BAYESIAN DETECTION AND UNCERTAINTY QUANTIFICATION OF THE FIRST CHANGE POINT OF THE COVID-19 CASE CURVE IN THE MIDWEST: TIMELINESS OF NON-PHARMACEUTICAL INTERVENTIONS.

Anonymous authors
Paper under double-blind review

ABSTRACT

In this work, we analyze by means of Bayesian methods the relationship between the first change point in the curve of COVID-19 cases in the Midwest and its position with respect to state executive orders, specifically the ”Face Mask” and ”Stay at Home” orders. We focus our attention on the twelve states of the Midwest. Estimation and uncertainty quantification of the first change point are provided. We further test the null hypothesis that the first change point arrives after those orders via the Savage–Dickey density ratio test. We find that the first qualitative change in the COVID-19 curve in the Midwest states precedes the ”Face Mask” and ”Stay at Home” orders, with the possible exclusion of Illinois, where our analysis sets the first change point in between the two state orders.

1 INTRODUCTION

COVID-19 hit the world in December 2019 and is still possibly the biggest public health concern (World Health Organization). The first case was identified in Wuhan, China, after which it spread throughout Europe and the US, leading to an ongoing pandemic, as officially determined by WHO in March 2020 (World Health Organization). Since the very first stages of the pandemic, the global research community mobilized and started to study the evolution of COVID-19 to understand its virology, pathophysiology, and epidemiology (Ackermann et al., 2020; Wiersinga et al., 2020; Vabret et al., 2020; Huang et al., 2020). The complexity of the problem requires the development of new methodologies and the collaboration of large interdisciplinary teams.

Our team joined this interdisciplinary research effort with the interest of understanding the dynamics of the disease from a machine learning perspective. We want to understand the time evolution of COVID-19 and in particular its changes with respect to non-pharmaceutical interventions (eg. lockdowns, social distancing, face mask, stay at home, and many others). In this manuscript, we will concentrate on one single aim: understanding the relationship between qualitative changes in the curve of COVID-19 cases and two government policy orders: ”Face Mask” and ”Stay at Home”.

Our analysis develops a Bayesian multivariate change point strategy (Verdinelli & Wasserman, 1995; Consonni & Veronese, 2008; Bürkner, 2017; Wagenmakers et al., 2010; Muggeo, 2003; Wetzels et al., 2010) to determine the distributional properties of the first change point in the evolution of COVID-19 cases in the twelve states of the Midwest. The Midwest is defined as that region situated in the north central United States that includes Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin (United States Census Bureau). We provide an estimate and uncertainty quantification of the first change point and we test the null hypothesis that the first change point occurs after ”Face Mask” and ”Stay at Home” orders via the Savage–Dickey density ratio test.

We find that the first qualitative change in the COVID-19 curve in the Midwest states precedes the ”Face Mask” and ”Stay at Home” orders, with the possible exclusion of Illinois, where our analysis sets the first change point in between the two state orders. Going forward, this analysis underlines the importance of prompt, reactive government decision-making in adopting non-pharmaceutical interventions. As possible qualitative and dramatic variations of COVID-19 can be hard to predict, it is critical that governments act rapidly to maximize effectiveness and avoid waste of possible advancements such as vaccinations, whose results have been promising at the time of writing of this paper (Polack et al., 2020; Knoll & Wonodi, 2021).

To be noted: We are not interested in the best overall fit for the evolution of the disease, while we want to describe precisely the first time point where the slope of the curve of COVID-19 cases changes substantially and position this with respect to two specific non-pharmaceutical interventions (”Stay at Home” and ”Face Mask” orders). For this reason, we found it reasonable to develop a simple segmented linear model (James et al., 2013), where the segments join at the change point.

The remaining part of this manuscript is organized as follows. Section 2 is dedicated to the methods, Section 3 to our results and a discussion, while in Section 4, we draw our conclusions.

2 METHODS

In this section, we describe the data sources, the Bayesian models of change points that we are considering, the Savage–Dickey density ratio test, and summarize our analysis.

2.1 DATA SOURCES AND SOFTWARE

The case counts by state were taken from the CDC (Centers for Disease Control and Prevention, COVID-19 Response), beginning with the first case in Washington reported on January 22, 2020 until February 21, 2021. The state policies, including dates and information on the ”Stay at Home” and ”Face Mask” orders, were taken from the COVID-19 US State Policy Database (CUSP) curated by Boston University (Raifman et al., 2020). The analysis was performed using the software R and its packages *mcp* and *patchwork*. All data is publicly available and code is available upon request.

