HANDLING AMBIGUITY OF INTERVAL CENSORED EVENTS

Jiahao Tian

A Thesis submitted to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

School of Data Science

University of Virginia

January 2024

Jeffery Blume, Chair
Michael D. Porter
Jonathan Kropko
Bill Basener
Negin Alemazkoor

Contents

List of Figures

List of Tables

Chapter 1

Introduction

It is of interest to statistical make inference of a certain events given data in the field of statistics. In the context of survival analysis, for example, researchers are interested in modeling occurrence and timing of events of interest. However, the event of interest may not be observed exactly, which compllicates the analysis of such data. This lack of complete data is called censoring, which refers to the case that where the specific timing of events of interest remains unknown, and we only have information about their occurrence within a specified interval.

1.1 Types of Censoring

An event is referred to as "left-censored" when it happens before a predefined time, known as the censoring time. This form of censoring frequently arises when a cohort of patients is enlisted to partake in a clinical trial, and the event of interest has already transpired for certain individuals prior to the initiation of the study.

In certain scenarios, the event times, denoted as T, are constrained to exist solely within a specified interval, denoted as $[L, R]$, where $L \leq T \leq R$. This occurrence typically arises when patients are subjected to periodic follow-ups, and it is established that the event has not been witnessed at time L, but it has indeed occurred by time R. In this context, such data is characterized as "interval censored."

Right-censored data pertains to events for which their actual occurrence remains unobserved up to a specified censoring time. This situation frequently arises when certain patients have not experienced the events by the conclusion of the study or due to instances of being lost to follow-up.

1.2 Wide Spread of Censoring

While we have primarily introduced and elucidated the concept of censoring from a biostatistical standpoint, it's important to note that censoring finds applicability in various domains beyond the medical context. In credit scoring, for instance, censoring is encountered when modeling time-to-default directly, as evidenced in studies by Narain, 2004; Dirick, Claeskens, and Baesens, 2017. Similarly, in the field of engineering, the accelerated failure model proves invaluable for modeling the failure time of machinery, as demonstrated in Wei, 1992; Newby, 1988. Even in the realm of politics, interval censored data can arise, particularly when pollsters exclusively provide aggregated statistics for conducted polls, as discussed by Tian and Porter, 2022a.

Despite its prevalence, the analysis of censored data often does not receive the same level of attention as it does in the field of biostatistics. Notably, there is a scarcity of research on crucial topics related to censored data in various domains. For instance, there is limited exploration on methods for detecting change points when dealing with censored data, and forecasting in the presence of censoring remains an underdeveloped area. These gaps in research highlight the need for further investigation and innovation in the analysis of censored data across a range of disciplines.

In this thesis, we delve into critical subjects that encompass change point detection,

intensity estimation, and forecasting within the context of political polling and crime data, all while contending with the challenges posed by interval-censored data.

1.3 Handling Censoring

Several straightforward methods have been devised for transforming interval-censored data, rendering them amenable to the same analytical techniques developed for exact data. Research has underscored the potential for bias in results when censoring is not appropriately managed, necessitating the utilization of suitable techniques and tools, as highlighted in studies by Turkson, Ayiah-Mensah, and Nimoh, 2021; Lindsey and Ryan, 1999.

One tempting approach is the straightforward imputation method, such as replacing the interval with its midpoint, which is favored for its simplicity, as noted by Lindsey and Ryan, 1999. Another noteworthy method is the Aoristic approach, as introduced by Ratcliffe, 2000 within the realm of crime analysis. This method treats each event individually and allocates a partial count to the corresponding bins covered by each interval. In this section, we would like to do a comparison to demonstrate how the likelihood-based approach is better at uncovering the true intensity curve compared with mid-point imputation and the aoristic method. Synthetic data has been generated to ensure a fair comparison. This dataset comprises 2000 observations, with a predetermined proportion of censored observations set at 0.8, mirroring the proportions observed in the real data we analyzed. Additionally, the average length of the intervals has been established in accordance with this proportion and is assumed to adhere to an exponential distribution with a mean value of 8. The experiment has been conducted a total of 5,000 times for each of the methods under consideration, en-

abling a robust and comprehensive assessment. In Figure 1.1, we present a summary

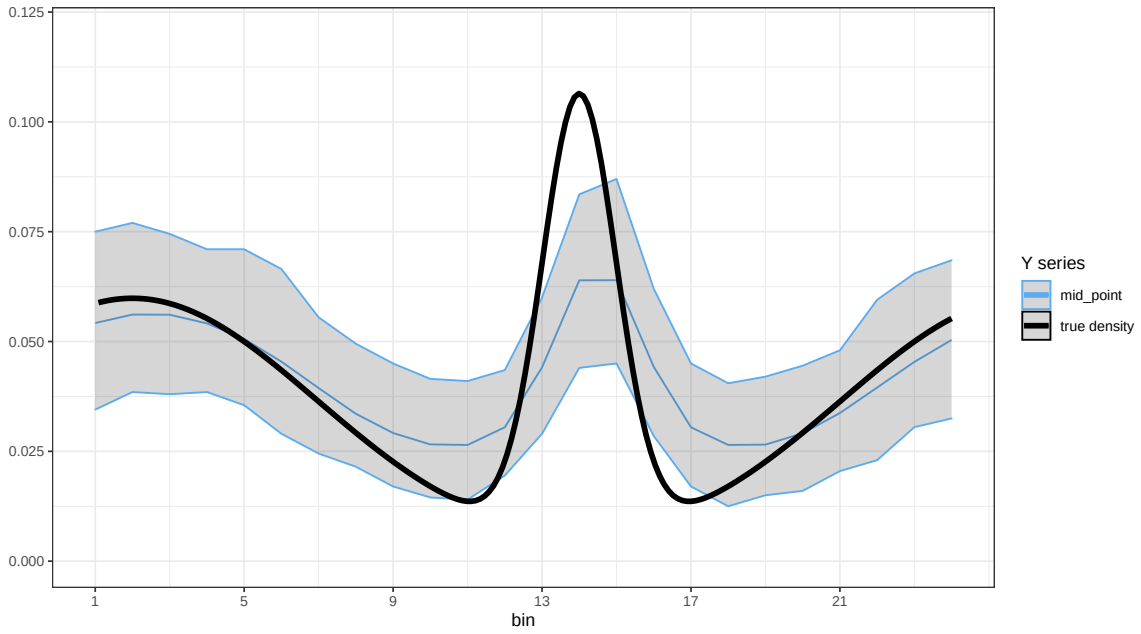

Figure 1.1: Estimated density with midpoint imputation

of the performance when applying mid-point imputation to analyze the aforementioned dataset. The solid black line represents the true density that every method aims to unveil. The blue line in the center denotes the average estimated density across 5,000 simulations. The remaining two lines correspond to the 2.5th percentile and 97.5th percentile of the estimated densities, providing a range that encapsulates the variation observed in the simulations. A comparison between the middle blue line, which represents the average estimated density, and the solid black line depicting the underlying density, leads us to the conclusion that mid-point imputation falls short in capturing the peaks and valleys within the density distribution. In Figure 1.2, we present the performance of applying the Aoristic method, as introduced earlier, to analyze the same synthetic data. As expected, the Aoristic method also falls short of fully capturing the true underlying density. However, the plot does indicate that the variation in the estimates is comparatively smaller when compared to the mid-point

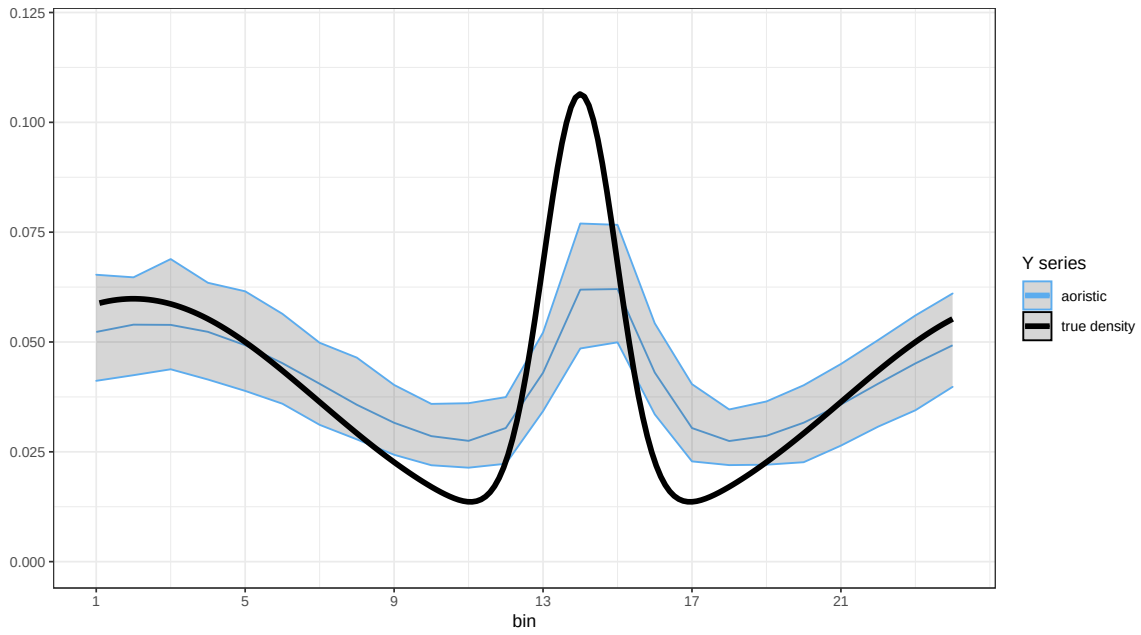

Figure 1.2: Estimated density with aoristic method

imputation method. Finally, we present the results here of applying a likelihood-based approach to analyze the same data. Specifically, the number of occurrences of the events is assumed to follow a Poisson distribution. We also derived the maximum likelihood estimate based on Expectation-Maximization Dempster, Laird, and Rubin, 1977. Figure 1.3 illustrates that the likelihood-based approach demonstrates remarkable capability in uncovering the true underlying density, particularly in capturing the peaks and valleys within the density curve. Additionally, the variation in estimates, as denoted by the grey-shaded area, is comparable to that observed with the Aoristic method, which is a significantly less complex model. Therefore, the experiment presented here underscores that the likelihood-based approach yields superior results compared to other methods considered, as it accurately uncovers the true density while achieving comparable variation.

Figure 1.4 illustrates the squared bias obtained through various approaches to han-

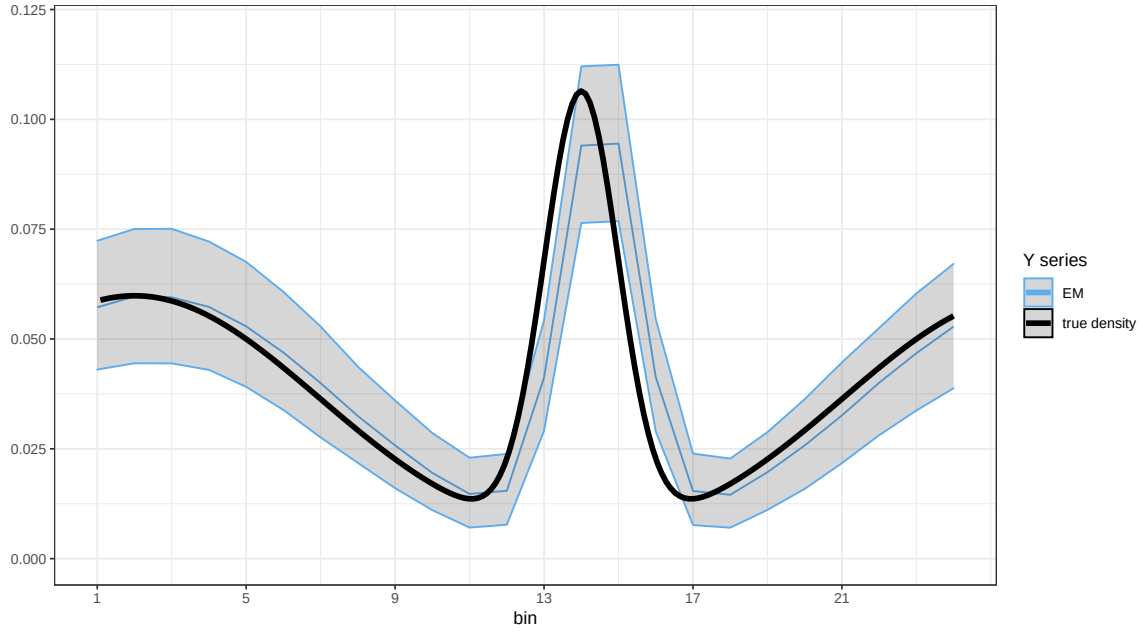

Figure 1.3: Estimated density with EM approach

dling censored data. A notable spike in squared bias occurs when the underlying intensity is at its peak or valleys. The reason why both the aoristic and midpoint methods fail to accurately capture the intensity is that they produce an overly smooth curve, leading to an increase in bias at intensity spikes.

The simulation study mentioned above illustrates that the utilization of simple imputation methods may result in significant bias, while the likelihood-based approach can yield better inference by considering the intensity covered by each interval. The subsequent sections of the dissertation will delve into important applications of censored data modeling, encompassing change point detection, intensity estimation, and forecasting. These applications highlight the significance of handling censored data appropriately to obtain sensible and accurate results.

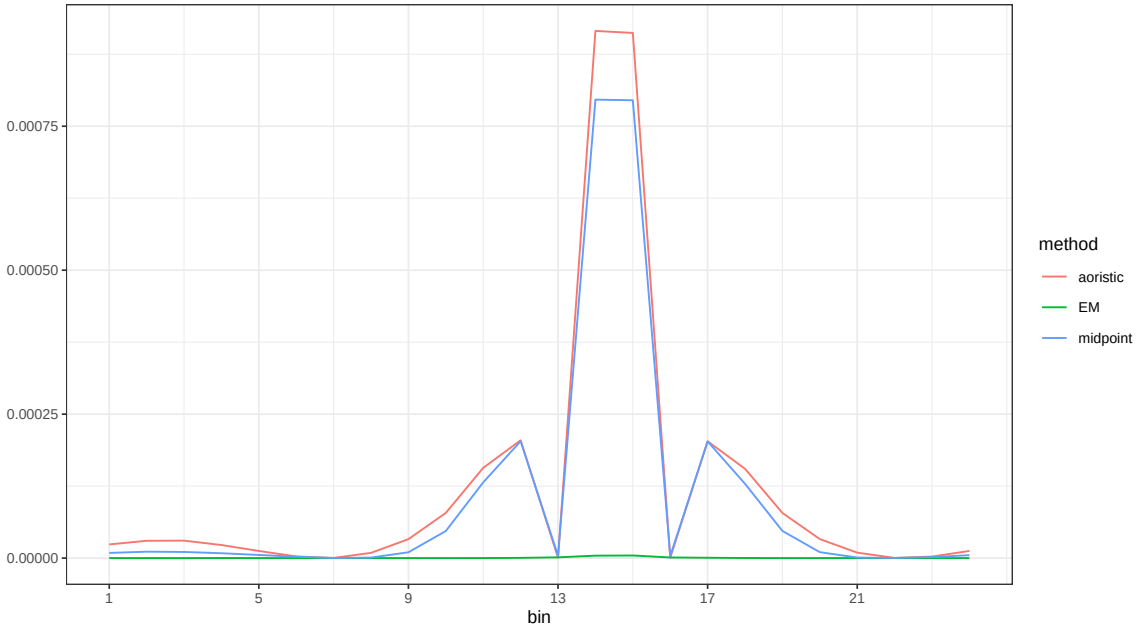

Figure 1.4: Squared bias of different approaches considered

Chapter 2

Change point detection

Change point detection (CPD) involves identifying abrupt changes in data, indicating a shift in a property of the time series. Depending on the nature of the methods, CPD can be achieved through likelihood ratio-based approaches, probabilistic methods, clustering methods, and so on. The detection of change points with probabilistic methods encompasses both frequentist and Bayesian approaches. Gaussian processes are often assumed and employed for change point detection by comparing data points with a reference distribution Chandola and Vatsavai, 2010. Bayesian Change Point Detection (BCPD) focuses on identifying abrupt changes using Bayesian-based methods by estimating the posterior distribution of the run length, which is the time elapsed since the last change point Tan et al., 2015; Ruggieri and Antonellis, 2016. These methods exhibit two distinct characteristics: 1) their primary goal is to find the point at which the underlying mechanism changes, and 2) they assume data is precisely observed. However, data such as political polling is often reported in an aggregated form, rendering methods suitable for exact data inapplicable. In this project, we introduce a method that leverages the combination of the joinpoint model with the Expectation-Maximization (E-M) framework to detect change points in the presence of censored data. Furthermore, we propose the use of Bayesian Model Averaging (BMA) to average the evidence for the change points and obtain the probability for each change point.

Changing Presidential Approval: Detecting and Understanding Change Points in Interval Censored Polling Data

AbstractUnderstanding how a society views certain policies, politicians, and events can help shape public policy, legislation, and even a political candidate's campaign. This paper focuses on using aggregated, or interval censored, polling data to estimate the times when the public opinion shifts on the US president's job approval. The approval rate is modelled as a Poisson segmented (joinpoint) regression with the EM algorithm used to estimate the model parameters. Inference on the change points is carried out using BIC based model averaging. This approach can capture the uncertainty in both the number and location of change points. The model is applied to president Trump's job approval rating during 2020. Three primary change points are discovered and related to significant events and statements.

Keywords:Bayesian Model Averaging, Change Point Detection, EM algorithm, Aggregated Data, Interval Censoring, Joinpoint Regression, Segmented Regression, Polling, Presidential Approval

2.1 Introduction

Polling data is an important source of information for evaluating how a society views issues, people, and policies (Weisberg, Krosnick, and Bowen, 1996). It is especially useful for understanding how politicians and political candidates are viewed by the voting public (Jonge, Langer, and Sinozich, 2018; Walther, 2015; Murray, Riley, and Scime, 2009; Prosser et al., 2020; Mostafavi et al., 2021) and can be used to predict election results (Lock and Gelman, 2010). Prior research has identified that the poll standing of a politician can change significantly after important events like party

conventions, political debates, policy decisions, and speeches (Campbell, Cherry, and Wink, 1992; McAvoy, 2006; Benoit, Hansen, and Verser, 2003; Willer, 2004; Druckman and Holmes, 2004).

This paper examines President Trump's job approval rating. Instead of testing a set of pre-determined events for their influence on the poll results, we focus on identifying the time points when public opinion shifts on Trump's approval and then relating those change points to potential explanatory events. The task of identifying the change points is complicated by the aggregated, or interval censored, nature of the polling data; a poll often spans multiple days but only the aggregated results are available.

We posit that change in approval will not be abrupt, but rather develop over time as news coverage and public opinion builds following decisive events. This leads us to consider identifying change points in the trend of approval rating. The hypothesis is that certain events will trigger a change in opinion, either positive or negative, that will lead to growing or shrinking job approval. Because we expect these changes to be gradual, we consider joinpoint models; these are change detection models that represent the underlying mean of a process as a set of piece-wise linear segments that are connected at the change, or join, points (Lerman, 1980; Hinkley, 1971; Kim et al., 2000). Joinpoint regression models are popular approaches for capturing trend changes in cancer, mortality, and epidemiological data (Martinez-Beneito, García-Donato, Salmerón, et al., 2011; Qiu et al., 2009; Dehkordi, Tazhibi, and Babazade, 2014; Wong et al., 2018; Puzo, Qin, and Mehlum, 2016). Sometimes called segmented regression, joinpoint models partition the data into segments and model the data in each segment with a (usually) simple regression function. A key aspect of joinpoint models is that they enforce the estimates to be equal at the joinpoints (i.e., no abrupt

changes are assumed). Mathematically, a joinpoint model can be expressed:

$$f(t;\beta,\tau) = \begin{cases} f_1(t|\beta_1) & for \quad t \leq \tau_1 \\ f_2(t|\beta_2) & for \quad \tau_1 \leq t \leq \tau_2 \\ ... & \\ f_k(t|\beta_k) & for \quad t \geq \tau_k \end{cases}$$

where $\tau_1 ... \tau_k$ are the unknown change points and each f_i is a known parametric function of the parameter(s) β_i. Lerman (1980) described a grid search to estimate both the change points and corresponding parameters. Hinkley (1971) proposed an estimation approach in the case of a single joinpoint. Ghosh et al. (2009) took a Bayesian approach to allow prior information about the number and position of change points.**(Siddiqa, Ali, and Shah, 2021) used an integer programming approach to detect common change points in multivariate time series with censored Gaussian observations. Assareh and Mengersen (2012) used a Bayesian hierarchical model to detect single change points in right censored survival data. Wang, Wang, and Song (2019) tackled the problem of single change point detection in the hazard function for interval censored survival data.** Polls are often conducted over a duration of several days and only the aggregated data are made available. This can also be viewed as interval censoring; the exact time of poll responses are not known, but only that they occurred sometime within the survey period. This introduces additional difficulties in accurately estimating change points as polls that span the change points will only provide information on the average responses weighted by how much of the poll was conducted before and after the change point. A common approach in estimating models using censored data is to formulate a two-stage procedure that iteratively estimates the

time of each censored event based on the current model parameters and then updates the model parameters using the estimated event times (e.g., the EM algorithm) (Yu et al., 2009). We take this approach and derive an EM algorithm for maximizing the likelihood of a joinpoint Poisson regression model with interval censored data.

