


Extended Bayesian endemic–epidemic models to incorporate mobility data into COVID-19 forecasting

Dirk DOUWES-SCHULTZ, Shuo SUN* , Alexandra M. SCHMIDT, and Erica E. M. MOODIE 

Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Canada

Key words and phrases: Autoregressive model; Bayesian prediction; distributed-lag model; time series.

MSC 2020: Primary 62M20; secondary 62F15.

Abstract: Forecasting the number of daily COVID-19 cases is critical in the short-term planning of hospital and other public resources. One potentially important piece of information for forecasting COVID-19 cases is mobile device location data that measure the amount of time an individual spends at home. Endemic–epidemic (EE) time series models are recently proposed autoregressive models where the current mean case count is modelled as a weighted average of past case counts multiplied by an autoregressive rate, plus an endemic component. We extend EE models to include a distributed-lag model in order to investigate the association between mobility and the number of reported COVID-19 cases; we additionally include a weekly first-order random walk to capture additional temporal variation. Further, we introduce a shifted negative binomial weighting scheme for the past counts that is more flexible than previously proposed weighting schemes. We perform inference under a Bayesian framework to incorporate parameter uncertainty into model forecasts. We illustrate our methods using data from four US counties.

Résumé: La prévision du nombre de cas quotidiens de COVID-19 est cruciale pour la planification à court terme de ressources hospitalières et d'autres ressources publiques. Les données de localisation des téléphones mobiles qui mesurent le temps passé à la maison peuvent constituer un élément d'information important pour prédire les cas de COVID-19. Les modèles de séries chronologiques endémiques-épidémiques sont des modèles auto-régressifs récents où le nombre moyen de cas en cours est modélisé comme une moyenne pondérée du nombre de cas antérieurs multipliée par un taux auto-régressif (reproductif), plus une composante endémique. Les auteurs de ce travail généralisent les modèles endémiques-épidémiques pour y inclure un modèle à décalage distribué, et ce, dans le but de tenir compte du lien entre la mobilité et le nombre de cas de COVID-19 enregistrés. Pour saisir les variations de temps supplémentaires, ils y incorporent une marche hebdomadaire aléatoire d'ordre supérieur. De plus, ils proposent un schéma de pondération binomiale négative décalée pour les dénombrements passés, qui est plus flexible que les schémas de pondération existants. Ils utilisent l'inférence bayésienne afin d'intégrer l'incertitude des paramètres aux prédictions du modèle et ils illustrent les méthodes proposées avec des données provenant de quatre comtés américains.

1. INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus that causes the coronavirus disease 2019 (COVID-19), is thought to spread mainly through close contact

* Corresponding author: shuo.sun@mail.mcgill.ca

through the respiratory route (CDC, 2020a). In March 2020, the number of cases began to increase at alarming rates, as did the number of hospitalizations and deaths. The COVID-19 outbreak has been declared a global public health emergency. Predicting the evolution of the pandemic is important for short-term planning and for policy-making to slow the spread of COVID-19. In March 2020, American states and territories began widely implementing various executive “stay-at-home” orders to mitigate the risk of virus transmission. The extent to which people comply with those orders (i.e., reduce population movement) may meaningfully impact the number of confirmed COVID-19 cases and may be useful in refining short-term forecasts.

Forecasting COVID-19 cases has attracted considerable attention among researchers and has been accomplished using a variety of forecasting techniques and data sources. Among related studies, susceptible-infected-recovered-dead (SIRD) and autoregressive integrated moving average (ARIMA) models have been widely used (e.g., Anastassopoulou et al., 2020; Benvenuto et al., 2020; Fanelli & Piazza, 2020; Yonar et al., 2020; Yousaf et al., 2020). However, none have incorporated population mobility into prediction and instead have focused only on point forecasts, that is, without an uncertainty interval attached to forecasts. Some authors have taken account of the effect of “social distance” on the dynamics of COVID-19. For example, Zhang, Ma & Wang (2020) incorporated government intervention factors (e.g., stay-at-home orders, lockdowns, and quarantines) through binary, segmented time windows into a segmented Poisson model to identify the time of the peak in the daily number of new cases and predict the spread of COVID-19. Chiang, Liu & Mohler (2021) used Hawkes processes and mobility data to model COVID-19 transmission in the United States and estimated its dynamic reproductive number; however, the model’s confidence intervals for out-of-sample projections failed to cover the reported number of cases. In the literature, the majority of works do not take into account week and/or day-of-the-week effects, or lagged effects. Celani & Giudici (2021) proposed an endemic–epidemic (EE) model with a negative binomial distribution to understand the temporal and spatial contagion dynamics of COVID-19 in Italy. The authors’ approach was motivated by policy containment measures that limit social mobility but overlooked the lagged effects of population mobility and the day of the week. Grimée et al. (2022) incorporated weekend effects into an EE model using shifted Poisson weights and maximum likelihood estimation. The authors included mobility data in the matrix of between-area weights but not in the autoregressive rate.

EE models (Held, Höhle & Hofmann, 2005) are autoregressive count time series models designed to account for both endemic and epidemic contributions to infectious disease incidence. Broadly, in these models, the expected number of new cases is a weighted sum of the number of previous cases multiplied by an autoregressive rate, known as the epidemic component, plus an endemic component that accounts for other contributions to incidence, such as imported infections. Bauer & Wakefield (2018) showed that if the serial interval corresponds to one time step, the autoregressive rate can be interpreted as the local effective reproductive number, which is the average number of secondary infections produced by a single infectious individual in a population made up of both susceptible and nonsusceptible hosts (Cori et al., 2013). However, the authors implicitly assumed that there is no under-reporting or reporting delay. The presence of either can lead to important differences between the autoregressive rate and the effective reproductive number (Gostic et al., 2020; Bracher & Held, 2021), which means that EE models need to be interpreted with some care.

The EE framework may serve as a useful basis for performing COVID-19 forecasting that is able to incorporate and quantify the influence of population mobility on the spread of COVID-19. We focus on predicting the dynamics of the total number of reported infections in four counties in the United States: Hennepin, MN; King, WA; New York, NY; and Miami-Dade, FL. Washington was the first US state to announce a confirmed case and death. By mid-March 2020, Washington had the highest absolute number of confirmed cases and the highest number of confirmed cases per capita of any state in the United States (Perlstein, 2020) until it was

surpassed on April 10, 2020 by New York state, which quickly became an epicentre of the pandemic. By April 10, New York had more confirmed cases than any other country outside of the United States (Dzhanova, 2020). From June to September, Florida saw a surge in coronavirus cases and Miami was declared a new high-case region of the pandemic. For this reason, we select three populous counties, King, New York, and Miami-Dade, from three states along the East and West Coasts for our analysis. We further select another populous Midwestern county, Hennepin, to compare the temporal evolution of COVID-19 in different geographic locations.

There is a growing body of literature extending the EE framework (see, Paul, Held & Toschke, 2008; Held & Paul, 2012; Meyer & Held, 2014; Meyer & Held, 2017). We build on the most recent development from Bracher & Held (2022), which extended the EE framework to include multiple lags of the case counts. We focus on forecasting new daily reported cases of COVID-19 by extending the EE framework through a Bayesian inference procedure. In particular, we use Markov chain Monte Carlo methods to obtain samples from the resultant posterior distribution. Previous applications of the Bracher & Held (2022) EE model used frequentist inference and prediction that did not incorporate parameter uncertainty into forecasting. Additionally, previous weighting schemes for past case counts have not been very flexible, either allowing for only a decay in the weights or leading to weights that are very concentrated around their mean. Therefore, in this work, we introduce shifted negative binomial weights for the past case counts, which provide a more flexible weighting scheme than in Bracher & Held (2022) and include their shifted Poisson and geometric weights as special cases.

In order to incorporate mobility data into EE forecasts, we model the autoregressive rate as a function of lagged values of the proportion of individuals staying at home, as determined by mobile device location data, using distributed-lag models (DLMs) (Gasparrini, Armstrong & Kenward, 2010). This is motivated by the hypothesis that less time spent at home will result in more secondary infections and that, in turn, this effect will be lagged due to delays in diagnosis and reporting. As population mobility (i.e., the stay-at-home rate) might not capture all of the variability in the autoregressive rate across time, we also include day-of-the-week effects and a first-order random walk term that evolves at the weekly level.

This article is structured as follows. Section 2 proposes the model framework. Section 3 describes the data sources and presents the model specification and the results. We conclude with a discussion in Section 4.

2. METHODOLOGY

2.1. Endemic–Epidemic Model

Let y_t be the reported COVID-19 case count on day t for $t = p + 1, \dots, T$, where p is the number of lags and is fixed a priori. Let $\mathbf{y}_{(t-1):(t-p)} = (y_{t-1}, \dots, y_{t-p})^\top$ be lagged values of the case counts. We model y_t as following a negative binomial distribution, that is,

$$y_t | \mathbf{y}_{(t-1):(t-p)}; \boldsymbol{\theta}, r \sim \text{NegBin} \left(u_t \left(\mathbf{y}_{(t-1):(t-p)}, \boldsymbol{\theta} \right), r \right), \quad (1)$$

where $u_t \left(\mathbf{y}_{(t-1):(t-p)}, \boldsymbol{\theta} \right)$ is a positive function of lagged values of the outcome, $\boldsymbol{\theta}$ is a vector of model parameters that we will introduce shortly, and r is an overdispersion parameter such that $V \left(y_t | \mathbf{y}_{(t-1):(t-p)}; \boldsymbol{\theta}, r \right) = u_t \left(\mathbf{y}_{(t-1):(t-p)}, \boldsymbol{\theta} \right) \left(1 + u_t \left(\mathbf{y}_{(t-1):(t-p)}, \boldsymbol{\theta} \right) / r \right)$. We will drop the arguments of the function $u_t(\cdot, \cdot)$ from this point on to simplify notation. We use a model for daily counts and not a model for weekly counts as the latter can only forecast the total number of cases in the next week and not their distribution throughout the week, which may be of interest. For example, policymakers may wish to know if cases will likely increase or decrease throughout the next week.

We use an EE specification for u_t (Bracher & Held, 2022),

$$u_t = v_t + \phi_t \sum_{d=1}^p [\omega_d] y_{t-d}, \quad (2)$$

where the weight $[\omega_d] = \omega_d / \sum_{c=1}^p \omega_c$ is normalized and restricted to be positive, v_t is the endemic component of the model, and $\phi_t > 0$ is the autoregressive rate of the disease. The weight $[\omega_d]$ represents a relative contribution of past cases to current incidence. Bracher & Held (2022) considered geometric and shifted Poisson weights. However, these weights offer little flexibility: geometric weights only permit a decay with the lags and force the largest weight to be placed at a lag of one (Bracher & Held, 2022), while shifted Poisson weights are highly concentrated around their mean. When considering weekly data, like in Bracher & Held (2022), decaying weights may be reasonable when the average serial interval is less than 1 week, such as with COVID-19 (Nishiura, Linton & Akhmetzhanov, 2020). However, with daily data, a more flexible weighting scheme should be considered. Therefore, we propose the shifted negative binomial weights

$$\omega_d = \frac{\Gamma(d-1+q)}{(d-1)!\Gamma(q)} (1-\kappa)^q \kappa^{d-1}, \quad (3)$$

for $d = 1, \dots, p$, where $0 < \kappa < 1$ and $q > 0$ are parameters to be estimated.

Shifted negative binomial weights allow the most weight to be placed at lags greater than one and are more flexible than shifted Poisson weights. In fact, both the geometric weights and the shifted Poisson weights of Bracher & Held (2022) are special cases of Equation (3). Geometric weights arise when $q = 1$ and the shifted Poisson weights arise when q grows large. However, we consider a truncated distribution for the weights and find that there can be large relative differences in the tails of the Poisson and negative binomial distributions even when q is large. By the properties of the negative binomial distribution, if $q \leq 1$, then the weights will have a mode at $d = 1$ and decay with increasing lags. However, if $q > 1$, then the weights may either have a mode at $\lfloor \kappa(q-1)/(1-\kappa) \rfloor + 1$, if $\lfloor \kappa(q-1)/(1-\kappa) \rfloor + 1$ is less than p , or at p if $\lfloor \kappa(q-1)/(1-\kappa) \rfloor + 1$ is greater than p . The first scenario represents weights concentrated around some central lag, which describes the serial interval distribution of COVID-19 (Nishiura, Linton & Akhmetzhanov, 2020) that is thought to be centred around 3–5 days. The second scenario represents increasing weights, which could arise from strong cyclical patterns in the case counts. Bracher & Held (2022) also considered unrestricted weights, which would involve adding p parameters to the model, versus two parameters for the restricted shifted negative binomial weights. As we consider $p = 7$, $p = 14$, and $p = 21$ and have relatively short time series, we prefer restricted weights as unrestricted weights could lead to overfitting or model instability.

The endemic component v_t in Equation (2) represents infection not due to local sources (e.g., imported infections) and can also account for differences in reporting across time. Following Bracher & Held (2022), we model v_t as a log-linear function of fixed effects $(\zeta_0, \zeta)^\top$ and temporal covariates z_t , i.e.,

$$\log v_t = \zeta_0 + z_t^\top \zeta. \quad (4)$$

2.2. Autoregressive Rate

The autoregressive rate ϕ_t is a multiplier on the weighted previous daily cases at time t . We assume that ϕ_t is affected by lagged versions of the stay-at-home rate. The rationale is that population mobility can increase human interaction, which raises the risk of COVID-19

transmission. Additionally, due to delays between infection, diagnosis, and reporting, the current number of reported cases reflects transmission events from some time in the past. Therefore, the effect of changes in population mobility on reported case counts should be lagged to account for the delay from infection to observation. We also assume that a mobile device was brought every time an (anonymized) individual made a trip.

We propose using the DLM to characterize ϕ_t . This allows the effect of the stay-at-home rate to be distributed over a specific period of time, with certain coefficients describing each lag’s contribution. However, due to high correlation between lags in adjacent days, simply using a linear combination of those lags would result in collinearity and poor precision (Zanobetti et al., 2002; Gasparriani, Armstrong & Kenward, 2010). We therefore propose extending the EE model from Bracher & Held (2022) by adding constraints to the distributed lags, such as by imposing a smooth curve via polynomial functions or splines, to gain more precision.

Let s_t denote the stay-at-home rate for day t . In the presence of delayed effects, we propose modelling the autoregressive rate ϕ_t at a given day t using a DLM in terms of past values of s (e.g., s_{t-l} , where l is the lag representing the days between a past stay-at-home rate and the outcome ϕ_t). In matrix notation, with fixed L lags of the stay-at-home rate, we assume that

$$\log(\phi_t) = w_{\text{week}(t)} + \mathbf{x}_t^\top \boldsymbol{\gamma} + \mathbf{s}_t^\top \mathbf{C}\boldsymbol{\eta}, \tag{5}$$

where \mathbf{x}_t is a vector of temporal covariates that are separate from the stay-at-home rate, $\boldsymbol{\gamma}$ is a vector of fixed effects, and $\mathbf{s}_t = (s_{t-L_0}, s_{t-L_0-1}, \dots, s_{t-L})^\top$, where L_0 defines the minimum lag and L defines the maximum lag, with $L \geq 1$, $L_0 \leq L$, and $L_0, L \in \{0, 1, \dots\}$. Here, \mathbf{s}_t can be interpreted as a set of stay-at-home rate histories at lags over the lag period from L_0 to L ; the lag length is $L - L_0 + 1$. The unknown parameter vector $\boldsymbol{\eta}$ is associated with basis variables for the lag vector \mathbf{s}_t and has a length of h , and \mathbf{C} is a $(L - L_0 + 1) \times h$ matrix of basis variables derived from the application of a specific function to the lag vector \mathbf{s}_t . For example, if $\mathbf{C} = \mathbf{1}$ (i.e., a vector of ones), Equation (5) is a moving average model and if \mathbf{C} is the $(L - L_0 + 1) \times (L - L_0 + 1)$ identity matrix, Equation (5) describes a simple linear model. The matrix \mathbf{C} could also be defined through a polynomial or spline function to describe a nonlinear curve across lags. The effect $\boldsymbol{\beta}$ of the lags is represented as $\boldsymbol{\beta} = \mathbf{C}\boldsymbol{\eta}$. The choice of \mathbf{C} can be viewed as applying a constraint to the shape of the distributed-lag curve described by $\boldsymbol{\beta}$ (Gasparriani, Armstrong & Kenward, 2010). Under Equation (5), $\phi_t > 0$ and, therefore, u_t in Equation (2) is always strictly positive.

The component $w_{\text{week}(t)}$, where $\text{week}(t)$ is the week number corresponding to time t , is a time trend that evolves at the weekly level. The weeks are measured from Monday to Sunday starting from $\text{week}(p + 1)$. We assume that $w_{\text{week}(t)}$ follows a first-order random walk so that

$$w_{\text{week}} | w_{\text{week}-1} \sim N(w_{\text{week}-1}, \sigma_w^2), \tag{6}$$

where σ_w is a standard deviation that needs to be estimated. Bracher & Held (2022) did not include a random walk and thus assumed that all variability in the autoregressive rate over time can be captured by the covariates. The situation around COVID-19 is quite volatile and can change quickly in unexpected ways due to changes in policy, viral variants, and so on; thus, we include the random walk to capture weekly variability in the autoregressive rate that is not explained by the stay-at-home rate or \mathbf{x}_t . We include a random walk in ϕ_t and not v_t as the endemic component of the EE model typically only accounts for a very small share of the overall incidence (Bauer & Wakefield, 2018).

2.3. Epidemiological Interpretation of the Parameters

If we assume that y_t is the true number of new infections at time t and that individuals are negligibly infectious after p days, then ϕ_t can be interpreted as the effective reproductive

number and the $[\omega_d]$ s can be interpreted as the serial interval distribution of the disease or the distribution of the time between when an infector and an infectee develop symptoms (Quick, Dey & Lin, 2021). However, in practice, the actual number of daily infections is not fully observable. Instead, the reported number of cases is subject to reporting delay and under-reporting. Due to reporting delay, we would expect the autoregressive rate to reflect transmission events from further in the past compared with events that impact the effective reproductive number. This should be considered when choosing the lags for the mobility data in Equation (5). Additionally, under-reporting will have the effect of biasing the autoregressive rate towards zero relative to the effective reproductive number (Bracher & Held, 2021). Furthermore, changes to reporting, such as in testing eligibility, interest, or availability, could affect the autoregressive rate, while we would not expect these changes to impact the effective reproductive number. For example, an increase in testing availability could lead to an increase in the autoregressive rate even if there is no change in the amount of interaction between individuals. Finally, the weight $[\omega_d]$ will reflect both the serial interval distribution and different effects of reporting.

The above differences should be accounted for in model building, such as with the DLM in Equation (5) or through x_t and z_t . One needs to be careful to not overinterpret the results of the EE model. Explicitly accounting for reporting delay and under-reporting is the subject of ongoing research, and there are several challenges in and limitations with existing methods (Bracher & Held, 2021; Quick, Dey & Lin, 2021), a point we return to in the discussion.

2.4. Inference Procedure

Let θ represent the vector of all model parameters specifying u_t (i.e., $\theta = (\kappa, q, \zeta^T, \gamma^T, \beta^T, w_{\text{week}(p+1)-1}, w_{\text{week}(p+1)}, \dots, w_{\text{week}(T)}, \sigma_w)^T$) and let $y = (y_{p+1}, \dots, y_T)^T$ be the vector of case counts starting after the first lag p . The likelihood function is

$$p(y | \theta, r) = \prod_{t=p+1}^T p(y_t | y_{(t-1):(t-p)}, \theta, r),$$

where $p(y_t | y_{(t-1):(t-p)}, \theta, r)$ is the probability function associated with the negative binomial distribution specified in (1). We take a Bayesian approach so that all uncertainty in the estimation of the parameters is accounted for in the prediction of COVID-19 cases for future instants of time. In contrast, the frequentist approach in Bracher & Held (2022) relies on plug-in forecasts, that is, forecasts with the estimated parameters plugged in, and therefore does not account for parameter uncertainty that will in turn affect forecasting precision (see Chapter 16.2 in Pawitan (2001), for example).

By Bayes' theorem, the posterior distribution $p(\theta, r | y)$ of $(\theta, r)^T$ is proportional to $p(y | \theta, r)p(\theta, r)$, where $p(\theta, r)$ is the prior distribution of $(\theta, r)^T$. Because we have a hierarchical structure in our model, the prior factors as,

$$p(\theta, r) = \left(\prod_{\text{week}=\text{week}(p+1)}^{\text{week}(T)} p(w_{\text{week}} | w_{\text{week}-1}, \sigma_w) \right) p(w_{\text{week}(p+1)-1})p(\sigma_w)p(\kappa, q, \zeta, \gamma, \beta).$$

We assign normal priors with means of zero and variances of 100 to all unbounded parameters. In order to avoid overfitting in the random walk, we assign a half-Cauchy prior distribution with a mode at zero to σ_w . The Poisson distribution is a special case of the negative binomial distribution where $r = \infty$. Therefore, in order to penalize the complexity of the model, we assign a half-Cauchy prior distribution to $1/r$, which gives a reasonable prior probability close to the Poisson model. Note that $\exp(\zeta_0)$ represents the rate of imported cases per day at a baseline of

the covariates. The posterior of ζ_0 often has a long, flat tail from which it is difficult to sample. Therefore, we use the slightly informative prior distribution $\zeta_0 \sim \text{Unif}(-2, 50)$, which represents a lower bound of one case imported every 10 days – a reasonable lower bound for a large county.

Some care is needed when specifying prior distributions for q and κ . Figure 1 (bottom) shows that using a $\text{gamma}(0.1, 0.1)$ prior distribution for q and a uniform prior distribution for κ leads to quite an informative prior distribution for the shifted negative binomial weights, where very little prior probability is put on the weights for moderate lags. Therefore, we use a mixture uniform prior distribution for q ,

$$p(q) = \frac{1}{p} \mathbb{I}[0 < q < 1] + \frac{p-2}{p(p-1)} \mathbb{I}[1 < q < p] + \frac{1}{p(L_{\max} - p)} \mathbb{I}[p < q < L_{\max}],$$

where p is the maximum lag, L_{\max} is the upper bound of the final uniform distribution of the mixture, and $\mathbb{I}[\cdot]$ is an indicator function. Based on the discussion of the properties of the shifted negative binomial weights in Section 2.1, the mixture uniform prior distribution for q places equal prior probability on the following three scenarios: the weights are decreasing; the weights have a mode between lags 1 and p ; and the weights are increasing. Empirically, it appears that a larger L_{\max} leads to more prior probability being placed on the weight for the last lag. As we believe the weights should decay to zero over time, we use $L_{\max} = 10$ for $p = 7$, $L_{\max} = 18$ for

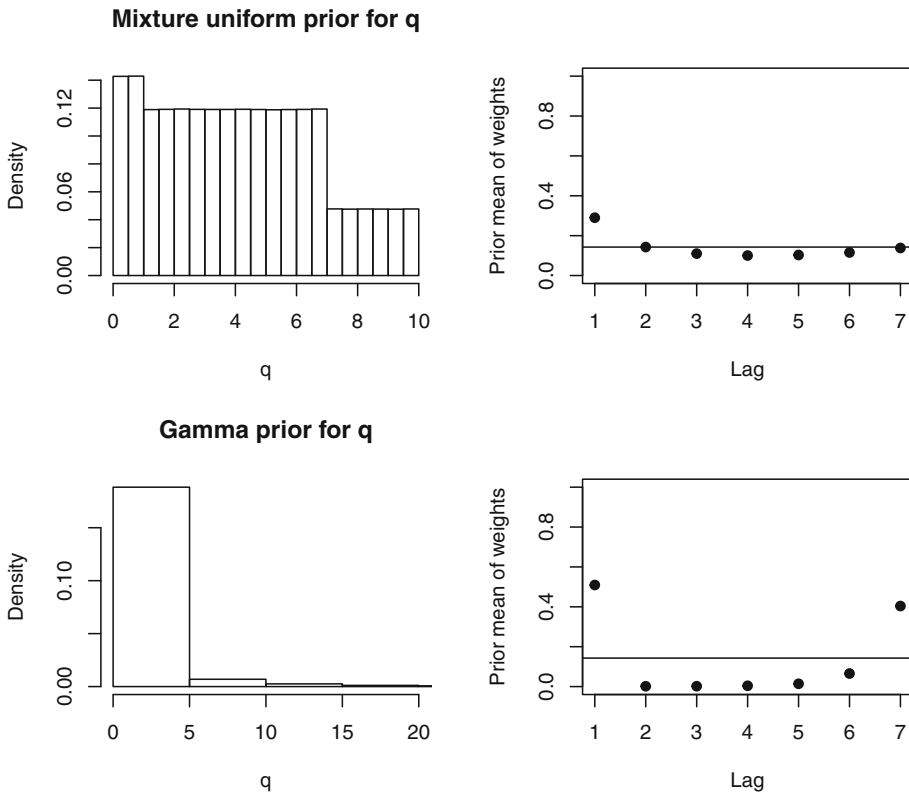


FIGURE 1: The left plots show the prior distribution for q under a mixture uniform prior distribution with $p = 7$ and $L_{\max} = 10$ (top), and a $\text{gamma}(0.1, 0.1)$ prior (bottom). The right plots show the resulting prior means of the shifted negative binomial weights $\{\omega_d\}$ for $d = 1$ (lag 1) to $d = 7$ (lag 7), where we assume that κ follows a uniform prior distribution on $(0, 1)$. The horizontal lines in the right plots are drawn at $1/7$, which represents a uniform prior distribution on the weights.

$p = 14$, and $L_{\max} = 27$ for $p = 21$, which keeps the value of L_{\max} close to p . Figure 1 shows that the mixture uniform prior distribution for q leads to a much more uniform prior distribution for $[\omega_d]$ compared with the gamma prior distribution. Therefore, the mixture uniform prior distribution is preferable to the gamma prior distribution as it is less informative regarding the weights.

As the posterior distribution $p(\theta, r | \mathbf{y})$ has no closed form, we use Markov chain Monte Carlo (MCMC) methods to obtain samples from the posterior distribution. More specifically, we use the R package Nimble (de Valpine et al., 2017), which relies on a Gibbs sampler with some steps of the Metropolis–Hastings and other sampling algorithms, to draw from the joint posterior distribution. Parameters that show high posterior correlations, such as q and κ , are sampled in blocks using automated factor slice sampling, which greatly improves mixing (Tibbits et al., 2014). Parameters that mix slowly but whose posterior distributions are not highly correlated are sampled using univariate slice sampling (Neal, 2003). We use Nimble as it accommodates custom distributions and can therefore easily accommodate our proposed mixture uniform prior distribution for q . The Nimble code used to fit the model and the data are available on Github (https://github.com/Dirk-Douwes-Schultz/extended_bayesian_EE_mobility).

2.5. Temporal Predictions

From a Bayesian point of view, predictions for future instants in time are obtained through the posterior predictive distribution, which naturally accounts for the uncertainty in the estimation of the parameter vector θ . Here the goal is to obtain K -step ahead forecasts of the number of cases. In this case, the posterior predictive distribution, $p(y_{T+K} | \mathbf{y})$, is

$$\begin{aligned} p(y_{T+K} | \mathbf{y}) &= \int p(y_{T+K} | y_{T+K-1}, \dots, y_{T+K-p}, \theta, r, w_{\text{week}(T+K)}) \\ &\quad \times p(y_{T+K-1} | y_{T+K-2}, \dots, y_{T+K-p-1}, \theta, r, w_{\text{week}(T+K-1)}) \\ &\quad \times \dots \times p(y_{T+1} | y_T, \dots, y_{T+1-p}, \theta, r, w_{\text{week}(T+1)}) \\ &\quad \times \prod_{\text{week}=\text{week}(T)+1}^{\text{week}(T+K)} p(w_{\text{week}} | w_{\text{week}-1}, \sigma_w) \\ &\quad \times p(\theta, r | \mathbf{y}) dy_{T+K-1} \dots dy_{T+1} d\theta dr dw_{\text{week}(T)+1} \dots dw_{\text{week}(T+K)}. \end{aligned}$$

The above integral can be approximated through Monte Carlo integration: once a sample from the posterior distribution of the parameters is available, we have that

$$\begin{aligned} p(y_{T+K} | \mathbf{y}) \\ \approx \frac{1}{Q-M} \sum_{m=M+1}^Q p\left(y_{T+K} | y_{T+K-1}^{[m]}, \dots, y_{T+K-p}^{[m]}, \theta^{[m]}, r^{[m]}, w_{\text{week}(T+K)}^{[m]}\right), \end{aligned} \quad (7)$$

where the superscript $[m]$ denotes a draw from the posterior distribution of the variable, M is the size of the burn-in sample, Q is the total MCMC sample size, and $y_t^{[m]} = y_t$ when $t \leq T$.

In practice, once a sample from the posterior distribution of the parameter vector is available, we use composition sampling to draw samples from the posterior predictive distributions. The sampling algorithm is described in Algorithm 1. For $k = 1, \dots, K$, Algorithm 1 obtains realizations from the resultant posterior predictive distribution $p(y_{T+k} | \mathbf{y})$. The prediction procedure requires $s_{T+K} = (s_{T+K-L_0}, s_{T+K-L_0-1}, \dots, s_{T+K-L})^\top$, so that we must have $K = L_0$. In our applications, we considered $L_0 = 7$ to predict cases in the next week using mobility data from the current and previous weeks.

Algorithm 1. Posterior predictive simulation

```

for  $m$  in  $M + 1 : Q$  do
  for week in  $(\text{week}(T) + 1) : \text{week}(T + K)$  do
    draw  $w_{\text{week}}^{[m]}$  from  $p(w_{\text{week}} | w_{\text{week}-1}^{[m]}, \sigma_w^{[m]})$ 
  end
  for  $k$  in  $1 : K$  do
    draw  $y_{T+k}^{[m]}$  from
     $p(y_{T+k} | y_{T+k-1}^{[m]}, \dots, y_{T+k-p}^{[m]}, \theta^{[m]}, r^{[m]}, w_{\text{week}(T+k)}^{[m]})$ , where
     $y_t^{[m]} = y_t$  for  $t \leq T$ 
  end
end

```

2.6. Model Comparison

To compare two models within the sample period, we use the Watanabe–Akaike information criteria (WAIC) (Gelman, Hwang & Vehtari, 2014). The model with the lowest WAIC is considered to have the best fit and, generally, a difference of five or more in the WAIC is considered meaningful. However, the WAIC is only an approximation of out-of-sample prediction error and might not accurately select the correct model. Similar to Bracher & Held (2022), to compare models out of sample, we use proper scoring rules (Gneiting & Raftery, 2007) approximated by draws from the posterior predictive distributions. Suppose that the model is fit using data up to time T . The logarithmic score for the k th-step-ahead prediction is

$$LS_{Tk} = -\log \left(p(y_{T+k}^{(\text{obs})} | \mathbf{y}) \right),$$

where $y_{T+k}^{(\text{obs})}$ is the observed future value and $p(y_{T+k}^{(\text{obs})} | \mathbf{y})$ is approximated by Equation (7) using draws from the posterior predictive distributions generated by Algorithm 1. The score is averaged across multiple time points to compute the comparison criterion

$$LS_k = \frac{1}{T_u - T_l + 1} \sum_{T=T_l}^{T_u} LS_{Tk},$$

where T_u and T_l denote the upper and lower bounds for T . The model with the lowest LS_k is considered the best at k -step ahead prediction (Bracher & Held, 2022). Because a model may perform better in short-term prediction but worse in long-term prediction, LS_k should be computed for a wide range of k . The initial time point T_l should be big enough so that all models being compared can produce stable estimates and T_u should be big enough to reduce random error.

3. DATA ANALYSIS

3.1. Data Description

We select four large counties in the United States for analysis: Hennepin, MN; King, WA; New York, NY; and Miami-Dade, FL. The USAFacts (<https://usafacts.org>) COVID-19 time series data are gathered from the Centers for Disease Control and Prevention (CDC) and state- and local-level public health agencies. County-level data are confirmed by referencing state and local

agencies directly. Hennepin, New York, and Miami-Dade counties' time series data are directly retrieved from the USAFacts website. For King county, county-level data from USAFacts are incomplete and inaccurately include negative numbers of daily new cases. Therefore, we collect King county time series data from the local public health department website (<https://www.kingcounty.gov/depts/health.aspx>). Cases, deaths, and per capita adjustments reflect cumulative totals since January 22, 2020. Confirmed daily cases are based on the date of report from January 2020 to January 2021. Data on the daily percentage of people staying at home are obtained from the Bureau of Transportation Statistics (BTS) and are estimated by the Maryland Transportation Institute and Center for Advanced Transportation Technology Laboratory at the University of Maryland. Daily travel estimates are from a mobile device data panel from multiple data sources that address geographic and temporal sample variation. The data only include mobile devices whose anonymized location data meet a set of data quality standards, which further ensures the overall quality and consistency of the data. These quality standards consider both the temporal frequency and spatial accuracy of anonymized location point observations, temporal coverage and representativeness at the device level, and spatial representativeness at the sample and county levels. A multilevel weighting method that employs both device- and trip-level weights is applied before travel statistics are computed, so the results are representative of the entire population in a nation, state, or county (Bureau of Transportation Statistics, 2020).

In the observed data, there are a small number of days with zero daily cases after the emergence date, which we posit is likely due to under-reporting or delayed reporting. We redistributed cases around days with zero reported cases by averaging the most recent 5 days ahead and then rounding to ensure that the number of cases remains integer-valued. Specifically, the average of the 5 days ahead of days with zero daily cases are deducted proportionally from these 5 days. The emergence date of COVID-19 is set as the time of the number of first nonzero cases for each county. Data from the 260 days (i.e., about 37 weeks) after the emergence date are used to fit a model for Hennepin, New York, and Miami-Dade, while 320 days (about 46 weeks) are used for King county. More days were chosen for King county because that data, collected from the local public health agency, are expected to have higher accuracy (CDC, 2020c). The number of new cases reported to the CDC each day fluctuates. There is generally less reporting during the holidays (e.g., Christmas) and more volatility from December 2020 to January 2021. The end date for each time series is chosen as the most recent time at which the time series is relatively stable (e.g., without containing a sequence of zeros).

The outcome of interest is the daily number of new, confirmed COVID-19 cases. The covariates are lagged daily case counts and lagged daily stay-at-home rates in each of the four counties. The primary aim of our analysis is to forecast the autoregressive rate and the number of new daily cases seven days ahead.

3.2. Model Specification and Fitting

The model in Equation (2) represents a general form with many possible specifications. For each model specification considered in this section, we run 100,000 MCMC iterations with three chains, each started from different points in the parameter space and each with a burn-in of 10,000 iterations. Convergence was checked using the Gelman–Rubin statistic (with an upper bound of 1.05 for all estimated parameters), the minimum effective sample size (with a lower bound of 1000), and trace plots (Plummer et al., 2006).

As there can be considerable variation in COVID-19 case counts across different days of the week (Grimée et al., 2022), we assume that z_t in Equation (4) and x_t in Equation (5) are composed of day-of-the-week indicators. Having an indicator for each day of the week in both z_t and x_t would add 12 parameters to the model, which is not small considering that our time series have lengths of 260 and 320. Therefore, we reduce the number of parameters associated with the day-of-the-week effects by fitting a model with full effects and then combining effects that are

TABLE 1: WAIC of the fitted EE models with different stay-at-home lag and weighting settings for each county.

County	Shifted Poisson weights	Shifted NB weights, lags of stay-at-home rate at time t (s_t)				
		None	s_t	s_{t-7}	$(s_{t-7}, \dots, s_{t-14})^T$	$(s_{t-7}, \dots, s_{t-21})^T$
Hennepin	2883	2907	2884	2883	<u>2873</u>	2880
King	3157	3165	3161	3165	<u>3151</u>	3161
New York	2577	2573	2575	<u>2550</u>	2558	2558
Miami-Dade	3479	3464	3465	3464	3464	<u>3458</u>

Note: The best model with the lowest WAIC for each county is underlined. As the model with shifted Poisson weights consistently fit more poorly than those with the shifted negative binomial weights, WAICs were computed for models with Poisson weights that correspond to the best-fitting (underlined) models with negative binomial (NB) weights model.

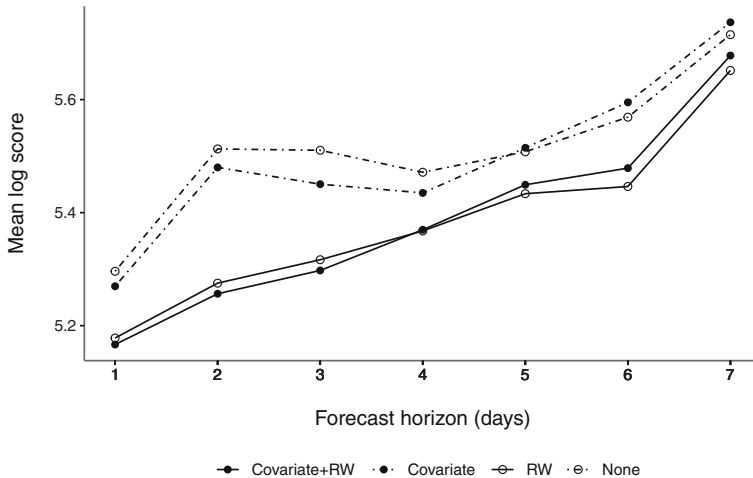


FIGURE 2: Mean log scores across forecast horizon (days) for King county. “RW” stands for the model with a weekly random walk effect. “Covariate” indicates the inclusion of the stay-at-home rate DLM in the autoregressive rate.

very similar and dropping insignificant effects. The other model specifications are determined by an initial version of Table 1 calculated without day-of-the-week effects. We then choose between the model with the reduced day-of-the-week effects and the model with no day-of-the-week effects using the WAIC. Based on this process, we include only a Monday indicator in x_t only for Miami-Dade, King, and New York counties, and a combined indicator of Monday and Friday in z_t only for Hennepin county.

The log score represents the performance of the model as if it were fit each day and used to forecast a week ahead, as described in Section 2.6. This is how we would expect the model to be used in practice, and, therefore, the log score is a useful tool for comparing models in this application. However, it is very computationally intensive to compute the log score, so we use the log score with the King county data to test the effectiveness of the random walk and the addition of the stay-at-home rate through the DLM. Figure 2 shows the mean log scores for each forecast horizon (i.e., 1, . . . , 7 days ahead) for the four models for the King county data. We compare a model that includes both the random walk and the stay-at-home rate DLM with

models that use exactly one or none of these components. For the DLM model, we choose a minimum lag of $L_0 = 7$ and a maximum lag of $L = 14$, as suggested by the WAIC (see Table 1). The lag of the case counts p in Equation (2) is set to 7 days for all models compared. We choose $T_l = 100$ and $T_u = 320$ for computing the mean log scores, which results in 221 fitted models for each specification (for 884 total models). From Figure 2, the random walk clearly improves forecasting performance at all horizons, especially when forecasting the first 5 days ahead. Because the random walk also provides a more accurate historical estimate of the autoregressive rate, we include the random walk component in the model for New York and Miami-Dade counties. The random walk component is not considered in the model for Hennepin county as the estimate of σ_w in (6) is approximately zero, which indicates no significant week effects and the WAICs with and without the random walk appear to be the same. There is uncertain evidence regarding the effectiveness of the stay-at-home rate DLM, based on the log score. For this reason, we use the WAIC to decide whether to include the DLM in the model for each county and to make decisions regarding other model specifications. We describe this process in more detail below.

First, we compare models with $p = 7$, $p = 14$, and $p = 21$ and find no significant difference in the WAIC for any county. Therefore, we use $p = 7$ for all counties as it allows the most data to be incorporated into the fitting. Table 1 gives some results of the model comparisons, specifically for determining the optimal lags in the DLM of the stay-at-home rate and for determining the optimal weighting scheme for past counts. We compared the shifted Poisson and shifted negative binomial weights and find the shifted negative binomial weights significantly improve the fit for all counties. We use a natural cubic spline for the basis matrix for the stay-at-home rate in Equation (5). For each county, we concentrate on four variants of Equation (5) that differ in $s_t = (s_{t-L_0}, s_{t-L_0-1}, \dots, s_{t-L})^\top$. According to the CDC, the incubation period of COVID-19 is thought to extend to 14 days, with a median time of 4–5 days from exposure to symptom onset (CDC, 2020b). Many categories of tests are used to detect COVID-19: some tests (e.g., antigen tests) provide results within minutes, while others (e.g., laboratory-based tests) require time for processing and may result in diagnostic delays due to the processing time and return times. For example, most nucleic acid amplification tests (NAATs) need to be processed in a laboratory and it may take 1–3 days for results to become available (CDC, 2021). According to an analysis of national survey data conducted by researchers at Northeastern University, Harvard University, Rutgers University, and Northwestern University, the average turnaround time for receiving COVID-19 test results dropped from 4.0 days (with a median of 3.0 days) in April 2020 to 2.7 days (with a median of 2.0 days) in September 2020 (Chwe et al., 2021). Considering the incubation period and delays in diagnosis and reporting, the observed number of cases y_t informs transmission some time before t . Therefore, it is reasonable to assume that ϕ_t is affected by lags of population mobility with a minimum lag of $L_0 > 0$. Specifically, we compare $(L_0 = 0, L = 0)$, $(L_0 = 7, L = 7)$, $(L_0 = 7, L = 14)$, and $(L_0 = 7, L = 21)$, and compare these to a null model that excludes the stay-at-home rate s_t as a covariate.

For the four counties, we find an optimal minimum lag (L_0) of 1 week for the stay-at-home rate in the DLM model in Equation (5) and an optimal maximum lag (L) of 1 week for New York county (i.e., only the past seventh day), 2 weeks for Hennepin and King counties, and 3 weeks for Miami-Dade county. This means that population mobility between 1 and 2 or 3 weeks in the past is the most relevant for determining current incidence.

3.3. Model Results

Table 2 gives the estimates of the posterior means for the best-fitting model in each county (underlined models in Table 1), where we exclude spline coefficients as they are not informative (see Figure 4 instead). King county has the largest overdispersion parameter while Hennepin

county has the smallest. This reflects the different case count trajectories in these counties: cases initially increased very quickly in King county, which led to a large amount of variability in the case numbers. The standard deviation of the random walk term (i.e., σ_w in Equation (6)) for King county is much larger than that for New York and Miami-Dade counties, indicating greater weekly variability in the autoregressive rate for King county (see Figure 5). By examining patterns in incidence by day of the week in all counties, we notice that Mondays often report fewer cases than the surrounding days; the daily cases appear to rise from Monday to Friday.

TABLE 2: Posterior means and 95% posterior credible intervals (in brackets) of the parameters (excluding spline coefficients) from the best-fitting model for each county.

Variable	Hennepin	King	New York	Miami-Dade
r in Equation (1)	5.42 (4.43, 6.59)	23.66 (19.09, 29.13)	9.13 (7.41, 11.18)	6.55 (5.44, 7.80)
κ in Equation (3)	0.94 (0.73, 1.00)	0.91 (0.69, 1.00)	0.88 (0.51, 1.00)	0.65 (0.24, 0.98)
q in Equation (3)	0.57 (0.37, 1.03)	7.59 (4.71, 9.90)	0.80 (0.45, 2.00)	1.95 (0.77, 7.02)
σ_w in Equation (6)	-	0.23 (0.17, 0.32)	0.09 (0.04, 0.17)	0.05 (0.01, 0.12)
ζ_0 in Equation (4)	-0.26 (-1.90, 1.45)	3.19 (2.82, 3.45)	0.65 (-1.83, 2.22)	1.22 (-1.83, 3.46)
z_r in Equation (4) (Monday and Friday)	2.67 (0.86, 4.46)	-	-	-
x_r in Equation (5) (Monday)	-	-0.34 (-0.46, -0.23)	-0.26 (-4.0, -0.12)	-0.32 (-0.47, -0.17)

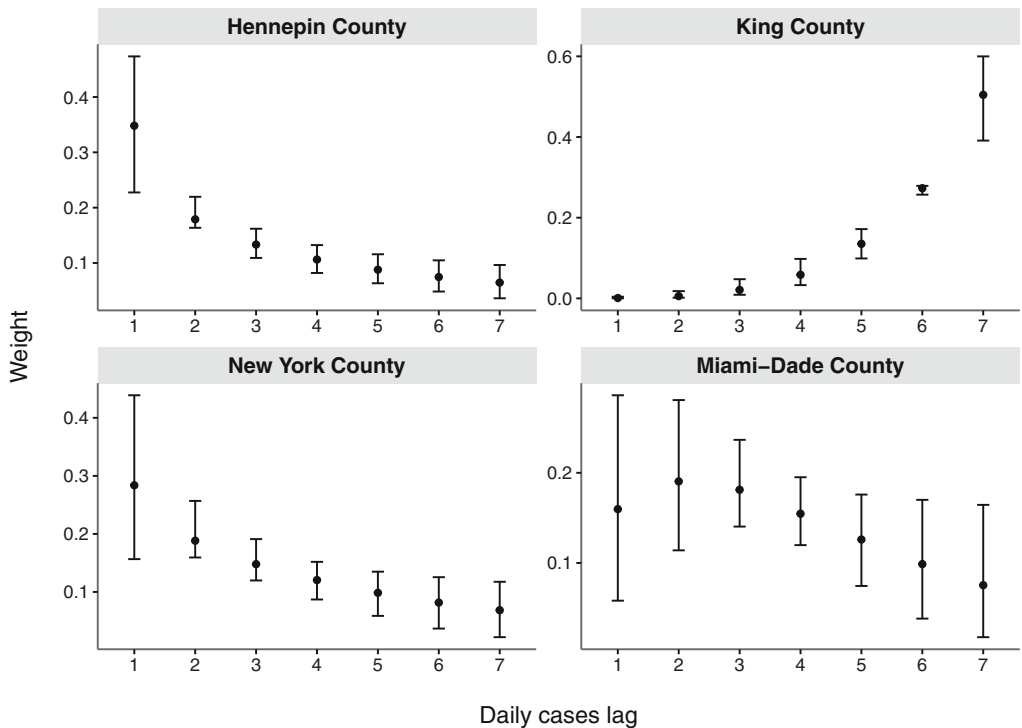


FIGURE 3: Posterior summaries (where solid circles and line segments denote posterior means and 95% credible intervals, respectively) of normalized weights $[\omega_d]$ from the EE models with optimal stay-at-home rate lags.

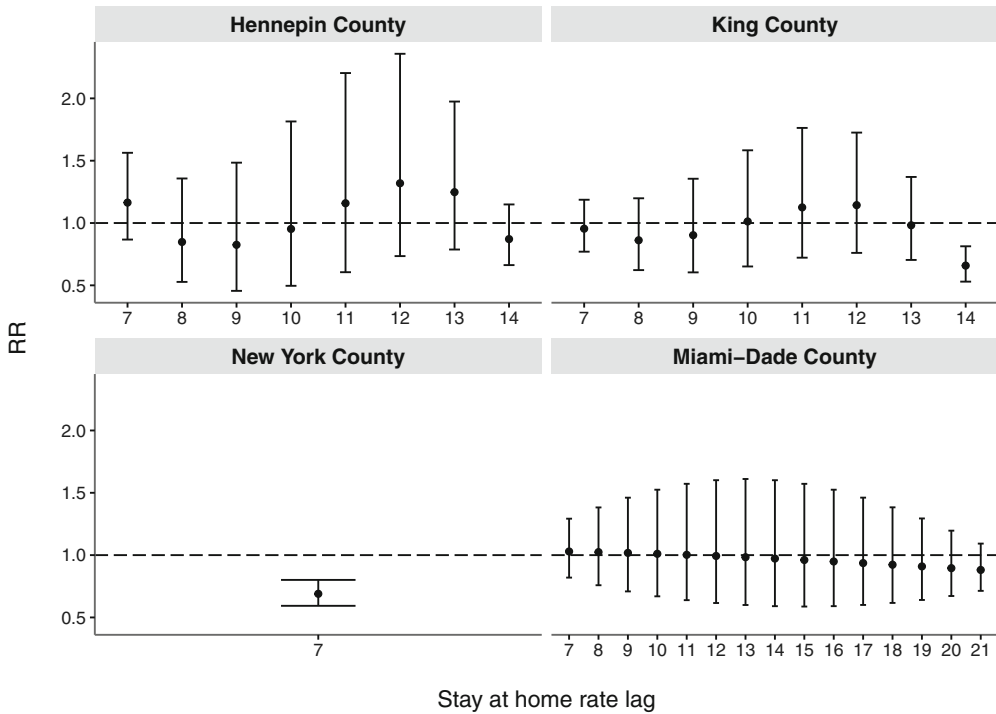


FIGURE 4: Lag-specific effects. Posterior summary of the autoregressive rate ratio (RR) for every 10% increase in the proportion of the population staying at home, as measured by mobility data (i.e., for a 0.1-unit increase in stay-at-home rate, $0.1 \times \beta = 0.1 \times C\eta$ in Equation (5)) with 95% posterior credible intervals.

This day-of-week pattern is consistent with the findings in Slater, Brown & Rosenthal (2021), which observed lower deaths on Sundays and Mondays. However, the day-of-week pattern tends to differ among counties. The estimated Monday and Friday effect z_t in the endemic component and the estimated Monday effects x_t in the epidemic component are summarized in Table 2. In Hennepin county, Monday and Friday have more infections not due to local sources (e.g., imported cases). In King, New York, and Miami-Dade counties, Monday appears to have a lower autoregressive rate (and therefore fewer cases) than the other days.

Figure 3 plots the posterior weights, given by $[\omega_d]$ in Equation (3), for the optimal EE model with shifted NB weights for Hennepin, King, New York, and Miami-Dade counties, including 95% posterior credible intervals for the estimated weights. Hennepin and New York counties display similar monotonically decaying curves, and Miami-Dade county presents a concave curve with a maximum at lag two; larger lags show lower weights, which suggests that the further from the current time, the less the forecasting power of previous daily reported cases is. On the contrary, King county displays a monotonically increasing curve that has the first four estimated weights close to zero, suggesting that daily reported cases for the past 1–4 days do not provide information for forecasting current case counts: this explains the weekly pattern of daily reported cases shown in Figure 6. Note that geometric weights would not be able to reveal an increase in the weights such as those observed in King county. When fitting the model with $p = 14$ and $p = 21$ in King county, we also observed most of the weight being put on the last lag, but with a significant increase in the WAIC. Therefore, this phenomenon of the most weight being estimated on the last lag in King county appears to be due to strong weekly patterns and not due to a need to increase p . The 95% posterior credible intervals show a non-monotonic pattern. For Hennepin, New York, and Miami-Dade counties, the first lag has

the largest posterior credible interval and the narrowest interval appears around the middle of the lags. In contrast, for King county, the largest posterior credible interval appears in the last lag. The serial interval distribution of COVID-19 is thought to have a mode at 3–5 days (Nishiura, Linton & Akhmetzhanov, 2020). Our weighting scheme does allow for this but instead estimated a constant decaying, concave or increasing pattern in the weights. This could be explained by the fact that our data (reported cases) are the result of a complex process that not only involves new infections but also variable diagnosis delays and changing testing eligibility and reporting procedures.

The lag-specific figure (Figure 4) shows the estimated autoregressive rate ratios for the stay-at-home rate at the optimal lags for the four counties. The overall significant lag effects of stay-at-home rate on daily cases are negative, while specific lag effects tend to be different. This supports the assumption that more travel outside the home increases the risk of COVID-19 infection. The relationship between the stay-at-home rate and the autoregressive rate ϕ_t seems to change with the lag. Figure 4 confirms the delayed effects of population mobility on observed cases: for King county, the 95% credible interval of lag 14 does not contain one, and for New York county, the 95% credible interval of the only lag does not contain one. In the best-fitting models selected by the WAIC, the credible interval of every lag contains one in Hennepin and Miami-Dade counties. Figure 4 displays the posterior distributions of lag-specific effects, with many of the lagged effects being not significant. A possible explanation is that we condition on past values of the case counts, which may already reflect much of the information about the stay-at-home rate.

Figure 5 shows estimated and 7-day-ahead predicted autoregressive rates (i.e., ϕ_t in Equation (5)) over time for the optimal EE models. The average ϕ_t is approximately 1.03 (range: 0.68–1.48)

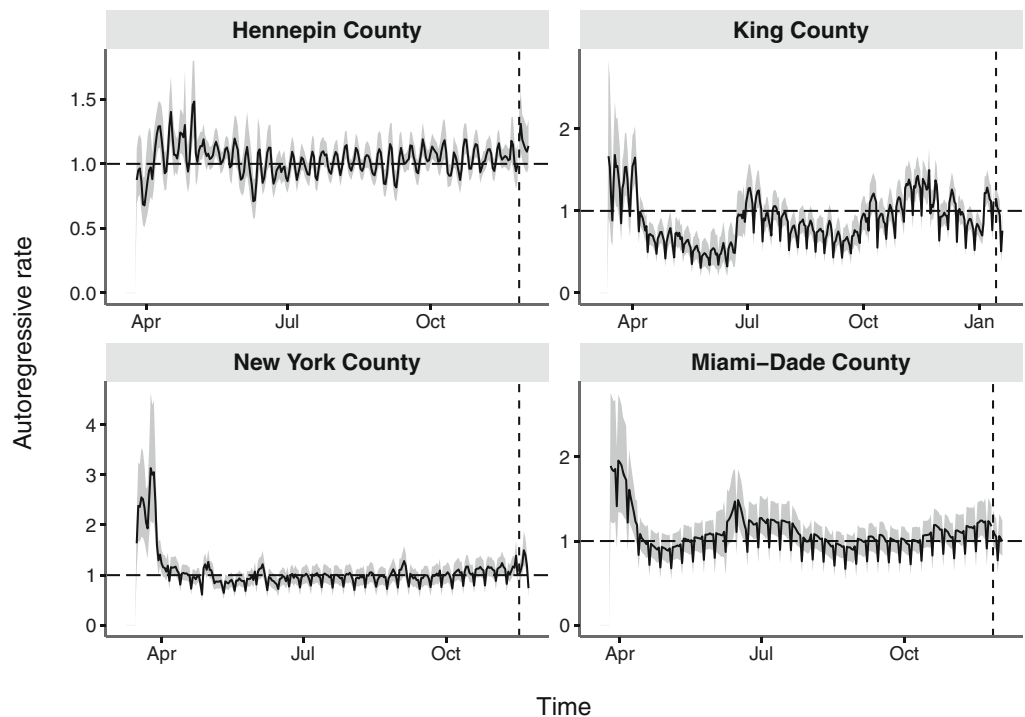


FIGURE 5: The autoregressive rate ϕ_t in Equation (5) over time with 95% posterior credible intervals from the EE models with optimal stay-at-home rate lags. The portion of the series to the left of the vertical dashed line shows fitted autoregressive rate values, while the series to the right of the line plots predicted values.

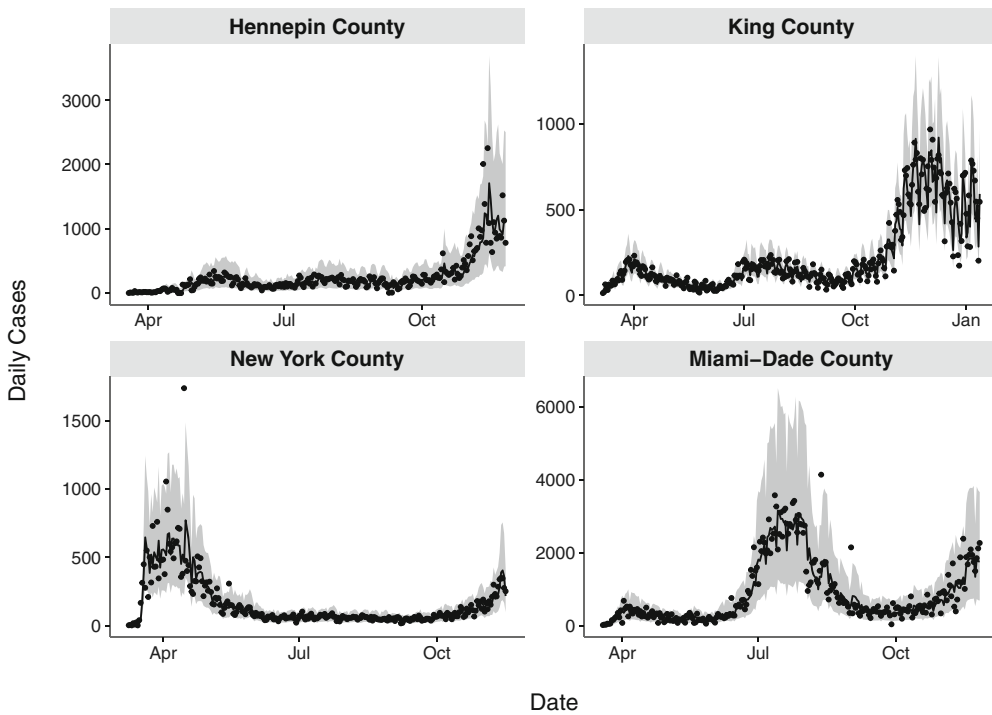


FIGURE 6: Fitted values (solid lines) with 95% posterior credible intervals (shaded areas) together with observed values (solid circles) from the EE models with optimal stay-at-home rate lags.

for Hennepin county, 0.88 (range: 0.31–1.68) for King county, 1.06 (range: 0.61–3.13) for New York county, and 1.09 (range: 0.67–1.95) for Miami-Dade county. Though on different scales, the two time series of autoregressive rates for King and New York display a similar trend that starts with a large value in March, and decreases in April. The largest ϕ_t appears around mid-March, the time when daily reported new cases increased steeply at the very beginning of COVID-19's emergence (see Figure 6). The autoregressive rate time series for Hennepin county started with a low value below one, but then increased and became relatively stable before finally going above one after October. This further explains the daily cases pattern of Hennepin in Figure 6: daily case counts reached a small peak around May, then plateaued before steeply increasing after October. The small ϕ_t value in late April to early May could be due to statewide community mitigation activities, including the closure of nonessential businesses and the issuance of orders encouraging (if not mandating) residents to stay at home. However, changes in the autoregressive rate could also be due to changes in testing eligibility or interest as well as changes in reporting procedures.

Figure 6 examines the time series of the fitted values together with the observed counts. The fitted values are constructed by drawing from $p(y_t | \mathbf{y}_{(t-1):(t-p)}, \boldsymbol{\theta}^{[m]}, r^{[m]})$ for $m = M + 1, \dots, Q$ and $t = p + 1, \dots, T$ and, therefore, represent draws from the posterior distribution of a new case count generated by the same parameters and past counts that generated y_t . Clearly, the fitted models capture well the structure of the different time series. Figure 7 displays the predicted values together with the actual observed counts for the following 7-day period, along with 95% credible intervals using the distributed observed daily cases. The predicted values represent draws from the posterior predictive distribution of future COVID-19 case counts, $p(y_{t+k} | \mathbf{y})$, and were generated using Algorithm 1. The figures suggest that the EE model captures the daily trend well. The good fit is even more pronounced for the 1-week ahead forecasts.

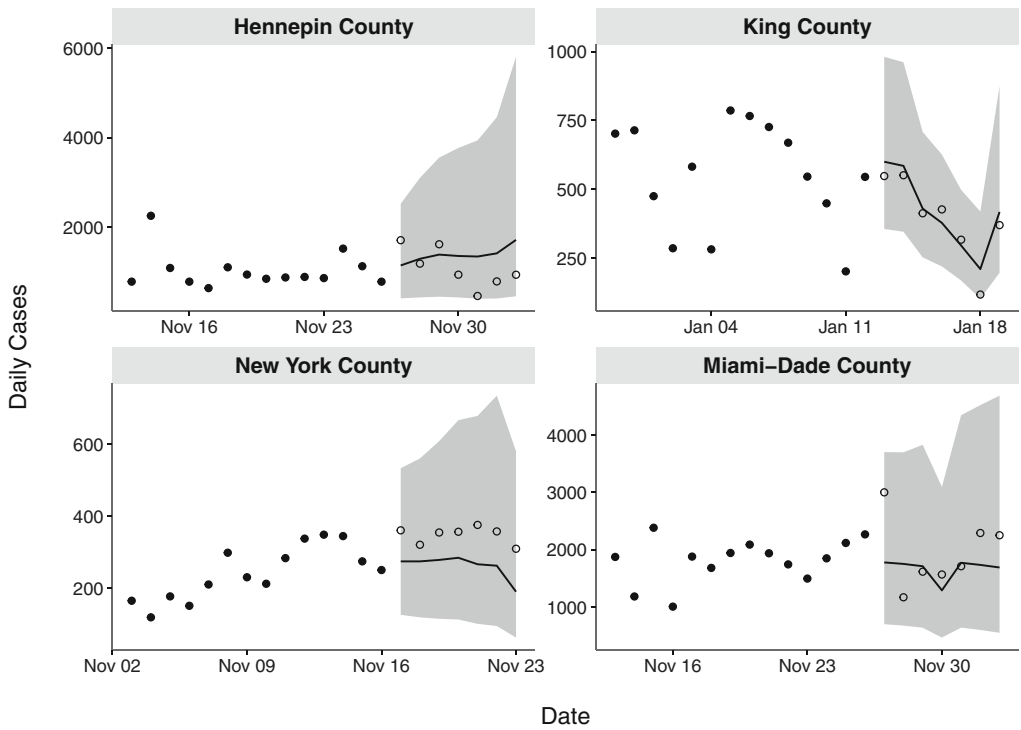


FIGURE 7: Predicted values (solid lines) with 95% posterior credible intervals (shaded areas) from the EE models with optimal stay-at-home rate lags. Open circles denote the observed cases left out from the inference procedure, whereas solid circles represent the last few observations used for model fitting.

To further assess model adequacy, we plot the autocorrelation function (ACF) of the posterior mean of the Pearson residuals $(y_t - u_t)/\sqrt{u_t(1 + r^{-1})}$ (Bracher & Held, 2022), in each county. There are no significant autocorrelations in any county except for New York at around 3 weeks. However, fitting the model with $p = 21$ in New York does not yield a significantly better model fit.

We also investigate the sensitivity of the model fitting to the choice of priors for r , σ_w , and q in King county. Uniform priors for r and σ_w with large upper bounds do not change the results noticeably. A gamma prior for q leads to a noticeably faster increase in the weights in King county. However, the gamma prior is very informative for the weights, as discussed in Section 2.4.

4. DISCUSSION

We extended EE time series models to include a distributed-lag model in the specification of the autoregressive rate in order to incorporate lagged effects of mobility data into the forecasting of daily reported COVID-19 cases. This method is based on sound epidemiological theory about the effect of mobility on the generation of secondary cases: less time spent at home should increase the number of secondary cases produced by the current infectious population. This effect should be lagged due to delays between infection and diagnosis. As the stay-at-home rate might not be able to explain all of the variability in the autoregressive rate over time, we also included a first-order weekly random walk and/or day-of-the-week effects. Additionally, we introduced a novel shifted negative binomial weighting scheme for past counts. Shifted negative binomial weights are more flexible than the weights previously proposed in

Bracher & Held (2022) and contain the shifted Poisson and geometric weights as special cases. Our use of Bayesian inference and prediction, which has the advantage of incorporating all uncertainty regarding unknown parameters into the forecasts, is another significant contribution of this work. The incorporation of the DLM led to significantly better fits in Hennepin, King, New York, and Miami-Dade counties, according to the WAIC, with stay-at-home rate at lags of, respectively, 1 or 2–3 weeks being most predictive of current incidence. The lag was particularly large in Miami-Dade county at 21 days while p was only seven. This could be due to undetected cases of COVID-19 that contribute significantly to the spread of the disease. However, in all counties, many of the lagged effects were not significant. Also, there was no clear evidence regarding the usefulness of the DLM according to the mean log score in King county, although the random walk greatly improves forecasting performance according to the same measure. Therefore, we have mixed evidence overall about whether mobility data improve forecasting performance. A possible explanation is that conditioning on past values of the case counts may already reflect much of the information about the stay-at-home rate, because human behaviour and government orders are often reactive to changes in cases. It is important to note that, as we are mainly concerned with forecasting, we are not focused on determining the causal effect of past mobility on current reported cases in this article, but rather on using mobility to predict future cases. In light of this, one should be cautious in interpreting this result, especially considering the complex effects of reporting bias. The estimated autoregressive rate ϕ_t began to increase again after May, which could be the result of increased trips: we hypothesize that this pattern is due to “COVID-19 fatigue” leading to poorer adherence to infection mitigation strategies. However, the increase could also be due to changes in testing interest or availability. Moreover, the random walk in ϕ_t can account for variability in transmission rates over time from unmeasured sources, such as the potential effects of wearing a face mask or adhering to social distancing guidelines. For example, we would expect attenuated stay-at-home lag-specific effects with raised awareness and adherence to social distancing guidelines.

Accounting for reporting delay and under-reporting in the estimation of the effective reproductive number is the subject of ongoing research. There are several limitations in and challenges with existing methods (Gostic et al., 2020; Bracher & Held, 2021; Quick, Dey & Lin, 2021). For example, the recently proposed method in Quick, Dey & Lin (2021), while reflecting important progress on this problem, requires the reporting delay distribution to be known and time-invariant; requires the use of seroprevalence surveys, which might not be widely available, to account for under-reporting; does not account for overdispersion; assumes that the serial interval distribution is known; and introduces several latent variables for each time point t that would greatly increase the computational complexity of a Bayesian analysis. Therefore, as we are mainly concerned with extending recently proposed EE models (Bracher & Held, 2022) to the Bayesian setting while incorporating a DLM for mobility data, negative binomial weights with uninformative priors, and a random walk, we consider explicitly accounting for reporting delay and under-reporting to be outside the scope of this article. Instead, we attempt to account for reporting effects through covariate design, such as with the DLM in Equation (5), which assumes a lagged effect of mobility data to account for delays between infection, diagnosis, and reporting and with weekday effects. We try not to overinterpret the results of the model. Another modelling decision we made was to use cross validation with proper scoring rules and/or the WAIC to choose between different model specifications. However, a shrinkage prior, such as the Bayesian lasso (Park & Casella, 2008), is an alternative to using proper scoring rules and/or the WAIC when selecting covariates such as when selecting the weekday effects or the stay-at-home rate DLM. There are advantages and disadvantages to using the lasso for covariate selection (Heinze, Wallisch & Dunkler, 2018): investigating these for our particular setting lies outside the scope of this article.

This case study was motivated by the desire to incorporate population mobility into COVID-19 forecasting. This required us to extend the EE model and propose a Bayesian approach to estimate the unknowns of the proposed model. In our application, the proposed models were able to capture different structures, though the mobility data might not be greatly informative in forecasting reported COVID-19 cases for the four time series. However, these extensions and developments of the EE model are not limited to the four selected counties and will undoubtedly be applicable to other regions and useful in wider contexts, such as incorporation of other covariates if not mobility data. One limitation to our approach is that our proposed model is only univariate, while EE models are often fit in a spatio-temporal setting where the disease is allowed to spread between areas through spatial autoregression (Bracher & Held, 2022). A natural extension of our proposed approach could consider multiple time series related to different counties that experienced the first wave of COVID-19 around the same time. Then, a hierarchical Bayesian approach, wherein parameters of the model change across counties but follow a common prior distribution, could be used. This would naturally impose a correlation structure among the different time series and possibly provide posterior predictive credible intervals with shorter ranges. Our model can help guide policymakers by quantifying likely spikes in cases due to increased mobility that is picked up in nearly real time by mobile devices.

ACKNOWLEDGEMENTS

An earlier version of this article was submitted to and won first prize in the special COVID-19 case study competition of the Statistical Society of Canada. The authors are grateful to the Biostatistics section of the Statistical Society of Canada and Dr. Aurelie Labbe for the opportunity to take part in this competition.

Douwes-Schultz is grateful for financial support from IVADO and the Canada First Research Excellence Fund/Apogée (PhD Excellence Scholarship 2021-9070375349). Alexandra M. Schmidt and Erica E. M. Moodie acknowledge funding from Discovery Grants from the Natural Sciences and Engineering Research Council of Canada. Alexandra M. Schmidt also acknowledges funding from Institut de valorisation des données (IVADO) (PRF-2019-6839748021). Erica E. M. Moodie holds a Canada Research Chair (Tier 1) and is supported by a career award from the Fonds de recherche du Québec - Santé.

This research was enabled in part by support provided by Calcul Québec (<https://www.calculquebec.ca>) and Compute Canada (<https://www.computeCanada.ca>).

REFERENCES

- Anastassopoulou, C., Russo, L., Tsakris, A., & Siettos, C. (2020). Data-based analysis, modelling and forecasting of the COVID-19 outbreak. *PLoS One*, 15(3), e0230405.
- Bauer, C. & Wakefield, J. (2018). Stratified space-time infectious disease modelling, with an application to hand, foot and mouth disease in China. *Journal of the Royal Statistical Society Series C*, 67(5), 1379–1398.
- Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., & Ciccozzi, M. (2020). Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data in Brief*, 29, 105340.
- Bracher, J. & Held, L. (2021). A marginal moment matching approach for fitting endemic-epidemic models to underreported disease surveillance counts. *Biometrics*, 77(4), 1202–1214.
- Bracher, J. & Held, L. (2022). Endemic-epidemic models with discrete-time serial interval distributions for infectious disease prediction. *International Journal of Forecasting*, 38(3), 1221–1233.
- Bureau of Transportation Statistics (2020). Trips by distance. <https://data.bts.gov/Research-and-Statistics/Trips-by-Distance/w96p-f2qv>
- CDC (2020a). How COVID-19 spreads. <https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-covid-spreads.html>

- CDC (2020b). Interim clinical guidance for management of patients with confirmed coronavirus disease (COVID-19). <https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html>
- CDC (2020c). About CDC COVID-19 data. <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/about-us-cases-deaths.html>
- CDC (2021). Overview of testing for SARS-CoV-2 (COVID-19). <https://www.cdc.gov/coronavirus/2019-ncov/hcp/testing-overview.html>
- Celani, A. & Giudici, P. (2021). Endemic-epidemic models to understand COVID-19 spatio-temporal evolution. *Spatial Statistics*, 49, 100528.
- Chiang, W. H., Liu, X., & Mohler, G. (2021). Hawkes process modeling of COVID-19 with mobility leading indicators and spatial covariates. *International Journal of Forecasting*, 38(2), 505–520.
- Chwe, H., Quintana, A., Lazer, D., Baum, M., Ognyanova, K., Perlis, R. H., Santillana, M., et al. (2021). The COVID states project #17: COVID-19 test result times. OSF Preprints.
- Cori, A., Ferguson, N. M., Fraser, C., & Cauchemez, S. (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics. *American Journal of Epidemiology*, 178(9), 1505–1512.
- de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Lang, D. T., & Bodik, R. (2017). Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, 26(2), 403–413.
- Dzhanova, Y. (2020). New York state now has more coronavirus cases than any country outside the U.S. CNBC News, Retrieved June 14, 2020.
- Fanelli, D. & Piazza, F. (2020). Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos, Solitons & Fractals*, 134, 109761.
- Gasparrini, A., Armstrong, B., & Kenward, M. G. (2010). Distributed lag non-linear models. *Statistics in Medicine*, 29(21), 2224–2234.
- Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24(6), 997–1016.
- Gneiting, T. & Raftery, A. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359–378.
- Gostic, K. M., McGough, L., Baskerville, E. B., Abbott, S., Joshi, K., Tedijanto, C., Kahn, R. et al. (2020). Practical considerations for measuring the effective reproductive number, R_t . *PLoS Computational Biology*, 16(12), e1008409.
- Grimée, M., Bekker-Nielsen Dunbar, M., Hofmann, F., & Held, L. (2022). Modelling the effect of a border closure between Switzerland and Italy on the spatiotemporal spread of COVID-19 in Switzerland. *Spatial Statistics*, 49, 100552. <https://doi.org/10.1016/j.spasta.2021.100552>
- Heinze, G., Wallisch, C., & Dunkler, D. (2018). Variable selection – A review and recommendations for the practicing statistician. *Biometrical Journal*, 60(3), 431–449.
- Held, L., Höhle, M., & Hofmann, M. (2005). A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Statistical Modelling*, 5(3), 187–199.
- Held, L. & Paul, M. (2012). Modeling seasonality in space-time infectious disease surveillance data. *Biometrical Journal*, 54(6), 824–843.
- Meyer, S. & Held, L. (2014). Power-law models for infectious disease spread. *The Annals of Applied Statistics*, 8(3), 1612–1639.
- Meyer, S. & Held, L. (2017). Incorporating social contact data in spatio-temporal models for infectious disease spread. *Biostatistics*, 18(2), 338–351.
- Neal, R. M. (2003). Slice sampling. *Annals of Statistics*, 31(3), 705–767.
- Nishiura, H., Linton, N. M., & Akhmetzhanov, A. R. (2020). Serial interval of novel coronavirus (COVID-19) infections. *International Journal of Infectious Diseases*, 93, 284–286.
- Park, T. & Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, 103(482), 681–686.
- Paul, M., Held, L., & Toschke, A. M. (2008). Multivariate modelling of infectious disease surveillance data. *Statistics in Medicine*, 27(29), 6250–6267.
- Pawitan, Y. (2001). *In All Likelihood: Statistical Modelling and Inference Using Likelihood*, Oxford University Press, Oxford.

- Perlstein, M. (2020). New Orleans is second only to Seattle in COVID-19 cases per capita. 4WWL, Retrieved March 16, 2020.
- Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, 6(1), 7–11.
- Quick, C., Dey, R., & Lin, X. (2021). Regression models for understanding COVID-19 epidemic dynamics with incomplete data. *Journal of the American Statistical Association*, 116(536), 1561–1577.
- Slater, J. J., Brown, P. E., & Rosenthal, J. S. (2021). Forecasting subnational COVID-19 mortality using a day-of-the-week adjusted Bayesian hierarchical model. *Stat*, 10(1), e328.
- Tibbits, M. M., Groendyke, C., Haran, M., & Liechty, J. C. (2014). Automated factor slice sampling. *Journal of Computational and Graphical Statistics*, 23(2), 543–563.
- Yonar, H., Yonar, A., Tekindal, M. A., & Tekindal, M. (2020). Modeling and forecasting for the number of cases of the COVID-19 pandemic with the curve estimation models, the Box-Jenkins and exponential smoothing methods. *Eurasian Journal of Medicine and Oncology*, 4(2), 160–165.
- Yousaf, M., Zahir, S., Riaz, M., Hussain, S. M., & Shah, K. (2020). Statistical analysis of forecasting COVID-19 for upcoming month in Pakistan. *Chaos, Solitons & Fractals*, 138, 109926.
- Zanobetti, A., Schwartz, J., Samoli, E., Gryparis, A., Touloumi, G., Atkinson, R., Le Tertre, A. et al. (2002). The temporal pattern of mortality responses to air pollution: A multicity assessment of mortality displacement. *Epidemiology*, 13(1), 87–93.
- Zhang, X., Ma, R., & Wang, L. (2020). Predicting turning point, duration and attack rate of COVID-19 outbreaks in major Western countries. *Chaos, Solitons & Fractals*, 135, 109829.
-

Received 27 June 2021

Accepted 23 April 2022