2.2 BAYESIAN CHANGE POINT ESTIMATION

To estimate the change point we will use a Bayesian perspective. Although, the methodology can be adapted to multiple change points, we will concentrate on the case of one single change point as it is the only case reported in this manuscript. Consider a sequence of observations of an outcome variable Y (in our case the COVID-19 case counts), given by $y_1, \dots, y_T$ with $T > 0$ the time extension of our study (January, 22nd 2020 to February 21st, 2021) and $t = 1, \dots, T$ the corresponding time component. We model the mean response $\mu = E[Y]$ with a piece-wise linear function such as $\beta_1 t + \beta_2 \left(t - \psi\right)_+$, where $\left(t-\psi\right)_+ := \left(t - \psi\right) I\left(t > \psi\right)$ and $I(\cdot)$ representing the indicator function (Muggeo, 2003; Lindeløv, 2020). Here β_1 is the slope at the left of the change point ψ and β_2 is the difference-in-slopes between the slopes at left and right sides of ψ. We will estimate change points and their level of uncertainty with the mean and standard deviation of their posterior distribution via Montecarlo Markov Chain methods. The priors of all parameters are uninformative, with the the exception of the prior for the change point which is restricted to be ordered monotonically while otherwise remaining uninformative (Lindeløv, 2020).

2.3 SAVAGE-DICKEY RATIO TEST

In this subsection, we describe the Savage–Dickey density ratio, which is a method for Bayesian hypothesis testing (Wagenmakers et al., 2010; Dickey & Lientz, 1970; Lindley, 1972; O’Hagan & Forster, 2004; Wetzels et al., 2010; Verdinelli & Wasserman, 1995).

Suppose you observe data D and have the vector of parameters $\theta = (\theta_1, \theta_2)$ with θ_1 the parameters of interest, and θ_2 nuisance parameters. Consider a null hypothesis, $H_0 : \theta_1 = h$, with h a fixed vector of hypothesized values of θ_1. The alternative hypothesis is $H_0 : \theta_1 \neq h$. Denote p_0 and p_1 the probability density distributions under H_0 and H_1, respectively. Suppose that $\lim_{\theta_1 \to h} p_1(\theta_2|\theta_1) = p_0(\theta_2)$, then $p_1(\theta_2|\theta_1 = h) = p_0(\theta_2)$. Consider the Bayes factor

$$BF_{01} := p(D|H_0)/p(D|H_1) = p_0(D)/p_1(D).$$

Then

$$p_0(D) = \int p_0(D|\theta_2) p_0(\theta_2) d\theta_2 = \int p_1(D|\theta_2, \theta_1 = h) p_1(\theta_2|\theta_1 = h) d\theta_2 = p_1(D|\theta_1 = h),$$

which by Bayes' rule leads to

$$p_0(D) = \frac{p_1(\theta_1 = h|D)p_1(D)}{p_1(\theta_1 = h).}$$

In this way, we obtain the Savage-Dickey density ratio, namely the ratio between posterior and prior distributions:

$$BF_{01} = \frac{p_0(D)}{p_1(D)} = \frac{p_1(\theta_1 = h|D)}{p_1(\theta_1 = h)}.$$

In our case, we are interested in the parameter $\theta_1 = \psi$, the change point, although other parameters (eg. the two intercepts and two slops) will be estimated as well. The observed data is $D = \{(t, y_t)\}_{t=1}^T$. Note also that the hypothesis we are interested in is actually one sided $H_0 : \psi > h_i$ with $i = 1, 2$. In particular, we want to test if the change point ψ arrives after the "Stay at Home" order h_1 or not, and if it arrives after the "Face Mask" order h_2 or not. For more information about the Savage-Dickey density ratio test, we refer to (Lindeløv, 2020).

2.4 Our Analysis

We ran the algorithm described in Subsection 2.2 with $K = 9000$ iterations and 3 chains to estimate the parameter ψ. Our outcome variable Y is taken on the log scale and represents the natural logarithm of the cumulative case counts. We will have one Y for each of the twelve states in the Midwest. We estimated the posterior distribution of the change point parameter ψ, computed its posterior mean and its corresponding 95% credible interval for each of the twelve states in the Midwest. We compared this with the dates of the first case detected in each state and the dates of the "Stay at Home" and "Face Mask" orders. We performed the Savage-Dickey density ratio test to make this comparison.