Instead of model selection, we use model averaging to assess the evidence for change points occurring at certain times. This approach fits separate models for a large collection of change points and uses the Bayesian Information Criterion (BIC) to estimate the posterior probability that changes occurred at those times.

The contribution of our work includes:

- an EM approach for estimating the model parameters in a joinpoint Poisson regression using aggregated polling data.
- a Bayesian model averaging approach to estimate the number of change points and their locations.
- a list of change points in President Trump's 2020 approval rating and their alignment with potential explanatory events.

2.2 Methodology

We consider observing p polls that were conducted over days $1, 2, \ldots, T$. Let $D_i = (L_i, R_i, N_i, Y_i)$ be the observed information from poll i where $1 \leq L_i \leq T$ is the start time, $R_i \geq L_i$ is the end time, N_i is the number of respondents, and Y_i is the number of those respondents that support the outcome of interest. Note that we only observe the aggregate counts, (N_i, Y_i), but not the daily responses (for polls that span more

than one day).

Our goal is to estimate the probability that a randomly selected respondent supports the president on days $t \in \{1, 2, \ldots, T\}$. Because we expect trend changes in response to significant events we model the support probability on day t as

$$\alpha_t = \exp\left(\beta_0 + \beta_1 t + \sum_{s \in \tau} \beta_s (t - s)_+\right) \tag{2.1}$$

where $\tau \subset \{2, 3, \ldots, T-1\}$ are the unknown change points, β are the changes in trend associated with the change points, and $(t-s)_+ = \max(0, t-s)$. Thus, the slope changes by β_s if there is a change point at time s. **We formulate the model as** $\log \alpha = X\beta$ **where the design matrix** X **has a column for the intercept, slope, and change points. For example, the design matrix for a two change point model** ($\tau = \{4, 10\}$) **is** $X = [X_0, X_1, X_4, X_{10}]$ **where** $X_0 = [1, 1, \ldots, 1]^T$**,** $X_1 = [1, 2, \ldots, T]^T$**,** $X_4 = [0, 0, 0, 0, 1, 2, \ldots T-4]^T$**, and** $X_{10} = [\underbrace{0, \ldots, 0}_{10}, 1, 2, \ldots T-10]^T$**.**

We model the number of supporting respondents for poll i as a Poisson random variable with intensity

$$\lambda_i = \sum_{t=1}^{T} n_{it} \alpha_t$$

where $n_{it} = N_i/(R_i - L_i + 1)\mathbb{1}(L_i \leq t \leq R_i)$ is the assumed number of respondents for poll i on day t. This specifies that the number of respondents is equally distributed over the duration of the poll and zero for the other days. The (observed) log-likelihood

is

$$\begin{aligned}\log L(\beta) &= \sum_{i=1}^{p} Y_i \log \lambda_i - \lambda_i \\ &= \sum_{i=1}^{p} Y_i \log \left(\sum_{t=1}^{T} n_{it}\alpha_t \right) - \sum_{t=1}^{T} n_{it}\alpha_t \\ &= \sum_{i=1}^{p} Y_i \log \left(\sum_{t=1}^{T} n_{it} e^{X_i^T \beta} \right) - \sum_{t=1}^{T} n_{it} e^{X_i^T \beta}\end{aligned}$$

where α_t is a function of the model parameters β as given in (2.1). This expression is not easy to optimize due to the sum inside the log. An alternative to direct optimization is the Expectation-Maximization (EM) algorithm (Dempster, Laird, and Rubin, 1977), an iterative algorithm to find the maximum likelihood estimates in the presence of unobserved data (latent variables). Let the latent variables be $\{Z_{it}\}$ which represent the unobserved number of supporters from poll i who responded on day t.

The EM algorithm approach is an iterative two-step procedure that first calculates the expected value of the complete log-likelihood with respect to the latent variables given a current estimate of β **(E-step)** and next updates β to maximize the expected log-likelihood **(M-step)**.

The expected (complete) log-likelihood is

$$\begin{aligned}E[\log L(Z, \beta) \mid \beta] &= \sum_{i=1}^{p} \sum_{t=1}^{T} \phi_{it}(X\beta + \log n_{it}) - n_{it} e^{X\beta} \\ &= \sum_{i=1}^{p} \sum_{t=1}^{T} \phi_{it} \log \alpha_t n_{it} - \alpha_t n_{it}\end{aligned} \tag{2.2}$$

where the conditional expectation, $\phi_{it} = E[Z_{it} \mid Y_i, \hat{\alpha}]$ is written

$$\phi_{it} = Y_i \frac{n_{it}\hat{\alpha}_t}{\sum_{s=1}^{t} n_{is}\hat{\alpha}_s} \tag{2.3}$$

as a function of the current estimate of $\hat{\alpha}$. The updated β values can be obtained from a Poisson regression with offset $\log n_{it}$ using ϕ_{it} as the outcome variables.

2.2.1 Bayesian Model Averaging

Our approach involves fitting many different change point models and combining the information from all models to quantify the evidence for changes in approval rates at certain times. Recall that we use $\tau = \{s : \hat{\beta}_s \neq 0, s > 1\}$ to represent a model. Each model includes the number of change points $|\tau|$ and the associated change times.

Following Neath and Cavanaugh (2012), we approximate the posterior probability of τ given the poll data D using the Bayesian Information Criterion (BIC)

$$\hat{p}(\tau \mid D) \propto e^{-B(\tau)/2}\, p(\tau)$$

where

$$B(\tau) = -2 \log \hat{L}(\tau) + d(\tau) \log p$$

is the BIC for model τ, $\hat{L}(\tau)$ is the observed likelihood using the MLE estimated coefficients of $\hat{\beta}$ obtained from the EM algorithm, $d(\tau) = 2 + |\tau|$ are the number of estimated model parameters, and p are the number of polls. The prior probability of model τ can be written

$$p(\tau) = p(|\tau| = k)\, p(\tau_1, \ldots, \tau_k \mid |\tau| = k)$$

where the first component corresponds to the prior on the number of change points and the second on the locations of the k change points.

Once we have have estimated $\hat{p}(\tau \mid D)$, then many useful estimates can be obtained. For example,

$$\begin{aligned} \widehat{\Pr}(k \text{ change points}) &= \sum_{\tau} \hat{p}(\tau \mid D) \cdot I(|\tau| = k) \\ \widehat{\Pr}(t \text{ is a change point}) &= \sum_{\tau} \hat{p}(\tau \mid D) \cdot I(t \in \tau) \end{aligned} \tag{2.4}$$

Similar to Lerman (1980), we take a grid search approach to obtain information on the evidence about $|\hat{\tau}|$ and $\hat{\tau}$. Considering how media (television, newspapers and social media) respond to political events or campaigns and appropriate time it takes for the voting public to react, we set the gap between the change point to one-week. Also, only those events of significance are able to shift people's option (Shaw, 1999). Hence, we set the maximum number of change points for the time period considered to five. **After the grid search, the evidence about the change points is combined using equations 2.4 to get the predicted number of joinpoints $|\hat{\tau}|$ and their locations τ.**

The complete algorithm for our proposed model is summarized in Algorithm 1.

2.3 Simulation

To illustrate the performance of our approach, we carried out a simulation study. With the aim of mimicking realistic data we used the observed US Presidential approval rating data (detailed in Section 2.4) to guide the simulation. **Specifically, we used the observed times and intervals of the 73 polls (spanning 315 days) and**

Algorithm 1 EM Algorithm for Estimating Change Point Model with Censored Data

```
Initialize the support rate α_t as a constant
for Each set of change points considered in the search space do
    while Convergence criteria is not met do
        for Each poll_i do
            Calculate expectation of latent vector using equation 2.3
        end for
        Use Poisson regression to find the β vector that maximizes the expected log
likelihood in equation 2.2
    end while
end for
```

only simulated the number of respondents and number of supporters. The simulated number of respondents of poll i was generated as a Poisson random variable

$$N_i \sim Pois\left(\sum_{t=1}^{T} n_{it}\right)$$

where n_{it} is from the observed data. The total number of supporters is generated from a binomial distribution

$$Y_i \sim Binom\left(N_i, \frac{\sum_{t=1}^{T} n_{it}\alpha_t}{\sum_{t=1}^{T} n_{it}}\right)$$

where α_t is a specified approval rate.

We considered four different scenarios corresponding to zero through 3 true change points. Figure 2.1 shows the underlying approval rates used for each scenario. For the three change point scenario, the underlying change point ares set to March 23rd, June 29th and August 31st, respectively. For the two change point scenario, the underlying change point are set to March 30th and April 27th, respectively. For the one change point scenario, it's set to June 29th. To evaluate how well our method can detect the real number of change points and how certain it is about

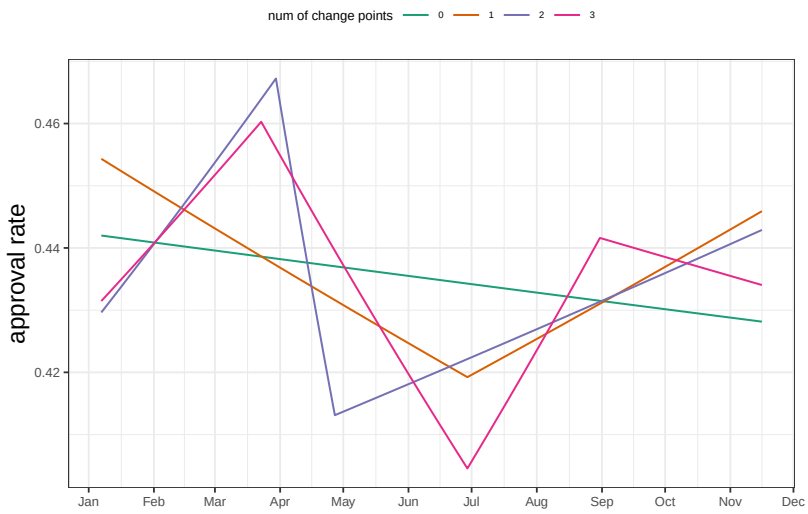

Figure 2.1: Approval rates used for simulation. Each line represents a simulation scenario.

the locations detected, **we simulated data from each scenario 100 times.**

k	% $(\hat{k} = k)$	Avg $\hat{P}(\tau \mid D) > \hat{P}(\tau_{\text{true}} \mid D)$
0	100%	0.00%
1	89%	0.00%
2	95%	0.01%
3	28%	0.31%

Table 2.1: Performance on simulated data. k is the true number of change points, $\%(\hat{k} = k)$ is the percentage of simulations that favored the true number of change points, and Avg $\hat{P}(\tau \mid D) > \hat{P}(\tau_{\text{true}} \mid D)$ is the average percentage of models that were favored over the true model.

Key performance metrics are summarized in Table 2.1. The result suggests that the algorithm is able to detect the true number of change points very well **in the case of no change points, one change point, and two change points.** In addition to estimating the correct number of change points, we are also interested in the proportion of models $\hat{\tau}$ that have larger model evidence than the true underlying model τ. If this value is small, the model is confident in finding the true change point locations. The mean number of models that are favoured over the true underlying

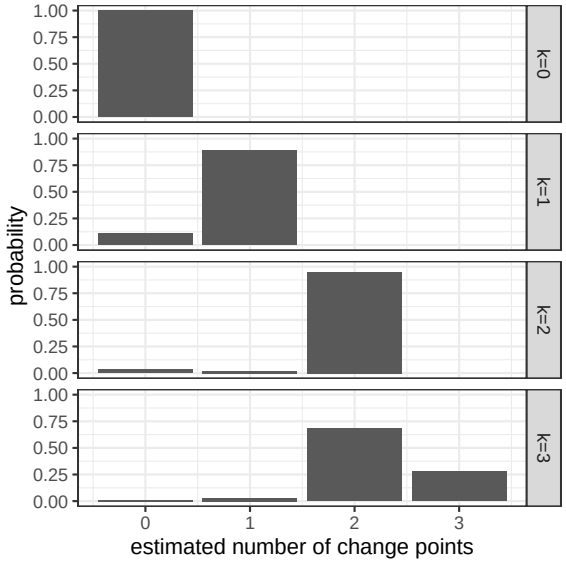

Figure 2.2: Distributions of most likely number of change points from simulated data.

change points is adequately small compared with the sheer number of models that have been considered. It was also found that the models that were preferred over true models often include combination of change points that are in the neighborhood of real change points used.

Figure 2.2 shows the **distribution of the estimated number of change points for each k considered in the** simulation study. **This shows that the model is able detect the true number of change points in the first three cases considered. For the three change point scenario, the model more often favors the two change point model. There are two reasons for this finding. First, we specified the prior distribution on the number of change points to be a truncated geometric with $p = .4$, which will favor $\hat{k} = 2$ over $\hat{k} = 3$. The other reason is that there are several two change point models that have similar likelihoods and even smaller BIC scores then the best three change point model.** Because the simulated data closely follows the real data,

this result indicates that we may have difficulty distinguishing between two and three change point solutions.

2.4 Results

We illustrate our method on the 2020 US Presidential approval rating. The polling data comes from 538 (Silver, 2020a). We consider the 73 polls conducted in 2020 that had an A- or better rating and used the bias adjusted approval rates (Silver, 2020b). The adjusted approval rates take into account all sorts of possible bias ranging from sample size, pollster to how surveys are conducted. **We used a truncated ($k \leq 4$) geometric prior, with $p = 0.4$ to model the *a priori* number of change points, used a uniform prior on the locations of the change points, and evaluated the model for change points every 7 days apart. This led to a total of 149986 models evaluated which took less than 1 hour to run. Code for replication can be found at: `github.com/mdporter/presidential-approval`.**

Figure 2.3 shows the estimated fit using the best single model (i.e., lowest BIC) for zero to four change points. Figure 2.4 shows the estimated distribution of the number of change points which gives a preference to the $k = 2$ change point models. Figure 2.5 shows the estimated distribution for the first two change points. While there is strong evidence for the first change point being around the end of March, the second change point is less certain with a dates around the end of April or end of June being favored. The upper panel of Figure 2.6 shows the distribution of the change point locations. The filled polygons show the contribution for all four potential change points. Again, this shows the uncertainly around the second change point but also indicates there is limited support for the three or four change point models.

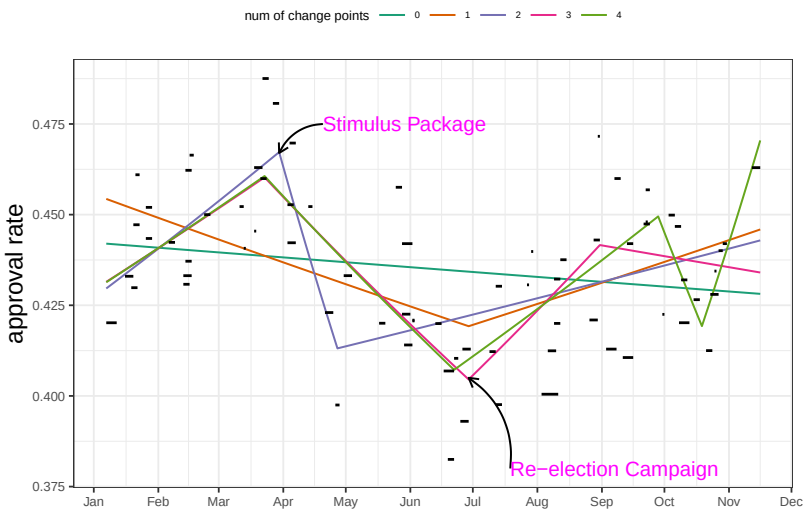

Figure 2.3: Estimated approval rates from the best models for each k change points. The black horizontal segments correspond to the observed aggregate polls.

We found three significant events in Trump's presidency and campaign that align with these three time periods. The end of March sees Trump's approval peak and begin a decline. This is during the COVID-19 pandemic when Trump signs the largest stimulus package in US history (CARES Act on March 27), announces estimates that 240K Americans are likely to die from the virus, and most Americans are under stay at home orders. The second mode around the end of April follows Trump's suggestion that COVID-19 could possibly be treated by injecting bleach or UV light into a human (April 23). The third mode occurs at the end of June when Trump's approval rating hits and all-time low and begins increasing. This change point corresponds to the start of Trump's re-election campaign and first rally in Tulsa, OK (June 30).

To further explore the potential change point explanations, we collected google trend keyword searches during 2020 (Google, 2020). The bottom of Figure 2.6 shows how relative search interest for the three events discussed above line up with the distribution of change point locations. The sharp peaks in the searches align with the

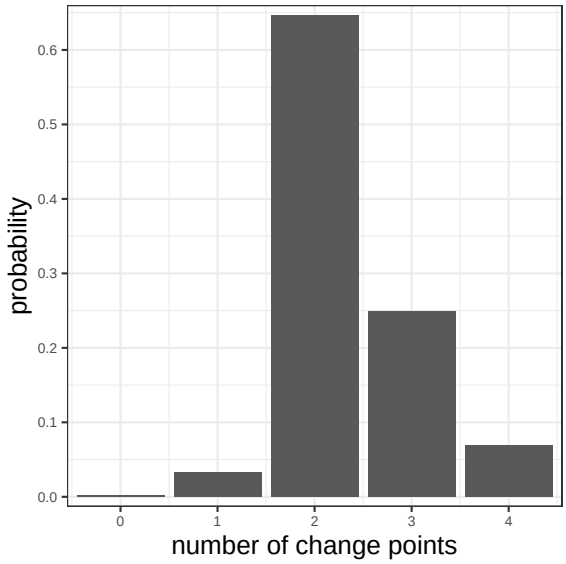

Figure 2.4: Estimated distribution of the number of change point for the observed data.

identified change points well, indicating that the public paid increased attention to the events listed here and these events have potential to trigger a change in voting public's opinion.

2.5 Discussion

This article introduces our approach to joinpoint regression modeling with interval-censored survey data. Our approach comprises three steps: (i) an efficient EM based algorithm for estimating model parameters in a Poisson regression with interval censored, or aggregated, survey data; (ii) a search over a large collection of possible change point models; and (iii) using Bayesian model averaging, based on BIC, to accumulate the information from all models about the number and location of change points. This approach was used to efficiently discover potential change points and model trends in the 2020 approval ratings of US president Donald Trump. This

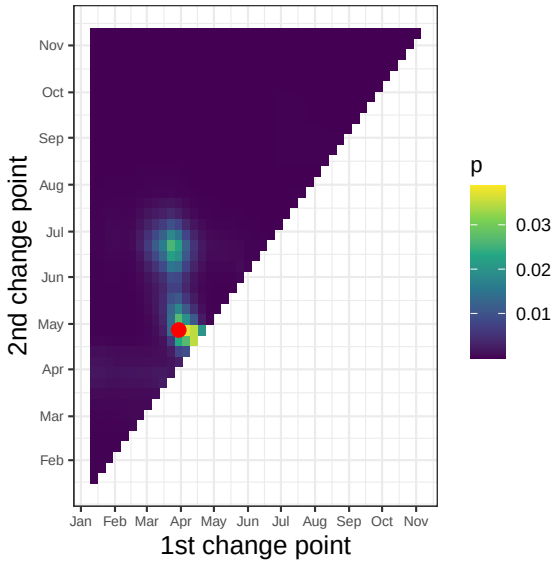

Figure 2.5: Estimated distribution of the location of the first two change points.

revealed three potential change points that we associated with significant events in Trump's presidency and campaign.

We evaluated the performance of our approach on realistic simulated data. By matching the properties of the actual data, our simulation performance can give a more accurate assessment of how well the model can identify the true change points. For the data we analyzed in this paper, the simulation exercise gives us confidence in the reported change points.

There are also some factors that could affect the performance of the purposed algorithm and the understanding of the results. Even though we filtered out polls with less trustworthy ratings, factors like sample size and bias introduced by pollsters also have effects on the data quality. Besides these, the length of polls affects the ability to capture the true structure. Longer intervals introduce extra variability that will lead to smoother (fewer change points) estimated support rates. In cases where few polls span the change points, it would be difficult for the algorithm to detect both

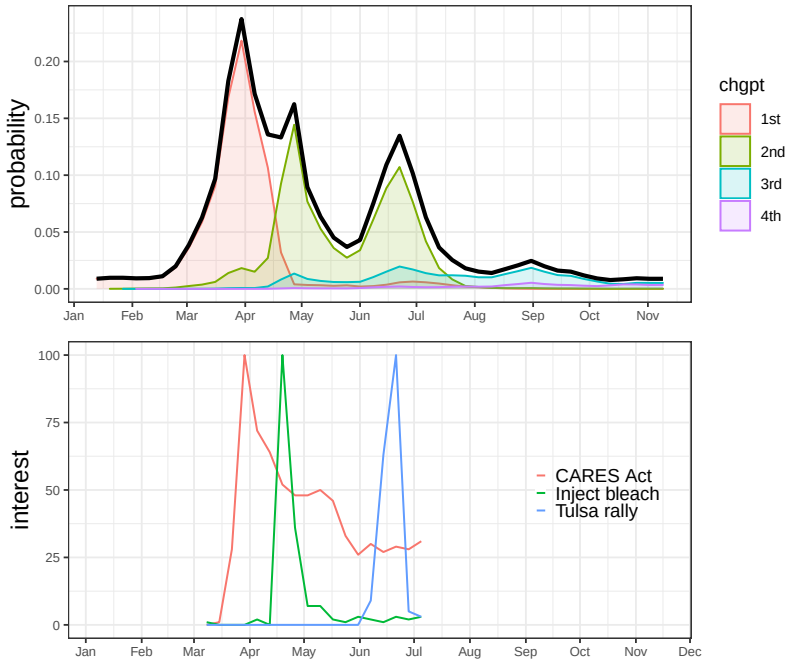

Figure 2.6: Distribution of change point location (top) and relative interest of Google searches by keyword (bottom).

the $\hat{k}$ and model $\hat{\tau}$.

