on the right) the original curve i_t as well as the estimation using the renewal equation $F_2(\hat{i}, \hat{R}, \Phi_s, t)$. In the output file named "Rn.csv" you will find in the first row, the final value of the efficiency measure $\mathcal{I}$ (see (9)) and the vector $\hat{\mathbf{q}}(i,\hat{R},\Phi)$. The elements of the vector are organized in such a way that the last one $\hat{q}_6$ corresponds to the multiplicative factor of the last element of i_t (that is i_{t_c}). In this way, to obtain $\hat{i}_t$ we have just to do $\hat{i}_{t_c} = \hat{q}_6 i_{t_c}$, $\hat{i}_{t_c-1} = \hat{q}_5 i_{t_c-1}$,..., and, in general: $\hat{i}_t = \hat{q}_{(t+6-tc\%7)\%7} i_t$. Moreover, in the first column of the output file "Rn.csv" you will find the values of $\hat{R}_t$, in the second column, $F_2(\hat{i}, \hat{R}, \Phi_s, t)$, in the third column, the original number of infected i_t, and in the forth column a measure of the variability in the estimation of $\hat{R}_t$, as explained in [1].