3 Results and Discussion

In this section, we discuss our results. In Table 1, we collect the posterior means of the first change point, together with the upper and lower limit of its 95% credible interval (all of them rounded to the closest integer date) and the dates of the first case and "Stay at Home" and "Face Mask" orders. Figure 1 illustrates the trajectory of COVID-19 cases, estimates of the change points, and the dates of "Stay at Home" and "Face Mask" policy implementation for each of the Midwest states.

State	Illinois	Indiana	Iowa	Kansas	Michigan	Minnesota
First Case	24-01	06-03	08-03	08-03	10-03	06-03
Stay at Home	21-03	25-03	NO	30-03	24-03	28-03
Mask	01-05	27-07	16-11	03-07	27-04	24-07
First CP	28-02	07-04	29-04	11-04	01-04	27-04
LB CI	22-03	06-04	27-04	10-04	31-03	21-04
UB CI	23-04	08-04	01-05	14-04	02-04	02-05
State	Missouri	Nebraska	North Dakota	Ohio	South Dakota	Wisconsin
First Case	07-03	06-03	12-03	10-03	10-03	03-03
Stay at Home	06-04	NO	NO	24-03	NO	25-03
Mask	NO	04-05	14-11	23-07	NO	01-08
First CP	04-04	02-05	14-04	06-04	21-04	01-04
LB CP	03-04	30-04	10-04	04-04	19-04	03-04
UB CP	05-04	04-05	18-04	07-04	23-04	05-04

Table 1: This table provides the 2020 dates for all 12 Midwest states for: First Case of COVID-19 (Row 1) , Stay at Home order (Row 2), Face Mask order (Row 3), First Change Point (CP) ψ (Row 4), Date of the Lower Bound (LB) of the 95% Credible Interval (CI) for ψ (Row 5), Date of the Upper Bound (UB) for the 95% CI for ψ (Row 6). NO indicates when an order was not executed.

Illinois is the only state where we cannot reject the null hypothesis h_1 and so we cannot exclude the possibility that the first change point is subsequent to the "Stay at Home" order. Note that Illinois

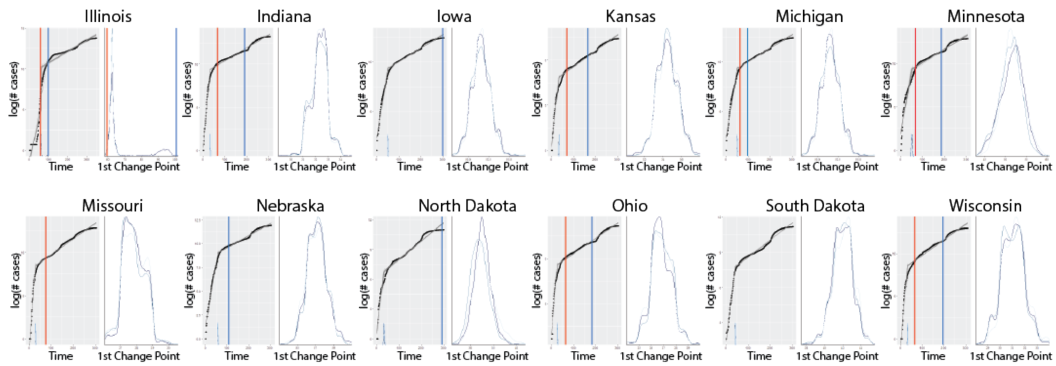

Figure 1: This figure illustrates the results of the Bayesian Change Point Analysis with comparison to the dates of the "Stay at Home" (Red Bar) and "Face Mask" (Blue Bar) orders for each of the 12 states in the Midwest. Each state has two plots. Left Plot: The horizontal axis represents the time variable, while the vertical represents the logarithm of the cumulative number of cases. Right Plot: Represents the posterior distribution of the first change point ψ.

saw the first case much earlier than the other states and registered a plateau soon after. Possibly related: Chicago is the biggest airline hub in the Midwest area by far, a fact that speculatively might be responsible for this impetus for the earlier crackdown on mask use and movement outside the home. The higher uncertainty of the estimate of the first change point in Illinois is possibly due to this plateau occurring at the beginning of the epidemic. The change points of Indiana, Kansas, Michigan, Minnesota, Ohio, and Wisconsin have been estimated to be before both governmental policies were put in place. Iowa and North Dakota did not execute a "Stay at Home" order, while the "Face Mask" order arrived much later than the estimated first change point. Missouri's policy recommended rather than required mask use, while its "Stay at Home" order was much later than the change point. Nebraska did not have a "Stay at Home" order and they mandated face mask use by employees only in public-facing businesses, and the first change point arrived before that. In South Dakota, there hasn't been any "Stay at Home" order, while masks were encouraged, but not required.