While we developed this approach for polling data, it can be modified to address joinpoint modeling of other aggregated, or interval censored, data. If the focus is on trend modeling, without change point detection, instead of the best-subsets type of search we used here, a lasso or ridge type penalty on adjacent coefficients could be added to the likelihood to encourage a smoother trend estimate (see Eilers and Marx (1996) and Kim et al. (2009) for details). The model can be used as a standalone method to detect both the number of change points and locations. The results can provide statistical evidence to findings in political science (Shaw, 1999).

Chapter 3

Intensity Estimation

The modeling of censored data is extensively explored in the biostatistical community, particularly in contexts where the focus lies in making inferences regarding time-to-event. However, the application of censored data modeling outside the field of biostatistics is less mature. Typically, methods adopted in non-biostatistical domains often result in biased inferences. Crime modeling is one such domain where prevailing practices in censored data modeling yield overly smooth curves, impeding the identification of peaks and valleys in intensities. Consequently, the derived intensity estimates become less useful to informed decision-making. In this project, we introduced a method for estimating intensity in the context of interval-censored data, aiming to provide law enforcement agencies with precise intensity estimates for the optimization of resource allocation. The method incorporates two distinctive penalty terms, coupled with hierarchical clustering, to ensure uniformity of estimates within the same cluster. This innovative approach provides law enforcement agencies with the flexibility to achieve the maximum level of crime reduction with the resources available. Our method yields intensity estimates subject to dual penalties and incorporates a clustering algorithm, resulting in an optimized and realistic patrol plan. It is noteworthy that our approach diverges from the two-stage methodology outlined by (Camacho-Collados and Liberatore, 2015), wherein intensity estimates are initially derived without constraints, followed by a subsequent optimization process to devise

a patrol strategy under realistic constraints. Our method provides a holistic approach to get accurate estimates with constraints imposed in the form of penalties.

Time of week intensity estimation from interval censored data with applications to police patrol planning

Abstract Law enforcement agencies are tasked with crime prevention and crime reduction under limited resources. Having an accurate temporal estimate of the crime rate would be valuable to achieve such a goal. However, estimation is usually complicated by the interval censored nature of crime data. We cast the problem of intensity estimation as a Poisson regression using an EM algorithm to estimate the parameters. Two special penalties are added that provide smoothness over the time of day and day of week. This approach provides accurate intensity estimates and can also uncover day of week clusters that share the same intensity patterns. Both simulated and real crime data gathered from the city of Cincinnati and the city of Dallas are used to demonstrate the effectiveness of the proposed model.

Keywords:Intensity Estimation; EM Algorithm; Cluster Detection; Interval Censoring; Patrol Planning; Smart Policing Initiative

3.1 Introduction

Anticipating where and when crimes might occur is a key element to successful policing strategies (National Academies of Sciences, Engineering, and Medicine and others, 2018). However, this task is complicated by the presence of interval censored data. The censored data refers to the type of data that the event time is only known to lie within an interval instead of being observed exactly. This type of data is prevailing in the field of criminology because of the absence of victims for certain types of crime. Despite its importance, the research in temporal analysis of crime has lagged behind the spatial component (Ashby and Bowers, 2013). Inspired by the success of solving

crime-related problems with statistics (Ratcliffe, 2000; Porter, 2016; Koch, Tian, and Porter, 2020), we proposed a statistical model for the temporal intensity estimation of crime with interval censored data. The model is built on Poisson regression and has special penalty terms added to the likelihood. The proposed model is able to yield accurate intensity estimates and generate meaningful day of week clusters compared with the competing method.

Our research is in line with the smart policing initiative (SPI) proposed by the Bureau of Justice of Assistance (BJA) as an effort to support law enforcement agencies in building evidence-based, data-driven law enforcement tactics. The goal is to identify strategic approaches that are effective in crime prevention and reduction. One of the key practices specified by SPI is Strategic Targeting which refers to the analysis that could help agencies focus on a small percentage of people or places with limited resources. In our case, we allow agencies to deploy their resources in a relatively short period of time to achieve the maximum level of crime reduction. By analyzing a particular area within cities where data are available, our proposed approach could not only provide an accurate estimate of intensities for the time unit considered, but a time-variation crime incidence pattern. Both will be helpful in the allocation of limited resources when the agencies design their patrol plan. We discussed how the plan could be improved with the estimates given by the model in Section 3.4.

The remainder article organizes as follows. In Section 3.2, we introduce the research in the area of criminology and censored data and describe the model of the penalized temporal intensity estimation method and a procedure for tuning the penalty parameter. We present both simulation study and real data analysis in Section 3.3 with the idea of repeated-cross validation to demonstrate how our proposed framework could generate more accurate results. The concluding remarks and the discussion are given

in Section 3.4 and 4.7.

3.2 Temporal analysis in crime and censored data

3.2.1 Temporal analysis

Research into the temporal patterns associated with different types of crimes provides us with insights into criminal behavior. For instance, offenders are more likely to commit crimes in the early evening (Tompson and Townsley, 2010). Weekends usually see a large number of residential burglaries. This illustrates how the knowledge of the victim's behavior affects the criminal's behavior. Research showed that crime rates are found to go up on holidays (Cohen and Felson, 1979). One possible explanation is the lack of capable guardians and suitable targets encourage offenders to commit crimes. It was also shown that crime rates are found to be related to both the type of crime and the type of holidays (Cohn and Rotton, 2003). It has been shown in the same paper that crime rates of violent crime go down on major holidays and the opposite is true for property crimes. The effects of the type of holiday on crime rates were also investigated using routine activity theory (Towers et al., 2018). This type of research is valuable to our research in the sense that it shows how crime rates could vary with time and provides us with an idea to describe the variation with statistical power.

Generating temporal maps of when crimes occur is quite difficult because of the nature of crimes and possible data issues. The research showed how circular statistics can be adapted to analyze crimes by time of day and day of week. In this project, we estimate the crime intensity for each hour in a week, leading to 168 estimated intensity (Brunsdon and Corcoran, 2006). To overcome the difficulties brought by

the indeterminate timing at which the crime actually occurs, aoristic analysis was proposed and it was shown that this method was able to uncover underlying temporal patterns in burglary crime data that would not be observed using other analysis methods (Ratcliffe and McCullagh, 1998). The aoristic approach assumes that the contribution of an event to each time unit covered by the time interval is equal. Each event is processed individually without considering all other events that happened around the same time. The research suggests that the temporal distribution generated by the aoristic method was smoother than other methods which showed peaks related to routine activities of burglary victims, reducing irregularities in the data set. (Ratcliffe, 2000).

Though the aoristic method has a variety of advantages like easy-to-understand, computationally inexpensive, the aoristic method also oversimplifies the problem. One important idea in supervised learning in the area of machine learning/data mining is bias-variance tradeoff (Hastie et al., 2009). The aoristic approach produces a model with a low variance but high bias by oversmoothing the rate, thus interesting patterns could be concealed. The deployment strategy could go wrong in two possible ways; when a crime occurs without police officers present or no crimes occur when police officers are on duty. The first situation adversely affects the effectiveness of developed strategies. The second situation, however, is a waste of valuable resources. Intensity estimates provided by the method that is over-smooth could conceal peaks and valleys in intensity. The incapability to detect peaks causes the absence of policing at crime scenes, and the inability to detect valleys could waste limited resources available. Therefore, such temporal analysis only provides limited help in terms of crime prevention. In other words, the performance of the fitted model on unseen data could be impaired by adopting a method that can't fully extract the information from

the data. The fact that the aoristic method doesn't consider the other events while generating temporal distribution increases bias. By acknowledging the weakness associated with the current practice, there is a need for novel approaches for intensity estimation when interval censoring is present which takes advantage of information that exists in the data to generate more accurate temporal intensity.

3.2.2 Censored data

Due to the uncertain nature of the data, we manage to solve the problem from the perspective of censored data analysis. Interval censored data could be observed in a variety of areas including epidemiological, financial, medical, and sociological studies. In politics, the interval censored data appear because pollsters only report aggregate statistics for polls conducted (Tian and Porter, 2022b). In medical research, the interva l censored data arise because the event of interest requires laboratory tests or a comprehensive examination (Goggins and Finkelstein, 2000). Also in crime analysis, the interval censored data were recorded because the victims were not present when certain types of crime happened and the event timing is only known to occur within an interval (Ratcliffe and McCullagh, 1998). Censored data has been well studied in the medical field where the patients' progression of disease is assessed periodically and the time to event is known to occur between two adjacent visits. Extensive literature on regression analysis for interval censored data is available. Most of the work has been focusing on the proportional hazard model(Cox, 1972). The use of the Cox model for interval-censored data with discrete hazard assumed is considered in (Finkelstein, 1986). Asymptotic properties of the Cox model and nonparametric estimation are investigated in (Huang and Wellner, 1997). Hazard regression for interval censored data using linear splines was studied in (Kooperberg and Clarkson, 1997). Penalized

spline approach for hazard estimation was proposed in (Cai and Betensky, 2003). The nonparametric method for estimating survival function was purposed in (Peto, 1973).

We consider the problem of temporal intensity estimation in the context of interval censoring for crime data. We adopted a piecewise Poisson model for each temporal unit considered, which is similar to the work in (Friedman et al., 1982) where a piecewise exponential model was assumed. An Expectation-Maximization (Dempster, Laird, and Rubin, 1977) is then used for the estimation of densities by optimizing the likelihood function incorporating penalty structure which is based on routine activity theory.

3.2.3 Notation and model

Let $T \in [0, 168)$ denote the time of the week, in hours, that an event of interest (e.g., a crime) occurred. The event time can either be observed exactly ($T = t$) or only observed within an interval ($T \in X$), where $X \subseteq [0, 168)$. We are interested in modeling the weekly crime intensity to help with police patrol planning. We discretize time into a set of J equally sized bins and let λ_j be the intensity in the jth bin denoted by b_j. Each bin has a width of $\text{bw} = 168/J$ hours with the first bin starting at time 0. Let N be the number of events and t_i the true event time of event i, which is not known exactly for interval censored events.

This specifies a piecewise constant model for the intensity $\lambda(t) = \sum_j \lambda_j I(t \in b_j)/\text{bw}$. We denote w_{ij} as the proportion of bin j that is contained in event i's interval:

$$w_{ij} = \begin{cases} \int_{X_i} I(t \in b_j)dt/\text{bw} & \text{censored events} \\ I(t_i \in b_j)/\text{bw} & \text{uncensored events} \end{cases} \tag{3.1}$$

The log-likelihood for our partially interval-censored data consists of three components. Following Kim, 2003, one component represents the density from the uncensored data and another the probability for the interval censored events. The third component represents the overall event rate. The notation introduced in (3.1) allows the first two components to be written in the same form giving the log-likelihood:

$$\begin{aligned}\log L &= \sum_{i\in\text{unc}} \log f(t_i) + \sum_{i\in\text{cen}} \log \Pr(T \in X_i) + \log \Pr(N = n) \\ &= \sum_{i=1}^{N} \log\left(\frac{\sum_j \lambda_j w_{ij}}{\sum_j \lambda_j}\right) + N \log\left(\sum_j \lambda_j\right) - \sum_j \lambda_j - \log N! \\ &= \sum_{i=1}^{N} \log\left(\sum_j \lambda_j w_{ij}\right) - \sum_j \lambda_j - \log N!\end{aligned} \tag{3.2}$$

Our model for the intensity is $\log \lambda_j = \beta_j$ for $j = 1, 2, \ldots, J$. While this model stipulates one parameter per bin, we will use two penalties during the estimation to discourage over-fitting. The first penalty is what we call *time of day* penalty, and encourages adjacent estimates (e.g., β_j and β_{j+1}) to be close to produce a smoother estimate. The other type of penalty is the *day of week* penalty, which forces estimates at the same time on similar days to be close (e.g., β_j and β_{j+24}). This encourages similar estimates for different days within the same day-of-week cluster, over all 24 hours. The addition of this penalty is backed by the routine activity theory (Andresen and Malleson, 2015) which relates offending behaviors to the daily patterns of social interaction. For example, the routine activities of potential victims are often similar on weekdays, but different on the weekend days.

The penalty takes the form:

$$\text{Pen}(\beta) = \phi_1 \beta^T K_1 \beta + \phi_2 \beta^T K_2 \beta \tag{3.3}$$

where $\beta^T = [\beta_1, \beta_2, \ldots, \beta_J]$, K_1 and K_2 are $J \times J$ matrices that enforce the two types of penalties. The ϕ_1 and ϕ_2 are the strength of the two penalties. The time of day penalty matrix is $K_1 = D_1^T D_1$, where D_1 is the first order difference matrix:

$$D_1 = \begin{bmatrix} -1 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & -1 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & -1 & 1 \end{bmatrix}$$

This difference matrix specifies the time of day penalty $\beta^T K_1 \beta = \sum_{j=1}^{J} (\beta_j - \beta_{j-1})^2$ to penalize differences in adjacent estimates.

Similarly, the day of week penalty matrix is defined as $K_2 = D_2^T D_2$. The difference matrix D_2 is more complex to incorporate a grouping structure that specifies the days of the week that should have more similar estimates. For example, if we have a group of days corresponding to weekdays, we want to encourage the estimates for 9 AM (and every other hour of the day) on Mondays - Fridays to be close. Specifically, we develop a difference matrix that penalizes the variance of the estimated coefficients for the same hour in each group.

As an example, consider a two-group structure that specifies a weekday group and a weekend group. Considering a one-hour binwidth, let $d_j \in \{1, 2, \ldots, 7\}$ indicate the day and $h_j \in \{1, \ldots, 24\}$ the hour of bin j, respectively. Also, let $\beta'_{d,h}$ denote β_j where $d_j = d$ and $h_j = h$. The penalty is

$$\beta^T K_2 \beta = \sum_{d=1}^{5} \sum_{h=1}^{24} \left(\beta'_{d,h} - \frac{1}{5} \sum_{d'=1}^{5} \beta'_{d',h} \right)^2 + \sum_{d=6}^{7} \sum_{h=1}^{24} \left(\beta'_{d,h} - \frac{1}{2} \sum_{d'=6}^{7} \beta'_{d',h} \right)^2$$

where the first term with $d \in \{1, 2, 3, 4, 5\}$ specifies the weekdays and the second term with $d \in \{6, 7\}$ the weekend days. The day of week difference matrix, D_2, for this scenario is illustrated in Figure 3.1. D_2 is a 168 x 168 matrix, with each row and column corresponding to a specific hour in a week. In the context of the weekday-weekend two-group structure, this plot is divided into two segments at hour 120, which marks the end of Friday. The diagonal line in the matrix represents the hours subject to penalization, while any other non-zero entries in the same row represent the same hour on different days within the same group. In addition to preventing

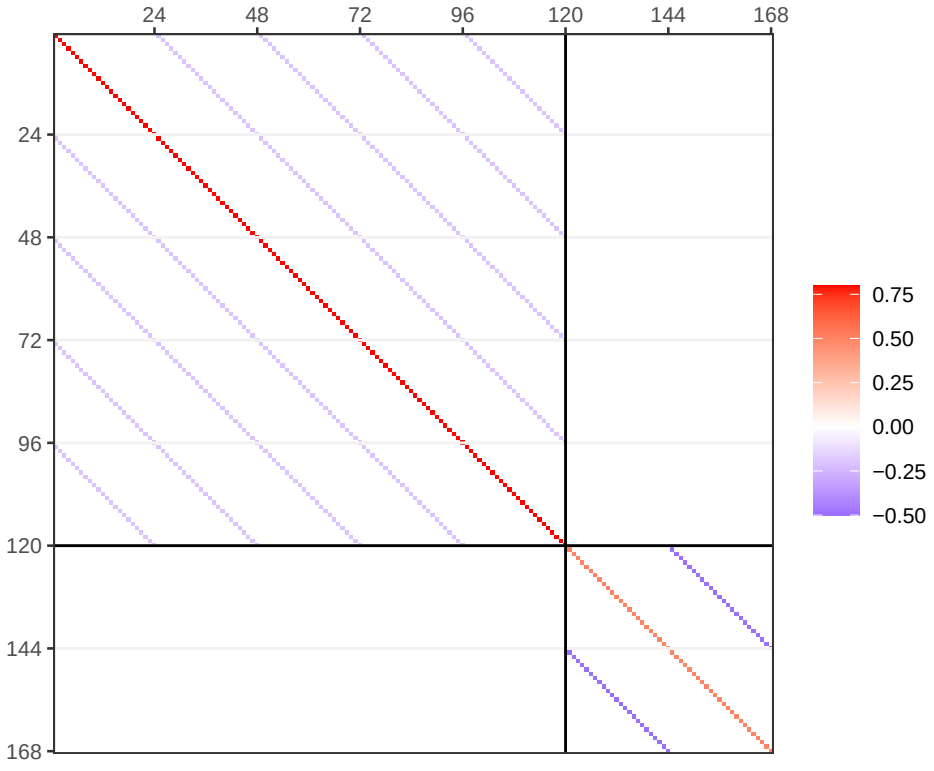

Figure 3.1: The day of week difference matrix, D_2, for a two-group (weekdays and weekend) scenario with one-hour bin width.

over-fitting, the penalties also allow estimation in bins in which no events occurred. As long as there is a small time of day penalty (i.e., $\phi_1 > 0$) then all intensity related parameters can be uniquely estimated (Eilers and Marx, 1996).

3.2.4 An expectation-maximization algorithm

The penalized log-likelihood we want to maximize is

$$\begin{aligned} J(\beta) &= \log L(\beta) - \text{Pen}(\beta) \\ &= \sum_{i=1}^{N} \log \left(\sum_j e^{\beta_j} w_{ij} \right) - \sum_j e^{\beta_j} - \left(\phi_1 \beta^T K_1 \beta + \phi_2 \beta^T K_2 \beta \right) \end{aligned} \tag{3.4}$$

Due to the summation inside the log this equation doesn't lend itself to direct optimization techniques. Therefore, we introduce a latent vector for each event $Z_i = [Z_{i1}, Z_{i2}, \ldots, Z_{iJ}]$ where Z_{ij} takes a value of 1 if the true event time t_i falls into bin j and 0 otherwise. The partially observed number of events occurring in bin j, denoted as y_j, is represented in terms of the latent variable Z_{ij}, where $y_j = \sum_{i=1}^{N} Z_{ij}$. If the latent vectors were known, the *complete* penalized log-likelihood could be written (up to a constant)

$$J_c(\beta) = \sum_{j=1}^{J} \left(y_j \beta_j - e^{\beta_j} \right) - \left(\phi_1 \beta^T K_1 \beta + \phi_2 \beta^T K_2 \beta \right) \tag{3.5}$$

which is easier to optimize. The details of the derivation are given in the appendix.

E-step

The conditional expectation of (3.5) at the $(r+1)$th iteration is :

$$\begin{aligned} Q(\beta, \beta^r) &= E(J_c(\beta)) \\ &= \sum_{j=1}^{J} \left(E[y_j]\beta_j - e^{\beta_j}\right) - \left(\phi_1 \beta^T K_1 \beta + \phi_2 \beta^T K_2 \beta\right)) \end{aligned} \tag{3.6}$$

The expected bin counts are given by

$$\begin{aligned} E(y_j|\beta^r, X_i) &= \sum_{i=1}^{N} E(I(Z_{ij}=1)|\beta^r, X_i) \\ &= \sum_{i=1}^{N} P(Z_{ij}=1|\beta^r, X_i) \\ &= \sum_{i=1}^{N} \frac{w_{ij}\lambda_j^r}{\sum_k w_{ik}\lambda_k^r} \end{aligned} \tag{3.7}$$

M-step

After the pseudo counts are estimated in (3.7), the expected penalized log-likelihood (3.6) can be maximized using the Newton-Ralphson algorithm with:

$$\beta^{r+1} = \beta^r - H^{-1}U \tag{3.8}$$

where H^{-1} and U correspond to the inverse Hessian and Score, respectively. They can be obtained through the equation below:

$$U = \frac{\partial Q}{\partial \beta} = \hat{Y} - e^{\beta^r} - P_\phi \beta^r - P_\phi^T \beta^r \tag{3.9}$$

$$H = \frac{\partial^2 Q}{\partial \beta^r \, \partial \beta^{rT}} = -W - P_\phi - P_\phi^T \tag{3.10}$$

where $\hat{Y} = E(y|\beta^r, X_i)$ is the vector expected bin counts and the matrix $P_\phi = \phi_1 K_1 + \phi_2 K_2$ lumps the two types of penalties together. Therefore, the updating formula for parameters β is:

$$\beta^{r+1} = (W + P_\phi + P_\phi^T)^{-1} W A \tag{3.11}$$

matrix W is a diagonal matrix with entries e^{β_k} on the diagonal and zeros everywhere else. A is

$$A = W^{-1}\hat{Y} - W^{-1}e^{\beta^r} + \beta^r \tag{3.12}$$

The EM algorithm iterates between equations 3.7 and 3.11 until successive estimates of parameters are close enough to meet the stopping criterion. In our implementation, the algorithm stopped when $max_k(\beta_k^{r+1} - \beta_k^r) < \epsilon$, where $\epsilon = 10^{-6}$.

Model selection

Given a model specified by hyper-parameters (ϕ_1 and ϕ_2) and the time of day penalty matrix (K_1) and day of week penalty matrix (K_2) defined, a criterion that allows a comparison between different models is necessary. We propose the use of the Akaike Information Criterion (AIC) to select an accurate and parsimonious model (Sakamoto, Ishiguro, and Kitagawa, 1986).

We define AIC as

$$\text{AIC} = -2 \log L + 2 \text{ edof}$$

The first term in the equation represents the observed log-likelihood defined in (3.2), and edof represents effective degrees of freedom, which is defined as the trace of the smoother matrix (i.e., the hat matrix defined below). The hat matrix S is an approximation to the degrees of freedom in the presence of penalty Hastie and Tibshirani,

1990. The hat matrix S can be obtained from (3.11) with algebraic manipulation:

$$S = (W + P_\phi + P_\phi^T)^{-1} W$$

3.2.5 Uncovering the grouping structure

Besides the regularization effects, the penalty terms added in the likelihood (3.3) also introduce a way to search for the underlying grouping structure given a particular city. In other words, instead of imposing a pre-specified grouping structure, the model allows us to find the appropriate grouping structure with given data.

Inspired by hierarchical clustering, the search starts with the scenario where each day is in its own group. We let $\mathbf{C}$ denote the number of clusters. At each iteration, the algorithm searches for all combinations of two clusters and merges the two clusters together that produces the minimum of resulting AIC, leading to $\mathbf{C - 1}$ clusters for the next iteration. The algorithm stops when there exists only one cluster. The grouping structure that yields the lowest AIC value is the most probable one.

Algorithm 2 EM algorithm for estimating temporal intensity with interval censored events.

Input: X, ϕ_1, ϕ_2, K_1, and day of week cluster
Output: Intensity estimate for all bins λ_t
Use uncensored observations to help initialize β with non-censored data and start EM algorithm
while Convergence criteria is not met **do**
  for Each bin j **do**
    Calculate pseudo count using equation $\sum_{i=1}^{N} \frac{w_{ij}\lambda_j}{\sum_k w_{ik}\lambda_k}$
  end for
  Maximize the β using equation $\beta^{r+1} = (W + P_\phi + P_\phi^T)^{-1} W A$
end while
return Intensity estimates λ_t

Algorithm 3 Hyper-parameter tuning and cluster detection

Input: X, K_1, upper and lower bound for ϕ_1 and ϕ_2, and maximum number of iteration

Output: Intensity estimate for all bins λ_t and day of week cluster

while Maximum number of iterations has not reached **do**
  Update ϕ_1 and ϕ_2 based on Bayesian optimization
  Set number of clusters $C = 7$
  while $C > 1$ **do**
    for $\binom{C}{2}$ possible clusters **do**
      use Algorithm 4 to fit model and obtain AIC
    end for
    merge the clusters with the lowest AIC
    set $C = C - 1$
  end while
  Return $\{$day of week structure, ϕ_1, ϕ_2, AIC$\}$ associated with the lowest AIC
end while
return Intensity estimates λ_t, ϕ_1, ϕ_2, day of week structure associated with best model

It's important to note that the shrinkage parameters ϕ_1, ϕ_2 are regarded as tuning parameters. We employed the `rBayesianOptimization` Snoek, Larochelle, and Adams, 2012 package to efficiently fine-tune and identify the optimal combination of ϕ_1 and ϕ_2. Bayesian optimization provides a sensible way to explore areas in parameter space that improve the objective function. Our experiments indicate that typically,

3.3 Real data analysis

Unless the criminal is captured at the scene or a witness observed the crime, it is unlikely that the exact event time of a crime is known. This makes interval censored data prevalent for many crime types (e.g. theft, burglary). To demonstrate the predictive performance and the capability of detecting the grouping structure of the proposed model, we applied it to real data that were gathered in the cities of Cincinnati and Dallas in the United States Department, 2023a; Department, 2023c.

To ensure consistency and avoid any changes introduced by the pandemic, we used one year's worth of data from the year 2019. Cincinnati has five districts in total. Since the patrol officers and shifts are assigned district-wise, we concentrate the analysis on district three which has the most observations in the year 2019 with offense types of Burglary and Breaking and Entering. The total number of observations is 996, out of which 34 observations have the exact event time recorded. For Dallas, we analyzed data on burglary and theft from the Northeast division, resulting in a total of 4293 observations, with 379 of them being non-censored. The crime types analyzed in this study are those that research has shown can be deterred by the presence of police in the right place at the right time Evans and Owens, 2007. Table 3.1 shows the number of each type of observation for the cities considered in this section. For the two types

City	**Censored**	**Uncensored**
Cincinnati	962(96.6%)	34(3.4%)
Dallas	3914(91.2%)	379(8.8%)

Table 3.1: Number of censored and uncensored crimes in 2019.

of offenses of interest, drastically different interval length pattern was observed. In Cincinnati, the median interval length for Breaking and Entering is 13.1, while for burglary it is 5.25. For Dallas, the corresponding values are 8 and 1.16, respectively. We hypothesize that models that analyze each event in isolation, without taking into account the intensity of the intervals they encompass during estimation, are less efficient in addressing interval censored data. This inefficiency arises due to the increased uncertainty associated with longer intervals.

We demonstrate how our model is used to cluster the days of the week such that each cluster corresponds to the days that have similar time of day crime patterns. This can help patrol planners simplify the schedules as the days in each cluster can be

assigned the same patrol times. We also provide an analysis of how well our model predicts crime rates without restricting the days into hard clusters.

3.3.1 Day of week cluster discovery

To understand how crime occurrences vary within the week, we only tune the parameter ϕ_1 and set the parameter ϕ_2 to a large value, forcing parameter estimates within the cluster to be nearly identical. Our analysis suggests that there should be three clusters. Monday, Thursday, and Friday are in a group, Tuesday and Wednesday are in a group, and Saturday and Sunday make the third group. The findings that weekdays and weekends belong to different group matches our expectation as different patterns exist within a week (Andresen and Malleson, 2015).

3.3.2 Intensity Estimation

There are also times when we care more about how accurate those estimated intensities are. In such cases, we lift restrictions of forcing all estimates within a cluster to be nearly identical and tune both penalty coefficients ϕ_1 and ϕ_2 and put less emphasis on the structure discovery. In other words, penalties are imposed to prevent over-fitting and improve the model's performance, rather than focusing on cluster discovery as in the previous section. In order to evaluate the performance of our proposed algorithm and compare it with that of the aoristic algorithm. We took twenty percent of the data as hold-out data. We then obtained intensity estimates from both algorithms and compare their performance in the following scenarios. After obtaining the estimated values of β, we used (3.7) to calculate the expected number of occurrences and determine the proportions of crimes that occurred in each bin. Next, we ranked the bins in decreasing order based on the estimated intensities derived from different

models. We then used the ranking orders from each model to compute the optimal cumulative proportions of captured crimes for each model as a function of the patrol hours.

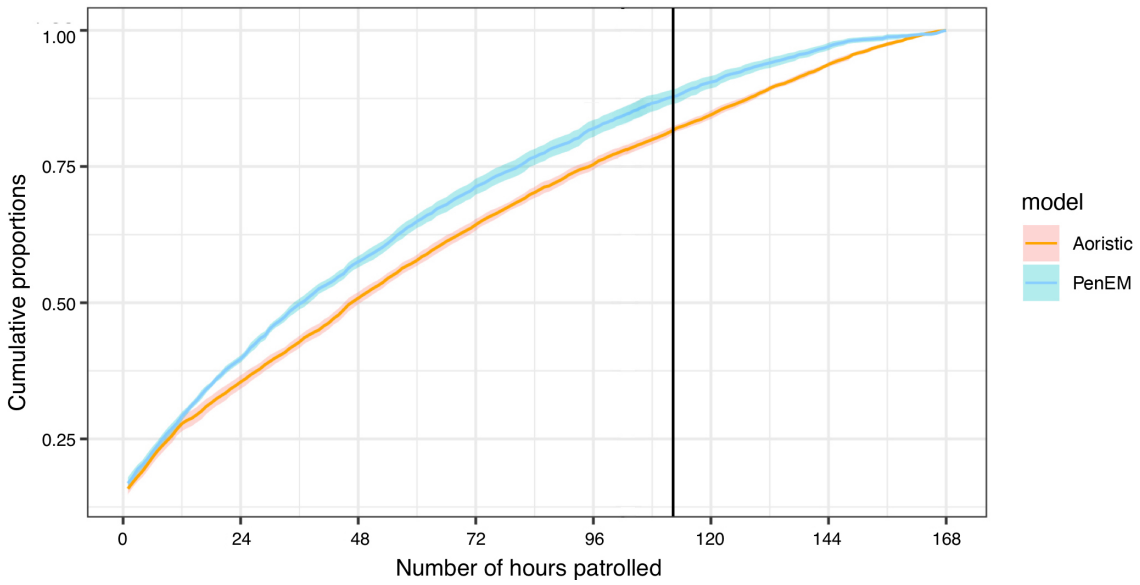

Figure 3.2: Cumulative proportions of crimes captured versus the number of hours patrolled for the city of Cincinnati. The reference line denotes a typical number of patrol hours in a week.

Figure 3.2 shows the performance of both models. The x-axis is the number of hours patrolled. This plot can be interpreted as the percentage of crimes that could be deterred by police patrolling a certain number of hours each week. This not only shows that our proposed model is better at deterring crimes because the area under the curve (AUC) is larger but the percentages are higher for those typical number of patrol hours (e.g., the Northeast division of Dallas police department patrols 112 hours per week Department, 2023b). We also calculate the information gain for the hold-out data for each algorithm. The information gain is the log-likelihood ratio between our fitted model and a reference model. The reference model is a homogeneous Poisson with the intensity of 1. The information gain can be calculated

as follows:

$$Gain(\lambda) = \sum_{i=1}^{n} \log\left(\frac{\sum_j \lambda_j w_{ij}}{\sum_j w_{ij}}\right) - \sum_j \lambda_j + J \tag{3.13}$$

Figure 3.3 displays the variation in performance obtained from conducting the hold-out analysis 20 times. We also carried out a one-sided paired t-test to test the

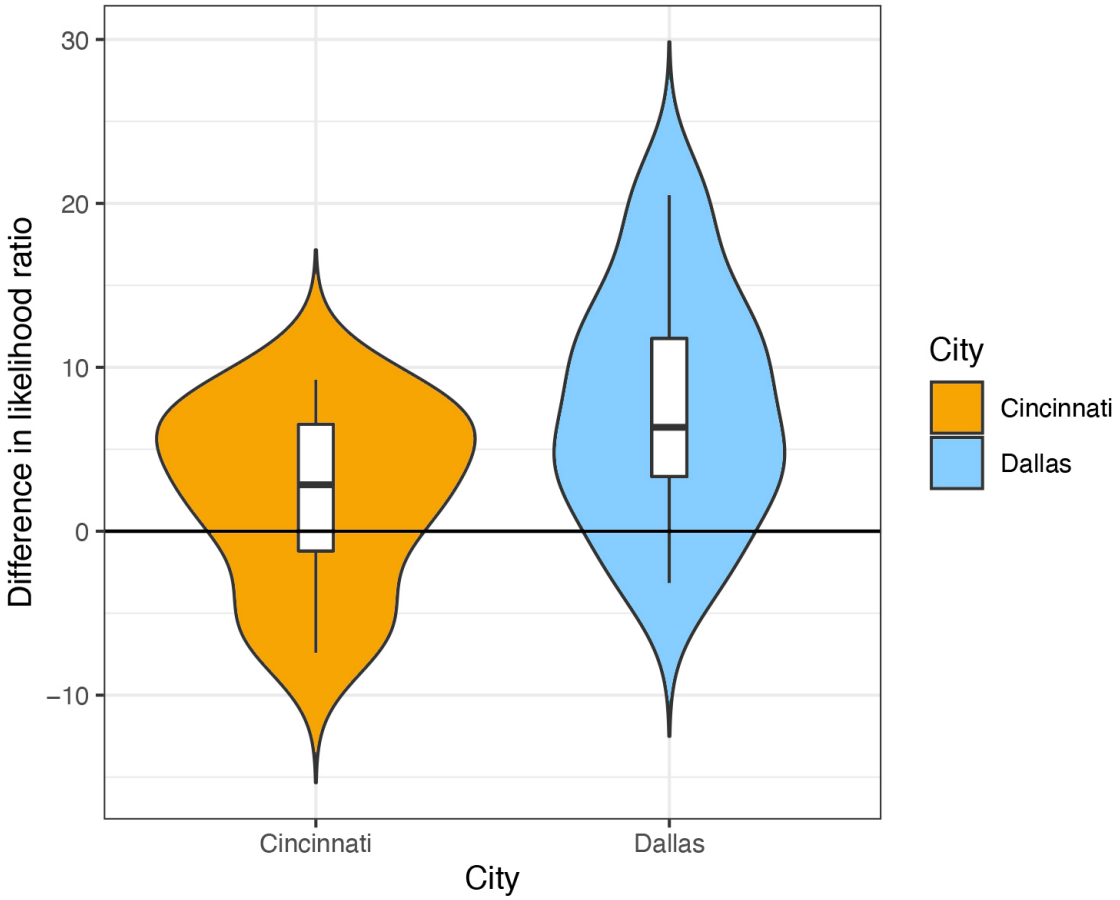

Figure 3.3: Violin plot showing variation of difference in likelihood ratio returned by two models for cities analyzed.

difference between the information gain returned by two methods for both cities. The p-values of 0.013 for Cincinnati and 0.006 for Dallas obtained from the paired t-tests suggest that the null hypothesis of equal performance can be rejected at a significance

level of 0.05, indicating that the proposed model has improved performance compared to the aoristic model.

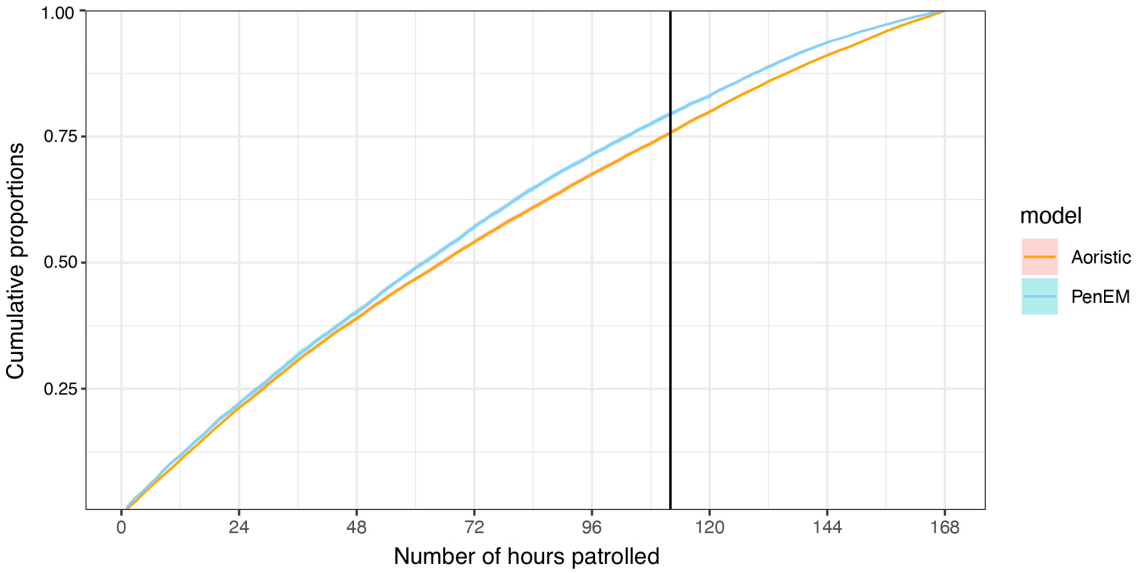

Figure 3.4: Cumulative proportions of crimes captured versus the number of hours patrolled for the city of Dallas. The reference line denotes a representative number of patrol hours in a week.

We created a plot similar to Figure 3.2 for Dallas, which also demonstrates the performance improvement achieved by our model, as shown in Figure 3.4. It is noteworthy that the benefit derived from employing the proposed algorithm is more pronounced in the case of Dallas compared to Cincinnati. This observation is supported by the density curve of the difference in likelihood ratio depicted in Figure 3.3. One possible explanation for this performance improvement is the disparity in the size of the datasets between Dallas and Cincinnati. The larger number of observations in the Dallas dataset allows for a more robust estimation of the model parameters. This is explored in more detail with a simulation study.

3.3.3 Simulated Data Analysis

In order to explore how data size and interval lengths affect performance, we performed a simulation study using simulated data that closely mimics the real data from Cincinnati. To generate the simulated data, we utilized the intensity estimates derived from our model, which includes three distinct day-of-week clusters. The exact observations were initially generated using inverse transform sampling and subsequently transformed into interval-censored observations, with the interval length determined by random draws from an exponential distribution whose mean was set to the average interval length observed in the actual data. We considered four different scenarios that varied in the size of the training data and the proportions of non-censored observations. Specifically, panels A and B in Figure 3.5 correspond to cases where the size of the training data was set to 1K and the non-censored proportion was set to 0.03 and 0.09, respectively. Increasing the percentage of non-censored observations, while keeping the size of the training data constant, reduces variance and yields better results due to less uncertainty in the data. Panel C and D in plot 3.5 refer to cases where the size of the training data was increased to 5K and the exact proportion was set to 0.03 and 0.09. It turns out that increasing the amount of data helps reduce the variance significantly. Additionally, it is worth mentioning that the proposed model's performance at hour 112 closely approximates the performance of the true underlying model, which achieves an upper bound of 0.88. In all four scenarios we considered, our proposed model was able to capture a higher percentage of crimes during typical patrol hours than the aoristic model. Furthermore, we conducted one-sided paired t-tests to assess the performance of the models specifically at hour 112. The results indicate that the null hypothesis of equal performance can be rejected at a significance level of five percent for all four simulation scenarios un-

der consideration. Hence, our estimated intensity provides a more precise temporal representation of crime intensity, offering the potential to capture a greater number of crimes if utilized effectively.

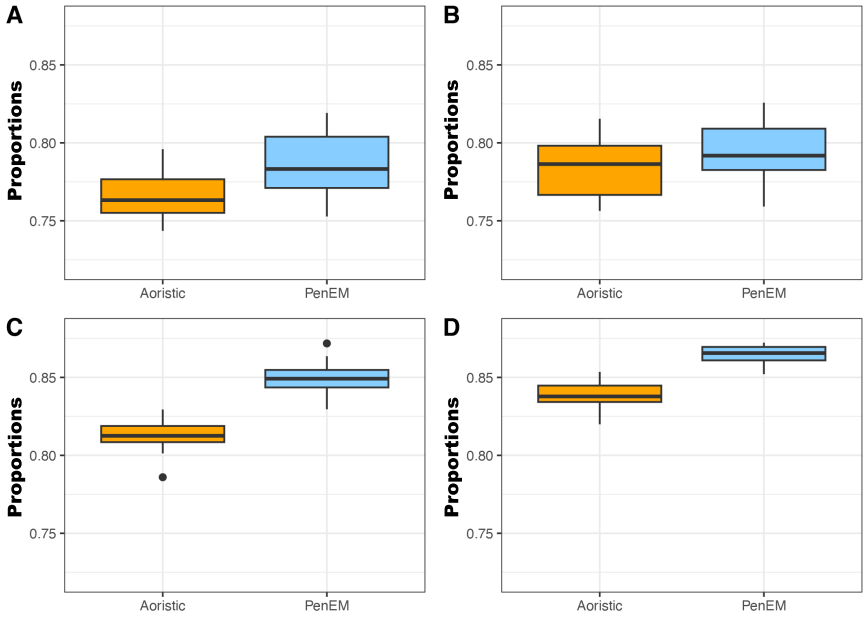

Figure 3.5: The box plots illustrate the cumulative percentage of crimes that occurred in the top 112 predicted hours. Panel A shows the results for a training size of 1K and a non-censored proportion of 0.03, while Panel B displays the corresponding data for a non-censored proportion of 0.09. Panel C and D exhibit the results for a training size of 5K, with non-censored proportions of 0.03 and 0.09, respectively.

3.4 Discussion

The aforementioned analysis demonstrates that our model is better at estimating crime intensities than existing approaches. It has the added benefit of being able to discover the day of week structure. Figure 3.2 shows the proportion of crimes in Cincinnati that would have occurred during a patrol if the optimal patrol scheduled from our model (PenEM) and the current state-of-art (Aoristic) was followed. The vertical reference line at 112 hours is Cincinnati's weekly number of patrol hours. If

112 hours were patrolled according to our model, 6.3 percent of more crimes could be potentially deterred. Also, the patrol plan could be designed with the help of day of week structures detected to minimize the disturbance to their routine plan. Figure 3.6 shows the estimated intensities for each of the three groups detected for the city of Cincinnati. Each group has different peak hours and time of day patterns. For Dallas, Figure 3.7 shows the estimated intensities for the two groups identified by our model.

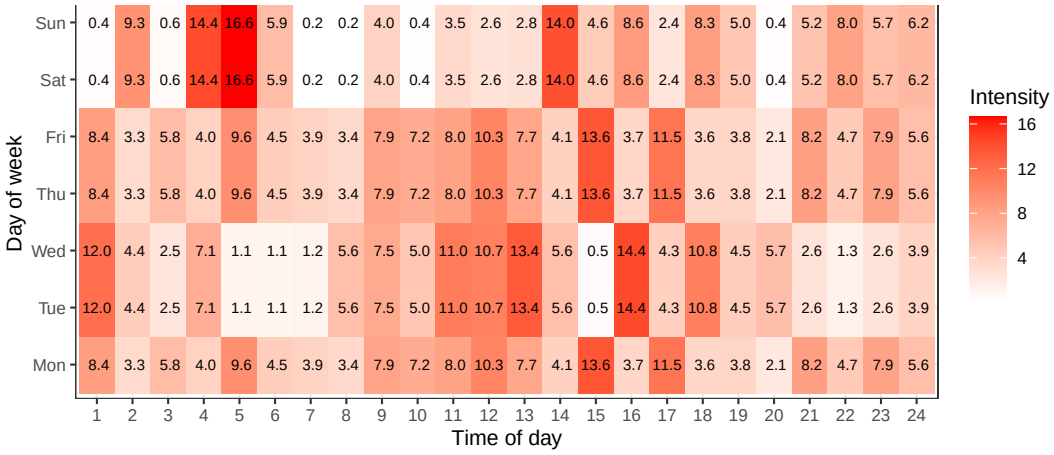

Figure 3.6: Heatmap of estimated intensity with grouping structure for each hour of the week for the city of Cincinnati

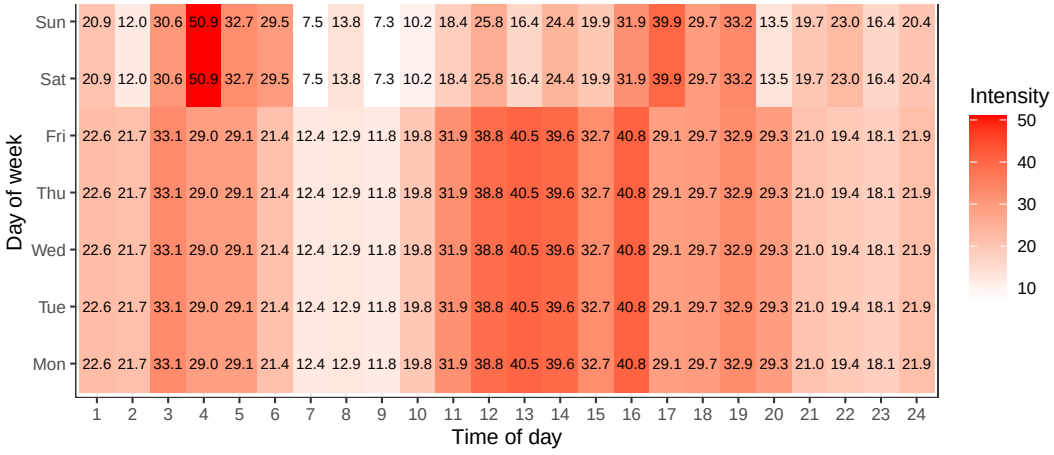

Figure 3.7: Heatmap of estimated intensity for with grouping structure each hour of the week for the city of Dallas

Depending on the resources each police station has, they could make a patrol plan that follows the intensity estimates given by the model or use a model to estimate intensities with any number of groups for day of week structure for ease of operation. The first option enables planners to create optimal schedules, but it may necessitate unique staffing and resource requirements for each day. On the other hand, the second option offers planners the flexibility to attain the best possible outcomes with a less complex schedule. For example, the best one-group model could be estimated to have a uniform patrol plan throughout the week. Figures 3.8 and 3.9 demonstrate cumulative probabilities for each day of the week versus hours patrolled. This type of information is particularly useful in the case where the enforcement agencies have extra resources to spend on patrolling and have to decide what day and time to patrol. The time and day could be picked based on the largest cumulative probability gain. These findings align well with the strategic targeting of the smart policing initiative by allowing law enforcement agencies to focus on the times with the highest expected occurrence of crimes.

Besides ARC analysis which demonstrates the gains when the proposed method is used for resource planning, we also investigated the performance of the intensity estimates. In our simulations, we have noticed that the estimated intensities generated by our model closely align with the actual underlying intensities used in creating synthetic data. In the scenario with 1,000 observations, the estimated intensities, on average, deviate from the true intensity by 10.76 percent. In the case of 5,000 observations, the average deviation of the estimated intensity from the true intensity is 6.23 percent. We observed a 30% performance improvement over the competing aoristic methods.

3.5 Conclusion

This article introduces our approach to estimating crime intensity from interval-censored data. Our model could serve three connected purposes with different levels of flexibility: (i) produce accurate intensity estimates and associated optimal patrol plan for the maximum crime reduction; (ii) improve existing patrol plan with the understanding of the discovery of day of week cluster, (iii) support better decision making for extra resources available. The proposed model was applied to both data gathered from the city of Cincinnati and the city of Dallas. Day of week structure was detected for each city and suggestions to modify the patrol plan based on the result were also made. Also, it turns out that the proposed model, if utilized, could deter more crimes by accurately estimating peak hours in crime occurrence (Braga, Papachristos, and Hureau, 2008) on hold-out data.

The proposed model achieves an accurate estimation of intensity by leveraging a structured statistical approach and specially designed penalties. The resulting penalized likelihood allows for a better estimation of intensity by considering the intensity estimate of adjacent bins as well as bins in the same cluster. Given the application setting, our proposed model assumes a piece-wise constant intensity for all discretized bins. However, the model could be easily extended to a more flexible model, such as penalized splines (Eilers and Marx, 1996; Cai and Betensky, 2003). The simulation study and real data analysis demonstrate the predictive performance compared with the competing method and its usefulness in the design of patrol for practical use. The performance of our model is affected by the number and length of the censored observations. Although our model effectively utilizes censored observations, the variance of the model will be proportional to the severity of censoring. While we developed the model for crime data, it can be modified to address the modeling of other interval

censored data.

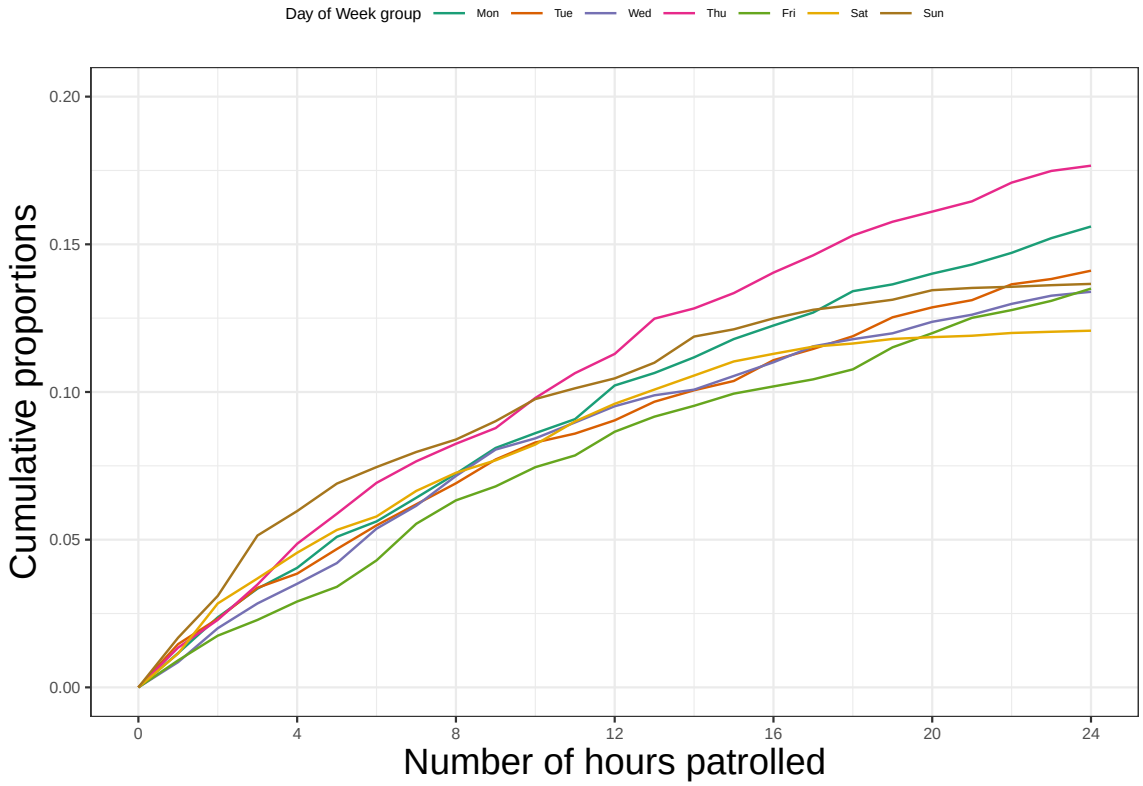

Figure 3.8: Cumulative proportions versus number of hours for each day of week for the city of Cincinnati.

3.6 Appendix

Let $P(X_i)$ represent the likelihood for event i. Following Kim, 2003, the density is used for the uncensored events and probability for the interval censored events:

$$P(X_i) = \begin{cases} f(t_i) & \text{uncensored events} \\ \Pr(T \in X_i) = \int_{X_i} f(t)dt & \text{censored events} \end{cases}$$

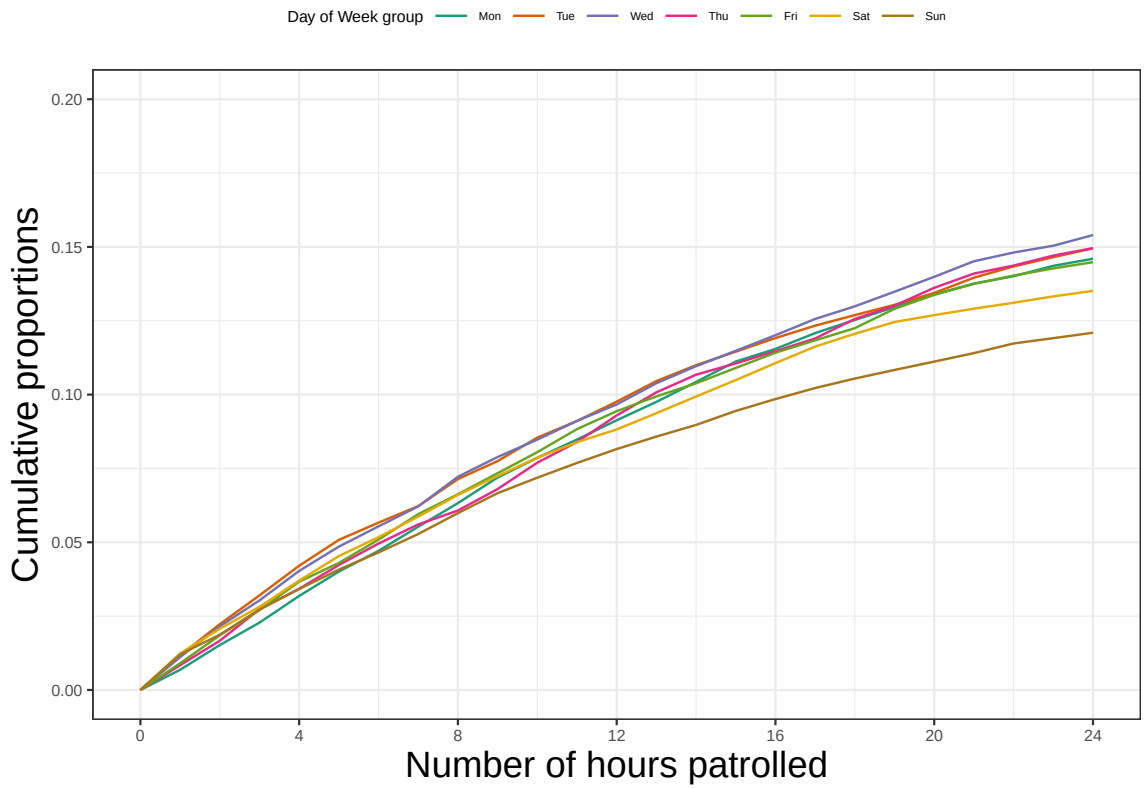

Figure 3.9: Cumulative proportions versus number of hours for each day of week for the city of Dallas.

The *observed* Poisson likelihood is

$$\Pr(X_1, X_2, \ldots, X_n) = \left(\prod_{i=1}^{n} P(X_i) \right) \Pr(N = n)$$

which combines the event likelihoods with the likelihood of observing n total events.

The *complete* likelihood is

$$\Pr((X_1, Z_1), (X_2, Z_2), \ldots, (X_n, Z_n)) = \left(\prod_{i=1}^{n} P(X_i, Z_i) \right) \Pr(N = n)$$

where $Z_i = [Z_{i1}, Z_{i2}, \ldots, Z_{iJ}]$ is a latent vector with Z_{ij} taking a value of 1 if the true event time t_i falls into bin j and 0 otherwise.

The likelihood for the latent vectors is

$$P(Z_i) = \frac{\prod_j \lambda_j^{Z_{ij}}}{\sum_j \lambda_j}$$

and the conditional likelihood of the observed event is

$$P(X_i \mid Z_i) = \prod_j w_{ij}^{Z_{ij}}$$

where w_{ij} is the proportion of bin j that is contained in interval X_i as defined in (3.1). The complete log-likelihood for event i can be written

$$\begin{aligned}\log P(X_i, Z_i) &= \log P(Z_i) + \log P(X_i \mid Z_i) \\ &= \sum_j Z_{ij} \log \lambda_j - \log\left(\sum_j \lambda_j\right) + \sum_j Z_{ij} \log w_{ij}\end{aligned}$$

with the sum over all n events

$$\begin{aligned}\sum_{i=1}^{n} \log P(X_i, Z_i) &= \sum_j \left(\sum_i Z_{ij}\right) \log \lambda_j - n \log\left(\sum_j \lambda_j\right) + \sum_j \sum_i Z_{ij} \log w_{ij} \\ &= \sum_j y_j \log \lambda_j - n \log\left(\sum_j \lambda_j\right) + \sum_j \sum_i Z_{ij} \log w_{ij}\end{aligned}$$

where the last line uses the notation $y_j = \sum_{i=1}^{n} Z_{ij}$ as the number of events falling in bin j. The last term $\sum_j \sum_i Z_{ij} \log w_{ij}$ can be ignored because it doesn't involve any model parameters.

To complete the derivation we note that the likelihood corresponding to the total

number of events is

$$\log \Pr(N = n) = n \log \left(\sum_j \lambda_j \right) - \sum_j \lambda_j - \log n!$$

The $\log n!$ can be dropped since it doesn't involve any model parameters.

These equations are used to write the complete penalized log-likelihood in (3.5).

Chapter 4

Forecasting

The project outlined in Chapter 3 proves effective in resource allocation and patrol planning under the assumption that there is no alteration in the underlying mechanism driving crime occurrences. However, this assumption may not hold in the face of significant events that influence how offenders engage in criminal activities. For instance, the FBI's annual crime report indicates a notable increase of 7.1% in property crime in the year 2022 (Investigation, 2023). Hence, in this chapter, we introduce an analytical framework that incorporates historical censored data to generate multi-horizon forecasts, considering trends and seasonalities evident in the historical data. Our forecasting framework, inspired by the sucecss of deep-learning based solutions in a variety of areas Vaswani et al., 2017; Mostafavi and Porter, 2021, produces probabilistic multi-horizon forecasts. It operates as a unified model for all interconnected time series, mitigating challenges associated with insufficient data in certain time series. The simulation results illustrate the model's capability to accurately capture trend changes and seasonality, yielding reliable forecasts. Furthermore, the model was evaluated using real data collected in Dallas.

In contrast to existing methods, the DeepCensored framework exhibits notable advantages, as outlined below:

- **Unified Model for Related Time Series**: The DeepCensored framework

stands out by generating a singular model capable of handling all interconnected time series. This consolidated approach streamlines the modeling process and promotes a cohesive understanding of diverse data sets.

- **Strength Borrowing**: One distinctive feature is the model's ability to leverage strength from related series, particularly in scenarios where certain time series lack sufficient observations.

- **Uncertainty Quantification**: DeepCensored adopts a probabilistic forecasting approach, enabling the quantification of uncertainty in predictions. This not only provides insights into the model's confidence levels but also enhances decision-making processes by offering a nuanced understanding of potential outcomes.

DeepCensored: Deep-Learning Based Probabilistic Forecasting Framework for Censored Data

Abstract Time series forecasting is an essential field where future values are predicted based on past observations, with applications ranging from demand forecasting to electricity usage prediction. While earlier research has mainly centered on real-valued data, our focus lies in forecasting count data, which poses additional challenges due to the inherent constraints on count values. What complicates the problem at hand even further is that certain events, such as crime, involve data that is not precisely observed, bringing an extra layer of complexity. It is crucial to develop accurate and robust forecasting algorithms in the presence of such data, providing valuable insights for resource allocation in crime deterrence for law enforcement agencies. In this paper, we propose a deep learning framework combined with the Expectation-Maximization (E-M) algorithm to enable probabilistic forecasting with censored data. Through both simulation and real data analysis, our proposed methods demonstrate superior performance compared to existing approaches.

Keywords Censoring, EM, Probabilistic Forecasting

4.1 Introduction

Crimes incur significant losses to both the victims and the entire society, negatively affecting the quality of life of people and the stability of society. The estimated annual cost of crimes in the U.S., including direct and indirect costs, amounts to $\$4.71-\$5.76$ trillion U.S. dollars Anderson, 2021. According to the Bureau of Justice Statistics, the median dollar value of financial loss due to burglary increased 54% from 1994 to 2011 Walters et al., 2013. Besides financial loss, research has shown that crime

victimization has implications for individual health and well-being Tan and Haining, 2016. Therefore, it's essential to have effective crime control measures to minimize the impacts of crime.

In an effort to reduce crime, the Bureau of Justice of Assistance (BJA) proposed the smart policing initiative (SPI) BJA, 2022 to support law enforcement agencies in building evidence-based, data-driven law enforcement tactics. The goal of the SPI is to identify strategic approaches that are effective in crime prevention and reduction. In the paper, we aim to develop a model that bridges the gap in current practices of crime forecasting in the spirit of the initiative.

Previous research has established that quantitative methods can play a beneficial role in providing insights into crime deterrence. A randomized block design was implemented to assess the impact of foot patrol on crime reduction in Philadelphia, with the results indicating a substantial decrease in crimes within the treatment areas after 12 weeks Ratcliffe et al., 2011. Additionally, randomized control trials were conducted in both Los Angeles and Kent to evaluate the effectiveness of predictive policing algorithms in comparison to hotspot maps given by crime analysts. The experiments demonstrated that employing predictive algorithms for determining policing patrols resulted in a significant reduction in crime Mohler et al., 2015.

The process of pinpointing hotspots empowers law enforcement agencies by allowing them to focus their efforts on specific small areas responsible for a substantial portion of crimes. However, an effective strategy must also include the identification of when these crimes are most likely to occur. The temporal aspect of the strategy plays a pivotal role in determining the optimal times of day for patrols, thereby achieving the greatest reduction in crime while working within realistic staffing constraints.

Forecasts given by predictive models are essential to the development of efficient and effective policing strategies, as SPI stated, allowing agencies to focus on a small percentage of people and places that account for most crimes Yu et al., 2011; Hunt, 2019; Rummens, Hardyns, and Pauwels, 2017. Our proposed model produces accurate multi-horizon forecasts, enabling the identification of times with high occurrences of crimes and facilitating more effective crime reduction strategies. Anticipating when crimes might occur is complicated by the presence of interval-censored data. Interval-censored data refers to situations where the exact event time is only known to fall within a specific interval, rather than being observed precisely Zhang and Sun, 2010. This type of data is prevalent in the field of criminology, especially in cases where there are no victims present when crimes occur. Forecasting with interval-censored data poses a significant challenge due to the heightened uncertainty involved. Previous research has predominantly focused on forecasting problems using exact data. However, certain crimes such as burglary and theft make it impossible to precisely record when these incidents occur. Prior research proposed the Aoristic method that individually addresses each event without taking into account variations in crime intensity at different times, as highlighted in references Ratcliffe, 2000; Camacho-Collados and Liberatore, 2015. This method involves assigning a partial count to each pre-defined bin that the time interval encompasses. In addition to its individual event focus, this method isn't directly applicable for forecasting purposes, as it assumes that future crime intensity remains the same as observed in the past. Given the frequency of such crimes, there is a pressing need to develop methods capable of handling interval-censored data while addressing the aforementioned limitations. These advanced methods have the potential to offer valuable insights for optimizing law enforcement patrols and strategies.

The key contribution of the work is to adapt a deep-learning-based probabilistic forecasting framework to make forecasts from historic crime data that are partially interval censored. Our proposed model demonstrates superior performance, surpassing the competing method by a 50% reduction in the Mean Absolute Error (MAE) of forecasts when tested on realistic simulations. Furthermore, our model exhibits the capability to detect emerging crime trends, thereby providing timely support to address and mitigate these evolving criminal activities.

The structure of this paper is organized as follows. Section 4.2 provides a review of the technical background, encompassing censored data modeling and the deep learning architecture utilized in this study. In Section 4.3, we introduce our deep learning framework tailored for handling censored data. Section 4.4 introduces how parameters are estimated. Section 4.5 presents empirical evidence of the effectiveness of our proposed approach to generating forecasts with highly uncertain censored data. Section 4.6 delves into a discussion of our findings within the context of providing support for public decision-making. Finally, in Section 4.7, we conclude the paper by summarizing the major findings and contributions.

