*Article*

# A Bayesian Analysis of the Inversion of the SARS-COV-2 Case Rate in the Countries of the 2020 European Football Championship

Luca Casini [ID] and Marco Roccetti *[ID]

Department of Computer Science and Engineering, University of Bologna, 40127 Bologna, Italy; luca.casini7@unibo.it
* Correspondence: marco.roccetti@unibo.it

**Abstract:** While Europe was beginning to deal with the resurgence of COVID-19 due to the Delta variant, the European football championship took place from 11 June to 11 July 2021. We studied the inversion in the decreased/increased rate of new SARS-COV-2 infections in the countries of the tournament, investigating the hypothesis of an association. Using a Bayesian piecewise regression with a Poisson generalized linear model, we looked for a changepoint in the timeseries of the new SARS-COV-2 cases of each country, expecting it to appear not later than two to three weeks after the date of their first match. The two slopes, before and after the changepoint, were used to discuss the reversal from a decreasing to an increasing rate of the infections. For 17 out of 22 countries (77%) the changepoint came on average 14.97 days after their first match (95% CI 12.29–17.47). For all those 17 countries, the changepoint coincides with an inversion from a decreasing to an increasing rate of the infections. Before the changepoint, the new cases were decreasing, halving on average every 18.07 days (95% CI 11.81–29.42). After the changepoint, the cases begin to increase, doubling every 29.10 days (95% CI 14.12–9.78). This inversion in the SARS-COV-2 case rate, which happened during the tournament, provides evidence in favor of a relationship.

**Keywords:** SARS-COV-2; Bayesian regression; changepoint detection; European football championship

## 1. Introduction

Europe, as well as other countries around the world, is seeing a resurgence in the COVID-19 pandemic, after a brief respite given by the effects of the vaccination that started in the first half of 2021. This new wave of the pandemic seems to be driven by a new strain of virus that has been referred to as the Delta variant. This is the scenario in which the European football championship has taken place, from 11 June to 11 July 2021 (one year later than it should have been). This 2020 edition, being a special celebration for the 60th anniversary of the tournament, has had the peculiarity of being hosted by several different countries, instead of just one as it normally happens.

The decision to allow such a massive event across the European continent, in such a delicate time, immediately triggered a debate on the problems it would cause. Nonetheless, the competition was held, leaving each hosting country some freedom on which restrictions to apply (e.g., the number of fans allowed at each football stadium). This resulted in very different behaviors, ranging from Hungary hosting its matches at full stadium capacity at Puskás Arena (~68 thousand seats) to Germany limiting the attendance to 22% of the maximum stadium capacity [1–4]. Obviously, there were more factors than just the stadium, with fans, massively gathering in pubs, squares, and public places, to watch the matches, thus leading to infection clusters that surged all around Europe, as witnessed by the media coverage of these events [5–8]. Not only that but even the gathering of teams and their staff may have given their contribution to the spread of the virus (given the itinerant nature of this edition), as the COVID-19 literature on football and other sports suggests [9–11].

On one side, one could conclude that those who considered this event to be a minor risk did not take into any consideration of those theories that maintain that, with COVID-19,

super-spreading events may be the main driver of an epidemic spread, under specific circumstances [12,13]. An example, on 19 February 2020, was the Champions League match, between Atalanta and Valencia, which attracted a third of Bergamo's population to Milan's San Siro stadium. In addition, more than two thousand and a half of Spanish supporters took part. Experts, now, point to that 2020 football match as one of most relevant reasons why the city of Bergamo had become the epicenter of the COVID-19 pandemic, during the first wave in Italy, with a very high death toll; not to mention, that the 35% of Valencia's team also became infected [14]. On the other hand, it is well known that the return of supporters to stadiums is the highest priority for football's business, and the financial impact of the COVID-19 pandemic on football depends, almost exclusively, on both the timing and the scale of supporters' return to stadiums [15].

Following this debate, this work focused on the European football championship and its matches, looking for a possible compatibility with the reversal of the decrease/increase trend of the SARS-COV-2 cases, observed in many countries participating in the tournament. To investigate the hypothesis of an association between those football matches and the resurgence of the virus, we searched for a changepoint in the daily timeseries of the new SARS-COV-2 cases registered in each country, expecting it to appear not later than 2–3 weeks after the date of the first match that the national team played. Upon finding such a changepoint, we investigated if that changepoint was coincidental with a change in the infection rate, from a decreasing trend to an increasing one. It should be noted that our type of analysis has been observational in nature, and it was used to determine if the exposure to the specific risk factor, given the frequent mass gatherings following the football events, might have correlated with the particular outcome of the virus resurgence in many European countries. With this type of study, we cannot demonstrate any cause and effect, but we can make preliminary inferences on the correlation between the participation in the European football championship of a given country and the inversion in the SARS-COV-2 case rate that may have hit, at a particular point in time, the population living in that country.

We can anticipate that 17 out of 22 countries (77%) had a reversal from a decreasing to an increasing rate of the infections, which is temporally coincident with their participation in the European football championship, thus providing evidence to the hypothesis of a link between the upturn of new cases and the tournament. Instead, only 4 out of 12 countries (33%) that did not take part in the tournament (subject of an additional investigation) followed the same pattern as above. This further confirms that, while it can be inferred that an increase in COVID-19 cases may have been an inevitable consequence of the general European situation in July 2021, the European football tournament, with its mass gatherings, has at least played an important role of the accelerator of this phenomenon for many of its participating countries.

The remainder of the paper is structured as follows: In the next section, we describe more precisely the data we used, their sources, and the methodologies we employed. Section 3 presents the results we obtained, while Section 4 discusses them, along with their limitations, and concludes the paper, presenting our final considerations.

## 2. Materials and Methods

In this section, we provide a description of the data on which our analysis is based, along with the methods used for its collection and the sources from which we collected them (Section 2.1). Then, we present the methodologies we have chosen to conduct our analysis (Section 2.2).

### 2.1. Data Collection

The timeframe for this study starts two weeks before the start of the tournament on 28 May, and it ends two weeks after the final match on 25 July 2021. All data regarding COVID-19 infections were collected from the online repository: Our World in Data [16], that in turn aggregates various sources. In particular, the confirmed cases were provided by the

COVID-19 Data Repository by the Center for Systems Science and Engineering at the Johns Hopkins University. The timeseries of daily confirmed cases was then smoothed using a rolling average with a 7 day-long window. This was useful for removing the periodicity patterns of the various testing and registering case processes, with some countries that unfortunately release numbers once every few days (e.g., Sweden) or slow down on weekends (e.g., Italy).

Data for the European football championship were collected from the relevant Wikipedia page [17]. We looked at the participating countries, their first and last matches in the competition, and their last hosted match (if they were a hosting country). These dates were then compared with the changepoints found with the Bayesian method described in the next section. For the sake of simplicity, given that the data for the United Kingdom were given as a whole in the dataset we used, we considered Wales, Scotland, and England as a single entity, even if the three countries participated individually.

We conclude this subsection by confirming that patients and/or the public were not involved in the design, conduct, reporting, or dissemination plans of this research. All data come from a publicly available repository where they are stored in an aggregated and anonymized format.