Altogether our results suggest that important government non-pharmaceutical interventions restricting movement outside the home and mandating the use of masks were put in place after a qualitative change in the COVID-19 case trajectory had already taken place. Thus, these government mandated policies were not a likely contributor to the observed first flattening in the curve of COVID-19 cases.

4 CONCLUSIONS

In this paper, we studied the problem of detecting the first change point in the curve of COVID-19 cases in the twelve Midwest states. We found evidence that there has been qualitative rate changes in the diffusion of COVID-19 before the "Stay at Home" and "Face Mask" orders were implemented, in all states but Illinois. This calls for possible quicker governmental actions. The analysis described in this manuscript is descriptive and not predictive, associative and not causal. Moving forward, we will attack those problems as well and to a larger scale. Note that although COVID-19 diffused with a similar timing in the Midwest, non-pharmaceutical interventions were largely heterogeneous. This would be a factor in extending our work to the detection of subsequent change points. A further frequentist analysis confirmed our results but was not reported here due to space limitations.

REFERENCES

Maximilian Ackermann, Stijn E. Verleden, Mark Kuehnel, Axel Haverich, Tobias Welte, Florian Laenger, Arno Vanstapel, Christopher Werlein, Helge Stark, Alexandar Tzankov, William W. Li, Vincent W. Li, Steven J. Mentzer, and Danny Jonigk. Pulmonary vascular endothelialitis, thrombosis, and angiogenesis in covid-19. *New England Journal of Medicine*, 383(2):120–128, 2020. doi: 10.1056/NEJMoa2015432.

Paul-Christian Bürkner. brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, 80:1–28, 2017.

Centers for Disease Control and Prevention, COVID-19 Response. United states covid-19 cases and deaths by state over time. URL `https://data.cdc.gov/d/9mfq-cb36`.

Guido Consonni and Piero Veronese. Compatibility of prior specifications across linear models. *Statistical Science*, 23:332–353, 2008.

James M. Dickey and B. P. Lientz. The weighted likelihood ratio, sharp hypotheses about chances, the order of a markov chain. *The Annals of Mathematical Statistics*, 41:214–226, 1970.

Chaolin Huang, Yeming Wang, Xingwang Li, Lili Ren, Jianping Zhao, Yi Hu, Li Zhang, Guohui Fan, Jiuyang Xu, Xiaoying Gu, Zhenshun Cheng, Ting Yu, Jiaan Xia, Yuan Wei, Wenjuan Wu, Xuelei Xie, Wen Yin, Hui Li, Min Liu, Yan Xiao, Hong Gao, Li Guo, Jungang Xie, Guangfa Wang, Rongmeng Jiang, Zhancheng Gao, Qi Jin, Jianwei Wang, and Bin Cao. Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. *The Lancet*, 395:497–506, 2020.

Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *Introduction to Statistical Learning*. Springer, 2013.

Maria Deloria Knoll and Chizoba Wonodi. Oxford-astrazeneca covid-19 vaccine efficacy. *The Lancet*, 397:497–506, 2021.

Jonas Kristoffer Lindeløv. mcp: An r package for regression with multiple change points. *OSF Preprints*, 2020. doi: 10.31219/osf.io/fzqxv.

D. V. Lindley. *Bayesian Statistics, A Review*. SIAM, 1972.

Vito M.R. Muggeo. Estimating regression models with unknown break-points. *Statistics in Medicine*, 22:3055–3071, 2003.

Anthony O'Hagan and Jonathan Forster. *Kendall's Advanced Theory of Statistics*, volume 2B of *Bayesian Inference*. Hodder Arnold, 2nd edition, 2004.