4.2 Background

4.2.1 Censored data modeling

Censored data arises when the occurrence of an event of interest is only known to have taken place within a specific interval. These kind of events are commonly encountered in various fields such as clinical research, finance, and sociology Halling and Hayden, 2006; Haibe-Kains et al., 2008; Guo, 1993; Tian and Porter, 2022a. Censored data can arise in clinical trials, particularly when patients are subject to

scheduled follow-up visits. In such cases, if the event of interest occurs, its timing is only known to fall within the interval between two consecutive visits. In another scenario, interval censored data is encountered in specific crime types, such as car theft and burglary. In these cases, the absence of on-site victims results in the reporting of intervals defined by the earliest and latest times when the crime could have been committed Ratcliffe, 2000. The increased uncertainty inherent in such data introduces complexities into its modeling. Improper handling of such data could potentially result in biased conclusions. As a consequence, diverse methodologies and frameworks have been explored to address the challenges posed by censored data. For example, a penalized EM framework is introduced for estimating crime rate intensity in the presence of censored data Tian, in review. Considerable research effort has been directed towards time-to-event analysis, which traditionally originated from the analysis of right-censored data. This method has been extended to encompass interval-censored data analysis, aiming to derive the cumulative distribution function, as proposed in Peto, 1973; Turnbull, 1976. A substantial body of literature is dedicated to regression analysis techniques tailored for interval-censored data. Much of this work has concentrated on the proportional hazard model Cox, 1972; Finkelstein, 1986; Huang and Wellner, 1997; Kooperberg and Clarkson, 1997; Cai and Betensky, 2003.

Event forecasting involves making predictions about future events based on historical data observations. The methods discussed earlier are ill-suited for addressing this particular challenge, as many of them are oriented toward density estimation and cannot handle count forecasting naturally. In this context, deep learning has gained considerable traction for time series forecasting involving events with recorded timestamps. Notably, the DeepAR framework has been introduced as a deep-learning-based probabilistic forecasting approach capable of generating forecasts for multiple time series

Salinas et al., 2020. Moreover, Temporal Fusion Transformers have harnessed attention mechanisms to capture long-term dependencies, resulting in accurate multi-horizon forecasts Lim et al., 2021. Deep-Learning based forecasting frameworks have already been successfully applied to solve a wide range of problems. A deep learning model has been devised to predict the citation count of academic papers by leveraging bibliometric features and metadata Ma et al., 2021. In the realm of genomic selection, a Poisson deep neural network has been applied to address the genomic selection problem Montesinos-Lopez et al., 2021. In this paper, our focus is on tackling the challenge of making multi-step predictions for crime incidences in pre-defined bins, while accounting for the presence of interval-censored data.

4.2.2 Deep Forecasting

DeepAR

DeepAR represents a state-of-the-art deep-learning forecasting framework that harnesses neural networks to effectively manage long-term dependencies in forecasting Salinas et al., 2020. It distinguishes itself as a powerful competitor due to its capacity to handle a multitude of interconnected time series and produce probabilistic forecasts. The concurrent modeling of multiple related time series is particularly advantageous, especially in the realm of crime modeling. Urban areas commonly employ a hierarchical organizational structure for law enforcement purposes. Cities are typically partitioned into distinct districts, each falling under the purview of its respective police department. To optimize resource allocation and management, these districts are further subdivided into finer units referred to as beats. Information is then recorded and reported at this more localized beat level, resulting in the creation of multiple interrelated time series. The geographical proximity of these series con-

tributes to their interconnected nature. Consequently, the joint modeling approach enhances the development of a more robust predictive model, leveraging the shared geographical context for improved accuracy. At the core of the DeepAR model lies the LSTM module, which is a specialized type of Recurrent Neural Network (RNN). The recurrent nature of RNN, allowing it to make predictions based on both the current input and the information processed in the past, renders them a robust choice for forecasting tasks. However, RNNs can encounter issues like gradient vanishing or exploding when dealing with extended sequences. To address this limitation, the Long Short-Term Memory (LSTM) architecture offers a remedy through the incorporation of a gating mechanism within the memory cell Hochreiter and Schmidhuber, 1997. This mechanism enables LSTMs to effectively mitigate the challenges associated with long-range sequences. DeepAR employs two identical LSTM models, one serving as an encoder and the other as a decoder. The encoder is responsible for processing all the information within the prediction range, while the decoder utilizes the information obtained from the encoder to make forecasts. The capacity of LSTM networks to capture dependencies enables the framework to take into account all previous information when generating forecasts in the decoder. DeepAR, being a probabilistic model, directly models the parameters of the assumed distribution. Probabilistic forecasting holds significant importance in various scenarios as it facilitates decision-making while quantifying the associated uncertainty.

Attention

The attention mechanism has been a pivotal breakthrough in recent advancements in the field of deep learning, serving as a cornerstone for many groundbreaking frameworks Bahdanau, Cho, and Bengio, 2014; Vaswani et al., 2017; Lim et al., 2021; Fan et al., 2021. The attention mechanism operates in a manner akin to the human

brain, directing greater attention toward pertinent information and minimizing focus on irrelevant data. For example, achieving state-of-the-art performance in machine translation has been made possible by enabling the model to concentrate on the source sentence relevant for predicting the next word Bahdanau, Cho, and Bengio, 2014. Inspired by the achievements in natural language processing and computer vision, deep learning methods for forecasting have increasingly adopted attention mechanisms, resulting in notable improvements in performance Yi et al., 2023; Hu and Xiao, 2022.

4.3 Censored data forecasting using deep learning

4.3.1 Notation and Data

Let T denote the actual time that an event of interest (e.g., a crime) occurred. The event time can either be observed exactly ($T = t$) or only known to occur within an interval $T \subseteq [L, R]$. where L and R are the reported left and right endpoints. We discretize time into equally sized bins, and each bin has a width of one hour, by default. In our problem, we utilize historical data consisting of N partially interval-censored events from K locations to generate one-week-ahead probabilistic forecasts (equivalent to 168 hours). Mathematically, the model outputs

$$P(c_{k,t_0+s}|\mathcal{D}) \qquad \text{where } s = 1, 2, \ldots, 168 \tag{4.1}$$

where c_{k,t_0+s} denotes the number of events that will occur at time $t_0 + s$ and location k, with t_0 representing the start of the forecasting period. D encompasses all crime events used for model training, along with any known covariates employed in the forecasting process.

4.3.2 Model Parameters

As demonstrated in Equation 4.1, our objective is to model the probabilistic distribution of count data at a given hour. We employ the Poisson distribution to model count data, which is parameterized by the mean value.

$$
\begin{aligned}
P(c_{k,t}|\lambda) &= \frac{e^{-\lambda_{k,t}} \lambda_{k,t}^{c_{k,t}}}{c_{k,t}!} \\
\lambda_{k,t}(h_{k,t}) &= \log(1 + \exp(w^T h_{k,t} + b))
\end{aligned}
\tag{4.2}
$$

where $c_{k,t}$ represents the count for timestamp t and series k, and w and b denote the weight matrix and bias associated with the fully connected layer. $h_{k,t}$ represents the hidden states outputted by the LSTM layer, which is a function of the count at the previous step and covariates. Following Salinas et al., 2020, the covariates include the count observed at the previous timestamp, the time of day, the day of the week, and the month of the year. It is important to note that the count is not precisely observed for censored data, and we substitute the estimated count. The estimation process is detailed in Section 4.4. To ensure the non-negativity of the predicted parameters for Poisson distribution, the soft plus activation function is applied to the output of the fully connected layer.

4.3.3 Scale handling

Modeling multiple series from different beats enables the creation of a more robust model, as series without enough observations can draw valuable information from other series. Utilizing a single unified model also eliminates the need to train separate models for each series. However, the task of creating a shared model for all related series also poses challenges, particularly when these series exhibit significant

differences in scale. Therefore, it becomes essential to scale the output of the network depending on the specific series being modeled. As elucidated in Salinas et al., 2020, addressing this issue involves rescaling the network's output using a factor specific to each time series. Following this approach, the scaling is achieved by multiplying the network's output by a series dependent factor, denoted as ν_k. Consequently, the scaling factor is applied to the results of the soft plus activation and the formulation takes the shape $\lambda_{k,t} = \nu_k \log(1 + \exp(w^T h_{k,t} + b)$. The introduction of the time-series dependent factor ν_k serves to appropriately adjust the scaling of the output, aligning it with the specific range requirements of each time series. $\nu_k = 1 + \frac{1}{t_0} \sum_{t=1}^{t=t_0} c_{k,t}$ works well following the practice recommended in Salinas et al., 2020.

4.3.4 Training window selection

For traditional time series forecasting tasks, selecting the training time window is a straightforward process. The division of the training and forecasting period is straightforward as it's always desirable to include as much data in the modeling as possible. However, when dealing with censored data modeling, special consideration is required when choosing the training window. Censoring introduces a challenge, as events can not be recorded until they are reported to the police department. If all training data are incorporated into the model, it would entail modeling time steps with insufficient censored event coverage, leading to a significant drop in predicted values near the end of the training window. To address the issue mentioned above, instead of including all censored events, we exclude those censored events that fall within one week leading up to the last covered time step. Figure 4.1 illustrates how training data is selected during the modeling phase. The grey dots represent timesteps not included in the modeling because not all censored events have been reported in

that window. Including them could introduce bias for those timestamps.

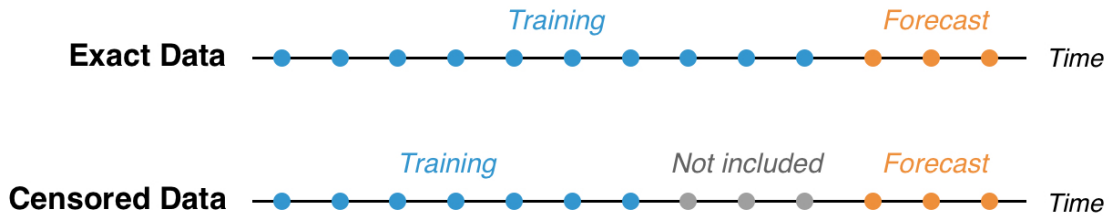

Figure 4.1: Training window selection for censored and exact data

4.3.5 Masked attention

The distinctive feature of LSTM lies in its gating mechanism, which grants the model the ability to discard information processed in preceding steps. Nonetheless, the introduction of the forget gate, while enhancing the network's adaptability, can inadvertently lead to the network disregarding information that was initially processed, particularly when dealing with lengthy sequences. This can significantly impede the model's capability to capture long-term dependencies within the data. Building upon the principles outlined in Vaswani et al., 2017, we introduced the use of an attention mechanism to allow the model to determine which time steps processed thus far are significant for predicting the current step. The attention mechanism is applied to the hidden states produced by the LSTM before they are fed into the fully connected layer for the prediction of distribution parameters at the current step. In our implementation, we utilized masked attention to ensure that forecasting is carried out exclusively using historical data, without incorporating future information.

4.3.6 Model Architecture

Figure 4.2 depicts the architecture of DeepAR with masked attention. The model takes into account both the response at the previous time step and the covariates mentioned earlier, generating hidden states as output. These hidden states are then passed through an attention layer, allowing the model to consider all available information from the past up to the respective timestamp and generate updated hidden states. These transformed hidden states are subsequently input into a fully connected layer, which produces the predicted values for the distribution parameters. Note that the paper Salinas et al., 2020 employed an encoder-decoder structure, and we adhered to their approach by maintaining the same structure for both the encoder and decoder in our implementation.

4.4 Parameter Estimation

The optimal parameters of the model are determined by maximizing the likelihood. However, censored data introduces complexities, rendering the count data inaccessible without appropriate transformation. Research has indicated that improper handling, such as midpoint imputation of censored data, may introduce bias into the estimates, as discussed in Turkson, Ayiah-Mensah, and Nimoh, 2021. To address this challenge, we combine the previously mentioned framework with the Expectation-Maximization (E-M) framework Dempster, Laird, and Rubin, 1977.

The E-M framework operates through an alternating process between the E-step, which entails calculating expectations of the likelihood based on the current model parameter values, and the M-step, which seeks to maximize the model parameters using the expectations computed. In the context of our problem, the E-step refers to the

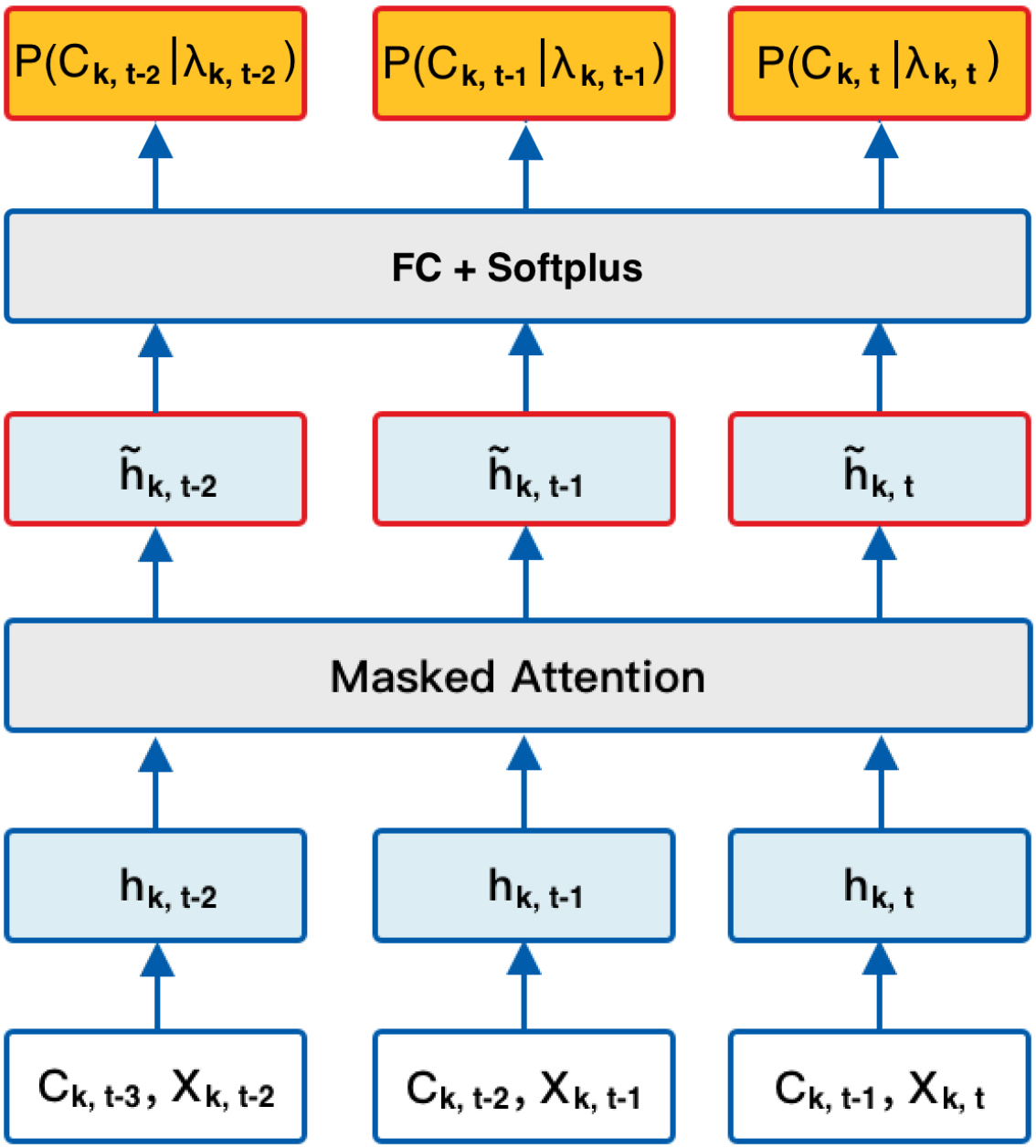

Figure 4.2: Architecture of the proposed deep learning forecasting framework with self-attention

calculation of the expected count $\hat{c}_{k,t}$ in the training range and the M-step optimizes the neural network-related parameters. The computation of the expected count of observations within each bin requires information on the proportion of bin j covered

in the ith interval, denoted by w_{ij}. The complete algorithm for the Expectation-Maximization (E-M) framework is outlined below.

Algorithm 4 EM algorithm for forecasting with interval censored events.

Input: Interval censored data, initial λ
Output: predicted count
while Convergence criteria is not met **do**
  for Each bin j **do**
    Calculate pseudo counts using equation $\hat{c}_{k,t} = \sum_{i=1}^{N} \frac{w_{it}\lambda_t}{\sum_J w_{ij}\lambda_j}$
  end for
  Maximize the λ using the DeepCensored
end while
Make forecast for the specified steps
return Predicted distribution parameters

Figure 4.3 demonstrates the general framework of our proposed model. Two components in the figure are associated with E-step and M-step described above. Note that the M-step pictured on the right is the DeepAR with attention model shown in Figure 4.2. The M-step could be replaced with any time series model that could make forecasts, but we chose DeepAR model for its ability to capture long-range dependency and trend changes.

4.5 Experiment results

To evaluate the model's forecasting capability and its capacity to capture changes in crime intensity patterns, simulations were conducted using synthetic data featuring related series. The simulation process begins by specifying the intensity and generating the events. The training window is set to ninety days. To create the synthetic intensity data, we started with one-week-long intensities, where each value corresponds to the intensity for each hour of the week. Additionally, we assume that there is an increasing trend in intensity after a certain point in time, driven by salient

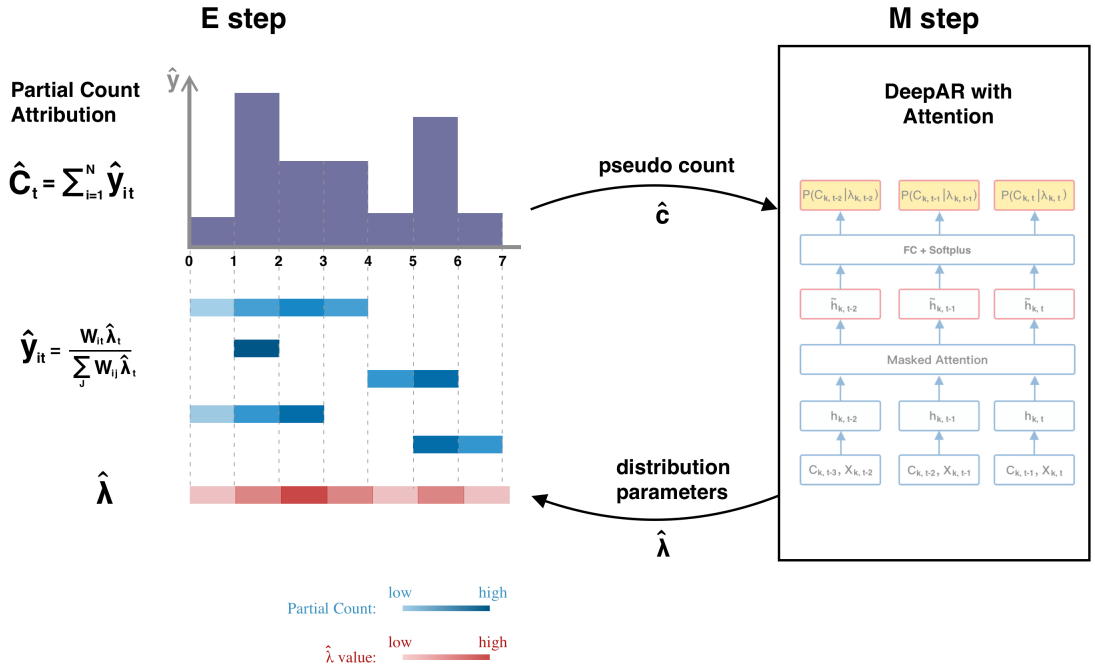

Figure 4.3: Framework of DeepCensored to handle interval censored data.

events such as the pandemic. To simulate the intensities over the specified length, we replicate the one-week intensity pattern to the desired length and generate intensity values that follow an increasing linear trend. This reflects the changing patterns, with intensities exhibiting an upward trend beyond a certain point. This approach allows us to incorporate the evolving nature of crime intensity and assess the model's ability to capture such changes accurately. Mathematically, the simulated intensities can be expressed as follows:

$$\lambda_{k,t} = Trend_{k,t} + Seasonality_{k,t}$$

where $Trend_{k,t} = constant_{series} * max(0, t - t_{chgpt})$. $constant_{series}$ is set to 0.01 for series A and 0.03 for series B, and the change point is fixed at 1200. In the simulations, the two series under consideration differ in the trend component while sharing the exact same seasonality component. With the intensity for each bin, we proceed to

simulate the number of observations falling into each bin by generating samples from the Poisson distribution. For events in bin j, we generate the exact time for those events from uniform distribution within the bin. To convert exact event times into censored data, the process begins by generating interval lengths from an exponential distribution $len \sim Exp(8)$, aligning with observations in real data. Censored data is then created by two steps: The left point is randomly selected to be $U(0, \text{len})$ from the true event time. The right endpoint is set accordingly based on the length of each interval and length between left endpoint and exact time created in the previous step.

We utilize three months' worth of data (90 days) to train the model to make forecasts of intensity for the subsequent 168 hours. For a single realization, Figure 4.4 illustrates the plot of fitted values provided our our DeepCensored model alongside the true intensities. Our model visually captures both the cyclic pattern and the increasing trend present in the training data, even in the presence of increased uncertainty due to censored data. This illustrates the effectiveness of our proposed model in discerning intricate patterns within the synthetic data. In addition, our model is a single, unified framework applicable to both related time series, enabling it to make accurate predictions even for series with diverse scales. To illustrate this, consider the intensities for series A, displayed in the upper panel of the plot, which are approximately half the magnitude of those for series B, shown in the lower panel.

The fitted time series serves as a demonstration of the proposed model's proficiency in identifying both trend changes and seasonality within the training data. However, of paramount importance for law enforcement agencies is the model's ability to provide precise forecasts for future time periods. As depicted in Figure 4.5, our model's forecasts closely align with the actual underlying mean. Furthermore, the model excels in accurately predicting both peaks and valleys in intensities. This capability

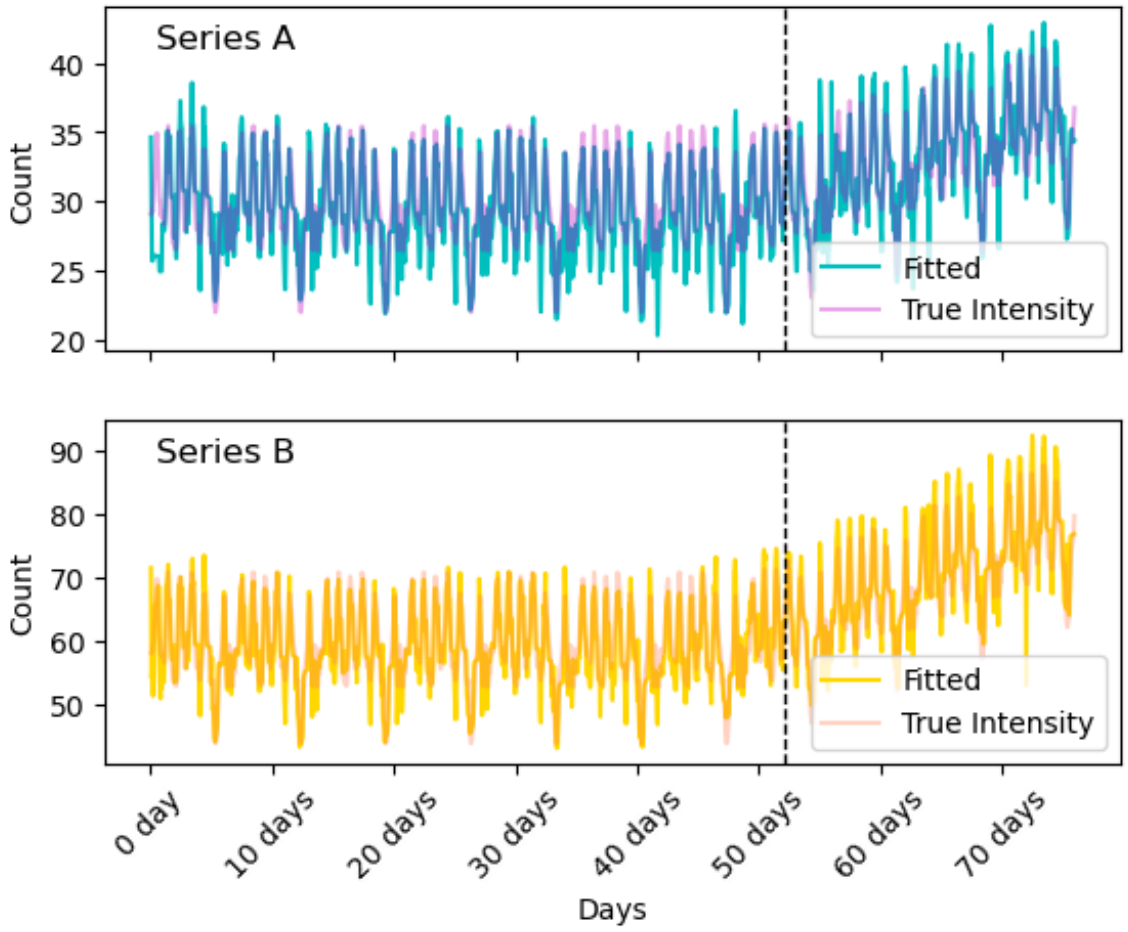

Figure 4.4: Predicted intensities versus true intensities for both series considered

empowers law enforcement agencies to design strategies that maximize efficiency and minimize resource wastage by deploying officers at the most appropriate times.

Figure 4.6 demonstrates the area reduction curve of different methods and ground truth for both series considered Tian, in review. This plot can be interpreted as the percentage of crimes that could be deterred by police patrolling a certain number of hours each week, assuming that patrols would follow the predicted peak hours. This metric bears resemblance to the hit ratio employed in the study in Kadar, Maculan, and Feuerriegel, 2019, albeit with a temporal perspective. To gauge the

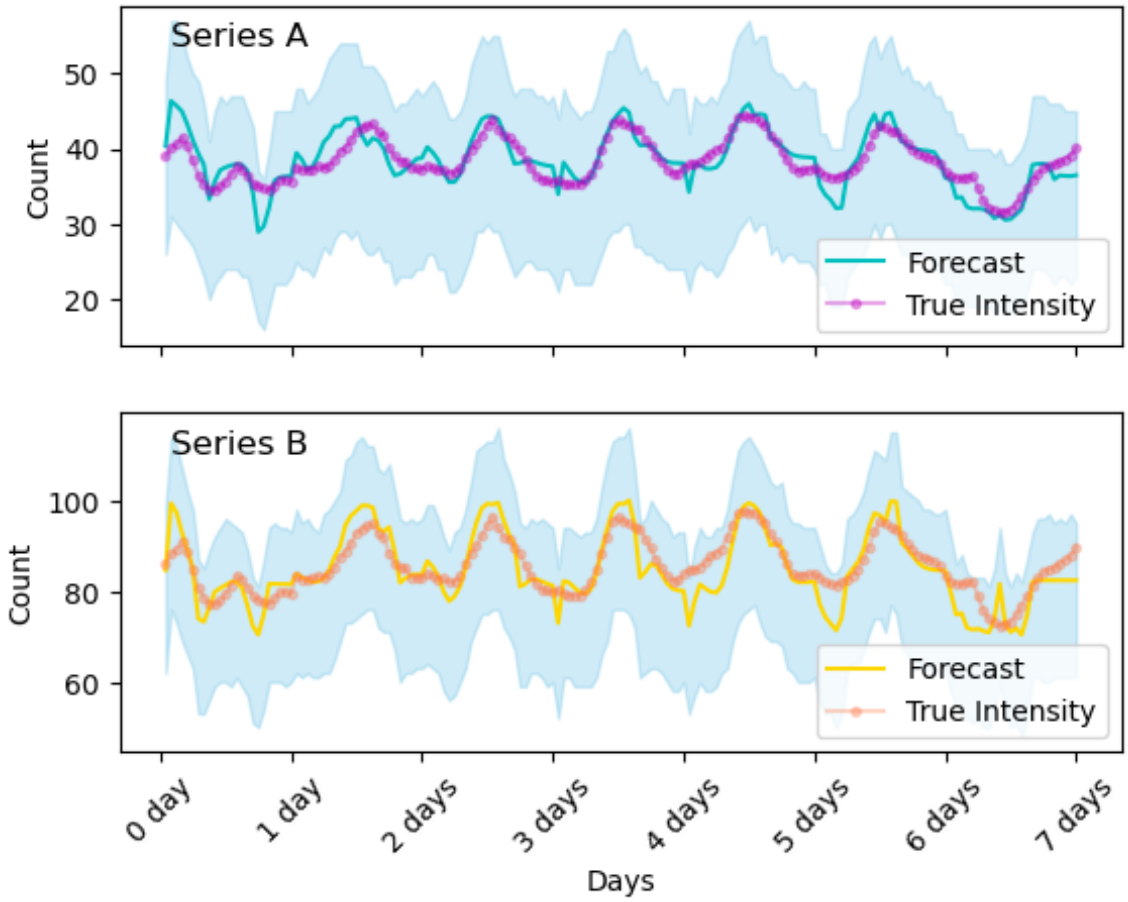

Figure 4.5: Forecast of intensities for one week period

effectiveness of our model, we compare its forecasting performance with that of the aoristic method Ratcliffe, 2000. The plot demonstrates that, across two time series in the simulation, the proposed model consistently outperforms the competing method on a global scale. This is evident from the larger Area Under the Curve (AUC), which closely approaches the upper bound established by the ground truth. Furthermore, the cumulative proportion at any given number of hours is consistently higher for DeepCensored compared to the Aoristic approach. This observation implies that the adoption of the proposed method could potentially result in a greater deterrence of crime, as it is associated with a higher preventive impact.

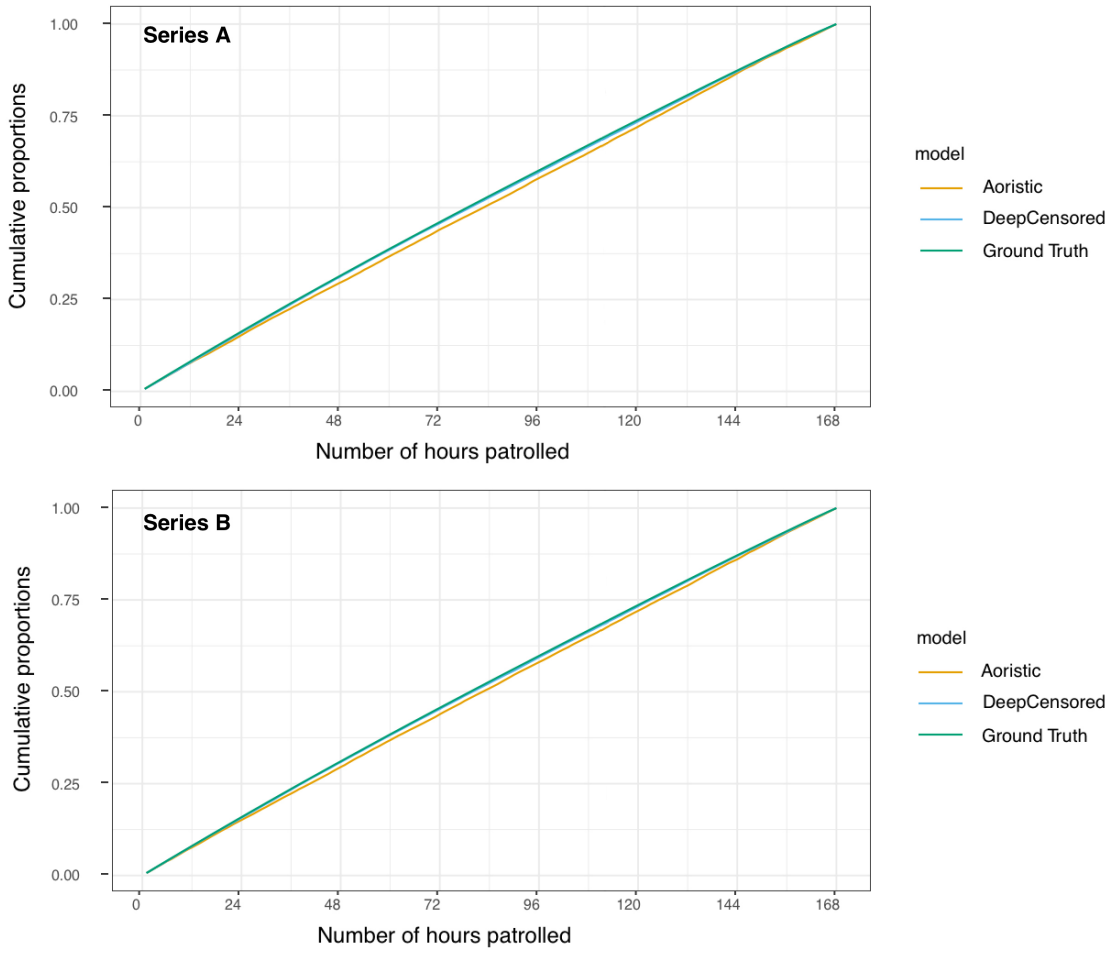

Figure 4.6: ARC plot of various methods under consideration compared with the ground truth.

In addition to the visual assessment provided by the plotted forecasts, we have employed numerical metrics to evaluate the performance of our proposed model on synthetic data. Table 4.1 summarizes the models' performance across five realizations and numbers after ± denote the standard error. This demonstrates the superior performance of our proposed approach compared to the competing methods we have consideredMelard and Pasteels, 2000; Ratcliffe, 2000. Note that for methods like Exponential smoothing and AutoARIMA that require exact observations, we used the resulting count obtained by applying the aoristic method. The non-competitive re-

Series	DeepCensored	Aoristic	Exponential Smoothing	AutoARIMA
A	$2.94(6.70\%) \pm 0.58$	$5.18(13.5\%) \pm 0.39$	$4.68(10.6\%) \pm 0.32$	$6.82(17.8\%) \pm 0.43$
B	$6.04(6.69\%) \pm 0.52$	$9.17(10.7\%) \pm 0.30$	$12.79(14.16\%) \pm 0.26$	$19.72(23.1\%) \pm 0.37$

Table 4.1: MAE of different methods considered in the simulation.

sults also underscore the challenges that traditional methods encounter when handling interval-censored data using transformed data that treats each event individually in a forecasting scenario.

The proposed model was also applied to real crime data. We use data gathered from the city of Dallas in the year 2019. This dataset includes detailed information indicating the specific beat and division in which each crime occurred. In Dallas, each division encompasses a considerable number of beats; for instance, the Southwest division comprises 33 beats. While it is conceivable to construct individual models for each beat through independent time series analysis, this approach presents several drawbacks: 1) certain beats may lack sufficient data, hindering the development of a robust model; 2) the correlated relationships among beats are overlooked; and 3) creating a model for each beat entails repetitive and labor-intensive work. This is precisely where our proposed model comes into play, as it leverages relationships among similar time series to construct a unified model, thereby mitigating the aforementioned limitations.

We applied our model to analyze the data collected from the Southwest division during the period spanning June 1st to September 1st in the year 2019. Before modeling, we took out those observations with interval lengths longer than one week. The resulting dataset contains a total of 17,668 observations from 33 different beats/units, with an average interval length of 6.8 hours. Among these, 1415 observations, or eight percent of the total, are uncensored. We built a unified DeepCensored model encompassing

all beats and generated forecasts for the subsequent week after the training period. To assess the efficacy of the forecasts on censored observations, we employed the observed likelihood of the data for the following week where the forecasts were generated. The observed likelihood characterizes the probability of observing the censored data, considering the estimated parameters provided by various models. It is computed as follows:

$$\log L = \sum_{k=1}^{K} \sum_{i=1}^{N} \log \left(\frac{\sum_t \hat{\lambda}_{k,t} w_{it}}{\sum_{t=1}^{168} \hat{\lambda}_{k,t}} \right)$$

The evaluation dataset contains 1472 observations. Our model achieved better results than the competing aoristic method and yields an average observed likelihood of -5.11, outperforming the -5.18 produced by the competing method for all observations considered in the dataset. Additionally, it is noteworthy that the proposed model exhibits superior performance in beats with limited observations. For instance, Beat 425, with the least number of observations at 38, demonstrates a more substantial performance gain over the competing model compared to beats with greater observations. Specifically, the observed likelihood produced by the proposed model is -180.5, surpassing the -185.7 from the competing method in this beat.

4.6 Discussion

The analysis presented in the preceding section underscores the effectiveness of the proposed model relative to other competing methods in the realm of time series forecasting with censored data. Beyond accurate estimates, the forecasts and associated further analyses offer tangible real-world benefits, such as crime reduction and improved resource allocation. This, in turn, leads to enhanced efficiency and a reduction in resource wastage.

To enhance crime prevention and deterrence efforts, applying our proposed model to the dataset and generating time series of fitted values enables the detection of intricate trends and patterns. For instance, the emergence of an increasing trend in crime occurrences may warrant the implementation of targeted programs or strategies for effective crime reduction. Research conducted by the National Institute of Justice (NIJ) has indicated that well-designed policing strategies can indeed be effective in deterring crime Haskins et al., 2019. Identifying areas requiring such interventions facilitates timely and proactive measures.

The allocation of human resources has become increasingly crucial, especially in light of the downsizing challenges that certain police departments are currently grappling with. This issue has been a central focus of prior research in decision support system in proactive policing Tian, in review . Similar to the ARC curves in Figure 4.6, Figure 4.7 creates an ARC for each day of the week. This type of plot will be particularly useful in the scenario where the enforcement agencies have extra resources to spend on patrolling and have to decide what day and time to patrol. The time and day could be picked based on the largest cumulative probability gain. As an example, if resources permit more than 12 hours of patrols in beat A, allocating additional resources to Thursday could result in more significant potential crime reduction. Similarly, assigning extra resources to Wednesday yields a higher Return on Investment (ROI) in beat B. These findings align well with the strategic targeting of the smart policing initiative by allowing law enforcement agencies to focus on the times with the highest expected occurrence of crimes.

The proposed model has demonstrated its efficacy in making accurate forecasts of future counts. This capability allows the police department to design strategies based on precise probabilistic forecasts provided by the model. Our next research step

involves crafting the optimal strategy by taking into account both the model forecasts and all practical constraints that law enforcement agencies may encounter.

4.7 Conclusion

In this research, we introduce a deep-learning-based model for forecasting a specific type of incident occurrence, wherein the actual event time is only known to fall within a given interval. By applying the model to both synthetic data and real data, we demonstrate two key findings:

Firstly, our proposed model demonstrates the capability to model all related time series with a single unified framework, effectively capturing trends within the data even in the presence of interval-censored observations.

Secondly, the proposed model exhibits superior predictive performance, as measured by the Area-reduction curves and MAE, when compared to previously proposed methods. This enhanced forecasting ability, particularly for multi-horizon forecasts, empowers practical resource allocation and planning well in advance. Overall, our model is well-suited for a diverse range of problems involving interval-censored data analysis.

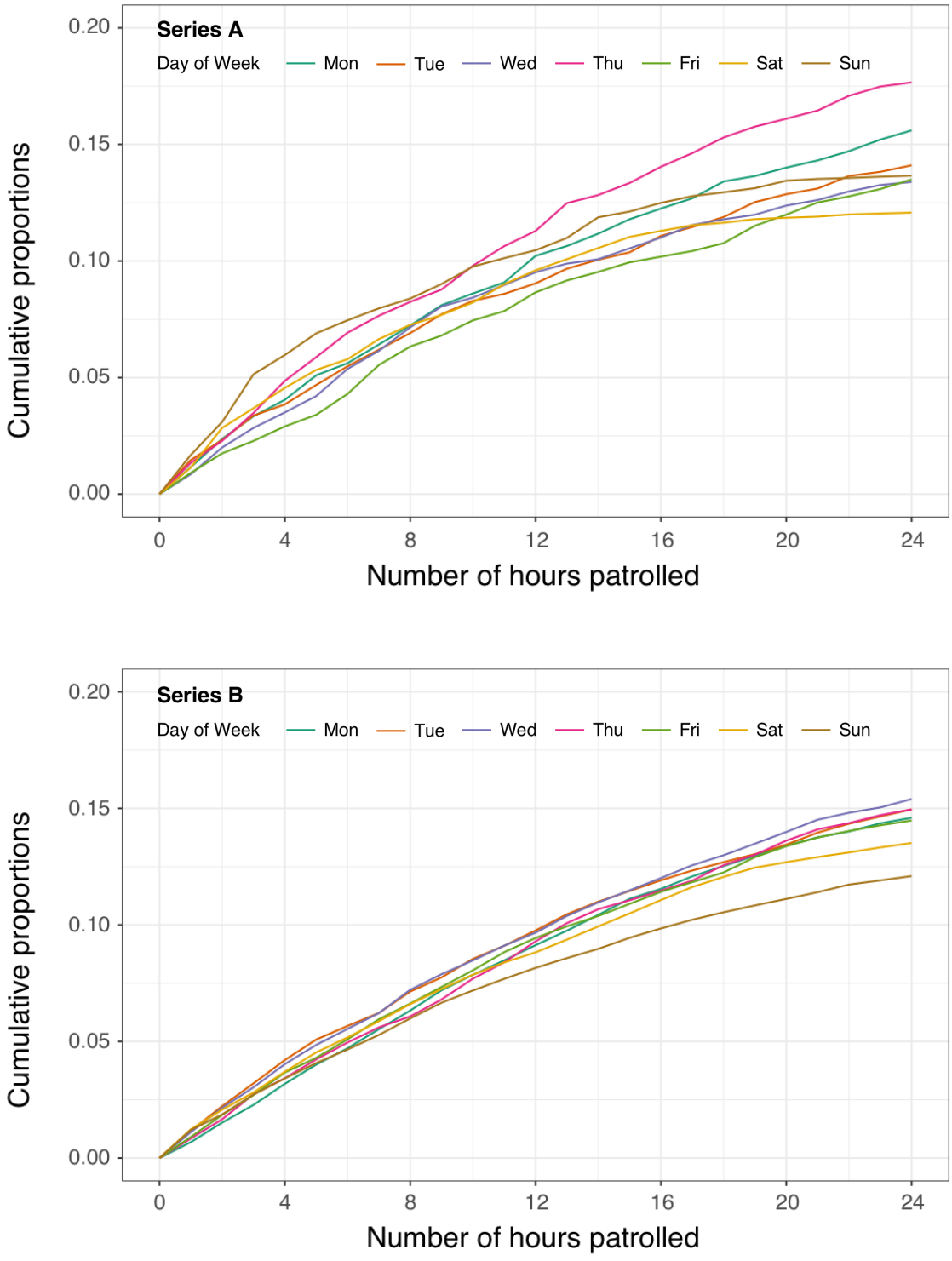

Figure 4.7: ARC for each day of the week.

Chapter 5

Conclusion

We initiated our discussion of modeling censored data by addressing the bias introduced due to the inappropriate handling of such data. We conducted a simulation study to underscore the bias associated with the straightforward mid-point imputation approach. The study confirmed the notion that the likelihood-based approach leads to smaller bias during inference because each censored event is processed in a way that considers all intensities of the corresponding bins it covers.