### 2.2. Bayesian Changepoint Detection and Analysis

Using a changepoint estimation technique, based on a Bayesian piecewise regression, we have looked for a changepoint in the trend of the infection curve, whether it was growing or falling. In particular, we fitted a Poisson generalized linear model where the dependent variable was the number of new daily confirmed SARS-COV-2 cases, and the independent variable was just the number of days since 28 May 2021 (until 25 July). The result was a model comprised of a changepoint and two segments, whose slopes represent, respectively, the increase/decrease in case rate before the changepoint and the increase/decrease in case rate after it. The fact that our interest was not in modeling the spread of the virus with the maximum precision but rather in finding the point in time when the infection rate inverted (or simply changed) its trend, with the added bonus of a Bayesian uncertainty estimation, is worth noting. The model takes the mathematical form below:

$$\ln(E(Y|x)) = a_1 + xb_1 \text{ if } x < \tau \quad \ln(E(Y|x)) = a_2 + xb_2 \text{ if } x > \tau \tag{1}$$

It is worth noting that our dependent variable (the confirmed daily cases) was modelled as a Poisson distribution, whose mean depends on the regression coefficients $a_1$ and $b_1$, respectively, along with the intercept and angular coefficient before the changepoint $\tau$ (while $a_2$ and $b_2$ play the same role after the changepoint). To be considered are the three following facts:

First, since the two regression lines are joined at the changepoint $\tau$; the second intercept term $a_2$ is not estimated as it is bound to be $a_2 = \tau(b_1 - b_2) + a_1$.

Second, the formula above returns the exponential growth/decay trend, both before and after $\tau$, as easily identifiable slopes.

Third, to compute the number of days needed to halve/double the number of cases before/after a changepoint, the following two formulas can be used: specifically, Formula (2) can be used to compute the halving ($H$) and doubling ($D$) time, before a changepoint:

$$H_b = \frac{\ln \frac{y_\tau}{2} - a_1}{b_1} - \tau \, ; D_b = \frac{\ln 2y_\tau - a_1}{b_1} - \tau \, ; \text{ where } y_\tau = e^{a_1 + \tau b_1} \tag{2}$$

Formula (3), instead, can be used to compute the halving and doubling time, after a changepoint.

$$H_a = \frac{\ln \frac{y_\tau}{2} - \ln y_\tau}{b_2} \, ; \, D_a = \frac{\ln 2y_\tau - \ln y_\tau}{b_2} \, ; \text{ where } y_\tau = e^{a_1 + \tau b_1} \tag{3}$$

To fit the model above, we used the R package mcp, using a Markov chain Monte Carlo method [18]. For starting the Bayesian estimation, the default priors for $\tau$, $a_1$, $a_2$, and $b_1$ were chosen as suggested in [18], thus considering the prior of $\tau$ as a uniform, and the parameters $a_1$, $b_1$, and $b_2$ as normally distributed, as reported in the following formulas:

$$\tau \sim \text{Uniform}\ (\min(x), \max(x)) \tag{4}$$

$$a_1, b_1, b_2 \ \backsim \ \mathcal{N}(0, 10) \tag{5}$$

It is now worth noting that the mean value of the computed changepoint posterior distribution was used to calculate the distance in time between the date of a given changepoint and that of the first match played the corresponding team. Similarly, the mean values for the coefficients $b_1$ and $b_2$ were used to compute the steepness of the two slopes, respectively, before and after the changepoint.

The values obtained from the Bayesian regression have 95% credible intervals, associated with them. The aggregated statistics we computed for the countries (average distance from changepoint, average doubling/halving time, etc.) have 95% confidence intervals, computed using bootstrap.

This completes the description of our method from a statistical viewpoint. Nonetheless, it is appropriate to motivate the reason behind the use of this statistical methodology. The intuition is as follows: we were interested in finding if there was a particular point in time (occurring during the championship) that had brought a change in the curve of the number of the new daily infections, something like: a before and an after. In such a case, we also wanted to have some clear mathematical representations describing the increase or the decrease in the number of cases, what we could call the growth/decay rates.

We have obtained this by fitting a regression model that is segmented (i.e., piecewise). The precise point of the change was found by looking for the place that yielded the best fit with the regression. Not only that, we have also chosen to use a Bayesian regression, as it makes the model more interpretable, especially in the case of a bad fit (e.g., multiple changepoints, when we look for just one).

At that point, once we have obtained our posterior distribution on the parameters of interest, we have then used the mean value of the changepoint distribution to compare it with the date of the first match that each given national team had played, to see the existence of some relationship. Here, the idea is that if: (i) no more than two or three weeks separate these two events (first match and changepoint), and (ii) the change returns an inversion in the infection rate from a decrease to an increase trend, then we can strengthen the suspicion that the tournament with its mass gatherings played the role of the accelerator of a broader infection increase trend in Europe.

To complete this informal description, it is worth mentioning that we have used the mean values of the parameters from the distribution to draw the two straight lines representing the rate of the new cases before and after the changepoint, and finally, we have used them to compute the number of days needed to double/halve the number of cases. This final computation gives one a more precise idea of the impact of the change.

As a final note, it is important to mention that while it is quite common that COVID-19 cases show their biggest single-day jumps two to three weeks after a particular mass event [19], we have extended the search space for a changepoint to four weeks, for the sake of reliability. Nonetheless, following the literature, we have considered to be of interest only those changes that occurred in the infection curves in the temporal interval from 5–6 to 22–23 days after the event of interest.

## 3. Results

This section is split into two different parts. The first one (Section 3.1) reports the results we obtained with the 22 countries that took part in the European Football Championship. The second one (Section 3.2) illustrates the results we obtained with some 12 European countries that did not participate in the tournament.

### 3.1. Countries That Participated in the Tournament

In total, 17 out of 22 (77%) countries taking part in the European football championship show a changepoint occurring not later than 2–3 weeks after their first match (i.e., during the tournament).

For all these 17 countries, the changepoint coincides with a reversal in the new daily SARS-COV-2 cases, from a decreasing to an increasing rate.

The group of all these countries provides an evidence in favor of the hypothesis. Precisely, the group is comprised of all the following countries: Austria, Belgium, Croatia, Czechia, Denmark, Finland. France, Germany, Hungary, Italy, Netherlands, North Macedonia, Poland, Slovakia, Spain, Switzerland, and Ukraine.

Table 1 provides the lists of those countries, where under the $\tau$ we listed, for each country, the mean value of the days passed before the changepoint was detected since 28 May 2021 (i.e., the beginning of the period of observation). Since we are working with a posterior distribution, the 95% CI is indicated in brackets.

**Table 1.** Countries with a changepoint coincidental with a reversal from a decrease to an increase in the SARS-COV-2 case rate that occurred during the European football championship.

