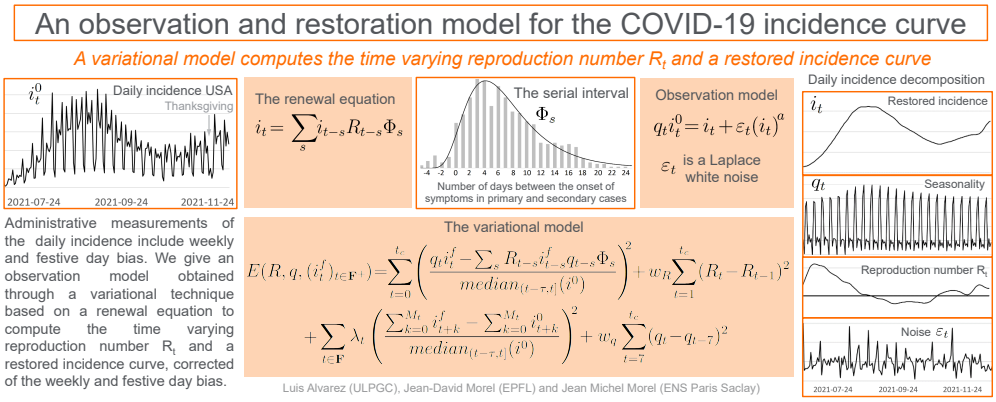
Graphical Abstract

An observation and restoration model for the COVID-19 incidence curves

Luis Alvarez, Jean-David Morel, Jean-Michel Morel



Highlights

An observation and restoration model for the COVID-19 incidence curves

Luis Alvarez, Jean-David Morel, Jean-Michel Morel

- Experimentally-validated model of COVID-19 incidence and reproduction.
- Adapts to weekly seasonality and festive day bias.
- Difference between model and bias corrected observed value is unbiased white noise.
- Noise follows an exponential distribution, multiplied by a power of incidence that is estimated.

An observation and restoration model for the COVID-19 incidence curves

Luis Alvarez[a], Jean-David Morel[b], Jean-Michel Morel[c]

[a]*Departamento de Informática y Sistemas, Universidad de Las Palmas de G.C., Las Palmas de G.C., 35017, Spain*
[b]*Laboratoire de Physiologie Intégrative et Systémique Ecole Polytechnique Fédérale de Lausanne, AI 1144 Station 1, 5, Lausanne , CH-1015, Switzerland*
[c]*ENS Paris-Saclay, CNRS, Centre Borelli, Université Paris-Saclay, Gif-sur-Yvette, F-91190, France*

Abstract

The sanitary crisis of the past two years has focused the public's attention on quantitative indicators of the spread of the COVID-19 pandemic. The daily reproduction number R_t, defined by the average number of new infections caused by a single infected individual at time t, is one of the best metrics for estimating the epidemic trend. In this paper, we give a complete observation model for sampled epidemiological incidence signals obtained through periodic administrative measurements. The model is governed by the classic renewal equation using an empirical reproduction kernel, and subject to two perturbations: a time-varying gain with a weekly period and a white observation noise. We estimate this noise model and its parameters by extending a variational inversion of the model recovering its main driving variable R_t. Using R_t, a restored incidence curve, corrected of the weekly and festive day bias, can be deduced through the renewal equation. We verify experimentally on many countries that, once the weekly and festive days bias have been corrected, the difference between the incidence curve and its expected value is well approximated by an exponential distributed white noise multiplied by a power of the magnitude of the restored incidence curve.

Keywords: Incidence curve, COVID-19, time dependent reproduction number, administrative noise, exponential distribution, variational inversion method.
2000 MSC: 92C60, 92C55, 45Q05, 65K10

---

1. Introduction

The renewal equation, first formulated for birth-death processes in a 1907 note of Alfred Lotka [1], establishes a model for epidemic propagation based on the individual infectiousness. The infectiousness of individuals at time t is characterized by the reproduction number R_t, defined as the average number of cases generated by an infected person at time t, and by the generation time, [2, 3], defined as the probability distribution of the time between infection of a primary case and infections in secondary cases. This probability distribution depends on the incubation time (a permanent biological factor) and on the detection time (which we assume stationary). For these reasons, the distribution of the generation time is supposed to be independent of t. In practice, the generation time is replaced by the observable serial interval Φ_s which represents the time distribution of the delay of the onset of symptoms between primary and secondary cases. In Fig. 1, we show the serial interval obtained in [4] using 689 observed pairs of primary and secondary cases.

The classic formulation of the case renewal equation (see for instance [5, 6]) is given by

$$i_t = \sum_s i_{t-s} R_{t-s} \Phi_s \quad \text{for} \quad t = 0, .., t_c, \tag{1}$$

where i_t represents the incidence curve of new daily cases, t_c is the current time. Government statistics of the observed incidence curve are affected by changes in testing and polling policies and by weekend reporting delays. These recording delays and subsequent rash corrections result in impulse noise, and a strong weekly periodic bias observable on the observed incidence curve i_t^0. In [3] this bias is corrected by a seven days sliding average and in [7] it is corrected by multiplying i_t^0 by a 7-day periodic factor q_t. These *bias correcting coefficients* q_t are learned by a variational method that we describe below. Our first purpose in this note is to resolve the festive day problem. We denote by $\mathbf{F}$ the set of festive days t, at which the i_t^0 curve is strongly affected by the reduction in the number of registered cases. This reduction is compensated by an increase in the number of registered cases the following days. No model has been proposed so far to address this problem, which creates strong impulse noise in any estimation of i_t and R_t. We tackle this problem by a variational method computing R_t, where both i_t and R_t are considered unknown on festive days and in the next few days. To that purpose, we shall denote by $\mathbf{F}_+$ the union of festive days and the ones

following them affected by the festive day (typically 2 or 3 days after the festive day).

Our second purpose is to give a noise model for the difference $\hat{i}_t - i_t^r$ between the signal $\hat{i}_t$ corrected of the week-end and festive effects, and its restored version i_t^r using the renewal equation, defined by

$$i_t^r = \sum_s \hat{i}_{t-s} R_{t-s} \Phi_s. \tag{2}$$

We give strong experimental evidence that the relation between $\hat{i}_t$ and i_t^r, can be empirically modeled by

$$\hat{i}_t = i_t^r + \varepsilon_t (i_t^r)^a, \tag{3}$$

where $a > 0$ and ε_t is a white noise.

This leads us to propose a signal processing version of the renewal equation model taking into account noise and bias and justifying *a posteriori* the variational method. The proposed observation model linking the observed signal i_t^0 to the ground truth incidence i_t is

$$q_t i_t^0 = i_t + \varepsilon_t (i_t)^a \quad \text{for } t \in [0, t_c] \setminus \mathbf{F}_+, \tag{4}$$

where q_t is a quasi-periodic gain with period 7 and ε_t is a white noise. The exceptional set $\mathbf{F}_+$ is introduced because festive days provoke perturbations of the observation model (4). Specifically, the 7 days period of q_t is broken for these groups of days.

We shall verify experimentally on 38 countries (and detail the results on USA, France and Germany) that the normalized error ε_t is indeed a white noise with a distribution that is well described by an exponential distribution. The observed noise model contradicts the classic stochastic formulation of the renewal equation where the first member i_t of equation (1) is assumed to be a Poisson variable, and the second member of this equation is interpreted as the expectation of this Poisson variable. Using this Poisson model leads to maximum likelihood estimation strategies to compute R_t [8, 9, 3, 10]. As we shall see, the Poisson model is not verified. Indeed, the empirically observed standard deviation of the noise follows a power law with exponent significantly larger than 0.5, which is incompatible with the Poisson model.

The proposed observation model (4) of the pandemic's incidence curve raises three challenges:

1. Computing the reproduction number R_t;
2. Correcting the weekend and festive days bias on i_t;
3. Modeling as noise the difference between the observed incidence curve after bias correction and its expected value using the renewal equation.

Plan. In section 2 we review the anterior variational method [7]. In section 3 we point out its three main limitations: its weekly bias correction is strongly periodic, which does not work on long periods; the festive days cause strong perturbations in the inversion, finally no residual noise model is proposed. We therefore modify its variational formulation. In section 4 we present the results of the statistical analysis of the residual noise on many countries. These examples lead to specify the noise model and to validate *a posteriori* the proposed inversion model.

Timely estimates of restored versions of i_t and R_t are extremely useful to tame a pandemic. The proposed restoration and inversion algorithm can be run through an online demo [11] for every day in every country and U.S. state. The demo plots the objects of this paper, namely the incidence curve i_t^0, its bias corrected version $\hat{i}_t$, its fully restored version i_t^r, finally the main pandemic index, the time-dependent reproduction number R_t. Fig. 2 illustrates the application of the variational method of section 3 to USA on February 1st, 2022, as displayed by the online demo. Fig. 3 compares the results of this inversion method, applied with and without festive day bias correction, obtained for France on January 6, 2022.

2. Anterior work

The EpiInvert variational method proposed in [7] is a deconvolution + denoising procedure estimating both R_t and a restored i_t corrected for the weekend bias. To remove the weekend effect, it computes a 7-day periodic multiplicative factor $q = (q_0, q_1, q_2, q_3, q_4, q_5, q_6)$. From the observed incidence curve and the serial interval, R_t and q are jointly estimated by minimizing

$$E(R,q) = \sum_{t=0}^{t_c} \left(\frac{q_{t\%7} i_t^0 - \sum_s R_{t-s} i_{t-s}^0 q_{(t-s)\%7} \Phi_s}{median_{(t-\tau,t]}(i^0)} \right)^2 + w \sum_{t=1}^{t_c} (R_t - R_{t-1})^2 \quad (5)$$

where $t\%7$ denotes the remainder of the Euclidean division of t by 7 and $median_{(t-\tau,t]}(i^0)$ is the median of i_t^0 in the interval $(t-\tau, t]$ used to normalize the energy with respect to the size of i_t. The total number of cases is

preserved by adding to (5) the constraint on q_t :

$$\sum_{t=t_c-T+1}^{t_c} i_t^0 = \sum_{t=t_c-T+1}^{t_c} q_{t\%7} i_t^0, \tag{6}$$

where T is a period of analysis empirically fixed to $T = 56$ days. The minimization of the above energy yields estimates of R_t, q and a restored incidence curve.

3. The proposed variational model

One limitation of using a 7-day periodic formulation to model the weekend effect is that it does not take into account the variation over time of the seasonal profile. To deal with this issue, we consider q_t for $t = 0, .., t_c$ allowing different correction factors q_t for every day but keeping the values $q_t - q_{t-7}$ small which forces q_t to be quasi-periodic.

In addition to the weekend bias, festive days can introduce a strong bias in the incidence values. On a festive day $t \in \mathbf{F}$, a sharp decrease in the number of registered incident cases is generally observed. This is compensated by increased incidence numbers in the next few days. Assuming that each festive day, $t \in \mathbf{F}$, mainly affects the value of the incidence curve in the festive day and in the next M_t days (where M_t is an algorithm parameter (by default we fix $M_t = \min\{2, t_c - t\}$)), we consider the values of $i_t^0, i_{t+1}^0, .., i_{t+M_t}^0$ as unknown. We denote by $\mathbf{F}_+$ the union of the festive days $t \in \mathbf{F}$ and the M_t days following them. We set $i_t^f = i_t^0$ for $t \notin \mathbf{F}_+$ and consider the values $(i_t^f)_{t\in\mathbf{F}_+}$ as unknowns. Then the new proposed inversion functional is

$$E(R, q, (i_t^f)_{t\in\mathbf{F}^+}) = \sum_{t=0}^{t_c} \left(\frac{q_t i_t^f - \sum_s R_{t-s} i_{t-s}^f q_{t-s} \Phi_s}{median_{(t-\tau,t]}(i^0)} \right)^2 + w_R \sum_{t=1}^{t_c} (R_t - R_{t-1})^2 +$$
$$\sum_{t\in\mathbf{F}} \lambda_t \left(\frac{\sum_{k=0}^{M_t} i_{t+k}^f - \sum_{k=0}^{M_t} i_{t+k}^0}{median_{(t-\tau,t]}(i^0)} \right)^2 + w_q \sum_{t=7}^{t_c} (q_t - q_{t-7})^2, \tag{7}$$

The values i_t^f for $t \in \mathbf{F}_+$ are set free in the minimization. Yet the third term in the functional ensures that the overall number of cases in the affected days remains unchanged. For each $t \in \mathbf{F}$, $\lambda_t \geq 0$ represents the weight we assign to this constraint for each festive day. We fix, experimentally, $\lambda_t = 2^{t_c-t-2}$ if $t_c > t$ and $\lambda_t = 0$ if $t_c = t$. In other terms, the value of λ_t is

adjusted according to the number of days that have passed since the festive day. To keep a smooth seasonality we add to the energy a regularization term where we penalize high values of $q_t - q_{t-7}$. The parameters w_R and w_q are regularization weights with default values $w_R = w_q = 5$. Their values are proven in [7] to be nearly optimal for Covid-19 incidence curves.

By minimizing this energy we obtain the reproduction number R_t, the seasonality q_t and i_t^f, which corresponds to the original incidence i_t^0 but with the optimized values in the festive days. The bias corrected incidence $\hat{i}_t$ defined in model (3) is given by $\hat{i}_t = q_t i_t^f$.

The estimated incidence curve must preserve the number of cases. In the original EpiInvert formulation this constraint is enforced by (6) on its analysis interval $(t_c - T, T]$. In the new formulation, the interval time of analysis is the whole time interval $[0, t_c]$. Extra conditions are required to keep i_t^0 close to $\hat{i}_t$ and i_t^r. Therefore, to preserve the number of cases we add to the energy (7) the constraints on q_t:

$$\sum_{t=0}^{t_c} i_t^f = \sum_{t=0}^{t_c} q_t i_t^f \quad ; \qquad \sum_{t_c-14(k+1)}^{t_c-14k} i_t^f = \sum_{t_c-14(k+1)}^{t_c-14k} q_t i_t^f \qquad \text{for } k = 0, 1, 2, ... \quad (8)$$

The first constraint corresponds to a global preservation of the number of cases in the whole period and the second one corresponds to a local preservation of the number of cases every 2 weeks. In particular, the second constraint ensures a good agreement between the epidemiological indicator given by the accumulated number of cases in the last 14 days of the original incidence curve and the estimated ones using the proposed method. This indicator is currently widely used to evaluate the current epidemic transmission.

The minimization of the energy (7) is obtained by alternating steps computing in turn R_t, q_t, and then i_t^f (for $t \in \mathbf{F}_+$) until convergence. The above constraints are added to the minimization by the Lagrange multiplier technique.

4. Results and Discussion

We used the incidence data published in [12]. We checked the observation model and its inversion on the 626 daily incidence data from March 24, 2020 to December 9, 2021 for 38 countries and will detail the results for France, Germany, and the USA. In general, for the festive days we fixed $M_t = 2$, so the method estimated the incidence value of the festive day and of the

next 2 days. However, not all festive days disturb the incidence in the same way. Parameter M_t allows us to adapt the number of days affected. To illustrate this option we set $M_t = 5$ for Thanksgiving in the USA in 2021 because this festive day causes in 2021 a longer perturbation in the number of registered cases. Figs. 4-6 show the minimization results for the energy (7). They display for each country (i) the original incidence curve i_t^0, (ii) the incidence curve after bias correction $\hat{i}_t$, (iii) the restored incidence curve i_t^r using the renewal equation (1), (iv) the weekly bias correction factors q_t, (v) the reproduction number estimation R_t and (vi) the normalized error defined by

$$\varepsilon_t = \frac{\hat{i}_t - i_t^r}{(i_t^r)^a}. \tag{9}$$

The power a was obtained through log-log linear regression. Indeed, if $|\hat{i}_t - i_t^r|$ is proportional to $(i_t^r)^a$, then $\log(|\hat{i}_t - i_t^r|) \approx a \cdot \log(i_t^r) + b$, and a and b can be estimated by a linear regression between $\log(|\hat{i}_t - i_t^r|)$ and $\log(i_t^r)$. Its results are illustrated for 38 countries in fig. 8 and table 1. The Pearson correlation p-values in this table confirm the linear relation. The estimated exponent a varies between 0.7 and 0.9, and the constant coefficient b varies between -0.11 and -2.6. For the world we have $a = 0.76$ and $b = -1.16$.