Throughout the dissertation, we addressed three crucial tasks in the presence of interval censoring. Change point detection (CPD) is a well-researched area, typically approached from various angles, with much of the research assuming that the data is precisely observed. In our project, we explored the utilization of the joinpoint model combined with the Expectation-Maximization (E-M) framework to detect change points. Additionally, we leveraged Bayesian model averaging to combine evidence and assign probabilities to each change point. We applied the proposed model to the 2020 Presidential approval rate and identified change points that are supported by various sources of evidence. Intensity estimation for crime data with interval censoring is an under-researched area, often addressed using data transformation methods. The oversmoothed intensity curve resulting from such methods could lead to resource wastage. In our work, we demonstrated a penalized Expectation-Maximization (E-

M) approach to intensity estimation, combined with hierarchical clustering. This approach produces an estimated intensity that is both realistic and has the potential to achieve optimal crime reduction. For example, this approach yields clustering results that group days of the week into clusters, where intensity estimates within the same group share identical values. Such results provide law enforcement agencies with flexibility in setting the number of different patrol plans they can afford, aiming to achieve the maximum level of crime reduction with available resources.

The estimated intensity provided by the previous project may work well for resource planning in a city where the seasonality remains consistent on a weekly basis. However, this may not hold if a significant event occurs, triggering a change in the underlying mechanism of crime. For instance, the COVID-19 pandemic altered the modus operandi(MO) of offenders due to the shift to remote work. The forecasting model can explicitly consider all historical information and make forecasts for future timesteps. We proposed a deep learning-based forecasting model that generates probabilistic forecasts in the presence of censored data. The model also offers estimates for all related time series simultaneously, instead of constructing separate models for each. The model generates multi-horizon forecasts, enabling the optimization of resources across multiple future steps.

Through a series of projects, we have demonstrated that machine learning/deep learning is superior to simple data transformation in terms of handling interval-censored data. As the next steps in research, some additional directions could be pursued. Our DeepCensored model leveraged a recurrent model for forecasting, capturing dependencies through time. However, it may face two potential issues: 1) The LSTM module alone might struggle with processing long-range dependencies, potentially relying on short-term memories for predictions. Addressing this could involve stor-

ing all hidden states and incorporating attention mechanisms, though this could be computationally and storage expensive. 2) The use of LSTM inhibits the possibility of parallelism in computation, which could be a concern when dealing with large amounts of data.

A potential solution could involve eliminating the recurrent module altogether and relying solely on a Transformer-like module with positional encoding to denote the temporal distance. Research has shown promising results with the replacement of LSTM Song et al., 2018; Lim et al., 2021, suggesting that this architecture might be beneficial in censored data prediction tasks. Moreover, our current approach primarily leverages tabular data for forecasting crime occurrences. Prior research has demonstrated the effectiveness of incorporating other sources of information in crime prediction. For instance, Twitter data has shown promise in providing additional signals for crime prediction tasks Chen, Cho, and Jang, 2015. Additionally, images such as Google Street View can offer valuable information about the environment for forecasting Kang and Kang, 2017. It is conceivable that the fusion of information from different modalities could lead to improved forecasting performance. Both directions show promise in terms of further improving the performance of modeling censored data.

Bibliography

Anderson, David A (2021). "The aggregate cost of crime in the United States". In: *The Journal of Law and Economics* 64.4, pp. 857–885.

Andresen, Martin A and Nick Malleson (2015). "Intra-week spatial-temporal patterns of crime". In: *Crime Science* 4.1, pp. 1–11.

Ashby, Matthew PJ and Kate J Bowers (2013). "A comparison of methods for temporal analysis of aoristic crime". In: *Crime Science* 2.1, p. 1.

Assareh, Hassan and Kerrie Mengersen (2012). "Change point estimation in monitoring survival time". In: *PLoS One* 7.3, e33630.

Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio (2014). "Neural machine translation by jointly learning to align and translate". In: *arXiv preprint arXiv:1409.0473*.

Benoit, William L, Glenn J Hansen, and Rebecca M Verser (2003). "A meta-analysis of the effects of viewing US presidential debates". In: *Communication Monographs* 70.4, pp. 335–350.

BJA (2022). *Smart Policing Initiative (SPI)*. https://bja.ojp.gov/program/smart-policing-initiative-spi/overview#core-spi-practices.

Braga, Anthony A, Andrew V Papachristos, and David M Hureau (2008). "Crime Prevention Research Review". In: *US Department of*.

Brunsdon, Chris and Jon Corcoran (2006). "Using circular statistics to analyse time patterns in crime incidence". In: *Computers, Environment and Urban Systems* 30.3, pp. 300–319.

Cai, Tianxi and Rebecca A Betensky (2003). "Hazard regression for interval-censored data with penalized spline". In: *Biometrics* 59.3, pp. 570–579.

Camacho-Collados, Miguel and Federico Liberatore (2015). "A decision support system for predictive police patrolling". In: *Decision support systems* 75, pp. 25–37.

Campbell, James E, Lynna L Cherry, and Kenneth A Wink (1992). "The convention bump". In: *American Politics Quarterly* 20.3, pp. 287–307.

Chandola, Varun and Ranga Raju Vatsavai (2010). "Scalable Time Series Change Detection for Biomass Monitoring Using Gaussian Process." In: *CIDU*, pp. 69–82.

Chen, Xinyu, Youngwoon Cho, and Suk Young Jang (2015). "Crime prediction using Twitter sentiment and weather". In: *2015 systems and information engineering design symposium*. IEEE, pp. 63–68.

Cohen, Lawrence E and Marcus Felson (1979). "Social change and crime rate trends: A routine activity approach". In: *American sociological review*, pp. 588–608.

Cohn, Ellen G and James Rotton (2003). "Even criminals take a holiday: Instrumental and expressive crimes on major and minor holidays". In: *Journal of Criminal Justice* 31.4, pp. 351–360.

Cox, David R (1972). "Regression models and life-tables". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 34.2, pp. 187–202.

Dehkordi, Zahra Fazeli, Mehdi Tazhibi, and Shadi Babazade (2014). "Application of joinpoint regression in determining breast cancer incidence rate change points by age and tumor characteristics in women aged 30–69 (years) and in Isfahan city from 2001 to 2010". In: *Journal of education and health promotion* 3.

Dempster, Arthur P, Nan M Laird, and Donald B Rubin (1977). "Maximum likelihood from incomplete data via the EM algorithm". In: *Journal of the royal statistical society: series B (methodological)* 39.1, pp. 1–22.

Department, Cincinnati Police (2023a). *PDI (police data initiative) crime incidents*. URL: https://data.cincinnati-oh.gov/safety/PDI-Police-Data-Initiative-Crime-Incidents/k59e-2pvf.

Department, Dallas Police (2023b). *Dallas Police Department Patch*. URL: https://dallaspolice.net/division/northeast/commandstaffandadmin.

— (2023c). *Police incidents: Dallas opendata*. URL: https://www.dallasopendata.com/Public-Safety/Police-Incidents/qv6i-rri7.

Dirick, Lore, Gerda Claeskens, and Bart Baesens (2017). "Time to default in credit scoring using survival analysis: a benchmark study". In: *Journal of the Operational Research Society* 68, pp. 652–665.

Druckman, James N and Justin W Holmes (2004). "Does presidential rhetoric matter? Priming and presidential approval". In: *Presidential Studies Quarterly* 34.4, pp. 755–778.

Eilers, Paul HC and Brian D Marx (1996). "Flexible smoothing with B-splines and penalties". In: *Statistical science* 11.2, pp. 89–121.

Evans, William N and Emily G Owens (2007). "COPS and Crime". In: *Journal of public Economics* 91.1-2, pp. 181–201.

Fan, Haoqi et al. (2021). "Multiscale vision transformers". In: *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6824–6835.

Finkelstein, Dianne M (1986). "A proportional hazards model for interval-censored failure time data". In: *Biometrics*, pp. 845–854.

Friedman, Michael et al. (1982). "Piecewise exponential models for survival data with covariates". In: *The Annals of Statistics* 10.1, pp. 101–113.

Ghosh, Pulak et al. (2009). "Semiparametric Bayesian approaches to joinpoint regression for population-based cancer survival data". In: *Computational statistics & data analysis* 53.12, pp. 4073–4082.

Goggins, William B and Dianne M Finkelstein (2000). "A proportional hazards model for multivariate interval-censored failure time data". In: *Biometrics* 56.3, pp. 940–943.

Google (2020). *Google Trends*. URL: https://trends.google.com/trends/.

Guo, Guang (1993). "Event-history analysis for left-truncated data". In: *Sociological methodology*, pp. 217–243.

Haibe-Kains, Benjamin et al. (2008). "A comparative study of survival models for breast cancer prognostication based on microarray data: does a single gene beat them all?" In: *Bioinformatics* 24.19, pp. 2200–2208.

Halling, Michael and Evelyn Hayden (2006). "Bank failure prediction: a two-step survival time approach". In: *Available at SSRN 904255*.

Haskins, Paul A et al. (2019). "Research will shape the future of proactive policing". In: *National Institute of Justice Journal Issue No* 281.

Hastie, Trevor et al. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer.

Hastie, Trevor J and Robert J Tibshirani (1990). *Generalized additive models*. Vol. 43. CRC press.

Hinkley, David V (1971). "Inference in two-phase regression". In: *Journal of the American Statistical Association* 66.336, pp. 736–743.

Hochreiter, Sepp and Jürgen Schmidhuber (1997). "Long short-term memory". In: *Neural computation* 9.8, pp. 1735–1780.

Hu, Yuntong and Fuyuan Xiao (2022). "Network self attention for forecasting time series". In: *Applied Soft Computing* 124, p. 109092.

Huang, Jian and Jon A Wellner (1997). "Interval censored survival data: a review of recent progress". In: *Proceedings of the First Seattle Symposium in Biostatistics*. Springer, pp. 123–169.

Hunt, Joel (2019). "From crime mapping to crime forecasting: The evolution of place-based policing". In: *National Institute of Justice. Available online: https://www. ncjrs. gov/pdffiles1/nij/252036. pdf (accessed on 1 December 2019)*.

Investigation, Federal Bureau of (2023). *FBI Releases 2022 Crime in the Nation Statistics*. https://www.fbi.gov/news/press-releases/fbi-releases-2022-crime-in-the-nation-statistics.

Jonge, Chad P Kiewiet de, Gary Langer, and Sofi Sinozich (2018). "Predicting state presidential election results using national tracking polls and multilevel regression with poststratification (MRP)". In: *Public Opinion Quarterly* 82.3, pp. 419–446.

Kadar, Cristina, Rudolf Maculan, and Stefan Feuerriegel (2019). "Public decision support for low population density areas: An imbalance-aware hyper-ensemble for spatio-temporal crime prediction". In: *Decision Support Systems* 119, pp. 107–117.

Kang, Hyeon-Woo and Hang-Bong Kang (2017). "Prediction of crime occurrence from multi-modal data using deep learning". In: *PloS one* 12.4, e0176244.

Kim, Hyune-Ju et al. (2000). "Permutation tests for joinpoint regression with applications to cancer rates". In: *Statistics in medicine* 19.3, pp. 335–351.

Kim, Jong S (2003). "Maximum likelihood estimation for the proportional hazards model with partly interval-censored data". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 65.2, pp. 489–502.

Kim, Seung-Jean et al. (2009). "\ell_1 trend filtering". In: *SIAM review* 51.2, pp. 339–360.

Koch, Andrew, Jiahao Tian, and Michael D Porter (2020). "Criminal Consistency and Distinctiveness". In: *2020 Systems and Information Engineering Design Symposium (SIEDS)*. IEEE, pp. 1–3.

Kooperberg, Charles and Douglas B Clarkson (1997). "Hazard regression with interval-censored data". In: *Biometrics*, pp. 1485–1494.

Lerman, PM (1980). "Fitting segmented regression models by grid search". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 29.1, pp. 77–84.

Lim, Bryan et al. (2021). "Temporal fusion transformers for interpretable multi-horizon time series forecasting". In: *International Journal of Forecasting* 37.4, pp. 1748–1764.

Lindsey, JC and L Ryan (1999). "Tutorial in biostatistics: methods for interval-censored data (vol 17, pg 219, 1998)". In: *STATISTICS IN MEDICINE* 18.7, pp. 890–890.

Lock, Kari and Andrew Gelman (2010). "Bayesian combination of state polls and election forecasts". In: *Political Analysis* 18.3, pp. 337–348.

Ma, Anqi et al. (2021). "A deep-learning based citation count prediction model with paper metadata semantic features". In: *Scientometrics* 126.8, pp. 6803–6823.

Martinez-Beneito, Miguel A, Gonzalo García-Donato, Diego Salmerón, et al. (2011). "A Bayesian joinpoint regression model with an unknown number of break-points". In: *The Annals of applied statistics* 5.3, pp. 2150–2168.

McAvoy, Gregory E (2006). "Stability and change: The time varying impact of economic and foreign policy evaluations on presidential approval". In: *Political Research Quarterly* 59.1, pp. 71–83.

Melard, Guy and J-M Pasteels (2000). "Automatic ARIMA modeling including interventions, using time series expert software". In: *International Journal of Forecasting* 16.4, pp. 497–508.

Mohler, George O et al. (2015). "Randomized controlled field trials of predictive policing". In: *Journal of the American statistical association* 110.512, pp. 1399–1411.

Montesinos-Lopez, Osval A et al. (2021). "Application of a Poisson deep neural network model for the prediction of count data in genome-based prediction". In: *The Plant Genome* 14.3, e20118.

Mostafavi, Moeen and Michael D Porter (2021). "How emoji and word embedding helps to unveil emotional transitions during online messaging". In: *2021 IEEE International Systems Conference (SysCon)*. IEEE, pp. 1–8.

Mostafavi, Moeen et al. (2021). "A tale of two metrics: Polling and financial contributions as a measure of performance". In: *2021 IEEE International Systems Conference (SysCon)*. IEEE, pp. 1–6.

Murray, Gregg R, Chris Riley, and Anthony Scime (2009). "Pre-election polling: Identifying likely voters using iterative expert data mining". In: *Public Opinion Quarterly* 73.1, pp. 159–171.

Narain, B (2004). "16. Survival analysis and the credit-granting decision". In: *Readings in Credit Scoring: Foundations, Developments, and Aims*, p. 235.

National Academies of Sciences, Engineering, and Medicine and others (2018). *Proactive policing: Effects on crime and communities*. National Academies Press.

Neath, Andrew A and Joseph E Cavanaugh (2012). "The Bayesian Information Criterion: background, derivation, and applications". In: *Wiley Interdisciplinary Reviews: Computational Statistics* 4.2, pp. 199–203.

Newby, M (1988). "Accelerated failure time models for reliability data analysis". In: *Reliability Engineering & System Safety* 20.3, pp. 187–197.

Peto, Richard (1973). "Experimental survival curves for interval-censored data". In: *Journal of the Royal Statistical Society: series C (Applied Statistics)* 22.1, pp. 86–91.

Porter, Michael D (2016). "A statistical approach to crime linkage". In: *The American Statistician* 70.2, pp. 152–165.

Prosser, Christopher et al. (2020). "Tremors but no youthquake: Measuring changes in the age and turnout gradients at the 2015 and 2017 British general elections". In: *Electoral Studies*, p. 102129.

Puzo, Quirino, Ping Qin, and Lars Mehlum (2016). "Long-term trends of suicide by choice of method in Norway: a joinpoint regression analysis of data from 1969 to 2012". In: *BMC public health* 16.1, pp. 1–9.

Qiu, Dongmei et al. (2009). "A Joinpoint regression analysis of long-term trends in cancer mortality in Japan (1958–2004)". In: *International journal of cancer* 124.2, pp. 443–448.

Ratcliffe, Jerry H (2000). "Aoristic analysis: the spatial interpretation of unspecific temporal events". In: *International journal of geographical information science* 14.7, pp. 669–679.

Ratcliffe, Jerry H and Michael J McCullagh (1998). "Aoristic crime analysis". In: *International Journal of Geographical Information Science* 12.7, pp. 751–764.

Ratcliffe, Jerry H et al. (2011). "The Philadelphia foot patrol experiment: A randomized controlled trial of police patrol effectiveness in violent crime hotspots". In: *Criminology* 49.3, pp. 795–831.

Ruggieri, Eric and Marcus Antonellis (2016). "An exact approach to Bayesian sequential change point detection". In: *Computational Statistics & Data Analysis* 97, pp. 71–86.

Rummens, Anneleen, Wim Hardyns, and Lieven Pauwels (2017). "The use of predictive analysis in spatiotemporal crime forecasting: Building and testing a model in an urban context". In: *Applied geography* 86, pp. 255–261.

Sakamoto, Yosiyuki, Makio Ishiguro, and Genshiro Kitagawa (1986). "Akaike information criterion statistics". In: *Dordrecht, The Netherlands: D. Reidel* 81.10.5555, p. 26853.

Salinas, David et al. (2020). "DeepAR: Probabilistic forecasting with autoregressive recurrent networks". In: *International Journal of Forecasting* 36.3, pp. 1181–1191.

Shaw, Daron R (1999). "The impact of news media favorability and candidate events in presidential campaigns". In: *Political Communication* 16.2, pp. 183–202.

Siddiqa, Hajra, Sajid Ali, and Ismail Shah (2021). "Most recent changepoint detection in censored panel data". In: *Computational Statistics* 36.1, pp. 515–540.

Silver, Nate (2020a). *How (un)popular is Donald Trump*. URL: https://projects.fivethirtyeight.com/trump-approval-ratings.

— (2020b). *Polls and policy FAQS*. URL: https://fivethirtyeight.com/features/polls-policy-and-faqs/.

Snoek, Jasper, Hugo Larochelle, and Ryan P Adams (2012). "Practical bayesian optimization of machine learning algorithms". In: *Advances in neural information processing systems* 25.

Song, Huan et al. (2018). "Attend and Diagnose: Clinical Time Series Analysis Using Attention Models". In: *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*. AAAI'18/IAAI'18/EAAI'18. New Orleans, Louisiana, USA: AAAI Press. ISBN: 978-1-57735-800-8.

Tan, Bien Aik et al. (2015). "Change-point detection for recursive Bayesian geoacoustic inversions". In: *The Journal of the Acoustical Society of America* 137.4, pp. 1962–1970.

Tan, Su-Yin and Robert Haining (2016). "Crime victimization and the implications for individual health and wellbeing: A Sheffield case study". In: *Social Science & Medicine* 167, pp. 128–139.

Tian, Jiahao (in review). "Time of week intensity estimation from interval censored data with applications to police patrol planning". In: *Journal of Applied Statistics*.

Tian, Jiahao and Michael D Porter (2022a). "Changing presidential approval: Detecting and understanding change points in interval censored polling data". In: *Stat* 11.1, e463.

— (2022b). "Changing presidential approval: Detecting and understanding change points in interval censored polling data". In: *Stat* 11.1, e463.

Tompson, Lisa and Michael Townsley (2010). "(Looking) Back to the Future: Using Space—Time Patterns to Better Predict the Location of Street Crime". In: *International journal of police science & management* 12.1, pp. 23–40.

Towers, Sherry et al. (2018). "Factors influencing temporal patterns in crime in a large American city: A predictive analytics perspective". In: *PLoS one* 13.10.

Turkson, Anthony Joe, Francis Ayiah-Mensah, and Vivian Nimoh (2021). "Handling censoring and censored data in survival analysis: a standalone systematic literature review". In: *International journal of mathematics and mathematical sciences* 2021, pp. 1–16.

Turnbull, Bruce W (1976). "The empirical distribution function with arbitrarily grouped, censored and truncated data". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 38.3, pp. 290–295.

Vaswani, Ashish et al. (2017). "Attention is all you need". In: *Advances in neural information processing systems* 30.

Walters, Jennifer Hardison et al. (2013). *Household Burglary, 1994-2011*. US Department of Justice, Office of Justice Programs, Bureau of Justice …

Walther, Daniel (2015). "Picking the winner(s): Forecasting elections in multiparty systems". In: *Electoral Studies* 40, pp. 1–13.

Wang, Bing, Xiaoguang Wang, and Lixin Song (2019). "Estimation in the single change-point hazard function for interval-censored data with a cure fraction". In: *Journal of Applied Statistics*.

Wei, Lee-Jen (1992). "The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis". In: *Statistics in medicine* 11.14-15, pp. 1871–1879.

Weisberg, Herbert, Jon A Krosnick, and Bruce D Bowen (1996). *An introduction to survey research, polling, and data analysis*. Sage.

Willer, Robb (2004). "The effects of government-issued terror warnings on presidential approval ratings". In: *Current research in social psychology* 10.1, pp. 1–12.

Wong, Martin CS et al. (2018). "The global epidemiology of bladder cancer: a join-point regression analysis of its incidence and mortality trends and projection". In: *Scientific reports* 8.1, pp. 1–12.

Yi, Shiyan et al. (2023). "A deep LSTM-CNN based on self-attention mechanism with input data reduction for short-term load forecasting". In: *IET Generation, Transmission & Distribution* 17.7, pp. 1538–1552.

Yu, Binbing et al. (2009). "Modelling population-based cancer survival trends by using join point models for grouped survival data". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 172.2, pp. 405–425.

Yu, Chung-Hsien et al. (2011). "Crime forecasting using data mining techniques". In: *2011 IEEE 11th international conference on data mining workshops*. IEEE, pp. 779–786.

Zhang, Zhigang and Jianguo Sun (2010). "Interval censoring". In: *Statistical methods in medical research* 19.1, pp. 53–70.