| Country (Participating in the Tournament) | $\tau$ (Changepoint, avg. Value and 95% CI) | Diff (Days Separating $\tau$ from First Match) | $b_1$ (Angular Coefficient before $\tau$, avg. Value and 95% CI) | $b_2$ (Angular Coefficient after $\tau$, avg. Value and 95% CI) | $a_1$ (Intercept before $\tau$, avg. Value and 95% CI) |
|---|---|---|---|---|---|
| Austria | 36.4 (35.8, 37.1) | 20 | −0.05 (−0.06, −0.05) | 0.08 (0.08, 0.09) | 6.28 (6.25, 6.32) |
| Belgium | 24.9 (24.6, 25.2) | 10 | −0.06 (−0.06, −0.06) | 0.04 (0.04, 0.04) | 7.70 (7.68, 7.71) |
| Croatia | 28.7 (27.4, 30.3) | 13 | −0.06 (−0.06, −0.06) | 0.02 (0.02, 0.03) | 5.87 (5.83, 5.92) |
| Czechia | 26.2 (25.3, 27.2) | 9 | −0.06 (−0.06, −0.05) | 0.02 (0.02, 0.03) | 6.30 (6.26, 6.34) |
| Denmark | 28.2 (27.8, 28.6) | 13 | −0.06 (−0.06, −0.06) | 0.05 (0.05, 0.06) | 7.13 (7.11, 7.15) |
| Finland | 19.8 (18.4, 21.0) | 5 | −0.03 (−0.04, −0.03) | 0.04 (0.04, 0.05) | 4.98 (4.90, 5.05) |
| France | 34.7 (34.6, 34.8) | 17 | −0.06 (−0.06, −0.06) | 0.11 (0.11, 0.11) | 9.28 (9.28, 9.29) |
| Germany | 35.1 (34.6, 35.5) | 17 | −0.07 (−0.07, −0.07) | 0.05 (0.05, 0.05) | 8.60 (8.59, 8.62) |
| Hungary | 40.3 (38.4, 42.0) | 22 | −0.06 (−0.06, −0.06) | 0.03 (0.02, 0.05) | 5.99 (5.95, 6.03) |
| Italy | 36.5 (36.3, 36.8) | 23 | −0.05 (−0.05, −0.05) | 0.09 (0.09, 0.10) | 8.25 (8.24, 8.26) |
| Netherlands | 26.4 (26.2, 26.6) | 10 | −0.06 (−0.06, −0.06) | 0.09 (0.09, 0.09) | 8.18 (8.17, 8.20) |
| N. Macedonia | 34.8 (31.9, 37.6) | 19 | −0.05 (−0.05, −0.04) | 0.05 (0.04, 0.07) | 3.59 (3.46, 3.72) |
| Poland | 35.2 (33.5, 36.8) | 18 | −0.07 (−0.07, −0.07) | 0.01 (−0.00, 0.01) | 6.92 (6.89, 6.94) |
| Slovakia | 39.4 (36.8, 42.1) | 22 | −0.05 (−0.05, −0.04) | 0.02 (0.00, 0.03) | 5.02 (4.96, 5.08) |
| Spain | 24.9 (24.8, 25.0) | 8 | −0.01 (−0.01, −0.01) | 0.07 (0.06, 0.07) | 8.45 (8.44, 8.46) |
| Switzerland | 32.9 (32.5, 33.5) | 18 | −0.07 (−0.07, −0.07) | 0.08 (0.08, 0.08) | 6.93 (6.91, 6.96) |
| Ukraine | 25.8 (25.1, 26.5) | 10 | −0.05 (−0.05, −0.05) | 0.00 (0.00, 0.01) | 8.06 (8.04, 8.07) |

In the *diff* column, instead, we listed the difference, in terms of days, between the point in time when the changepoint occurred and the date of the first match played by that given national team.

The fourth and fifth columns of Table 1 show the mean values (with the corresponding 95% CI) for the coefficients $b_1$ and $b_2$, that have been used to compute the steepness of the slopes, respectively, before and after the changepoint.

The sixth column, finally, reports the average value of the first intercept $a_1$, with its 95% CI.

We further worked with the numbers comprised in Table 1 by rounding the mean changepoint value for all the 17 countries and then by calculating the difference, in terms of days, between that value and the date when they played their first match.

This way, we obtained that the average date of the changepoint, for all the 17 countries of interest, falls 14.97 days (95% CI 12.29–17.47) after the beginning of their participation in the tournament (approximately two weeks).

Finally, we made a step further and, taking the mean values for the coefficients $b_1$ and $b_2$, we estimated how the slopes for the two lines changed, on average, before and after the changepoint. We gathered that all the 17 countries had a decreasing number of daily cases until the changepoint and ended up with a reversed trend afterwards.

Table 2 shows the halving time before, and the doubling time, after the changepoint, for each given country of this group.

**Table 2.** Quantifying the inversion from a decrease to an increase in the SARS-COV-2 case rate for the countries of Table 1.

| Country | Days Needed to Halve the Number of Cases (before $\tau$) | Days Needed to Double the Number of Cases (after $\tau$) |
|---|---|---|
| Austria | 12.69 | 8.18 |
| Belgium | 11.05 | 17.60 |
| Croatia | 11.77 | 28.32 |
| Czechia | 12.05 | 28.22 |
| Denmark | 11.49 | 12.71 |
| Finland | 21.81 | 15.84 |
| France | 11.86 | 6.32 |
| Germany | 10.14 | 13.17 |
| Hungary | 11.10 | 22.10 |
| Italy | 13.56 | 7.43 |
| Netherlands | 12.11 | 7.69 |
| N. Maced. | 14.88 | 13.20 |
| Poland | 9.67 | 92.80 |
| Slovakia | 14.91 | 41.59 |
| Spain | 103.50 | 10.62 |
| Switzerland | 10.03 | 8.56 |
| Ukraine | 14.43 | 159.00 |

More precisely, the mean halving time before the changepoint is 18.07 days (95% CI 11.81–29.42), while the mean doubling time after the changepoint is 29.10 days (95% CI 14.12–49.78).
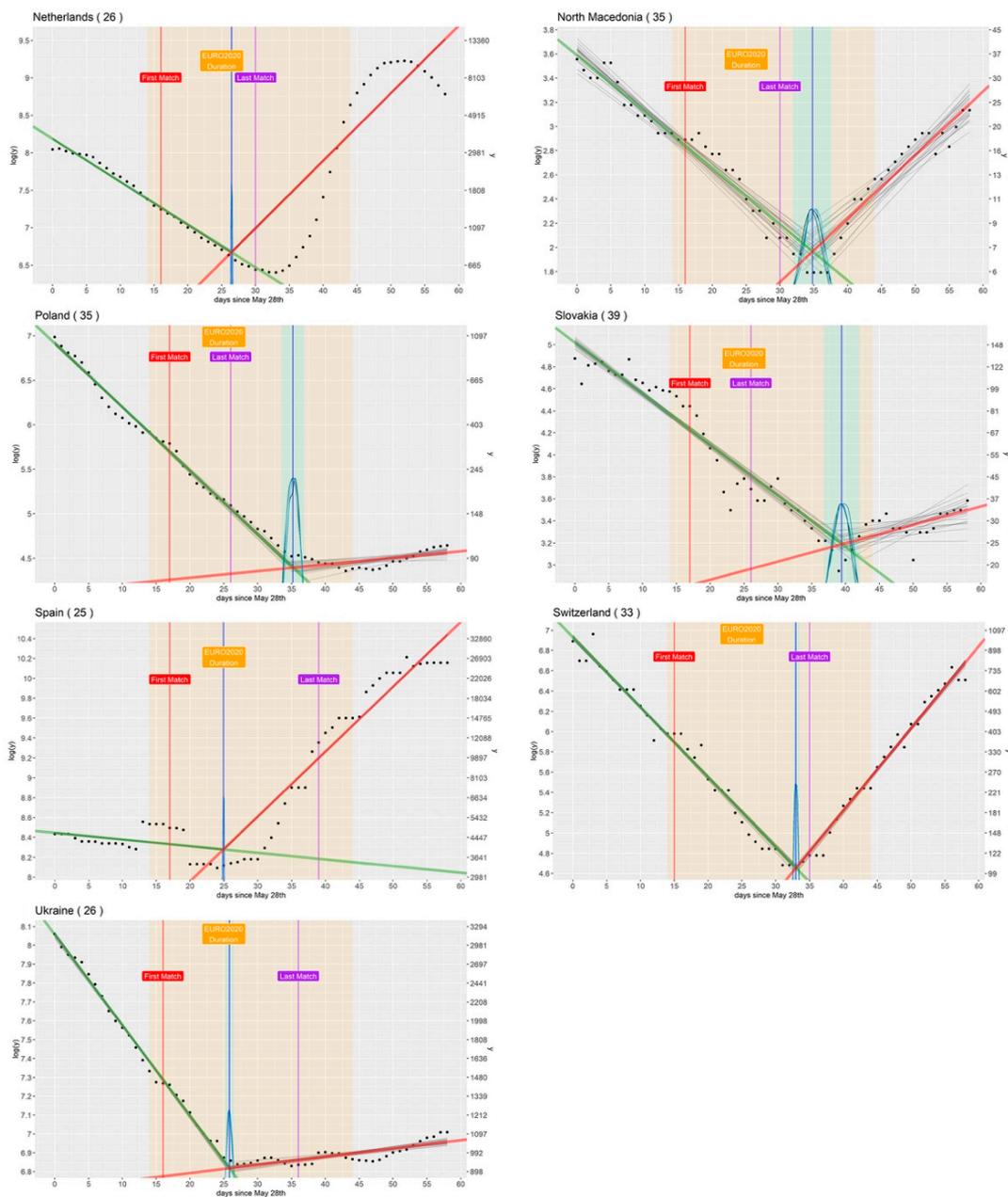
The credible intervals are quite wide, but if we better investigate the values reported in Table 2, we recognize that most of the deviation depends on just three countries, namely: Spain, Ukraine, and Poland, with their exceptionally large values.