Fernando P. Polack, Stephen J. Thomas, Nicholas Kitchin, Judith Absalon, Alejandra Gurtman, Stephen Lockhart, John L. Perez, Gonzalo Pérez Marc, Edson D. Moreira, Cristiano Zerbini, Ruth Bailey, Kena A. Swanson, Satrajit Roychoudhury, Kenneth Koury, Ping Li, Warren V. Kalina, David Cooper, Robert W. Frenck, Laura L. Hammitt, Özlem Türeci, Haylene Nell, Axel Schaefer, Serhat Ünal, Dina B. Tresnan, Susan Mather, Philip R. Dormitzer, Uğur Şahin, Kathrin U. Jansen, and William C. Gruber. Safety and efficacy of the bnt162b2 mrna covid-19 vaccine. *New England Journal of Medicine*, 383(27):2603–2615, 2020. doi: 10.1056/NEJMoa2034577. PMID: 33301246.

Julia Raifman, Kristen Nocka, David Jones, Jacob Bor, Sarah Lipson, Jonathan Jay, and Philip A. Chan. Covid-19 us state policy database. 2020. URL `http://www.tinyurl.com/statepolicies`.

United States Census Bureau. Census regions and divisions of the united states. URL `https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf`.

Nicolas Vabret, Graham J. Britton, Conor Gruber, Samarth Hegde, Joel Kim, Maria Kuksin, Rachel Levantovsky, Louise Malle, Alvaro Moreira, Matthew D. Park, Luisanna Pia, Emma Risson, Miriam Saffern, Bérengère Salomé, Myvizhi Esai Selvan, Matthew P. Spindler, Jessica Tan, Verena van der Heide, Jill K. Gregory, Konstantina Alexandropoulos, Nina Bhardwaj, Brian D. Brown, Benjamin Greenbaum, Zeynep H. Gümüş, Dirk Homann, Amir Horowitz, Alice O. Kamphorst, Maria A. Curotto de Lafaille, Saurabh Mehandru, Miriam Merad, Robert M. Samstein, Manasi Agrawal, Mark Aleynick, Meriem Belabed, Matthew Brown, Maria Casanova-Acebes, Jovani Catalan, Monica Centa, Andrew Charap, Andrew Chan, Steven T. Chen, Jonathan Chung, Cansu Cimen Bozkus, Evan Cody, Francesca Cossarini, Erica Dalla, Nicolas Fernandez, John Grout, Dan Fu Ruan, Pauline Hamon, Etienne Humblin, Divya Jha, Julia Kodysh, Andrew Leader, Matthew Lin, Katherine Lindblad, Daniel Lozano-Ojalvo, Gabrielle Lubitz, Assaf Magen, Zafar Mahmood, Gustavo Martinez-Delgado, Jaime Mateus-Tique, Elliot Meritt, Chang Moon, Justine Noel, Tim O'Donnell, Miyo Ota, Tamar Plitt, Venu Pothula, Jamie Redes, Ivan Reyes Torres, Mark Roberto, Alfonso R. Sanchez-Paulete, Joan Shang, Alessandra Soares Schanoski, Maria Suprun, Michelle Tran, Natalie Vaninov, C. Matthias Wilk, Julio Aguirre-Ghiso, Dusan Bogunovic, Judy Cho, Jeremiah Faith, Emilie Grasset, Peter Heeger, Ephraim Kenigsberg, Florian Krammer, and Uri Laserson. Immunology of covid-19: Current state of the science. *Immunity*, 52(6):910–941, 2020. ISSN 1074-7613. doi: https://doi.org/10.1016/j.immuni.2020.05.002.

Isabella Verdinelli and Larry Wasserman. Computing bayes factors using a generalization of the savage–dickey density ratio. *Journal of the American Statistical Association*, 90:614–618, 1995.

Eric-Jan Wagenmakers, Tom Lodewyckx, Himanshu Kuriyal, and Raoul Grasman. Bayesian hypothesis testing for psychologists: A tutorial on the savage–dickey method. *Cognitive Psychology*, 60(3):158–189, 2010. ISSN 0010-0285. doi: https://doi.org/10.1016/j.cogpsych.2009.12.001. URL https://www.sciencedirect.com/science/article/pii/S0010028509000826.

Ruud Wetzels, Raoul P.P.P. Grasman, and Eric-Jan Wagenmakers. An encompassing prior generalization of the savage–dickey density ratio test. *Computational Statistics and Data Analysis*, 54:2094–2102, 2010.

W. Joost Wiersinga, Andrew Rhodes, Allen C. Cheng, Sharon J. Peacock, and Hallie C. Prescott. Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. *JAMA*, 324(8):782–793, 08 2020. ISSN 0098-7484. doi: 10.1001/jama.2020.12839.

World Health Organization. Rolling updates on coronavirus disease (covid-19). URL https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen.