Nonparametric estimation of recursive point processes with application to mumps in Pennsylvania.

Andrew Kaplan[1] Junhyung Park[2] Frederic P. Schoenberg[3]

[1] Department of Statistics, 8142 Math-Science Building, University of California, Los Angeles, 90095-1554, U.S.A., amkaplan@ucla.edu

[2]Department of Statistics, 8142 Math-Science Building, University of California, Los Angeles, 90095-1554, U.S.A., junhyung@ucla.edu

[3]Department of Statistics, 8142 Math-Science Building, University of California, Los Angeles, 90095-1554, U.S.A., frederic@stat.ucla.edu

Abstract

"The self-exciting Hawkes point process model (Hawkes, 1971) has been used to describe and forecast communicable diseases. A variant of the Hawkes model, called the recursive model, was proposed by Schoenberg et al. (2019) and has been shown to fit well to various epidemic disease datasets. Unlike the Hawkes model, the recursive model allows the productivity to vary as the overall rate of incidence of the disease varies. Here, we extend the data-driven non-parametric E-M method of Marsan and Lengliné (2008) in order to fit the recursive model without assuming a particular functional form for the productivity. The non-parametric recursive model is trained to fit to weekly reported cases of mumps in Pennsylvania during the January 1970 - September 1990 time frame and then assessed using one week forecasts for the October 1990 - December 2001 time period. Both its training and predictive ability are evaluated compared to that of other candidate models, such as Hawkes and SVEILR (Susceptible, Vaccinated, Exposed, Infected, Lightly infected, Recovered) compartmental models."

1. Introduction

The self-exciting Hawkes point process model (Hawkes, 1971) has been commonly used to describe clustered phenomena, including earthquakes (Ogata 1988, 1998), crimes (Mohler et al., 2011), invasive species (Balderama et al., 2012) and contagious diseases (Meyer et al., 2012), (Meyer & Held, 2014). The model is specified by a conditional rate $\lambda(t)$ satisfying

$$\lambda(t) = \mu + K \int_0^t g(t - t^{'})dNt^{'}, \tag{1}$$

where μ is the background rate, the parameter K is called the productivity, and the density function g is called the *triggering function*. The simple Hawkes model makes the assumption that the productivity, K, which is the expected number of future infections directly transmitted by each infected individual, is constant.

Schoenberg et al. (2019) argued that the expected number of transmissions for a subject infected at time t may depend on the conditional rate at time t. For instance, early in the outbreak of a disease, when the prevalence of the disease is low, the rate of transmission may be much higher than at later times when the virus has already spread, due to differences in awareness, human mitigation efforts, and prior exposure of the population to the disease. Schoenberg et al. (2019) thus introduced the recursive model, which allows for changes in the productivity over time as the rate of incidence varies:

$$\lambda_t = \mu + \int_0^t H(\lambda_{t'})g(t - t^{'})dN_t^{'} \tag{2}$$

Note that here and in what follows, we alternate between the notation $\lambda(t)$ and λ_t which are interchangeable. As in the simple Hawkes model, the conditional intensity is dependent on the background rate $\mu(t)$ and the triggering function g governing temporal distribution of transmissions from an individual as a function of time since the individual's infection. However, the conditional rate λ in the recursive model also depends on the function H, which determines how the productivity varies with the conditional rate. If H is constant, then the recursive model is equivalent to the Hawkes model. Schoenberg et al. (2019) showed improved fit for the recursive model compared to that of the simple Hawkes model in describing known cases of Rocky Mountain Spotty Fever in California between 1960 and 2011. Wang (2019) showed improved fit of the recursive model relative to Hawkes models for Coccidioidomycosis in California, and Yang (2019) found improved fit of the recursive model over the Hawkes

and SEIR models using data on Pertussis in Nevada from 1940 to 2017.

The purpose of this paper is to estimate the recursive model non-parametrically, so that the forms of both functions g and H are purely data driven, allowing for more flexibility when fitting the recursive model to data, which in turn should result in more accurate forecasts. In Schoenberg et al. (2019), Wang (2019), and Yang (2019), both the Hawkes and recursive models were estimated parametrically after assuming a specific form for the functions g and H. The fitting is performed by extending the non-parametric estimation technique for the Hawkes model developed by Marsan and Lengliné (2008), but modified to take into account the additional recursive component, H. The non-parametric recursive model is evaluated using epidemiological data by comparing the accuracy of both the model fit and forecast to those of other point process models, as well as more widely used compartmental models such as the SEIR (Susceptible, Exposed, Infected, Recovered) compartmental model first introduced by Kermack and McKendrick (1927) and specifically its extension, the SVEILR model. The SVEILR model is a system of differential equations that can be adjusted to account for varying methods of transmission and rates of exposure, vaccination and recovery. To date, non-parametric methods for point process models have shown promise. The non-parametric version of the Hawkes model outperformed the generalized SEIR model in fitting and forecasting the spread of Ebola (Kelly & Park et. al., 2019), (Park et. al., 2019). In addition, it has been extended to applications in other areas such as renewal immigration (Wheatley et. al., 2014), online learning algorithms (Yang et. al., 2017) and finance (Kirchner & Bercher, 2018).

Of particular interest here is whether the proposed non-parametric version of the recursive model can outperform both the Hawkes and SVEILR models in its ability to fit epidemic data and to forecast future cases. We compare the models using 32 years of reported cases of mumps in Pennsylvania, fitting using training data from January 1970 to September 1990 and then assessing their fit using data from October 1990 to December 2001.

2. Methods

2.1 Nonparametric Hawkes Estimation Method of Marsan and Lengliné (2008)

Parametric Hawkes models are conventionally estimated by maximum likelihood estimation (MLE) (e.g. Ogata 1988), and the resulting estimates have desirable asymptotic properties such as consistency, asymptotic normality and efficiency (Ogata 1978). However, in practice often the likelihood function is quite flat around its maximum, and optimization methods may fail to converge or depend greatly on the choice of starting values (Schoenberg 2013). As an alternative, Veen and Schoenberg (2008) suggested a method based on the expectation-maximization (EM) algorithm to approximate maximum likelihood estimation. The key to such an approach is that the information containing which event triggers which later event is unknown. The probabilities of such triggerings, for each pair of points, are updated iteratively as parameter estimates are also updated.

In 2008, Marsan and Lengliné extended this EM-based method to the case of estimating the triggering function of a Hawkes process non-parametrically. Marsan and Lengliné (2008) assumed that the triggering function is a step function and proposed using the EM-based method to estimate the step heights as though they were parameters in a parametric model. In this setting, the expectation step of the E-M algorithm involves computing an updated matrix of triggering probabilities given the conditional intensity. In the maximization step, estimates of λ are recalculated given the updated matrix of triggering probabilities.

2.2 Non-Parametric Estimation of the Recursive Model

In fitting the recursive model (2), both the triggering function g and the productivity function H must be estimated. We propose an iterative EM-based procedure similar to that in Marsan and Lengliné (2008), where the productivity function H is estimated based on initial estimates of the background rate μ and the triggering function g. This estimated productivity function is then used to update estimates of the background rate μ and the triggering function g, and so on until a level of

convergence is reached.

A few details should be given before describing the algorithm. We suppose purely temporal point process data of the form $\tau_1, \tau_2, ..., \tau_n$, where n is the number of events observed in the time interval $[0, T]$. P is defined as a matrix of estimated probabilities such that P_{ij} is the estimated probability that event i was directly triggered by event j. This probability matrix is by definition lower-triangular since an event j can only trigger later events. Each diagonal entry P_{ii} represents the probability that infection i is a background event, not triggered by any previous event. The sum of any row of P must therefore equal 1, since each event must either be a background event or have been triggered by some prior event. As in Marsan and Lengliné, g is assumed to be a step function indicating the density of triggered points with time lags in some prespecified bins, and similarly we estimate H as a step function with predefined bins corresponding to intervals of the conditional rate, λ. We assume here that the background rate μ is constant and estimate this parameter μ as well.

After a quick initialization step (first E-step) where initial estimates of P_{ij} and H are defined, the following E-M algorithm can be run until a level of convergence or number of iterations is reached. Note that the initial value for H is a guess constant $\sum_i \sum_{j<i} \frac{\hat{P}_{ij}}{n}$.

Maximization Step:

Part 1: Estimate $\hat{g}$. First determine the interevent times, $\tau_i - \tau_j$, for all positive integers $j < i < n$. For each bin B_l, an interval of the real line containing some of the interevent times, set

$$\hat{g}_l = \frac{\sum_i \sum_{j<i} I(\tau_i - \tau_j \in B_l)(\hat{P}_{ij})}{\sum_i \sum_{j<i} w_l \hat{P}_{ij}} \tag{3}$$

where w_l is the width of bin B_l. The bins B_l need not be of equal width.
We suggest setting the bins so that they span the entire range of interevent times. To be consistent with the methodology used for the non-parametric version of the Hawkes model (Gordon, 2017), a loglinear approach using base 10 is applied here.

Part 2: Update $\hat{\mu}$.

$$\hat{\mu} = \frac{1}{T}\sum_{i}^{n} \hat{P}_{ii}. \tag{4}$$

Part 3: Update $\hat{\lambda}(\tau_i)$ using $\hat{\mu}$ and the most recently updated estimates of $\hat{g}$ and $\hat{H}$, letting

$$\hat{\lambda}(\tau_i) = \hat{\mu} + \sum_{j=1}^{i} \hat{H}_j \hat{g}(\tau_i - \tau_j). \tag{5}$$

Part 4: Update estimates of the productivities $H_i = H(\lambda(\tau_i))$.

For each bin C_k, an interval of the real line containing some of the values of $\hat{\lambda}(\tau_j)$, set

$$\hat{H}_k = \frac{\sum_{j=1}^{n-1}\sum_{i=j+1}^{n} I(\hat{\lambda}_j \in C_k)(\hat{P}_{ij})}{\sum_{j=1}^{n} I(\hat{\lambda}_j \in C_k)}, \tag{6}$$

provided $\sum_{j=1}^{n} I(\hat{\lambda}_j \in C_k) > 0$, and $\hat{H}_k = 0$ otherwise.

Thus the sum of the columns of $\hat{P}_{ij}$ where $\hat{\lambda}_j \in C_k$ are averaged.

Then, for any j such that $\hat{\lambda}(\tau_j) \in C_k$, set $\hat{H}(\lambda(\tau_j)) = \hat{H}_k$.

Expectation Step:

Update $\hat{P}$ using the values of $\hat{g}$, $\hat{H}$ and $\hat{\mu}$ obtained in the Maximization Step.

$$\hat{P}_{ij} = \frac{\hat{g}(\tau_i - \tau_j)\hat{H}_j}{\hat{\mu} + \sum_{k=1}^{i-1} \hat{g}(\tau_i - \tau_k)\hat{H}_k}. \tag{7}$$

Running all the parts of the expectation and maximization steps results in one complete iteration of the recursive Algorithm. While it may be customary to run until convergence, due to small changes from iteration to iteration in the large matrix P_{ij} of size $n \times n$, the algorithm often fails to converge completely. In such cases, we terminated the algorithm after 100 iterations.

2.3 Fitting and Forecasting Point Process Models

In the case of the parametric versions of the point process models, the values for the parameters in the triggering functions are estimated via maximum likelihood estimation

(Ogata, 1978). For the parametric Hawkes process,

$$\lambda(t) = \mu + K \sum_{i:\tau_i<t} g(t - \tau_i) \tag{8}$$

two common forms for the triggering function g are the Pareto

$$g(u) = (p-1)c^{p-1}\frac{1}{(u+c)^p} \tag{9}$$

and the exponential

$$g(u) = \beta e^{-\beta u}. \tag{10}$$

When fitting the parametric recursive model, the productivity K varies by

$$K_i = \frac{c}{(\lambda_i)^p} \tag{11}$$

To fit the non-parametric Hawkes model, we used the version of the Marsan and Lengliné (2008) E-M algorithm implemented by Gordon (2017), which is available in the *R* package *nphawkes*.

To assess and compare competing models, one may inspect the likelihood as well as the root mean squared error (RMSE) over each week. The latter is particularly relevant when making comparisons over a testing period for simulations of the models fit using separate training data. Because of occasional outliers in the simulations, we compared the trimmed mean of the simulated totals, with the top and bottom 10% of these values removed, to the observed weekly totals in the testing data. Another statistic useful for goodness of fit assessment is the scaled Stoyan-Grabarnik statistic (Stoyan & Grabarnik, 1991), $\sum_{i=1}^{n} \frac{1}{T\hat{\lambda}(\tau_i)}$, which should ideally be close to one if the estimates $\lambda(\tau_i)$ are close to the true conditonal rate, since the expected value of the statistic is easily seen to equal one by the martingale formula (Baddeley et al. 2005).

Another diagnostic tool useful for comparing the fit of point process models is superthinning (Clements et al. 2012). Given a constant b selected by the user, such as the mean of the estimated conditional intensities at the observed points (Gordon et al.

2015), one thins the data keeping each observed point τ_i independently with probability $\min\{1, \frac{b}{\hat{\lambda}(\tau_i)}\}$ and then superposing simulated points according to a Poisson process with rate $\max\{0, b - \hat{\lambda}(t)\}$. If the estimated conditional rate $\hat{\lambda}(t)$ is correct, then the resulting residual process is a stationary Poisson process of rate b. The superthinned residuals may thus be inspected for trend, clusters, gaps or other patterns as evidence of lack of fit of the model.

2.4 The SVEILR Model

Fitting the SVEILR model requires that parameters such as the rate of exposure, method of spread and human intervention must be prescribed. The model is designed as a closed system of differential equations where subjects go from one state, i.e. susceptible, infected or recovered to another at a rate determined by terms that have a simple interpretation. A single vaccine model is warranted since the original mumps vaccine was introduced before this period in 1967 and the second dose was not mandated until 1989 (CDC, 2019).

We followed Li et al. (2018), who developed a one vaccine model for mumps in mainland China, where the disease is still prevalent. Although there is sufficient availability of the second MMR dose, China only provides one free dose of MMR and there is no push from National Health and Family Planning Commission of the People's Republic of China (NHFPC) to require children to take a second dose (Li et al., 2018). The proposed expanded SVEILR (Susceptible, Vaccinated, Exposed, Infected, Light Infection, Recovered) model adds two necessary nodes to take into account the vaccinated state (V) as well as a lightly infected state (L) since not all cases are symptomatic (Li et al., 2018), (CDC, 2019).

In this analysis, the differential equation model used to model mumps in Pennsylvania during the 1970-1990 time period is quite similar to the one applied in Li et al. (2018). As in Li et al. (2018), the values optimized by minimizing the sum of squares include the transmission rate (β), waning immunity rate (λ), vaccine coverage of the susceptible/exposed (ϵ), proportion of people seeking medical advice (γ) and

initial proportions susceptible and vaccinated $(S(0), V(0))$.

The system of differential equations developed by (Li et al., 2018) for this SVEILR model is as follows:

$$\begin{aligned}
\frac{\partial S}{\partial t} &= \mu - p\mu E - \beta S(I+L) + \lambda V - (\epsilon + \mu)S \\
\frac{\partial V}{\partial t} &= \epsilon S + \epsilon_1 E - \lambda V - \kappa\beta V(I+L) - \mu V \\
\frac{\partial E}{\partial t} &= \beta S(I+L) + \rho\mu E + \kappa\beta V(I+L) - (\alpha + \epsilon_1 + \mu)E \\
\frac{\partial I}{\partial t} &= \alpha\gamma E - (\delta_1 + \mu)I \\
\frac{\partial L}{\partial t} &= \alpha(1-\gamma)E - (\delta_2 + \mu)L \\
\frac{\partial R}{\partial t} &= \delta_1 I + \delta_2 L - \mu R
\end{aligned}$$

The set of differential equations establish the rate that the numbers in each state change over time. In addition, each term in this model represents an effect which can be interpreted independently. As described in Li et al. (2018),

$$\begin{aligned}
\beta &= \text{transmission rate} \\
\lambda &= \text{waning immunity rate} \\
\epsilon &= \text{vaccine coverage of the susceptible} \\
\epsilon_1 &= \text{vaccine coverage of the exposed} \\
\kappa &= \text{invalid vaccination rate} \\
\alpha &= \text{rate moving from exposed to severe or mild infectious} \\
\gamma &= \text{proportion of the severe infections seeking medical advice} \\
\delta_1 &= \text{rate moving from severe infectious to recovered} \\
\delta_2 &= \text{rate moving from light infectious to recovered.}
\end{aligned}$$

The model was then applied to monthly mumps data acquired from the CDC of China. Since mumps spreads the most rapidly amongst children in school, Li et al. (2018) fit the SVEILR model for half year periods since peak cases tend to occur after major holiday breaks. Li et al. (2018) fit the SVEILR model using data from February 2009 to September 2014 by minimizing the sum of squares between the number of severe infections (state I) accounted for by the model and the actual case count and then forecast from October 2014 to

September 2015.

The only difference in our fitting of the model to mumps in Pennsylvania versus that of Li et al. (2018) is that peaks in mumps cases occur in Pennsylvania annually corresponding with the beginning of the academic year when schoolchildren return from break rather than twice per year as reported in Li et al. (2018). As a result, the SVEILR model is fit here using weekly data for each year separately beginning from the first week in October and ending in the last week in September.

Forecasting the SVEILR model was performed using standard exponential smoothing. Parameter estimates are obtained by taking a weighted average for all the years during the training period (Shmueli & Lichtendahl, 2016). To smooth, a reasonable value of the smoothing parameter α is chosen and for each of the six parameters of interest $(\beta, \lambda, \gamma, \epsilon, S(0), V(0))$, the value of each parameter estimate is determined by $(\frac{1-\alpha}{\alpha})(\alpha p_{i-1})(\alpha^2 p_{i-2})...(\alpha^k p_{i-k})$ where p_i is the parameter estimate from year i.

3. Pennsylvania Mumps Data

Mumps is a contagious viral disease, usually spread through airborne transmission, causing symptoms including fever, headache, fatigue, loss of appetite, and swelling of the salivary glands (CDC, 2019). While morbidity from mumps is low (2 out of 10,000 cases in the United States in the pre-vaccine era), mumps has been known to cause orchitis (testicular inflammation), encephalitis and even permanent hearing loss in a small percentage of cases (CDC, 2019). Recent interest has been in investigating the occasional outbreak of mumps since the second dose of MMR was mandated and whether there is waning immunity over time to the disease (Lewnard & Grad, 2018), (Porter & Oleson, 2013).

Recorded cases of mumps statewide in Pennsylvania between January 1970 and December 2001 were obtained courtesy of Project Tycho (Van Panhuis, Cross & Burke, 2018) and are shown in Figure 1. The dataset consists of weekly statewide case totals during this time period. Weeks with no data over this timeframe were treated as having no confirmed cases. We selected January 1, 1970 - September 30, 1990 as the

training portion of the dataset and October 1, 1990 - December 31, 2001 as the testing portion.

When fitting point process models to the data, the estimated onset time for each infection within a given week period is randomly drawn from a uniform distribution covering the 7 day time interval as in Park et al. (2019) and Schoenberg et al. (2019). For each point process model, 3 iterations of uniformly drawn event times were used, for each iteration, parameters were estimated by MLE, and of the three sets of parameters, those yielding the smallest RMSE over the training set were selected.

In 1967, the still currently used mumps vaccine was introduced in the United States. In the U.S., the mumps vaccine is now available in combination with the measles and rubella vaccines, known as MMR (CDC, 2019). In 1977, one dose of the MMR vaccine was recommended for all children 12 months and older. In 1989, the vaccine policy was updated to include a second dose at 4-6 years of age after it was discovered that waning immunity to mumps was a real possibility (CDC, 2019). As a result, the incidence of mumps in the U.S. has dropped from 55.5 cases per 100,000 people to less than 2 cases per 100,000 in 2017 (Elflein, 2019), and this decrease is also evident from Figure 1.

Despite the fact that the MMR vaccine has appeared to greatly diminish MMR incidence, occasional outbreaks still occur. In 2006, there was an outbreak which affected many midwestern university students, most of whom lived in dormatories (CDC, 2019). Another occurred in 2009 when 3502 cases of mumps were reported in the Orthodox Jewish communities in New York City, an area where 90% of the children at the time received one or more doses of MMR (CDC, 2019). These mini-resurgences of mumps have led researchers to question whether those vaccinated are still prone to losing immunity to the disease over their lifetimes (CDC, 2019).

4. Results

4.1 Model Fitting

4.1.1 SVEILR Model

Table 1 lists the parameter estimates for the SVEILR model for each of the twenty years fit individually during the course of the training period. The decrease over time in the annual estimated incidence of infection is obvious. This may be due to corresponding increases in the vaccination rate which rose from 58.4% in 1970 to a high of 67.6% in 1982 (CDC, 2011). However, data from 1986 to 1990 was not recorded and the survey methods to collect such data changed after 1993 resulting in dramatically higher vaccination rates being reported (CDC, 2011). However, some of the reduction in reported cases is also reflected by a drop in the estimates of the proportion of infections that are severe, γ. This is especially apparent in the precipitous drop in $\hat{\gamma}$ from 1970 to 1971.

In addition, Figure 1 shows the fit of the SVEILR model to the Pennsylvania mumps training data on the left side of the dotted line. The SVEILR model appears to adequately account for both the seasonality of mumps infections as well as the decreasing overall trend during the twenty year period. In addition, the model is largely identifying the peak times correctly. The RMSE for the SVEILR model over the training period is 15.58 cases per week.

4.1.2 Point Process Models

For the simple Hawkes model given by (8) and with triggering function (9), the following parameter estimates are obtained using MLE: $\hat{\mu} = 0.089$, $\hat{\kappa} = 0.951$ points per observed event, $\hat{c} = 1.82$ events per day, and $\hat{p} = 2.50$. For the recursive model with triggering function (10) and with recursive component (11), the estimates are $\hat{\mu} = 0.1279$, $\hat{c} = 0.8623$ points per observed event, $\hat{\beta} = 0.7027$ events per day, and $\hat{p} = -0.04486$.

Fitted estimates of the triggering function, g, for the parametric and non-parametric Hawkes and recursive models are shown in the left panel of Figure 2. The right panel shows the corresponding estimated productivity functions, $\hat{H}(\lambda)$, for the four models. There appear to be rather substantial variations in productivity as the estimated conditional rate

varies.

Figure 3 shows the weekly numbers of observed cases on mumps in Pennsylvania along with the fitted rates from the fitted parametric and non-parametric Hawkes and recursive models. All four of the point process models appear to fit quite closely on the training data, despite underestimating the largest peaks, especially in 1970-1972.

As shown in Table 2, the non-parametric version of the recursive model achieves the lowest RMSE of five models considered. All four of the point process models fit the training data substantially better than the SVEILR model. The Stoyan-Grabarnik statistics, for the parametric Hawkes model, the nonparametric Hawkes model, the parametric recursive model, and the nonparametric recursive model, respectively, were 0.9992, 0.9988, 1.0035 and 1.0003, indicating no noticeable lack of fit for any of the four point process models.

4.1.3 Residual Analysis

Figure 4 shows the superthinned residuals over time for the non-parametric recursive model. The number of superthinned residuals from 1970 to 1975 is slightly lower than for subsequent years, indicating overestimation of the rate during this early period. Figure 4 also shows a histogram and lag plot of the standardized interevent times for the superthinned residuals from the non-parametric recursive model. There is a noticeably higher frequency than expected of the longest interevent times for the non-parametric recursive model. However, there is no noticeable clustering of the interevent times.

4.2 Out-of-Sample Forecasting

4.2.1 Forecasting using the SVEILR Model

Figure 1 shows forecasts using the SVEILR model with parameters exponentially smoothed using $\alpha = 0.05$ or with no smoothing ($\alpha = 0$). The weekly reported case counts during the testing period resemble the last few years in the training period more than the earlier years. To note, the RMSE is lower using $\alpha = 0$ than when using $\alpha = 0.05$.

4.2.2 Forecasting using point process models

The parametric Hawkes, parametric recursive and non-parametric Hawkes models forecast the number of cases accurately from 1990-1994, but substantially overestimate the cumulative number of infections during later years. On the other hand, the forecast using the non-parametric recursive model slightly underestimates the case counts during the first several years of the testing data but is relatively accurate from 1996-2001. Overall, as shown in Table 2, the non-parametric recursive model has the smallest RMSE of 1.905 cases/week, followed by the non-parametric Hawkes model with an RMSE of 2.253 cases/week. The SVEILR model, whether with $\alpha = 0$ or $\alpha = 0.05$, had substantially higher RMS errors.

The non-parametric recursive model has both the lowest RMSE within the training portion of the dataset as well as improved accuracy in forecasting weekly cases of mumps in the testing dataset. The next best in fitting was the parametric recursive and the runner up for forecasting was the non-parametric Hawkes model. The RMS error for all four point process models was smaller than that of the SVEILR model for both fitting and forecasting.

5. Discussion

The nonparametric recursive model appears to outperform the alternative models for forecasting weekly mumps cases in Pennsylvania and is also substantially less prone to overfitting as evidenced by the improved performance during the testing period relative to competing point process and SVEILR models.

An important item for further research would be to explore how all of these models might be improved by estimating a nonstationary background rate, incorporating the decreasing trend in cases as well as their seasonality, while simultaneously estimating the triggering portions of the models as in Zhuang and Mateu (2019), for example. Additionally, future work should focus on ideal selection of bin widths and bin numbers for fitting the nonparametric forms of the recursive and Hawkes models, especially for the estimation of the productivity

function, H. Also important is to find optimal ways to smooth the simple binned estimates considered here, perhaps via kernel smoothing as in e.g. Mohler (2014).

The methods explored here should be applied to other diseases as well, including emerging epidemics. It is unclear as to whether the techniques of fitting and forecasting would remain the same if the incidence of a disease is static or even increases over time.

Another important potential use of the non-parametric recursive model is to forecast the prevalence of various diseases based on the rate of vaccine coverage. This is particularly relevant in light of recent anti-vaccination trends in many developed nations, some of which may be attributable to misguided fears about unknown side-effects from vaccines (Dubé et al., 2015). According to Dubé et al. (2015), 5% of parents in the United States refuse for their children to be vaccinated and up to one-third of children in the U.S. lack the recommended protection from easily avoidable diseases.

6. Acknowledgements

This material is based upon work supported by the National Science Foundation under grant number DMS 1513657. Computations were performed in R (https://www.r-project.org). Thanks to the US CDC for supplying the data and to Project Tycho for making it so easily available.

7. References

Althaus, C.L. (2014). Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa. In: *PLOS Current Outbreaks* 6, 1-11.

Balderama, E., Schoenberg, F.P., Murray, E., and Rundel, P.W. (2012). Application of branching point process models to the study of invasive red banana plants in Costa Rica. *JASA* 107(498), 467-476.

Brauer, F. (2017). Mathematical epidemiology: past, present and future, *Infectious Disease Modeling*.

Brémaud, P. *Point Processes and Queues: Martingale Dynamics*. SpringerVerlag,

New York, 1981.

Centers for Disease Control and Prevention, *Vaccine Coverage Levels – United States, 1962-2009*, The Pink Book, 11th Edition, February, 2011.
Centers for Disease Control and Prevention, *Epidemiology and Prevention of Vaccine-Preventable Diseases*, The Pink Book, 13th Edition, April, 2019.

Clements, R.A., Schoenberg, F.P., and Veen, A. (2012). Evaluation of space-time point process models using super-thinning. *Environmetrics*, 23(7), 606-616.

Dubé, E., Vivion, M., MacDonald, N. (2015). Vaccine hesitancy, vaccine refusal and the anti-vaccine movement: influence, impact and implications. *Expert Rev. Vaccines*, 14(1), 99–117.

Elflein, J. (2019). New cases of mumps per 100,000 population in the U.S. from 1970 to 2017. *Statista*.

Gordon, J. (2017). nphawkes R package, Version 0.1.

Gordon, J.S., Clements, R.A., Schoenberg, F.P., and Schorlemmer, D. (2015). Voronoi residuals and other residual analyses applied to CSEP earthquake forecasts. *Spatial Statistics*, 14b, 133-150.

Hawkes, A. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*. 58(1), 83-90.

Huppert, A. and Katriel, G. (2013). Mathematical modelling and prediction in infectious disease epidemiology. *Clinical Microbiology and Infection*. 19(11), 999-1005.

Kelly, D., Park, J., Harrigan, R., Hoff, N., Lee, S., S.R. Wannier, Selo, B., Massoko, M., Njoloko, B., Okitolonda-Wemakoy, E., Mbala-Kingebeni, P., Rutherford, G., Smith, T., Ahuka-Mundeke, S., Muyembe-Tamfum, J.J., Rimoin, A.W. and Schoenberg, F.P. (2019). Real-time predictions of the 2018-2019 Ebola virus disease outbreak in the Democratic Republic of Congo using Hawkes point process models. *Epidemics* 28, 100354.

Kermack, W.O. and McKendrick, A.G. (1927). A contribution to the

mathematical theory of epidemics. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. 115, 700-721.

Kirchner, M. and Bercher, A. (2018). A nonparametric estimation procedure for the Hawkes process comparison with maximum likelihood estimation. *Journal of Statistical Computation and Simulation*, 88:1106–1116.

Lewis, P.A.W and Shedler, G.S. (1979). Simulation of nonhomogeneous poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3), 403-413.

Lewnard, J.A. and Grad, Y.H. (2018). Vaccine waning and mumps re-emergence in the United States. *Sci Transl Med.* 10(433).

Li, Y., Liu, X., Wang, L. (2018). Modelling the transmission dynamics and control of mumps in mainland China. *Int J Environ Res Public Health*. 15(1): 33.

Marsan, D. and Lengliné, O. (2008). Extending Earthquakes' Reach Through Cascading. *Science*. 319(1076).

Meyer, S., Elias, J., and Hohle, M., "A Space-Time Conditional Intensity Model for Invasive Meningococcal Disease Occurrence", Biometrics, June 2012.

Meyer, S., Held L. (2014), Power-law models for infectious disease spread. AoAS, Vol. 8, No.3, p.1612-1639.

Meyer, S., Held, L., Hohle, M., "twinstim: An endemic-epidemic modeling framework for spatio-temporal point patterns", Journal of Statistical Software, Section 3, 2017.

Mohler, G.O. (2014). Marked point process hotspots maps for homicide and gun crime prediction in Chicago. *International Journal of Forecasting* 30, 491.

Mohler, G.O., Short, M.B., Brantingham, P.J., Schoenberg, F.P., and Tita, G.E. (2011). Self-exciting point process modeling of crime. *JASA*, 106(493), 100-108.

Park J, Chaffee A., Harrigan R., and Schoenberg F.P. A non-parametric Hawkes model of the spread of Ebola in West Africa. *J Appl. Stat.*, submitted Sep19.

Ogata, Y. (1981). On Lewis' simulation method for point processes.

IEEE Transactions on Information Theory, 27(1).

Ogata, Y. (1998). Space-time point process models for earthquake occurrences. *Ann. Inst. Statist. Math.* 50(2), 379-402.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association.* 83(401), 9-27.

Ogata, Y. (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics.* 30(2), pp 243–261.

Park J., Chaffee A., Harrigan R., and Schoenberg F.P. A non-parametric Hawkes model of the spread of Ebola in West Africa. J. Appl. Stat., submitted Sep19.

Porter, A.T. and Oleson, J.J. (2013). A path-specific SEIR model for use with general latent and infectious time distributions. *Biometrics.* 69(1):101-8.

Schoenberg, F.P. (2013). Facilitated estimation of ETAS. *Bulletin of the Seismological Society of America*, 103(1), 601-605.

Schoenberg, F.P., Hoffmann, M., and Harrigan, R. (2019). A recursive point process model for infectious diseases. AISM 71(5), 1271-1287.

Schoenberg, F.P. (2003). Multi-dimensional residual analysis of point process models for earthquake occurrences. *Journal of the American Statistical Association*, 98:789–795.

Shmueli, G., Lichtendahl, K., "Practical Time Series Forecasting with R", Second Edition, 2016.

Stoyan, D. and Grabarnik, P. Second order characteristics for stochastic strutures connected with Gibbs point processes. *Mathematische Nachrichten*, 151:95–100, 1991.

The World Bank, "Life expectancy at birth, total (years) - United States", 2019.

Van Panhuis W., Cross A. and Burke D., Counts of Mumps reported in UNITED

STATES OF AMERICA: 1923-2017 (version 2.0, April 1, 2018): Project Tycho data release, DOI: 10.25337/T7/ptycho.v2.0/US.36989005

Veen, A. and Schoenberg, F.P. (2008). Estimation of space-time branching process models in seismology using an EM-type algorithm. *JASA*, 103(482), 614-624.

Wang, J. (2019). Point Process Models for the spread of Coccidioidomycosis, an infectious disease, in California. Masters Thesis, Department of Statistics, University of California, Los Angeles.

Wheatley, S., Filimonov, V., and Sornette, D. (2014). Estimation of the Hawkes process with renewal immigration using the EM algorithm. *Swiss Finance Institute Research Paper*, 14–53.

Yang, A.S. (2019). Modeling the Transmission Dynamics of Pertussis Using Recursive Point Process and SEIR model. Masters Thesis, Department of Statistics, University of California, Los Angeles.

Yang, Y., Etesami, J., and Kiyavash, N. (2017). Online learning for multivariate Hawkes processes. Advances in Neural Information Processing Systems, 4938–4947.

Zhuang, J. and Mateu, J. (2019). A semiparametric spatiotemporal Hawkes-type point process model with periodic background for crime data. *Journal of the Royal Statistical Society Series A*, Royal Statistical Society, 182(3), 919-942, June.

8. Appendix

Academic Year	β	λ	γ	ϵ	$S(0)$	$V(0)$	I/100000
Jan 1970 - Sep 1970	0.55211	0.01426	0.16893	0.00100	0.08833	0.61189	68
Oct 1970 - Sep 1971	0.34540	0.01142	0.19802	0.00100	0.07234	0.75687	51
Oct 1971 - Sep 1972	0.56523	0.00521	0.05487	0.00100	0.02342	0.82246	40
Oct 1972 - Sep 1973	0.48028	0.00615	0.08134	0.00100	0.02063	0.87002	49
Oct 1973 - Sep 1974	0.38864	0.00915	0.10813	0.00100	0.07885	0.75271	41
Oct 1974 - Sep 1975	0.32220	0.00858	0.04812	0.00100	0.04734	0.72389	36
Oct 1975 - Sep 1976	0.87266	0.00618	0.07130	0.00100	0.00391	0.54681	35
Oct 1976 - Sep 1977	0.58590	0.00502	0.01516	0.00100	0.06679	0.82386	12

Academic Year	β	λ	γ	ϵ	$S(0)$	$V(0)$	I/100000
Oct 1977 - Sep 1978	0.52773	0.00452	0.03159	0.00621	0.00001	0.78110	33
Oct 1978 - Sep 1979	0.33159	0.00704	0.02371	0.00100	0.07724	0.86552	16
Oct 1979 - Sep 1980	0.41620	0.01561	0.05648	0.00943	0.00001	0.93898	17
Oct 1980 - Sep 1981	0.42398	0.02703	0.05925	0.03364	0.00001	0.98814	63
Oct 1981 - Sep 1982	0.48071	0.02221	0.05562	0.04158	0.00001	0.95661	13
Oct 1982 - Sep 1983	0.38916	0.02290	0.01558	0.02730	0.00001	0.92277	5
Oct 1983 - Sep 1984	0.31616	0.00602	0.00599	0.00100	0.20079	0.74775	4
Oct 1984 - Sep 1985	0.38391	0.01970	0.05065	0.03394	0.08606	0.89321	8
Oct 1985 - Sep 1986	0.69370	0.06991	0.13761	0.11711	0.88178	0.11712	5
Oct 1986 - Sep 1987	0.25167	0.02881	0.02315	0.02672	0.00001	0.95429	4
Oct 1987 - Sep 1988	0.48329	0.00521	0.01040	0.00100	0.06763	0.83424	8
Oct 1988 - Sep 1989	0.12743	0.02214	0.09552	0.00790	0.15226	0.82994	8
Oct 1989 - Sep 1990	0.79982	0.00534	0.01344	0.00100	0.00001	0.58238	8

Table 1: Annual estimates of the transmission rate (β), waning immunity rate (λ), vaccine coverage of those who are susceptible or exposed (ϵ), proportion of people with severe infections (γ), initial proportions susceptible and vaccinated ($S(0), V(0)$) and rate of severe infection per 100,000 people at the end of each year ($\frac{I}{100000}$).

Model	Training RMSE	Stoyan-Grabarnik	Forecasting RMSE	Standard Error
SVEILR ($\alpha = 0.05$)	15.580	—	4.136	0.095
SVEILR ($\alpha = 0$)	15.580	—	3.059	0.087
Param. Hawkes	5.901	0.9992	2.380	0.090
Param. Recursive	5.058	0.9988	2.502	0.089
N.P. Hawkes	4.744	1.0035	2.253	0.089
N.P. Recursive	4.692	1.0003	1.905	0.078

Table 2: RMSE for each model using the training data and the testing data. The standard error reported is for the forecasting RMSE.

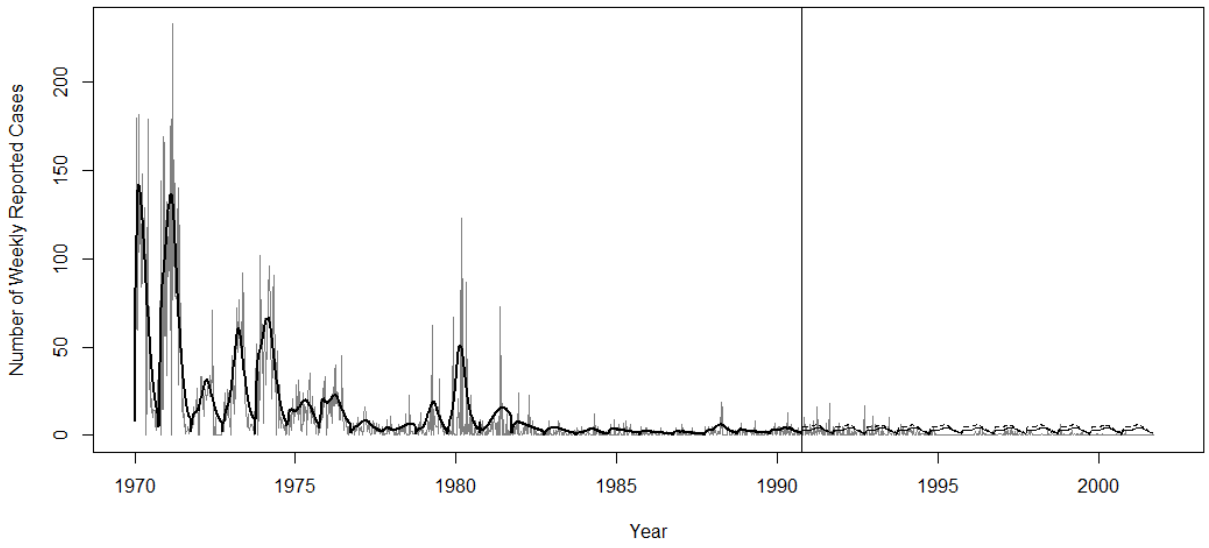


Figure 1: Reported weekly total cases of Pennsylvania mumps (gray) with fitted SVEILR model estimates for the training period (solid black) and SVEILR forecasts for the testing period using exponential smoothing with $\alpha = 0.05$ (dash line) or $\alpha = 0$ (solid line).

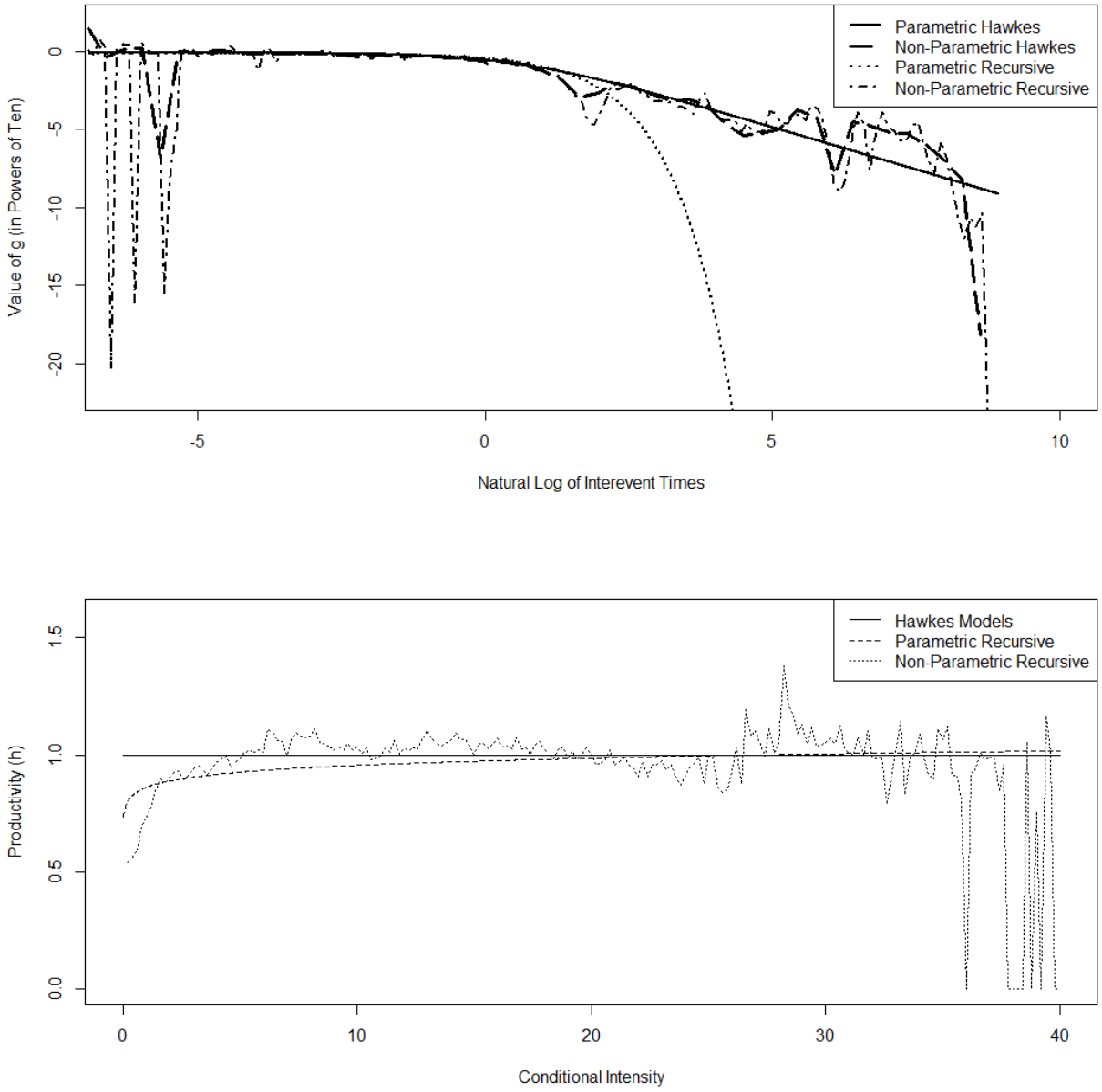


Figure 2: Estimated triggering density, g, for the parametric Hawkes, non-parametric Hawkes, parametric recursive and non-parametric recursive models (top). Estimated productivity function, $\hat{H}(\lambda)$, for the Hawkes, parametric recursive and non-parametric recursive models (bottom).

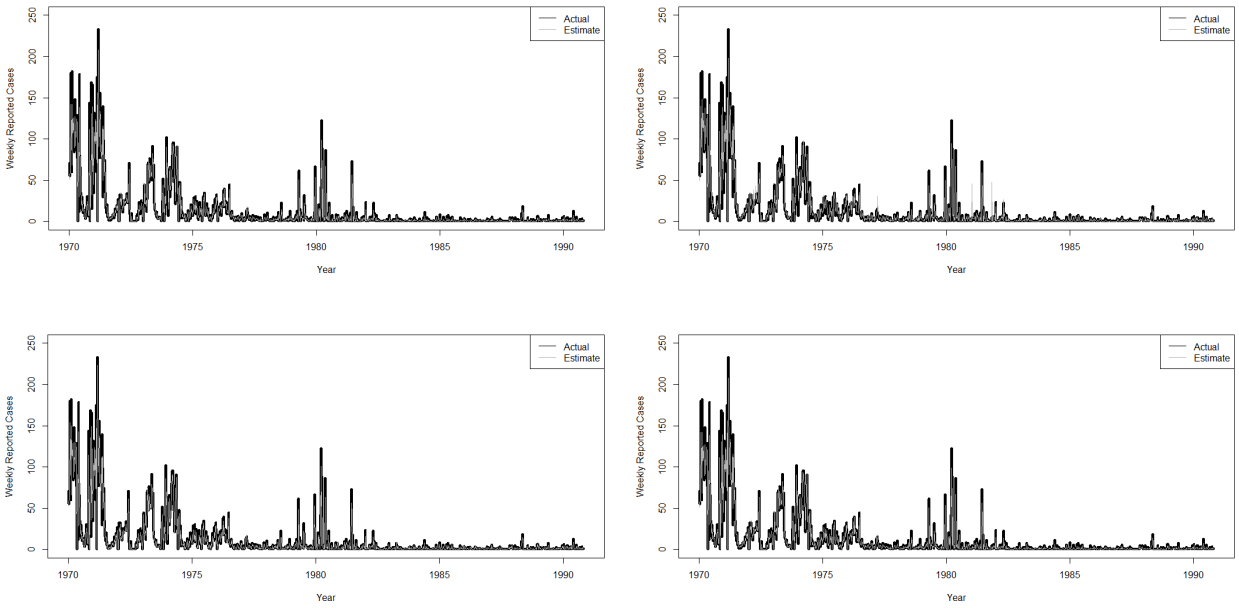
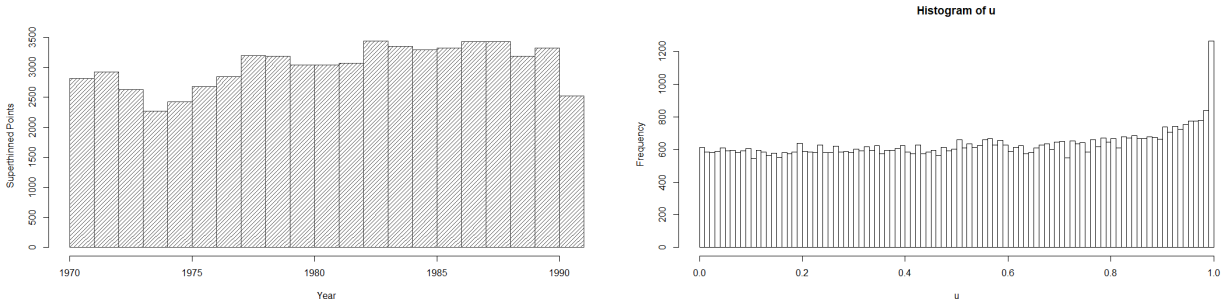


Figure 3: Estimated conditional rates and observed weekly number of cases of mumps in Pennsylvania over the 1970-1990 training period. The fits for the parametric Hawkes (top left), non-parametric Hawkes (top right), parametric recursive (bottom left) and non-parametric recursive (bottom right) models are shown along with the observed number of reported cases.



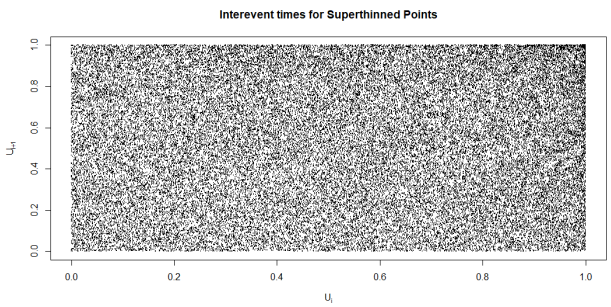


Figure 4: For the non-parametric recursive model: (a) Count of superthinned residuals over time using $b = 100$, (b) histogram of standardized times between consecutive events, $u_i = F^{-1}(\tau_i - \tau_{i-1})$, where F is the cumulative distribution function of the exponential with rate $b = 100$, and (c) lag plot of standardized times between consecutive events, for superthinned residuals.