FORECASTING WAVED DAILY COVID-19 DEATH COUNT SERIES WITH A NOVEL COMBINATION OF SEGMENTED POISSON MODEL AND ARIMA MODELS

Journal of Applied Statistics, 2023
Zhang, Xiaolei & Ma Renjun

JOURNAL OF APPLIED STATISTICS
2023, VOL. 50, NOS. 11–12, 2561–2574
https://doi.org/10.1080/02664763.2021.1976119

By Armin Mahdavi 31 July 2025

Introduction

- Article’s key focus was 2020 COVID-19 deaths for USA and Canada
- Proposes to employ two different models, for various reasons.
- Segmented Poisson model, to “capture mean structures of data in waves” and to respect count aspect of data.
- And, ARIMA model, commonly used for infectious diseases
- Proposes combining both model types in two-stage approach

Data Collection, Key Features

- Data will need to consider key policy and govt intervention dates

Table 1. Segmentation points of Canada and the USA.

Country	Governments intervention	Mass demonstrations	Winter outbreak rebound
Canada	March 16		September 7
USA	March 16	May 25	September 7

Data Collection, Key Features

- Cyclical seasonal nature of data for the year -> time series
- Unequal wavelengths of the data, time-varying variances
- Typical regression modeling for this global mean structure unlikely to fit appropriately.

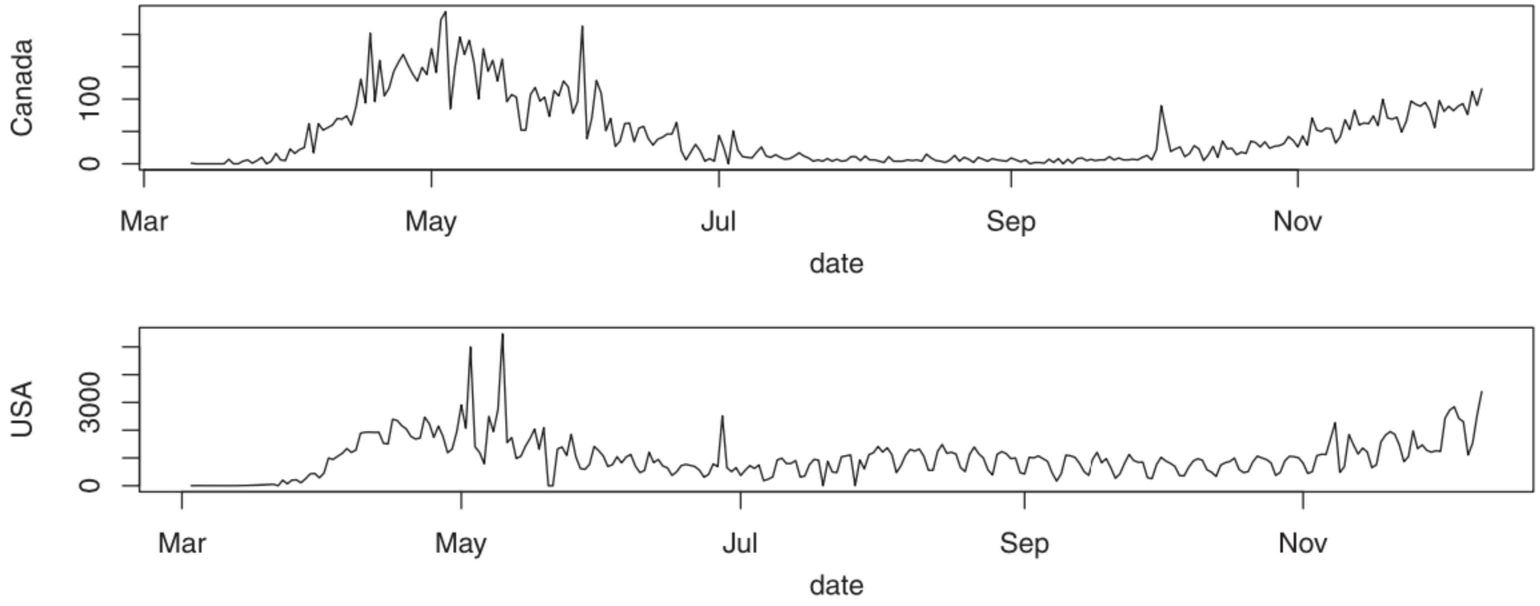

Figure 1. Daily COVID-19 deaths in Canada and the USA.

Data Collection, Key Features

- Very skewed data
- Some zero counts + High proportion of low counts
- Methods that assume approximate normality would not be appropriate

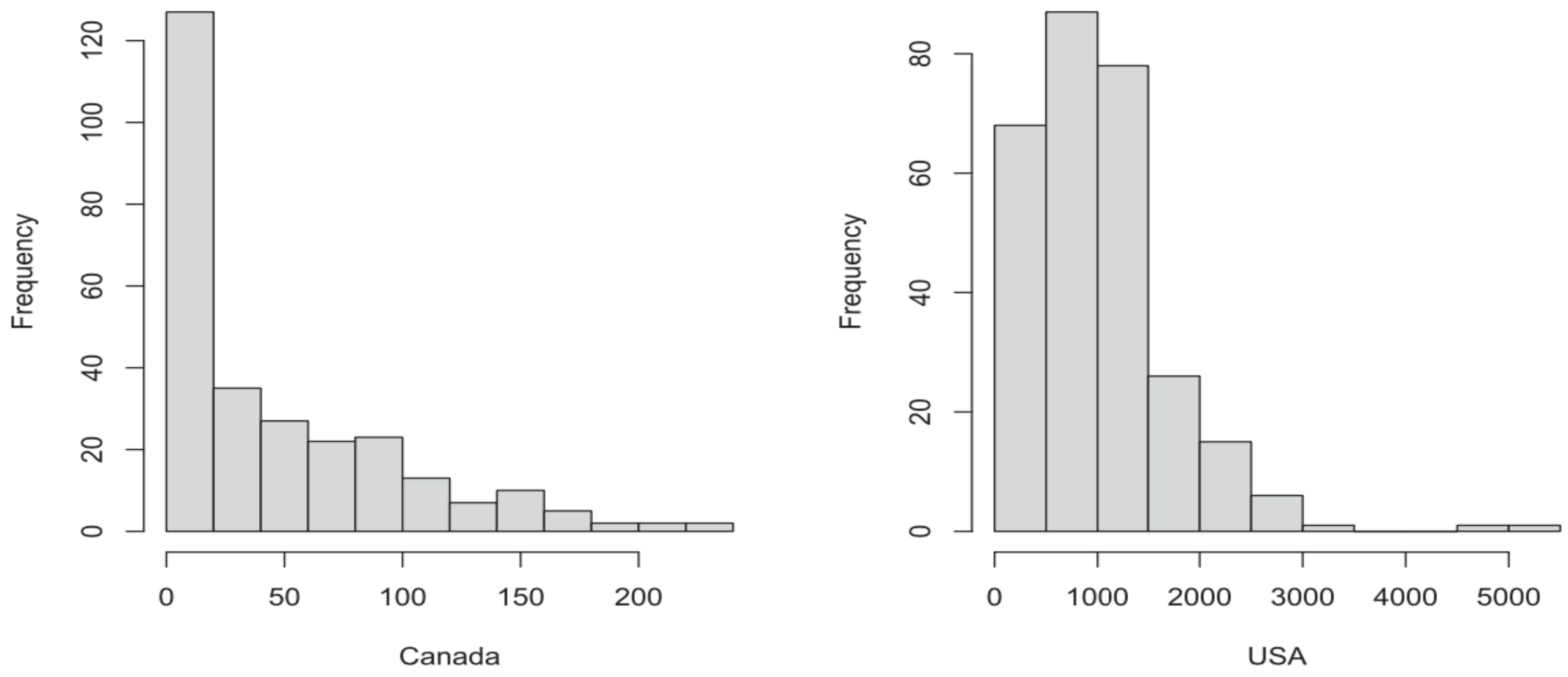

Figure 2. Histogram of daily COVID-19 deaths in Canada and the USA.

Statistical Methods

- How to resolve the three key features?
- Segment the data based on those important policy dates, then focus on each chunk of data.
- Apply Poisson regression to detrend the data by removing overall mean structure/trend
  - Captures piecewise trends in data
- Detrended data’s raw residuals can then be used and modeled by ARIMA

Segmented Poisson Model

the segmented Poisson model is defined as a single Poisson model below fitted with
.ata

$$Y_t \sim Poisson(\mu_t), \tag{1}$$

- Tau1 = 1 week or 7 days after starting data of major government intervention.

but with piecewise mean structure given by

$$\mu_t = \begin{cases} \alpha_1 t^{\beta_1} e^{-\gamma_1 t} & \text{for } t = 1, \ldots, \tau_1 \\ \alpha_j t^{\beta_j} e^{-\gamma_j t} & \text{for } t = \tau_{j-1} + 1, \ldots, \tau_j \; j = 2, \ldots, J \\ \alpha_{J+1} t^{\beta_{J+1}} e^{-\gamma_{J+1} t} & \text{for } t = \tau_J + 1, \ldots, T \end{cases} \tag{2}$$

where parameters α_j, β_j and γ_j are regression parameters over the jth segment,

ARIMA Model

- ARIMA modeling assumes stationarity, key assumption
  - (mean and variance shouldn’t change over time)

goal is to model and forecast the daily COVID-19 death count series data; therefore, we go further to account for correlation left in the raw residuals defined below.

$$r_t = y_t - \widehat{y_t} \tag{4}$$

where y_t and $\widehat{y_t}$ represent the observed and fitted daily death counts at time t, respectively. As raw residuals r_t form a single time series and are no longer count data, we may employ the ARIMA models to tackle the complex correlation structure of this time series at the

ARIMA Model

is given by:

$$
\begin{aligned}
X_t = {} & \overbrace{\sum_{i=1}^{p} \varphi_i X_{t-i}}^{\text{Regular AR}} + \overbrace{\sum_{j=1}^{P} \Phi_j X_{t-js}}^{\text{Seasonal AR}} - \sum_{i=1}^{p} \sum_{j=1}^{P} \varphi_i \Phi_j X_{t-i-js} \\
& + \epsilon_t + \underbrace{\sum_{k=1}^{q} \theta_k \epsilon_{t-k}}_{\text{Regular MA}} + \overbrace{\sum_{l=1}^{Q} \Theta_l \epsilon_{t-ls}}^{\text{Seasonal MA}} - \sum_{k=1}^{q} \sum_{l=1}^{Q} \theta_k \Theta_l \epsilon_{t-k-ls}, \qquad (5)
\end{aligned}
$$

In this expression, parameters P, D and Q are similar to p, d and q defined in an ARIMA(p, d, q) model. Here P indicates the number of seasonal autoregressive terms, D indicates the number of seasonal differences, Q indicates the number of seasonal moving average terms. Similarly, $D = 1$ refers to seasonal difference series $\nabla_s r_t = r_t - r_{t-s}$. The key assumption of the ARMA models is stationarity. A less explicit assumption of the

Limitations and Considerations

- Must manually segment data based on segmentation points in order to set it up for modeling.
  - Policy dates, like in Table 1
- Requires stationarity and approximate symmetry for ARIMA
- Segmented Poisson modeling needs a specific function form, the power-exponential trend
- Won’t adapt to new structural changes, govt interventions, new COVID waves

Conclusion

- Improve segmentation from manual to automation
- This two-stage approach can be adapted to ”model and forecast development of other infectious diseases as well.”
- Their novel approach captured wave-based COVID-19 daily death data with good forecasting power, by their own account