To better highlight and summarize all the results we have discussed so far, we also present Figures 1 and 2, where the same results are portrayed from a clear graphical viewpoint.

In particular, Figure 1 takes into account the inversion of the SARS-COV-2 case trend of the following countries: Austria, Belgium, Croatia, Czechia, Denmark, Finland, France, Germany, Hungary, and Italy.
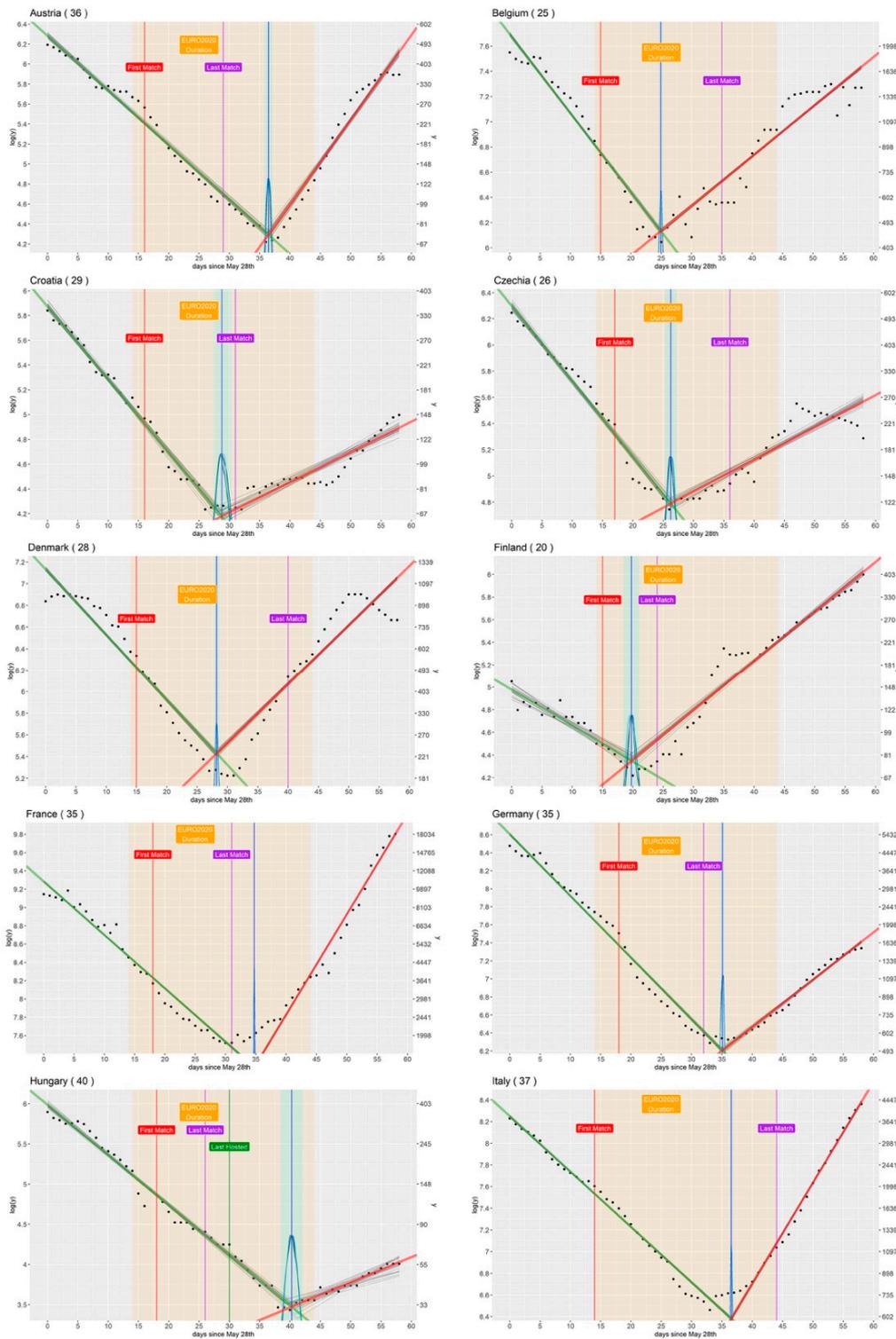
**Figure 1.** Inversion of the SARS-COV-2 case trend for Austria, Belgium, Croatia, Czechia, Denmark, Finland. France, Germany, Hungary, and Italy, occurring not later than 2–3 weeks after their first match. Yellow space: duration of the tournament. Red vertical line: first match. Purple vertical line: last match. Green vertical line: last hosted match. Blue vertical line: changepoint. Blue space: CI amplitude for the changepoint. Blue bell-shaped peaks: peaks of the probability density function for the changepoint. Green segment: case rate trend before the changepoint. Red segment: case rate trend after the changepoint. Grey segments: fitted lines drawn randomly from the posterior distribution, based on the corresponding CI. Black dots: number of daily CARS-COV-2 cases. Rightmost *y* axis: number of cases. Leftmost *y* axis: logarithm of the number of cases.

**Figure 2.** Inversion of the SARS-COV-2 case trend for the Netherlands, North Macedonia, Poland, Slovakia, Spain, Switzerland, and Ukraine, occurring not later than 2–3 weeks after their first match. Yellow space: duration of the tournament. Red vertical line: first match. Purple vertical line: last match. Green vertical line: last hosted match. Blue vertical line: changepoint. Blue space: CI amplitude for the changepoint. Blue bell-shaped peaks: peaks of the probability density function for the changepoint. Green segment: case rate trend before the changepoint. Red segment: case rate trend after the changepoint. Grey segments: fitted lines drawn randomly from the posterior distribution, based on the corresponding CI. Black dots: number of daily CARS-COV-2 cases. Rightmost *y* axis: number of cases. Leftmost *y* axis: logarithm of the number of cases.

Figure 2, instead, shows the inversion of the SARS-COV-2 case trend of the Netherlands, North Macedonia, Poland, Slovakia, Spain, Switzerland, and Ukraine. All the relevant information needed to interpret the two figures was inserted in the corresponding captions.

We used two separate figures, just for the sake of manageability. At the end, also based on an analysis of these figures, we can maintain that these results are fully compatible with the tournament being a factor.

All the five remaining countries (i.e., Portugal, Russia, Sweden, Turkey, and the UK), instead, break the pattern and cannot be considered an evidence in favor of the research hypothesis. In particular: (i) Portugal, Russia, and the UK show a robust increasing trend in the SARS-COV-2 infection case, starting well before the beginning the tournament; hence, the detected changepoints, as well as the relative slopes, cannot considered to be evidence in favor the hypothesis; (ii) Turkey seems to show quite a regular pattern, with a well identifiable changepoint and the usual inverting trend in the case rate; nonetheless, the problem is that that changepoint happens well after the team left the competition, more than four weeks since its first match; and (iii) finally, for Sweden, the model fails to fit because there seem to be two different changepoints, that are either before or after the tournament, making them irrelevant. The situations mentioned above are illustrated in Figure 3, where it is evident that all those five countries break the pattern. Again, all the relevant information needed to interpret Figure 3 was inserted in the corresponding caption.
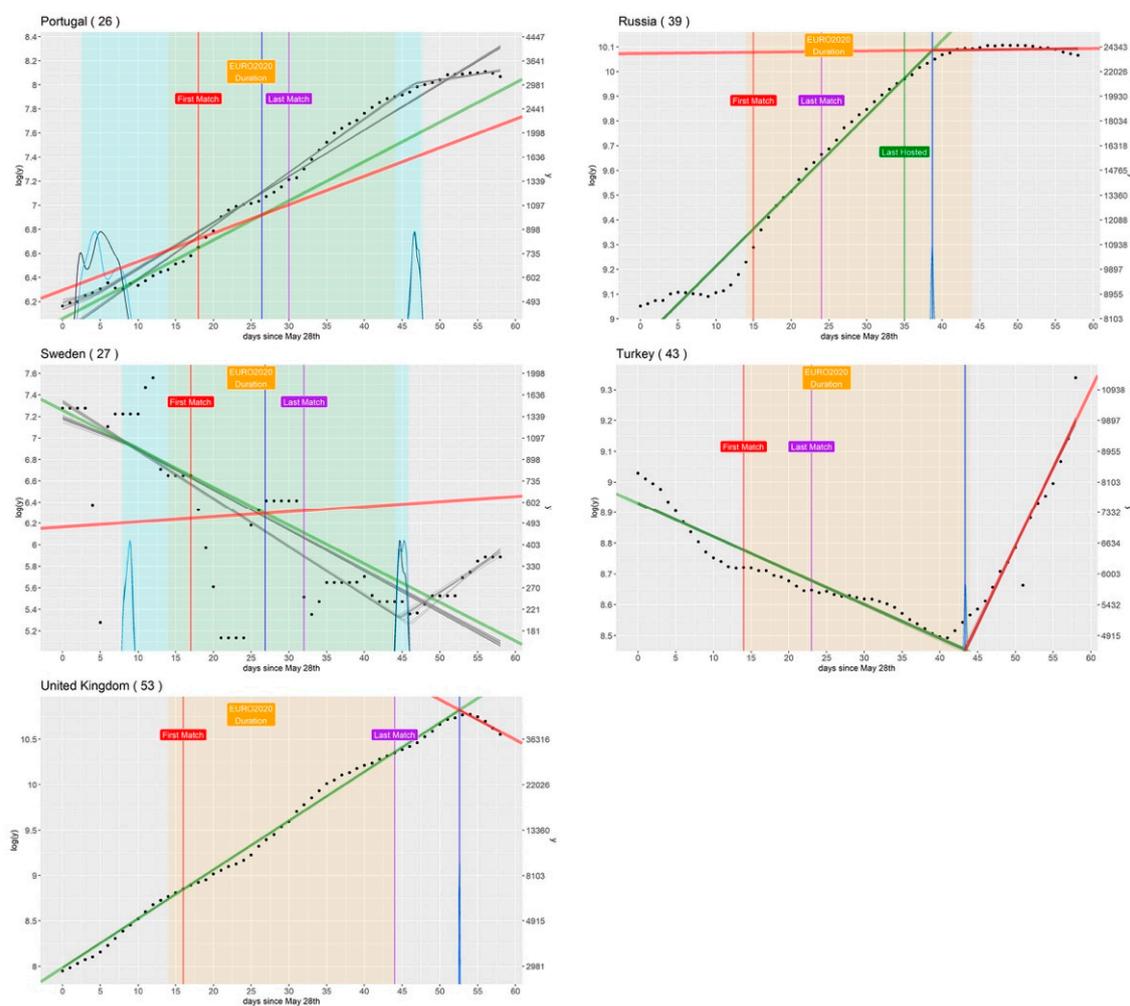


**Figure 3.** Portugal, Russia, Sweden, Turkey, and the UK break the pattern, without: (i) a well-recognizable changepoint and (ii) a reversal from a decrease to an increase in the SARS-COV-2 case rate, occurring not later than 2–3 weeks after the beginning of the tournament. Yellow space: duration of the tournament. Red vertical line: first match. Purple vertical line: last match. Green vertical line: last hosted match. Blue vertical line: changepoint. Blue space: CI amplitude for the changepoint. Blue bell-shaped peaks: peaks of the probability density function for the changepoint. Green segment: case rate trend before the changepoint. Red segment: case rate trend after the changepoint. Grey segments: fitted lines drawn randomly from the posterior distribution, based on the corresponding CI. Black dots: number of daily CARS-COV-2 cases. Rightmost $y$ axis: number of cases. Leftmost $y$ axis: logarithm of the number of cases.

Finally, Table 3 reports the value of τ, *diff*, and of all the other parameters, with the corresponding 95% CI. Of particular interest, here, is the large excursion in the CIs for Sweden and Portugal that witnesses the peculiarity of that situation.

**Table 3.** Regression parameters for the five countries that break the pattern.

| Country (Participating in the Tournament) | τ (Changepoint, avg. Value and 95% CI) | Diff (Days Separating τ from First Match) | $b_1$ (Angular Coefficient before τ, avg. Value and 95% CI) | $b_2$ (Angular Coefficient after τ, avg. Value and 95% CI) | $a_1$ (Intercept before τ, avg. Value and 95% CI) |
|---|---|---|---|---|---|
| Portugal | 26.4 (2.4, 47.7) | 8 | 0.03 (0.01, 0.05) | 0.08 (0.08, 0.09) | 6.28 (6.25, 6.32) |
| Russia | 38.7 (38.4, 39.0) | 24 | 0.03 (−0.03, −0.03) | 0.00 (0.00, 0.00) | 8.91 (8.90, 8.91) |
| Sweden | 26.9 (7.9, 45.8) | 10 | −0.04 (−0.05, −0.02) | 0.00 (−0.04, 0.05) | 7.26 (7.15, 7.36) |
| Turkey | 43.4 (43.1, 43.6) | 29 | −0.01 (−0.01, −0.01) | 0.05 (0.05, 0.06) | 8.93 (8.92, 8.94) |
| UK | 52.6 (27.8, 28.6) | 37 | 0.05 (0.05, −0.05) | −0.04 (−0.05, −0.04) | 7.99 (7.98, 7.99) |

For the sake of conciseness, we did not repeat, here again, the exercise to compute the halving/doubling times for those countries. Nonetheless, an interested reader could easily obtain those values by exploiting Formulas (2) and (3) in Section 2.2 and by using the correspondent data reported in Table 3.

*3.2. Countries That Did Not Participate in the Tournament*

While maintaining the pure observational nature of the inferences of our analysis about the effect of the tournament, we took advantage of another natural experiment, by observing what happened, during the tournament, in some 12 additional European countries that did not take part in the European football championship (considering the beginning of the tournament as the basis of our statistical observations).

This group was comprised of the following countries (with motivations for their choice reported in brackets): Greece and Ireland (great football traditions), Romania and Azerbaijan (hosting countries), Norway and Iceland (representatives of Northern Europe), Bulgaria and Moldova (representatives of Eastern Europe), Serbia and Bosnia (representatives of Balkans), and Latvia and Lithuania (largest countries representatives of Baltic Europe).

Needless to say, many other countries were left out. The motivations were manifold, ranging from their limited geographical dimensions (e.g., Malta, Faroe Islands, San Marino, Cyprus, Andorra, Montenegro, Kosovo, etc.) to geopolitical considerations, also in relationship with the game of football. For example: Georgia, Armenia, Kazakhstan, and Belarus are not famous for their international football traditions. Moreover, they are also well aligned with the contagion dynamics of one of their most influential neighboring countries, that is, Russia, which we had already examined.

The results of the application of our method to the above 12 countries are presented in Table 4. The 12 countries are listed based on the increasing value of *diff* (i.e., the number of days that separate the changepoint from the beginning of the tournament).

Here, it is important to remind what was already stated at the end of Section 2.2, that is: COVID-19 cases can show their biggest daily jumps 2–3 weeks after a particular mass event; hence, only those countries with inverting changes occurring in the time interval from 5–6 to 22–23 days after the beginning of the tournament were considered as those that have followed the pattern. This group is comprised of Greece, Azerbaijan, Ireland, Serbia—just 4 countries out of 12 (33%).