We performed a control test on a Brownian motion simulated by starting from 10000 and sampling $i_{t+1} - i_t \simeq \mathcal{N}(0, 100)$. The obtained exponent a is negative ($a = -1.01$) and we have $b = 13.4$. Both values are far away from the group of coefficients of real incidence curves. The p-value for the control is anyway non significant (0.0844), compared to the extremely small p-values for the real incidence curves. Fig. 7 shows the results of the variational inversion method on the Brownian control. For this control, both R_t and the weekly seasonality correction coefficients stay very close to 1 as should be expected, with means 1.001 and 1.00002, and standard deviations 1.7% and 0.3% respectively.

Next, we looked for a stochastic model of the normalized error ε_t defined by (9). Figs. 4-6 visually support a stationarity assumption for ε_t in France, Germany and USA. In fig. 9 we show the autocorrelation function for these three countries. For most non-zero shifts, its value stays inside the 95% confidence interval for the stationarity assumption. (This interval is indicated by horizontal blue lines in the plot.) Similar results were obtained on 33 more countries, as illustrated in fig. 13. These results support a white noise assumption for ε_t.

We finally estimated the parameters of the distribution of ε_t assuming an exponential power distribution with density

$$\frac{\beta}{2\alpha\Gamma(1/\beta)} e^{-\left(\frac{|x-\mu|}{\alpha}\right)^{\beta}}, \tag{10}$$

where μ is the location, α the scale and β the shape. These parameters to approximate ε_t by an exponential power distribution were estimated by the R-package *normalp* [13].

In Fig. 10, we plot for these three countries the histogram of the distribution of ε_t and its approximation by a normal ($\beta = 2$) and by the obtained optimal exponential distribution. We display the same result for 33 more countries in fig. 12.

Table 2 gives the results for all countries. Columns 5 to 8 in the table give the parameters of the optimal exponential law: location, scale, shape. In all cases the exponent remains close to 1. Fig. 11 displays a quantile-quantile plot comparing ε_t with the estimated exponential distribution for three countries: France, Germany, USA. The linear fit is excellent, and this goodness of fit is confirmed for 33 more countries in Figure 14.

Conclusion: The above experiments seem to confirm the validity of the observation model (4). As we saw, this model can be inverted by minimizing the energy (7). This minimization yields three signals: a restored incidence on the festive days, the administrative bias correcting coefficients q_t that are quasi-periodic with period 7, and the time varying reproduction number R_t, arguably the pandemic's most useful control parameter. Last but not least, the renewal equation deduces a restored incidence i_t^r by (2) from the bias compensated incidence $\hat{i}_t$. The modeling loop was closed by verifying that the normalized error defined by (9) is a white noise. We also found that this noise follows an exponential distribution. This analysis discards the Poisson model for the pandemic's case count. A pure case count *should* be a Poisson noise, but we saw that the main perturbation was an administrative bias which, once compensated, leaves behind a noise with standard deviation proportional to a power larger than 0.5 of the case count i_t. Under the Poisson model this standard deviation would have been equal to the square root of i_t.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] A. J. Lotka, Relation between birth rates and death rates, Science 26 (653) (1907) 21–22.

[2] J. Wallinga, P. Teunis, Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures, American Journal of epidemiology 160 (6) (2004) 509–516.

[3] A. Cori, N. M. Ferguson, C. Fraser, S. Cauchemez, A new framework and software to estimate time-varying reproduction numbers during epidemics, American journal of epidemiology 178 (9) (2013) 1505–1512.

[4] S. Ma, J. Zhang, M. Zeng, Q. Yun, W. Guo, Y. Zheng, S. Zhao, M. H. Wang, Z. Yang, Epidemiological parameters of coronavirus disease 2019: a pooled analysis of publicly reported individual data of 1155 cases from seven countries, Medrxiv (2020).

[5] H. Nishiura, Time variations in the transmissibility of pandemic influenza in Prussia, Germany, from 1918–19, Theoretical Biology and Medical Modelling 4 (1) (2007) 20.

[6] H. Nishiura, G. Chowell, The Effective Reproduction Number as a Prelude to Statistical Estimation of Time-Dependent Epidemic Trends, Springer Netherlands, Dordrecht, 2009, pp. 103–121.

[7] L. Alvarez, M. Colom, J.-D. Morel, J.-M. Morel, Computing the daily reproduction number of COVID-19 by inverting the renewal equation using a variational technique, Proceedings of the National Academy of Sciences 118 (50) (2021) 1–10.

[8] 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) 100356.

[9] Q.-H. Liu, M. Ajelli, A. Aleta, S. Merler, Y. Moreno, A. Vespignani, Measurability of the epidemic reproduction number in data-driven contact networks, Proceedings of the National Academy of Sciences 115 (50) (2018) 12680–12685.

[10] T. Obadia, R. Haneef, P.-Y. Boëlle, The r0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks, BMC medical informatics and decision making 12 (1) (2012) 147.

[11] L. Alvarez, M. Colom, J. D. Morel, J. M. Morel, EpiInvert Online Interface, IPOL : Image Processing On Line.
URL https://ipolcore.ipol.im/demo/clientApp/demo.html?id=55555000009

[12] H. Ritchie et al., Coronavirus Pandemic (COVID-19), OurWorldInData.org, Accessed December 9, 2021.
URL https://ourworldindata.org/coronavirus

[13] A. M. Mineo, On the estimation of the structure parameter of a normal distribution of order p, Statistica 63 (1) (2003) 109–122. doi:10.6092/issn.1973-2201/342.
URL https://rivista-statistica.unibo.it/article/view/342

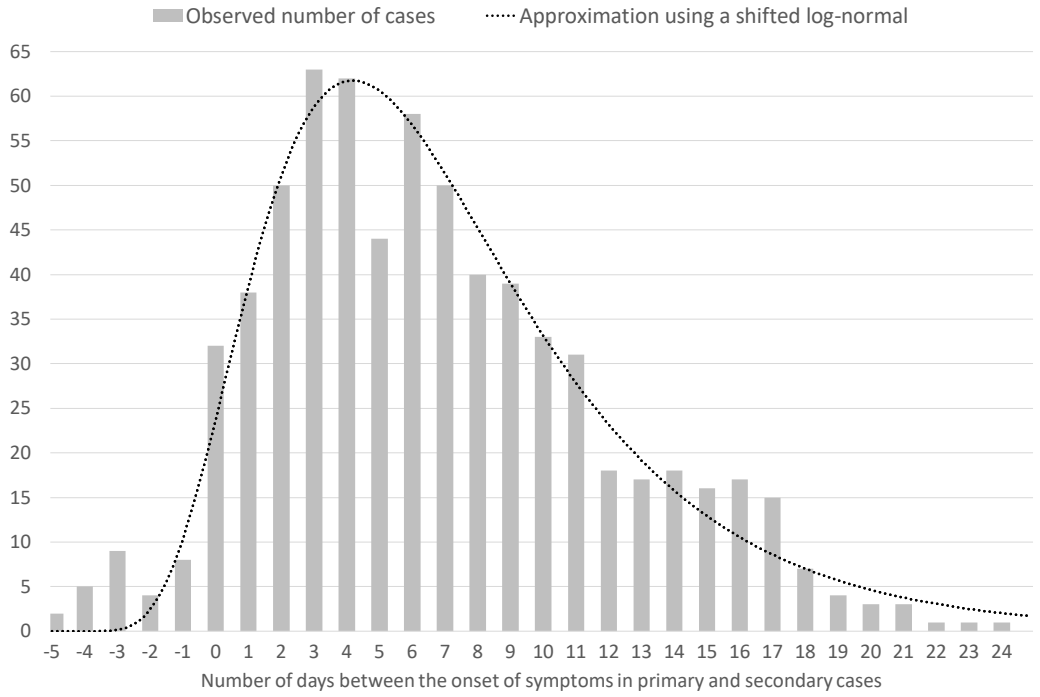


Figure 1: The serial interval Φ_s obtained by [4]. The bars represent the observed number of cases in function of the number of days between the onset of symptoms in primary and secondary cases. The dotted line is its approximation by a scaled and shifted log-normal distribution.