**Table 4.** Regression parameters for some 12 countries that did not participate in the tournament.
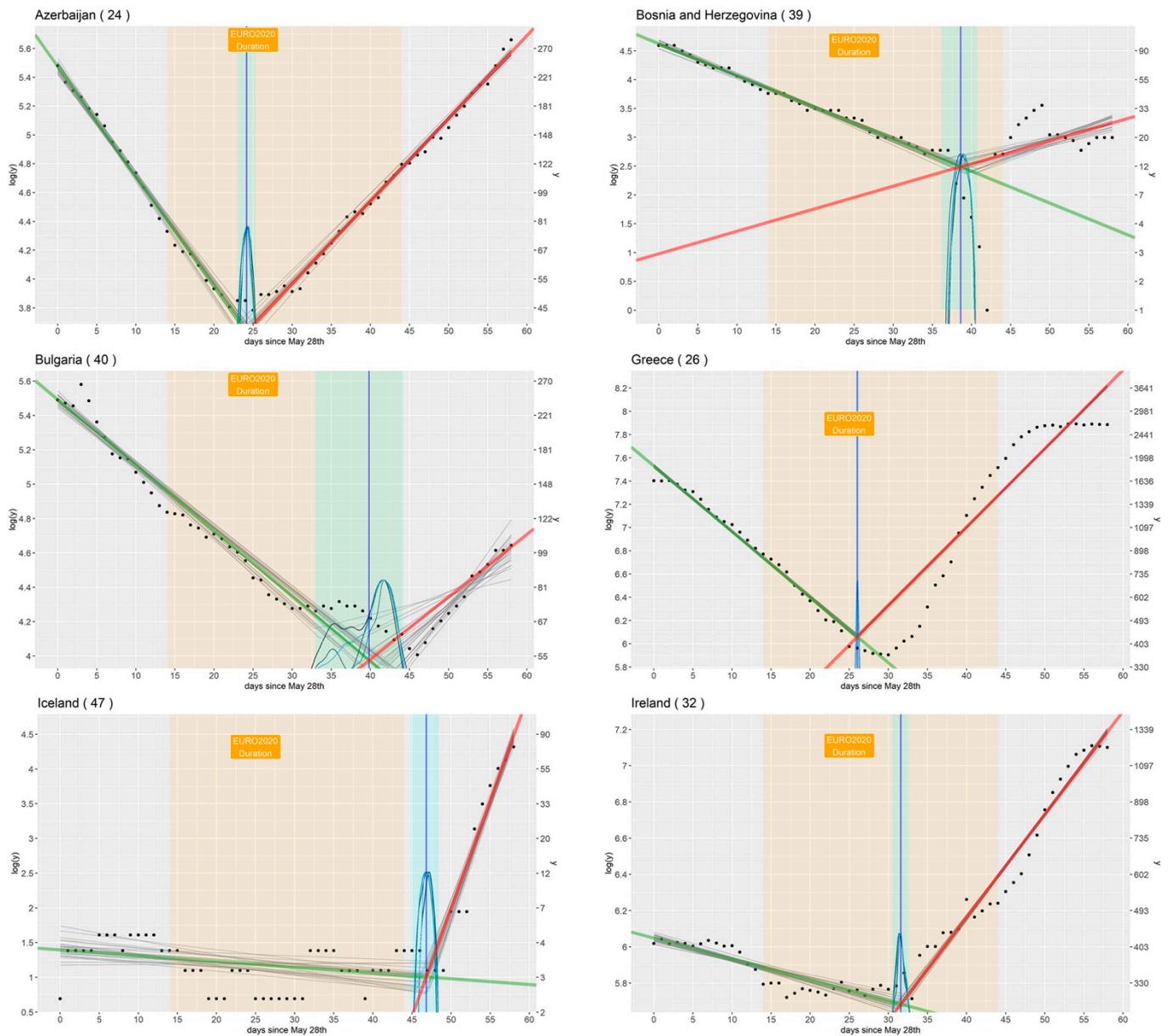
| Country (Participating in the Tournament) | τ (Changepoint, avg. Value and 95% CI) | Diff (Days Separating τ from Beginning of Tournament) | $b_1$ (Angular Coefficient before τ, avg. Value and 95% CI) | $b_2$ (Angular Coefficient after τ, avg. Value and 95% CI) | $a_1$ (Intercept before τ, avg. Value and 95% CI) |
|---|---|---|---|---|---|
| Moldova | 9.63 (6.52, 12.52) | −4 | −0.06 (−0.10, −0.03) | 0.01 (0.01, 0.02) | 4.36 (4.21, 4.50) |
| Norway | 16.56 (15.09, 17.98) | 3 | −0.05 (−0.06, −0.04) | −0.00 (−0.00, 0.00) | 6.03 (5.98, 6.07) |
| Azerbaijan | 24.16 (23.00, 25.32) | 10 | −0.08 (−0.08, −0.07) | 0.06 (0.05, 0.06) | 5.47 (5.41, 5.54) |
| Greece | 26.03 (25.77, 26.31) | 12 | −0.06 (−0.06, −0.06) | 0.07 (0.07, 0.07) | 7.53 (7.51, 7.55) |
| Ireland | 31.58 (30.53, 32.68) | 18 | −0.01 (−0.01, −0.01) | 0.06 (0.05, 0.06) | 6.05 (6.01, 6.08) |
| Serbia | 34.84 (33.49, 36.20) | 21 | −0.05 (−0.05, −0.04) | 0.05 (0.04, 0.06) | 5.83 (5.79, 5.87) |
| Lithuania | 39.12 (38.17, 40.08) | 25 | −0.08 (−0.08, −0.08) | 0.09 (0.08, 0.10) | 6.42 (6.39, 6.45) |
| Latvia | 45.07 (41.54, 48.44) | 31 | −0.05 (−0.05, −0.05) | 0.02 (−0.01, 0.05) | 5.91 (5.87, 5.94) |
| Romania | 37.60 (35.21, 39.90) | 24 | −0.06 (−0.06, −0.06) | 0.04 (0.03, 0.05) | 5.77 (5.72, 5.82) |
| Bosnia and Herzegovina | 38.61 (36.16, 40.76) | 25 | −0.06 (−0.06, −0.05) | 0.04 (0.02, 0.06) | 4.63 (4.55, 4.70) |
| Bulgaria | 39.82 (32.97, 44.20) | 26 | −0.04 (−0.04, −0.03) | 0.04 (0.01, 0.06) | 5.49 (5.44, 5.55) |
| Iceland | 46.82 (44.85, 48.43) | 33 | −0.01 (−0.02, 0.00) | 0.31 (0.27, 0.36) | 1.40 (1.11, 1.70) |

For all the other eight countries (67%), either their changepoint was premature (Norway and Moldova) or it came too late, precisely more than 23 days after the beginning of the tournament (Latvia, Lithuania, Romania, Bosnia, Bulgaria, and Iceland, in some cases, even without a clear case trend inversion, e.g., Bosnia).

As usual, with Figures 4 and 5, we portrayed a graphical representation of the same data of Table 4 for all the 12 countries of interest. Yet again, all the relevant information needed to interpret Figures 4 and 5 were inserted in the corresponding captions. We have used two separate figures, just for the sake of simplicity.

Finally, it is worth noting that we have not provided here again all the statistical information that we had computed for the countries participating in the tournament (e.g., various statistics, halving and doubling times, etc.). There is no precise motivation but that of brevity. Any interested reader could easily compute those statistics, with the data from Table 4. For example, halving and doubling times can be obtained by using the data from Table 4 along with the Formulas (2) and (3) of Section 2.2.

In conclusion, these final numbers have clearly shown that, while one could suppose that an increase in COVID-19 cases may have been an inevitable consequence of the general European situation in July 2021, the European football tournament, with its mass gatherings, played the important role of accelerator of this phenomenon, for many of its participating countries.

**Figure 4.** Azerbaijan, Bosnia, Bulgaria, Greece, Iceland, and Ireland not participating countries. Yellow space: duration of the tournament. Red vertical line: first match. Purple vertical line: last match. Green vertical line: last hosted match. Blue vertical line: changepoint. Blue space: CI amplitude for the changepoint. Green segment: case rate trend before the changepoint. Red segment: case rate trend after the changepoint. Blue bell-shaped peaks, grey segments, black dots, and rightmost and leftmost *y* axis: same as in previous figures.
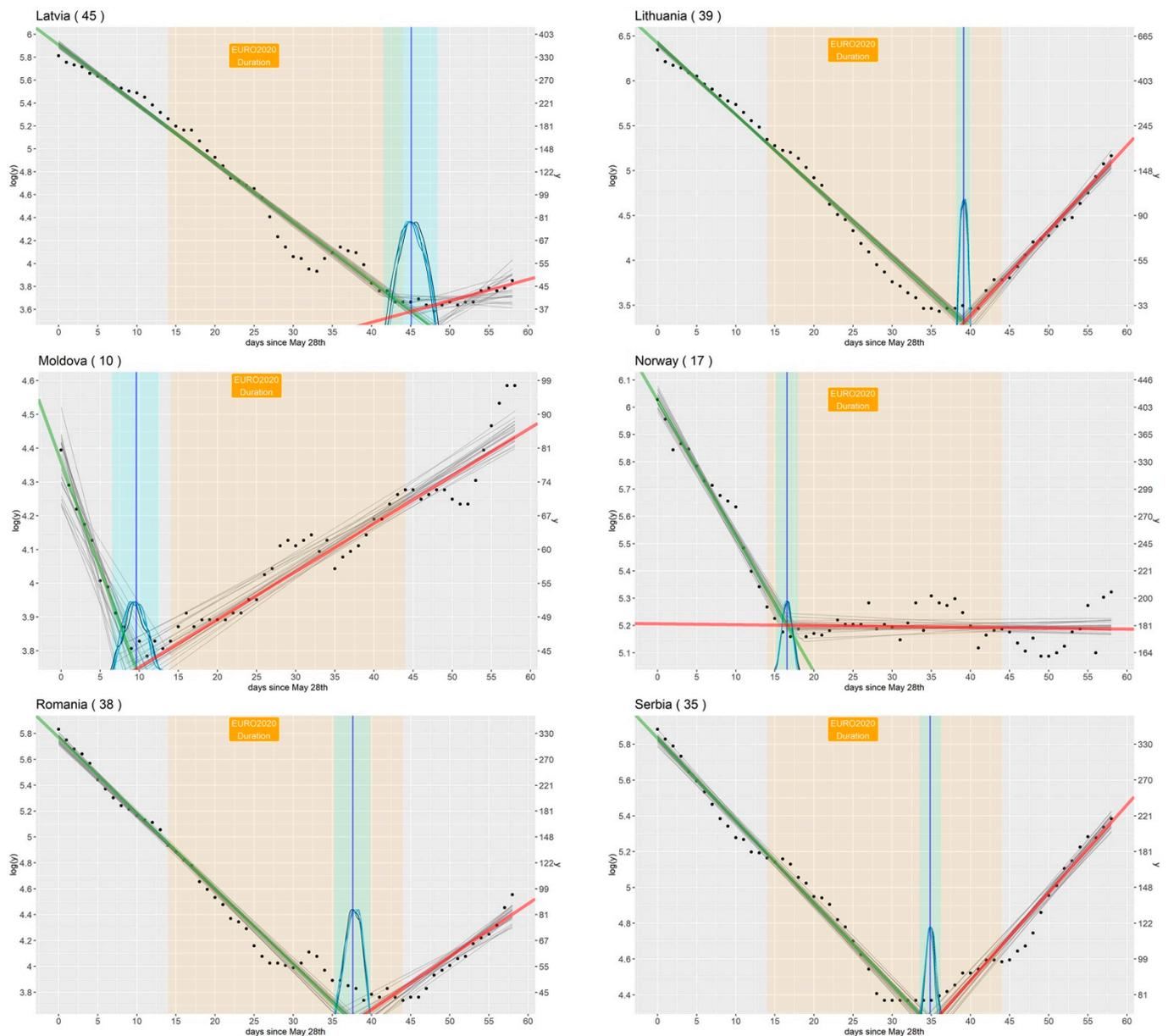
**Figure 5.** Latvia, Lithuania, Moldova, Norway, Romania, and Serbia not participating countries. Yellow space: duration of the tournament. Red vertical line: first match. Purple vertical line: last match. Green vertical line: last hosted match. Blue vertical line: changepoint. Blue space: CI amplitude for the changepoint. Green segment: case rate trend before the changepoint. Red segment: case rate trend after the changepoint. Blue bell-shaped peaks, grey segments, black dots, and rightmost and leftmost *y* axis: same as in previous figures.

## 4. Discussion and Conclusions

With this study, we found that, in 17 out of 22 (77%) countries involved in the 2020 European football championship, there has been a changepoint in the number of daily new SARS-COV-2 cases during the tournament, falling on average 14.97 days (95% CI 12.29–17.47) after the first match they played. Not only that, the case rate of the new daily infections was inverted for all these 17 countries, changing from a decreasing trend to an increasing one. We have quantified this inversion by measuring, for each national infection curve, the halving time before the change and the doubling time after it; they are respectively, on average: 18.07 days (95% CI 11.81–29.42) days and 29.10 days (95% CI 14.12–49.78).

There are five countries that break the pattern, and the presence of which could be seen as a first limitation of this study. Nonetheless, a careful consideration of the situation of these countries could provide a plausible explanation to this behavior. For example, it is evident that, for many of them (the UK, Russia, and Portugal), it is not possible to detect a changepoint in the infection rate which is coincidental with their participation in the tournament. This is because the inflation of the new COVID-19 cases was already in effect, in all these countries, when the tournament started, with the effect of the football championship probably absorbed into that inflation. The causes for this premature upturn of SARS-COV-2 cases are quite clear for the UK, which was the first European country to face the Delta variant. Portugal, instead, could have been the first European country to face the tourism impact, with many early tourists coming just from the UK. The situation in Russia, because of its enormous geographical extension, is, instead, too complex to look for a single explanation. Sweden, which reports COVID-19 numbers four days a week, entails a difficult interpretation, with our model not able to spot plausible changepoints. In regards of Sweden, it should not be forgotten that this was a country where very different strategies for managing the pandemic were adopted, without resorting, for example, to national lockdowns. Obviously, at the current stage of our research, no inference can be drawn regarding the existence of a relation between this fact and the results we achieved concerning this country. The situation for Turkey is different. It seems to follow the pattern, with an easily identifiable changepoint, coincidental with a reversion in the decrease/increase trend of the new COVID-19 cases. Nonetheless, this changepoint comes a bit too late (29 days after its first match). Hence, our decision was not to consider it as a further evidence in favor of the investigated link.

A second limitation of this study is that it ignored the possible effects of other confounding factors that could have played a role. Unfortunately, there are too many, and they are also too country specific, in many cases, to be considered as a whole. Nonetheless, the following two facts should also be considered. A general trend toward the decrease in the new daily SARS-COV-2 cases had already begun during the beginning of the 2021 spring, in almost all the considered countries, as an effect of the vaccination. In response to the benefits of the vaccines, almost all these European countries had consequently begun to lift the restrictions that were imposed to combat the third wave of the contagion. This happened well before the beginning of the tournament, and without any evident effect in terms of an upturn of new SARS-COV-2 cases (with the only exception of the already-discussed situation in the UK). It is a matter of fact, instead, that many infection clusters have surged in Europe during the football tournament. At the end, despite many possible country-specific confounding factors that could have played a role, our study has revealed that the temporal coincidence between the tournament and the inverting trend of the infections in many participating countries is an issue that cannot go unnoticed.

A third limitation touches more upon the mathematical and statistical nature of our analysis. We have already anticipated that our study is purely observational, without any possibility to demonstrate the existence of a clear relationship of cause and effect. Our intent was simply that of enquiring if all the mass gatherings following the football matches could have correlated with the virus resurgence in many European countries. For this reason, to study the plausibility of the correlation of interest, we have developed a simple model (similar to that employed in [20]) that does not possess the ambition of being exhaustive in the representation of the COVID-19 dynamics [21]; instead, it is very effective in detecting a changepoint in the infection curves, with the two corresponding slopes (before and after it) with which the decrease/increase case trends can be analyzed.

Finally, with an additional experiment, we have also demonstrated that the number of countries that follow the pattern falls down from 77% to 33% if we consider European countries that did not take part in the tournament. At the end, we can conclude that the results of our analysis are compatible with the hypothesis that most of the countries involved in the European football championship have seen a rise in the number of new SARS-COV-2 cases, or a slowdown in the fall, temporally coincident with their participation.

While this study has no ability to establish a final causal relationship, we think that the tournament, with its mass gatherings inside and outside the stadiums, has surely had an acceleration effect, that, coupled with the release of restrictions, could have given a contribution to ignite a new wave of the COVID-19 spread.

## References

1. Thomasson, E. German Minister Chides 'Irresponsible' UEFA over Euro 2020 Crowds. Reuters. 2021. Available online: https://www.reuters.com/world/europe/german-minister-slams-uefas-decision-fuller-stadiums-2021-07-01/ (accessed on 28 July 2021).
2. World Health Organization. Statement by Dr Hans Henri P. Kluge, WHO Regional Director for Europe. 2021. Available online: https://www.euro.who.int/en/media-centre/sections/statements/2021/statement-covid-19-the-stakes-are-still-high (accessed on 28 July 2021).
3. Henley, J.; Rankin, J. COVID: Euro 2020 Crowds 'a Recipe for Disaster', Warns EU Committee", The Guardian. 2021. Available online: https://www.theguardian.com/world/2021/jul/01/covid-euro-2020-crowds-a-recipe-for-disaster-warns-german-minister0 (accessed on 28 July 2021).
4. UEFA. Euro 2020 Key Information for Spectators. 2021. Available online: https://www.uefa.com/uefaeuro-2020/news/025b-0ef33753d7d0-100629325be2-1000--key-information-for-euro-spectators/ (accessed on 28 July 2021).
5. Italian Associated Press Agency (ANSA). Cluster of 91 COVID-19 Cases Linked to Euro 2020 Game. 2021. Available online: https://www.ansa.it/english/news/general_news/2021/07/16/cluster-of-91-covid-19-cases-linked-to-euro-2020-game_84349124-e130-453b-ade2-8b7136bd8993.html (accessed on 28 July 2021).
6. Kington, T. Italy's Euro 2020 Victory Tour Sent Rome Cases Rocketing. *The Times*, 22 July 2021, pp. 1–4. Available online: https://www.thetimes.co.uk/article/italys-euro-2020-victory-tour-sent-rome-cases-rocketing-r6m667r0b (accessed on 28 July 2021).
7. Peltier, E. Crowds for European Championship Soccer Games Are Driving Infections, the W.H.O. Says. *New York Times*, 1 July 2021, pp. 1–2. Available online: https://www.nytimes.com/2021/07/01/world/europe/euro-2020-covid-outbreak.html (accessed on 28 July 2021).
8. Skydsgaard, N.; Gronholt-Pedersen, J. Euro 2020 Crowds Driving Rise in COVID-19 Infections, Says WHO. Reuters. 2021. Available online: https://www.reuters.com/world/europe/who-warns-third-coronavirus-wave-europe-2021-07-01/ (accessed on 28 July 2021).
9. Schumacher, Y.O.; Tabben, M.; Hassoun, K.; Al Marwani, A.; Al Hussein, I.; Coyle, P.; Abbassi, A.K.; Ballan, H.T.; Al-Kuwari, A.; Chamari, K.; et al. Resuming professional football (soccer) during the COVID-19 pandemic in a country with high infection rates: A prospective cohort study. *Br. J. Sports Med.* **2021**, 1–11. [CrossRef]
10. Kochańczyk, M.; Grabowski, F.; Lipniacki, T. Super-spreading events initiated the exponential growth phase of COVID-19 with $\mathcal{R}0$ higher than initially estimated. *R. Soc. Open Sci.* **2020**, *7*, 200786. [CrossRef] [PubMed]
11. Weed, M.; Foad, A. Rapid Scoping Review of Evidence of Outdoor Transmission of COVID-19. *medRxiv* **2020**, 1–16. Available online: https://www.medrxiv.org/content/10.1101/2020.09.04.20188417v2 (accessed on 16 August 2021).
12. Cereda, D.; Tirani, M.; Rovida, F.; Demicheli, V.; Ajelli, M.; Poletti, P.; Trentini, F.; Guzzetta, G.; Marziano, V.; Barone, A.; et al. The early phase of the COVID-19 outbreak in Lombardy, Italy. *arXiv* **2020**, arXiv:2003.0932.
13. Mercker, M.; Betzin, U.; Wilken, D. What influences COVID-19 infection rates: A statistical approach to identify promising factors applied to infection data from Germany. *medRxiv* **2020**, 1–12. Available online: https://www.medrxiv.org/content/10.1101/2020.04.14.20064501v1 (accessed on 16 August 2021).
14. Signorelli, C.; Odone, A.; Riccò, M.; Bellini, L.; Croci, R.; Oradini-Alacreu, A.; Fiacchini, D.; Burioni, R. Major sports events and the transmission of SARS-CoV-2: Analysis of seven case-studies in Europe. *Acta Biomed.* **2020**, *91*, 242–244. [CrossRef]
15. El Hassan, N. Deloitte's Sports Business Group Estimates That Football Money League Clubs Will Miss out on Revenue of over €2 Billion by End of the 2020/21 Season Due to the COVID-19 Pandemic. 2021. Available online: https://www2.deloitte.com/xe/en/pages/about-deloitte/articles/deloittes-sports-business-group-estimates-football-money-league-miss-out-revenue-over-2euros-billion-end-202021-due-covid.html (accessed on 28 July 2021).

16. Our World in Data. COVID-19 GitHub Repository. 2021. Available online: https://github.com/owid/covid-19-data/blob/master/public/data/README.md (accessed on 28 July 2021).
17. Wikipedia. European Football Championship 2020. Available online: https://en.wikipedia.org/wiki/UEFA_Euro_2020 (accessed on 28 July 2021).
18. Lindeløv, J.K. mcp: An R Package for Regression with Multiple Change Points. 2020. Available online: https://osf.io/fzqxv/ (accessed on 28 July 2021).
19. Sebastiani, G.; Palù, G. COVID-19 and School Activities in Italy. *Viruses* **2020**, *12*, 1339. [CrossRef] [PubMed]
20. Casini, L.; Roccetti, M. Reopening Italy's schools in September 2020: A Bayesian estimation of the change in the growth rate of new SARS-CoV-2 cases. *BMJ Open* **2021**, *11*, e051458. [CrossRef]
21. Tosi, D.; Campi, A. How Data Analytics and Big Data Can Help Scientists in Managing COVID-19 Diffusion: Modeling Study to Predict the COVID-19 Diffusion in Italy and the Lombardy Region. *J. Med. Internet Res.* **2020**, *22*, e21081. [CrossRef] [PubMed]