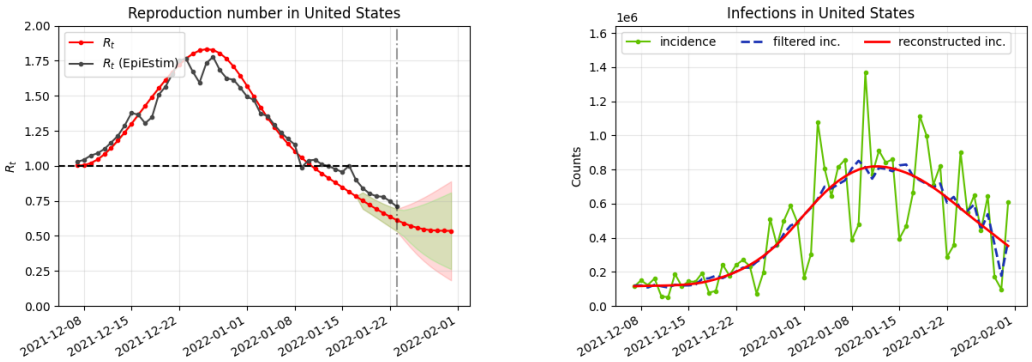


Figure 2: Illustration of the online inversion method [11]. On the left in red, the obtained reproduction number R_t and in black its estimate obtained by the classic EpiEstim method. On the right in green, the original incidence curve i_t of new cases, in blue the incidence curve $\hat{i}_t$ corrected of the weekend and festive biases, and in red the final reconstructed incidence curve i_t^r obtained from R_t by the application of the renewal equation. Estimate obtained for USA on February 1st, 2022.

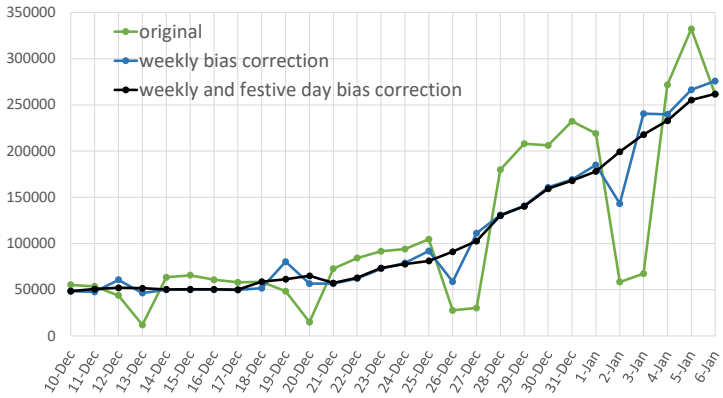


Figure 3: Incidence curve (in green) of France up to January 6, 2022. In blue, the incidence corrected of the weekly bias, in black the incidence corrected of the weekly and festive day. The Christmas holidays introduce a distortion in the weekly bias corrected incidence that is corrected by the festive day bias correction.

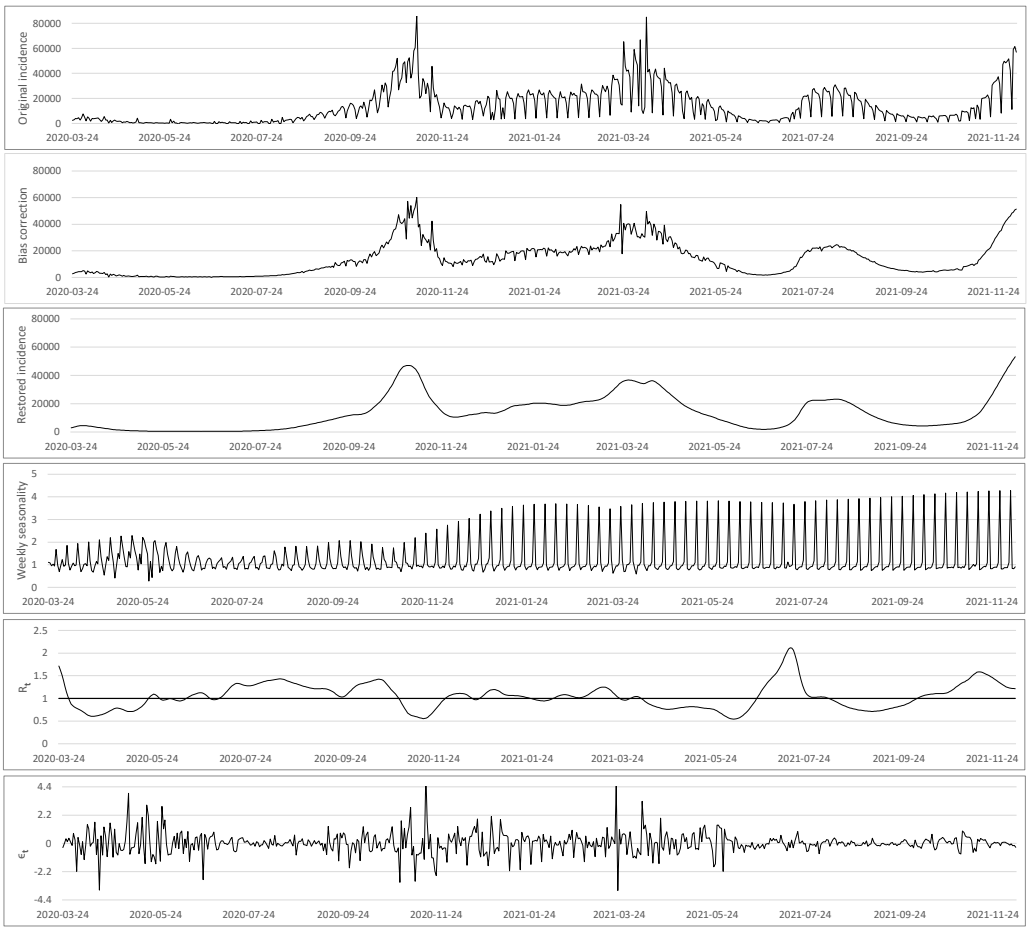


Figure 4: From top to bottom : (i) the original incidence curve i_t^0 of France, (ii) the incidence curve after bias correction $\hat{i}_t$, (iii) the restored incidence curve using the renewal equation i_t^r, (iv) the weekly bias correction factors q_t, (v) the reproduction number estimation R_t and (vi) the normalized error $\varepsilon_t = (i_t^r - \hat{i}_t)/(i_t^r)^a$, where a is the optimal exponent obtained by regression (see table 1).

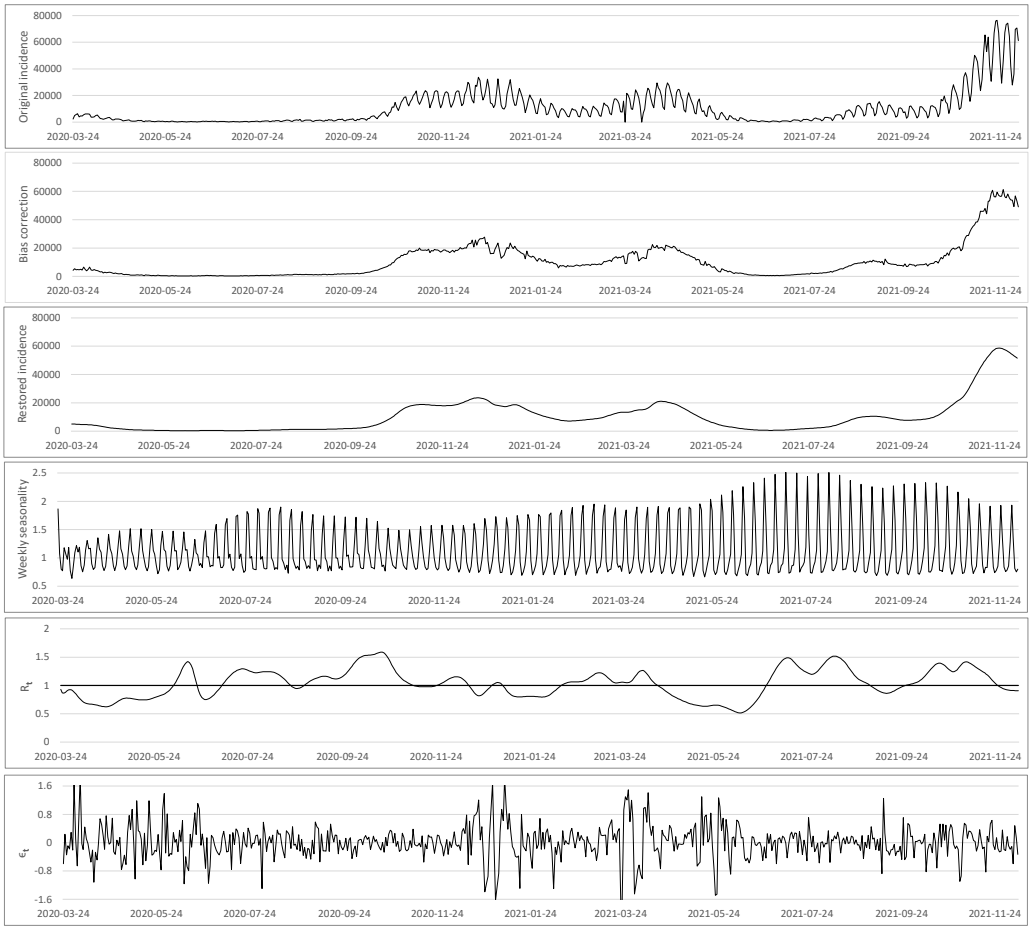


Figure 5: From top to bottom : (i) the original incidence curve of Germany i_t^0, (ii) the incidence curve after bias correction $\hat{i}_t$, (iii) the restored incidence curve using the renewal equation i_t^r, (iv) the weekly bias correction factors q_t, (v) the reproduction number estimation R_t and (vi) the normalized error $\varepsilon_t = (i_t^r - \hat{i}_t)/(i_t^r)^a$, where a is the optimal exponent obtained by regression (see table 1).

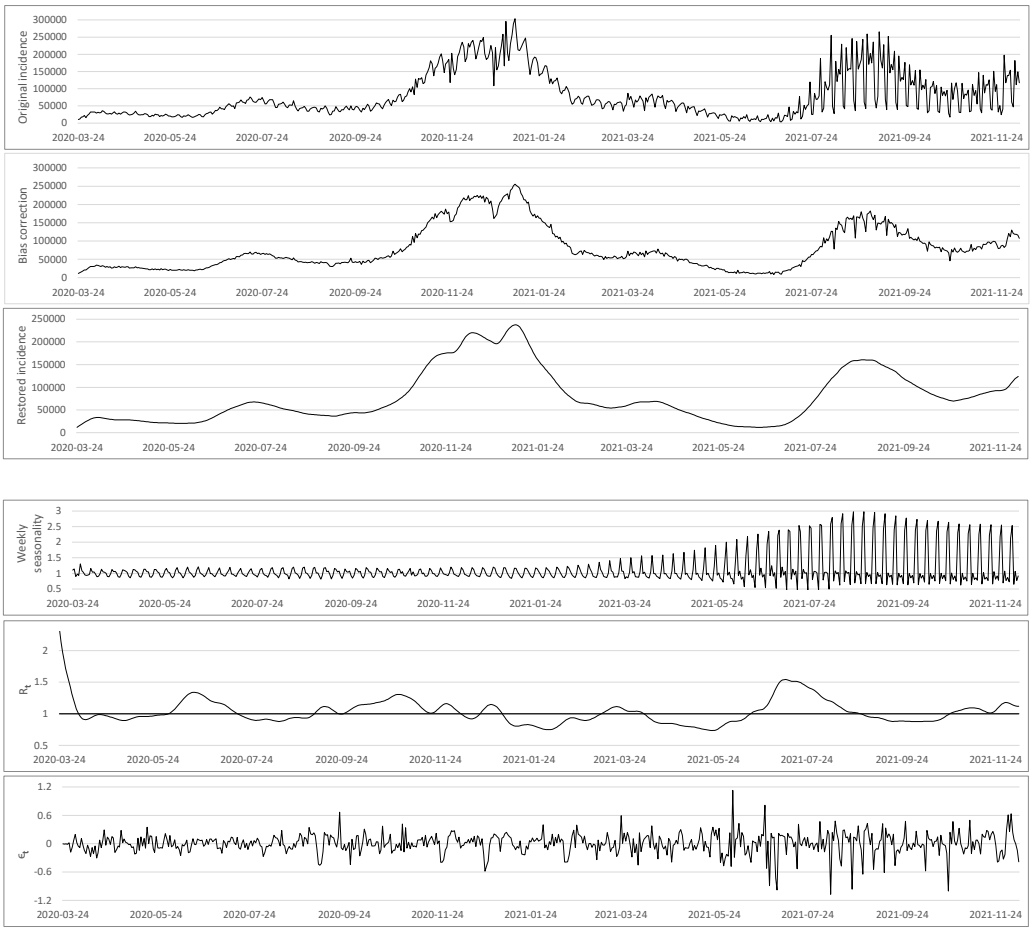


Figure 6: From top to bottom : (i) the original incidence curve i_t^0 of USA, (ii) the incidence curve after bias correction $\hat{i}_t$, (iii) the restored incidence curve using the renewal equation i_t^r, (iv) the weekly bias correction factors q_t, (v) the reproduction number estimation R_t and (vi) the normalized error $\varepsilon_t = (i_t^r - \hat{i}_t)/(i_t^r)^a$, where a is the optimal exponent obtained by regression (see table 1).

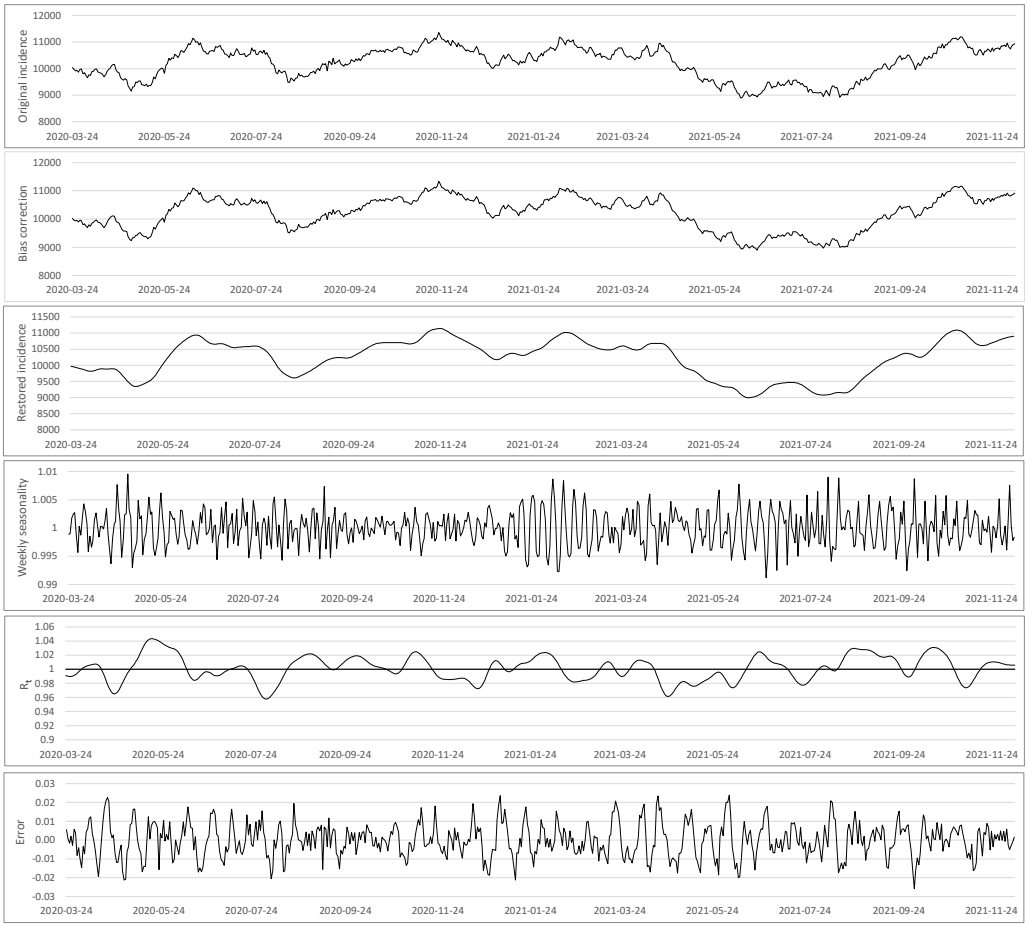


Figure 7: Control test, from top to bottom : (i) the test incidence curve i_t^0 which is a Brownian motion, (ii) the test curve after bias correction $\hat{i}_t$, (iii) the restored incidence curve using the renewal equation i_t^r, (iv) the weekly bias correction factors q_t, (v) the reproduction number estimation R_t and (vi) the relative error $(i_t^r - \hat{i}_t)/i_t^r$. Note that both R_t and the weekly seasonality correction coefficients stay very close to 1, with means 1.001 and 1.00002, and standard deviations 1.7% and 0.3% respectively.

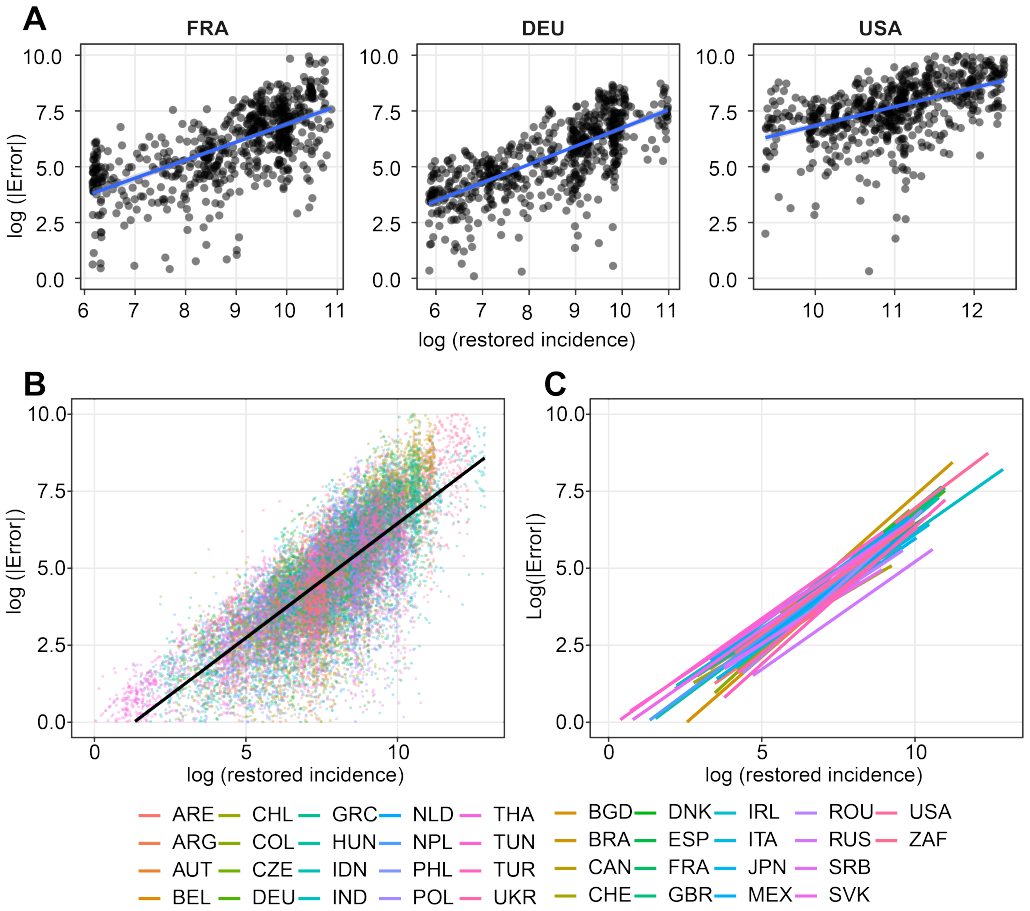


Figure 8: Worldwide log-log correlations between restored incidence i_t^r and the residual $|\hat{i}_t - i_t^r|$ (defined as restored incidence - bias-corrected incidence). The plot presents the log(error) as a function of the log(incidence). The regression parameters were computed through robust linear regression by the R package MASS. A: Correlation in France, Germany, and USA, with festive day correction. B: Spread of the values for 38 countries, without festive corrections. C: Robust linear regression curves for all countries. The linear regression coefficients a and b can be found in table 1. The worldwide coefficients are $a = 0.76$ and $b = -1.16$.

Country	a	b	p-value	Country	a	b	p-value
FRA*	0.8074272	-1.164141	2.01E-75	FRA	0.8136197	-1.1710322	2.76E-71
DEU*	0.8233846	-1.496739	5.99E-95	DEU	0.8235076	-1.5057318	3.01E-92
USA*	0.9076139	-2.264255	3.16E-42	USA	0.8638492	-1.7287377	6.37E-37
ARG	0.8340299	-1.5574878	1.71E-101	AUT	0.6628437	-0.5661912	3.45E-86
BGD	0.9104934	-2.5672893	6.14E-56	BEL	0.7184413	-0.6589731	3.65E-61
BRA	0.8906214	-1.536314	1.03E-58	CAN	0.7240632	-0.6726824	2.96E-44
CHL	0.8349688	-1.9543089	2.64E-40	COL	0.9175985	-2.2638884	3.03E-112
CZE	0.8520268	-1.4708978	2.88E-133	DNK	0.6900743	-0.6284769	2.78E-68
GRC	0.6555842	-0.5683038	2.58E-102	HUN	0.7838904	-1.3618843	4.47E-142
IND	0.7042499	-0.8457334	8.20E-68	IDN	0.8406915	-1.7674138	5.30E-97
IRL	0.7043354	-0.5484242	1.35E-89	ITA	0.6964125	-0.8659193	2.53E-71
JPN	0.7222903	-1.2445353	5.65E-85	MEX	0.725394	-0.4661005	1.76E-32
NPL	0.7548857	-1.0559482	1.42E-55	NLD	0.7494921	-1.1280471	3.35E-96
PHL	0.6715338	-0.1103984	1.90E-47	POL	0.9306078	-2.6041615	3.02E-133
ROU	0.6920366	-1.0282145	4.11E-77	RUS	0.7212814	-2.0048746	4.05E-26
SRB	0.628712	-0.65103	6.76E-92	SVK	0.7381511	-0.7853881	8.53E-164
ZAF	0.7275793	-0.7811203	9.48E-69	ESP	0.6806819	-0.3916179	2.03E-42
CHE	0.6138378	-0.5491828	1.38E-75	THA	0.7110685	-0.4672682	1.63E-222
TUN	0.7539949	-0.503523	3.03E-163	TUR	0.8998264	-2.658924	1.32E-68
UKR	0.8172308	-1.8996555	1.75E-70	ARE	0.7511088	-1.5460453	3.80E-52
GBR	0.8705096	-1.9395546	1.70E-96	**World**	0.7631129	-1.1389749	0.0000000
Brownian	-1.0155743	13.3969412	0.0844				

Table 1: Coefficients a and b for 38 countries of the log-log linear regression $ax+b$ between restored incidence i_t^r and the residual $|\hat{i}_t - i_t^r|$ as displayed in Fig. 8. The Pearson correlation p-values given by the *stats* R package confirm a linear relation. The exponent a varies between 0.7 and 0.9. Stars* indicate countries with festive correction. Note that the pvalues are slightly better with festive correction than without. The last row shows the results on the control curve, simulated as a Brownian process. Its large p-value discards a linear log-log relation, and the estimated values of a and b also stand far away from the estimated values for real incidence curves.

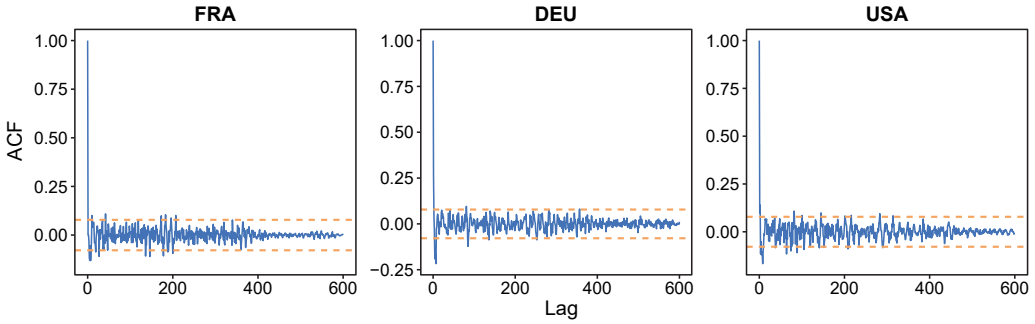


Figure 9: For France, Germany and USA, autocorrelation of the normalized error ε_t, using the festive day correction, obtained with the R-software functionalities (acf() function). The orange dotted line gives the 95% confidence interval for non-correlation. Similar plots for the same countries and 33 more countries, without using the festive day correction, are displayed in fig. 13.

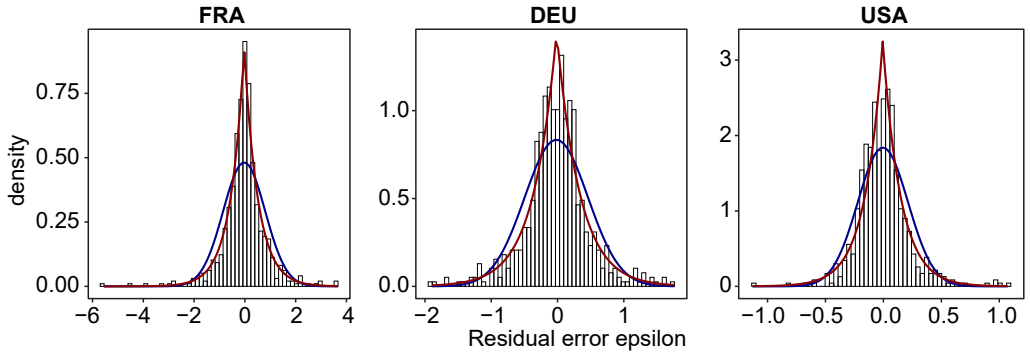


Figure 10: For France, Germany and USA, histogram of the normalized error ε_t, using the festive day correction, its normal approximation (blue line) and its optimal approximation using an exponential distribution (red line) (we use the R-package *normalp* to approximate ε_t by an exponential distribution). See fig. 12 for the results for the same countries and 33 more countries without using the festive day correction.

Country	**Mean**	**Std**	**Location**	**Scale**	**Shape (β)**
			Exponential	*Exponential*	*Exponential*
FRA*	-0.0283	0.8290	-0.0286	0.5394	1.0000
DEU*	-0.0178	0.4785	-0.0135	0.3433	1.0144
USA*	-0.0044	0.2169	-0.0059	0.1537	1.0000
FRA	0.0109	1.0024	-0.0316	0.6026	1.0000
DEU	0.0091	0.5143	0.0050	0.3458	1.0000
USA	0.0032	0.4779	-0.0097	0.3003	1.0000
ARG	0.0025	0.4430	-0.0286	0.3153	1.0000
AUT	0.0419	1.1030	-0.0041	0.9035	1.2701
BEL	0.0413	1.2175	-0.0366	0.8304	1.0000
BRA	-0.0018	0.4825	-0.0368	0.3312	1.0000
CAN	0.0068	1.2720	-0.0252	0.8290	1.0000
CHL	0.0019	0.2960	-0.0082	0.2138	1.0252
COL	-0.0026	0.2006	-0.0107	0.1490	1.0751
CZE	0.0116	0.5671	-0.0415	0.3755	1.0000
DNK	0.0278	1.2446	-0.0298	0.8126	1.0000
GRC	0.0218	1.2764	-0.0410	0.8847	1.0000
HUN	0.0069	0.6600	-0.0267	0.4410	1.0000
IND	0.0419	0.9891	-0.0084	0.6786	1.0000
IDN	-0.0015	0.3374	-0.0140	0.2607	1.1466
IRL	0.0030	1.1778	-0.0748	0.8252	1.0000
ITA	0.0368	1.1441	0.0141	0.7130	1.0000
JPN	0.0243	0.6647	-0.0254	0.4515	1.0000
MEX	-0.0318	1.7329	-0.0955	1.1091	1.0000
NPL	0.0035	0.8994	0.0005	0.5652	1.0000
NLD	0.0437	0.7185	-0.0404	0.4910	1.0000
PHL	-0.0196	2.0401	-0.0930	1.4011	1.0000
POL	-0.0017	0.1911	-0.0043	0.1268	1.0000
ROU	0.0063	0.9465	-0.0011	0.5798	1.0000
RUS	0.0107	0.3383	0.0066	0.2270	1.0000
SRB	0.0675	1.0140	0.0758	0.7932	1.1728
SVK	0.0024	1.3671	-0.0778	0.8194	1.0000
ZAF	0.0139	0.9110	-0.0320	0.7059	1.1497
ESP	0.0637	1.6068	-0.0047	1.0840	1.0000
CHE	0.0528	1.2228	0.0017	0.8667	1.0000
THA	0.0299	1.3738	-0.0312	0.9374	1.0000
TUN	0.0123	1.3033	-0.0845	0.9224	1.0000
UKR	0.0034	0.4117	-0.0215	0.2586	1.0000
ARE	0.0108	0.4192	-0.0127	0.3265	1.1588
GBR	0.0085	0.3304	-0.0171	0.2163	1.0000

Table 2: Table with the mean and standard deviation of ε_t and the parameters of the best fit to the exponential distributions for 36 countries. The data of starred countries in the first three rows have undergone the festive bias correction.

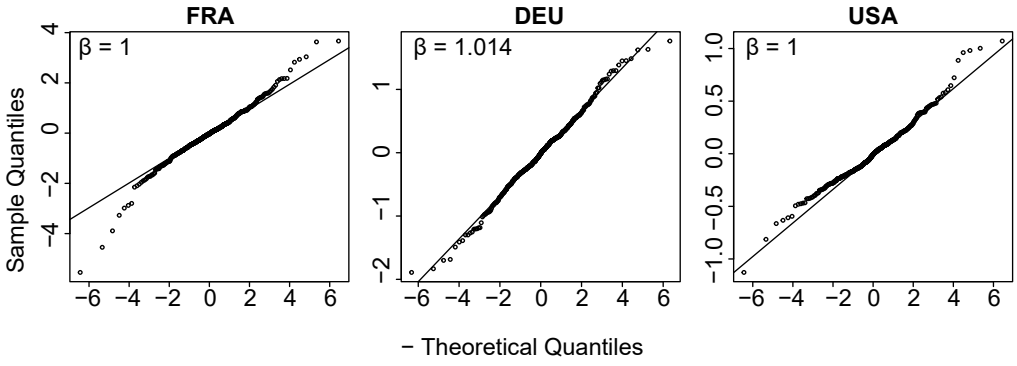


Figure 11: Quantile-quantile plot with France, Germany and the USA comparing ε_t, using the festive day correction, with the optimal exponential distribution using the R-package *normalp*.

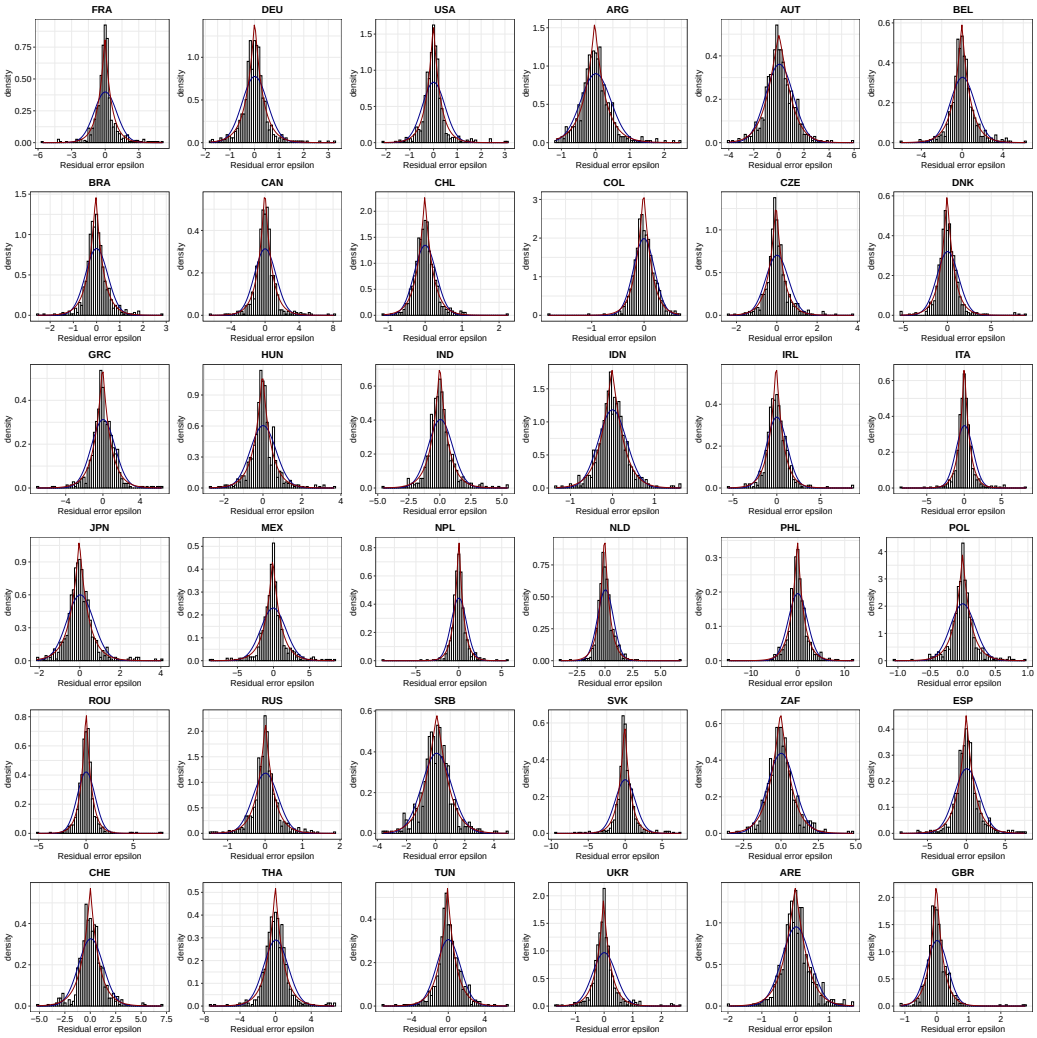


Figure 12: Fit of exponential distributions for 36 countries. In red, the best fitting exponential distribution with shape larger or equal to 1. In black, the best fitting normal law.

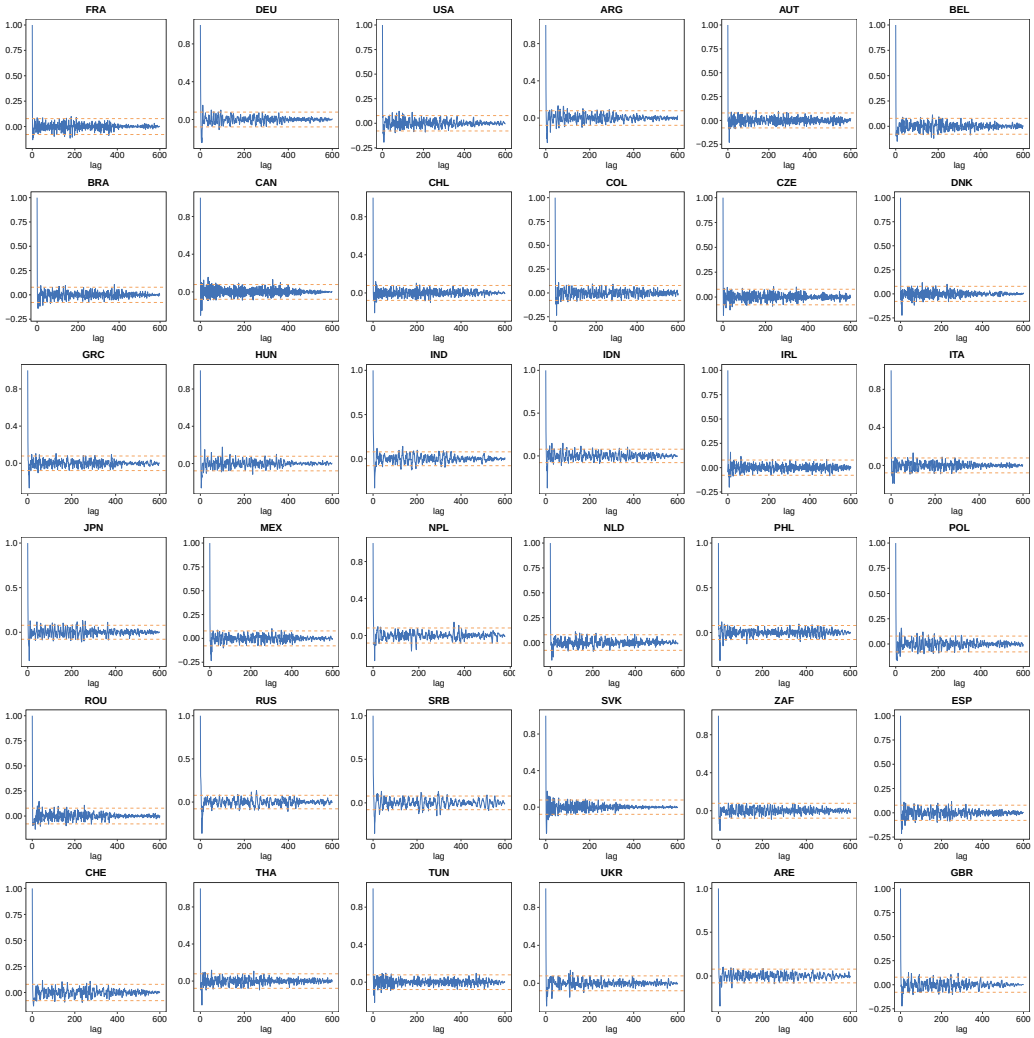


Figure 13: Autocorrelation of the normalized error ε_t using the R-software functionalities (acf() function) for 36 countries. The dotted lines give the 95% confidence interval for non-correlation.

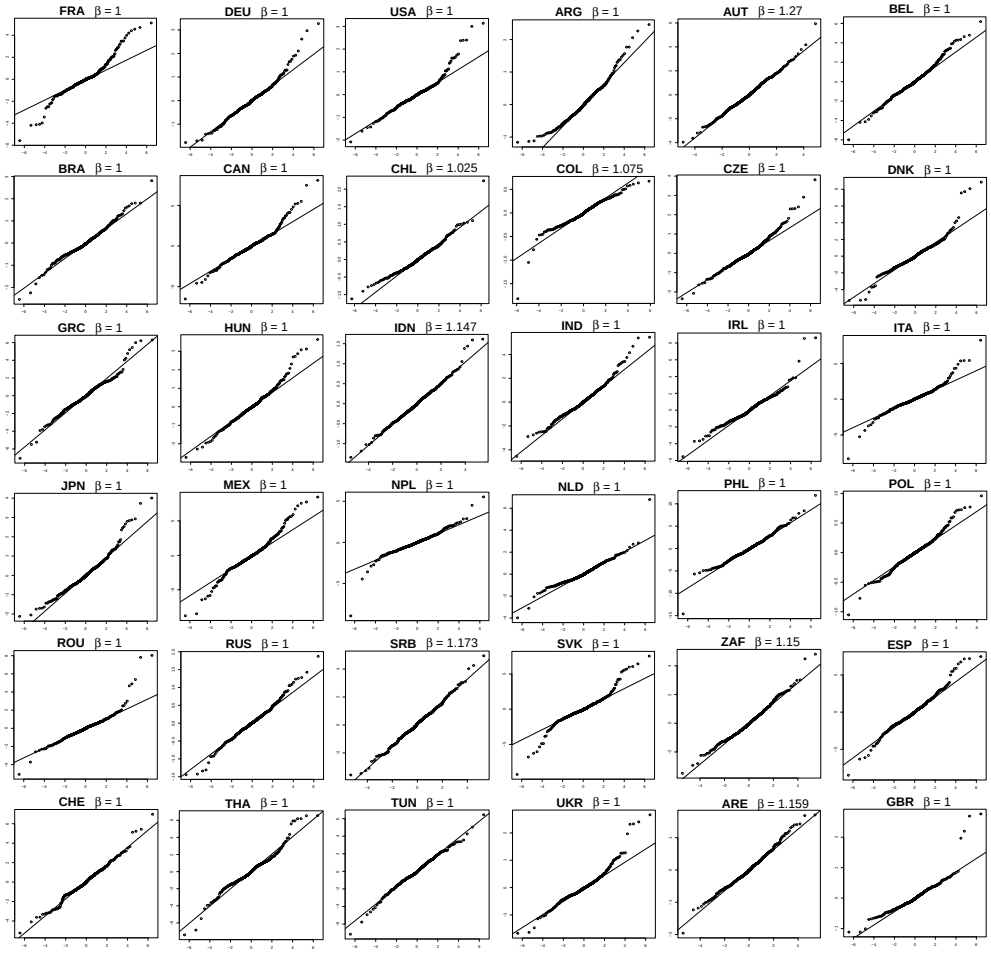


Figure 14: Quantile-quantile plot with 36 countries comparing ε_t (without using the festive day correction) with the optimal exponential distribution using the R-package *normalp*. In the horizontal axis we show the theoretical quantiles and in the vertical axis, the sample quantiles. Note that the exponential distribution shape parameter β, indicated on the graphs can have values >1.